Spectral Models for Air Transportation Networks

by

Max Zhaoyu Li

Bachelor of Science in Engineering, University of Pennsylvania (2018)
Master of Science in Engineering, University of Pennsylvania (2018)

Submitted to the Department of Aeronautics and Astronautics
in partial fulfillment of the requirements for the degree of

Doctor of Philosophy

at the

MASSACHUSETTS INSTITUTE OF TECHNOLOGY

September 2021

Author .....
Department of Aeronautics and Astronautics
September 22, 2021

Certified by .....
Hamsa Balakrishnan
William E. Leonhard (1940) Professor of Aeronautics and Astronautics
Thesis Supervisor

Accepted by .....
Jonathan How
Professor of Aeronautics and Astronautics
Chair, Graduate Program Committee

This doctoral thesis has been examined by a Committee comprising of the following members:

Professor Hamsa Balakrishnan .....
Thesis Committee Chair and Thesis Supervisor
William E. Leonhard (1940) Professor of Aeronautics and Astronautics

Professor Alexandre Jacquillat .....
Member, Thesis Committee
1942 Career Development Professor and Assistant Professor of Operations Research and Statistics

Professor Megan S. Ryerson .....
Member, Thesis Committee
UPS Chair of Transportation and Associate Professor of City and Regional Planning and Electrical and Systems Engineering
University of Pennsylvania

Professor Caroline Uhler .....
Member, Thesis Committee
Henry L. and Grace Doherty Associate Professor of Electrical Engineering and Computer Science

Spectral Models for Air Transportation Networks

by

Max Zhaoyu Li

Submitted to the Department of Aeronautics and Astronautics
on September 22, 2021, in partial fulfillment of the
requirements for the degree of
Doctor of Philosophy

Abstract

From natural disasters to power outages, these events, even geographically-localized ones, often result in widespread disruptions across the air transportation network. In order to engineer resilience and design better proactive mitigation strategies, it is important to identify, characterize, and control the effects of such disruptions. A more resilient and well-prepared air transportation system directly translates to mitigated delay costs and increased service quality.

Current delay performance metrics reflect only the magnitude of incurred flight delays at airports. In the first half of the thesis, we show that it is also important to consider the *spatial distribution* of delays across a network of airports. We analyze graph-supported signals, leveraging techniques from spectral graph theory and graph signal processing to compute analytical and simulation-driven bounds for identifying outliers in spatial distribution. We then apply these methods to analyze 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. 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, as well as compile an inventory of outlier days. This inventory could guide future aviation system planning efforts and research. We demonstrate the generalizability of this outlier identification and characterization framework through a comparative analysis of US and Chinese airport networks.

After establishing the framework of modeling and analyzing airport delays as graph-supported signals, in the second half of the thesis we focus on two applications enabled by this framework: Examining commonly-occurring disruption-recovery cycles in the US airport network, and proposing an approximate network control scheme. In regards to the first application, we study these disruption and recovery cycles through a state-space representation that captures the severity and spatial impact of airport delays. In particular, using US airport delay data from 2008-2017, we first identify representative disruption and recovery cycles. These representative cycles provide insights into the common operational patterns of disruptions and recoveries

in the system. We also relate these representative cycles to specific off-nominal events such as airport outages, and elucidate the differing disruption-recovery pathways for various off-nominal events. Finally, we explore temporal trends in terms of when and how the system tends to be disrupted, then subsequently recovers. For the second application, we consider the problem of designing control strategies for high-dimensional systems that lack a detailed model. To do so, we leverage the ability of *copulas* to represent dependent structures in high-dimensional data, and approximate the state space of airport delays through inverse sampling. We demonstrate the use of the control policies obtained from our methodology through a case study of controlling flight delays within the US air transportation network.

We conclude this thesis with some directions for future work, an example of which is a new hierarchical approach towards air traffic management procedures such as airport ground holding. We also comment briefly on the applicability of the methods developed in this thesis for other transportation and networked systems.

Thesis Supervisor: Hamsa Balakrishnan
Title: William E. Leonhard (1940) Professor of Aeronautics and Astronautics

Acknowledgments

People told me to expect the journey towards a doctoral thesis to be (1) fun and at times frustrating, (2) rewarding and at times challenging, (3) exhilarating and at times mind-numbing. No one told me that I will (4) finish this journey in the middle of a global pandemic, but hey, last time I checked, 3 out of 4 is a passing grade. In the midst of this global tragedy, a tiny silver lining: For me, it brings into sharp focus how truly fortunate and privileged I am to have *my own village* of family, friends, and mentors. Think of this writing as the village directory: A flailing, perhaps poorly-written attempt by me to convey my deepest gratitude for every person, named and unnamed.

First, to Prof. Hamsa Balakrishnan, who took a chance on me the summer of my senior year to work as a research intern in her group, and then again, as my doctoral advisor. She gave me the intellectual freedom to pursue an array of wild ideas, while mentoring and advising me with kindness, genuine care for my well-being, and a good dose of humor. If I could become half of what she was to me to my own future students, I would ask for nothing more.

This thesis is about air transportation, a field that I had always been fascinated by since I was a child. Only after Prof. Megan S. Ryerson's literally life-changing presentation about her *research* in air transportation, did I begin seeing myself as becoming a researcher in this field. She guided and polished my passion for aviation into myriad questions, models, results. All while, akin to Prof. Balakrishnan, guiding and caring about me as a person, too.

Labmates! It has been a blast sharing in group meeting pastries, lunch outings, sailing adventures, and ski outings with you all, Sandeep B., Chris C., Karthik G., Jackie B., Kailing N., Gerard D., Joao C., Lindsey B., Lauren C., Titi F., Matt K., Siddharth N., Simran P., Takuya K., and Baris B. – Y'all have made the days in Cambridge fly, and even when we switched to remote work, y'all made the group feel connected. I am lucky to have collaborated, and continue to collaborate, with some of y'all, and even more so to call all of y'all dear friends and colleagues. And to the rest of the ICAT group and CEE friends, we better go on bubble tea runs at Abide again! Kevin W., Bazyli S., Tim L., Alex P., Akash R., Pedro T., and Nick L., the conversations and the many times y'all had to let me into your office space were always bright spots in my day, and Shraddha R., one day we'll complete that Panda Express plus Gong Cha run. Perhaps multiple times.

Along with Prof. Balakrishnan and Prof. Ryerson, it has been a pleasure having Prof. Alex Jacquillat and Prof. Caroline Uhler as thesis committee members. I am still ecstatic that Prof. Uhler agreed to join my committee after I approached her towards the end of her Graphical Models course; she is a wealth of knowledge and always brought fresh insights and questions. The same goes for Prof. Jacquillat, who has been a constant source of support and mentor since I cold-emailed you when you were still at CMU! Not only was that still one of the best brunch spots I've ever been to, but your dedication to optimization models is unmatched.

One of the biggest reasons I look forward to attending in-person events and con-

ferences again is to catch up with everyone who has been a mentor and friend to me in this cozy field of air transportation research. First off, I would like to thank Dr. Andrew Churchill and Dr. Lixia Song for serving as readers for this thesis. To other colleagues, mentors, role models, and friends: Lu Dai, Kristyn Pantoja (Penn buddy shout-out!), Dr. Daniel Suh, Dr. Xiaoxia Dong, Prof. Yanjun Wang, Prof. Amedeo Odoni, Prof. Lavanya Marla, Prof. Seth Young, Prof. Xiaoqian Sun, Prof. Dave Lovell, Prof. Amber Woodburn-McNair, Prof. Amy Kim, Prof. Mark Hansen, Prof. Man Liang, Prof. Ma Ji, Prof. Yash Pant, Prof. Rahul Mangharam, Dr. Jungwoo Cho, and Prof. Alex Estes, it has truly been too long, and I can't wait to see y'all again.

Once COVID hit, one shining beacon that I look forward to, like clockwork, every Thursday at 7 pm central, has been the good ol' EH4 Zoom calls with Yusr G., Arabella U., Helen F., Denice A., Judy B., Stephanie G., Abhishek R., Alex Z., Ivy Z., Annie S., Ally C., Ethan F., Jack K., Michael S., Kanika M., Krish M., Will T., among others. At the time of writing, we're about to hit our 68^{th} weekly call. Absolutely insane, in many ways. I will never be able to thank Penn's housing lottery enough for EH4. And to the #51Porter crew, special shout-out to Gloria H., I've been so lucky to have known you since literally subbie year, and I'm forever grateful we had a blast reconnecting in Cambridge. Niharika G. and David G., y'all know what y'all mean to me, and the literal hundred of hours of FaceTime never seem enough.

There are so many other people that I want to acknowledge: From something as simple as grabbing boba to literally letting me crash on your couch when I'm in town for work, each and every moment has, as Marie Kondo puts it, sparked joy. Especially for the people who I went to high school with, who stuck with me after a 5 (!) year high school, and many more years of undergrad and grad school after that. Y'all always were happy to hang when I was back in Champaign-Urbana, or when I happened to drop by wherever y'all are now: Ananth N., Roberto C., Diego G., Samah A., Tanisha A., Divya J., Chas N., Alice R., Surya L., Mary Evans D., Raebekkah P.-C., Thea K., Nathan B., and lots of other Uni kids. Truly, Uni High 'til I die?!

◇◇◇

Mom, Dad, and my lil' bro Michael.

My protectors. Heroes. Cheerleaders. Providers. Supporters.

My everything.

I love you.

Thank you.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

1.1 Background and context

The aviation industry, fueled by technological, economic, and societal advances, has irreversibly become a cornerstone of the domestic and international transportation system. It is now not only temporally feasible, but also economically viable to conduct far-flung domestic and international travel. The elimination of what otherwise would have been a daunting logistical constraint has no doubt played a critical role in enabling everything from leisure travel microcosms to an entire macrocosm of accelerating globalization [20]. Both of these paradigms – leisure versus business passenger segments – are crucial in fueling the continued growth of the aviation industry, along with larger, industry-specific economic and societal trends such as emerging low-cost carriers (LCCs) and ultra low-cost carriers (ULCCs), continued deregulation through bilateral open-sky agreements, and a rising middle class with the financial means to fly in developing nations [23]. The socio-technical and cyber-physical infrastructures that comprise the foundation of the aviation industry is a complex amalgamation of many disparate components, intricately connected by the common goal of enabling safe and efficient movement of people and goods through the skies. As such, the scope of this thesis will be limited to addressing inefficiencies within the *operational* component. Specifically, as a capacity-constrained system operating with relatively well-matched demands, there will be times during which flight delays and flight cancellations are

unavoidable. However, the mismanagement of demand-capacity imbalances can lead to *avoidable* inefficiencies within the system: In this thesis, we propose new models to identify, better characterize, and manage these inefficiencies.

As the demand for air travel continues to grow both nationally and internationally, this capacity-constrained system is burdened with increasing amounts of traffic. To provide some context regarding the current prevalence and utilization of the global aviation system, in 2018 this system served more than 4.3 billion passengers worldwide, representing a 6.4% increase from the previous year [121]. Even when the scope is restricted to US airlines, more than 2.4 million passengers *per day* were served by the aviation industry in 2018, along with approximately 58,000 tons of cargo. Still restricting ourselves to the scope of US airlines, the medium of transport for these peoples and goods, i.e., the aircraft themselves, performed more than 28,000 flights on a daily basis in order to satisfy the travel needs of people and cargo. All of these statistics discount the humans working at the forefront and behind-the-scenes to ensure the safe and efficient operation of the air transportation system: In the US alone, the aviation industry requires over 740,000 direct employees managing every aspect of this complex system [7].

Two important and intertwined factors that together drive such high rates of utilization are continued increases in connectivity along with continued decreases in airfares. Over 22,000 unique city pairs were serviced by airlines worldwide in 2018; more importantly, this represents an increase of 1,300 city pairs from 2017 [118], hereby providing a larger selection of personal, leisure, and business travel opportunities. However, merely increasing the choice set of destinations would not be enough in and of itself – the second complementary ingredient resulting in increasing air travel is that the cost for air transport, when adjusted for inflation, have dropped by more than 50% over the past 20 years [118]. Figure 1-1 displays the upward trend of unique city-pairs, as well as the downward trend of the real price of air transport from 1998 to 2018. What these two factors have induced, as hinted previously, is an upswell of demand, particularly from emerging markets. For example, origin-destination (OD) markets in China experienced 10% year-over-year growth rates, posting an additional

50 million passengers in 2018 alone. Moreover, the two largest OD air passenger market is China and the US, together accounting for approximately 30% of all world-wide passengers. Given the impact and scale with regards to the sheer number of passengers that operational factors – route networks, delays, cancellations, air traffic control and air traffic flow management – would have within the air transportation networks of the US and China, this is one central motivation for why many of our completed analyses and proposed future work focus on the air transportation system in these two countries [118].

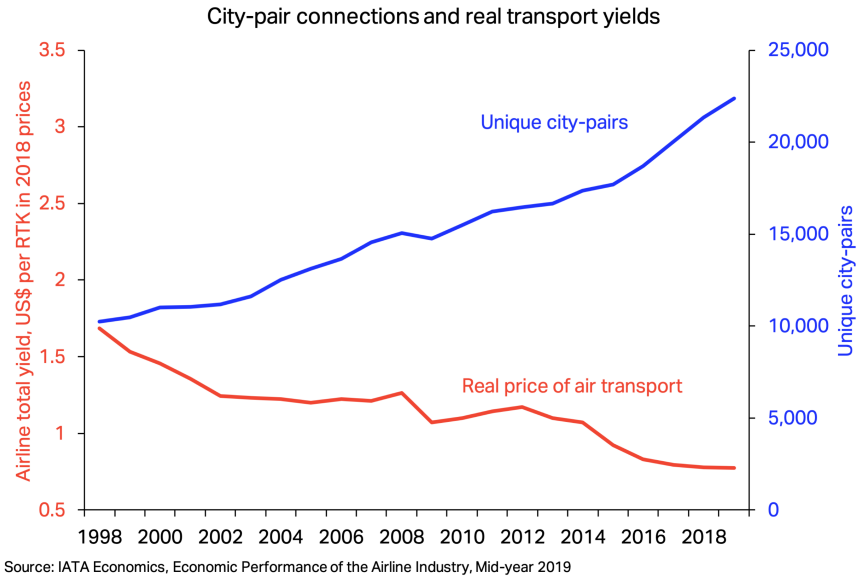


Figure 1-1: Number of unique city-pairs connected by the global aviation system, superimposed with the real price of air transportation, expressed in US dollars per revenue ton kilometers (US $ per RTK), adjusted to 2018 prices [118].

1.2 System inefficiencies: Flight delays and cancellations

Unsurprisingly, one consequence of increased system demand is the greater prevalence and impact of inefficiencies in the system. Such a statement holds true not just for the aviation system, but for other networked systems and infrastructures as well. In 2018, average system load factors (a percentage that measures how many seats

in an average aircraft is filled by passengers) reached 82%, up from 77% in 2009 [121]. A quick, back-of-the-envelope calculation would dictate that a delayed 100-seat airplane in 2009 resulted in 77 delayed passengers, whereas in 2018 an additional five passengers would have been impacted. While there may be an easily agreed-to definition for flight cancellations, i.e., a scheduled commercial flight ceases operations on a short-term, temporary basis, the definition of flight delays are more nuanced. For the purposes of these introductory remarks, we use measures of delay obtained from the US Department of Transportation (US DOT) Bureau of Transportation Statistics (BTS), wherein a particular flight is considered to be delayed if it arrived at or departed from its gate 15 minute or more after its scheduled arrival or departure time, respectively [35]. More specifically, the scheduled departure and arrival times for a flight is standardized and reflected within a Computerized Reservation System, or CRS [37].

Equipped now with standardized definitions of flight delays and cancellations, we motivate the need to characterize, model, and better manage inefficiencies in the system by first discussing their pervasiveness and prevalence. First, some positive signs: Despite the dramatic increases in US air travel, the overall on-time performance of the US aviation system has remained steady over the past decade. Specifically, in 2018, 78.9% of arriving flights and 79.7% of departing flights were on time, compared to averages of 79.0% and 80.0% for 2010 through 2018, respectively. However, these relatively stable on-time performance figures should be evaluated in the context of an 11.8% growth in the number of flight operations from 2010 to 2018, or from 4.84 million to 5.42 million flight operations. Thus, 23.3% more flight operations were delayed in 2018, compared to the 2010-2018 average. This comes as a result of steady on-time performance, combined with continued air travel growth. A parallel story is reflected in cancellation statistics, with a 1.8% cancellation rate in 2018 compared to an average of 1.7% for 2010 through 2018. Overall, more than 1 million arriving flights as well as 1 million departing flights were delayed in 2018 in the US alone.

As with any phenomenon exhibited by such a complex, networked infrastructure, the specific set of scenarios and circumstances that caused a particular flight to be de-

layed or canceled can vary widely. In order to standardize delay cause classifications, BTS aggregates delay causes into the following five categories [38]:

1. *Air carrier:* The cause of the flight inefficiencies was due to airline-specific operational issues, or "circumstances within the airline's control." This includes delays and cancellations caused by maintenance issues or crew delays and absences, as well as disruptions during the turnaround process (i.e., delays attributable to cleaning crews, baggage handling, refueling, etc.)

2. *Extreme weather:* The cause of the flight inefficiencies was due to ongoing or forecasted meteorological disruptions that are deemed severe (tornado, blizzard, hurricane, etc.) Note that this does *not* include more typical disruptions caused by inclement (but not extreme) weather.

3. *National Aviation System (NAS):* This is a "catch-all" condition where the cause of the flight inefficiencies was due to operational factors that could be infrastructure-related (e.g., taxiway congestion, runway construction, airport equipment outage) or inclement weather-related.

4. *Late-arriving aircraft:* The cause of flight inefficiencies was due to inefficiencies affecting the inbound flight sharing the same aircraft (i.e., "sharing the same tail", as the specific airframe of the aircraft is uniquely identified through its aircraft registration, or tail number). In this work, we will broadly refer to this as tail-propagated or propagated delay.

5. *Security:* The cause of the flight inefficiencies was due to transit safety-related aspects, such as airport evacuations, security breaches, issues with screening equipment, and excessive security checkpoint queues.

We plot the distribution of flight delay and cancellation causes in Figures 1-2 and 1-3; the former accounts for all US airports with reporting carriers, whereas the latter accounts for a 30-airport subset designated by the Federal Aviation Administration (FAA) as the "Core 30" airports. The Core 30 airports are the busiest airports in

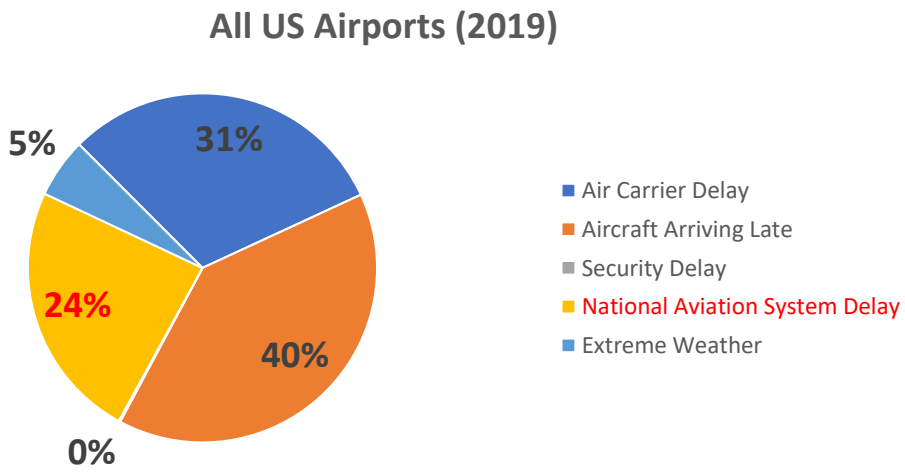


Figure 1-2: Reported cause of delay at all US airports in 2019 [36].

the US, typically measured in terms of passenger enplanements. We list the Core 30 airports, along with their designated International Civil Aviation Organization (ICAO) and International Air Transport Association (IATA) airport codes, in Table C.1 in the appendix. Henceforth, we will refer to any of these airports by their three-letter IATA code. Note that for both the all-encompassing airport group as well as the Core 30 airports, the dominant cause of delay is due to the late arrival of an inbound aircraft, at 40% and 38% of all causes, respectively. This illustrates the network effects at play in the system: Disruptions and inefficiencies do not stay isolated in one region of the system, but instead behave dynamically and propagate.

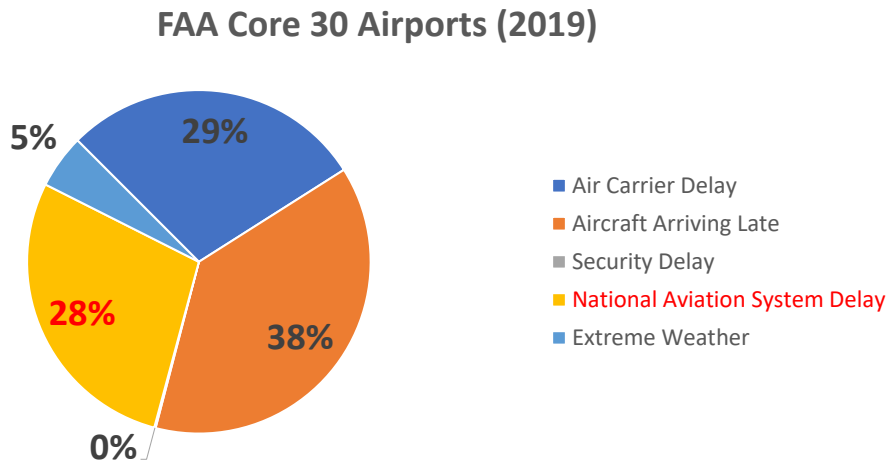


Figure 1-3: Reported cause of delay at the FAA Core 30 US airports in 2019 [36].

We emphasize that even though late-arriving aircraft is the predominant cause of delays, the origins of that propagated delay could have been due to any of the other factors, including NAS-type delays, which make up the second- and third-largest causes of delays for the Core 30 airports and all US airports, respectively. Furthermore, we note that while some delays and inefficiencies are unavoidable (e.g., the primary delays due to extreme weather), others could have been better managed and mitigated (e.g., reducing the impact of propagated delays through schedule adjustments). NAS-type delays are furthered separated into five categories by BTS, comprising of weather, volume (or congestion), airport equipment issues, closed runways, and other miscellaneous reasons that do not fall into any of the other NAS-type or larger-scoped delay categories. In Figure 1-4 we plot the breakdown of the NAS-type delays for both groups of airports. Here, we see that by far the largest cause of NAS-type delays is inclement weather resulting in the issuance of Traffic Management Initiatives (TMIs) such as Ground Delay Programs (GDPs), Airspace Flow Programs (AFPs), and aircraft separation-based flow control mechanisms such as Miles-in-Trail (MIT) and Minutes-in-Trail (MINIT) restrictions [78].

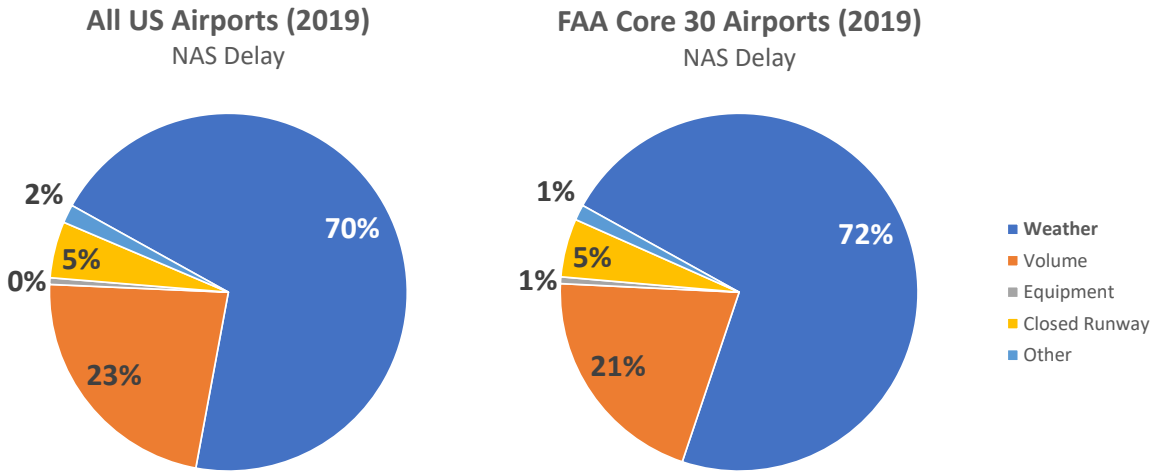


Figure 1-4: Reported cause of NAS-type delays at all US airports (*left*) and the FAA Core 30 airports (*right*) in 2019 [36].

1.3 Impact of delays and cancellations

We end this motivational examination of inefficiencies within the aviation system with a discussion on the quantifiable impacts of delays and cancellations with respect to two primary stakeholders: Airlines and the passengers. In 2007, a comprehensive report on total delay impacts conducted by NEXTOR and the Brattle Group estimated that the total direct cost of flight delays to airlines, to passengers, as well as accounting for lost demand, was $27.2 billion dollars, with a further $4 billion dollars-worth of negative impact on the US Gross Domestic Product (US GDP) [15]. Given that approximately 880,000 arriving and 1 million departing flights were delayed in 2007, this amounts to a direct cost of $14,000 per delayed flight, as well as a negative impact of $2,000 per delayed flight on the US GDP. On a per-passenger basis, one way to estimate the cost of delays is through estimating a given passenger's willingness-to-pay (WTP) for alleviating one hour of flight delays. This approach was taken by [149], where the authors estimated the WTP for leisure and business travelers in the context of 2012, with further segmentation based on the income level of an individual. Overall, as expected, the study found that business travelers were more time-sensitive, and thus had a higher WTP than leisure travelers when it comes to flight delays. In fact, one hour of flight delays was worth anywhere from $186 to $559 for business travelers, whereas an equivalent hour of flight delays was worth between $107 to $340 for leisure travelers [149]. These quantifiable and nontrivial costs attributable to flight delays further motivate the need to better understand and model the behavior of delays within air transportation networks.

In terms of environmental impact, the additional fuel usage and resultant emissions that can accompany flight delays (e.g., from aircraft waiting in taxi or runway queues with engines on [194, 76], or airborne holding delays resulting in additional distances flown [32]) add to the overall environmental impacts of the airline industry [41]. In 2018, intra-North America flights accounted for approximately 136 million metric tons of carbon dioxide (CO_2) emissions, with the US ranking first in terms of CO_2 emissions by departing flights [104]. Globally, the CO_2 emissions from air

passenger transport amount to 24% of total CO_2 emissions [104]; this does not begin to factor in other pollutants, such as sulfur dioxide and particular matters (i.e., PM10 and PM2.5) [77]. Another strong motivation for better understanding aviation delays and inefficiencies stems directly from the fact that, due to operational and predictability reasons, short-haul flights tend to be the most susceptible to traffic management actions [129, 175, 278]. Moreover, with short- to medium-haul routes (i.e., routes less than 1,500 kilometers) contributing to approximately one-third of total passenger CO_2 emissions [104], any traffic management improvements can have an out-sized impact on reducing additional environmental impacts due to operational inefficiencies.

Finally, we have thus far ignored the cost of cancellations, even though they constitute an important control action on the part of airlines in order to drive down excessive delays and inject buffers into schedules at airports with demand-capacity imbalances. Due to a wide variety of passenger compensation practices worldwide regarding airline cancellations, we only provide cancellation cost estimates for US carriers. The estimate of $6,000 direct cost to airlines per canceled flight in the US is a very broad average, primarily due to different flights having different values, as well as two broad cancellation compensation regimes: Controllable versus uncontrollable events resulting in the flight cancellation. The canonical practice for US carriers is to compensate for passenger accommodations such as hotels and meals only when flight cancellations occur due to events within an airline's control, examples of which fall under the "air carrier" category as specified in Section 1.2. Thus, for these airline-controllable events, the per-flight cancellation costs range from $2,700 for a flight operated by regional carriers to $42,900 for a long-haul flight operated on a wide-body aircraft. This is in comparison to per-flight cancellation costs of $1,000 to $13,000 for a canceled flight with similar characteristics, but in the case of an uncontrollable event such as extreme weather [180, 94].

The remainder of this chapter will first focus on network structures, attributes, and characteristics that appear in the air transportation system. In particular, we discuss the complications these factors bring into any analysis of the system, and

the accompanying explosion of data being generated and collected from the system. We then move to an introduction and motivation of the central research questions addressed in this thesis, and conclude with an outline of how the remainder of the thesis is organized.

1.4 Networks and data: A double-edged sword

Along with the set of complexities that arise from network interactions, the aviation industry and its constituent subsystems also operate under increasingly rapid rates of innovation, particularly when it comes to the acquisition, storage, and processing of data. The former complicate nuanced and rigorous analyses of the system, but the latter provides the foundations for rich, data-driven models. In this section, we first provide an overview of six complicating factors found within the air transportation network (Section 1.4.1). Our contributions in this thesis revolve around models that incorporate various elements of these aforementioned factors. We then reflect on the so-called "Big Data" advancements in aviation (Section 1.4.2).

1.4.1 Network interactions and complexity in air transportation systems

Agent interactions and shared resources. Even if we were to consider a subset of the possible agents or agents-like factors within the aviation system, there already exists many multi-way interactions between agents. This is particularly noticeable when agents interact in conjunction with shared, possibly constrained resources. For example, gates at an airport can be considered as resources that are shared by agents (aircraft), heterogeneous (different gates belong to different airlines, some gates are owned by the airport; gates also have aircraft equipment constraints), and constrained (there is a minimum ground turnaround time for aircraft occupying a gate) [109]. A non-exhaustive list of other shared infrastructure resources include airspace sectors, navigational fixes, runways, taxiways, and aprons [80]. The aircraft itself also falls

within many of these categorizations, given the fact that one aircraft typically flies multiple trips per day – the average departures per aircraft day across all US airlines in 2019 is slightly under 4 departures, with certain LCCs like Southwest Airlines (WN) operating an average of more than 5 departures per aircraft day [182].

Exogenous inputs. We could also observe the aviation system as a dynamical system that is constantly impacted by exogenous inputs from a variety of disparate sources. Take, for example, a relatively common scenario of convective weather resulting in irregular operations (IROPs) within an air transportation network. The convective weather impact itself can be thought of as – and often modeled as – an exogenous input with some degree of randomness [107]. In response to the effects and aftershocks – typically an induced mismatch between demand and capacity – of this initial disruption, a centralized authority such as an air navigation service provider (ANSP) may initiate TMIs in an attempt to correct demand-capacity imbalances within the system. As we have hinted at earlier, even if such corrective actions occur at a relatively localized scale, due to the networked nature of the system, it has the potential to cascade and propagate throughout the system. Finally, there is another class of exogenous inputs that act on individual sub-networks embedded within the larger network: Airline-specific actions. These airline-specific actions could range from the cancellation of a few flights, to requesting a ground stop for all flight operations at a certain airport [128, 82]. All of these exogenous inputs mix with the already-complex natural evolution of the aviation system state to eventually bring the system back to some nominal condition, i.e., the recovery phase after an initial disruption phase.

Nodal heterogeneity. One complicating feature that exists within air transportation networks, particularly in the US National Airspace System (US NAS), is the heterogeneity at airports due to different airline operating strategies. This heterogeneity can already be seen superficially when looking at delay causes. For example, in 2018, late-arriving aircraft is responsible for 9.3% more delays at BOS than at

SFO, whereas NAS-type delays affect SFO at a rate 12.7% higher than BOS [36]. This already indicates that different nodes are impacted by propagated delay versus TMIs at different scales. One prominent example of this airport-based, or nodal (if airports are thought of as the nodes within the network) heterogeneity arises from the different hub-and-spoke architecture of US network legacy carriers (NLCs), specifically American Airlines (AA), Delta Air Lines (DL), and United Airlines (UA). Some airport nodes are *weaker* hubs than others, such as JFK and LGA for AA, BOS and SEA for DL, and LAX and IAD for UA – these weak hub nodes carried the least number of departing seats in 2017 [202]. In Chapter 5 we will show that these weak hub nodes often behave differently within an airline's sub-network when it comes to airport network delay dynamics [158] .

Another interesting node-based feature is *dual-* and *tri-hub* airports, where there is a hub presence for multiple major airlines. Examples of these include ORD for AA and UA, JFK for AA and DL, and LAX for all three US NLCs, along with being a focus city for WN. We already see involvement of this particular nodal heterogeneity in some of our analyses dealing with airline sub-networks, where disruptions in one sub-network often leave signatures in other airline sub-networks through these shared hubs.

Sub-network heterogeneity. An air transportation network can be thought of as a multi-layer network, with each layer comprising of a specific airline's sub-network. Thus, the system can be decomposed into interacting airline sub-networks that can be analyzed individually. These sub-networks can be quite different from each other, driven by the fact that airlines can choose to operate by focusing more on hub-and-spoke route structures, or incorporate more point-to-point service. Even though the trend in the US airline industry is towards further network consolidation and hub strengthening [227, 276], as well as the emergence of alliance-based hubs and intercontinental hub-to-hub service, there are still airlines in the US (e.g., WN and B6) that incorporate more point-to-point flying within their sub-networks [55]. Hence, it is crucial to study the behavior of these individual sub-networks as well.

Temporal and spatial variance. To throw more complications into the mix, along with the fact that so many aspects of the aviation system is intertwined (via interactions), heterogeneous, and constrained in some way, all of these factors have temporal and spatial variations as well. The most apparent of these would be short- to long-term temporal trends. For example, the arrival and departure demand at airports fluctuate on an hour-to-hour basis, the ebb and flow of which is exaggerated at the larger hub airports where airlines will typically operate banks of arrivals and departures in order to facilitate connecting passengers and the possibility of aircraft swaps in the event of irregular operations [260, 23]. In terms of spatial variation, the physical distance between airports has its own set of implications. In many of our correlation-based analyses, we see that, e.g., the historical delay time series at many East Coast airports in the US are typically highly positively correlated. Such correlations result from a combination of geographic proximity (e.g., the same weather system affecting BOS will probably affect LGA as well) as well as mutual traffic flows (e.g., AA and DL shuttle flights linking many East Coast hub airports). Furthermore, the terminal airspace surrounding collocated airports within the same city, or *metroplexes*, presents its own set of challenges in terms of airspace complexity, geometries, and shared arrival and departure resources [162, 191].

Variance of stakeholder goals. The final complicating factor within the aviation system that we will take into consideration in this thesis is the variety of stakeholders operating within the aviation system. The goals of one stakeholder might be orthogonal to, or even detrimental to the goals of another stakeholder. As we have foreshadowed in the discussion regarding the diversity of exogenous inputs that influence the aviation system on a regular basis, some of the basic stakeholders include the passengers themselves, the airlines, the airports, as well as ANSPs such as the FAA in the US. Of course, even within one of these stakeholder groups, there are further, "sub"-stakeholders (take, for example, the variety of airport-specific stakeholders [234, 223, 165, 81]), but many of our analyses will target comparing results between one stakeholder group at-large versus another. For instance, in this thesis

we will oftentimes give airline-specific results in addition to system-wide results, then compare and contrast between the two.

1.4.2 An abundance of data from air transportation systems

On the other hand, the air transportation system is becoming more data-driven overall. From the microscale-level (flight-specific, passenger-specific, etc.) to the mesoscale- (airport-specific, airline-specific, etc.) and macroscale-level (US National Airspace, transatlantic and transpacific networks, etc.), advancements in data collection, storage, and distribution capabilities dominate the current conversation in aviation. One prominent example is the recent System Wide Information Management (SWIM) initiative by the FAA [79]. A single day of air traffic operations within the US National Airspace System (NAS) generates (1) $O(10^4)$ reports pertaining to arrival and departure information, boundary crossing updates, and oceanic reports, planning positions, and actual positions; (2) $O(10^5)$ records of flight management information, flight plan-specific data, aircraft specifications, and flight route data; and (3) more than $O(10^6)$ observations relating to flight tracks, airways, centers, fixes, sectors, and waypoints [10]. In short, we argue that aviation, along with aviation research, is entering an era of *Big Data*. This current phase of development for aviation can be characterized by the applicability and usage of data-driven models to analyze, control, optimize, and predict aviation processes [163].

Figure 1-5 illustrates that even in the limited example of weather-related impacts on the NAS, the complexity of the system is evident. The forward layers in Figure 1-5 that describe various NAS airspace partitions are highly organized and structured, while the last layer containing raw meteorological data and weather radar returns is comparatively less structured, but with higher data dimensions and quantities. Taking the layers described in Figure 1-5 into account with other features within the aviation network (airports as nodes, air routes as edges, origin-destination delays, ground stops, Ground Delay Programs, Airspace Flow Programs, etc.), a highly complex and high-dimensional network begins to surface.

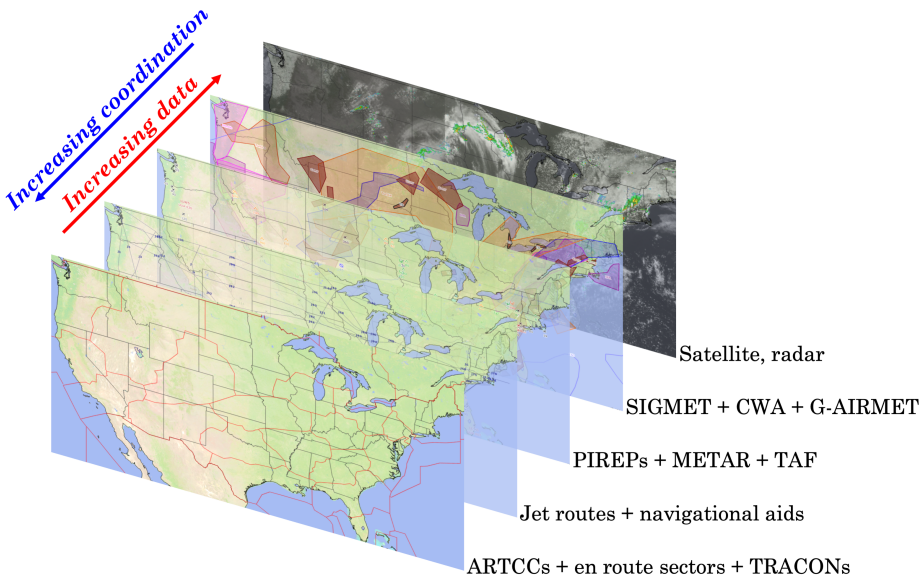


Figure 1-5: The architectural layers of the NAS with respect to convective weather; each layer provides its own set of data that are closely related to data from another layer. *Reprinted from [164].*

1.4.3 Data-driven aviation research

We recently conducted a large-scale literature survey of data-driven aviation research in [163]; two important metrics we used to explore the quality of data usage and diversity in the 200 reviewed aviation research articles are (1) measuring the distribution of articles that fall into each of 16 data categories, and (2) measuring the distribution of data categories per article. The former provides a sense of data diversity in the field of aviation research as a whole, whereas the latter details the amount of data diversity contained within each article. Figures 1-6 and 1-7 give the distribution of articles that fall into each of the 16 data categories, and data categories per article, respectively.

In terms of data diversity, we see that all 16 data categories are represented in our collection of 200 data-driven aviation research articles, with certain categories being more prevalent than others. We also see high levels of data diversity on an article-specific level – 76% of the articles (152 articles out of 200) were classified as using aviation data sourced from two or more data categories, with 45 articles out of the 152 articles (22.5% of the original 200 articles) using four or more data categories.

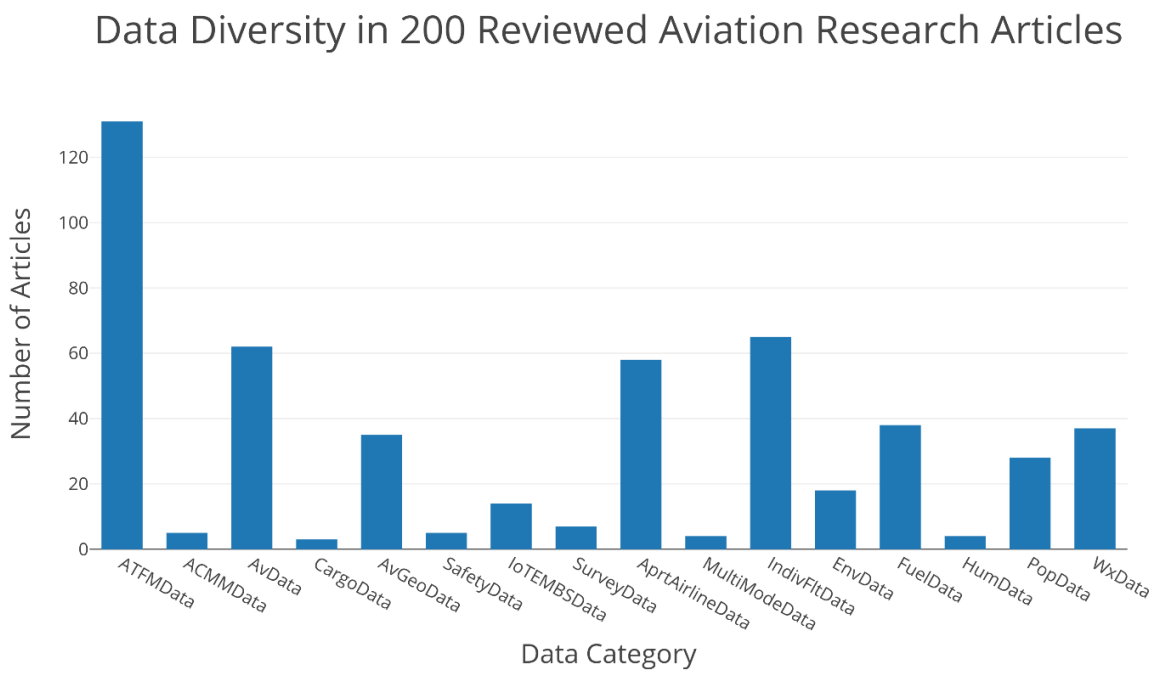


Figure 1-6: Data categories represented by the 200 reviewed aviation research articles. *Reprinted from [163].*

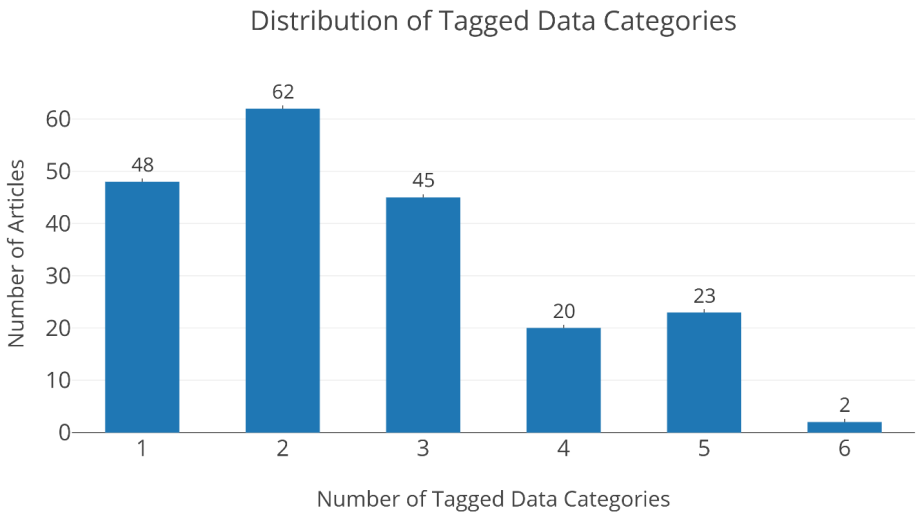


Figure 1-7: Distribution of data categories among all 200 reviewed aviation research articles. *Reprinted from [163].*

This indicates not only a proliferation of aviation data, but also a multidisciplinary approach to compiling and utilizing the data.

1.5 Motivating the central research problems

We have built up the aviation system as a prototypical example of a heavily-utilized piece of networked infrastructure, and one that is steadily becoming more connected and easier to access. We then introduced the variety of inefficiencies – with real consequences that are quantifiable financially and environmentally – within the system, as well as the complexities (interactions, heterogeneity, etc.) that complicate efforts to analyze and improve the system. However, we note that a parallel trend of increasing data acquisition, storage, and usage indicates that the field is ripe for advanced, data-driven models. In this section, we provide a high-level summary of the thesis contributions, organized by the different research problems that Chapter 4 through Chapter 7 will address.

1.5.1 Spatial distribution of airspace disruptions

Intuitively, the spatial distribution of airport delays describe where in the system the delay is located. This is distinctly different from observing the impact, or magnitude of the delays within the network. For example, the total delay of D_{tot} minutes within a system of N airports could be exactly the same, but distributed in completely different ways: One extreme would be equal amounts of delay (D_{tot}/N minutes) at all airports, whereas another extreme would be zero delay at all but one airport, where that one airport has D_{tot} minutes of delay. The metric of total delay, or total system impact, or even aggregated delay in a small subset of airports (e.g., northeast US airports) have been the focus of air traffic management and previous network models. As our example with N airports shows, this total delay metric is not sufficient in terms of differentiating between the two extreme scenarios. Furthermore, it is clear that traffic flow managers would handle the first extreme scenario very differently than the second extreme scenario. Of course, even though such idealistic extremes

may not occur in practice, scenarios that lie between the two extremes do. These scenarios, all with the same, or approximately the same, total delay, would not be identifiable through classical total impact-based metrics for delay.

We depict an illustrative example of the spatial variation, or spatial distribution of airport delays in Figure 1-8. Specifically, in this illustration, both (a) and (b) scenarios could have the same total delay, but in scenario (a), airport A1 is severely congested, whereas in scenario (b), all airports have equal levels of moderate congestion. In Chapter 4, we have focused significant efforts into the problem of *identifying whether or not* a given spatial distribution of airport delays is *unexpected*. To do so, we have defined new notions of signal outliers on graphs, as well as how to compute analytical and empirical bounds to evaluate outlier statuses [100]. With this new framework of outlier detection, which we formalized specifically for airport delays in [161], we have analyzed the airport delay networks of the Core 30 US airports [157, 158], including an airline-specific analysis [158]. These analyses have led to new insights into the behavior of delay – spatially and temporally – in the US NAS. We have also demonstrated the generalizability of this framework to other national airspace by performing a comparative analysis of the US and Chinese airspace in [161], as well as a deep dive into Chinese airline-specific sub-networks [160].

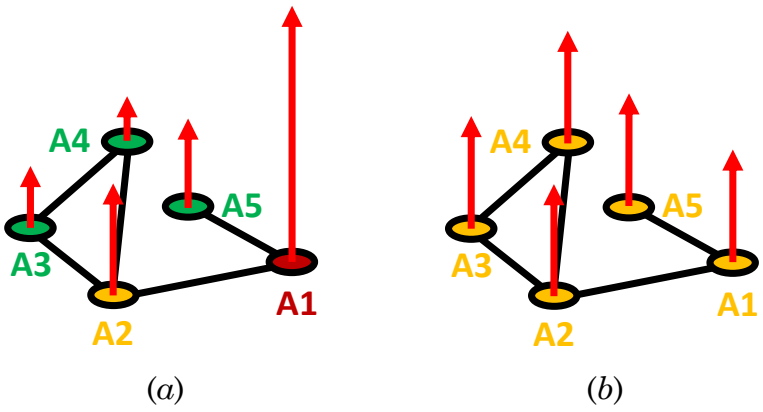


Figure 1-8: An illustration of the *spatial variance and distribution* of airport delays; note that the total delay in both (a) and (b) could be identical, but (a) has an extremely non-uniform delay distribution, with airport A1 incurring severe delays.

1.5.2 Airport delays and graph signal processing

Our primary focus in Chapter 4 is to develop the definitions and tools to be able to distinguish between, e.g., Figure 1-8(a) from its counterpart in Figure 1-8(b). The methods explicated in Chapter 4 would label the delay distribution in Figure 1-8(a) as *less smooth* compared to the distribution in Figure 1-8. Visually, we see that this is due to the large delay signal at airport A1, even though "closely related" airports A2 and A5 (by way of edges existing between A1 and A2 as well as A1 and A5) have comparatively smaller delay signals. However, the tools from Chapter 4 do not identify *which* airport, or groups of airports, whose delays resulted in a particular delay distribution being characterized as less smooth with respect to the underlying graph.

This question of identifying the specific airport, or groups of airports, whose delays contribute to an uneven (or less smooth) delay distribution motivates Chapter 5, which closely complements the spatial distribution outlier detection tools that are the focus of Chapter 4. In particular, we will use methods from *graph signal processing*, which are analogous to Fourier transforms, except that the domains of interest upon which signals are supported are non-Euclidean manifolds, e.g., undirected graphs. The so-called graph Fourier transforms, whose theory was developed in seminal papers such as [230, 231, 232], allow for the decomposition of graph-supported signals into linear combinations of the eigenvectors of graph shift operators such as the combinatorial graph Laplacian matrix. In terms of airport delays supported on nodes representing airports, with edges denoting some relation between airport delays (e.g., the historical correlations between pairs of airports), the eigenvectors can be viewed as "typically encountered" delay modes. For example, one such eigenvector ("$v_{\text{energetic}}$") could describe the circumstance where airports A1 and A2 have increased delays in comparison to A3, A4, and A5, whereas another eigenvector describes the circumstance where delays at all airports A1 through A5 are equal ("v_{constant}"). A decomposition of Figure 1-8(a) would see increased contributions from $v_{\text{energetic}}$, whereas a decomposition of Figure 1-8(b) would see increased contributions from v_{constant}.

We will formalize all of these notions that we have talked loosely about (e.g., graph Fourier transforms, eigenvector decomposition, energetic versus constant eigenvectors, graph frequencies, etc.) in Chapter 4 and Chapter 5. The focus of Chapter 5 in particular will be on applications to the US airspace, as well as airline-specific sub-network decomposition analyses. We also study the airport network of China, its delays, and compare them through these spectral lenses to the US airport network and airspace. Finally, we make explicit the intimate links between Chapter 4 and Chapter 5 in Section 5.6, where we tie together the outlier detection framework with airport delay eigenvector modes.

1.5.3 Low-dimensional models for airspace disruptions and recoveries

We now discuss and motivate two questions that we will broadly address in Chapter 6. The first relates to defining and formalizing periods of disruptions and recoveries comprehensively. A simple way to do so may be to consider the total delay as a measure of system disruptions, and then define any time interval in which delays exceed some threshold as a disrupted period. However, as emphasized in Chapter 4, this approach is unable to capture spatial information regarding the geographical extent of disruptions. Suppose only one strongly-connected airport – and no other airport in the system – is experiencing high delays. The total delay metric may not classify this as a disruption, even though this scenario is spatially unexpected, and may indicate an impending propagation of delays. Similar approaches that monitor temporal trends in delays at specific airports or origin-destination (OD) pairs are also unable to account for network connectivity-based information. Finally, in the context of extreme events such as hurricanes and nor'easters, the start of an event may not always coincide with the start of the system disruption. For example, airlines may proactively delay or cancel flights before the event, in which case the disruption precedes the event. On the other hand, airlines may opt to continue operations that progressively deteriorate, in which case the opposite order occurs. Hence, we

address this question by providing a comprehensive method (i.e., the disruption-recovery trajectory framework) in Chapter 6 to identify both disruptions and the subsequent recovery phases, based on the magnitude and spatial distribution of delays.

The second motivation relates to understanding broader trends and patterns in historical disruption-recovery cycles in order to improve system predictability and resilience. As discussed earlier, a significant challenge to analyzing past disruptions and recoveries is their inherently large variability. Thus, seemingly simple questions – what is the typical duration of disruptions due to specific off-nominal events; do two different events with similar delay impacts recover in different ways; is the recovery phase longer than the disruption phase; can we predict the onset and duration of the recovery phase; and so on – become very difficult to answer. Addressing these questions requires not only precise definitions for the onset and progression of disruptions and recoveries, but also the identification of "typical" or "representative" patterns that disruption-recovery cycles tend to follow.

Finally, the need to accurately define, quantify, and record a disruption within the air transportation system also plays an crucial role when it comes to predictive models for flight delays and disruptions. Since the training process for these predictive models is critical to their accuracy and performance, a natural question is how to generate the required training data. For instance, if a learning-based model seeks to predict the length of a disruption, it first requires historical observations of such a disruption. These are some of the central background motivating factors for the work we pursue in Chapter 6, wherein we formalize the notion of *disruption-recovery trajectories* (DRTs) for airline operations within a network of airports. These DRTs use an intuitive representation of the network state as a *trajectory* between discrete regions, capturing information regarding both the magnitude and spatial impact of air transportation system disruptions.

1.5.4 Towards hierarchical traffic flow management

Practical motivation: A new feedback in the air traffic flow management process

The current structure of the air traffic management system in the US can be considered to be essentially linear, starting with strategic-level airline-specific flight planning to real-time tactical decisions between the pilots and air traffic controllers. Replanning given feedback from the current state of the system occurs at various granularities and timescales, but can be particularly ad hoc at the flow planning level. During periods of low activity and/or nominal system performance, there may not be a need to adjust flow planning at a national level. On the other hand, during periods where capacity constraints throughout the system are constantly shifting (e.g., perhaps due to convective activities), it may be necessary to re-plan and adjust traffic management initiatives multiple times within a short time frame.

From a system synthesis perspective, the US National Airspace System (NAS) could be viewed as multiple feedback systems taking inputs from one level up, e.g., facility-specific traffic controllers from facility-specific traffic planners, who receive flow rates from individual traffic management units, and so on. Indeed, this model predictive approach has been used in previous research to model individual blocks of the air traffic flow management problem [253, 181, 45, 9]. Analogous to how the actual air traffic management system functions with multiple feedback at multiple granularities, these past models also work at different scopes, ranging from modeling individual agents to entire bundles of traffic flows. This is the context within which the work in Chapter 7 best fits: Proposing and describing a new way to model and influence system behavior. Specifically, the *hierarchical* approach proceeds as follows: Periodically observe the end product of air traffic management (i.e., the delays within the system, or another near-real time measure of system performance), decide whether or not there is a more desirable network performance goal (i.e., are delays becoming increasingly worse at particular airports, and where, if given idealistic circumstances, would you like to shift them), then leverage a model of the system plus an optimization

routine to construct a target trajectory for your system (i.e., this is how we want our system to perform).

Up to this point is where we depart slightly from the classical workflow of model predictive control: We do not directly suggest control inputs for the system, as this is unrealistic given the scope we are working with (i.e., you cannot just "suggest" for LGA to dissipate 300 minutes worth of delays) as well as the system we are working with (i.e., there is no direct nor easy translation between whatever control action is proposed and an air traffic flow management decision, although ongoing work in [49] links a *high-level planner* and a *low-level controller* together within this exact hierarchical control framework). Instead, we accept the target trajectory as a *performance upper bound* for an ideal system, suggest constraints on *how much* delays can accrue at an airport, and feed this into a traffic flow management problem (TFMP) [28] or a multi-airport ground holding problem (MAGHP) formulation [220, 8]. For the majority of Chapter 7, we do not focus on the problem of integration (i.e., the ongoing work in [49]), but instead on the problem of generating and characterizing the high-level model and system trajectory.

Theoretical motivation: An approximate network control strategy

Many systems in the transportation, energy, robotics, and communication network domains have state spaces that are extremely high-dimensional and contain intricate dependencies. In other words, the system state exists in an N-dimensional vector space $\mathbf{x} \in \mathbb{R}^{N \times N}$ with N being large, in addition to certain dependence structures between different components. Such large-scale networked systems, where N represents the number of nodes in the system, are naturally captured using high-dimensional state space representations, where each element of the state vector denotes the signal generated at a particular node.

Systems with high-dimensional state spaces have a few unique challenges associated with them. Firstly, it is difficult to develop accurate, high-fidelity models for such large-scale complex systems, with no well-understood physical mechanisms governing system interactions. Furthermore, these systems typically operate under

noisy conditions, with temporally-dependent dynamics. Secondly, owing to the lack of reliable models, predicting the future state of the system is challenging, and often results in the use of non-interpretable black box models. Lastly, it may not be practically feasible to control each element of the system. System operators may only be able to prescribe *high-level* requirements, or set system performance targets in some lower-dimensional projection of the system state.

Motivated by the practical problem of designing airport network delay redistribution strategies (e.g., the hierarchical framework we laid out in Section 1.5.4), we leverage an *approximate projection-based network control framework* to analyze and control a high-dimensional networked system such that performance targets in a low-dimensional projected space are satisfied. Note that in cases where the system dynamics are well-known and the performance requirements are of certain parametric forms (e.g., $\mathbf{x}^\top Q\mathbf{x}$ quadratic costs), the problem is analytically solvable and well-studied (e.g., through LQR and LQG controllers). However, our focus is on situations in which the lack of an accurate model is coupled with low-dimensional performance measures. The task of prescribing control actions becomes ambiguous when the dynamics and control costs are unclear. Furthermore, if the performance target is $h(\mathbf{x}) \in \mathcal{F}$, where $\mathcal{F}$ is some feasible set, then the optimization problem is significantly more challenging, and often analytically intractable. In Chapter 7, we investigate one way to design reasonable control strategies for a high-dimensional, heterogeneous system in the absence of a good model, given performance metrics that are defined on a lower-dimensional projection of the state space.

1.6 Outline of thesis structure

Chapter 2 and Chapter 3 bind the entire thesis together, collating the literature reviews for, and contributions of, every chapter that follows. Specifically, Chapter 2 contains non-exhaustive surveys of relevant past research for Chapter 4 through Chapter 7. With a review of pertinent previous literature in hand, we state the contributions of Chapter 4 through Chapter 7 in Chapter 3.

We begin in Chapter 4, where we examine graph signal outliers in the context of spatial airport delay distributions. Chapter 4 is closely complemented by Chapter 5, in which we address the successive question of identifying airports that contribute to certain spatial delay distributions using graph signal processing. With the framework of examining airport delays through the lenses of signals on graphs, we propose low-dimensional models for disruptions and recoveries within airport networks in Chapter 6, as well as approximate network delay control models in Chapter 7. We conclude in Chapter 8 by summarizing the work done in this thesis, and offering some directions for future work.

We note that all proofs to theorems, propositions, as well as associated corollaries and lemmas can be found in Appendix A. Supplementary figures and tables can be found in Appendix B and Appendix C, respectively.

Chapter 2

Literature Review

For each of the research areas introduced in Section 1.5, we survey a range of pertinent past research literature to better contextualize the work and contributions of this thesis. Each section of this chapter is dedicated to a specific research area chapter: Section 2.1 reviews literature related to spectral graph theory, signal processing on non-Euclidean manifolds, outlier detection methods, and airport network delay dynamics, all of which are relevant for Chapters 4 and 5. We then move to Section 2.2, wherein we summarize past work related to graph processes, identification of aviation disruptions, and predictive models for air transportation on-time performance and operations, e.g., predicting airport capacities and configurations; this literature review sets up the work in Chapter 6. To prepare for our work on approximate network delay control models in Chapter 7, in Section 2.3 we review past literature on air traffic flow management models at different scopes, standard airport ground holding models, and ways to sample representative airport delay distributions. Since this chapter serves as the centralized literature review portion of the thesis, the focus in Chapters 4 through 7 will be on our methods, implementation details, results, and discussions.

2.1 Graph signal processing, spectral analysis, and airport network dynamics

2.1.1 Spectral graph theory, graph signal processing, and wavelets

We provide a non-exhaustive overview of literature related to the field of discrete signal processing over graphs, or graph signal processing (GSP) for short. One foundational element leveraged by GSP is the connection between graphs, signals supported on graphs, and the matrices that characterize the graph (or more generally, so-called graph shift operators). The graph Laplacian, along with other related matrices such as the adjacency and the degree matrix, is an example of a matrix that represents an underlying graph. Specifically, the study of the relationship between the structure of a graph and the eigenvectors and eigenvalues of the graph Laplacian is known as *spectral graph theory* [51]. GSP is one application of a subset of ideas and notions stemming from spectral graph theory: A recent example of another practical application is the use of the Szemerédi regularity lemma for separating noise from important structural information in large graphs [88].

Central to our methods is the GSP framework [230, 244, 231, 232, 245], which extends the notion of a Fourier decomposition to a graph setting, and provides a toolkit to analyze signals on graphs. To demonstrate not only the applicability of the GSP framework but also the breadth of its applications, we provide a sampling of research literature whose methodology hinges on GSP and performing analyses within the graph Fourier domain. GSP has proved particularly useful in the analysis of high-dimensional medical data; GSP and graph filters have been used for analyzing human brain imaging data [115], and to better understand networked interactions in the brain when affected by neurodegenerative diseases [114]. Other settings include abstracting fMRI data of brain activation as a graph, and performing wavelet-based transforms on the fMRI graph in order to capture interacting patterns of brain activity [21]. A similar study in [116] uses GSP and fMRI brain signal data to characterize

task learning in humans. Finally, along the same lines of detecting outliers in airport delay distributions on networks, [258] analyzed atrial epicardial electrograms as graph-supported signals, and examined its temporal as well as spatial variations in the graph Fourier domain.

One of the earliest applications of GSP was for image processing and other computer vision tasks, with newer studies focusing on how GSP can shed light on machine learning and neural networks. Specifically, a regularization of the graph Laplacian was explored as a prior for image denoising tasks [206], and in general, GSP has been shown to be quite effective at image compression, restoration, filtering, as well as image segmentation [47]. GSP has also been used to derive insights into the intermediary stages of deep neural networks [105], often described as a "black box" model. Given the natural occurrence of structured as well as ad hoc networks in settings such as wireless sensor systems, power distribution systems, and smart grids, GSP has emerged as a powerful new tool for analyses. [70] studies the usage of a GSP-based framework for detecting anomalies in wireless sensor networks, demonstrating advantages over traditional PCA- and clustering-based anomaly detection methods. Similarly, anomalies within smart grid systems that might be caused by malicious activities are analyzed by [64] using the spectrum of a graph signal. The problem of designing energy-efficient configurations within wireless sensor networks is also given a GSP-flavored solution in [134]. The computationally complex problem of monitoring individual "smart" appliances connected to an electrical grid can also be more efficiently tackled using GSP; specifically [111] leverages the total variation metric of a graph signal, which we will also exploit for airport delay signal outlier detection purposes.

The final area of past research we will survey is closely related to applications of GSP in analyzing air transportation networks. GSP and graph wavelets have been used in a wide range of applications within civil and environmental engineering. Air pollution data from major metropolitan cities were analyzed in [126] using graph signal representations and GSP. More related to transit and mobility, spectral graph wavelets are leveraged in [62] to infer people's mobility patterns with only privacy-

retaining aggregate information. Spectral methods have been used to study surface traffic congestion [57, 184] as well as air traffic flows in Air Route Traffic Control Centers (ARTCCs) [65, 66]. In particular, [65, 66] 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.

2.1.2 Modeling the air transportation system and aviation delay dynamics

There has been a wide range of previous research focused on modeling and predicting delays in the air transportation network. The diversity of techniques reflects the complexity of the system at hand. Starting at the airport itself, aircraft congestion on the surface resulting in taxi-in and taxi-out delays has been modeled [246, 247, 110], along with sophisticated tools to optimize and control airport surface operations [248, 136, 11]. Extending just past the surface and to the terminal airspace surrounding a major airport, this busy airspace can be broadly divided into the terminal departure airspace and the terminal arrival airspace. Both of these subsystems have been examined extensively in terms of delays and flight inefficiencies [201, 167, 165, 271]. Zooming out to the perspective of entire air transportation networks – and into settings more similar to that of this thesis – several previous works have focused on modeling the dynamics of flight delays within airport networks, using approaches ranging from queuing theory [212, 213], network models [98, 99], discrete event simulators [4], contagion-based susceptible-infected-susceptible (SIS) models [277], agent-based simulations [216], and machine learning [143, 284, 289]. Staying with the larger, system-wide perspective, another large body of previous literature deals with modeling the underlying distribution of airport and flight delays, along with any secondary (or reactionary) delays that are generated. A sampling of some of these approaches include analytical econometric models that first separates primary versus secondary delays, then tries to determine the independent variables which ini-

tiate and perpetuates secondary, propagated delays [138, 132]. Other studies directly attempt to faithfully recreate the probability distribution of incurred delays, either in an aggregate sense [266, 215] or during specific operational segments, e.g., during the en route phase of flight [22]. While the new insights developed by these prior work help explain, for example, why one might observe certain historical correlations between delays at different airports, they do not characterize the spatial behavior of delay across a network in a succinct and interpretable manner.

Prior efforts have also considered grouping weather phenomena in the US National Airspace System (NAS) [233, 103, 188], clustering similar airport arrival capacity profiles [168, 40] and traffic management initiatives such as Ground Delay Programs (GDPs) [147, 102, 218], identifying anomalous aircraft trajectories [241, 152, 153], and more pertinently, clustering delay networks [217, 101] and irregular operations scenarios [140]. A more recent line of work focuses on network topology, connectivity, and spectral properties such as eigencentrality: A comprehensive overview of this line of work is given in [286]. For example, [219] focuses on the connectivity of airports in terms of non-stop flights. Studies have also analyzed the topology of various air transportation networks to identify key nodes [67, 174, 106, 142, 268] and edges [68], evaluate network stability [287], and characterize the role of airspace structure on flight conflicts [288]. However, besides the aforementioned fact that these studies still only consider the magnitude, and not the spatial distribution, of disruptions or delays, there are two additional common limitations: Typically, these post-hoc, network and complexity analyses do not allow for tactical, hour-by-hour insights, whereas we are able to do so by adjusting the temporal scale of our graph signals. Secondly, these previous studies typically examine, and provide insights into, only particular sub-components of the system such as a specific airport, a small group of airports, or individual flight trajectories.

2.1.3 Outlier detection methods

The identification of unexpected spatial delay distributions is related to the broader problem of outlier detection in multivariate data sets. A common approach to outlier

detection in such settings involves clustering, where data points that are far away from clusters are labeled as outliers [108]. 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 [221, 87]. Information-theoretic approaches have also been considered to identify structural outliers [242, 69]. Previous literature related to graph-based data have used information theory to identify structural features [69] and spatial outliers [242]. One particular characteristic, *Total Variation (TV)*, a measure of the smoothness of a signal supported on a graph, has been used as a feature vector for classification [5]. While the outlier detection problem has been addressed in several contexts (for example, time-series data [262, 92]), it has received little attention in the context of graph signals. One theory-oriented goal of the work in Chapter 4 attempts to fill this gap by using TV as a metric to identify graph signals with off-nominal or unexpected spatial distributions. The philosophy of our approach is most similar to that explored by [70, 111, 258]; as an example, recall that [64] leverages the spectrum of the graph signal to detect false data injection attacks in a hypothetical smart grid power distribution environment.

Furthermore, while these outlier detection approaches have been successfully deployed, their usefulness for our aviation-specific problem is limited since (1) they do not identify outliers based on spatial distributions; (2) they cannot decouple the detection of outliers based on magnitude versus spatial characteristics; and (3) there may not be an interpretable explanation as to why a particular point was classified as an outlier. In particular, the interpretability of our results is critical to providing implementable insights for a variety of stakeholders, ranging from air traffic flow managers to airline flight dispatchers and network analysts.

2.2 Low-dimensional representations of network processes

The study of dynamical processes supported on graphs is a rich field, and we will only scratch the surface with this literature review, as well as with the initial ideas for the work in Chapter 7. In terms of well-known processes on graphs, these include diffusion [90] and cascade [273] processes, with newfound applications in, e.g., graph inference [209]. The relationship between the structure of the underlying graph and the supported dynamical process was examined in [235], with recent extensions to edge-based signal processing [237] and higher-order relationships by way of lifting graph Laplacians to generalized Hodge Laplacians [236].

In Chapter 7, we are interested in *projecting* graph processes (in our case, airport delays supported on nodes of an airport network graph) to some low-dimensional space of on-time performance measures. In terms of dimensionality reduction and projections of graph processes, or more generally, data and signals supported on graphs, a closely-related motivation is determining *controllability* of multi-agent systems that can be abstracted as graphs. These often involve some sort of partitioning or clustering of the original graph abstraction, then proving controllability properties on this reduced-order representation. This network controllability question was addressed for the case of directed graphs with linear Laplacian dynamics of the form $\dot{\mathbf{x}} = -\mathcal{L}\mathbf{x} + B\mathbf{u}$ in [3]. Stability results via block-diagonal Lyapunov functions for dissipative systems were addressed in [122]. In [122], as well as related works in [229, 186, 187], a graph clustering approach was often deployed, wherein the original graph abstraction is reduced by clustering nodes together based on some similarity measure. A modified projection-based approach in [185] uses graph partitions to preserve the structure of the original network, so that it is not lost during a dimensionality-reducing projection operation via, e.g., Petrov-Galerkin. Other dimensionality reduction approaches leverage graph embeddings [280], with extensions to non-linear embedding types and kernel-based PCA on graphs [243].

There is also a range of previous work that leverages graph signal processing

for purposes of dimensionality reduction and projections onto a lower-dimensional subspace. We were interested in surveying these works given the usage of measures such as the total variation of graph signals and graph Fourier transforms in this thesis, particularly in Chapter 4 and Chapter 5. In [224], the dominant eigenvector of the graph Laplacian is used to learn a low-dimensional linear subspace; similar graph filter-based dimensionality reductions were used in [193] to process neuroimaging data sets, and extended to dimensionality reduction in an online setting [238]. Once these reductions are achieved, the inverse problem of reconstructing signals that are known to lie in some lower-dimensional subspace can be examined [267], with potential improvements in the reconstruction process via the minimization of the total variation of a graph signal [130].

2.2.1 Identification and analysis of aviation disruptions

Prior work has defined aviation system recovery in terms of flight delays, displaced or delayed crews, and disrupted or delayed passengers [16, 54]. However, these recovery definitions and strategies are airline-specific and of limited use in defining and measuring system-wide characteristics. We expand on this line of work by incorporating *system-wide spatio-temporal information*, allowing us to construct our comprehensive definitions of disruptions and recoveries in Chapter 6. Analyzing disruptions and recoveries in a networked system is a growing field of study, and is closely related to understanding system resilience. Significant prior work has focused on developing models for such systems, and then analyzing them theoretically or through simulations [189, 127]. In [53], a system-level analysis was performed to model different types of synthetic disruptions and recoveries for simplified network models; they found that in simple aviation networks, there was asymmetry in the disruption and recovery timescales. Our work in Chapter 6 complements this literature by developing a framework that enables a data-driven analysis of historical disruption-recovery cycles.

Broadly, previous work related to analyzing disruptions in the air transportation network focused on aspects such as identifying common patterns, examining the impact of singular irregular operations (IROPs) events, or computing new performance

measures. In terms of extracting common patterns or trends, clustering-based techniques have been used to identify time periods with similar weather conditions and traffic management initiatives (TMIs) [102, 239, 101]. More recently, [140] applies k-means clustering directly to segment IROPs events of similar duration and severity, but do not address the initial identification problem. On the other hand, in [1] and [255], the authors pursue similar motivations as ours in Chapter 6 with regards to identifying sequences of disruptions, but the sequences are either restricted to be periodic [1], or the criteria for determining what is a disruption sequence need to set with manual heuristics [255]. Furthermore, neither works explicitly take into account the underlying network connectivity [222, 158]: We apply the models and results from Chapter 4 and Chapter 5 to integrate information regarding the underlying network.

We also point out previous research that seeks to identify more macro-level precursors of flight delays and disruptions, e.g., airline capacity decisions and jet fuel prices [183], as well as micro-level issues such as specific sub-groups of "deviant" flights [211], how flights operated by the same aircraft might proactively recover from an earlier disruption [2], and individual one-off disruption events [226, 177, 205, 225, 141]. However, micro-level analyses like the one in [2] often require detailed, potentially proprietary information such as crew and flight schedules. Furthermore, examining individual events, or specific tactical disruption management strategies [256] after disruptive events, provide useful case studies, but not a generalizable method for identifying future aviation disruptions. In summary, there remains a lack of a formal framework for defining disruptions and recoveries in networked systems, a gap that we address in Chapter 6.

2.2.2 Predictive models in air traffic management

We focus the remainder of this literature review section on a variety of predictive models in air transportation research. In particular, we are interested in models that map predictors (e.g., weather, time-of-day, etc.) to various on-time performance indicators. This non-exhaustive review of predictive models in aviation serves two purposes: First, in surveying this wide variety of predictive models, we emphasize that

any predictive model for on-time performance requires a baseline definition of what a disruption and recovery within an airport network is in order to be appropriately trained. Formalizing such a definition is a central motivation and contribution of Chapter 6. Second, this literature review subsection provides context for ongoing and future work related to predictive models of on-time performance and airline IROPs, further discussed in Section 6.9.

Since imbalances between system capacity and demand drive deteriorating on-time performance and flight delays, much attention has been given to predicting arrival capacities at airports, i.e., Airport Arrival Rates (AARs) [270, 56, 192, 150, 291], as well as departure-side activities such as the turn-time of a flight [292]. The flight delay and cancellation prediction problem has been studied from the demand side as well [148]. Another prediction target for researchers has been the TMIs, such as Ground Delay Programs (GDPs), that are activated once demand-capacity imbalances occur [170, 74]. Finally, with the appropriate data fidelity, previous works also examine predictions based on aircraft traffic flows [191], aircraft trajectories within the terminal airspace or terminal maneuvering area [265, 190], individual flight reroutes [75], as well as aircraft surface taxi times [13, 61, 112] and en route flight times [290]. For these wide-ranging prediction tasks, it is oftentimes useful to distinguish between time periods when the air transportation system is known to be disrupted, versus time periods when the system is operating nominally.

In terms of some methods used to map between predictors and predicted values, a popular approach is clustering-based techniques, combined with a variety of regression models. These types of methods have been used for predicting flight delays at an origin-destination level [217, 95], at the level of individual flights [50] or airports [239], as well as on-time performance within the aviation system more generally [261, 46, 73]. Furthermore, deep learning-based models such as deep neural networks and deep belief networks [259, 284], autoencoder [283], long short-term memory recurrent neural networks (LSTM RNNs) [143, 265], and reinforcement learning-based approaches [13, 112] have also been deployed, generally trading off interpretability for improvements in predictive accuracy. However, we will discuss using more inter-

pretable multi-dimensional kernel regression-based methods, in addition to learning-based approaches such as LSTM RNNs, in Section 6.9 to address prediction problems that naturally arise from our disruption-recovery trajectory (DRT) setting formalized in Chapter 6. For example, these problems include predicting the *trend*, e.g., increasing versus decreasing delays, of various performance measures for an airline's network in the next hour, and predicting *whether or not* an airline's network is about to become disrupted.

2.3 System models and traffic flow management

2.3.1 System modeling and copula distributions

High-dimensional, potentially interconnected systems are difficult to accurately model and control. The task of system identification examines developing system models based on measurements of input and output signals from the system in question [249, 171]. One difficulty in applying these system identification methods to our airport delay redistribution problem is the lack of sufficient numbers of observations for the given level of complexity, noise, and dimensionality within the system. Alternatively, in lieu of the entire system state, we could consider a reduced-order model, i.e., a model for a lower-dimensional representation of the system [137, 207]. The control of systems using such reduced-order models has been widely studied, such as in the field of computational mechanics [24, 228], as well as aforementioned literature on the control of graph processes in Section 2.2. Although such reduced-order models could be used in the analysis of any high-dimensional system, part of our work in Chapter 7 focuses on using an optimization formulation to map a desired control trajectory from a low-dimensional space to a multivariate Gaussian copula approximation for the higher-dimensional state space.

Copula distributions can be used to represent any multivariate distribution by de-coupling dependence from marginals [250]; they have been widely-used in finance, particularly in risk management and option pricing (see, e.g., [72, 144]). Copula-

based models have been leveraged recently for regression estimation and inference [200], information retrieval [71], and generating latent space models [272]. For a comprehensive introduction to copulas, see [199]. The use of copulas for representing and modeling the states of systems is also a more recent line of work. Methods to explicitly construct state spaces from copulas [145] and using inversion copulas to model parametric state space models (e.g., linear Gaussian) [251] have been explored.

2.3.2 Traffic flow management and airport ground holding

The goal of the approximate network control model we propose in Chapter 7 is to serve as a *high-level planner* for NAS delay states. This high-level planner proposes an unconstrained reference plan for NAS delays to be given to a complementary *low-level controller* such as a multi-airport ground holding problem (MAGHP) model, which then constrains the plan against a baseline flight schedule to determine ground and airborne delay assignments. While we defer the integration of the high-level planner and low-level controller to ongoing work in [49], since we consider modifications to the MAGHP that investigates the ability to redistribute delays at the airport level while maintaining a feasible schedule, we first provide a review of past work related to the airport ground holding problem (AGHP) and the air traffic flow management problem (ATFMP).

A comparative analysis of various MAGHP implementations, as well as simplified instances of the ATFMP [28], in [8] observed "smoothing" effects on the input flight schedule, a phenomenon that we will try to purposefully induce at the airport level via delay redistribution requirements. Additionally, previous work examines traffic management solution sensitivities with respect to factors such as ground to airborne delay cost ratios [42] and sector utilization costs [279].

Extensions to the MAGHP include scenario-based, stochastic implementations [220], along with inclusion of airline scheduling behaviors such as departure and arrival banks [113]. More recent work addresses the question of equitable delay distributions in traffic flow management; these include restrictions on allowable delays on a flight-by-flight basis [18] (trades off aggregation versus distribution of delay), across

airlines [26] (trades off slot swap flexibility versus satisfaction of network connectivity constraints), as well as examining trade-offs between ration-by-schedule (RBS) deviations versus total delay [146].

A key feature of the high-level planner in Chapter 7 examines the question of airport delay *redistribution*. This perspective of explicitly focusing on delay equity at airports stands in contrast to the flight-specific [18] or airline-specific [26] perspective. We propose implementing delay redistribution requirements in two ways: (1) Impose upper bounds on delays at individual airports within the MAGHP; (2) ensure "smoothness" of airport delays with respect to an underlying network structure [100, 158] by adding penalties for uneven airport delay distributions to the objective function. The latter approach is similar to ongoing work in [48] in terms of penalizing specific delay characteristics in the objective function of the baseline MAGHP.

Chapter 3

Thesis Contributions

3.1 Contributions of Chapter 4

The collection of methods we present in Chapter 4 seeks to identify outlier NAS days containing disruptions with unexpected airport delay dynamics. We consider multiple delay distribution characteristics during the outlier identification process, and ensure that outliers can be interpreted in the context of NAS operations. 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 and procedures, hereby reducing the need for tactical re-planning. Similar benefits specific to airline sub-networks can be derived from our airline-specific analysis. Furthermore, the methods proposed in this chapter may be adapted and applied to a number of networked systems, including other transportation networks, the Internet-of-Things, power systems, communication networks, and biological systems. The major contributions of Chapter 4 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. 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.

3. We further categorize days in terms of the severity of delay and cancellation impacts, and correlate these standard notions of disruptions with new spatial metrics, i.e., total variation.

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 these sub-networks and the system-wide network.

The thesis work documented in Chapter 4 has been published in part in [100, 161, 160, 158].

3.2 Contributions of Chapter 5

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. The contributions of Chapter 5 is complementary to Chapter 4: The proposed methods leverage GSP to elucidate on outlier delay signals identifiable via the collection of methods from Chapter 4. This in turn provides interpretable explanations for why certain data points are classified as outliers. To showcase applicability, we conduct four data-driven analyses focusing on: (1) The US Core 30 airport network; (2) the sub-networks of four US domestic airlines; (3) the sub-networks of four Chinese domestic airlines; and (4) differences in operational characteristics between the US and China networks. The main findings and contributions from this part of Chapter 5 are as follows:

1. *Identifying critical subsets of airports.* We find that the delay distributions at a small set of geographically proximate US East Coast airports are indicative of whether the system-wide delay distribution is expected or unexpected. By

contrast, such an equivalent set of airports in China is spread out over a large portion of the Chinese airport network.

2. *Variability in spatial delay distributions.* The baseline variability in terms of spatial distribution for delays in China is higher than in the US.

3. *Scale versus spatial delay distribution outliers.* We see that even though the baseline delay magnitudes are higher in China, the US experiences more outlier days in terms of delay magnitudes. On the other hand, China experiences more outliers in terms of spatial delay distributions, even though it has a higher baseline level of variability in its spatial delay distribution. In terms of temporal trends, we find significantly more outliers in winter months for both China and the US.

4. *Framework for interpreting outliers.* We provide a two-step process for interpreting outliers that considers both the type of outlier and the operational implications. We demonstrate this framework by analyzing specific outlier days in China and the US. Our analysis not only identifies the specific airport delays contributing to the outlier classification, but also corroborates findings with operational factors.

The thesis work documented in Chapter 5 has been published in part in [100, 161, 160, 158].

3.3 Contributions of Chapter 6

Our main contributions of Chapter 6 lie in formalizing a framework for examining disruptions and recoveries in networks, drawing from graph signal processing and the state-space representation of dynamical systems. Specifically, we consider disruptions and recoveries not only in terms of signal magnitudes, but also their spatial distribution and temporal evolution. Our framework is of potential use to air traffic managers who might be interested in characterizing and improving the resilience of air

transportation systems. Airlines and passengers may also benefit from an improved understanding of disruptions and subsequent recoveries. Finally, our work generates a reference data set of disruption-recovery cycles that can be used to benchmark system recovery. With regards to the motivation for Chapter 6 and a survey of relevant previous work in Section 2.2, the contributions of Chapter 6 are as follows:

1. We leverage techniques from graph signal processing to comprehensively define the start, progression, and end of flight delay disruption-recovery cycles in a network of airports. Our method not only considers the magnitude of delays, but also their spatial distribution, their relation to historical delay patterns, and temporal trends.

2. We identify *disruption-recovery trajectories* using operational data, and develop appropriate features in order to cluster them into representative groups. One of our key technical contributions is the choice of incorporating spectral graph-theoretic and temporal features to describe disruptions and recoveries in airport networks.

3. We uncover and interpret two interesting observations related to: (1) The behavior of disruption and recovery during off-nominal events (e.g., airport outages), and (2) the temporal trends in disruption-recovery trajectories. We also interpret representative system behaviors, extracted via clustering disruption-recovery trajectories.

4. We refine the disruption-recovery trajectory framework, and prove structural properties about these objects. We also provide operational interpretations for these structural properties of interest.

5. We extend the disruption-recovery trajectory framework to flight cancellations, and classify the interactions between delay- and cancellation-type disruption-recovery trajectories.

The thesis work documented in Chapter 6 has been published in part in [159], and is part of ongoing work in [156].

3.4 Contributions of Chapter 7

Our primary methodological contribution in Chapter 7 centers on the copula-based approximate network control framework. Given performance targets and control requirements that lie in some lower-dimensional space spanned by predetermined, human-interpretable performance metrics, we construct a *desired system trajectory* based on goals set within the lower-dimensional space. Then, through encoding these goals as constraints in an optimization routine, we construct the corresponding full-dimensional state trajectory by extracting candidate states from a copula-generated approximate state space. With regards to this approximate network control framework, our contributions are three-fold:

1. We develop a method to identify a sequence of future states and associated control actions for difficult-to-model high-dimensional systems based on performance targets in a lower-dimensional space.

2. We demonstrate how our approach can be used to drive the evolution of airport delays in accordance with aggregate performance targets.

3. We describe the generalizability of our framework and its relevance to networked systems.

We then discuss controlling NAS delays with hierarchical objectives, wherein the approximate network control model serves as a high-level planner, which is then constrained by a low-level controller (e.g., an implementation of the TFMP or MAGHP). While we defer the integration step to ongoing work in [49], we detail characteristics of the high-level planner via a case study of disruptions focused on the three major New York City airports.

1. We show that the high-level planner is able to successfully provide potential delay state trajectories that follow pre-specified delay redistribution and conservation constraints.

2. We detail how *delay absorption costs* can be quantified using this approach, where an airport with a low absorption cost is more likely to see increases in delays post-redistribution, and vice versa.

The thesis work documented in Chapter 7 has been published in part in [155], and is part of ongoing work in [49].

Chapter 4

Spatial Distribution of Airspace Disruptions

4.1 Identification and interpretation of graph signal outliers

4.1.1 Setup, notation, and definitions

Let us consider a multidimensional real signal $\mathbf{x} \in \mathbb{R}^{N\times 1}$. The N elements of the signal may not be 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 node $i \in V$ to the i^{th} element, $x_i \in \mathbb{R}$, 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 the $(i, j)^{\text{th}}$ element is A is set to be the weight w_{ij}, i.e., $[A]_{ij} = w_{ij}$. We restrict ourselves to undirected graphs, where $w_{ij} = w_{ji}$ and $A = A^\top$. This multidimensional signal $\mathbf{x}$ can therefore be represented by a *graph signal vector* supported on the nodes of graph G.

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 $\mathbf{x}^{(k)} = \left(x_1^{(k)}, \ldots, x_N^{(k)}\right)^\top \in \mathbb{R}^{N\times 1}$. 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) with respect to $\mathcal{O}_M$ is

$$r_{ij|\mathcal{O}_M} = \frac{\sum_{k=1}^{M}\left(x_i^{(k)} - \hat{\mu}_i\right)\left(x_j^{(k)} - \hat{\mu}_j\right)}{\sqrt{\sum_{k=1}^{M}\left(x_i^{(k)} - \hat{\mu}_i\right)^2}\sqrt{\sum_{k=1}^{M}\left(x_j^{(k)} - \hat{\mu}_j\right)^2}}. \tag{4.1}$$

We will primarily examine the undirected graph where we set $[A]_{ij} = w_{ij} \triangleq r_{ij|\mathcal{O}_M}$. Our graph can be referred to as a *correlation network*, since the edge weights correspond to the correlation computed via 4.1 between the signals at two nodes. The signals $\mathbf{x}$ are assumed to be realizations of a multivariate Gaussian random variable $\mathbf{X} = (X_1, \ldots, X_N)^\top \in \mathbb{R}^{N\times 1}$, where $\mathbf{X} \overset{\text{i.i.d.}}{\sim} \mathcal{N}(\boldsymbol{\mu}, \Sigma)$ with mean vector $\boldsymbol{\mu} = (\mu_1, \ldots, \mu_N)^\top \in \mathbb{R}^{N\times 1}$ and positive semi-definite covariance matrix $\Sigma \in \mathbb{R}^{N\times N}$, $\Sigma \succeq 0$. The correlation matrix $\mathbf{C} = [\rho_{ij}] \in \mathbb{R}^{N\times N}$ is given by:

$$\rho_{ij} = \frac{\mathbb{E}[(X_i - \mu_i)(X_j - \mu_j)]}{\sqrt{\mathbb{E}[(X_i - \mu_i)^2]\mathbb{E}[(X_j - \mu_j)^2]}}. \tag{4.2}$$

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 for our purposes. We will now define some preliminary concepts and objects from spectral theory that we will use for setting up the graph signal outlier detection problem.

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 $d_{ii} = \sum_{j=1}^{N} w_{ij}$.*

Throughout this thesis, the graph Laplacian will be denoted by either $\mathfrak{L}$ or $\mathcal{L}$, to be consistently defined and used for each chapter. 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 $v_i^\top v_j = \delta_{ij}$, where

$$\delta_{ij} = \begin{cases} 1, & \text{if } i = j \\ 0, & \text{otherwise} \end{cases}. \tag{4.3}$$

All of the eigenvalues by definition satisfy $\mathfrak{L}v_i = \lambda_i v_i$, and we will sometimes refer to a specific eigenvalue-eigenvector pair $(\lambda_i, v_i), i \in \{1, \ldots, N\}$. We sort the eigenvalues such that $\lambda_1 \leq \lambda_2 \leq \cdots \leq \lambda_N$. Since the graph Laplacian has row sums of 0 (Definition 1 gives $\mathfrak{L}_i^\top \mathbf{1} = (D_i - A_i)^\top \mathbf{1} = D_i^\top \mathbf{1} - A_i^\top \mathbf{1} = d_{ii} - \sum_{j=1}^N w_{ij} = 0$, where $\mathfrak{L}_i$, D_i, and A_i denote the i^{th} row of $\mathfrak{L}$, D, and A, respectively), $v_1 = \mathbf{1}$ is the *constant* eigenvector corresponding to the eigenvalue $\lambda_1 = 0$. Furthermore, the algebraic 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 $0 = \lambda_1 < \lambda_2 \leq \cdots \leq \lambda_N$, and $\text{span}(\{v_1, \ldots, v_N\}) \cong \mathbb{R}^{N \times 1}$. The set of eigenvectors of $\mathfrak{L}$ forms an orthogonal basis for $\mathbb{R}^{N \times 1}$. Thus, any vector $\mathbf{x} \in \mathbb{R}^{N \times 1}$ can be written as a linear combination of $\{v_1, \ldots, v_N\}$; i.e., there exist scalars α_i such that $\mathbf{x} = \sum_{i=1}^N \alpha_i v_i$. We can draw loose connections back to airport delay signals: Suppose $\mathbf{x}$ contains delay signals for airports $i = 1, \ldots, N$ and $\{v_1, \ldots, v_N\}$ can be thought of as a set of informative, "common" delay modes. Then, given an observation $\mathbf{x}$ of delays across a network of airports, we can decompose $\mathbf{x}$ into modes $\{v_1, \ldots, v_N\}$ by $\mathbf{x} = \sum_{i=1}^N \alpha_i v_i$, where α_i can be interpreted as the "contribution" of mode v_i. This begins to formalize some of the notions first mentioned in Sections 1.5.1 and 1.5.2.

We now explicate on the set of scalars $\{\alpha_1, \ldots, \alpha_N\}$ as they play a crucial role in graph signal processing and the graph Fourier transform:

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

Recall that for Fourier transforms, a time domain signal can be decomposed into linear combinations of sinusoidal signals. For the GFT, the eigenvectors of the graph Laplacian are analogous to the sinusoidal signals, but supported on the nodes of a graph. For each eigenvector-eigenvalue pair, the "graph frequency" associated with the eigenvector is given by its eigenvalue. The contribution of each eigenvector in the

linear combination which comprises the original graph signal is encoded via scalars $\alpha_i \in \{\alpha_1, \ldots, \alpha_N\}$. Similar to how sinusoidal signals with higher frequencies are considered to be more energetic, we can interpret an analogous notion of spectral energies for the GFT:

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

If we collect the scalars $\{\alpha_1, \ldots, \alpha_N\}$ from Definition 2 into a vector according to their indices $\boldsymbol{\alpha} = (\alpha_1, \ldots, \alpha_N)^\top$, and collect the set of normalized Laplacian eigenvectors column-wise to form the $N \times N$ matrix

$$\mathbf{V} = \begin{bmatrix} | & & | \\ v_1 & \cdots & v_N \\ | & & | \end{bmatrix},$$

since the set of normalized Laplacian eigenvectors forms an orthonormal basis for $\mathbb{R}^{N\times 1}$, we have that $\mathbf{V}$ is Hermitian, and $\mathbf{V}\mathbf{V}^\top = I_N$, where I_N is the $N \times N$ identity matrix. Now, the graph Fourier transform can be written compactly as $\boldsymbol{\alpha} = \mathbf{V}^\top \mathbf{x}$, and an inverse graph Fourier transform $\mathbf{x} = \sum_{i=1}^N \alpha_i v_i$ can be written as $\mathbf{x} = \mathbf{V}\boldsymbol{\alpha}$, since $\mathbf{V}\boldsymbol{\alpha} = \mathbf{V}\mathbf{V}^\top \mathbf{x} = I_N \mathbf{x} = \mathbf{x}$. Similarly, the energy conservation claim in Definition 3 follows directly as well: With the newly defined vector $\boldsymbol{\alpha}$, we note that $\sum_{i=1}^N \alpha_i^2 = \|\boldsymbol{\alpha}\|_2^2 = \boldsymbol{\alpha}^\top \boldsymbol{\alpha} = \left(\mathbf{V}^\top \mathbf{x}\right)^\top \mathbf{V}^\top \mathbf{x} = \mathbf{x}^\top \mathbf{V}\mathbf{V}^\top \mathbf{x} = \mathbf{x}^\top I_N \mathbf{x} = \|\mathbf{x}\|_2^2$.

The graph Laplacian can also be used to compute a measure of the "smoothness" of a graph signal $\mathbf{x}$, which can be interpreted as how much $\mathbf{x}$ varies with respect to the connectivity and weights in the underlying graph. Specifically, this measure is a quadratic form in $\mathbf{x}$ and $\mathfrak{L}$ called the *total variation*.

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

$$\mathrm{TV}(\mathfrak{L}, \mathbf{x}) = \frac{1}{2} \sum_{i \neq j} w_{ij} \left(x_i - x_j\right)^2 = \mathbf{x}^\top \mathfrak{L} \mathbf{x}. \tag{4.4}$$

For brevity, we write $\mathrm{TV}(\mathbf{x})$*, 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* $\mathrm{TV}(\mathbf{X})$*.*

The total variation provides a metric map $\mathrm{TV} : \mathbb{R}^{N\times N} \times \mathbb{R}^{N\times 1} \to \mathbb{R}$ that measures the smoothness of a graph signal. A low value of TV corresponds to a graph signal that is said to be *smooth* relative to a graph signal with a higher TV, assuming that the TV of both graph signals are with respect to the same graph Laplacian $\mathfrak{L}$. This notion of smoothness is related to how much the graph signal varies across adjacent nodes, modulated by the edge weight connecting adjacent nodes. When the difference in nodal signal values is large across a given, strongly-weighted edge, then its contribution to the TV is larger. The following proposition helps to interpret the GFT in terms of the TV and signal smoothness:

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

(i) $\mathrm{TV}\left(\mathbf{x}^{(k)}\right) = \left(\mathbf{x}^{(k)}\right)^\top \mathfrak{L}\mathbf{x}^{(k)}$,

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

where $\lambda_1, \ldots, \lambda_N$ *are the eigenvalues of the graph Laplacian* $\mathfrak{L}$*.*

Proof. See Appendix A.1. □

We see from Proposition 1 that the total variation measure of a graph signal $\mathbf{x}^{(k)}$, in addition to being written as a quadratic form on $\mathbf{x}^{(k)}$ and the graph Laplacian, can also be written as a weighted sum of each eigenvalue λ_i of the graph Laplacian, weighted by the spectral energy of the associated i^{th} eigenvector v_i. This ties together notions of graph signal smoothness (i.e., the TV) and spectral energies: In particular, the more a high-energy (i.e., higher graph frequency λ_i) eigenvector v_i contributes to the signal, the graph signal's TV will be higher as well. This is why we pre-sort the eigenvector-eigenvalue pairs in increasing order, with respect to index $i \in \{1, \ldots, N\}$, so that we can conveniently compare via indices i in lieu of eigenvalues.

For the remainder of Chapter 4, we focus on using TV to define graph signal outliers, with applications to airport delays in networks. With regards to the graph Fourier transform and decomposing airport delays into common delay modes, this will be the focus of Chapter 5.

4.1.2 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 the term $r_{ij|\mathcal{O}_M}(x_i - x_j)^2$. We consider the following possible scenarios of the correlation value and the contribution of the specific term $r_{ij|\mathcal{O}_M}(x_i - x_j)^2$ to the TV:

> **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. In other words, 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 $r_{ij|\mathcal{O}_M}(x_i - x_j)^2$ 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 $r_{ij|\mathcal{O}_M}(x_i - x_j)^2$ 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 $r_{ij|\mathcal{O}_M}(x_i - x_j)^2$ is small.

We note that the interpretations we provide in Case 1 and Case 2 hold true if we consider the absolute magnitude of TV, i.e., $|\mathrm{TV}(\mathbf{x})|$, along with having $r_{ij|\mathcal{O}_M} \geq 0$

or $r_{ij|\mathcal{O}_M} \leq 0$, for all pairs i and j. We note that these interpretations are still valid even if a small percentage of correlations $r_{ij|\mathcal{O}_M}$ have a different sign compared to all other correlation coefficients, as we could perform an affine transformation $\max\left(r_{ij|\mathcal{O}_M}, 0\right)$ or $\min\left(r_{ij|\mathcal{O}_M}, 0\right)$, depending on the sign of the other majority of correlation coefficients. We will explicate on this point in Section 4.1.7. An interesting question for future research may be to ascertain a maximal proportion $\xi^\star$ of correlation coefficients that differ in sign before the above interpretations no longer hold true: Formally, we have that

$$\xi^\star = \underset{\xi \in [0,1]}{\operatorname{argmax}} \left\{ \xi = \frac{(1/2)\mathbf{1}_{(i,j),(k,l)\in E}\left(\operatorname{sign}\left(r_{ij|\mathcal{O}_M}\right) \neq \operatorname{sign}\left(r_{kl|\mathcal{O}_M}\right)\right)}{|E|} \right\},$$

before the magnitude $|\mathrm{TV}(\mathbf{x})|$ no longer provides consistent interpretations. We denote by $\mathbf{1}(\mathcal{A})$ the indicator function for event $\mathcal{A}$.

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*

$$\mathrm{TV}(\mathbf{x}) \notin \left[\mathbb{E}\left[\mathrm{TV}(\mathbf{X})\right] - k\sqrt{\mathrm{Var}\left[\mathrm{TV}(\mathbf{X})\right]},\, \mathbb{E}\left[\mathrm{TV}(\mathbf{X})\right] + k\sqrt{\mathrm{Var}\left[\mathrm{TV}(\mathbf{X})\right]}\right], \qquad (4.5)$$

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.*

Note that since $\mathbf{X}$ is a random vector, 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.

Note that if the graph Laplacian is simply the identity, we would have that $\mathrm{TV}(\mathbf{x}) = \mathbf{x}^\top\mathbf{x} = \|\mathbf{x}\|_2^2$. 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*

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

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.*

Here, for a graph signal $\mathbf{x}$, the definition of outliers in scale examines the magnitude $\|\mathbf{x}\|$ of the graph signal, which is distinct from that of Definition 5, i.e., examining the spatial distribution of $\mathbf{x}$. However, since $\mathrm{TV}(\mathbf{x})$ is directly related to $\|\mathbf{x}\|$, we need a richer definition for outliers in spatial distribution that *isolate* the impact of $\|\mathbf{x}\|$ on the total variation. Thus, we modify Definition 6 by replacing the expectation and variance of the total variation with conditional versions, where the conditioning is on the realized norm $\|\mathbf{X}\| = \|\mathbf{x}\|$ of the graph signal.

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* $\mathrm{TV}(\mathbf{x}) \notin [\mathcal{A}, \mathcal{B}]$*, where*

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

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* $\|\mathbf{X}\| = \|\mathbf{x}\|$*.*

For analytical tractability we consider only the 1-norm for all graph signal vectors, i.e., $\|\mathbf{x}\| = \sum_{i=1}^{N} |x_i|$ unless otherwise specified. Furthermore, we assume that all graph signals are non-negative, so the 1-norm for these graph signal vectors becomes equivalent to $\|\mathbf{x}\| = \sum_{i=1}^{N} x_i$. We justify this particular assumption as many relevant physical parameters such as airport delays are non-negative signals supported on nodes. To summarize our various definitions of outliers, Figure 4-1 presents a graphical description of these definitions. Our goal is now to derive analytical bounds for both outliers in scale and weak outlier in distribution. Deriving analytical bounds for strong outliers in distribution remains an open problem since the conditioning on the realized norm complicates constructing analytical bounds. We therefore develop an algorithm for computing *empirical* bounds for strong outliers in distribution via simulation.

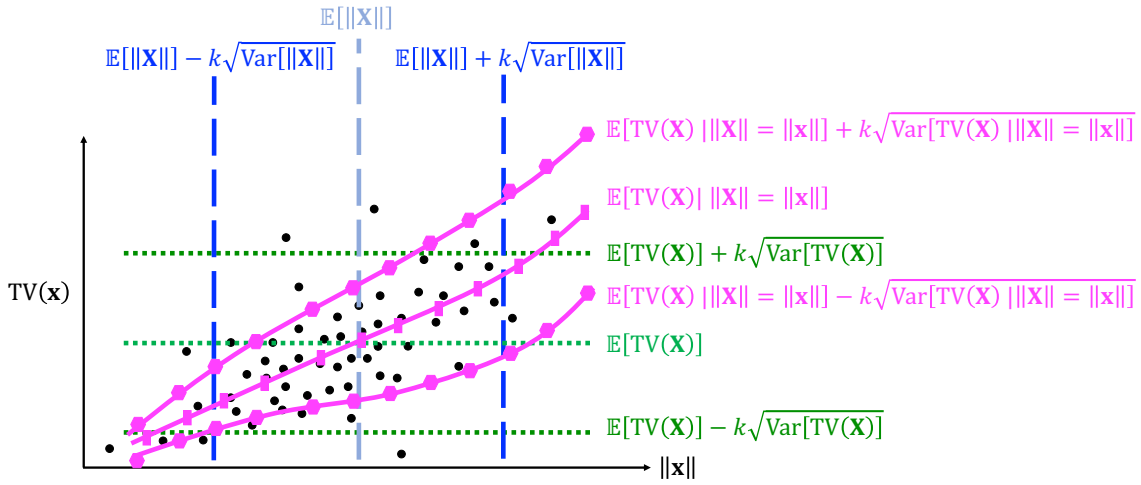


Figure 4-1: Notional representation of bounds that we will derive analytically (outliers in scale and weak outliers in distribution) and empirically via simulation (strong outliers in distribution). *Reprinted from [158].*

We emphasize that although the mean and variance for $\|\mathbf{X}\|$ and $\mathrm{TV}(\mathbf{X})$ can be approximated empirically from $\mathcal{O}_M$, our contribution lies in deriving analytical closed-form expressions for $\|\mathbf{X}\|$ and $\mathrm{TV}(\mathbf{X})$, as well as for outliers in scale and weak outliers in distribution. These analytical expressions offer two critical advantages over simulations: First, 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. Second, there may be an insufficient number of data observations in $\mathcal{O}_M$ to reliably estimate the bounds. With these motivations in mind, we present the bounds for outliers in scale (Section 4.1.3), followed by the bounds for weak outliers in distribution (Section 4.1.4). Finally, we propose an algorithm for computing empirical bounds for strong outliers in distribution via simulation (Section 4.1.6). We present a related discussion in Section 4.1.7 regarding the practical scenario in which 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.

4.1.3 Bounds for outliers in scale

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

$$\begin{aligned} \mathbb{E}\left[\|\mathbf{X}\|\right] &= \mathbb{E}\left[\sum_{i=1}^{N} X_i\right] = \sum_{i=1}^{N} \mathbb{E}\left[X_i\right] = \sum_{i=1}^{N} \mu_i, \\ \mathrm{Var}\left[\|\mathbf{X}\|\right] &= \mathrm{Var}\left[\sum_{i=1}^{N} X_i\right] = \sum_{i=1}^{N} \mathrm{Var}\left[X_i\right] + \sum_{i \neq j} \rho_{ij}\sigma_i\sigma_j, \end{aligned} \tag{4.6}$$

where σ_i and σ_j are the standard deviations of X_i and X_j, respectively. If the correlations between the nodal signals X_i and X_j are known precisely, then the first term can be rewritten as the sum of all elements in the known covariance matrix Σ, i.e., $\sum_{i=1}^{N} \mathrm{Var}\left[X_i\right] = \mathbf{1}^\top \Sigma \mathbf{1}$.

4.1.4 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 unconditional mean explicitly as:

$$\begin{aligned}\mathbb{E}[\mathrm{TV}(\mathbf{X})] &= \mathbb{E}\left[\frac{1}{2}\sum_{i\neq j}\left\{\rho_{ij}(X_i - X_j)^2\right\}\right] \\ &= \frac{1}{2}\sum_{i\neq j}\left\{\rho_{ij}\left(\mathbb{E}\left[X_i^2\right] + \mathbb{E}\left[X_j^2\right] - 2\,\mathbb{E}\left[X_i X_j\right]\right)\right\}.\end{aligned} \tag{4.7}$$

From (4.2), we have that $\rho_{ij} = \mathbb{E}\left[(X_i - \mu_i)\left(X_j - \mu_j\right)\right] / \left(\sigma_i \sigma_j\right)$, which gives $\mathbb{E}\left[X_i X_j\right] = \mu_i \mu_j + \rho_{ij}\sigma_i\sigma_j$ since $\mathbb{E}\left[(X_i - \mu_i)\left(X_j - \mu_j\right)\right] = \mathrm{Cov}\left[X_i, X_j\right] = \mathbb{E}\left[X_i X_j\right] - \mu_i\mu_j$. Furthermore, it follows that $\mathbb{E}\left[X_i^2\right] = \mu_i^2 + \sigma_i^2$. Now, (4.7) simplifies to

$$\mathbb{E}[\mathrm{TV}(\mathbf{X})] = \frac{1}{2}\sum_{i\neq j}\left\{\rho_{ij}\left[(\mu_i - \mu_j)^2 + (\sigma_i^2 + \sigma_j^2 - 2\rho_{ij}\sigma_i\sigma_j)\right]\right\}. \tag{4.8}$$

We can examine a few special cases for the parameters in (4.8):

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

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

$$\mathbb{E}[\mathrm{TV}(\mathbf{X})] = \sum_{i\neq j}\left\{\rho\sigma^2(1-\rho)\right\} = N(N-1)\rho\sigma^2(1-\rho). \tag{4.9}$$

We now detail 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:

$$\mathrm{Var}\left[\mathrm{TV}(\mathbf{X})\right] = \mathbb{E}\left[\mathrm{TV}(\mathbf{X})^2\right] - \mathbb{E}\left[\mathrm{TV}(\mathbf{X})\right]^2. \tag{4.10}$$

We have already derived the expression for the second term, since from (4.8) we can square the entire expression, and 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 in the right-hand side of (4.10) representing the expectation of the square of the TV can be expanded and rewritten as:

$$\mathbb{E}\left[\mathrm{TV}(\mathbf{X})^2\right] = \frac{1}{4}\,\mathbb{E}\left[\left(\sum_{i\neq j}\left\{\rho_{ij}(X_i-X_j)^2\right\}\right)^2\right]. \tag{4.11}$$

If we take $\mathbf{X}$ to be a multivariate Gaussian random variable, expanding (4.11) further will produce terms that depend on products of dependent Gaussian random variables. Specifically, we have that 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 $i,j,l,m \in V$. While the expression for $\mathbb{E}\left[X_i^4\right] = \mu_i^4 + 6\mu_i^2\sigma_i^2 + 3\sigma_i^4$ is easy to construct, generic analytical expressions for products of dependent Gaussian random variables have been derived in [123] and [133]:

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

$$\mathbb{E}\left[\prod_{i=1}^N X_i^{s_i}\right] = \sum_{\nu_1=0}^{s_1}\cdots\sum_{\nu_N=0}^{s_N}\sum_{r=0}^{[s/2]}\binom{s_1}{\nu_1}\cdots\binom{s_N}{\nu_N}\times\left\{\frac{\left(\frac{\boldsymbol{h}^\top\Sigma\boldsymbol{h}}{2}\right)^r\left(\boldsymbol{h}^\top\boldsymbol{\mu}\right)^{s-2r}}{r!(s-2r)!}\right\}, \tag{4.12}$$

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

Proof. See [133]. □

As we have previously mentioned, due to the large number of terms present in the full expansion of (4.11), 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 (4.12). There are also expressions for the unconditional expectation and variance of TV, with Gaussian assumptions on $\mathbf{x}$, based on the trace of the graph Laplacian and covariance matrix; see (4.17) for these expressions, which stem from Theorem 3.2b.2 of [178].

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 $\boldsymbol{\mu}$. Even with, e.g., (4.17), we see that the dependence of the variance on various parameters is complicated. However, one sufficient condition for the TV to be equal to zero is as follows:

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

Proof. See Appendix A.2. □

Note that $\mathbb{E}[\mathrm{TV}(\mathbf{X})] = 0$ is not a necessary condition for the variance of the TV to be 0. This can be seen from the following counterexample: Suppose $\rho_{ij} = 1$, $\mu_i \neq \mu_j$, and $\sigma_i = \sigma_j, \forall i, j \in V$. Then, we have that in general $\mathbb{E}[\mathrm{TV}(\mathbf{X})] = \frac{1}{2}\sum_{i \neq j}(\mu_i - \mu_j)^2 \neq 0$, which follows from Case 2 of (4.8). However, since we have perfect correlation with differing means, $X_i - X_j$ will always be constant $\forall i, j \in V$, and thus $\mathrm{Var}[\mathrm{TV}(\mathbf{X})] = 0$.

4.1.5 Gaussian quadratic forms and saddlepoint approximations

Quadratic forms appear in many engineering applications, and random quadratic forms are rich in terms of their applications as well as analyses. In addition to the fact that, if we take the airport delay vector $\mathbf{x}$ to be random, $\mathrm{TV}(\mathbf{x})$ is a random quadratic form, to motivate our exploration of these objects, we provide an example from stochastic control, with that of a stochastic linear-quadratic regulator:

Example 1 (Linear quadratic stochastic control) *Suppose we have the following finite horizon, discrete-time linear dynamical system, with dynamics*

$$\mathbf{x}(t+1) = A\mathbf{x}(t) + B\mathbf{u}(t) + \mathbf{w}(t), \qquad t = 0, \ldots, T-1, \tag{4.13}$$

where the initial condition $\mathbf{x}(0)$ *is drawn from* $\mathcal{N}(\mathbf{0}, \Sigma_{\mathbf{x}})$*, and the system is perturbed by additive noise* $\mathbf{w}(t) \overset{i.i.d.}{\sim} \mathcal{N}(\mathbf{0}, \Sigma_{\mathbf{w}})$*. The system matrix* A *may also be random, along with the matrix* B *modulating control inputs* $\mathbf{u}(t)$*, which are assumed to be deterministic. For stability, we set* $\max_i |\lambda_i(A)| = 1$*.*

The objective is to obtain a state-feedback controller $\mathbf{u}(t) \triangleq \psi_t(\mathbf{x}(t))$ *for all time steps, where* ψ_t *is the* control policy *at time* t*. We do this by solving the* linear quadratic stochastic control problem*, which seeks to minimize the cost, given by*

$$\mathcal{J} = \mathbb{E}\left[\mathbf{x}^\top(T)Q_f\mathbf{x}(T) + \sum_{t=0}^{T-1} \mathbf{x}^\top(t)Q\mathbf{x}(t) + \mathbf{u}^\top(t)R\mathbf{u}(t)\right]. \tag{4.14}$$

This problem can be solved via dynamic programming, and the optimal policies $\psi_t^\star(\mathbf{x}(t))$ *will have the form*

$$\psi_t^\star(\mathbf{x}(t)) = \arg\min_{\mathbf{v}} \left\{\mathbf{v}^\top R\mathbf{v} + \mathbb{E}\left[\min_{\psi_{t+1},\ldots,\psi_{T-1}} \mathcal{J}\right]\right\}. \tag{4.15}$$

These random quadratic forms appear in the cost function (4.14) for the linear quadratic stochastic control problem. The value functions that dictate the optimal policy are also quadratic forms of the state (random vector) and a symmetric (random) matrix that is a function of A, B, Q, and R (specifically, in the case of Example 1, this matrix takes the form $P_t = A^\top P_{t+1}A - A^\top P_{t+1}B\left(B^\top P_{t+1}B + R\right)^{-1}B^\top P_{t+1}A + Q$, and is known as the discrete-time algebraic Riccati equation). We first survey some results for the moments of Gaussian quadratic forms, and observe that the densities for general Gaussian quadratic forms tend to be "messy," in the sense that they can only be written as infinite expansions of other unwieldy densities. We then explore using saddlepoint approximations to obtain a closed-form density for Gaussian

quadratic forms in a limited setting.

Recall from Section 4.1.4 that we were interested in constructing bounds that could be used to detect outlying values of TV with respect to its mean. In order to construct these bounds, or leverage any concentration inequalities that might be useful, we require expressions for the moments of these Gaussian quadratic forms. The following theorem from [178] gives a way to evaluate the r^{th} moment of the Gaussian quadratic form:

Theorem 1 *For $\mathbf{x} \sim \mathcal{N}(\boldsymbol{\mu}, \Sigma)$ with valid $m \times m$ covariance matrix Σ, denote by $Q = \mathbf{x}^\top A\mathbf{x}$ where A is a $m \times m$ symmetric, real matrix. The r^{th} moment of Q is given by*

$$\mathbb{E}\left[Q^r\right] = \sum_{r_1=0}^{r-1} \binom{r-1}{r_1} g\left(r-1-r_1\right) \sum_{r_2=0}^{r_1-1} \binom{r_1-1}{r_2} g\left(r_1-1-r_2\right) \cdots \tag{4.16}$$

where $g(k) = 2^k k! \left(\operatorname{tr}(A\Sigma)^{k+1} + (k+1)\boldsymbol{\mu}^\top (A\Sigma)^k A\boldsymbol{\mu}\right)$ for $k \in \mathbb{N}_{\geq 0}$.

Proof. Theorem 3.2b.2 of [178]. □

From Theorem 1, the interesting moments of $Q = \mathbf{x}^\top A\mathbf{x}$ for our purposes, i.e., the mean and the variance, can be written explicitly:

$$\begin{aligned} \mathbb{E}\left[Q\right] &= \operatorname{tr}(A\Sigma) + \boldsymbol{\mu}^\top A\boldsymbol{\mu}, \\ \operatorname{Var}\left[Q\right] &= 2\operatorname{tr}(A\Sigma)^2 + 4\boldsymbol{\mu}^\top A\Sigma A\boldsymbol{\mu}. \end{aligned} \tag{4.17}$$

Unfortunately, unlike for the moments of $Q = \mathbf{x}^\top A\mathbf{x}$, no convenient, closed-form expression exists for a general Gaussian quadratic form when there are no assumptions on centrality and identity covariance. In particular, [178] provides several convergent expressions of the density $f_Q(q)$ for Q as infinite series expansions. These expansions can be written in terms of power series expansions, Laguerre series expansions, expansions in central χ^2 densities, confluent Hypergeometric functions, zonal polynomials, and densities of Gamma variates. As an example, we will detail the Laguerre series expansion.

Let $\mathbf{x} \in \mathbb{R}^{m \times 1}$ be a multivariate Gaussian random vector drawn from $\mathcal{N}(\boldsymbol{\mu}, \Sigma)$ with

possibly non-zero mean $\boldsymbol{\mu} \in \mathbb{R}^{m\times 1}$ and positive definite covariance matrix $\Sigma \in \mathbb{S}_{\succ 0}^{m\times m}$. We are interested in the distribution of the derived random variable $Q = \mathbf{x}^\top A\mathbf{x}$, where A is a $m \times m$ symmetric, real, and positive semi-definite matrix. We will refer to $Q = \mathbf{x}^\top A\mathbf{x}$ as a Gaussian quadratic form. From [178], the density of Q is surprisingly complicated, and can be written in terms of a series expansion using generalized Laguerre polynomials $L_k^{(\alpha)}(x)$, expressed via the Rodrigues formula as

$$L_k^{(\alpha)}(x) = \frac{1}{x!}e^x x^{-\alpha}\left[\frac{d^k}{dx^k}\left(e^{-x}x^{k+\alpha}\right)\right], \qquad \alpha > -1, \quad k = 0, 1, \ldots. \tag{4.18}$$

Specifically, by way of Theorem 4.2c.1 from [178], the density of Q using the Laguerre series expansion is

$$f_Q\left(\boldsymbol{\lambda}; \mathbf{b}; q\right) = \sum_{k=0}^{\infty} c_k \frac{k!}{2\beta\Gamma\left(\frac{m}{2}+k\right)}\left(\frac{q}{2\beta}\right)^{\frac{m}{2}-1} e^{-\frac{q}{2\beta}} L_k^{\left(\frac{m}{2}-1\right)}\left(\frac{q}{2\beta}\right), \tag{4.19}$$

for $q \in (0, \infty)$, where β is an arbitrary positive constant, c_k are power series expansion coefficients with $c_0 = 1$ and are dependent on $\boldsymbol{\lambda} = (\lambda_1, \ldots, \lambda_m)$ and $\mathbf{b} = (b_1, \ldots, b_m)$. The eigenvalues in $\boldsymbol{\lambda}$ belong to $\Sigma^{1/2}A\Sigma^{1/2}$ diagonalized by an orthogonal $m\times m$ matrix P, i.e., $P^\top\left(\Sigma^{1/2}A\Sigma^{1/2}\right)P = \text{diag}(\boldsymbol{\lambda})$. The vector $\mathbf{b}$ is dependent on the mean and covariance of $\mathbf{x}$, and is computed as $\mathbf{b}^\top = P^\top\Sigma^{-1/2}\boldsymbol{\mu}$.

Numerical experiments: Saddlepoint approximations

Saddlepoint approximation provides an accurate, closed-form expression of an unknown probability density (mass) function for a continuous (discrete) random variable, given that the mass generating function and cumulant generating function for the random variable are known [58, 39]. Compared to the three popular alternatives of approximations – enumerating exact probabilities, approximation via Gaussian densities, and brute force simulation with kernel density estimators – saddlepoint approximation oftentimes can overcome issues related to intractability, inaccuracy, and

convergence.

Let X be an univariate, continuous random variable supported on $\mathcal{X} \subseteq \mathbb{R}$ with an unknown probability density function $f_X(x)$ that is non-zero on $\mathcal{X}$. Let $M_X(s) = \mathbb{E}\left[e^{sX}\right]$ be the moment generating function of X, and $K_X(s) = \ln M_X(s)$ be the cumulant generating function of X. The saddlepoint approximation $\widehat{f}_X(x)$ to $f_X(x)$ is given by

$$\widehat{f}_X(x) = \left(2\pi \frac{d^2}{ds^2} K_X(\hat{s})\right)^{-1/2} \exp\left(K(\hat{s}) - \hat{s}x\right). \tag{4.20}$$

$\widehat{f}_X(x)$ is known as the *saddlepoint equation*, and $\hat{s} \triangleq \hat{s}(x)$ is the *saddlepoint associated with* x, where $\hat{s}(x)$ is the solution to the following differential equation in the cumulant generating function:

$$\frac{d}{ds} K(\hat{s}) = x. \tag{4.21}$$

We note that the saddlepoint equation serves as a meaningful approximation for $f_X(x)$ only for $x \in \mathcal{X}$, and that in general, the saddlepoint equation is not a true density, as in general we have that $\int_{\mathcal{X}} \widehat{f}_X(x)\, dx \neq 1$. However, if we set a normalizing constant $c = \int_{\mathcal{X}} \widehat{f}_X(x)\, dx$, then $\widetilde{f}_X(x) \triangleq c^{-1}\widehat{f}(x)$ is a valid density over the support $\mathcal{X}$. Finally, the discrete version of the saddlepoint approximation mirrors the continuous version, with the additional caveat that the (discrete) saddlepoint equation $\widehat{p}_X(x)$ is meaningful as an approximation to some probability mass function $p_X(x)$ only on integer-valued arguments.

To construct the saddlepoint approximation, we refer to the following theorem from [178] that provides the moment generating function for Gaussian quadratic forms that we are interested in.

Theorem 2 *Let A be a real, symmetric $m \times m$ matrix, and $\mathbf{x} \in \mathbb{R}^{m\times 1}$ with $\mathbf{x} \sim \mathcal{N}(\boldsymbol{\mu}, \Sigma)$. Then, the moment generating function $M_Q(s)$ of $Q = \mathbf{x}^\top A\mathbf{x}$ is*

$$M_Q(s) = \det\left(I - 2sA\Sigma\right)^{-1/2} \exp\left(-\frac{1}{2}\boldsymbol{\mu}^\top \left(I - (I - 2sA\Sigma)^{-1}\right)\Sigma^{-1}\boldsymbol{\mu}\right), \tag{4.22}$$

where I is the $m \times m$ identity. In the central case where $\boldsymbol{\mu} = \mathbf{0}$, the moment generating function reduces to

$$M_Q(s) = \det\left(I - 2sA\Sigma\right)^{-1/2}. \tag{4.23}$$

Furthermore, $M_Q(s)$ can be written in a scalar form involving the eigenvalues of $\Sigma^{1/2}A\Sigma^{1/2}$ and constants that depend on the mean $\boldsymbol{\mu}$ and precision matrix Σ^{-1}. Specifically, let $\lambda_1, \ldots, \lambda_m$ be eigenvalues of $\Sigma^{1/2}A\Sigma^{1/2}$, and define the vector of constants $\mathbf{b} = (b_1, \ldots, b_m)^\top \triangleq P^\top \Sigma^{-1/2}\boldsymbol{\mu}$, where P is any $m \times m$ orthogonal matrix that diagonalizes $\Sigma^{1/2}A\Sigma^{1/2}$. Then, $M_Q(s)$ can be rewritten as follows:

$$M_Q(s) = \begin{cases} \exp\left(s\sum_{j=1}^{m} \dfrac{b_j^2\lambda_j}{1-2s\lambda_j}\right)\prod_{j=1}^{m}\left(1-2s\lambda_j\right)^{-\frac{1}{2}}, & \text{if } \boldsymbol{\mu} \neq \mathbf{0}, \\ \prod_{j=1}^{m}\left(1-2s\lambda_j\right)^{-\frac{1}{2}}, & \text{if } \boldsymbol{\mu} = \mathbf{0}. \end{cases} \tag{4.24}$$

Proof. Theorems 3.2a.1, 3.2a.2, and Corollary 3.2a.1 of [178]. □

We note here that for Theorem 2, the A matrix is only required to be real and symmetric, e.g., the graph Laplacian. In order to deploy saddlepoint approximation, we first compute the cumulant generating function, as well as its first and second derivatives with respect to s. Starting with the simpler case of $\boldsymbol{\mu} = \mathbf{0}$, we have that

$$\begin{aligned} K_Q(s) &= \ln\left(\prod_{j=1}^{m}\left(1-2s\lambda_j\right)^{-\frac{1}{2}}\right) = \sum_{j=1}^{m}\ln\left(\left(1-2s\lambda_j\right)^{-\frac{1}{2}}\right), \\ \frac{d}{ds}K_Q(s) &= \sum_{j=1}^{m}\frac{\lambda_j}{1-2s\lambda_j}, \\ \frac{d^2}{ds^2}K_Q(s) &= \sum_{j=1}^{m}\frac{2\lambda_j^2}{\left(1-2s\lambda_j\right)^2}. \end{aligned} \tag{4.25}$$

For the non-central case where $\boldsymbol{\mu} \neq \mathbf{0}$, the cumulant generating function and its first two s derivatives are

$$
\begin{aligned}
K_Q(s) &= \sum_{j=1}^{m} \left(\frac{b_j^2 \lambda_j s}{1 - 2s\lambda_j} + \ln\left((1 - 2s\lambda_j)^{-\frac{1}{2}} \right) \right), \\
\frac{d}{ds} K_Q(s) &= \sum_{j=1}^{m} \frac{\lambda_j \left(1 + b_j^2 - 2s\lambda_j \right)}{(1 - 2s\lambda_j)^2}, \\
\frac{d^2}{ds^2} K_Q(s) &= 2 \sum_{j=1}^{m} \frac{\lambda_j^2 \left(-1 - 2b_j^2 + 2s\lambda_j \right)}{(-1 + 2s\lambda_j)^3}.
\end{aligned}
\tag{4.26}
$$

The difficulty with saddlepoint approximations arises from the fact that solving for the saddlepoint requires one to solve successively higher-degree polynomials. For the central case, the saddlepoint for the case of $m = 1$ is

$$
\hat{s} = \frac{1}{2} \left(\frac{1}{\lambda_1} - \frac{1}{x} \right), \tag{4.27}
$$

and for $m = 2$, we have the saddlepoints given by:

$$
\hat{s} = \frac{1}{4\lambda_1 \lambda_2 x} \left(\pm \sqrt{\lambda_1^2 x^2 - 2\lambda_1 \lambda_2 x^2 + \lambda_2^2 \left(4\lambda_1^2 + x^2 \right)} + \lambda_2 x + \lambda_1 \left(x - 2\lambda_2 \right) \right). \tag{4.28}
$$

For the central case with $m = 3$, a "solution" is technically obtainable (e.g., via the `Reduce` function in `Mathematica` specifying solutions to cubic polynomials), but it is not convenient for computation. For the non-central case, the one-node (i.e., $m = 1$) case has saddlepoints given by

$$
\hat{s} = \frac{\pm \sqrt{4b_1^2 \lambda_1^3 x + \lambda_1^4} - \lambda_1^2 + 2\lambda_1 x}{4\lambda_1^2 x}, \tag{4.29}
$$

with larger values of m being analogously difficult to obtain.

We plot in Figure 4-2 the saddlepoint approximation of the Gaussian quadratic form $Q = \mathbf{x}^\top A \mathbf{x}$ for a 2-node graph in the central case, and observe that the (non-normalize) saddlepoint approximation seems to follow the distribution of Q quite well. The approximation is close for both an identity covariance matrix, as well as for any arbitrary valid covariance matrix. We also explored a randomized version of this approximation by repeatedly sampling covariance matrices (e.g., via a Wishart distri-

bution), and fitting a single saddle approximation using the average of the eigenvalues of all sampled covariance matrices. This fit can be seen in Figure 4-3.

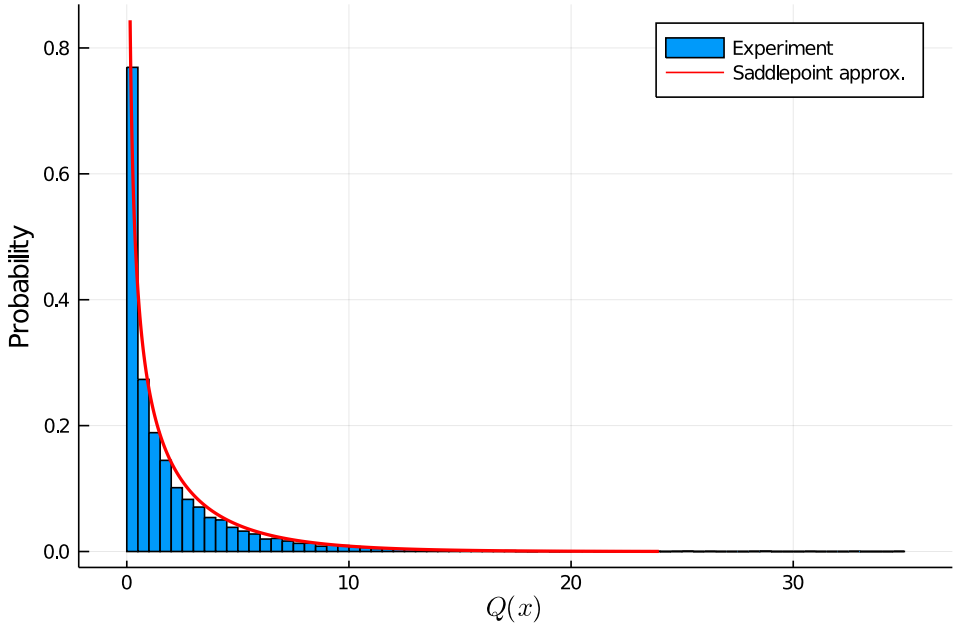


Figure 4-2: Saddlepoint approximation to Q in the central case for $m = 2$.

4.1.6 Empirical bounds for strong outliers in distribution

The final bounds that we evaluate involve outliers in distribution conditioned on the magnitude of the observed graph signal vector. This mitigates the interdependence between $\|\mathbf{x}\|$ and $\mathrm{TV}(\mathbf{x})$ by conditioning the expectation and variance of $\mathrm{TV}(\mathbf{X})$ on the magnitude. While we would like to analytically evaluate the full 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. From Section 4.1.5, we see that in general $\mathrm{TV}(\mathbf{x})$ is a very complicated distribution, even with Gaussian assumptions on $\mathbf{x}$. Instead, we propose the following simulation-based method (Algorithm 1) that uses an interval-based estimation scheme to empirically evaluate the two conditional random variables over only $\delta < \infty$

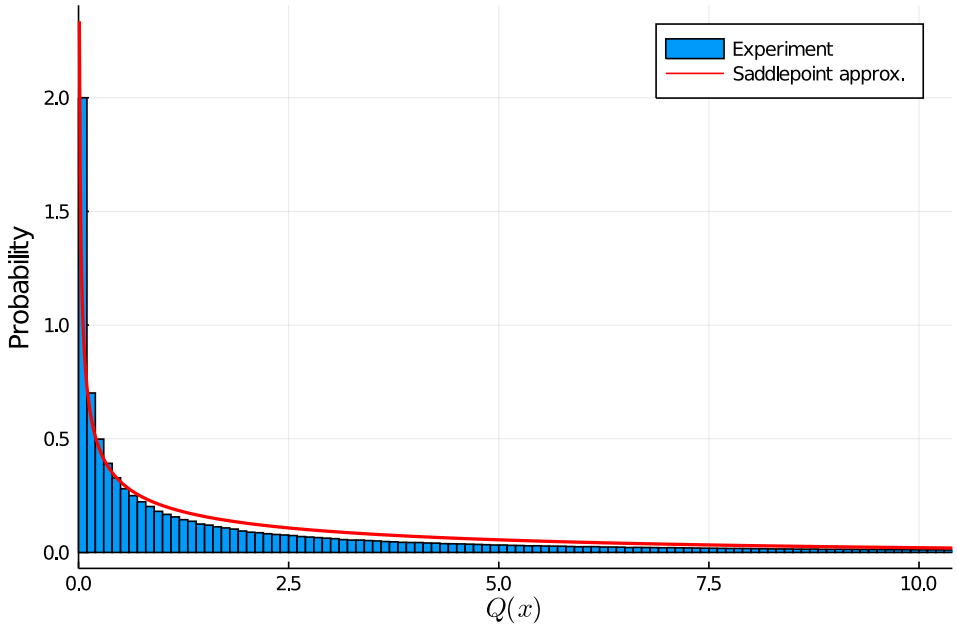


Figure 4-3: Saddlepoint approximation to Q in the central case for $m = 2$, with randomized covariance matrices.

pre-specified disjoint covers $\dot{\bigcup}_{\ell=1}^{\delta} \mathcal{U}_\ell \equiv [\min\{\|\mathbf{x}\|\}, \max\{\|\mathbf{x}\|\}] \subset \mathbb{R}_{\geq 0}$. Figure 4-4 provides an illustration of the empirical bound computation process used in Algorithm 1.

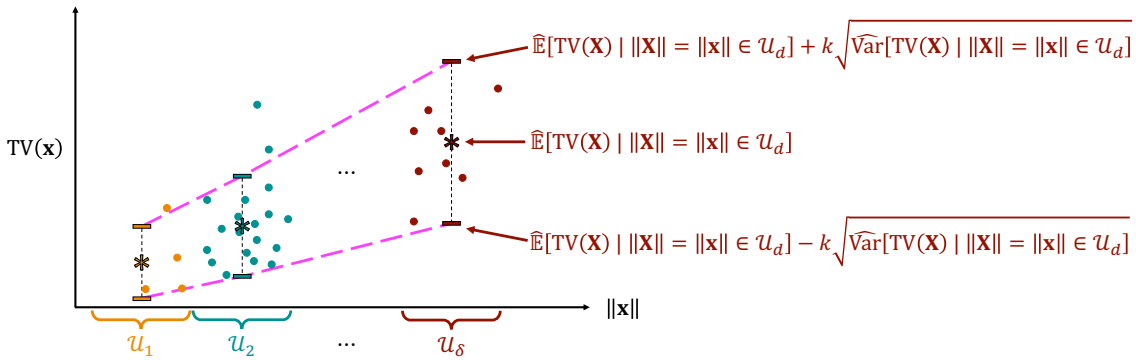


Figure 4-4: Empirical strong outliers in distribution bound generating process given simulated observations $(\|\mathbf{x}\|, \mathrm{TV}(\mathbf{x}))$ generated from $\left\{\hat{\boldsymbol{\mu}}, \hat{\Sigma}, \mathbf{C}\right\}_{\mathcal{O}_M}$ partitioned via $\dot{\bigcup}_{\ell=1}^{\delta} \mathcal{U}_\ell \equiv [\min\{\|\mathbf{x}\|\}, \max\{\|\mathbf{x}\|\}] \subset \mathbb{R}_{\geq 0}$. This approximates the magenta bounds shown in Figure 4-1. *Reprinted from [158].*

It is important to note that we can use any type of distribution in Algorithm 1,

Algorithm 1 Computing empirical bounds for strong outliers in distribution.

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

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

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

  $\mathbf{x} \leftarrow \max\{\mathbf{x}, \mathbf{0}\}$

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

end

$\widetilde{\Delta} \leftarrow \delta^{-1}\left(\max\left\{\boldsymbol{\mathcal{V}}_{\|\mathbf{x}\|}\right\} - \min\left\{\boldsymbol{\mathcal{V}}_{\|\mathbf{x}\|}\right\}\right)$

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

for *Interval ℓ of 1:δ* **do**

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

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

end

as long as there are sufficient observations in $\mathcal{O}_M$ to estimate $\hat{\boldsymbol{\mu}}, \hat{\Sigma}$, and $\mathbf{C}$ reliably. We now give a definition for empirical strong outliers in distribution:

Definition 8 (Empirical version of Definition 7) *An observation* $\mathbf{x}$ *is considered to be an* empirical strong outlier in distribution of level k *if* $\mathrm{TV}(\mathbf{x}) \notin \left[\widehat{\mathcal{A}_\ell}, \widehat{\mathcal{B}_\ell}\right]$, *where:*

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

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

4.1.7 Bounds for weak outliers in distribution: Partial information case

In Section 4.1.4, 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 discuss two such scenarios. In the first example, 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 correlation values.

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 $\Sigma \in \mathbb{R}^{N \times N}_{\succeq 0}$ (or equivalently the correlation matrix $\mathbf{C}$), but the precise values of ρ_{ij} are 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 on the correlation coefficients $\rho_{ij} \in (\nu_{ij}, \epsilon_{ij}) \subseteq [0, 1]$, 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 would 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 $N_\rho^{N^2-N}$ 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. 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 Figure 4-5 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, which is the setting for Figure 4-5.

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

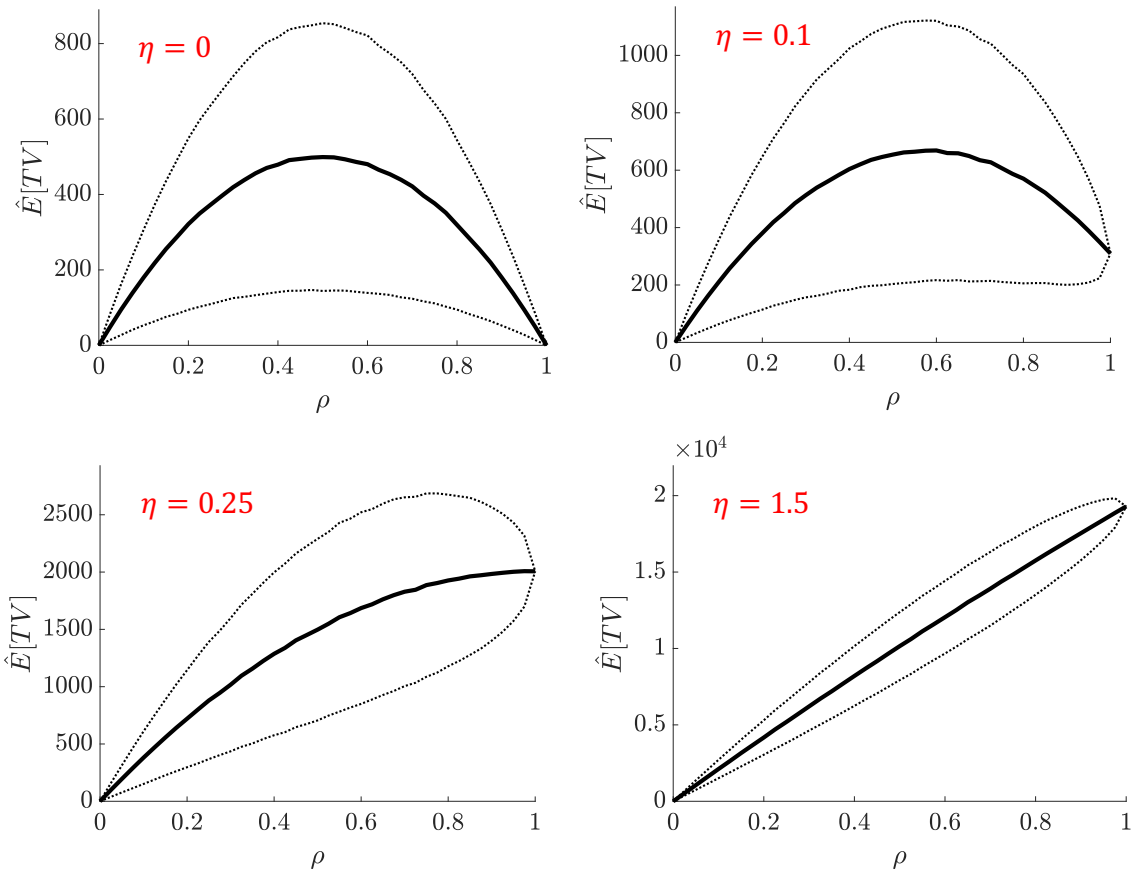


Figure 4-5: Empirically-derived curves for the expectation and variance of TV as a function of correlation, parameterized by $\boldsymbol{\mu}$ via η. We draw $M = 5 \times 10^4$ data points from a multivariate Gaussian distribution with $N = 5, \sigma_i = 10, \ \forall i$, and $\rho_{ij} = \rho, \ \forall i, j$. We set $\eta \in \{0, 0.1, 0.25, 1.5\}$ and $\mu_i \overset{\text{i.i.d.}}{\sim} 100(1-\eta)+200\eta X_U$, where $X_U \overset{\text{i.i.d.}}{\sim} \text{Unif}(0, 1)$. Higher η indicates that signals have higher baseline difference in terms of magnitudes across a pair of nodes. *Reprinted from [100].* © *2019 IEEE*

to the computationally prohibitive exploration of the search space $\rho_{ij} \in (\nu_{ij}, \epsilon_{ij}) \subseteq [0, 1]$. Furthermore, there is no reliance on estimation intervals and discretizations. Specifically, our two propositions quantify the change in $\mathbb{E}[\mathrm{TV}(\mathbf{X})]$ (Proposition 4) and $\mathrm{Var}[\mathrm{TV}(\mathbf{X})]$ (5) due to the uncertainty in ρ_{ij}.

For the propositions we construct and prove in this section, we require all the correlation coefficients to have the same sign, i.e., all $\rho_{ij} \geq 0$ or all $\rho_{ij} \leq 0,\ \forall i, j \in V$. We consider the former, and introduce the following projections of the correlation coefficients into the non-negative half-plane: $\rho_{ij}^{+} = \max\{0, \rho_{ij}\}$ and $r_{ij|\mathcal{O}_M}^{+} = \max\left\{0, r_{ij|\mathcal{O}_M}\right\}$. A similar projection can be defined for non-positive correlations, and all results follow analogously.

In accordance with the above projection operations, we will assign a projected weight $r_{ij|\mathcal{O}_M}^{+}$ for every edge (i, j). We note that the projected correlation coefficient $r_{ij|\mathcal{O}_M}^{+}$ turns out to be a consistent estimator of the underlying ρ_{ij}^{+}, due to the fact that the projection operation above (non-negative half-plane projection) can be rewritten in terms of piece-wise affine transformations [240]. One caveat is that we must take into account an offset, since this estimator will be biased: The offset derived in [240] can be written as

$$\mathbb{E}[r_{ij|\mathcal{O}_M}] = \rho_{ij}\left(1 - \frac{1 - \rho_{ij}^2}{2M} + O\left(\frac{1}{M^2}\right)\right). \tag{4.30}$$

We note that the expectation on $r_{ij|\mathcal{O}_M}$ is not constant, as $r_{ij|\mathcal{O}_M}$ is itself a random variable [89, 203]. However, the density of $r_{ij|\mathcal{O}_M}$ is only dependent on M. We will make use of this observation in the following proofs for propositions related to this case of partial information of correlations.

Given the above discussion regarding projections of correlation coefficients and consistent estimators, we can examine the definition of total variation for an *unobserved* graph signal vector $\mathbf{X} \sim \mathcal{N}(\boldsymbol{\mu}, \Sigma)$. In particular, the total variation will be taken with respect to graph Laplacian $\mathfrak{L} \in \mathbb{S}^{N \times N}$ corresponding to the correlation network estimated from $\mathcal{O}_M$. Now, we have that the total variation $\mathrm{TV}(\mathbf{X})$ is a derived random variable. Specifically,

$$\mathrm{TV}(\mathbf{X}) = \frac{1}{2}\sum_{i\neq j}\left\{r^{+}_{ij|\mathcal{O}_M}(X_i - X_j)^2\right\}. \tag{4.31}$$

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

$$\sum_{i=1}^{N}\sigma_i^2 + \sum_{i\neq j}\nu_{ij}\sigma_i\sigma_j < \mathrm{Var}\left[\|\mathbf{X}\|\right] = \mathrm{Var}\left[\sum_{i=1}^{N}X_i\right] < \sum_{i=1}^{N}\sigma_i^2 + \sum_{i\neq j}\epsilon_{ij}\sigma_i\sigma_j. \tag{4.32}$$

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 $\mathrm{Var}\left[\|\mathbf{X}\|\right] = \mathbf{1}^\top\Sigma\mathbf{1} + \sum_{i\neq j}\rho_{ij}\sigma_i\sigma_j$ 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* $0 \leq \nu_{ij} < \rho^{+}_{ij} < \varepsilon_{ij} \leq 1$ *for all unique pairs of nodes* $i, j \in V$*. Then, we can evaluate scalars* δ_1 *and* δ_2*, with* $\delta_2 \geq 0$*, such that* $\max\{0, \delta_1\} \leq \mathbb{E}[\mathrm{TV}(\mathbf{X})] < \delta_2$*.*

Proof. See Appendix A.3. □

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

Proof. See Appendix A.4. □

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*

$$\mathrm{TV}(\mathbf{x}) \notin \left[\max\left\{0, \delta_1 - k\sqrt{\delta_4}\right\},\ \delta_2 + k\sqrt{\delta_4}\right],\ \textit{for some } k \geq 0. \tag{4.33}$$

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. Let us denote by $\lambda_{\max}$ the largest eigenvalue of $\mathfrak{L}$. Then, we have the following bound via the Rayleigh quotient:

$$\mathrm{TV}(\mathbf{x}) = \mathbf{x}^\top \mathfrak{L}\mathbf{x} \leq \lambda_{\max} \|\mathbf{x}\|_2^2 \leq \lambda_{\max} \|\mathbf{x}\|_1^2 . \tag{4.34}$$

The second inequality comes from the standard fact that $\|\mathbf{x}\|_2 \leq \|\mathbf{x}\|_1$, and that $\lambda_{\max}$ is non-negative in our settings. 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.

4.2 Strong and weak bounds on TV in simulated networks

We now compute the bounds on strong distribution outliers using simulations, and compare them against the theoretically-derived weak distribution outlier bounds. We observe that the difference between these two bounds depends on the underlying $\boldsymbol{\mu}$ and Σ of the data, and we show two examples to highlight that dependency.

Using two simulations, we examine the performance gap between bounds on strong outliers in distribution versus bounds on weak outliers in distribution. We also demon-

strate the utility of distinguishing outliers using TV rather than the underlying distribution, as the former can provide much more useful interpretations, particularly in a networked setting. The idea for both simulations is to assume a fixed mean vector $\boldsymbol{\mu}$ and covariance matrix Σ, then generate M normally-distributed observations. We fix the number of generated observations to be $M = 1 \times 10^6$, a large value. The data points are plotted on a $\text{TV}(\mathbf{x})$ versus $\|\mathbf{x}\|$ plot, and the quantities $\hat{\mathbb{E}}\left[\text{TV}(\mathbf{X}) \mid \|\mathbf{X}\| = \|\mathbf{x}\|\right]$ and $\widehat{\text{Var}}\left[\text{TV}(\mathbf{X}) \mid \|\mathbf{X}\| = \|\mathbf{x}\|\right]$ required for obtaining the bounds on strong outliers in distribution are computed empirically by binning $\|\mathbf{x}\|$ and conditioning on each bin. The theoretically-derived bounds for weak distribution outliers and scale outliers are also plotted. Additionally, we color each observation by the density obtained from evaluating the probability density function (multivariate Gaussian) at the value of the realized $\mathbf{x}$. Note that if only the underlying distribution was used for outlier detection, all black-colored trials could be considered as outliers. This stands in contrast to the bounds provided by our derivations.

For parameters, we select $N = 2$, $\sigma_1 = \sigma_2 = 1$, $\rho_{12} = 0.5$, and mean vector $\boldsymbol{\mu} = (545.34,\, 582.13)^\top$ to use in the first simulation. We plot the trials and bounds in terms of $(\|\mathbf{x}\|, \text{TV}(\mathbf{x}))$ in Figure 4-6. If we examine the weak and strong outlier in distribution bounds, we see that they are quite well-aligned. In this simple 2-node setting, it may be the case that the weak outliers in distribution bounds are all that is needed, and serve as reasonable approximations for strong outlier in distribution bounds.

In the next simulation, we will show that the weak and strong bounds for outliers in distribution can deviate significantly in more complicated settings. To motivate the second simulation, we will increase the number of nodes from $N = 2$ to $N = 30$, and examine the US airport correlation network which we detail in Section 4.4.1. In lieu of selecting artificial parameters, we estimate the sample covariance matrix Σ from a set of historical airport delay data, and we also estimate a sample mean vector $\boldsymbol{\mu}$. We refer readers to Section 4.4.1 for a detailed description of the data set. We again plot the trials and bounds in terms of $(\|\mathbf{x}\|, \text{TV}(\mathbf{x}))$ for this airport delay network setting in Figure 4-7.

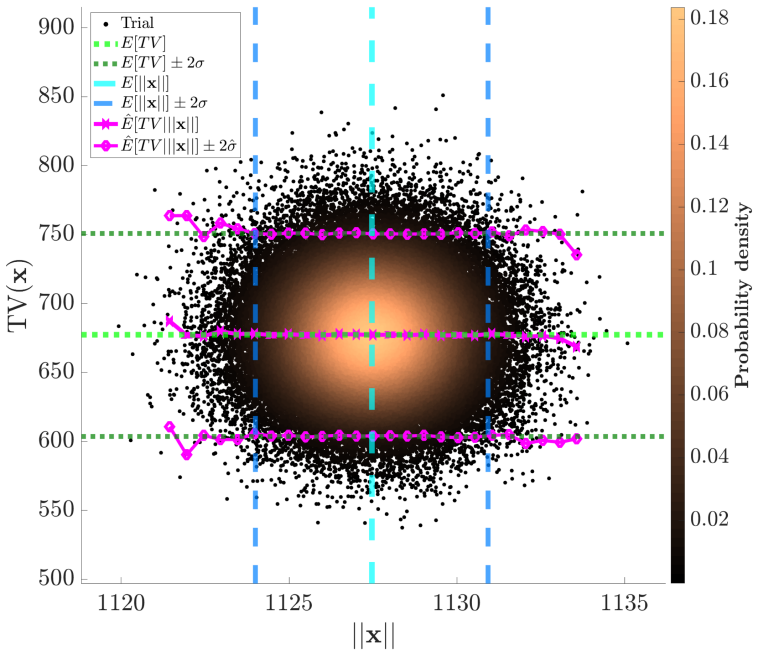


Figure 4-6: TV versus 1-norm of the graph signal for a generic bi-vertex graph, with scale outlier, weak outlier in distribution, and empirically-derived strong outlier bounds. Each observation $(\|\mathbf{x}\|, \mathrm{TV}(\mathbf{x}))$ is colored by its probability density $f_{\mathbf{X}}$. *Reprinted from [100]. © 2019 IEEE*

From the simulation results shown in Figure 4-7, unlike the 2-node case in the previous simulation, the weak and strong bounds do not agree. In fact, the nature of the disagreement between the two bounds changes depending on the value of $\|\mathbf{x}\|$. Specifically, for smaller values of $\|\mathbf{x}\|$, we see that the weak bounds are higher than the strong bounds, indicating that the strong bounds are tighter, and the weak bounds are overly-liberal. As the magnitude of the signal $\|\mathbf{x}\|$ increases, as expected, the weak bounds do not capture the natural increase in TV as $\|\mathbf{x}\|$ increases. Thus, the weak bounds are now overly-conservative, and would mis-classify graph signal observations as outliers. The strong bounds, conditioning on $\|\mathbf{x}\|$, adjust for the increase in $\|\mathbf{x}\|$.

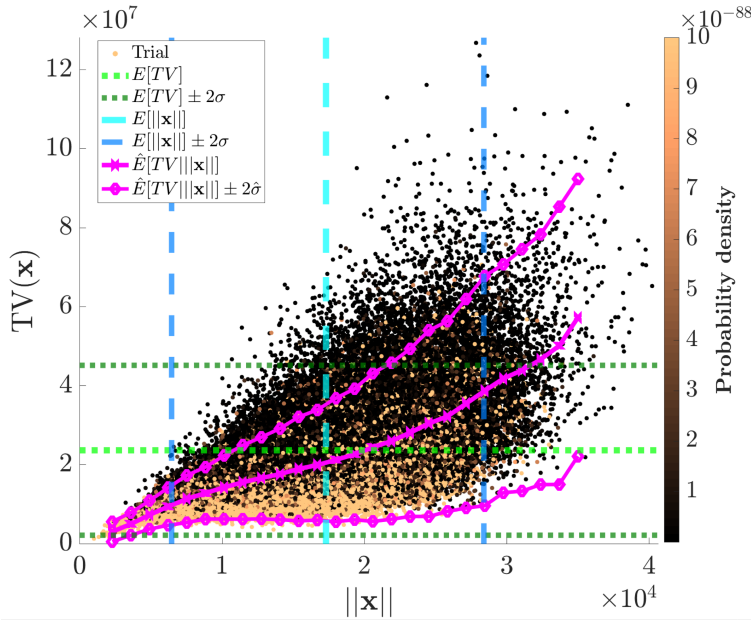


Figure 4-7: TV versus 1-norm of simulated graph signals within a 30-vertex graph; data set from the US air transportation network. *Reprinted from [100].* © *2019 IEEE*

4.3 Non-parametric identification of graph signal outliers

In Section 4.1, we typically assume that the graph signals are generated from a parametric underlying distribution with a well-defined mean and variance. In particular,

we assume that the underlying distribution is a multivariate Gaussian distribution with mean vector $\boldsymbol{\mu}$ and covariance matrix Σ. In lieu of parametric approaches, we will use the interquartile range (IQR) of the data to define desired outlier bounds in a non-parametric manner. Furthermore, we note that our previous assumptions of a Gaussian distribution implicitly assume also that our TD and TV distributions are symmetric: However, given that both TD and TV are non-negative quantities in our airport delay network applications, this symmetry assumption may be too specific. Thus, we will also explicitly take into consideration skewness in our non-parametric bound generation process by adjusting the resultant bounds based on the *medcouple* statistic, as proposed in [33, 117].

Suppose that we are given a univariate sample set $\{y_1, y_2, \ldots, y_K\}$. We denote by Q_1 the value of the first quartile of $\{y_1, y_2, \ldots, y_K\}$. Similarly, we denote by Q_3 the value of the third quartile of $\{y_1, y_2, \ldots, y_K\}$. The median of $\{y_1, y_2, \ldots, y_K\}$ we will denote by Q_2. The IQR for this data set is given by $IQR = Q_3 - Q_1$. We now have all of the summary statistics required to define the medcouple statistic:

$$MC = \underset{y_i \leq Q_2 \leq y_j}{\text{med}} h(y_i, y_j). \tag{4.35}$$

Observe that for the medcouple statistic, it takes a median conditioned on $y_i \leq Q_2 \leq y_j$, and for all well-ordered $y_i \neq y_j$, the kernel function $h(\cdot, \cdot)$ is given explicitly by:

$$h(y_i, y_j) = \frac{(y_j - Q_2) - (Q_2 - y_i)}{y_j - y_i}. \tag{4.36}$$

This medcouple statistic, as proposed by [33, 117], is a robust measure of the skewness exhibited by a univariate data set $\{y_1, y_2, \ldots, y_K\}$. We will leverage the medcouple statistic to adjust the IQR, with the goal of constructing outlier bounds that are better suited for asymmetric, skewed distributions. We note in our analysis that all TD and TV distributions are right-skewed, i.e., we have that $MC > 0$. This is the expected direction of the skew, since TD and TV are non-negative quantities for our airport delay network application. Since the adjusted box-plot formulas from [117] differ for $MC > 0$ and $MC < 0$, we use only the adjusted box-plot formulas

for the right-skewed case $MC > 0$. We refer readers to [117] for the left-skewed case. As with before, we restrict ourselves to the 1-norm of graph signals when we talk about the TD at an airport. Furthermore, we assume non-negative signals at each node, which is reasonable since airport delays are always a non-negative quantity. Thus, we again have that $\|\mathbf{x}\| = \|\mathbf{x}\|_1 = \sum_i x_i$. We now re-define the various outlier definitions (i.e., in-scale, in-distribution, both weak and strong versions) with respect to the skew-adjusted IQR. Figure 4-8 summarizes the three different non-parametric skew-adjusted outlier bounds from Definitions 10, 11, and 12.

Definition 10 *A data point* $\mathbf{x}$ *is classified as an* outlier in scale (OIS) *if*

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

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

$$\boldsymbol{\mathcal{V}}_{\|\mathbf{x}\|} = \left\{\mathcal{V}_{\|\mathbf{x}\|,1} = \left\|\mathbf{x}^{(1)}\right\|, \ldots, \mathcal{V}_{\|\mathbf{x}\|,M} = \left\|\mathbf{x}^{(M)}\right\|\right\}. \tag{4.37}$$

Definition 11 *A data point* $\mathbf{x}$ *is classified as a* weak outlier in distribution (weak OID) *if*

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

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

$$\boldsymbol{\mathcal{V}}_{\mathrm{TV}(\mathbf{x})} = \left\{\mathcal{V}_{\mathrm{TV}(\mathbf{x}),1} = {\mathbf{x}^{(1)}}^{\top}\mathcal{L}\mathbf{x}^{(1)}, \ldots, \mathcal{V}_{\mathrm{TV}(\mathbf{x}),M} = {\mathbf{x}^{(M)}}^{\top}\mathcal{L}\mathbf{x}^{(M)}\right\}. \tag{4.38}$$

Definition 12 *A data point* $\mathbf{x}$ *is classified as a* strong outlier in distribution (strong OID) *if*

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

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

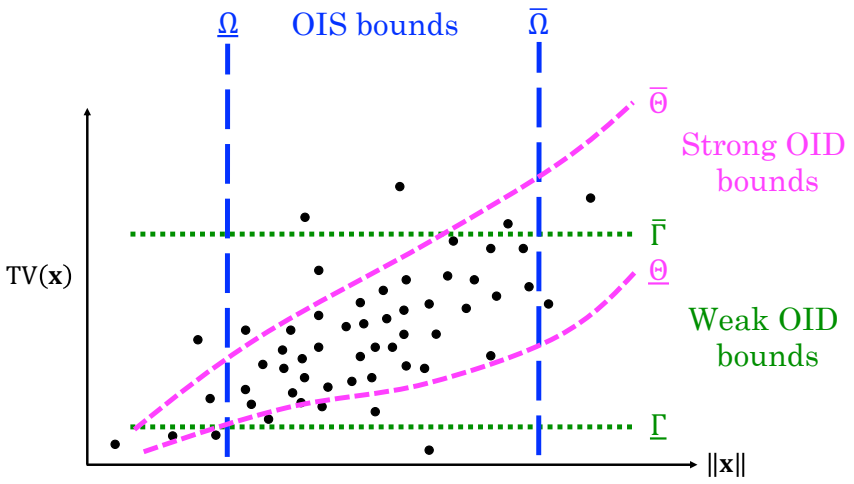


Figure 4-8: Depiction of the non-parametric, skew-adjusted IQR-based outlier detection bounds. *Reprinted from [161].* © *2020 IEEE*

In practice, due to a finite number of data points in $\mathcal{O}_M$, it is not possible to condition exactly on $\|\mathbf{x}\|$. Analogous to Section 4.1, we compute a relaxed, discretized bound instead. Algorithm 2 describes the steps required to compute the discretized non-parametric bounds $\left\{\overline{\widehat{\Theta}}, \underline{\widehat{\Theta}}\right\}$ for strong outliers in distribution. The idea behind Algorithm 2 is to condition the TV on a discrete interval around $\|\mathbf{x}\|$, rather than a particular value for the 1-norm. To obtain robust estimates, we dynamically vary the interval widths to ensure an equal number of data points n in each interval, with the possible exception of the last bin. The TV of each data observation within the (sub)set that falls into a particular interval is used to compute the IQR and MC, and then the outlier bounds for that particular interval. Finally, the output from Algorithm 2 is linearly interpolated for points between the discrete bins, and extrapolated at

Algorithm 2 Computing non-parametric strong OID bounds (IQR).

Input: Minimum bin size n; TD set $\boldsymbol{\mathcal{V}}_{\|\mathbf{x}\|}$; TV set $\boldsymbol{\mathcal{V}}_{\mathrm{TV}(\mathbf{x})}$

Output: Outlier bound bins $\widetilde{\mathbf{LB}}_{\|\mathbf{x}\|}$; Upper outlier bound $\overline{\Theta}$; Lower outlier bound $\underline{\Theta}$

$\widetilde{\boldsymbol{\mathcal{V}}}_{\|\mathbf{x}\|} \leftarrow$ Sort $\boldsymbol{\mathcal{V}}_{\|\mathbf{x}\|}$ s.t. $\mathcal{V}_{\|\mathbf{x}\|,i} \leq \mathcal{V}_{\|\mathbf{x}\|,j}, \forall i < j$ and $i, j \in \{1, \ldots, M\} \times \{1, \ldots, M\}$

$\mathbf{LB}_{\|\mathbf{x}\|} \leftarrow \left\{ \widetilde{\mathcal{V}}_{\|\mathbf{x}\|,i} \,|\, i \in \left\{ 1, \ldots, \min\left\{ M, 1 + \left\lfloor \frac{M}{n} \right\rfloor + n \right\} \right\} \right\}$

$\Delta_{\max} \leftarrow \max_i \left\{ \widetilde{\mathcal{V}}_{\|\mathbf{x}\|,i+1} - \widetilde{\mathcal{V}}_{\|\mathbf{x}\|,i} \right\}$

for *Lower bound index* $i_{\mathbf{LB}} = 1 : \left|\mathbf{LB}_{\|\mathbf{x}\|}\right|$ **do**

  if $i_{\mathbf{LB}} < \left|\mathbf{LB}_{\|\mathbf{x}\|}\right|$ **then**

    $\ell_b \leftarrow \mathbf{LB}_{\|\mathbf{x}\|,i_{\mathbf{LB}}}$; $\ell_u \leftarrow \mathbf{LB}_{\|\mathbf{x}\|,i_{\mathbf{LB}}+1}$

    $\boldsymbol{\mathcal{V}}^{i_{\mathbf{LB}}}_{\mathrm{TV}(\mathbf{x})} \leftarrow \left\{ \mathcal{V}_{\mathrm{TV}(\mathbf{x}),i} \,|\, \ell_b \leq \mathcal{V}_{\mathrm{TV}(\mathbf{x}),i} < \ell_u \right\}$

    $\widetilde{\mathbf{LB}}_{\|\mathbf{x}\|,i_{\mathbf{LB}}} \leftarrow \frac{1}{2}\left(\mathbf{LB}_{\|\mathbf{x}\|,i_{\mathbf{LB}}} + \mathbf{LB}_{\|\mathbf{x}\|,i_{\mathbf{LB}}+1} \right)$

  else

    $\ell_b \leftarrow \mathbf{LB}_{\|\mathbf{x}\|,i_{\mathbf{LB}}}$

    $\boldsymbol{\mathcal{V}}^{i_{\mathbf{LB}}}_{\mathrm{TV}(\mathbf{x})} \leftarrow \left\{ \mathcal{V}_{\mathrm{TV}(\mathbf{x}),i} \,|\, \ell_b \leq \mathcal{V}_{\mathrm{TV}(\mathbf{x}),i} \right\}$

    $\widetilde{\mathbf{LB}}_{\|\mathbf{x}\|,i_{\mathbf{LB}}} \leftarrow \frac{1}{2}\left(\mathbf{LB}_{\|\mathbf{x}\|,i_{\mathbf{LB}}} + \max\left\{ \boldsymbol{\mathcal{V}}_{\|\mathbf{x}\|} \right\} \right)$

  end

  $\left\{ Q_1^{i_{\mathbf{LB}}}, Q_3^{i_{\mathbf{LB}}} \right\} \leftarrow \left\{ Q_1 \text{ of } \boldsymbol{\mathcal{V}}^{i_{\mathbf{LB}}}_{\mathrm{TV}(\mathbf{x})}, Q_3 \text{ of } \boldsymbol{\mathcal{V}}^{i_{\mathbf{LB}}}_{\mathrm{TV}(\mathbf{x})} \right\}$

  $IQR^{i_{\mathbf{LB}}} \leftarrow Q_3^{i_{\mathbf{LB}}} - Q_1^{i_{\mathbf{LB}}}$

  $MC^{i_{\mathbf{LB}}} \leftarrow$ compute medcouple for $\boldsymbol{\mathcal{V}}^{i_{\mathbf{LB}}}_{\mathrm{TV}(\mathbf{x})}$

  if $MC^{i_{\mathbf{LB}}} \geq 0$ **then**

    $\overline{\theta}^{i_{\mathbf{LB}}} \leftarrow Q_3^{i_{\mathbf{LB}}} + 1.5\left(IQR^{i_{\mathbf{LB}}}\right) \exp\left(3 \times MC^{i_{\mathbf{LB}}}\right)$

    $\underline{\theta}^{i_{\mathbf{LB}}} \leftarrow Q_1^{i_{\mathbf{LB}}} - 1.5\left(IQR^{i_{\mathbf{LB}}}\right) \exp\left(-4 \times MC^{i_{\mathbf{LB}}}\right)$

  else

    $\overline{\theta}^{i_{\mathbf{LB}}} \leftarrow Q_3^{i_{\mathbf{LB}}} + 1.5\left(IQR^{i_{\mathbf{LB}}}\right) \exp\left(4 \times MC^{i_{\mathbf{LB}}}\right)$

    $\underline{\theta}^{i_{\mathbf{LB}}} \leftarrow Q_1^{i_{\mathbf{LB}}} - 1.5\left(IQR^{i_{\mathbf{LB}}}\right) \exp\left(-3 \times MC^{i_{\mathbf{LB}}}\right)$

  end

end

$\widetilde{\mathbf{LB}}_{\|\mathbf{x}\|} \leftarrow \left\{ \widetilde{\mathbf{LB}}_{\|\mathbf{x}\|,1}, \ldots, \widetilde{\mathbf{LB}}_{\|\mathbf{x}\|,\left|\mathbf{LB}_{\|\mathbf{x}\|}\right|} \right\}$

$\overline{\widehat{\Theta}} \leftarrow \left\{ \overline{\theta}^1, \ldots, \overline{\theta}^{\left|\mathbf{LB}_{\|\mathbf{x}\|}\right|} \right\}$; $\underline{\widehat{\Theta}} \leftarrow \left\{ \underline{\theta}^1, \ldots, \underline{\theta}^{\left|\mathbf{LB}_{\|\mathbf{x}\|}\right|} \right\}$

the edges up to the boundary data points. This process is illustrated in Figure 4-9, where the strong OID bounds $\left\{\widetilde{\mathbf{LB}}_{\|\mathbf{x}\|}, \left\{\overline{\widehat{\Theta}}, \underline{\widehat{\Theta}}\right\}\right\}$ are retrieved from Algorithm 2. The discretized estimates of the strong OID bounds converge to the actual bounds as the number of data points increases and the maximum bin width $\Delta_{\max}$ (see Algorithm 2) decreases, i.e.,

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

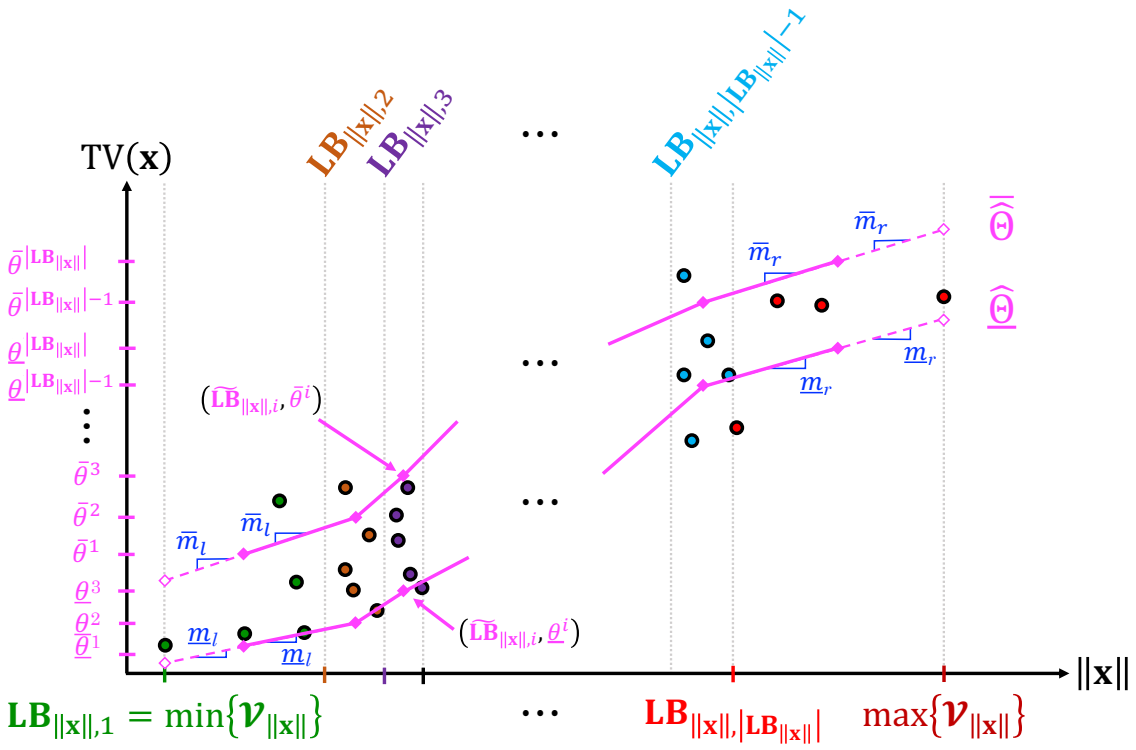


Figure 4-9: Illustration of the end-point interpolation scheme used to extend the non-parametric strong OID bounds $\left\{\widetilde{\mathbf{LB}}_{\|\mathbf{x}\|}, \left\{\overline{\widehat{\Theta}}, \underline{\widehat{\Theta}}\right\}\right\}$ retrieved from Algorithm 2. *Reprinted from [161]. © 2020 IEEE*

We emphasize that the bounds in this section does not assume any underlying distributions for the graph signals, the TV, or the conditional distribution of the TV. In addition, we also explicitly consider the skew of these empirical distributions, and the practical challenges associated with estimating robust bounds from a possibly

small, finite data set $\mathcal{O}_M$.

4.4 US NAS system-wide analysis

In this section, we present our graph signal outlier analysis of the US NAS system-wide network, aggregated over all airlines. We detail the data setup and processing in Section 4.4.1 In Section 4.4.2, 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 4.4.4.

4.4.1 Data setup and processing

We retrieve airport delay data from FAA ASPM, with the time frame of interest starting at January 1, 2008, through December 31, 2017 [85]. We restrict analysis to the 30 busiest airports in the US, collectively known as the FAA Core 30. Using the ASPM delay data, we generate the node-supported signals, i.e., the total delay at each of the Core 30 airports. We generate one graph signal vector per day, where a day is defined as 0000Z to 2359Z. We define the total delay at an airport to be the sum of arrival and departure delays; specifically, we compute the average hourly arrival and departure delays, and then sum across 24 hours. Given our time frame of interest, we generate 3,653 graph signal vectors (one graph signal vector per day), where each vector is of size 30×1. We then generate the airport correlation network, where we assign the edge weights of the graph to be the sample Pearson correlation coefficient (4.1), computed from the 10-year set of airport delay data from ASPM. We observe that all sample correlations are strictly positive: Thus, we do not need to perform a half-plane projection, such as the one discussed in Section 4.1. The resultant airport correlation network is a graph containing 30 nodes (one per airport). The graph is also complete and undirected, with $\binom{30}{2} = 435$ edges.

Figure 4-10 depicts the resultant correlation matrix as a heat map (Figure 4-10(a)) and as a geographical map (Figure 4-10(b)). Two distinct sub-networks demarcating major East Coast and West Coast airports can be seen, 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 that these airports will be 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.

4.4.2 Evaluation of outliers using total variation and total delay

Recall from Section 4.1 that TV and TD provide a low-dimensional projection for analyzing multivariate graph signals. Figure 4-12 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 the weak and strong outliers in distribution are also computed and plotted. $k = 4$ was chosen as the threshold level for outlier detection. The choice of the level k is important for this outlier detection method: As with any outlier detection method (e.g., interquartile range-based box-and-whiskers plots, etc.), there is a parameter that one must set to determine how to delineate outliers versus non-outliers. The important question is the *robustness* of this parameter, i.e., the parameter should be chosen such that small perturbations in the chosen parameter do not impact the outlier detection sensitivity significantly. In this application, our choice of k should be robust enough such that choosing $k \pm \varepsilon$ with ε being some small perturbation should not affect the outlier classification to a large extent. We can verify this through Figure 4-11, where we plot the percentage of system-wide

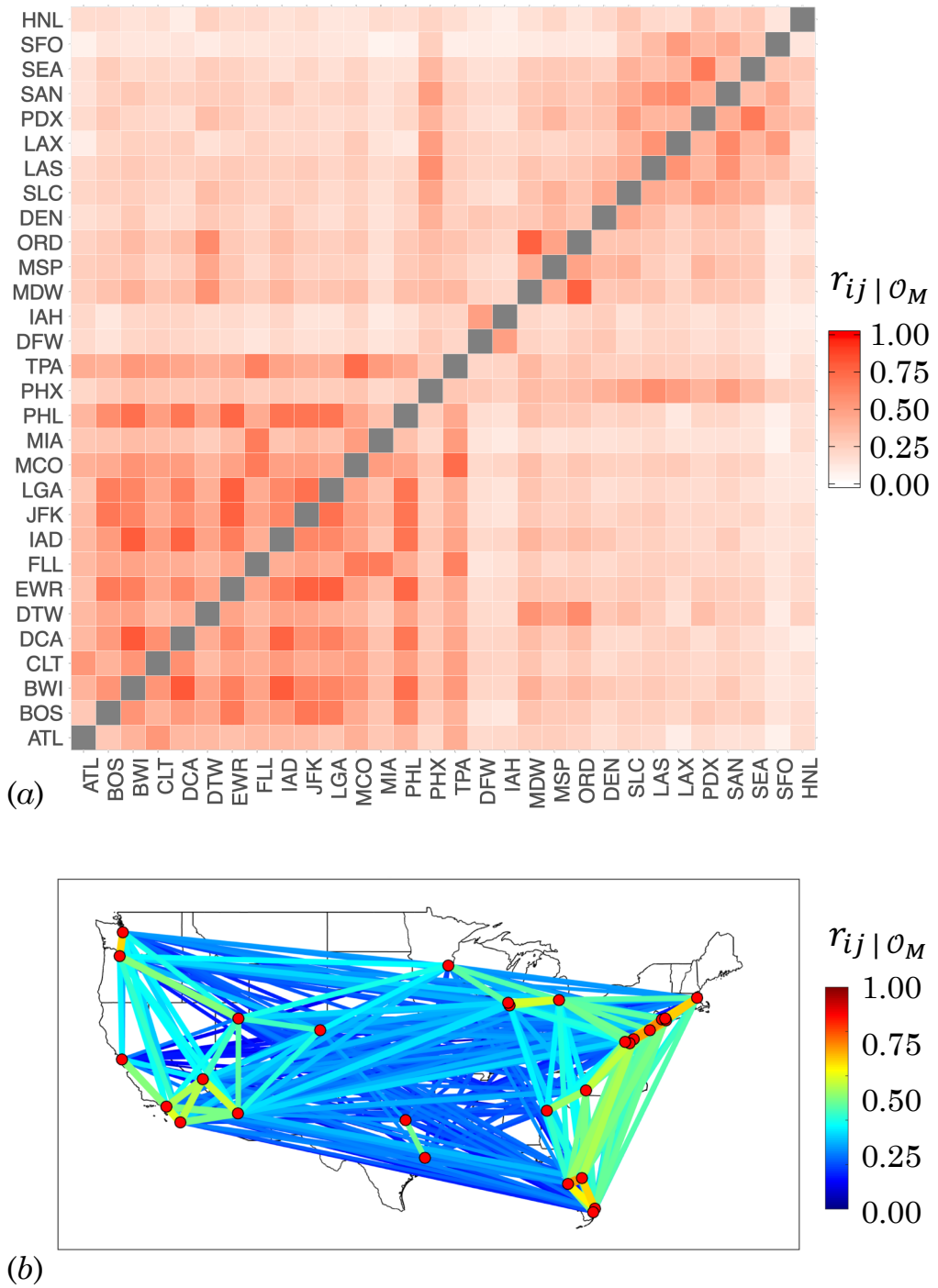


Figure 4-10: (a) Heat map displaying the delay correlation between the top 30 airports; (b) Correlations shown with geographical context. Higher correlations are also emphasized with wider lines in (b). Note that HNL is not shown in (b) for simplicity. *Reprinted from [158].*

days classified as a strong distribution outlier against the parameter k. Note that at around $k = 3$, the outlier detection stabilizes. Our choice of $k = 4$ is robust in the sense that choosing $k \in [4 - \varepsilon, 4 + \varepsilon]$ for some $\varepsilon > 0$ does not significantly alter the percentage of outliers. This would not be true if, for example, a selection of $k = 2$ was made. The behavior seen in Figure 4-11 can be explained by the fact that our data set in the TV-TD plots has a dense core region, where many low TV and TD days are observed, with increasing sparsity as we move away from the core region.

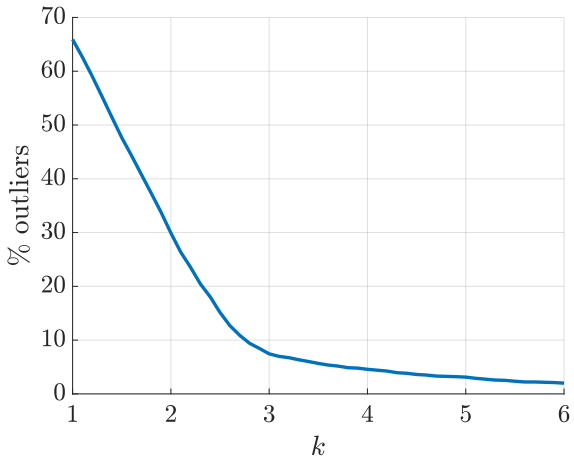


Figure 4-11: Percentage of system-wide days classified as strong outliers in distribution as a function of k.

In Figure 4-12, the lower bound for weak outliers in distribution lie below the horizontal axis, rendering it a vacuous lower bound as TV is a non-negative quantity in our settings (since all correlation coefficients are non-negative). We do not plot this vacuous lower bound: This re-emphasizes the need for the strong outliers in distribution bounds. The bounds for weak outliers in distribution are not able to dynamically adjust to different values of TD, resulting in it being unable to capture the increasing mean value and variance in TV($\mathbf{x}$) as TD increases.

Using the bounds shown in Figure 4-12, we can count the number of observations, i.e., number of days, that were classified as various types of outliers. Out of 3,653 total days, we have that 167 days (4.6 % of all days) were classified as strong outliers in distribution, and 221 days (6.0% of all days) were classified as weak outliers in distribution only. For this data set, no days were classified exclusively as outliers

in scale. Finally, we found that 14 days (0.4% of all days) were classified as both outliers in scale as well as weak outliers in distribution. Another observation from Figure 4-12 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 relationship. For all subsequent discussions and results in this section, we only examine strong outliers in distribution, as it is the tightest in terms of its bounds.

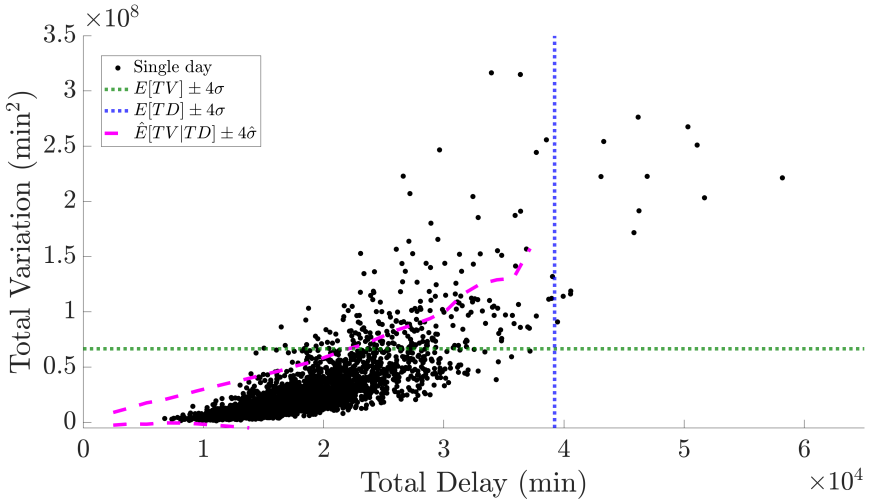


Figure 4-12: TV versus TD for all days in 2008-2017 with level $k = 4$ weak and strong outlier bounds demarcated. *Reprinted from [158].*

4.4.3 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 4-12 that experiences a particular type of disruption. Such an analysis would help 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 categorization is based on delays and cancellations.

In the first categorization of days, we identified four specific disruptions: Thun-

derstorms, nor'easters, hurricanes, and airport or airline outages. A total of 178 days out of the 10-year period was labeled with one of the four types of disruptions (see Table C.2 in the appendix for a list of these dates). The metrics and criteria used for identifying these disruption 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. Nor'easters are reasonably 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 [252], along with an estimate of financial damage to identify 60 days in our data set which are affected by nor'easters [198].

Hurricanes: We consider only Atlantic hurricanes that primarily impact the southern and southeastern coastal regions of the US, as well as 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 [196], (2) the geographic region of impact must include the contiguous US [195], and (3) the resultant financial costs [274]. Similar to nor'easters, hurricanes are typically 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. Another less common reason is due to security and safety-related events. 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 (GDS) and computer reservation systems (CRS). 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 [282].

Thunderstorms: Unlike the other three types of disruptions, thunderstorms are quite common in comparison, occur over very localized regions (on the order of a few hundred miles), are rapidly evolving, last only for a couple of hours, and can be 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 [97], then cross-referenced with weather radar maps to confirm the presence of convective activity [197]. Subsequently, 35 days with severe system-wide disruptions due to thunderstorms are identified. We emphasize that this set of 35 days for our time frame is obviously not exhaustive.

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 from [97, 101] 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 [97]. The four labels that any day can be assigned to, generated from possible delay-cancellation status combinations, are:

- **Low delay; Low cancellation (DLCL):** This is the most common type of day (74.7% occurrence) with relatively normal operations throughout the system.
- **Low delay; High cancellation (DLCH):** These days (3.9% occurrence) are typically indicative of proactive cancellations by airlines in anticipation of severe disruptions (e.g., a nor'easter). The large reduction in flight volumes provides ample schedule buffer and results in low delays.
- **High delay; Low cancellation (DHCL):** Such days (12.2% occurrence) may be indicative of an unplanned disruption such as pop-up thunderstorms, giving airlines little chance to proactively cancel.

- **High delay; High cancellation (DHCH):** The most severe unplanned disruptions typically lead to significant delays and cancellations. These are the days (9.3% occurrence) with the worst system impact.

Figure 4-13 depicts the 178 days classified as nor'easters, hurricanes, thunderstorms or airport outages. Note that we use the same bounds for outliers in scale, as well as the weak and strong outliers in distribution as Figure 4-12, since we are still searching for outliers in the context of the entire system across the 10-year span. Figure 4-14 presents all of the days from January 1, 2014 through October 31, 2017 partitioned into one of the four delay-cancellation groups. Because of the limited availability of the complete cancellation data set used for clustering in [97, 101], we are restricted to a shorter time span instead of the whole time frame of 2008 through 2017. The coordinates of the centroid for each of the subset of the days are also plotted to provide a high-level overview. The counts for outliers in all of these cases, as well as a discussion and interpretation of the results are presented in the next subsection.

4.4.4 The role of disruption in spatial delay distributions

We present the strong outliers in distribution statistics for the four types of disruptions (hurricanes, thunderstorms, nor'easters, and airline- or aiport-specific outages) in Table 4.1, 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 4.3.

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

Table 4.1: Outliers for the four types of disruptions. *Reprinted from [158].*

We see a clear distinction in the occurrence of strong distribution outlier days

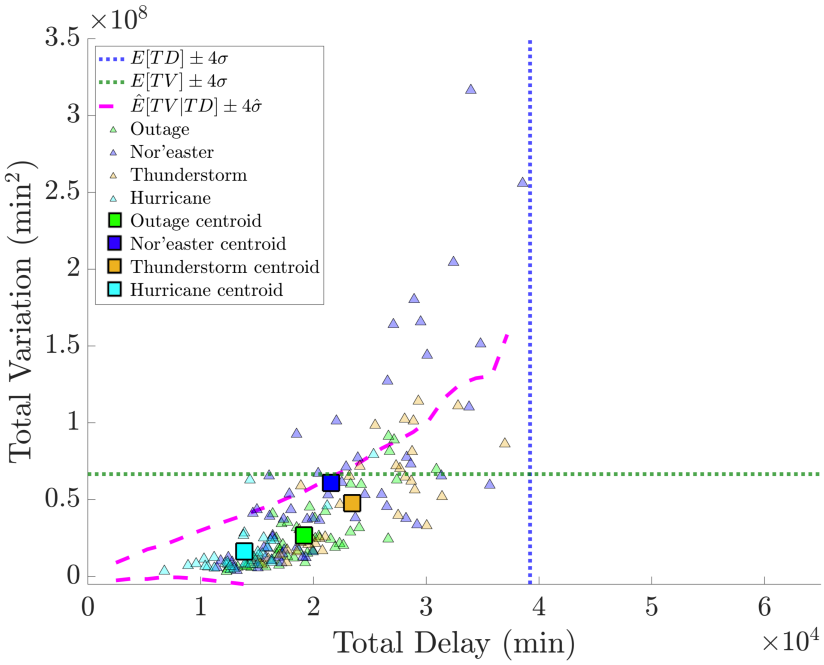


Figure 4-13: TV versus TD plot for a subset of days in 2008-2017 with four disruptions. The average value (centroids) for each category is also shown. *Reprinted from [158].*

for the four disruption categories (Table 4.1). 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 impacts 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 spatially 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

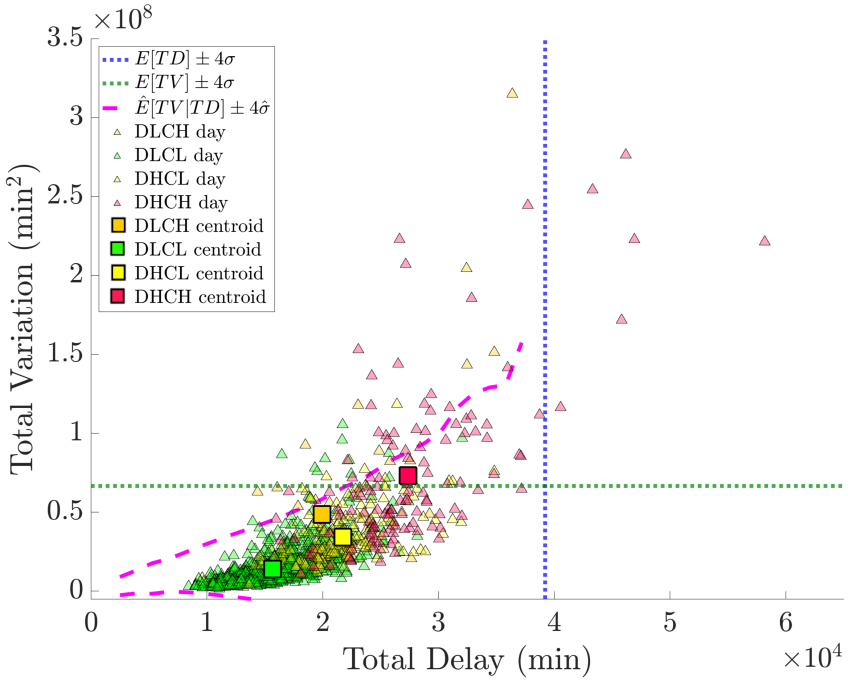


Figure 4-14: TV versus TD plot for a subset of days in 2014-2017 with four system-wide delay and cancellation conditions. The average value (centroids) for each category is also shown. *Reprinted from [158].*

that cancellations [36], which are not accounted for in our analysis, may offer an explanation. However, this is not the case: In Table 4.2, 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 for the four major airlines, which are the subjects of analysis in Section 4.5. We see 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 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 spatial 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 spatial distribution of delays. It is also possible that the regions typically affected by these nor'easter-type storms contain highly congested airports (e.g., New York City-area airports) which are already operating at their capacity limits, further exacerbating the problem. Our data-driven analysis highlights the current challenges faced by airlines regarding proactive management of these nor'easters, and motivates the need to develop more advanced 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 higher incurred airport delays than hurricane days. However, since outages tend to be isolated to one particular airport or airline, its overall impact within the entire system is limited, potentially 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 typically 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 in spatial distribution.

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 conclusion does *not* hold in the airline-specific analysis (Section 4.5).

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

Table 4.2: Percentage of flights canceled across the entire system as well as for each of the four airlines under different disruption categories. *Reprinted from [158].*

We also observe an insightful relationship that links flight cancellations with strong outliers in distribution from Table 4.3. We note 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. However, when the number of cancellations are high, almost 22% of the days are outliers. A possible explanation is that when flights are not canceled, they typically propagate delays based on their route structure and connectivity, and spread the delays across the system, akin to diffusion processes on a graph [90]. While this increases the system-wide delay, it is more spatially homogeneous, and hence decreases the TV. On the other hand, cancellations isolate parts of the network and prevent the propagation of delays. This can result in significantly lower delays downstream in the schedule, since the canceled 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.

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 4.3: Outliers for the delay-cancellation categories. *Reprinted from [158].*

In Section 4.5, we present the spectral analysis, outlier identification, and disruption impact assessment individually for four major US carriers. Extending into Chapter 5, we will also analyze the complex relationship between the system as a whole in comparison to the individual sub-networks of these carriers.

4.5 US airline-specific analysis

Several results from the previous section, as well as later on in Chapter 5, hint at the necessity to zoom in at an airline-specific level. First, some of the eigenvector modes that we will examine in Chapter 5 focus on hub airports for specific airlines, indicating the presence of airline-specific effects. 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 "smooths out" these variations and does not capture the subtleties and nuances of airline operations. We detail the data setup in Section 4.5.1 and the identification of outliers using TV versus TD plots in Section 4.5.2. As mentioned previously, we will defer the analysis of the spectral modes (Section 5.2) and the comparisons between system-wide versus airline-specific results (Section 5.3) to the complementary Chapter 5.

4.5.1 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 [36]. The data pre-processing involves filtering for flights arriving at or departing from our Core 30 airport list, aggregating delays over each specific 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 [34]. Specifically, these four airlines are 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 serve as airline-specific adjacency matrices. For brevity, airline-specific correlation heat maps are presented in Figures B-1 and B-2 in Appendix B. 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 (both do not have operations at MDW), while DL services all 30 airports. Thus, in our ordered indices i for eigenvector modes v_i and eigenvalues λ_i, we have that the highest i for AA, DL, UA, and WN will be 29, 30, 29, and 24, respectively.

We briefly discuss some interesting correlation patterns for each of the four airlines. 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 largest hub at DFW. In contrast, the DL network reflects a much stronger presence of airport delay correlations in the East Coast, and is more similar to the correlation network for the system-wide case. UA's network highlights correlations in the West Coast and the 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. This emphasizes the intrinsically different network structures, routing strategies, and tail assignments used by WN compared to the three other network legacy carriers.

We will now 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 4.5.2), and interpret the relationships between the airline-specific versus system-wide analyses (Section 5.3). In terms of specific disruptions, we consider the same set of 178 days from Section 4.4.2 in which each day was labeled as one of the four disruption categories.

4.5.2 Discussion of airline-specific outliers

We first compute airline-specific strong distribution outlier bounds; the TV versus TD plots with these bounds for each of the four airlines can be found in Figures B-3 (AA), B-4 (DL), B-5 (UA), and B-6 (WN) in Appendix B. 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 within airline sub-networks, compared to the system-wide network. Furthermore, while nor'easters had the greatest effect on system-wide spatial delay distributions, their impact at the airline sub-network level is diminished. While AA, UA, and WN disruption centroids all remained within the empirical strong outlier bounds, the DL centroids for thunderstorms and outages are outside of 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 4.4, along with corresponding system-wide outlier statistics. 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 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 thunder-

storms. 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 outliers in distribution.

Category (*System-wide outlier %*)	**Airline**	**Outlier counts**	**%**
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%
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%
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%
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%
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 4.4: Outlier counts and percentages for each type of disruption, split by airline. *Reprinted from [158].*

Finally, airline- and airport-specific outages result in lower system-wide outlier occurrences (4.1%) in comparison to airline-specific outliers (average of 15.8%). 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. Note that the airline-specific TV-TD plots, restricted to the 178 disruption days, can be found in Figures B-7 (AA), B-8 (DL),

B-9 (UA), and B-10 (WN) in Appendix B.

4.6 Chinese airspace delay analysis: Spatial and magnitude characteristics

We shift now to applying the outlier detection methods from earlier in this chapter to the Chinese airport network. For context, China is the second-largest aviation market in the world, with approximately 560 million passengers transported in 2017 [119]. This number is comparable in magnitude to the US, which is the largest commercial aviation market serving almost 890 million passengers. With a projected growth rate of 5.3% per year for the next 20 years, the growth rate for Chinese aviation is significantly higher than the global average of 3.8% [119]. To give a sense of scale, China contributed to the largest increase in domestic air traffic among all parts of the world, with an increase of 48.8 million passenger journeys in 2018 [120]. Such unprecedented growth, along with the massive scale of the system, motivates our analysis of inefficiencies such as flight delays in the Chinese aviation system. Currently, the average arrival delay for flights at the top 10 airports (based on traffic volumes) in China is 26 minutes [204]. Note that this is a substantial amount of arrival delays compared to other mature markets such as the US, which has an average arrival delay of 17 minutes per flight among US top 10 airports [264]. Faced with increasing traffic levels, the problem of excessive flight delays is also expected to grow in China.

Two main factors result in the Chinese air transportation system to be more vulnerable to delays: First, the limited availability of airspace resources creates a capacity-side constraint. Secondly, high-traffic routes are concentrated in a relatively small geographical region. To expand upon the first constraint, we note that in China, a vast majority of the airspace is blocked for military usage, resulting in only about 25% of the airspace available for civilian use. As a consequence, commercial flights do not have the flexibility to follow optimal routes, and create congestion at airway choke points in order to navigate around restricted airspace [219].

Expanding upon the second point of concentrated traffic densities, most of the traffic are distributed in the eastern part of China, resulting in collocated regions of congested airspace. Among the top 15 domestic passenger traffic origin-destination pairs, six of them belong to China (Beijing-Capital and Shanghai-Hongqiao; Chengdu and Beijing-Capital; Beijing-Capital and Shenzhen; Shanghai-Hongqiao and Shenzhen; Guangzhou and Shanghai-Hongqiao; Guangzhou and Beijing-Capital) [263]. We note that all of these are routes are in the eastern part of China. Thus, systemic high traffic volumes in the eastern part of China, combined with adverse meteorological phenomena that typically affect the same region, result in large disruptions of airline and airport operations. Additionally, we also note that certain social events such as the Golden Week of national holidays have been shown to significantly increase demand as well as delays and cancellations [161].

In this section, we focus our analysis on four major air carriers in China. These air carriers of interest are Air China (CA), China Southern Airlines (CZ), Hainan Airlines (HU), and China Eastern Airlines (MU). Together, these four full-service carriers represent nearly 70% of the market share in China. Of these four, China Southern Airlines is the largest air carrier, comprising of almost a quarter of the market share in China. We provide a high-level summary of these four airlines in Table 4.5. An important point worth noting is that all four of these carriers operate a hub-based network structure, with Beijing Capital International Airport (PEK) being a hub for three out of four airlines. While each of the four carriers have on-time performance metrics that could be improved upon, our work highlights the differences among these four airlines in terms of the temporal and spatial extent of delays they experience. Several of our results also provide comparisons between airlines in terms of temporal trends (i.e., when outlying disruption days occur), as well as which group of airports are typically affected in such cases.

	Air China (CA)	China Southern (CZ)	Hainan Airlines (HU)	China Eastern (MU)
Passengers (2019, Mil.)	115	152	82	130
Fleet size	429	619	223	567
Primary hubs (IATA code)	PEK, PVG, CTU	CAN, PEK	PEK, HAK, XIY, SZX	PVG, SHA, KMG
Carrier type	Full service	Full service	Full service	Full service
Market share (2017)	18%	23%	14%	17%
Average canc. (2017)	4.2%	5.2%	4.6%	5.4%

Table 4.5: Overview of network and operating characteristics for the four Chinese airlines we consider in this analysis. Please refer to Table C.6 in Appendix C for a list of the IATA codes and full airport names. *Reprinted from [160].*

4.6.1 Previous research on airline networks and flight delays in China

We will now provide a brief overview of some existing literature related to data-driven characterizations of Chinese airline networks. [287] investigates the evolution of the Chinese airport network (CAN), including its topology and traffic: The authors found that the traffic grows at an exponential rate with seasonal fluctuations, while the network topology does not change significantly. A k-core decomposition method was used in [67] to divide the CAN into multiple layers. The *core layer* contained 17.7% of airports within the CAN, most of which were located in economically developed cities. Furthermore, this core layer was found to be a densely connected network, sustaining most of the traffic flows. Comparisons between the Chinese air transportation network and its US counterpart have also been made [219, 161], albeit at different granularities: [219] focused on using flight-specific trajectory data to characterize air traffic networks, whereas [161] analyzed differences in airport-level delay distributions. From these comparative studies, some notable conclusions include the fact that compared with the US network, China has more restricted airspace availability for commercial flights and faces a greater risk of en route congestion [219]. At the level of all airlines, the airports responsible for unusual spatial delay distributions in China were found to be much less geographically clustered than similar airports in

the US [161]. A key extension that we pursue in this work is deriving airline-specific insights at the level of airport delay distributions for the CAN.

The use of graphs as a helpful abstraction for network analysis has been studied in the context of the CAN and the Chinese aviation system: [43] conducts community detection on Chinese airport delay correlation networks in which edges are built using correlations of the delay time series between two airports to reveal implicit delay propagation relationships. [269] found that inter-country (US and China) differences in how scheduled block times (SBTs) are set can account for large differences in on-time performance between the two countries. The authors of [269] indicate that there are restrictions on SBTs for flights passing through congested airways in peak hours, due to concerns that longer SBTs will lead to excessive airspace occupancy. Finally, [268] developed a number of operational resilience measures, and used them to assess how different scales and intensities of disruptions at a given airport affect aviation operations throughout the network.

4.6.2 Data processing

We obtain airport delay data, conditioned on the four Chinese airlines from Table 4.5, for the 30 Chinese airports considered in our analysis (see Table C.6 for a list of these airports and their IATA codes, as well as Figure B-11 for a geographic plot of these airports) from the Operations Monitoring Center of the Civil Aviation Administration of China (CAAC). The time frame of the airline-specific airport delay data spans January 1, 2012 through December 31, 2015. The data granularity is one observation per day, equaling $N = 2,192$ observations per airline, or $8,768$ daily airport delay observations in total. We can abstract each daily delay observation for an airline as a non-negative multidimensional vector $\mathbf{x}^{(t)} \in \mathbb{R}_{\geq 0}^{30\times 1}$, where the i^{th} component x_i gives the delay (in minutes) at airport i for day t.

We will need to build out the graph abstraction in order to perform the spatial delay distribution analysis in Section 4.6.3, i.e., we need to construct how the dependencies within our airport network G are modeled. As described earlier in this chapter, we use the sample Pearson correlation coefficient as a measure of dependence

between pairs of airports in V. Specifically, this gives us a first-order measurement of linear dependencies between airport delays. One way to represent this is through an *adjacency matrix* A that describes the 30×30 dependencies within the airport network G. This adjacency matrix A will have 30 rows and columns, with each element a_{ij} of A being the sample Pearson correlation coefficient between airports i and j. Note that since Pearson correlations are symmetric quantities, we have that $a_{ij} = a_{ji}$. Note also that we will have four different correlation networks, each corresponding to one of the four airlines we analyze. After computing all a_{ij}'s, we can plot them geographically in Figure 4-15 to get a sense of high and low correlation areas within each airline's graph.

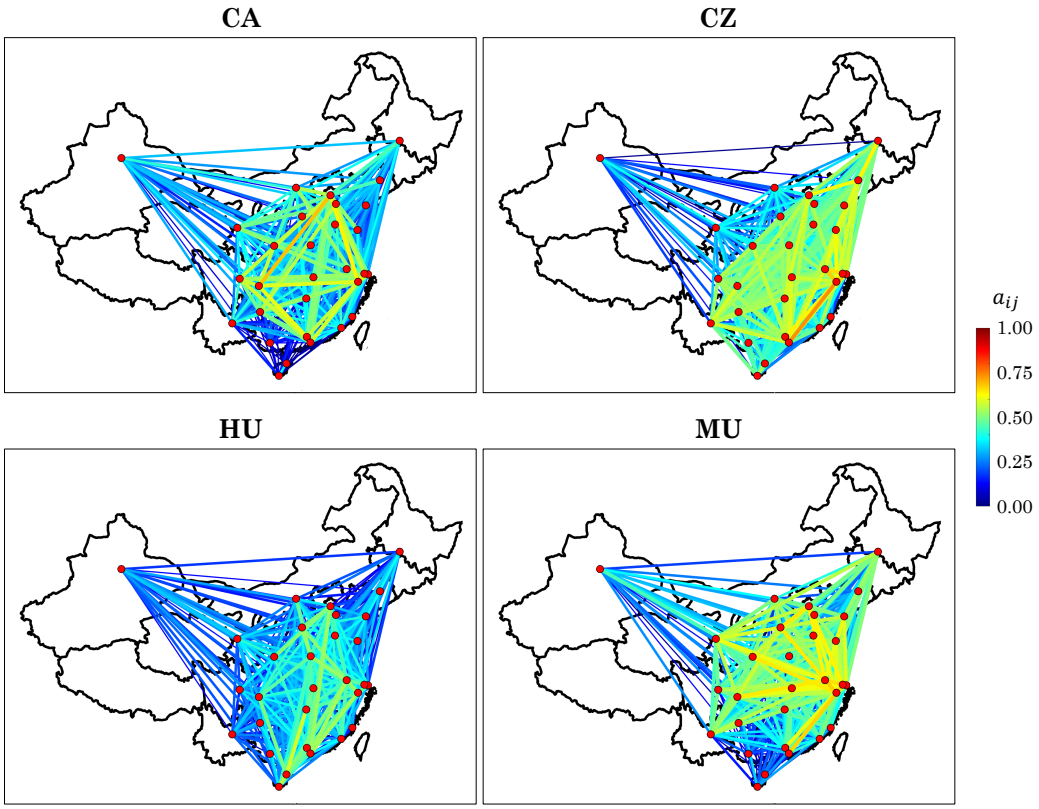


Figure 4-15: Correlation networks superimposed on a geographical map of China for the four Chinese airlines we analyze. Note that higher correlation values are also emphasized with wider lines. *Reprinted from [160].*

4.6.3 Spatial delay distribution analysis

We now analyze the spatial distribution of airport delays in the Chinese airport network, with a focus on *unexpected* spatial delay distributions. We identify individual outlier days in terms of their spatial delay distribution and examine monthly distributions of days that are considered spatial delay distribution outliers. We will contrast this with the monthly distribution of airport delay magnitudes: In particular, we will highlight the differences in monthly distributions when airport delays are observed from a magnitude versus spatial distribution standpoint, and comment on its operational implications.

To examine the day-by-day occurrences of unexpected spatial delay distributions for each airline's sub-network, we utilize the skewed interquartile range-based non-parametric outlier bounds detailed in Section 4.3 to classify the three different types of outliers summarized in Section 4.1.2. These outlier definitions include outliers in scale (OIS), weak outliers in distribution (weak OID), and strong outliers in distribution (strong OID). Recall that OIS describe days where the magnitude of delay – now taking into account network effects – is higher (or lower) than normal, whereas both weak and strong OID measure the spatial distribution of airport delays across an airline's sub-network, with the latter being a magnitude-independent measure of spatial variation. We plot the total delay and total variation for each of the 2,192 days between 2012-2017, for each airline, in Figure 4-16. We also plot the bounds demarcating whether or not a day is OIS, weak OID, or strong OID. Note that these outlier classifications are not mutually exclusive: For example, a day could be both an OIS as well as a strong OID, indicating that an extremely severe disruption occurred in that airline's sub-network, affecting not only the magnitude but also the spatial distribution of delays.

Based off of how many data observations (each observation represents a day in an airline's network) lie outside of the various outlier bounds from Figure 4-16, we record the number of OIS, weak OID, and strong OID days for each of the four Chinese airlines, and report the outlier statistics in Table 4.6. Note that the figures quoted

in Table 4.6 are in number of days, with percentages given by dividing the number of days by 2,192 total days, i.e., the total number of observations between 2012-2017. We see that HU has the highest number of OIS days and weak OID days, as well as the second-highest number of strong OID days. If we focus solely on the spatial distribution of delays independent from its magnitude, then CZ claims the highest number of strong OID days, with approximately 5% of all days between 2012-2017 being classified as a strong OID. Referring back to Figure 4-15, we note that CZ has several edges with high correlations, concentrated in the southeastern region of China. Recall that since "less smooth" distributions of delays can only occur across edges with high correlation (i.e., two highly correlated airports where one airport has low delays and the other has high delays, or vice versa), this indicates that CZ often experiences uneven delay distributions across highly correlated airports, leading to a high number of strong OID days. Since we are examining strong OID days, this observation holds true for CZ independent of how high the delays were, i.e., the magnitude of the delays.

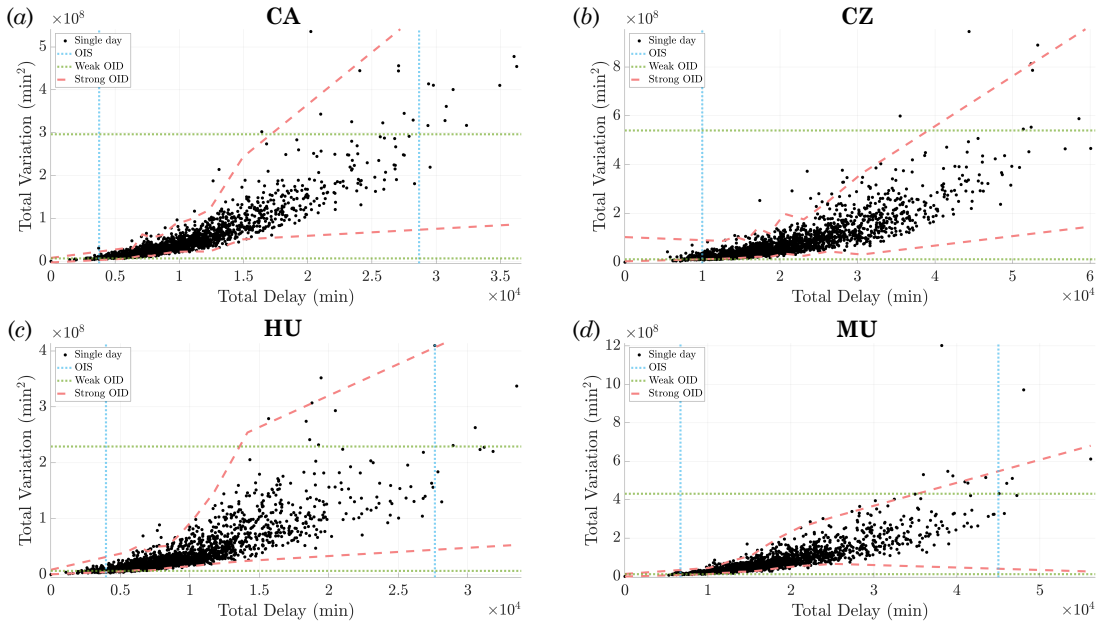


Figure 4-16: TV versus TD for all days in 2012-2017 for (*a*) Air China (CA), (*b*) China Southern Airlines (CZ), (*c*) Hainan Airlines (HU), and (*d*) China Eastern Airlines (MU). *Reprinted from [160].*

Outlier Type	Air China (CA)	China Southern (CZ)	Hainan Airlines (HU)	China Eastern (MU)
OIS	17 (0.8%)	44 (2.0%)	**52 (2.4%)**	13 (0.6%)
Weak OID	37 (1.7%)	24 (1.1%)	**61 (2.8%)**	32 (1.5%)
Strong OID	47 (2.1%)	**116 (5.3%)**	76 (3.5%)	54 (2.5%)

Table 4.6: Airline-specific outlier statistics; the number of outlier days as well as the percentage out of 2,192 days (2012-2017) are provided. *Reprinted from [160].*

We can also assess the temporal trends of unexpected spatial delay distributions occurring in the sub-networks of the four Chinese airlines we analyze. Specifically, we will examine the number of strong OID days by months, and contrast this with the average delay magnitudes by month, as shown in Figure 4-17. Note that we will only consider the strong OID counts in this analysis, because we want to focus solely on the effects of spatial variance, and explicitly exclude the effects of delay magnitudes. For the delay magnitudes in Figure 4-17, we see a distribution with peak average delays in the summer months. In particular, for the 2012-2017 time period, the month of July saw peak average delays across the four airline sub-networks. Interestingly, this is not the case with strong OID counts, as shown in Figure 4-18. This contrasting observation suggests that characterizing delays according to their magnitude versus their spatial variance can result in highly orthogonal measurements of two different kinds of network performance: For the Chinese airline sub-networks, the periods of high delay magnitudes do not seem to coincide with periods of unexpected spatial delay distributions. In other words, whereas delays are typically high during summer months, they are spatially located at "expected" airports throughout each of the airlines' network. On the other hand, during fall and winter months, and especially in October, even if the magnitude of delays are lower compared to the peak summer months, the *spatial location* of these delays within each airline's network is increasingly unexpected, resulting in delay distributions with higher variations across the correlation networks for each airline. One potential operational factor is that in the summer months, heavy convective activity cause severe levels of disruptions, but the resultant airport delays diffuse throughout the network, and were distributed within the network in an expected manner. On the other hand, unusual demand profiles in

months such as October (e.g., due to societal-scale holidays) could coincide with less severe disruptions, but result in delays that appear in unexpected locations throughout each airline's network. An example of such an unusual demand profile could be during the Golden Week holidays in October, where the Chinese aviation system typically experiences large surges in travelers [281, 161].

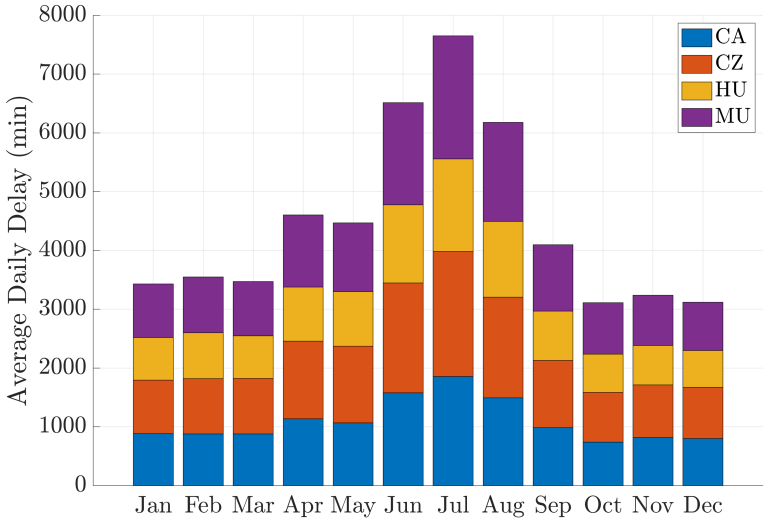


Figure 4-17: Monthly distribution of average daily delays for all four Chinese airlines. *Reprinted from [160].*

4.7 Future work

4.7.1 Control of graph signals and processes

There is wide range of previous research that deals with the controllability of networks [172, 169, 285] and optimal control of graph signals and processes, particularly within the framework of graph signal processing. We review some research advancements that may be particularly relevant to our problem of achieving magnitude- and spectral-based control of airport delay signals on top of our correlation network. In [19], the classical linear quadratic optimal controller (LQR) is re-cast as a graph filter through the GFT, acting on time-varying (in the auto-regressive sense, i.e., the state

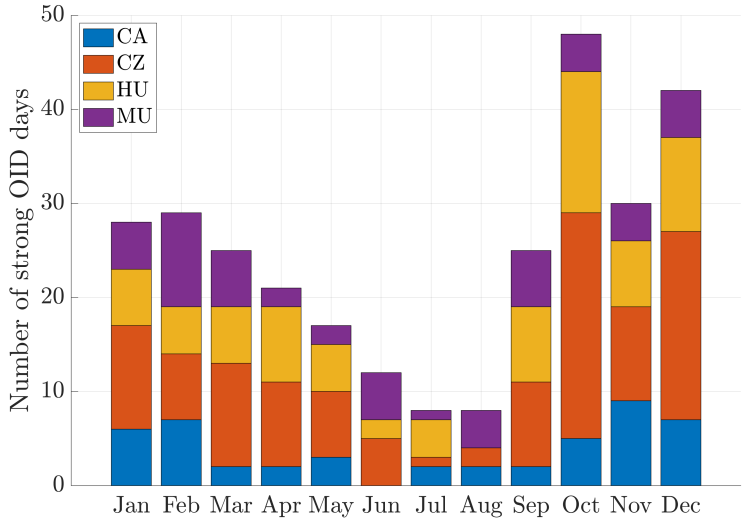


Figure 4-18: Monthly distribution of strong OID days for all four Chinese airlines. *Reprinted from [160].*

of the nodal signal in time $t+1$ depends linearly on neighboring nodal signals from the previous instant t) graph-frequency components. Specifically, the temporal state evolution equation is:

$$\mathbf{x}_{t+1} = \mathfrak{L}\mathbf{x}_t + b\mathbf{u}_t + \mathbf{w}_t, \tag{4.39}$$

where b is constant, with zero-mean Gaussian noise $\mathbf{w}_t \sim \mathcal{N}(\boldsymbol{\mu}, \Sigma)$ and control input $\mathbf{u}_t \in \mathcal{U} \subseteq \mathbb{R}^{N\times N}$. Here, [19] takes the state evolution matrix to be some diagonalizable graph shift operator (the graph Laplacian $\mathfrak{L} = \mathbf{V}\Lambda\mathbf{V}^{-1}$ is one example, where Λ stores the eigenvalues λ_i and $\mathbf{V}$ stores the eigenvector modes v_i, for $i = 1, \ldots, N$). This may be a drawback for us, since the evolution of airport delays cannot fully be represented solely by the weights of the correlation network. However, we could use, e.g., the Markov Jump Linear System model from [96], wherein the state matrix directly influences the evolution of airport delays. Recall from Definition 2 that the GFT of $\mathbf{x}$ can be written as $\hat{\mathbf{x}} = \mathbf{V}^\top\mathbf{x}$; hence, the graph frequency representation of (4.39) can be written as:

$$\hat{\mathbf{x}}_{t+1} = \Lambda \hat{\mathbf{x}}_t + b\hat{\mathbf{u}}_t + \hat{\mathbf{w}}_t, \tag{4.40}$$

and since $\Lambda = \text{diag}\,(\lambda_1, \ldots, \lambda_N)$, we can de-couple the graph frequency state evolution at each node $i = K$ as:

$$\hat{x}_{t+1,K} = \lambda_K \hat{x}_{t,K} + b\hat{u}_{t,K} + \hat{w}_{t,K}. \tag{4.41}$$

The LQR controller attempts to find some optimal policy $\mathbf{u}_t^*$ that minimizes the expected cost (optimal in expectation due to $\mathbf{w}_t$) across a finite time horizon $t = 0, \ldots, T$ as well as a terminal cost at $t = T$, constrained by the system evolution (4.39). In other words, we want to solve:

$$\begin{aligned} &\min_{\{\mathbf{u}_t\}_{t=0}^{T-1}} \mathbb{E}\left[\sum_{t=0}^{T-1} \left(\mathbf{x}_t^\top \mathfrak{L}\mathbf{x}_t + \mathbf{u}_t^\top R\mathbf{u}_t\right) + \mathbf{x}_T^\top \mathfrak{L}\mathbf{x}_T\right] \\ &\text{s. t.} \quad \mathbf{x}_{t+1} = \mathfrak{L}\mathbf{x}_t + b\mathbf{u}_t + \mathbf{w}_t, \qquad t = 0, \ldots, T-1. \end{aligned} \tag{4.42}$$

A main contribution of [19] is showing that the optimization problem to be solved in order to retrieve the optimal LQR controller (4.42) can be re-written in the graph frequency domain via the GFT. To do so, the system evolution matrix $\mathfrak{L}$ and control cost matrix R must be chosen such that:

$$\begin{aligned} \mathfrak{L} &= \mathbf{V}^{-H} D_Q \mathbf{V}^{-1}, \\ R &= \mathbf{V}^{-H} D_R \mathbf{V}^{-1}, \end{aligned} \tag{4.43}$$

where $D_Q = \text{diag}\,(q(\lambda_1), \ldots, q(\lambda_N))$ and $D_R = \text{diag}\,(r(\lambda_1), \ldots, r(\lambda_N))$. The functions q and r are positive cost functions that assign a cost to each eigenvector λ_i; note that since we assume the system evolution matrix to be the graph Laplacian itself (i.e., the cost directly penalizes the TV $\mathbf{x}_t^\top \mathfrak{L}\mathbf{x}_t$ at each time instance t), then $q\,(\lambda_i) = \lambda_i$. The graph frequency-version of (4.42), written node-wise at node $i = K$, is:

$$\begin{aligned} &\min_{\left\{\hat{u}_{t,K}\right\}_{t=0}^{T-1}} \mathbb{E}\left[\sum_{t=0}^{T} \lambda_K \left|\hat{x}_{t,K}\right|^2 + \sum_{t=0}^{T-1} r\left(\lambda_K\right) \left|\hat{u}_{t,K}\right|^2\right] \\ &\text{s. t.} \quad \hat{x}_{t+1,K} = \lambda_K \hat{x}_{t,K} + b\hat{u}_{t,K} + \hat{w}_{t,K}, \qquad t = 0, \ldots, T. \end{aligned} \tag{4.44}$$

(4.44) can now be solved for each node $i = 1, \ldots, N$ within the graph frequency component, since at each node $i = K$, we have that $\lambda_K > 0$ and $r(\lambda_K) > 0$ are positive scalars, and (4.44) becomes the standard linear-quadratic dynamic program which admits a solution through backward induction [25]. The optimal controller is given in [19] by the following proposition:

Proposition 6 (Proposition 1 from [19]) *Let* $\mathbf{u}_t^* = u_t^*(\mathbf{x}_t)$ *denote the optimal policy in* (4.42) *and* $\hat{\mathbf{u}}_t^* = \mathbf{V}^{-1}\mathbf{u}_t^*$ *its GFT with graph frequency components* $\hat{u}_{t,K}^*$ *at node* $i = K$*. Then,*

$$\hat{u}_{t,K}^* = -\frac{b\lambda_K \gamma_t(\lambda_K)}{b^2\gamma_t(\lambda_K) + r(\lambda_K)}\hat{x}_{t,K} \tag{4.45}$$

for all K *such that* $\lambda_K > 0$*, and* $\hat{u}_{t,K}^* = 0$ *otherwise. The function* $\gamma_t(\lambda_K)$ *in* (4.45) *is defined by the Riccati backwards recursion starting at* $\gamma_T(\lambda_K) = \lambda_K$*:*

$$\gamma_t(\lambda_K) = |\lambda_K|^2 \left(\frac{r(\lambda_K)\gamma_{t+1}(\lambda_K)}{b^2\gamma_{t+1}(\lambda_K) + r(\lambda_K)}\right) + \lambda_K \geq 0. \tag{4.46}$$

The foundation laid out by [19] is a good starting point for formulating any kind of optimal control problem for the airport delay signals case, since [19] focused in-depth on controlling the graph frequency (i.e., the spatial variation) at the level of an individual node (i.e., airport). One drawback of the scheme proposed by [19] is the assumption that all nodes can be used to control the system, whereas in reality, this may not be the case. For example, we might only be able to influence the airport delays at a limited subset of airports within our system through active TMIs. This limited-controllable node setting is studied in the context of *bandlimited* graph signals [17, 124], i.e., the graph signal $\mathbf{x}$ can be expressed by $M < N$ number of eigenvector modes. Specifically, [44, 131] all examine ways of constructing a *bandlimited prior* in order to pick the appropriate set of driving nodes, and a further study [93] examines the more nuanced case of controlling bandlimited graph processes (i.e., graph signals with temporal dynamics) supported on top of a random, time-varying graph topology. The latter aspect of [93] is particularly interesting, since most other works that

examine the problem of controlling graph signals assume time-invariant graph topologies, even though in reality, the actual node and edge set might vary randomly with time. For example, airport nodes could go offline according to ground stop TMIs, and various edges could be removed in the sense of OD-specific cancellations.

In particular, in [93], given a discrete-time diffusion process $\mathbf{x}_t = (I_{N\times N} - \epsilon\mathfrak{L})\,\mathbf{x}_{t-1}$ where $I_{N\times N}$ is the identity matrix, $\mathfrak{L}$ is the graph Laplacian, and ϵ is the system stability parameter that guarantees stability if $0 < \epsilon \leq 1/\left\|\mathfrak{L}\right\|_2$, we have the following N-state linear system:

$$\mathbf{x}_t = (I_{N\times N} - \epsilon\mathfrak{L})\,\mathbf{x}_{t-1} + B\mathbf{u}_{t-1}, \tag{4.47}$$

Since this is the limited driving node case, the control input $\mathbf{u}_t \in \mathbb{R}^{M\times 1}$ acts on $M < N$ nodes, and $B \in \mathbb{R}^{N\times M}$ is the corresponding control input matrix. Setting $A = I_{N\times N} - \epsilon\mathfrak{L}$, we can write (4.47) in the graph frequency domain as:

$$\hat{\mathbf{x}}_{t,\widetilde{M}} = \hat{A}_{\widetilde{M}}\hat{\mathbf{x}}_{t-1,\widetilde{M}} + \mathbf{V}_{\widetilde{M}}^{H} C^{\top}\mathbf{u}_{t-1}, \tag{4.48}$$

where $\widetilde{M}$ now represent the discrete number of eigenvector modes (i.e., bandlimited bandwidth) that we restrict our graph signals to, $\mathbf{V}_{\widetilde{M}}$ is the set of eigenvector modes, restricted to $\widetilde{M}$, and C is a combinatorial, binary matrix that picks the appropriate driving nodes. An interesting contribution of [93] is given in the following proposition, which provides a necessary condition on the minimum number of driving nodes needed in order to control the system in (4.47) in the graph frequency domain within a finite time interval T:

Proposition 7 (Proposition 1 from [93]) *Consider the linear system* (4.48) *describing a process over a deterministic graph* $\mathcal{G}$*. A necessary condition to control the system in a finite time* T *towards a target frequency content* $\hat{\mathbf{x}}^{*}_{\widetilde{M}}$ *over the* $\widetilde{M}$ *frequencies of interest is to select* $M \geq \lceil \widetilde{M}/T \rceil$ *driving nodes.*

Additional propositions dealing with the case of random, time-varying graphs are also provided in [93].

4.7.2 Graph sparsification and estimating better airport delay networks

We would also like to improve upon the actual choice of airport delay correlation networks that we use in our various analyses, specifically from the viewpoint of *sparsity*. In our various analyses, we always used a fully-connected, complete network, where for every unique pair of airports i and j, we always compute and retain the sample Pearson correlation coefficient between the delays at airport i and j. However, this may inadvertently introduce and retain spurious correlations. Another line of future work involves investigating and applying rigorous sparsification methods to introduce sparsity in the sample correlation (or equivalently, the sample covariance) matrix of airport delays. Along this line of thought, we present two relevant methods to examine: Thresholding and regularization of the sample covariance matrix [29, 30], and using graphical LASSO [91, 179].

Thresholding and regularization of the sample covariance matrix

Suppose that, similar to how we retrieve the sample covariance and correlation matrices from airport delay data, we observe n i.i.d. p-dimensional observations $\mathbf{X}_1, \ldots, \mathbf{X}_n$ distributed according to a Gaussian distribution F, where without loss of generality, F is zero-centered and with covariance Σ. Define the sample covariance matrix as

$$\hat{\Sigma} = \frac{1}{n} \sum_{k=1}^{n} \left(\mathbf{X}_k - \overline{\mathbf{X}} \right) \left(\mathbf{X}_k - \overline{\mathbf{X}} \right)^{\top}, \tag{4.49}$$

and denote individual covariance between $\mathbf{X}_i$ and $\mathbf{X}_j$ to be $\hat{\Sigma} = [\hat{\sigma}_{ij}]$. One simple way to introduce sparsity into $\hat{\Sigma}$, and thus removing an edge between nodes i and j in the corresponding correlation network, is to set a threshold $s \in \left[\min_{1 \leq i \leq j \leq p} \{\hat{\sigma}\}, \max_{1 \leq i \leq j \leq p} \{\hat{\sigma}\} \right]$ and remove all sample covariance matrix entries below that threshold. We briefly go over some results from [29] along with their related work in [30] that give guarantees on the consistency of such a thresholding approach. Furthermore, more relevant to our desired application, [29, 30] also proposes a method for constructing a threshold in a rigorous way.

We introduce some brief notation used in [29] to introduce one of their main results: A theorem that, loosely speaking, informs you that thresholding in a manner similar to what is described above will produce a sparser sample covariance matrix that is still "close" to the original sample covariance matrix, with useful preservation of characteristics such as label permutations and positive (semi-)definiteness of the sample covariance matrix. In particular, given a matrix $M = [m_{ij}] \in \mathbb{R}^{N \times N}$, denote the smallest and the largest eigenvalues by $\lambda_{\min}(M)$ and $\lambda_{\max}(M)$, respectively. Define a *thresholding operator* $T_s : \mathbb{R}^{N \times N} \to \mathbb{R}^{N \times N}$ that maps $M \mapsto T_s(M) = [m_{ij} \times \mathbf{1}(|m_{ij}| \geq s)]$. To guarantee that the sparser matrix $T_s(M)$ is still positive definite, construct the thresholding operator such that

$$\max_{1 \leq j \leq N} |\lambda_j (T_s - M)| \leq \varepsilon < \lambda_{\min}(M).$$

Finally, let $\|M\|_F^2 = \operatorname{tr}\left(MM^\top\right)$ denote the Frobenius matrix norm of M. Ignoring some mathematical technicalities regarding the uniformity of the covariance matrix Σ, a main result from [29] is the following:

Theorem 1 (Theorem 1 from [29]) *For sufficiently large M', if*

$$t_n = M' \sqrt{\frac{\log p}{n}} \tag{4.50}$$

and $\log p / n = o(1)$, then the operator norm difference between thresholding $\hat{\Sigma}$ at t_n and Σ is bounded in probability, i.e.,

$$\left\| T_{t_n}\left(\hat{\Sigma}\right) - \Sigma \right\| = \max_{1 \leq j \leq p} \left| \lambda_j \left(T_{t_n}\left(\hat{\Sigma}\right) - \Sigma \right) \right| = O_P\left(c_o(p) \left(\frac{\log p}{n} \right)^{(1-q)/2} \right), \tag{4.51}$$

where $c_0(p)$ and q are parameters on the uniformity class of Σ.

What is more readily applicable for our line of work is a data-driven method proposed by [29, 30] that constructs the thresholding value s for the T_s thresholding operator, with attractive boundedness in probabilty properties with respect to the

Frobenius norm. This proposed method is as follows: Split the sample set into two pieces of size n_1 and n_2 (one suggestion of sizing is $n_1 = n\left(1 - 1/\log n\right)$ and $n_2 = n/\log n$), and repeat this splitting process K times. Define the empirical covariance matrices $\hat{\Sigma}_{1,k}$ and $\hat{\Sigma}_{2,k}$ for the subset of size n_1 and n_2, respectively, at the k^{th} split out of K. Then, pick s^* where:

$$s^* = \underset{\substack{s \geq \varepsilon_n \to 0 \\ \varepsilon_n \asymp \log p/n}}{\arg\min} \left\{ \frac{1}{K} \sum_{k=1}^{K} \left\| T_s\left(\hat{\Sigma}_{1,k}\right) - \hat{\Sigma}_{2,k} \right\|_F^2 \right\}. \tag{4.52}$$

Another advantage of the method in (4.52) is that interpretability is retained in the sense that you know exactly the threshold value s for your threshold operator T_s, and there are no surprises regarding the values within the sparser sample covariance matrix $T_s\left(\hat{\Sigma}\right)$. This is an advantage that the next method, graphical LASSO, lacks.

Graphical LASSO

Graphical LASSO is a convex optimization-based way to estimate a sparse underlying undirected graphical model by penalizing the estimation of the precision matrix Σ^{-1} in a ℓ_1 sense [91, 179]. Suppose as per usual that we observe n i.i.d. p-dimensional observations $\mathbf{X}_1, \ldots, \mathbf{X}_n$ distributed according to a Gaussian distribution $\mathcal{N}(\boldsymbol{\mu}, \Sigma)$. Let us denote the precision matrix as $\Theta = \Sigma^{-1}$, and the empirical covariance matrix by $S = \hat{\Sigma}$. Graphical LASSO attempts to maximize the Gaussian log-likelihood over all positive (semi-)definite matrices $\Theta \in \mathbb{S}^p_{\succeq 0}$. As we have mentioned previously, there is a sparsity parameter γ that tunes the ℓ_1 penalty in the log-likelihood maximization problem:

$$\Theta^\star = \underset{\Theta \in \mathbb{S}^p_{\succeq 0}}{\arg\max} \left\{ \log\det\Theta - \operatorname{tr}\left(S\Theta\right) - \gamma \left\|\Theta\right\|_1 \right\}. \tag{4.53}$$

We could also think about using graphical LASSO as a way to achieve a sparser airport delay correlation network, with the hope that with enough observations of the delay signal, we could hope to estimate the true set of non-zero elements within Σ^{-1}, i.e., retain the important edges between airports in our correlation network. For

our application, there is a number of implementations that carry out the optimization problem in (4.53), such as the `glasso` package in `R`.

Chapter 5

Airport delays and graph signal processing

The focus of Chapter 4 was on establishing definitions of graph-supported outliers, and interpreting outlier results for various instances of delays in airport networks. While this allows us to categorize a specific day (or hour, or some other time discretization) as being an outlier or not, a natural question is which airport, or group of airports, had delays contributing to the overall delay distribution being categorized as spatially unexpected. This is the focus of Chapter 5, where we use graph signal processing and decomposition via Laplacian eigenvectors to identify the aforementioned groups of airports. The mental picture for the reader is illustrated in Figure 5-1: Given an observation of delays across an airport network, we can *decompose* this airport delay graph signal vector as linear combinations of Laplacian eigenvectors. These eigenvectors can be thought of as commonly-occurring "ingredients" that constitute any given airport network delay distribution. In particular, the ingredient components consist of relative delay differences between airports. We note that, unless otherwise stated, we follow the setup, conventions, and notations adopted in Chapter 4.

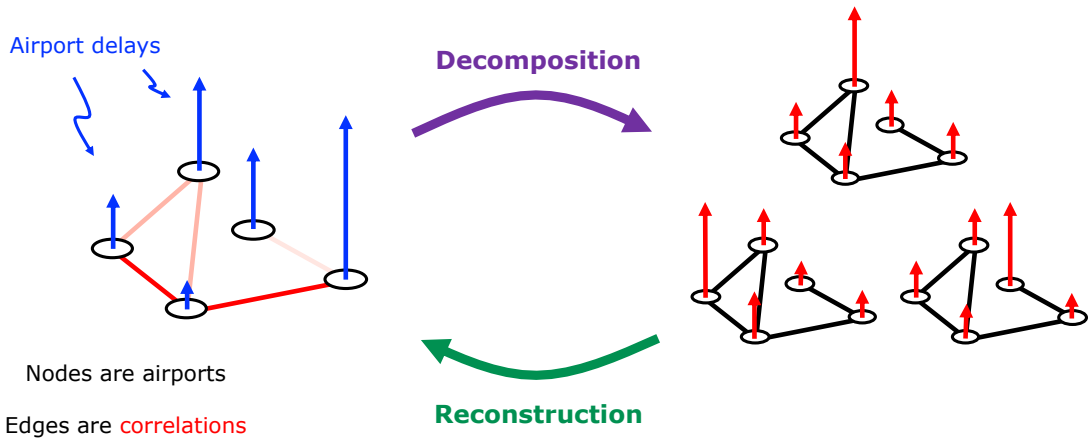


Figure 5-1: Notional representation of our graph setup (*left*), and the decomposition to Laplacian eigenvectors (*right*). *Reprinted from [160].*

5.1 US network-wide spectral analysis

We compute the graph Laplacian matrix $\mathfrak{L}$ from the adjacency matrix as per Definition 1, and retrieve its eigenvalues $(\lambda_1 < \cdots \leq \lambda_{30})$ and eigenvector modes $(v_1, \ldots, v_{30})$. Recall from Chapter 4 that these eigenvector modes are a basis for the space of airport delay signals. Note that since we order the eigenvalue-eigenvector pairs with increasing indices i in accordance to the eigenvalue magnitude, eigenvector modes v_i with larger indices i are said to be more energetic and have a higher TV (Section 4.1).

We retrieve these eigenvector modes for the US network-wide correlation network described in Section 4.4.1. Table 5.1 presents a qualitative description of all 30 eigenvector modes and their corresponding eigenvalues, while Figure 5-2 provides a visualization of the first and last three eigenvector modes, with respect to index i. The key feature of interest in the modes is the sign of eigenvector mode component 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 in 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, i.e., 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). 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.

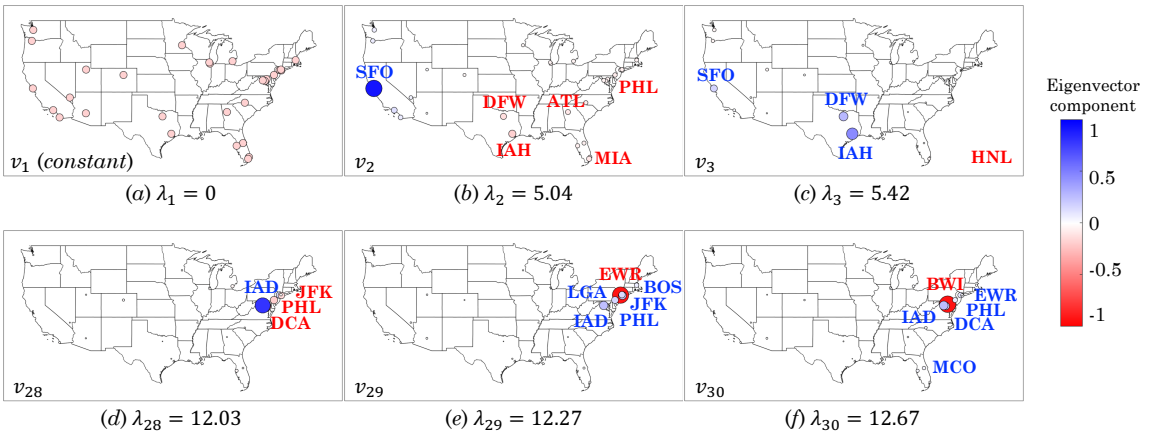


Figure 5-2: Most (v_{30}, v_{29}, v_{28}) and least (v_1, v_2, v_3) energetic eigenvector modes of the system-wide graph Laplacian. *Reprinted from [158].*

Recall that an airport delay graph signal vector for any day can be decomposed into linear combinations of graph Laplacian eigenvectors v_1 through v_{30}. Furthermore, from Definition 3, we can relate graph signal smoothness (i.e., TV) with notions such as the spectral energy of the graph signal. We compute the contribution α_i^2 to the total spectral energy from eigenvector mode v_i, across all 30 eigenvector modes. We calculate this mode-specific spectral energy contribution for each day in 2008-2017. We then plot in Figure 5-3 the contribution for each eigenvector mode v_i, averaged across the 10-year time period. As we noted in Chapter 4, there is a constant mode corresponding to the lowest eigenvalue: This constant mode v_1 accounts for approximately 80% of the total spectral energy on average, but is not operationally interest-

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 5.1: Description of eigenvector modes, delay trends (trends 1 and 2 move in opposite directions), and the airports involved. *Reprinted from [158].*

ing. In order to highlight the differences in the other eigenvector modes, we restrict Figure 5-3 to eigenvector modes v_2 through v_{30}. Recall that if the spectral energy of a particular eigenvector mode comprises a higher percentage of the total energy, then the particular delay pattern corresponding to the eigenvector mode contributes more to the overall delay dynamics. Since the result in Figure 5-3 is aggregated across 2008-2017, this removes any day-to-day variability in spectral energy contributions.

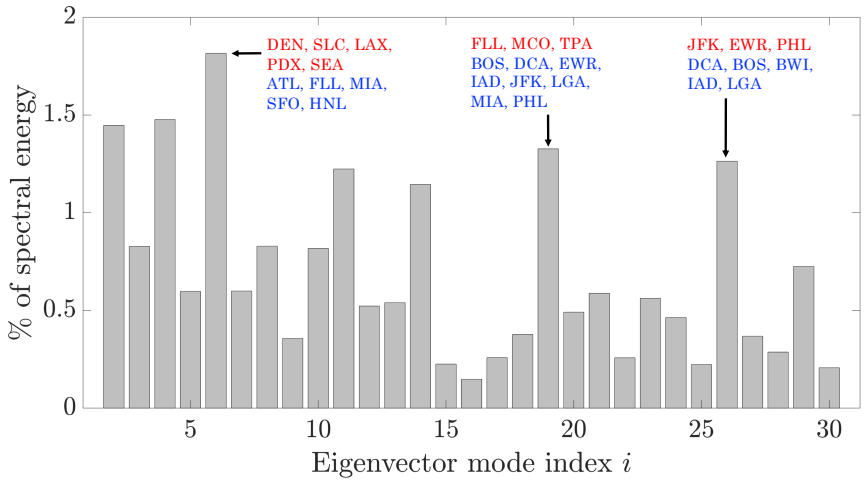


Figure 5-3: Average spectral energy across each system-wide eigenvector mode; eigenvector modes v_2 through v_{30} are shown, with the constant mode v_1 removed for fair comparisons. *Reprinted from [158].*

We interpret four eigenvector modes with relatively high spectral energy contributions based on Figure 5-3. Eigenvector mode v_{26} (contributing approximately 1.2% of the spectral energy) indicates unexpected delay dynamics when the two major New York City airports (JFK and 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 differing from EWR and JFK), but also at a regional level (New York City and Philadelphia differing from Boston and the DC area). Another mode that accounts for just 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, as well as 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 analysis discussed currently), 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 delays at ATL, MSP, and SLC trend opposite to delays at BOS, JFK, LGA, and SEA.

5.2 US airline sub-network spectral analysis

We now shift our attention to the correlation networks for the four US airlines we introduced in Section 4.5. We plot geographically the most energetic and the second-most energetic eigenvector modes for each airline in Figure 5-4. 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 4.1.

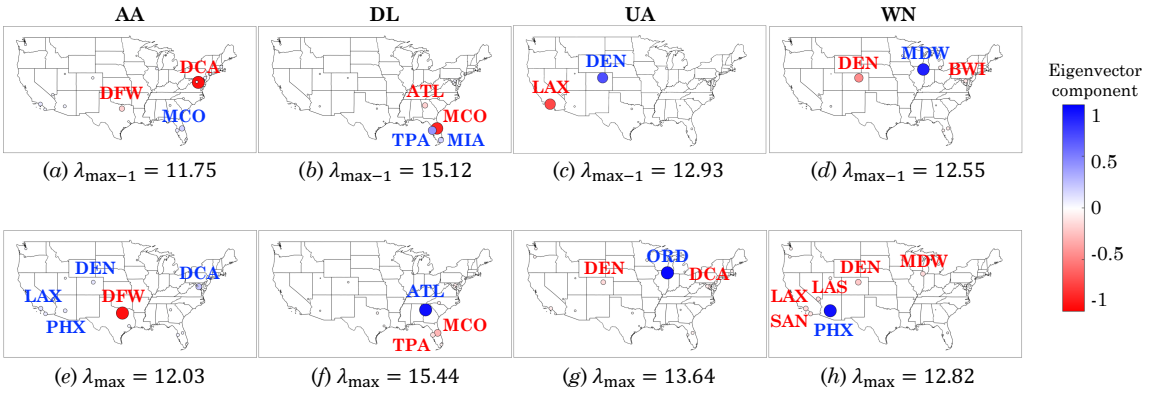


Figure 5-4: Most (λ_{max}) and second-most (λ_{max-1}) energetic eigenvector modes for AA, DL, UA and WN. *Reprinted from [158].*

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 C.3 in Appendix C for an annotated table of US 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.

Connecting back to Chapter 4 in regards to examining specific disruption categories, 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 outliers, 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 regions. However, unlike the system-wide case, the eigenvector modes with the highest spectral energies for airlines (Figure 5-4) 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 we will discuss in Figure 5-5, these high energy eigenvector modes also tend to be frequently triggered.

We plot in Figure 5-5 the distribution of spectral energy across each airline's eigenvector modes. Similar to the system-wide case, the first, constant eigenvector mode 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). Hence, for a fair comparison of the other eigenvector modes, 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 such that $v_{\max}$ contributes to a high percentage of the 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.

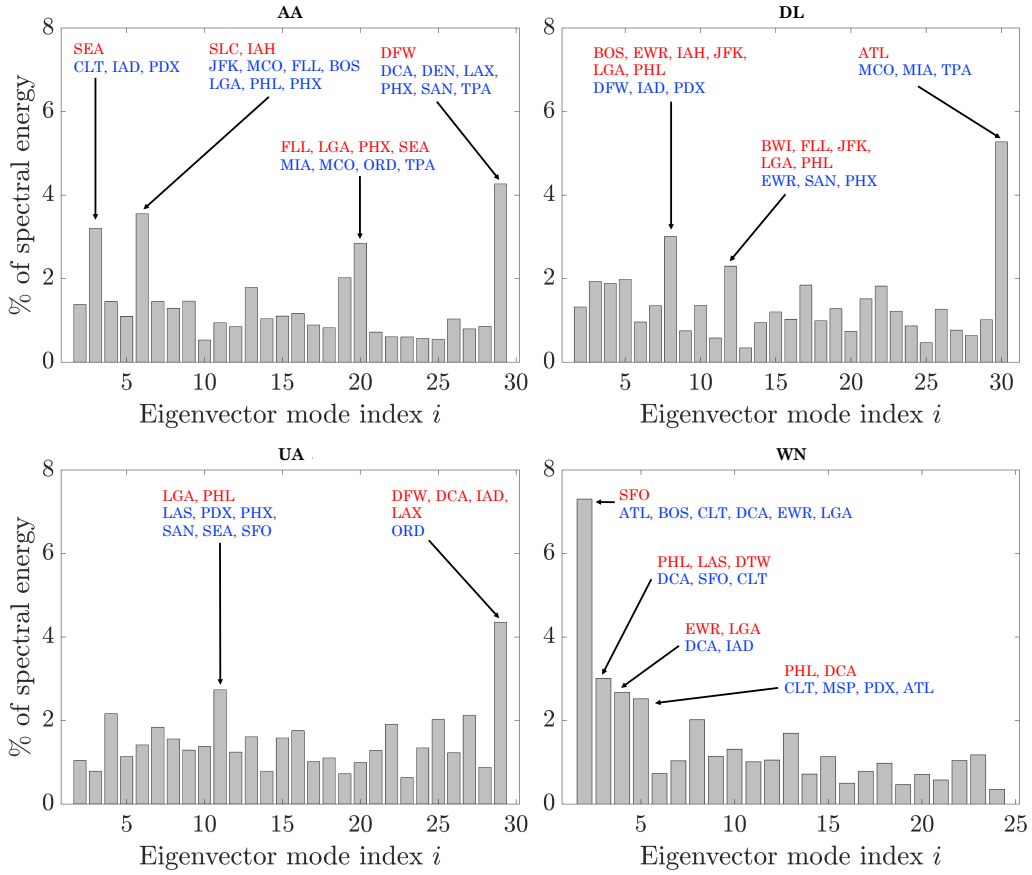


Figure 5-5: Average spectral energy across each eigenvector mode for all four airlines. v_1 is removed for all airlines. *Reprinted from [158].*

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 that specific airline, but also many other airports that tend to be hubs for other network legacy carriers. We will see in Section 5.3 that these lower-energy modes play a dominant role on days when some airline sub-networks are outliers, but others are not.

5.3 Connecting outliers and eigenvector modes: System-wide versus airline-specific

In this discussion, we connect the system-wide outlier results from Section 4.4.2 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. For instance, that 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. As an 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 contain at least four distinct days, ranging from the most common combination of (, , , ,) with 2,817 occurrences, to (×, , ×, ×, ×) with 4 occurrences. We list the tuple statistics in Table 5.2.

We now remark on some of the day-types from Table 5.2 that have interesting operational implications. The first day-type of interest denotes the case where only one airline's sub-network experiences 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 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. Analyzing these days have system resiliency implications, since the unexpected delay distributions were not confined to 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 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 5-6,

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 5.2: Counts of the number of days belonging to each of the 32 tuple types. *Reprinted from [158].*

where we plot the average spectral energy distribution for days when 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 patterns of low-energy eigenvector mode activation may indicate that when unexpected spatial delay distributions arise in an isolated airline's sub-network, the flights operated by unaffected airlines at shared hubs with the affected airline are impacted in an expected manner in terms of spatial delay distributions.

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 unusual 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 5.2 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, as well as 23 days in which the system is an outlier but none of the airlines are. This observation

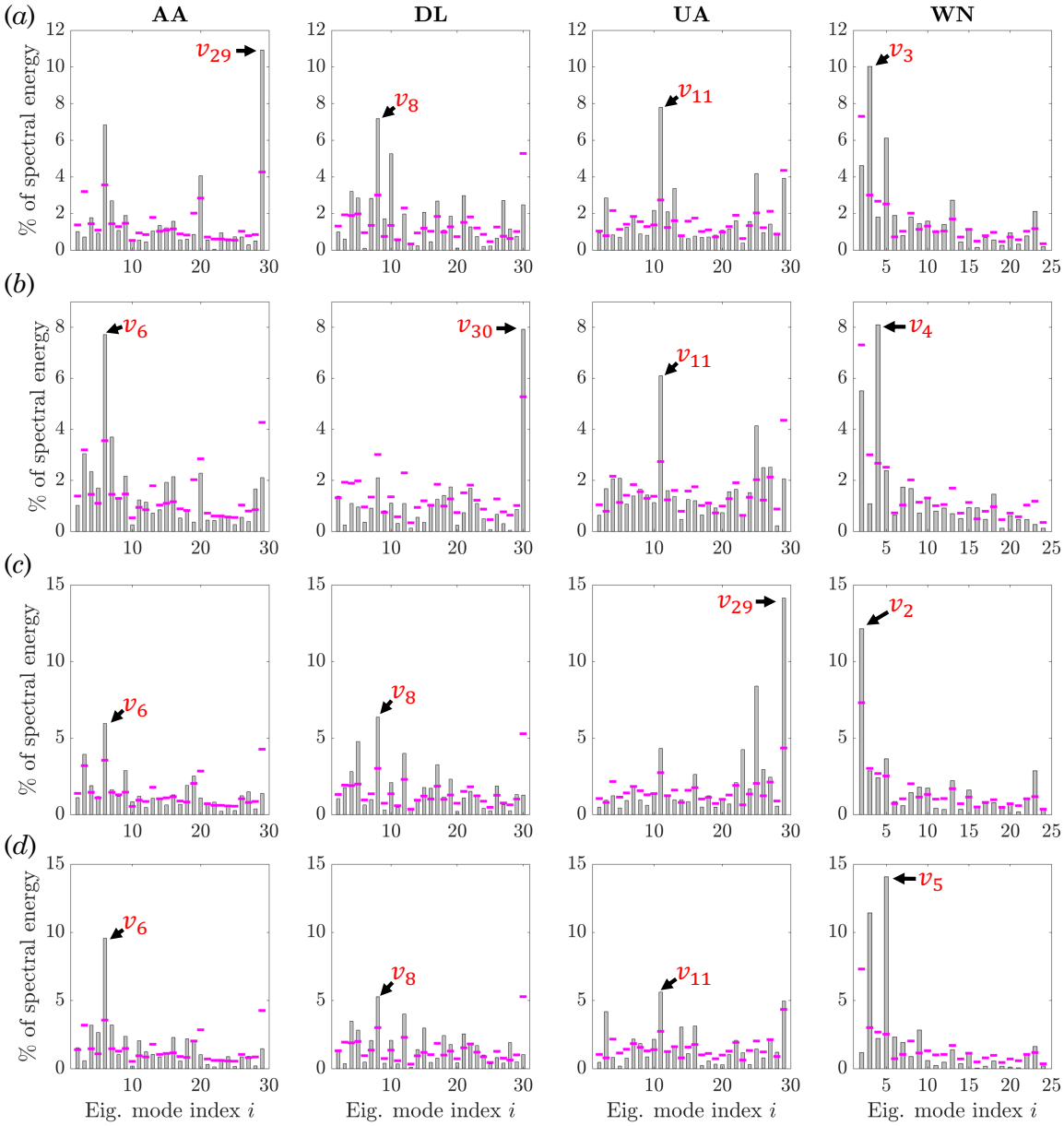


Figure 5-6: Spectral energy averaged across days when only the system and one specific airline (AA, DL, UA, or WN in (a) through (d), respectively) has airport delays that are strong outliers in distribution (gray bars). The airline-specific average spectral energy across the 10-year time frame is shown for benchmarking purposes (magenta bars). Mode v_1 is not plotted in order to highlight the other modes. *Reprinted from [158].*

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 5.2 can be found in Tables C.4 and C.5 in Appendix C.

5.3.1 Summary of US NAS system-wide and airline-specific analysis

In Chapter 4 and continuing thus far into Chapter 5, we applied our graph signal outlier detection techniques to study airport delays in the US. Specifically, 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 with an inventory of days where the delay distributions were unexpected, either spatially, in magnitude, or both, for further performance analyses. 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. For the remainder of Chapter 5, we carry out an eigenvector decomposition analysis for the Chinese airspace, as well as a comparison between the US and Chinese airspace from the perspective of eigenvector modes. We conclude by integrating Chapter 4 and Chapter 5, emphasizing the contextual interpretability of outliers via the eigenvector modes. Our work herein provides 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.

5.4 China airspace spectral analysis

We now shift our focus onto the Chinese airspace, using the same data set as described in Section 4.6.2. We will identify and examine sub-groups of Chinese airports that are commonly implicated in unexpected spatial distributions of delays, providing insights into which airports to strategically monitor the delays at. Specifically, when it comes

to delay magnitudes, one could monitor the total delay in the system in comparison to some threshold to determine whether or not the system is in a high-delay state. Alternatively, an individual airline could monitor its subset of hubs and focus city airports. The selection of airports is less obvious when it comes to the question of monitoring for unexpected spatial distribution of airports. We now characterize the airline-specific *eigenvector modes* corresponding to each airlines' correlation network to get a sense of airline-specific airport subsets whose delays are often implicated in unexpected spatial delay distributions. W

We are interested in the *higher energy* eigenvector modes, which we plot in Figure 5-7 for each of the four Chinese airlines. These modes describe how the delays at particular airports are behaving with respect to delays at other airports. We provide an example of how to interpret eigenvector modes through examining Figure 5-7(a), i.e., the highest-energy mode for Air China (CA). The two airports highlighted in Figure 5-7(a) are PEK and HGH, the former is CA's main hub, whereas the latter is a focus city for CA. While it is not surprising that hubs and focus cities would be implicated in any sort of CA-specific disruptive event, what the eigenvector mode analysis tells us is that an extremely unusual spatial delay distribution in CA's airline network consists of delays at PEK trending *opposite to* delays at HGH. Specifically, it is an unusual situation when delays at PEK are increasing, but the delays at HGH are decreasing, or vice versa. Hence, for CA, monitoring changes in the relative difference between delays at PEK versus HGH is an indicator for whether or not the delays are becoming more (or less) unusually distributed across CA's network. The second-most energetic mode for CA is the situation where PEK and HGH delays are trending opposite to CKG, another focus city for CA.

Note that besides CA, all high-energy modes for the other three airlines are concentrated in the southern and southeastern region of China. This is reflective of the network and hub structures of China Southern (CZ), Hainan (HU), and China Eastern (MU). One interesting observation is that HGH appears in every high-energy mode depicted in Figure 5-7, regardless of the airline. This indicates that regardless of the airline sub-network in consideration, it is likely that delays at HGH are either

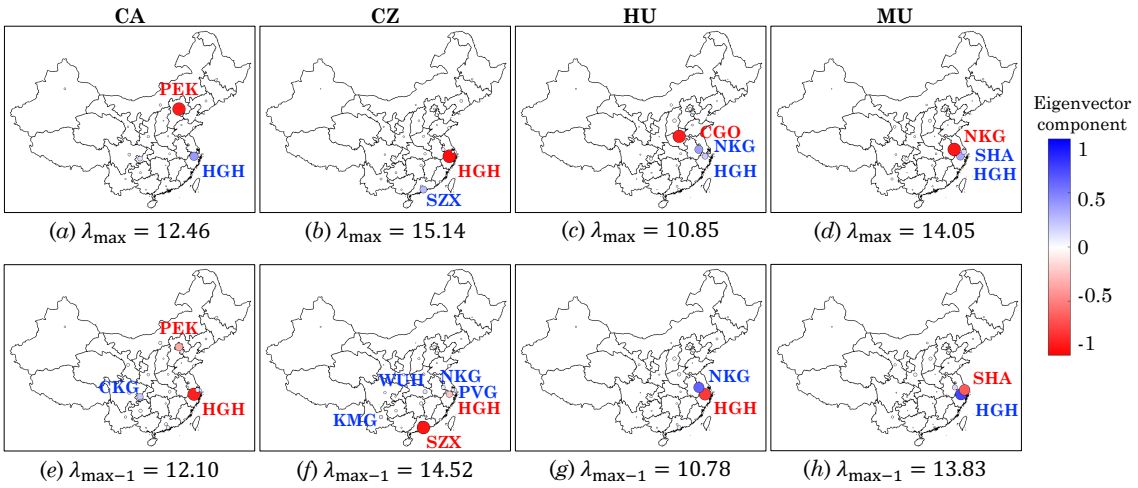


Figure 5-7: Most (a)-(d) and second-most (e)-(h) energetic eigenvector modes for each of the four Chinese airlines. *Reprinted from [160].*

unusually high or unusually low if the network is in a state of having unexpected spatial delay distributions. This might be important at a central traffic management level, particularly since the importance of HGH seems to not be airline-specific. We note that HGH is a rare case where it is a focus city for all four airlines we consider in this study: It is possible that this attribute of being a commonly-shared resource is being reflected in the eigenvector mode analysis. Finally, the airport subsets for HU and MU's high-energy modes are quite closely collocated, whereas some of CZ's airports are quite distributed throughout the southern regions of China.

5.5 US and Chinese airspace analysis

Our work in this section characterizes and compares the spatial delay patterns in the US and mainland China airport networks. There are several factors that motivate our choice of these two regions. In terms of similarities, both networks cover vast geographical extents; for example, the longest intra-region flight for both networks is around 6 hours. Both networks carry a significant volume of passenger and cargo (777 million passengers in the US [34] and 126 million passengers in China during 2018), experience severe delays (20.6% of flights are delayed in the US [36] and 19.9%

of flights are delayed in China during 2018 [52]), and have limited strategic measures such as airport slot controls that mitigate long term demand-capacity imbalances. There are also significant differences in the rate of growth (passenger growth of 4.9% in the US [118] and 10.2% in China [204]), maturity of the infrastructure, seasonal effects and weather patterns, and air traffic management procedures. For example, China has a limited number of fixed airways that are regularly restricted due to military activities [219, 63]. These similarities and differences provide an interesting opportunity for comparisons, in particular using data-driven methods to study the impact of these factors on operational performance.

5.5.1 Delay data and pre-processing

Airport delay data is obtained for the US from the Aviation System Performance Metrics (ASPM) database maintained by the Federal Aviation Administration (FAA), and for China from the Operations Monitoring Center of the Civil Aviation Administration of China (CAAC). We analyze data from 2012-17, and restrict the airport network to 30 airports in both cases. The US airports are chosen based on the FAA Core 30 list, and for China based on the traffic volume (see Table C.6 and Figure B-11 in the appendix for a list and map of all 30 airports for both networks). We eliminate canceled and diverted flights from both data sets and construct a graph signal of airport delays for each day in the 5-year period for both countries. The graph signal at each airport is equal to the total delay at that airport. The total delay for an airport is defined as the sum of the mean inbound and outbound delays in minutes, sampled at at every hourly interval for the day, where a day is defined as a 24-hour period. After the data pre-processing, we obtain $M = 2,192$ graph signals $\mathbf{x} \in \mathbb{R}_{\geq 0}^{30\times 1}$ for both networks. The edge weights a_{ij} in both graphs are computed as the correlation coefficient between the delay signals at airports via (4.1). We note that all the correlations are strictly positive, and the graph has only one connected component. Hence, the graph Laplacian will have only one zero eigenvalue, and all of the other eigenvectors and eigenvalues will be real.

We now analyze the US and China airport delay correlation networks, the eigende-

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

5.5.2 US and China correlation networks

We plot in Figures 5-8 (heatmap) and 5-9 (geographical map) the correlation coefficients for our network of 30 Chinese and US airports, computed from the 5-year data described in Section 5.5.1. The resultant correlation network for China shows a much larger subset of airports exhibiting high pairwise correlations in terms of their total delay time series. This is in contrast with the US, where there are two distinct airport subsets with correlation coefficients higher than the rest of the network: East Coast airports, and to a lesser extent, West Coast airports. Typically, these high correlations are due to geographic proximity, which leads to common weather impacts, combined with operational factors such as traffic flow and airline hub characteristics. For example, airport pairs such as ORD-MDW (US) and SZX-CAN (China) do not have any traffic flows, but are collocated at the same city, thus resulting in high correlations. On the other hand, the delay correlations at airport pairs such as SHA-SZX and BOS-LGA are influenced more by the high volume of shared traffic flows, along with potential geographic factors.

These correlation networks indicate that, compared to the US, the Chinese network has a larger and more geographically diverse set of airports whose delays are closely coupled. We explore this further in the next subsection by interpreting the eigendecomposition of the graph Laplacians corresponding to these correlation networks. The resultant eigenvector modes help us in identifying *specific groups of airports* which contribute to unexpected delay distributions.

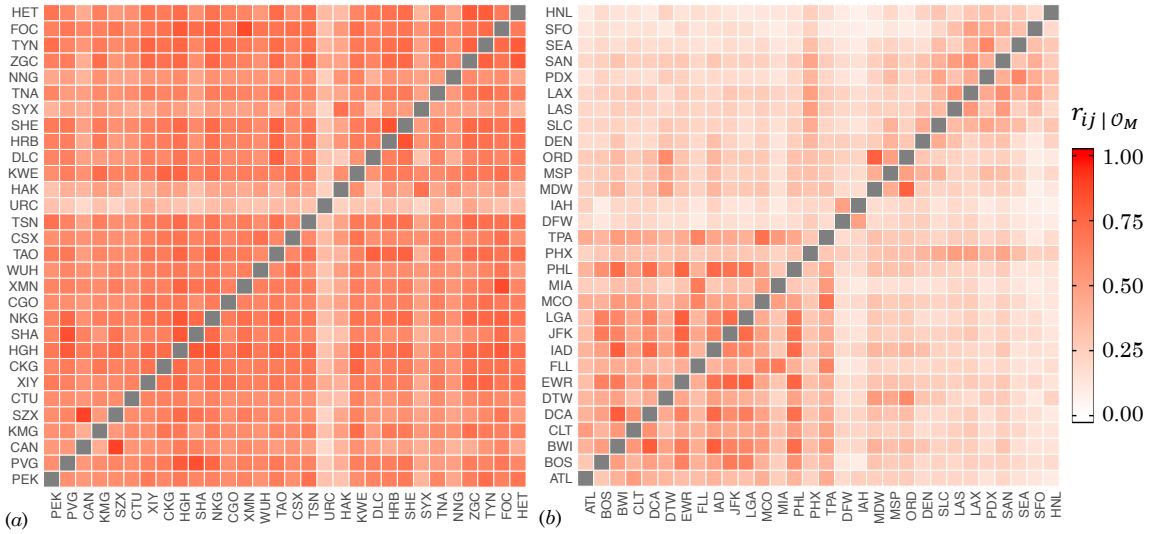


Figure 5-8: Airport delay correlations shown in heatmap format for the (a) Chinese and (b) US airspace.

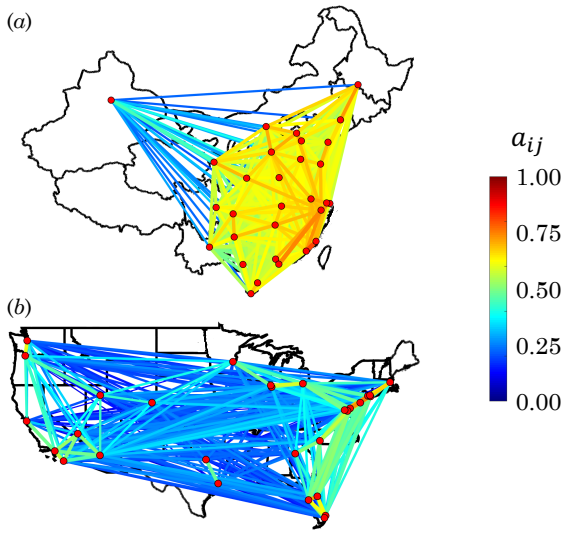


Figure 5-9: Airport delay correlations shown with geographical context for the (a) Chinese and (b) US airspace. Higher correlations are also emphasized with wider lines. Note that HNL is not shown in (b) for simplicity. *Reprinted from [161].* © *2020 IEEE*

5.5.3 US and China eigenvector modes

Recall that the eigenvector modes allows for the decomposition of any airport delay signal vector into linear combinations of these modes, where higher-indexed modes (corresponding to larger eigenvalues) are more "energetic" and result in higher TV. We order the set of 30 eigenvalues for China and the US airport network in ascending order according to their magnitude. Thus, Figure 5-10 provides a geographic depiction of the top 5 most energetic eigenvector modes for both networks.

Eigenvector modes have interesting operational interpretations based on the sign of each of the 30 components in v_i. In Figure 5-10 we represent positive components in blue and negative components in red; the important characteristic is the *difference* in signs between two or more airports, not so much the sign of an airport itself, since any scalar multiple of the eigenvector is also an eigenvector. For example, in mode v_{27} for China, the signs on NKG and TYN are the opposite of TAO and FOC. This mode portrays a scenario where airport delays at NKG and TYN are decreasing, and airport delays at TAO and FOC are increasing, *or vice versa*. This point regarding comparing different signs, and not the sign of one particular airport, is crucial to keep in mind when analyzing these eigenvector modes.

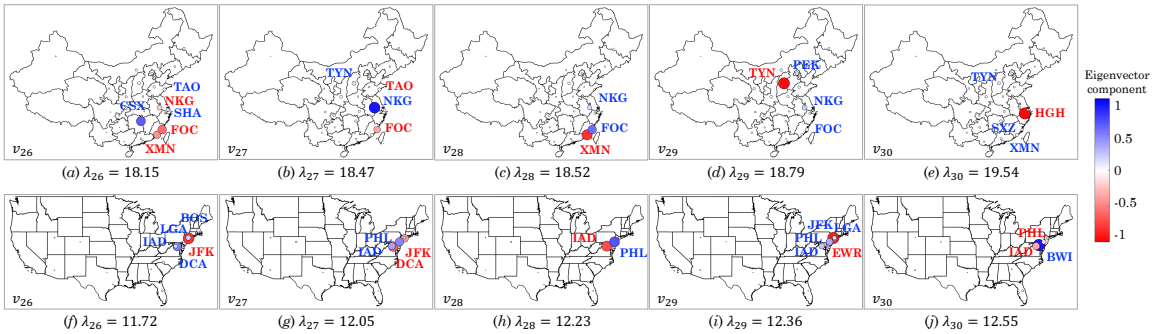


Figure 5-10: Top 5 most energetic eigenvector modes (i.e., v_{26} through v_{30}) of the graph Laplacian for China (a)-(e) and the US (f)-(j). *Reprinted from [161].* © *2020 IEEE*

Visually, the most energetic eigenvector modes for China are significantly different from those of the US (see the top and bottom rows in Figure 5-10) in terms of the geographic concentration of the highlighted airports. Most of the airports in China

are widely distributed across the eastern part of the country. With the exception of PEK in mode v_{29} and SXZ in v_{30}, none of the airports in these high-energy modes are ranked in the top five Chinese airports in terms of traffic volumes. Furthermore, there is only one high energy (and thus, high TV) mode in China (v_{28}) where geographically proximate airports (XMN and FOC) are experiencing opposite delay trends, whereas this scenario occurs for all US eigenvector modes (i.e., East Coast and Mid-Atlantic airports).

It is interesting that our data-driven delay analysis identifies the same airports from another study that identified critical airports for system resilience [268]. Note that such a list of airports is different from those arising out of studies on operational dynamics or traditional network measures such as degree and betweenness centrality. In particular, FOC has low traffic volumes (ranked 27^{th} in 2018), and is not considered to be a central airport in terms of its connectivity. However, it is determined to be critical in a simulation-based study [268] that quantify the resilience of the system. Analogously, our GSP analysis based on empirical data also identified FOC in four of the five high-energy modes as a crucial airport in determining whether the system-wide delay distribution is spatially expected or not.

Furthermore, the high-energy US modes involve airports within the same multi-airport system having opposing delay trends: For example, LGA and JFK in New York City have opposing delay trends in v_{26}. This is not the case in China. As an illustrative example, the two Shanghai airports (PVG and SHA) never have opposing delay trends in the high-energy modes. For the US, we see that the most energetic eigenvector modes correspond to cliques of East Coast airports with delays trending opposite to each other. These include airports serving the same metropolitan area having opposing delay trends; examples of this include modes v_{27} (IAD versus DCA), v_{29} (JFK and LGA versus EWR), and v_{30} (IAD versus BWI). Even if one of these modes has a high spectral energy on a given day, we expect a higher TV for that day due to the relatively larger eigenvalues associated with these modes (Proposition 1). Going one step further, we note that for the US, checking whether the delay distribution of the system is spatially expected is approximately equivalent to checking

if the delay distributions in the East Coast are spatially expected. Similarly, it is likely that the system is an outlier in distribution if the East Coast airports experience an unexpected spatial distribution of delays.

While it may be sufficient to primarily monitor a geographically localized subset of airports in the US to analyze spatial delay distributions, the same cannot be said for the China airport network. We see a lack of geographic consistency in the airports highlighted by the five most energetic eigenvector modes for China. In particular, for China, we see airports as far north as TYN and PEK appearing in some of the modes, along with airports in the southeast such as XMN and FOC. Although this is expected given the correlation networks for the two countries (Figures 5-8 and 5-9), the eigenvector modes provide specific cliques of airports that cannot be identified through a simple ranking of correlation coefficients, since these are inherently limited to pairwise interactions.

5.5.4 US and China spectral energies

Recall from Chapter 4 that the eigenvalue corresponding to an eigenvector mode is a measure of the "graph frequency" of this mode, whereas its spectral energy – specifically, the percent contribution of its spectral energy to the spectrum for an average day, averaged over the entire 10-year data set – is a measure of its *contribution*, or *impact* within the network. We plot these two quantities in Figure 5-11 for both the US as well as the Chinese network. We observe a clear distinction in the magnitude of the eigenvalues between the two countries. The average magnitude of US eigenvalues is 9.17, and they are lower than all but one eigenvalue from the China airport network. The average magnitude for the Chinese eigenvalues is 16.93, indicating that the average TV of delay signals in China is significantly higher than the US.

The differences in terms of the spectral energy distribution is less pronounced between China and the US. More than 80% of the average spectral energy is contained in the constant mode for both countries (80.92% for the US and 87.87% for China) due to the fact that typical days in both networks do not experience significant disruptions or unexpected spatial distributions of delay. Hence, we remove this constant mode for

a nuanced comparison between the more operationally interesting modes v_2 through v_{30} in Figure 5-11. We also note that in the China airport network, there appears to be a couple of modes that dominate the average eigendecomposition, whereas the distribution is more even in the US.

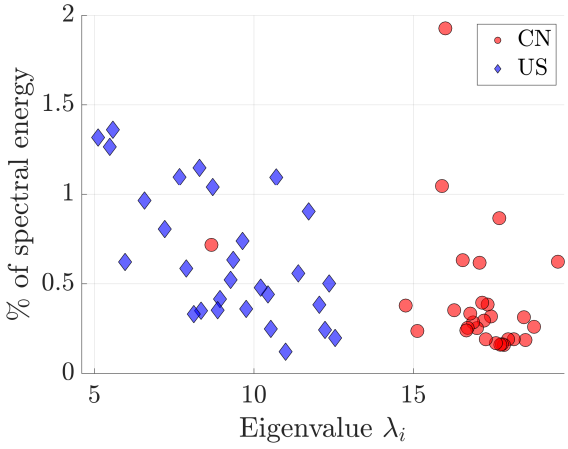


Figure 5-11: Percentage of spectral energy per eigenvector mode versus associated eigenvalue for the airport delay graph of the US and China (CN). *Reprinted from [161]. © 2020 IEEE*

5.5.5 Identifying outliers in US and China

For each day with an associated airport delay graph signal $\mathbf{x}$, we can compute the TD and TV, and visualize it as a point $(\|\mathbf{x}\|, \mathrm{TV}(\mathbf{x}))$ on a *TV versus TD plot*. Using the skew-adjusted IQR method for detecting outliers detailed in Section 4.3, we compute the outlier in scale (OIS), weak outlier in distribution (OID), and strong OID bounds for the China and US airport networks. We present the TV-TD plots for China and the US in Figure 5-12, with OIS, weak OID, and strong OID bounds demarcated.

Note that the observed delay within the China airport network is higher than the US, resulting in a significantly higher TV for China due to the quadratic-dominated relationship between TD and TV. However, since the outlier bounds are trained with respect to each country's data, the outlier statistics can be compared across the two countries. We summarize the outlier statistics in Table 5.3. We observe that in terms of the TD, which is the simplest measure of the severity of a disruption, the number

of OIS in the US is higher than in China. This indicates that even though the TD may be higher on average for China compared to the US, there are more days in the US where the TD was *unexpectedly* high or low. The statistics for OID highlight the importance of using the strong OID in lieu of the weak OID definition. While the US has more weak OID than China, the reverse is true when the strong OID bounds are used, which takes into account the fact that TV grows quadratically with TD. The weak OID bounds, even though they are easier to compute and use, is not a very consistent metric to identify spatial distribution outliers, since its conclusions may even be contrary to ones obtained from the strong OID definition. We conclude that the Chinese airspace not only incurs more severe delays than the US, but the delays also tend to be spatially distributed at unexpected sets of airports. Given the lack of geographic consistency in the higher-energy eigenvector modes for China, this indicates that the unexpected spatial delay distributions are also likely to be geographically dispersed throughout the airport network of China.

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

Table 5.3: Number of outlier days out of 2,192 days (2012-2017) for China and the US, categorized by outlier type. *Reprinted from [161]. © 2020 IEEE*

5.5.6 Monthly distribution of outliers in US and China

We plot the total number of strong OID days in each month for China and the US in Figure 5-13 to analyze seasonal and temporal patterns. While one might presume that there are more unexpected spatial delay distributions in the summer months due to disruptions such as thunderstorms, we found that May, June, and July actually contain the least number of strong OID days for both countries. The temporal distribution of outliers from Figure 5-13 suggests that unexpected spatial delay distributions are much more common in the winter, at least from a system-wide

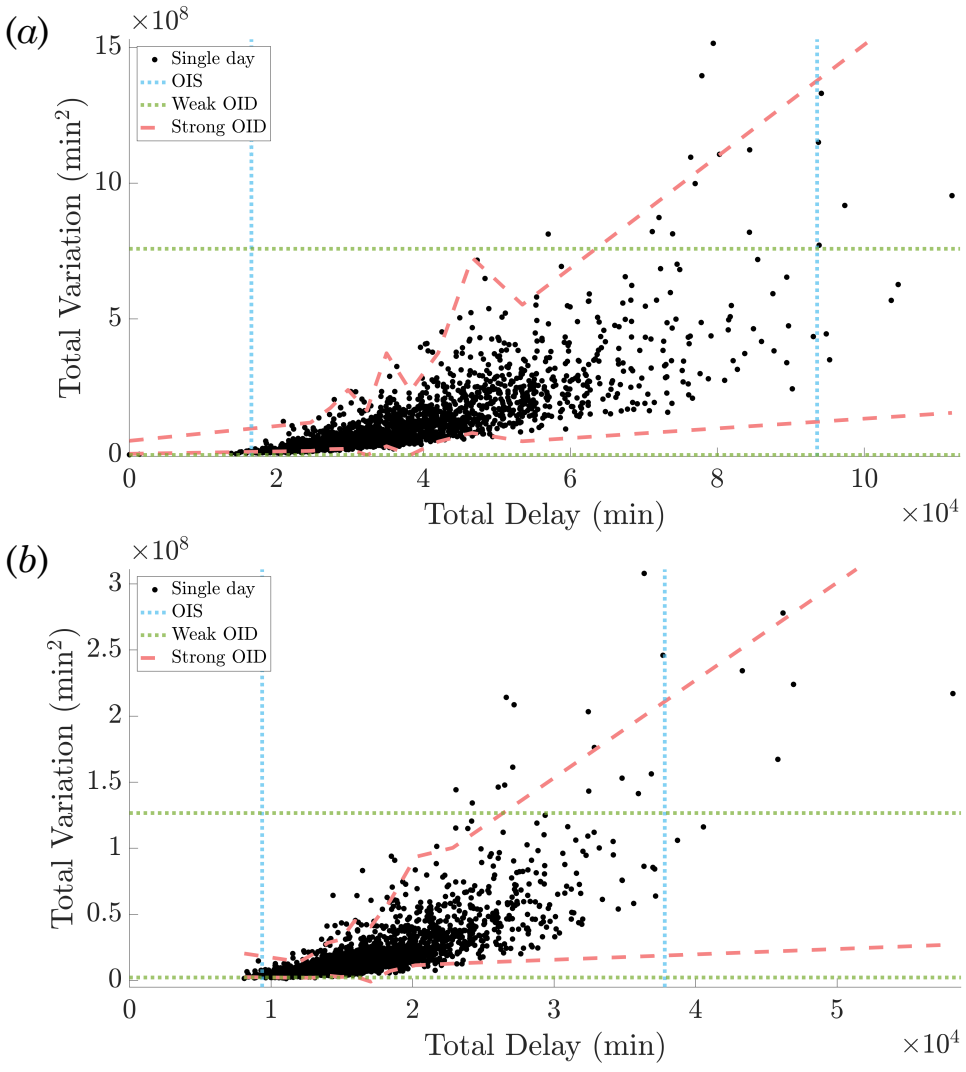


Figure 5-12: TV versus TD for all days in 2012-2017 for China (a) and the US (b), with outlier in scale (OIS) and distribution (OID) bounds marked. *Reprinted from [161]. © 2020 IEEE*

perspective for both countries.

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

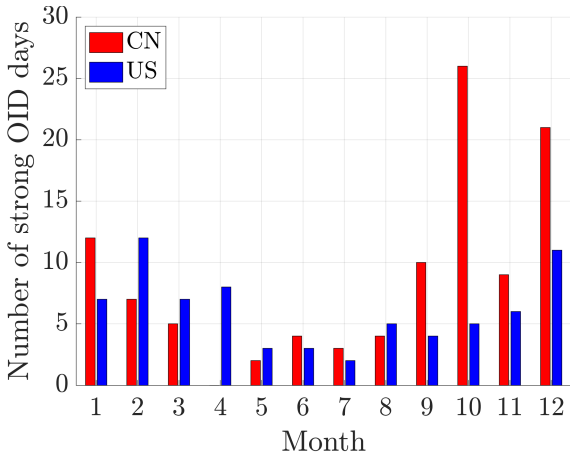


Figure 5-13: Comparison between the number of strong OID days in the US and China (CN) across all months. *Reprinted from [161]. © 2020 IEEE*

5.6 Unified framework of outlier detection

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

Our interpretation framework can be thought of as a workflow: We visualize this workflow in Figure 5-14. Note the dual layer of interpretations available through this framework: The first layer of interpretability distinguishes between different *types* of outliers, i.e., strong OID versus OIS. We do not interpret weak OID days, as we previously demonstrated the tighter performance of strong OID bounds. The second layer of interpretability allows for *operational* insights to be derived from the GFT and spectral modes. In particular, if the input day $\mathbf{x} \in \mathbb{R}_{\geq 0}^{N \times 1}$ is determined to be a strong OID via the appropriate bounds, then we can compute the spectrum of the signal $\{\alpha_1^2, \ldots, \alpha_N^2\}$. Using the spectral energy percentage contained in each eigenvector mode, we can retrieve the dominant eigenvector modes by selecting those with significant energy contributions. We can then identify subsets of airports implicated in a particular strong OID day, and use those airports as the basis for an in-depth operational analysis.

5.6.1 Outlier interpretation examples for China

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

We select 6 strong OID days from the China airport network and present them in Table 5.4 along with their TD, TV, and the fraction of spectral energy $\alpha_{\max}^2 / \sum_i \alpha_i^2$

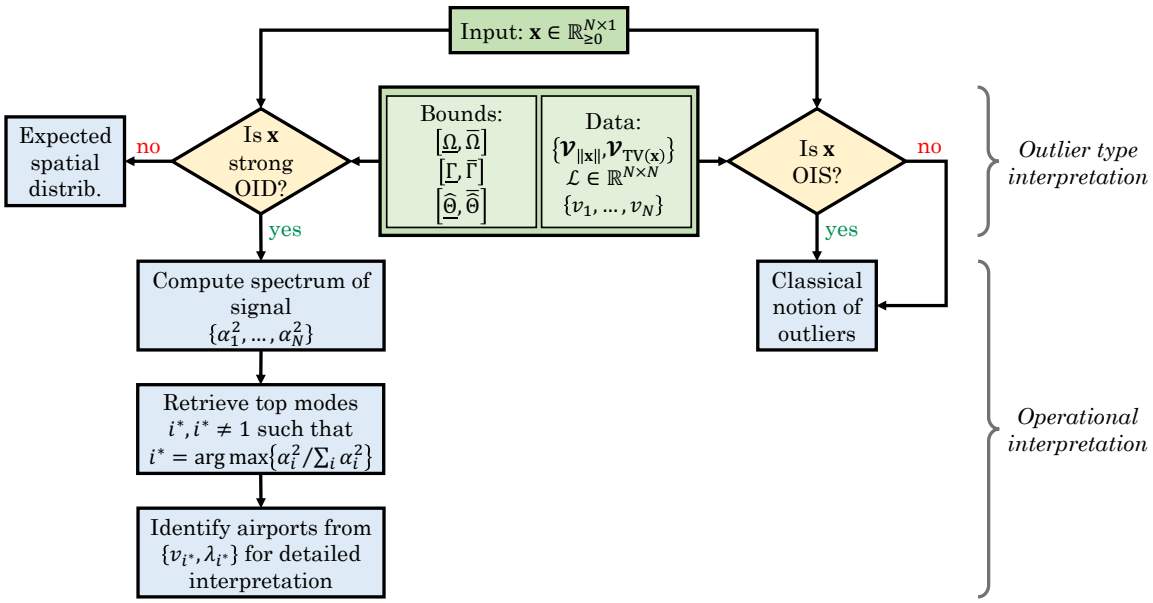
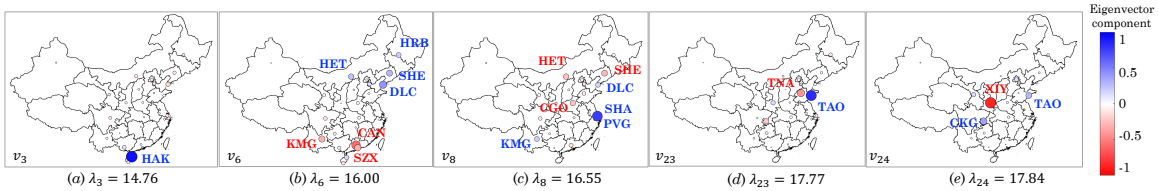


Figure 5-14: Flowchart depiction of our framework for the two-step process of interpreting strong OID days in an air transportation setting. *Reprinted from [161]. © 2020 IEEE*

contributed by the dominant eigenvector mode $v_{\max}$. We plot the relevant modes in Figure 5-15.



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

Figure 5-15: Dominant eigenvector modes observed during the six strong OID days in the Chinese airspace. *Reprinted from [161]. © 2020 IEEE*

- *6/10/2017 and 9/25/2017:* The dominant eigenvector mode v_8 shown in Figure 5-15(c) indicates that these two days were classified as outliers because the two Shanghai airports (SHA and PVG) had significantly higher delays than airports in the north, specifically HET, SHE, and CGO. This is operationally interpretable as heavy rain and thunderstorms affected operations on both days in the Shanghai area. In particular, September 25 involved more than 100

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

Table 5.4: The six strong OID days in the Chinese network with the highest TD. The date is given in month/date/year format. *Reprinted from [161].* © *2020 IEEE*

cancellations at PVG, and deteriorating conditions on June 10 resulted in a 50% capacity reduction at SHA and PVG.

- *12/29/2012:* The dominant mode v_{23} indicates that this day was classified as an outlier because of two geographically proximate airports (TAO and TNA) having substantially different delay magnitudes. Heavy snowfall at TAO (resulting in 145 canceled flights) with relatively no noticeable impact at TNA provides an operational interpretation for this unexpected delay distribution.

- *1/18/2012:* The eigenvector mode v_{24} highlights unexpected spatial delay distributions where delays at TAO and CKG move in opposite directions to XIY. Severe ice accumulation and fog at XIY forced a major airline, China Eastern, to cancel 186 flights. Interestingly, the delays were contained at XIY and did not spread to the other China Eastern focus city of TAO. This is operationally unexpected, as delays typically propagate within an airline's sub-network during disruptions at a major hub.

- *6/21/2012:* Eigenvector mode v_6 indicates that airports in the north – HET, HRB, SHE, DLC – had delay magnitudes opposite to airports in the south (KMG, CAN, SZX). This resulted in an unexpected delay distribution attributable to geographically-localized disruptions. Specifically, adverse weather south of the Yangtze river resulted in flood emergencies, disrupting airport and airspace operations for all of the southern airports. The northern airports

remained largely unaffected, and experienced low delays. Again, this is unexpected given the typical tendency of delays to propagate and spread.

- *1/28/2012:* The activation of eigenvector mode v_3 indicates that the delays at HAK were significantly higher (or lower) than other airports in China, resulting in the classification of this day as a strong OID. Upon closer inspection, we identify two operational factors that may have contributed to such a delay pattern. First, January 28, 2012 was the last day of the Spring Festival holidays in China, resulting in higher-than-normal scheduled flights at HAK (61,698 passengers on January 28 versus 36,142 on January 22). Second, heavy fog on January 28 exacerbated the already-strained demand-capacity imbalance and led to severe flight delays at HAK.

5.7 Summary of US and China graph signal analysis

In this comparative analysis of the US and Chinese airspace, we (1) identified critical subsets of airports in the US and Chinese airport networks that should be monitored for unexpected spatial delay distributions; (2) observed higher baseline variability in spatial delay distributions in China as compared to the US; (3) compared OIS, weak OID, and strong OID outlier statistics between China and the US, as well as examined temporal trends; and (4) demonstrated the theoretical as well as operational interpretability of our outlier identification results. In particular, the outlier interpretability framework unifies the outlier definitions from Chapter 4 with the eigenvector mode analyses in Chapter 5.

Chapter 6

Low-Dimensional Models for Airspace Disruptions and Recoveries

Recall from, for example, Section 4.4.4, the large variety of disruptive events (weather, equipment outage, or congestion, etc.) that conspire to create an imbalance between system capacity and demand, leading to flight delays and cancellations. These cycles of disruptions and subsequent recoveries can be viewed from a dynamical systems perspective: Exogenous inputs *disrupt* the system, inducing delays and inefficiencies from which the system eventually *recovers*. In Chapter 6, we study these disruption and recovery cycles through a state-space representation that captures the severity and spatial impact of airport delays. We also explore cancellations as a performance metric, which we have thus far overlooked in Chapter 4 and Chapter 5.

6.1 Clustering analysis of disruption-recovery trajectories (DRTs)

Disruptions and subsequent recoveries in the air transportation system vary in their geographical extent (number of airports affected), intensity (severity of resultant de-

lays and cancellations), and duration (ranging from hours to days). The inherent variability of factors such as weather in the operating environment, along with the complex interconnectivity of the system, make it difficult to extract actionable insights from past events. The work in Chapter 6 focuses on formalizing and analyzing disruptions and recoveries in the air transportation system by leveraging the spectral and graph signal processing techniques from Chapter 4.

In particular, using US airport delay data from 2008-2017, we first identify representative disruption and recovery cycles. These representative cycles provide insights into the common operational patterns of disruptions and recoveries in the system. We also relate these representative cycles to specific off-nominal events such as airport outages, and elucidate the differing disruption-recovery pathways for various off-nominal events. Finally, we explore temporal trends in terms of when and how the system tends to be disrupted, along with the subsequent recovery.

6.2 Defining and clustering disruption-recovery trajectories: Data and methodology

6.2.1 Data sources and processing

We use hourly airport delay data for the years 2008 to 2017 from the FAA ASPM database for our analysis, focusing on the US Core 30 airports (see Figure B-11(b) in Appendix B for a geographical overview), which were responsible for 72% of all US enplanements in 2017 [85]. For each hour in this data set, we construct a graph with the airports as nodes, and the signal at each node representing the total average arrival and departure delay experienced by all scheduled flights in that hour at that specific airport. The adjacency matrix for these graphs are the resultant hourly 30×30 correlation matrices evaluated by considering the hour-by-hour subsets of the 10-year airport delay data set. Thus, there are 24 unique adjacency matrices, corresponding to each hour of the day. For each graph, the graph Laplacian is the difference between the degree and the adjacency matrix (Definition 1). With the

hourly graph Laplacian, we can compute the total variation (TV) for each hour. The graph signal vector $\mathbf{x}(t) = [x_{i,t}] \in \mathbb{R}^{30\times 1}$ represents the delay $x_{i,t}$ at time t for airport i. The total delay (TD) is the 1-norm of $\mathbf{x}(t)$, and the total variation is given by the quadratic form $\mathrm{TV}(\mathbf{x}(t)) = \mathbf{x}(t)^\top \mathcal{L}\mathbf{x}(t)$, where $\mathcal{L}$ is the Laplacian for the hour-of-day of time t. Formally, each hour t is indexed by hours $\{1, \ldots, 24\}$, and we use the corresponding graph Laplacian $\{\mathcal{L}_1, \ldots, \mathcal{L}_{24}\}$ in the TV computation. For brevity of notation, we drop the hour index on the graph Laplacian when the context is clear. These spectral graph-theoretic notions follow directly from their usage in Chapter 4. Recall our intuition built up from Chapter 4 that the TV is a measure of signal *smoothness* with respect to the underlying graph. In the case where airport delays are node-supported signals and correlation coefficients are the edge weights, a high TV value indicates an imbalance between delays at airports with historically highly-correlated delays.

6.2.2 Formal definition of disruption-recovery trajectories

We represent disruptions and the subsequent recoveries as *disruption-recovery trajectories* (DRTs) in the TV-TD state space. We define a DRT $\mathbb{T}_{t^*}$ to be a chronologically ordered set of TD and TV values, capturing the evolution of the magnitude (TD) and spatial distribution (TV) of airport delays. DRTs project the state of the system in a qualitatively interpretable manner. For example, we can assess how the system evolves in terms of airport delays from t_1 to t_2 by looking at the progression of the TV-TD state space trajectory $(\|\mathbf{x}(t_1)\|, \mathrm{TV}(\mathbf{x}(t_1))) \rightarrow (\|\mathbf{x}(t_2)\|, \mathrm{TV}(\mathbf{x}(t_2)))$. We also note that, unlike Chapter 4 where we used day-by-day time intervals, we use hour-by-hour time intervals in this DRT analysis. The day-by-day view is not granular enough to describe the delay dynamics at the resolution we are interested in for Chapter 6.

We further divide the TV-TD space into *regions* according to regimes that are operationally interesting. We illustrate one potential (disjoint) partition that we utilize for our analysis in Figure 6-1(a). This partition distinguishes a nominal region (i.e., nominal TD and TV levels), a high-TD region (i.e., delay magnitudes are high),

and an unexpected TV region (i.e., the spatial distribution of delays is unexpected). As we derived in Chapter 4, we know that unexpected spatial distributions correspond to regions with very high or very low TV. In particular, given an observation $\mathbf{x}(t)$ belonging to a certain hour, we can compute the bounds on the TV for identifying delay distributions that are spatially perturbed at time t, denoted as $\left[\underline{\widehat{\Theta}}, \overline{\widehat{\Theta}}\right]_{\text{hour}(t)}$, using the methods proposed in Chapter 4. Recall from Chapter 4 that there are several different bounds we can define, use, and construct from data: Bounds for outliers in scale (OIS) distinguish outliers by the delay magnitude (i.e., the TD), whereas bounds for outliers in distribution (OID) use spatial characteristics of the delay graph signal as measured by the TV. We also re-emphasize that OID bounds can be further refined into weak and strong OID bounds, where the latter takes into account the realized delay magnitude. To complete the partition we use in this analysis, we also compute a delay threshold $f_{\text{TD}}^{\text{hour}(t)} \in \mathbb{R}_{\geq 0}$ for each hour to identify delay distributions that have a very high magnitude of TD. The regions in Figure 6-1(a) are defined as follows:

Definition 1 ($\mathcal{N}$ Region (nominal)) *The $\mathcal{N}$ region is characterized by* $\text{TV}\left(\mathbf{x}(t)\right) \in \left[\underline{\widehat{\Theta}}, \overline{\widehat{\Theta}}\right]_{\text{hour}(t)}$ *(i.e., spatial distribution is nominal), and* $\|\mathbf{x}(t)\|$ *is less than* $f_{\text{TD}}^{\text{hour}(t)}$ *(i.e., the magnitude of delay is not abnormally high).*

Definition 2 ($\mathcal{S}$ Region (scale)) *The $\mathcal{S}$ region is characterized by* $\text{TV}\left(\mathbf{x}(t)\right) \in \left[\underline{\widehat{\Theta}}, \overline{\widehat{\Theta}}\right]_{\text{hour}(t)}$ *(i.e., the spatial distribution is nominal), but* $\|\mathbf{x}(t)\|$ *is greater than* $f_{\text{TD}}^{\text{hour}(t)}$ *(i.e., the magnitude of delay is currently elevated).*

Definition 3 ($\mathcal{D}$ Region (distribution)) *The $\mathcal{D}$ region is characterized by* $\text{TV}\left(\mathbf{x}(t)\right) \notin \left[\underline{\widehat{\Theta}}, \overline{\widehat{\Theta}}\right]_{\text{hour}(t)}$ *(i.e., the spatial distribution of delays is unexpected).*

Algorithm 3 describes a method for constructing an operationally-significant DRT $\mathbb{T}_{t^*}$, given the 3-region partition from Figure 6-1(a). Specifically, it *anchors* a DRT at a particular time index t^* such that $\left(\|\mathbf{x}(t^*)\|, \text{TV}\left(\mathbf{x}(t^*)\right)\right) \in \mathcal{D}$. Algorithm 3 is $O(T)$, where T is the total number of hours, and it constructs t^*-anchored DRTs forward in time. This algorithm identifies minimal-length trajectories that have at

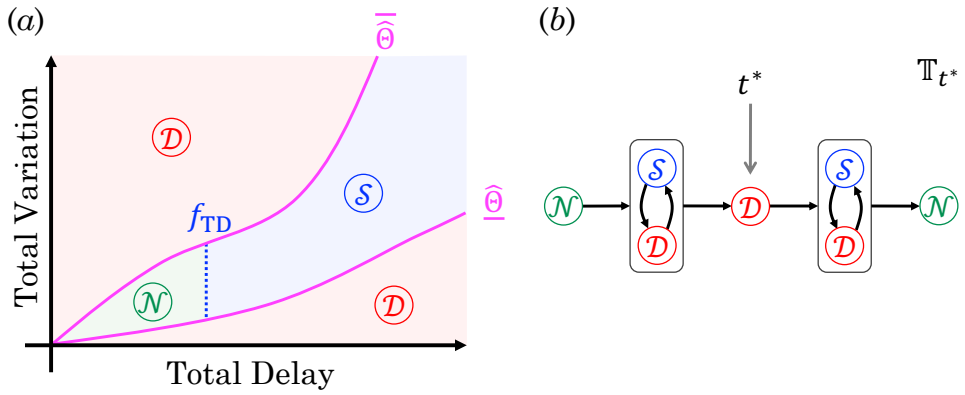


Figure 6-1: (*a*) A disjoint partition of the TV-TD state space into three regions; (*b*) Schematic representation of a disruption-recovery trajectory $\mathbb{T}_{t^*}$ constructed via Algorithm 3 anchored in time at t^*. *Reprinted from [159].*

least one state in the region with unexpected spatial delay distributions, i.e., region $\mathcal{D}$, as well as having start and end states in region $\mathcal{N}$.

Algorithm 3 Constructing DRTs given a 3-region disjoint decomposition of the TV-TD state space.

Input: Labeled states indexed by time $t \in [0 : \Delta t : T]$, where $\Delta t = 1$ hour; Region labels $\mathcal{R}(t) \in \{\mathcal{N}, \mathcal{S}, \mathcal{D}\}$

Output: Set of DRTs $\mathbb{T}_{t^*} \in \mathbb{T}$

$t_s \leftarrow \emptyset$; $\mathbb{T} \leftarrow \emptyset$; $\mathbb{1}_{\text{DRT}} \leftarrow$ FALSE
for $t \in [0 : \Delta t : T]$ **do**
  if $\mathcal{R}(t) = \mathcal{N}$ **then**
    $t_s \leftarrow t$
  end
  if $\mathcal{R}(t) = \mathcal{D} \wedge t_s \neq \emptyset$ **then**
    $\mathbb{1}_{\text{DRT}} \leftarrow$ TRUE
    $t^* \leftarrow t$
  end
  if $\mathcal{R}(t) = \mathcal{N} \wedge \mathbb{1}_{\text{DRT}} =$ TRUE **then**
    $\mathbb{T}_{t^*} := \{(\|\mathbf{x}(\tau)\|, \text{TV}(\mathbf{x}(\tau))) \mid \tau \in [t_s, t]\}$
    $t_s \leftarrow t$
    $\mathbb{1}_{\text{DRT}} \leftarrow$ FALSE
  end
end

6.2.3 DRT nomenclature

Given the many moving pieces within the DRT object itself, we will take some time to establish a common nomenclature for referring to and working with DRTs. For the remainder of Chapter 6, we will adhere to this nomenclature.

The *state* at time t is the TD and TV of the system, i.e., the non-negative tuple $(\|\mathbf{x}(t)\|, \mathrm{TV}(\mathbf{x}(t)))$. We refer to a sequence of consecutive states as a *trajectory*, and a trajectory of length 2 (i.e., consisting of two consecutive states) as a *maneuver*. A DRT of length N consists of $N-1$ *maneuvers* between N consecutive states. The TV-TD state space is partitioned into *regions* $\mathcal{N}$, $\mathcal{S}$, and $\mathcal{D}$. Hence, every point on the TV-TD space (i.e., every state) belongs to one of three regions. A *transition* is a maneuver where the two consecutive states are in different regions. Note that a DRT is the minimal trajectory (shortest-length trajectory) that starts and ends in region $\mathcal{N}$ and contains at least one state in region $\mathcal{D}$. We further classify each maneuver into two categories, depicted in Figure 6-2: A *symbiotic* maneuver is one in which both TD and TV are increasing, or that both are decreasing. On the other hand, a *trade-off* maneuver is one in which the TV and TD change in opposite directions. Symbiotic maneuvers indicate pure disruptions or recoveries, in the sense that both measures are increasing, or both are decreasing. Trade-off maneuvers are more nuanced, as one quantity is recovering at the detriment of the other. For example, a trade-off maneuver could indicate that although the system delay is decreasing, its spatial variability is increasing.

6.2.4 Formalizing disruptions and recoveries segments

In order to define disruptions and recoveries, we take into account the maneuver type (i.e., symbiotic versus trade-off) and whether or not a transition has occurred. For a DRT of length N, we define the *start of a disruption* to be a transition out of $\mathcal{N}$, and the *end of a recovery* to be a transition into $\mathcal{N}$. Among the remaining maneuvers, symbiotic maneuvers with increasing TV and TD are defined as a *disruption segment*, and symbiotic maneuvers with decreasing TV and TD are defined as a *recovery*

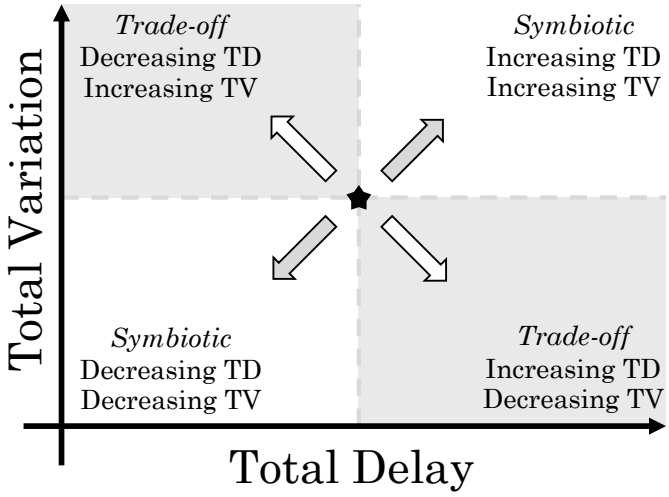


Figure 6-2: Symbiotic and trade-off maneuvers in TV-TD space. The star at the center indicates the current state. *Reprinted from [159].*

segment. Trade-off maneuvers inherit disruption or recovery classifications from the previous maneuver. The four possible trade-off maneuvers are defined as follows:

i. *Disruption-in-TV segment:* Increasing TV and decreasing TD following a disruption.

ii. *Disruption-in-TD segment:* Increasing TD and decreasing TV following a disruption.

iii. *Recovery-in-TD segment:* Increasing TV and decreasing TD following a recovery.

iv. *Recovery-in-TV segment:* Increasing TD and decreasing TV following a recovery.

Trade-off maneuvers are, by definition, a recovery along one axis and a disruption along the other. Our convention therefore assumes that a trade-off maneuver predominantly follows the trend of the preceding maneuver.

6.2.5 DRT clustering features

We cluster DRTs using twelve features that capture various operational characteristics:

DRT length: This feature, denoted by $|\mathbb{T}_{t^*}|$, captures the duration of a DRT. Note that since our time discretization is in hours, accordingly $|\mathbb{T}_{t^*}|$ will have units of hours. The minimum DRT length is 3 hours, since the shortest DRT is given by $\textcircled{\mathcal{N}} \to \textcircled{\mathcal{D}} \to \textcircled{\mathcal{N}}$.

Duration in $\textcircled{\mathcal{S}}$ and $\textcircled{\mathcal{D}}$ regions: These features represent the number of hours in a DRT during which the system is either experiencing high delays (region $\textcircled{\mathcal{S}}$) or unexpected spatial delay distributions (region $\textcircled{\mathcal{D}}$). Note that by construction, the minimum number of states in region $\textcircled{\mathcal{D}}$ is 1.

Average TD and TV intensity: For each DRT, we calculate the average TV and TD, and normalize them by their respective maximum values. The resultant features reflect the *intensity* in terms of the magnitude or spatial distribution of delay. A DRT where every hour attains a TD and/or TV value close to the maximum is considered to be more intense, with TD and/or TV intensity values close to 1. Explicitly, we have that the average TD intensity $\xi_{\text{TD}}(\mathbb{T}_{t^*})$ and average TV intensity $\xi_{\text{TV}}(\mathbb{T}_{t^*})$ are given by:

$$\xi_{\text{TD}}(\mathbb{T}_{t^*}) = \frac{\frac{1}{|\mathbb{T}_{t^*}|}\sum_{i=1}^{|\mathbb{T}_{t^*}|}\|\mathbf{x}(t_i)\|}{\max_{i=1,\ldots,|\mathbb{T}_{t^*}|}\|\mathbf{x}(t_i)\|}, \qquad \xi_{\text{TV}}(\mathbb{T}_{t^*}) = \frac{\frac{1}{|\mathbb{T}_{t^*}|}\sum_{i=1}^{|\mathbb{T}_{t^*}|}\mathbf{x}(t_i)^\top \mathcal{L}\mathbf{x}(t_i)}{\max_{i=1,\ldots,|\mathbb{T}_{t^*}|}\mathbf{x}(t_i)^\top \mathcal{L}\mathbf{x}(t_i)}.$$

Signed enclosed area: Figure 6-4 illustrates how in some DRTs, the disruption phase is characterized by higher spatial variability in delays, whereas in others, the spatial variability is higher during recovery. We use the *signed* enclosed area of a DRT as a feature:

$$\frac{1}{2}\sum_{i=1}^{|\mathbb{T}_{t^*}|-1}\left(\|\mathbf{x}(t_{i+1})\| - \|\mathbf{x}(t_i)\|\right)\left(\text{TV}\left(\mathbf{x}(t_{i+1})\right) + \text{TV}\left(\mathbf{x}(t_i)\right)\right),$$

where the summation is over all maneuvers in the DRT. This computation is shown

notionally in Figure 6-3. If the total summed area is negative, then more unexpected spatial delay distributions are associated with decreasing TD, whereas if the area is positive, these unexpected spatial delay distributions are associated with increasing TD. Note that this is an *aggregate* measure over an entire DRT.

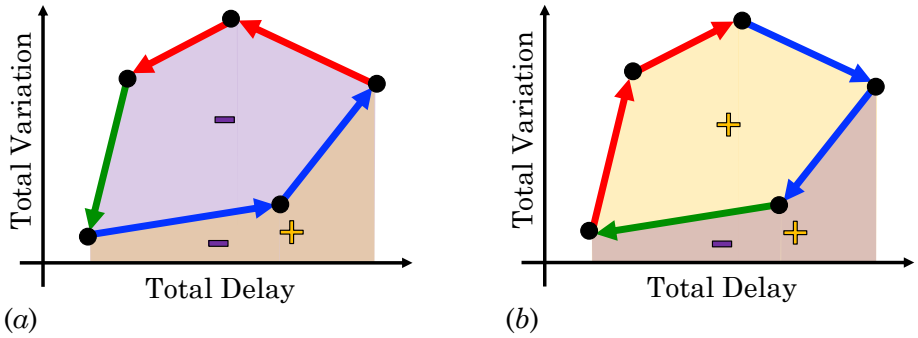


Figure 6-3: A notional DRT where higher TV is associated with recovery (*left*), and a DRT where higher TV is associated with disruption (*right*). *Reprinted from [159].*

Maximum TD and TV values: These features are the maximum observed TD values $\max_{i=1,\ldots,|\mathbb{T}_{t^*}|} \|\mathbf{x}(t_i)\|$ and TV values $\max_{i=1,\ldots,|\mathbb{T}_{t^*}|} \mathbf{x}(t_i)^\top \mathcal{L}\mathbf{x}(t_i)$ for each DRT. Unlike the intensity, these features incorporate the actual scale of the disruption.

Number of symbiotic and trade-off maneuvers: These features are the counts of each type of maneuver in a DRT. Recall from Section 6.2.4 the characteristics of symbiotic versus trade-off maneuvers.

Length of symbiotic and trade-off maneuvers: The length of a maneuver is defined as the Euclidean norm of the maneuver in $\mathbb{R}^2$. We use the total length of the symbiotic and trade-off maneuvers as features that indicate the dominance of each maneuver type. For example, given a maneuver from (1, 1) to (10, 10) and a maneuver from (1, 1) to (100, 100), both are symbiotic maneuvers, but the latter is a more pronounced, dominant evolution within the TV-TD state space. Explicitly, let us take the positive horizontal axis to be $\theta = 0$, and denote by $\measuredangle(\mathbf{v}(t_i), \mathbf{v}(t_{i+1})) \in [0, 2\pi)$ the angle of the 2-dimensional maneuver (represented as a vector in $\mathbb{R}^2$) with respect to the positive horizontal axis, where $\mathbf{v}(t_i) = [\|\mathbf{x}(t_i)\|, \mathrm{TV}(\mathbf{x}(t_i))]^\top \in \mathbb{R}^2$. Then, the length of symbiotic maneuvers $l_{\mathrm{sym}}(\mathbb{T}_{t^*})$ and the length of trade-off maneuvers $l_{\mathrm{to}}(\mathbb{T}_{t^*})$ are computed as

$$l_{\text{sym}}\left(\mathbb{T}_{t^*}\right) = \sum_{i=1}^{\mathbb{T}_{t^*}-1} \left\{\left\|\mathbf{v}\left(t_{i+1}\right) - \mathbf{v}\left(t_i\right)\right\|_2 \cdot \mathbf{1}\left(\measuredangle\left(\mathbf{v}\left(t_i\right), \mathbf{v}\left(t_{i+1}\right)\right) \in [0, \pi/2) \cup [\pi, 3\pi/2)\right)\right\},$$

$$l_{\text{to}}\left(\mathbb{T}_{t^*}\right) = \sum_{i=1}^{\mathbb{T}_{t^*}-1} \left\{\left\|\mathbf{v}\left(t_{i+1}\right) - \mathbf{v}\left(t_i\right)\right\|_2 \cdot \mathbf{1}\left(\measuredangle\left(\mathbf{v}\left(t_i\right), \mathbf{v}\left(t_{i+1}\right)\right) \in [\pi/2, \pi) \cup [3\pi/2, 2\pi)\right)\right\},$$

where $\mathbf{1}\left(\mathcal{A}\right)$ is the indicator function for event $\mathcal{A}$.

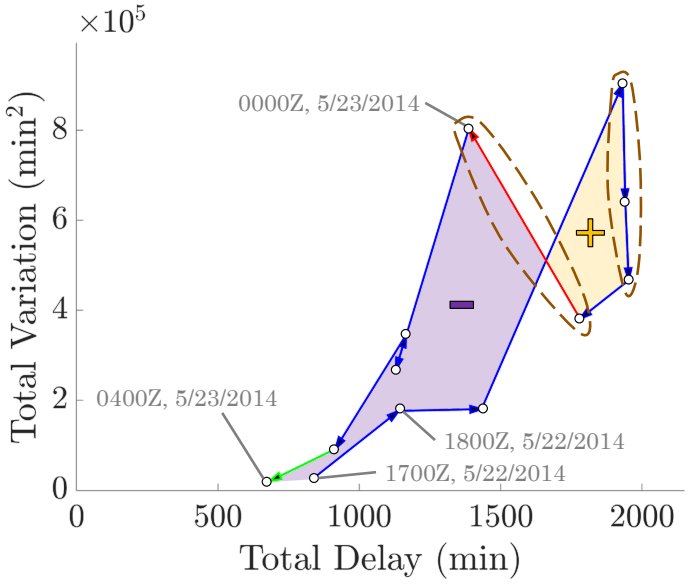


Figure 6-4: A 12-hour long DRT transitions out of the nominal region at 1700Z. Arrows denote maneuvers, and their colors denote the succeeding region (see Figure 6-1). Select DRT features from Section 6.2.5 such as signed enclosed area and trade-off maneuvers (brown dashed indicators) are annotated. Note that each state is a one hour interval. *Reprinted from [159].*

6.3 Average DRT characteristics

We first examine *average* DRT characteristics across the entire set of generated DRTs. Algorithm 3 yields 2,322 DRTs composed of 12,350 hours (approximately 14% of all hours) within the 10-year span contained in our data set. The average length of a DRT, i.e., the average time between the system state leaving and returning to the nominal $\left(\mathcal{N}\right)$ region, is 5.3 hours. Hence, the average duration during which the system is either in the disruption or recovery phase is 3.3 hours (subtracting the start and

end nominal hours). On average, two of these hours are in the high-delay region $\textcircled{\mathcal{S}}$, and one hour is in the unexpected distribution region $\textcircled{\mathcal{D}}$. In other words, although most of the duration of a typical DRT involves only high magnitudes of delay (the conventional measure of a disruption), one-third of the duration is associated with unusual spatial distributions of delay, independent from its magnitude. Note that this independence comes from the fact that we use strong OID bounds from Chapter 4 when computing $\left[\underline{\widehat{\Theta}}, \overline{\widehat{\Theta}}\right]_{\text{hour}(t)}$.

Recall that the TD and TV intensities measure how the values for each hour within a DRT compare to the maximum TD or TV values for that DRT. If the average intensity is close to 1, this indicates that a majority of the DRT was spent close to the maximum TD and/or TV value. Operationally, a higher intensity indicates that both the disruption *as well as* the recovery of the system happened in a shorter time span, or in other words, most of the hours were spent close to the peak disruption state. Figure 6-5 shows the histograms of the average TD and TV intensities for each DRT. The distribution of $\|\mathbf{x}\|$ (i.e., the TD) is left-skewed with a mean of 0.83, whereas the distribution of TV values is more symmetric with a mean of 0.57. The figure implies that when disruptions (and subsequently, recoveries) occur, the TD increases (and decreases) rapidly in time, but the effect on the spatial distribution of delay is more variable and evolves slowly. In other words, DRTs display faster changes along the horizontal (TD) axis than the vertical (TV) axis.

Recall that the signed enclosed area of a DRT reflects whether the TV was higher during increasing delays (positive area) or decreasing delays (negative area), aggregated across the entire DRT. A DRT with positive area indicates that when the delay in the system is increasing (i.e., during the disruption phase), this is also when the spatial variance is higher relative to the rest of the DRT. On the other hand, a DRT with negative area indicates that the spatial variance of delays was higher when total system delay is decreasing (i.e., during recovery). The average signed area is positive (6.72×10^6 min^3), indicating that the spatial distribution of delays tends to be more varied and unexpected during the disruption phase as compared to the recovery phase.

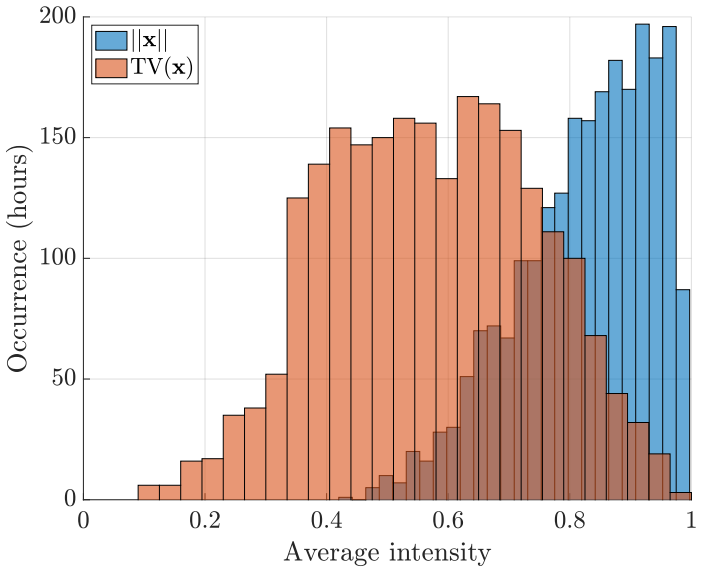


Figure 6-5: Distribution of TD and TV intensity values. *Reprinted from [159].*

The last feature that we discuss in an average sense across all 2,322 DRTs is the number of symbiotic and trade-off maneuvers. Since the average length of a DRT is approximately 5 hours, the average number of maneuvers in the TV-TD state space is $5 - 1 = 4$. Out of these, the average number of symbiotic maneuvers is 3, with 1 maneuver being a trade-off between TD and TV. Although the system prefers to evolve such that both TD and TV are increasing or decreasing, 25% of the times the system state exhibits a decrease in TD and an increase in TV, or vice versa. Since $\text{TV}(\mathbf{x}) = \mathbf{x}^\top \mathcal{L}\mathbf{x}$, there is a positive quadratic relationship between TV and TD, indicating that the system typically will evolve symbiotically. The 25% of times when the system state exhibits trade-off maneuvers form an interesting set of airport delay behaviors, possibly reflecting the influence of external inputs (Traffic Management Initiatives or TMIs, airline recovery actions, etc.) in the disruption-recovery process. Future work could focus on examining this set of trade-off maneuvers, and correlate these time intervals with known external interventions.

6.4 Clustering DRTs

We used k-means clustering with the 12 features from Section 6.2.5 in order to determine *representative* DRTs from the set of 2,322 DRTs. The distance metric we used for k-means is the standard squared Euclidean distance. While other clustering methods such as DBSCAN could be used, we chose k-means clustering for its interpretable parameter choice (i.e., number of clusters representing number of common DRT patterns) and simplicity. Prior to clustering, we standardize all feature observations by the feature mean and standard deviation. We select $k = 7$ clusters, taking into account the within cluster sum-of-square (WCSS) error, the cluster populations, and cluster interpretability. We plot the WCSS error as a function of k in Figure 6-6. Each cluster centroid provides an average representation of the DRTs that belong to that cluster. We list the centroids, along with pertinent DRT features, descriptive labels, and cluster population in Table 6.1.

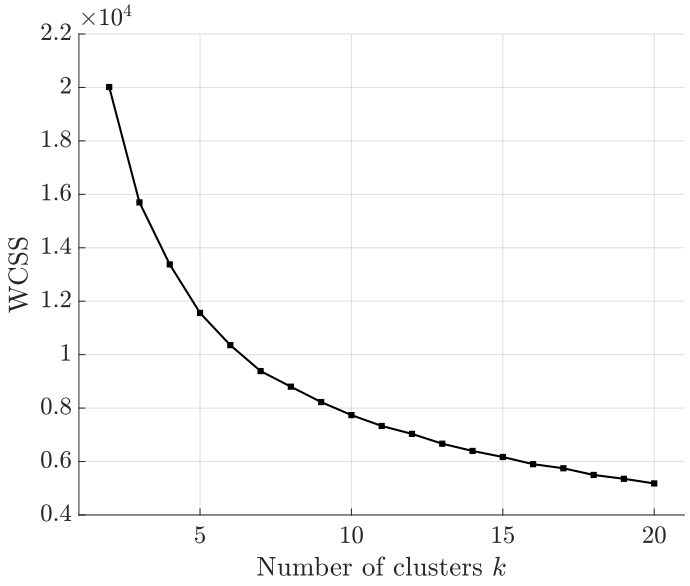


Figure 6-6: Within cluster sum-of-square (WCSS) error versus the number of clusters. *Reprinted from [159].*

DRT name	$\lvert\mathbb{T}_{t^*}\rvert$ (hours)	Region $\mathcal{S}$ (hours)	Region $\mathcal{D}$ (hours)	Avg. TD intensity	Avg. TV intensity	Area (min^3)	Max. TD (min)
Short_Dis	3	0	1	0.91	0.70	1.04×10^6	6.50×10^2
Short_Rec	4	1	1	0.76	0.47	-1.11×10^6	6.65×10^2
Med	6	2	2	0.76	0.43	-5.74×10^6	1.19×10^3
OpsDay_Dis	15	11	2	0.62	0.26	5.43×10^8	2.21×10^3
OpsDay_Rec	18	15	1	0.68	0.35	-4.98×10^6	1.68×10^3
MultiDay	55	49	4	0.59	0.22	2.20×10^8	2.28×10^3
Dec08Event	229	221	6	0.57	0.23	8.46×10^8	3.09×10^3

Table 6.1: The seven representative DRTs and their features. Note that clusters are sorted in increasing order by the average DRT duration, i.e., $\lvert\mathbb{T}_{t^*}\rvert$. Due to spacing, the table continues column-wise in Table 6.2. *Reprinted from [159].*

DRT name	Max. TV (min^2)	Symbiotic	Trade-off	Pop. (%)
Short_Dis	4.66×10^4	1	1	1163 (*50%*)
Short_Rec	1.05×10^5	2	1	777 (*33%*)
Med	5.50×10^5	4	1	196 (*8%*)
OpsDay_Dis	2.19×10^6	12	2	21 (*1%*)
OpsDay_Rec	5.02×10^5	13	4	142 (*6%*)
MultiDay	1.07×10^6	40	14	22 (*1%*)
Dec08Event	1.21×10^6	173	55	1

Table 6.2: The seven representative DRTs and their features. Note that clusters are sorted in increasing order by the average DRT duration, i.e., $\lvert\mathbb{T}_{t^*}\rvert$. Due to spacing, the table continues column-wise in Table 6.1. *Reprinted from [159].*

6.5 Analyzing representative DRTs (i.e., the identified DRT clusters)

For each representative DRT listed in Table 6.1, we discuss the operational characteristics that describe the disruption and subsequent recovery. Furthermore, these representative DRTs help identify when a disruption begins, when the recovery begins, and when the event ends, using historical data.

We now list the representative DRTs, along with a shortened tag that we will use to refer to them:

Short DRTs with spatially-perturbed disruption segments (`Short_Dis`): This type of DRT is the most prevalent (50% of all DRTs), and is a short-duration (3 hours) event. `Short_Dis` DRTs indicate brief disruptions: For example, a transient pop-up

thunderstorm around the vicinity of a major airport. In comparison to the other short representative DRT (`Short_Rec`), the average area for `Short_Dis` is positive, indicating that the airport delays were spatially distributed in a more unexpected manner during disruption than recovery. Furthermore, the maximum TV value observed for `Short_Dis` is significantly higher than `Short_Rec`, even though their maximum TD values are comparable.

Short DRTs with spatially-perturbed recovery segments (`Short_Rec`): This DRT type accounts for 33% of all DRTs. Similar to `Short_Dis`, these DRTs represent transient off-nominal conditions, with an average length of 4 hours. The average TD and TV intensity values for `Short_Rec` are smaller than those of `Short_Dis`, indicating that the system state does not typically attain the maximum TD and TV values. Furthermore, the area is negative, but of the same magnitude as `Short_Dis`, meaning that the spatial distribution of airport delays was more unexpected during recovery segments than disruption segments.

Medium-length DRTs (`Med`): These DRTs have an average length of around 6 hours, indicating that these disruptions and subsequent recoveries account for significant portions of an operational day in the US airspace system. The relative rarity of these longer-duration events are reflected in its cluster population: Only 196 out of 2,322 DRTs (about 8%) are classified as `Med`. We also note that, similar to `Short_Rec`, the airport delays were spatially distributed in a more unexpected manner during recovery segments than disruption segments.

Operational day-long DRTs with spatially-perturbed disruption segments (`OpsDay_Dis`): With average DRT lengths of approximately 15 hours, these disruptions and subsequent recoveries account for a major portion of an operational day. For example, a DRT in `OpsDay_Dis` beginning in the morning would not recover back into the nominal $(\mathcal{N})$ region until well into the evening. Similar to the difference between `Short_Dis` and `Short_Rec` DRTs, `OpsDay_Dis` and `OpsDay_Rec` DRTs differ by the signed enclosed area. The spatial distribution of airport delays was more unexpected during disruption segments for `OpsDay_Dis` DRTs. This DRT type accounts for less than 1% of all DRTs.

Operational day-long DRTs with spatially-perturbed recovery segments (`OpsDay_Rec`): The temporal persistence of these DRTs is similar to `OpsDay_Dis`, with an average length of 18 hours. As we have noted in `OpsDay_Dis`, the spatial delay distribution for `OpsDay_Rec` is more unexpected during recovery segments. Furthermore, the maximum TV value is significantly lower than `OpsDay_Dis`. Both `OpsDay_Rec` and `MultiDay` DRTs tend to occur in winter months, as we will discuss further in Section 6.6 when we combine information regarding specific off-nominal events (nor'easters, hurricanes, etc.) and month-of-occurrence.

Multi-day DRTs (`MultiDay`): This cluster of DRTs represents a prolonged disruption and subsequent recovery event, with average lengths of over 2 days (55 hours). The maximum observed TD and TV values are also some of the highest among all clusters, indicating that these lengthy DRTs impact the system severely in terms of both magnitude and spatial distribution of delays. We also note that the spatial distribution of delays tend to be more unexpected during disruption segments for `MultiDay` DRTs, as signified by the positive average area. The unique characteristic of these `MultiDay` DRTs is that there was no recovery back to a nominal $(\mathcal{N})$ region even during the overnight hours, when the system typically has low traffic and enough slack to reset the disruption.

We refer to the last cluster in Table 6.1 as `Dec08Event`; the fact that one unique DRT was placed in a cluster by itself indicates that it differs significantly from the other representative DRTs. Since it is a singular, extreme disruption-recovery event spanning almost 10 days (229 hours) in December 2008, we analyze it separately and present it as a case study.

In Section 6.3, we saw that the average DRT was composed of 75% symbiotic maneuvers and 25% trade-off maneuvers. Even though there is no reason by construction that this average needs to hold within each representative DRT cluster, the ratio of symbiotic to trade-off maneuvers appears to be robust to variations in the actual length of the DRT, assuming that it is long enough to observe such behavior. This indicates that even during prolonged disruption-recovery events, the preferred evolution of the system state is still in symbiotic directions, with maneuvers occur-

ring in the trade-off direction at a frequency of only 20-25%. A more theory-oriented direction for future work is to evaluate standard graph processes (diffusion, cascade, etc.) from the perspective of symbiotic and trade-off maneuvers, to see if there are general conditions under which the system dynamics evolve in one way or another.

6.6 Evaluating off-nominal events and temporal trends with regards to DRT clusters

6.6.1 Mapping off-nominal events to DRTs

We use the same set of 178 days from Section 4.4.3, along with the same off-nominal event types of nor'easters, hurricanes, thunderstorms, and airline- or airport-specific outages, and cross-reference these days with the set of DRTs. This allows us to examine what type of DRTs are common during each of these events. Figure 6-7 shows a normalized bar plot depicting the DRT type breakdown for each of the four off-nominal events.

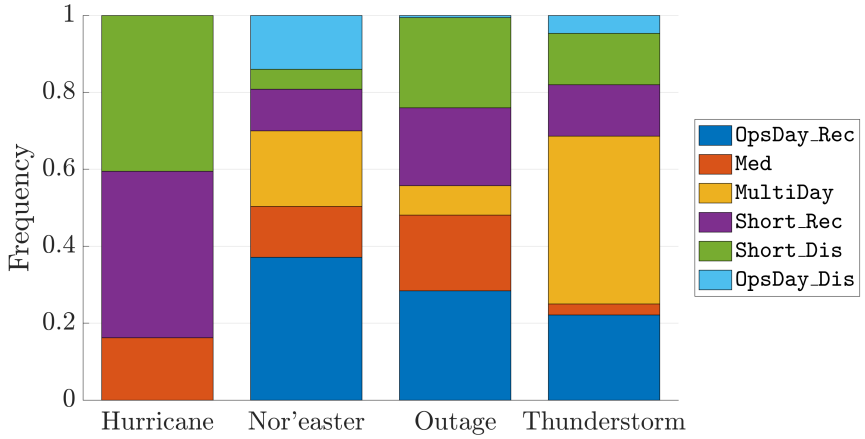


Figure 6-7: Frequency of representative DRTs, given the occurrence of an off-nominal event. The extreme `Dec08Event` cluster is not shown. *Reprinted from [159].*

Long-lasting DRTs (i.e., `OpsDay_Rec`, `OpsDay_Dis`, and `MultiDay`) are present in over 70% of nor'easter- and thunderstorm-type days, but only around 37% of days with an airline- or airport-specific outage. In particular, 37% of DRTs dur-

ing nor'easters are of type `OpsDay_Rec`, and 44% of DRTs during thunderstorms are `MultiDay`. The spatial distribution of airport delays during nor'easter days, particularly during recovery segments, tends to be more unexpected than during disruption segments. This could be indicative of airline-specific recovery efforts that result in airport delays at unusual combinations of airports. Examining the average representative DRT lengths, the time it takes for the system to be disrupted and recover from nor'easter-type days tends to be shorter than for thunderstorm-type days, which are dominated by `MultiDay` DRTs. This may be explained by the more volatile and disruptive nature of thunderstorm squall lines compared to large winter storms, resulting in more unpredictable DRTs. We note that an immediate line of future work will be to extend this DRT analysis to an airline-specific setting.

63% of DRTs are `Med`-length or shorter on outage-type days, indicating that the disruption and recovery of the system during these events are short-lived. Similar to nor'easters, the spatial distribution of airport delays is higher during recovery segments, with 68% of DRTs having a negative area. Finally, we note that for many hurricane-type days, due to pre-emptive cancellations and airport closures, both TD and TV values are suppressed. Hence, most DRTs (84%) during hurricane-type days are short-term disruptions and recoveries.

6.6.2 Monthly distribution of DRTs

In order to observe temporal trends in DRT occurrences, we plot the frequency of occurrence of representative DRT types in Figure 6-8, splitting the data set into a 2008-2016 subset, and a 2017 subset. The reason for this split is that certain representative DRTs in the year 2017 behaved differently than in the preceding 9 years. Specifically, `MultiDay` DRTs primarily appeared only in the winter months prior to 2017. By contrast, 42%, 30% and 23% of all DRTs in April, July, and August of 2017 were `MultiDay` DRTs. Furthermore, `MultiDay` DRTs in April and July 2017 are predominantly thunderstorm-type off-nominal days. This indicates an increased vulnerability of the system to thunderstorms in the summer of 2017. Further investigation outside of the scope of the work in this section would be needed

to determine what specific initiatives and policies might have caused this shift in disruption-recovery dynamics in 2017.

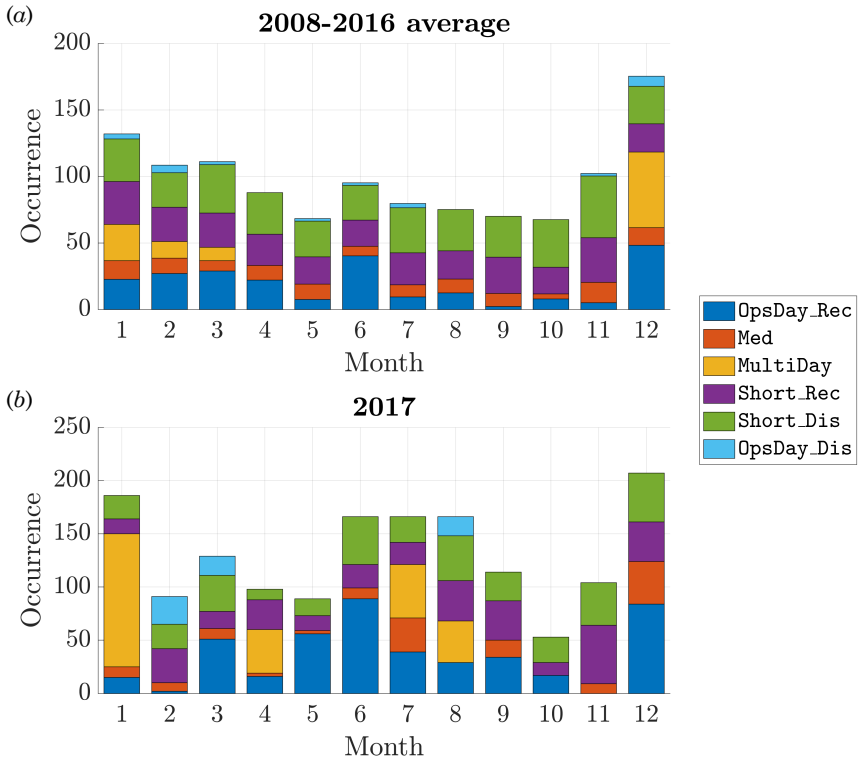


Figure 6-8: Occurrence counts of DRT hours for each month, split by representative DRT clusters; counts (a) averaged across 2008-2016 and (b) for 2017. *Reprinted from [159].*

6.6.3 December 2008 DRT: Case study

`Dec08Event` was an extremely long DRT (229 hours) with a sequence of disruptions and subsequent partial recoveries, occurring between December 15 and December 25, 2008. To better understand this DRT, we superimpose FAA-issued advisories related to Airspace Flow Programs (AFP), Ground Stops (GS), and Ground Delay Programs (GDP) for the duration of the `Dec08Event` DRT (Figure 6-9). The combined number of GS- and GDP-related advisories, a measure of airport capacity reductions and traffic flow management actions, remained at, or above, 29 for most of the `Dec08Event` DRT. There was a brief drop in the number of GS- and GDP-related advisories on December 22, but the continuity in the `Dec08Event` DRT indicates that the system was

unable to return to a nominal TV-TD state before undergoing another disruption-recovery event between December 23 to December 25. The system returned to a nominal state for about 48 hours, before entering into a `OpsDay_Rec`-type DRT between December 27 and 28.

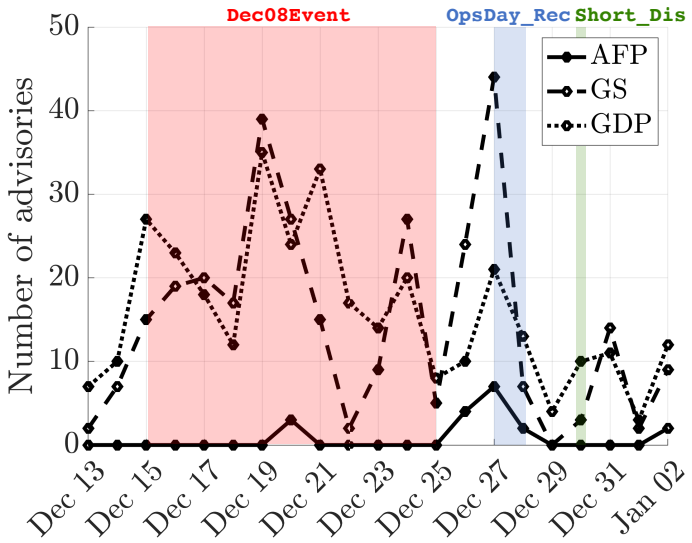


Figure 6-9: Plot of AFP-, GS-, and GDP-related advisories issued by the FAA during the primary December 2008 DRT and subsequent shorter DRTs. *Reprinted from [159].*

The `Dec08Event` DRT captures a series of disruptions caused by winter weather. This series of disruptions began with widespread ice storms throughout the US on December 11-12, and was followed by a separate, larger weather system that resulted in heavy rain in the West Coast, before transforming into a disruptive winter storm over the Midwest. The inability of the system state to return to the nominal $(\mathcal{N})$ region, even during late evening and early morning periods when the system typically resets, was extremely pronounced during this event.

6.7 Defining and clustering DRTs: Summarizing remarks

Thus far in Chapter 6, we defined disruptions and subsequent recoveries using the delay magnitude (TD) and the spatial distribution (TV) of the delays, in conjunction

with a low-dimensional state-space trajectory representation. We presented a partition of the TV-TD state space into various regions, representing nominal conditions, high-delay conditions, and conditions with unexpected spatial distributions of delay. We then focused on the problem of finding representative DRTs, which we did by clustering DRTs. The seven representative DRT clusters identified had interpretable characteristics in terms of lengths (i.e., the duration of disruptions and subsequent recovery), intensities, and delay behavior during the disruption or recovery segments.

With a formal framework of DRTs in hand, we build upon this foundation through a number of succeeding analyses: We will now seek to examine each of the individual DRTs at a more microscopic level, with a focus on time periods involving trade-off maneuvers. Doing so will help reveal whether these trade-off maneuvers correspond to the implementation of certain TMIs, driving the TD and TV values in a direction not normally traversed by them. Another direction for expanded analysis is to leverage the representative DRTs as features for predicting future system behavior, both at the system-wide and airline-specific levels. For instance, if we observe an ongoing disruption and can map it to some representative DRT, we may be able to better predict the future recovery segments for this ongoing disruption. This prediction problem would also be interesting, and potentially more applicable, if we restrict ourselves to specific airline sub-networks.

6.8 Refining the framework of DRTs

For the remainder of Chapter 6, we will refine the DRT framework introduced in Section 6.2, with two motivating goals: Provide consistent definitions of disruptions within airline sub-networks, and set the stage for predicting the evolution of disruptions and recoveries using DRTs. The refinements we make consist of two major components: A more operationally-revealing partition of the TV-TD state space, shown in Figure 6-10, and proving structural properties of DRTs along with discussing their interpretations in Section 6.8.3. Throughout the remainder of Chapter 6 dedicated to refining the DRT framework, we use the same nomenclature and conventions adopted

earlier in Section 6.2.3.

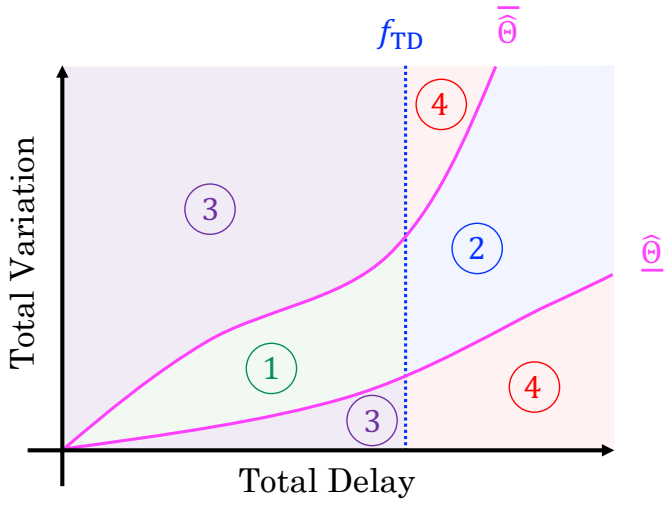


Figure 6-10: The refined partition of the TV-TD state space that we use from Section 6.8 onward.

Figure 6-10 depicts a 4-state partition of the TV-TD state space. We will demonstrate what it means for a particular airport delay observation to belong in one of the four regions $\{(1), (2), (3), (4)\}$ through a real-world example with hourly delay states from the AA sub-network. We plot these four AA hourly examples in Figure 6-11. For delay states belonging in (1), this region is characterized by low total delay, along with relative smooth delay distributions across the graph, given that the TV is within some bounds $\left[\underline{\Theta}, \overline{\Theta}\right]$, where the bounds can be derived through any of the methods from Chapter 4. Correspondingly, in the top-left panel of Figure 6-11, we have that AA airport delays were classified as belonging to region (1) in hour 1700Z on January 19, 2015: We see that delays at all Core 30 airports have close to zero delays within the AA sub-network, reinforcing its classification to region (1).

If we shift to the right on the TD axis such that we cross some delay threshold f_{TD}, but remain within $\left[\underline{\Theta}, \overline{\Theta}\right]$, we are now in region (2). In the top-right panel of Figure 6-11, we show an example from the AA sub-network of hourly delays that were classified as belonging to region (2). During 0000Z on April 17, 2013, we see high delays at almost all Core 30 airports for the AA sub-network. In particular, we note relatively similar delay impacts at airports that are historically highly-correlated,

e.g., East Coast airports such as those in Washington, DC and New York City. Thus, this delay state observation has high total delay, but distributed smoothly across the network.

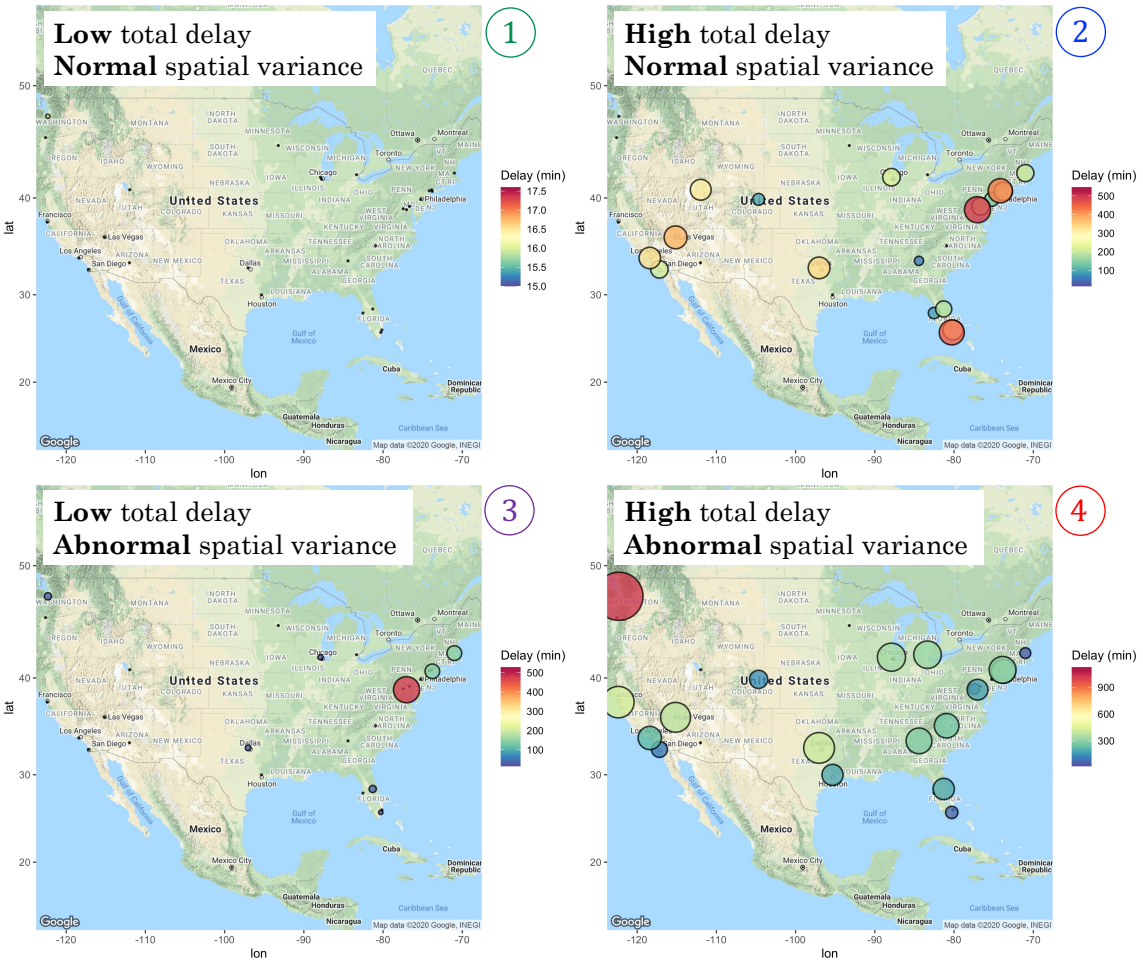


Figure 6-11: AA network delay signals observed at 1700Z on January 19, 2015 (region 1), 0000Z on April 17, 2013 (region 2), 0100Z on April 4, 2010 (region 3), and 2300Z on April 16, 2013 (region 4). Size and color of the circles indicate airport delays; note the different color scale magnitudes between the four sub-figures.

For regions (3) and (4), we leave the regime of smooth signal distributions, signified by leaving the $\left[\underline{\Theta}, \overline{\Theta}\right]$ bounds in the TV-TD state space. In region (3), we see low total delays, but the spatial distribution of delays are no longer expected given historical correlations. In the bottom-left panel of Figure 6-11, we see that, with the exception of DC area airports, the AA sub-network during 0100Z on April 4, 2010 had very little airport delays. However, the severe delays affecting DCA is spatially

unexpected, as other East Coast airports do not show similar levels of flight delays. Finally, in the bottom-right panel of Figure 6-11, the AA sub-network at 2300Z on April 16, 2013 incurred not only high delays spread throughout the Core 30 airports, but also experienced severely impacted operations at SEA.

6.8.1 State timelines and trajectory regimes

With this concrete motivation for the new 4-state partition drawn from hourly delays observed by the AA sub-network, we proceed to establishing some definitions that will be useful in proving structural characteristics regarding DRTs:

Definition 4 (State timeline) *Given the disjoint partition of the TV-TD state space as illustrated in Figure 6-10, a state timeline $\mathcal{H}$ is a list that describes an ordered sequence of countable, possibly infinite number of states drawn from the set* $\{\textcircled{1}, \textcircled{2}, \textcircled{3}, \textcircled{4}\}$.

With the language of an all-encompassing state timeline $\mathcal{H}$, we now define three DRT *trajectory regimes*, each describing different evolution patterns within the 4-state partitioned TV-TD state space:

Definition 5 (Trajectory regimes) *We define three trajectory regimes $\mathcal{S}(\mathcal{H})$, $\mathcal{D}(\mathcal{H})$, and $\mathcal{I}(\mathcal{H})$ with respect to state timeline $\mathcal{H}$ as follows:*

(i) Trajectory regime $\mathcal{S}(\mathcal{H})$ consists of minimal-length trajectories with terminal states in regions $\{\textcircled{1}, \textcircled{3}\}$ *and at least one anchor state in regions* $\{\textcircled{2}, \textcircled{4}\}$.

(ii) Trajectory regime $\mathcal{D}(\mathcal{H})$ consists of minimal-length trajectories with terminal states in regions $\{\textcircled{1}, \textcircled{2}\}$ *and at least one anchor state in regions* $\{\textcircled{3}, \textcircled{4}\}$.

(iii) Trajectory regime $\mathcal{I}(\mathcal{H})$ consists of minimal-length trajectories with terminal states in region $\{\textcircled{1}\}$ *and at least one anchor state in regions* $\{\textcircled{2}, \textcircled{3}, \textcircled{4}\}$.

To simplify notation, when it is clear that the regimes are defined with respect to the same state timeline $\mathcal{H}$, we omit the $\mathcal{H}$ and simply write $\mathcal{S}, \mathcal{D}$, and $\mathcal{I}$. See Figure 6-12 for a visualization of these trajectory regimes.

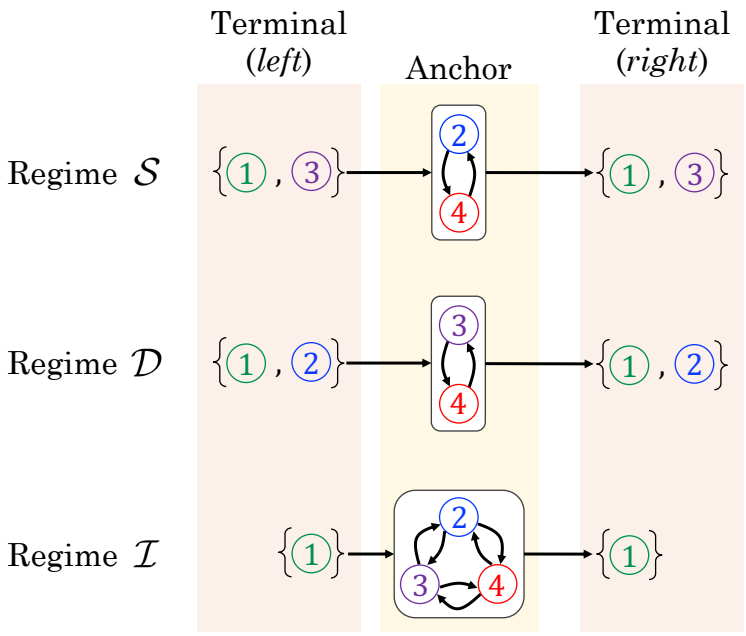


Figure 6-12: Schematic representation of the three different trajectory regimes $\mathcal{S}, \mathcal{D}$, and $\mathcal{I}$.

Given some state timeline realization $\mathcal{H}'$, we note that regimes $\mathcal{S}(\mathcal{H}'), \mathcal{D}(\mathcal{H}')$, and $\mathcal{I}(\mathcal{H}')$ are supersets of individual trajectories τ, whereas τ is an ordered sub-list of $\mathcal{H}'$. We denote the sub-list relationship by $\tau \subseteq_{\triangle} \mathcal{H}'$. Similarly, any other analogous set-theoretic operations extendable to list relationships are denoted by their usual set-theoretic relational symbols, with an additional $\triangle$ subscript. When we refer to $\mathcal{S}, \mathcal{D}$, and $\mathcal{I}$ without reference to a timeline $\mathcal{H}$, we are referring to a generic supserset of trajectories that all satisfy the above definitions. On the other hand, $\mathcal{S}(\mathcal{H}), \mathcal{D}(\mathcal{H})$, and $\mathcal{I}(\mathcal{H})$ indicate that, for example, a trajectory $\tau \in \mathcal{S}(\mathcal{H})$ satisfies both Definition 5(i) and that $\tau \subseteq_{\triangle} \mathcal{H}$.

6.8.2 Structural properties of DRTs

For the following theorems, lemmas, corollaries, and their associated proofs, we will use the following rules regarding notation when describing a trajectory τ: Possible states that could occupy a single position are surrounded by $\{\cdot\}$, whereas an ordered, non-empty permutation of possible states is denoted by $\mathcal{P}(\cdot)$. To emphasize that

order must be respected, we use $\rightarrow$ to connect states. For example, a trajectory τ given by

$$\tau: \quad \{\textcircled{1}\} \rightarrow \mathcal{P}\left(\textcircled{3}, \textcircled{4}\right) \rightarrow \{\textcircled{1}, \textcircled{2}\} \tag{6.1}$$

could be realized in the following (non-exhaustive) ways:

$$\begin{aligned} &\{\textcircled{1}\} \rightarrow \{\textcircled{3}\} \rightarrow \{\textcircled{1}\}, \\ &\{\textcircled{1}\} \rightarrow \{\textcircled{4}\} \rightarrow \{\textcircled{2}\}, \\ \text{or } &\{\textcircled{1}\} \rightarrow \{\textcircled{4}\} \rightarrow \{\textcircled{4}\} \rightarrow \{\textcircled{3}\} \rightarrow \{\textcircled{2}\}. \end{aligned}$$

In other words, for our example τ in (6.1), the left terminal state must be $\textcircled{1}$, and the right terminal state must be either $\textcircled{1}$ or $\textcircled{2}$. Between the left and right terminal states, there must be at least one of either state $\textcircled{3}$ or $\textcircled{4}$, with possibly countably infinite numbers of states $\textcircled{3}$ or $\textcircled{4}$ with no fixed order.

We first want to show that DRTs based solely on the magnitude of delay (i.e., $\mathcal{S}$-DRTs) can *always* be extended to an integrated definition of a DRT, i.e., $\mathcal{I}$-DRTs. We prove this in Theorem 3, with the help of the following definition formalizing the *extension* of a trajectory:

Definition 6 (Trajectory extension) *For any trajectory τ, denote an extension of τ by $\tau \subseteq_{\triangle} \text{Ext}(\tau) \subseteq_{\triangle} \mathcal{H}$. Furthermore, an empty extension is the only case where sub-list equality holds, i.e., $\tau \neq_{\triangle} \text{Ext}(\tau) \Leftrightarrow \tau \subset_{\triangle} \text{Ext}(\tau)$.*

Theorem 3 (Extension from $\mathcal{S}$ to $\mathcal{I}$) *Any regime $\mathcal{S}$ trajectory can be extended to a regime $\mathcal{I}$ trajectory, with the possibility of an empty extension.*

Proof. The approach is to split an arbitrary regime $\mathcal{S}$ trajectory into different cases depending on the left- and right-terminal states, then construct the appropriate extensions based on the terminal states. For the full proof, please refer to Appendix A.5. □

Corollary 1 (Extension from $\mathcal{D}$ to $\mathcal{I}$) *Any regime $\mathcal{D}$ trajectory can be extended to a regime $\mathcal{I}$ trajectory, with the possibility of an empty extension.*

Proof. The proof follows identically from the proof of Theorem 3 in Appendix A.5 by relabeling $\textcircled{2} \mapsto \textcircled{3}$ and $\textcircled{3} \mapsto \textcircled{2}$. □

Analogous to the extensions of $\mathcal{S}$- and $\mathcal{D}$-DRTs to $\mathcal{I}$-DRTs, we will now show that there is a natural decomposition from $\mathcal{I}$-DRTs back to $\mathcal{S}$- and $\mathcal{D}$-DRTs.

Theorem 4 (Decomposition from $\mathcal{I}$ to $\mathcal{S}$) *Any trajectory $\tau \in \mathcal{I}(\mathcal{H})$ can be decomposed into $|\mathcal{S}(\tau)| = k+1-m$ disjoint trajectories in $\mathcal{S}(\mathcal{H})$, where k is the number of $\textcircled{3}$ states in τ, m is the number of occurrences of pairs*

$$\{\textcircled{1}\} \to \{\textcircled{3}\},$$
$$\{\textcircled{3}\} \to \{\textcircled{1}\},$$
$$\textit{and } \{\textcircled{3}\} \to \{\textcircled{3}\}$$

in τ, and $|\cdot|$ denotes the cardinality, i.e., number of trajectories in the superset of decomposed trajectories $\mathcal{S}(\tau) \subseteq \mathcal{S}(\mathcal{H})$.

Proof. The main idea for the proof is to keep track of the number of terminal states, as well as the number of *pairs* of valid terminal states for $\mathcal{S}$ that do not terminate a trajectory in regime $\mathcal{I}$. For the full proof, please refer to Appendix A.6 □

Corollary 2 (Decomposition from $\mathcal{I}$ to $\mathcal{D}$) *Any trajectory $\tau \in \mathcal{I}(\mathcal{H})$ can be decomposed into $|\mathcal{D}(\tau)| = k+1-m$ disjoint trajectories in $\mathcal{D}(\mathcal{H})$, where k is the number of $\textcircled{2}$ states in τ and m is the number of occurrences of pairs*

$$\{\textcircled{1}\} \to \{\textcircled{2}\},$$
$$\{\textcircled{2}\} \to \{\textcircled{1}\},$$
$$\textit{and } \{\textcircled{2}\} \to \{\textcircled{2}\}$$

in τ.

Proof. The proof follows identically from the proof of Theorem 4 in Appendix A.6 by relabeling $\textcircled{2} \mapsto \textcircled{3}$ and $\textcircled{3} \mapsto \textcircled{2}$. □

Remark 1 (Regime $\mathcal{I}$ containment) *By way of Theorem 4 (or Corollary 2), note that since both left and right terminal states of $\tau \in \mathcal{I}(\mathcal{H})$ are also terminal states for $\mathcal{S}(\mathcal{H})$ (or $\mathcal{D}(\mathcal{H})$), any decomposition of τ into regime $\mathcal{S}$ trajectories by way of Theorem 4 (or into regime $\mathcal{D}$ by way of Corollary 2) is entirely contained within τ, i.e., $\tau_{\mathcal{S}} \subseteq_{\triangle} \tau, \forall \tau_{\mathcal{S}} \in \mathcal{S}(\tau)$, and similarly $\tau_{\mathcal{D}} \subseteq_{\triangle} \tau, \forall \tau_{\mathcal{D}} \in \mathcal{D}(\tau)$.*

Finally, given the extension and decomposition properties, it follows that there are relationships between the *number* of trajectories belonging to the three trajectory regimes, given a fixed state timeline. Analogously, there is also a relationship between the number of states within a specific trajectory. We state and prove both of these assertions in Theorems 5 and 6, respectively.

Theorem 5 (Regime cardinalities) *Given the decomposition relationship proved in Theorem 4, we have that*

$$|\mathcal{I}(\mathcal{H})| \leq |\mathcal{S}(\mathcal{H})| + |\mathcal{D}(\mathcal{H})|\,, \tag{6.2}$$

where $|\cdot|$ indicates the cardinality of the regime set, i.e., the number of trajectories belonging to that regime.

Proof. This proof first requires an additional lemma that shows the existence of at least one decomposition. Then, the proof follows by examining different decomposition cases. For the full proof, please refer to Appendix A.7. □

Theorem 6 (Trajectory cardinalities) *The relation in* (6.2) *from Theorem 5 holds at the level of trajectories as well:*

$$\sum_{\tau \in \mathcal{I}(\mathcal{H})} |\tau|_{\triangle} \leq \sum_{\tau \in \mathcal{S}(\mathcal{H}) \cup \mathcal{D}(\mathcal{H})} |\tau|_{\triangle}\,, \tag{6.3}$$

where $|\cdot|_{\triangle}$ indicates the cardinality of the trajectory, i.e., the length or number of states belonging to that trajectory.

Proof. This proof is similar to the proof for Theorem 5, except that we first work at the level of one individual trajectory, then aggregate appropriately. For the full proof, please refer to Appendix A.8. □

6.8.3 Operational interpretations

The three trajectory regimes given in Definition 5 reflect different ways of demarcating disruptions and recoveries within an air transportation network, Specifically, regime $\mathcal{S}$ delineates disruptions and recoveries solely based off of the magnitude of delays, as a disruption begins by leaving low-delay regions (1) or (3), and recoveries end after returning to (1) or (3). On the other hand, trajectories in regime $\mathcal{D}$ only considers the spatial distribution of delays within the network, since disruptions begin when the TV is aberrant (i.e., exits (1) or (2)) and recoveries end when the TV returns to normal (i.e., re-enters (1) or (2)). We then introduce an *integrated* regime $\mathcal{I}$, which delineates air transportation network disruptions by considering both deviations in delay magnitudes and spatial distribution.

Since it is clear that regime $\mathcal{S}$ could be considered as a canonical measure of inefficiencies and IROPs within the air transportation network (it simply measures disruptions and recoveries based on the magnitude of airport delays), we want to ensure that the integrated regime $\mathcal{I}$ remains *consistent* with the usual ways of measuring disruptions. For example, if $\tau_{\mathcal{S}}$ is a DRT under regime $\mathcal{S}$, a prototypical inconsistent behavior would be the failure of $\tau_{\mathcal{S}}$ to also be a DRT under regime $\mathcal{I}$ – in other words, our new disruption definitions should never nullify previous DRTs identified using canonical disruption measures. The theorems, propositions, and corollaries from the previous subsection precisely demonstrate these consistency guarantees between regimes $\mathcal{S}, \mathcal{D}$, and $\mathcal{I}$.

In particular, we first note that any DRT identified with respect to regimes $\mathcal{S}$ or $\mathcal{D}$ can be thought of as the "beginning" of a DRT in both delay magnitude and spatial distribution, i.e., in the regime $\mathcal{I}$ sense (Theorem 3 and Corollary 1). Combining this with the fact that any DRT from regime $\mathcal{I}$ can be separated into shorter disruptions as measured by its delay or spatial distribution (Theorem 4), we have the following,

somewhat intuitive guarantee: Using delay thresholds or the spatial variance of delays as a measure of disruptions is a strict subset of integrating the two performance measures. Along with the last guarantee that an integrated DRT can never contain an *incomplete portion* of a magnitude- or spatial distribution-based DRT (Remark 1), our three regime definitions are indeed self-consistent. Finally, we note that each of the theoretical results can be empirically verified during the construction of integrated DRTs in $O(N)$ time, where N is the total number of hours in the original data set. We can use the previously-stated Algorithm 3, with a new region partition $\mathcal{R}(t) \in \left\{①, ②, ③, ④\right\}$, to construct these refined DRTs.

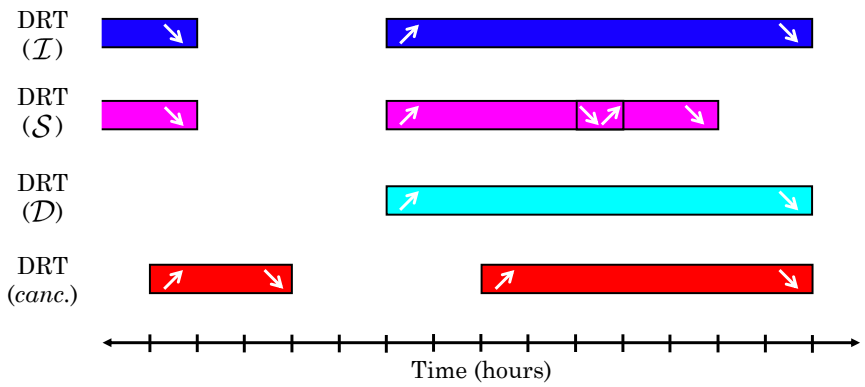


Figure 6-13: Notional representation of DRTs and their interactions.

6.8.4 Delay and cancellation DRTs

Previously in this section, we construct and prove properties of DRTs, which can be seen as encoding disruptions into discrete, ordered sequences of states. With this framework in mind, we focus on integrated delay DRTs, i.e., DRTs in trajectory regime $\mathcal{I}(\mathcal{H})$, which we will refer to as $\mathcal{I}$-DRTs for brevity. As we have seen via Theorem 4, an $\mathcal{I}$-DRT incorporates disruptions and recoveries both in terms of the magnitude as well as the spatial distribution of airport delay. The goal in this subsection is to examine the *interaction* between airport delays and cancellations within an airline network, taking advantage of the rigid structure of DRTs. Specifically, we define interactions between delays and cancellations as the *overlap* or *inclusion* between

$\mathcal{I}$-DRTs (e.g., the dark-blue segment within the barcode representation in Figure 6-13) and cancellation DRTs (e.g., the red segment in Figure 6-13). We focus on four categories of $\mathcal{I}$-DRT and cancellation DRT overlap or inclusion:

(i) **Cancellations *lead* delays:** In the middle of a cancellation DRT, an $\mathcal{I}$-DRT begins, and the cancellation DRT ends strictly earlier than the $\mathcal{I}$-DRT.

(ii) **Delays *lead* cancellations:** In the middle of an $\mathcal{I}$-DRT, a cancellation DRT begins, and the $\mathcal{I}$-DRT ends strictly earlier than the cancellation DRT.

(iii) **Cancellations *include* delays:** The entirety of an $\mathcal{I}$-DRT coincides with the entirety of, or some portion of, a cancellation DRT.

(iv) **Delays *include* cancellations:** The entirety of a cancellation DRT coincides with the entirety of, or some portion of, an $\mathcal{I}$-DRT.

These four categories can be visualized in Figure 6-14.

The structure provided by DRTs allows us to easily measure when $\mathcal{I}$-DRTs and cancellation DRTs begin (i.e., left-terminates, as shown in Figure 6-12) and end (i.e., right-terminates, as shown in Figure 6-12). With this in mind, delay-cancellation interactions in categories (i) and (ii) describe the scenarios where flight cancellations due to disruptions and demand-capacity imbalances precede delays, or vice versa, respectively. For example, suppose that an airline proactively cancels a subset of flights in anticipation of a certain level of capacity reductions; however, the realized capacity reductions are more severe than anticipated by previous forecasts, resulting in the non-cancelled flights experiencing significant delays. Such a scenario would fall under category (i), where cancellations precede delays. On the other hand, reactive flight cancellations in response to worsening delays would fall under category (ii). The other categories (iii) and (iv) encapsulate scenarios where there is no clear order of precedence between $\mathcal{I}$-DRTs and cancellation DRTs: Instead, they describe scenarios where flight cancellations happen completely within the time span of a high-delay period, or vice versa. We can describe categories (iii) and (iv) in similar interpretative flavors as (i) and (ii), with the caveat that the leading indicator of disruptions (i.e.,

either delays or cancellations) persist even after the appearance and resolution of the other indicator.

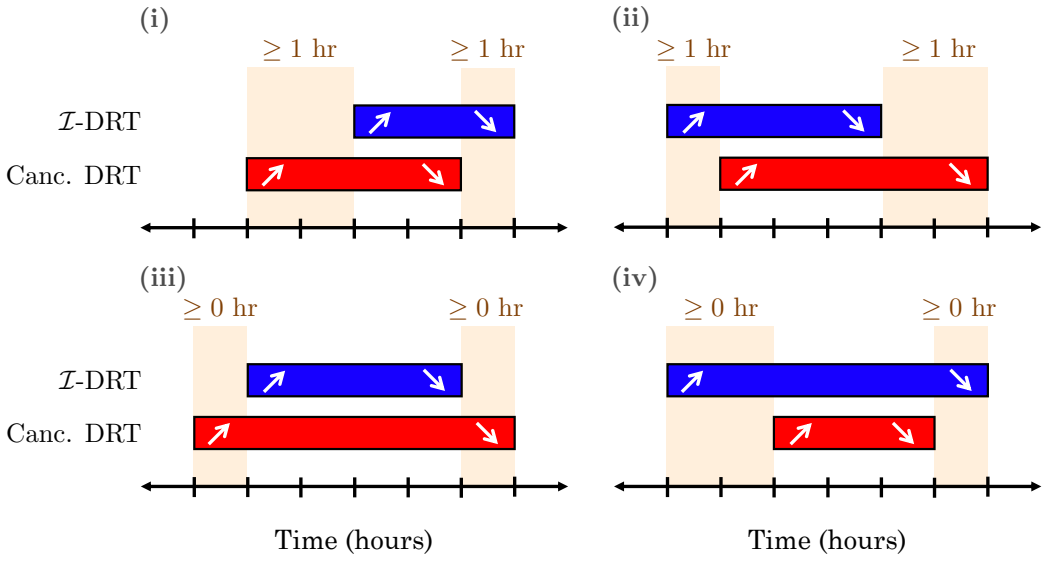


Figure 6-14: Search criterion for the delay-cancellation interaction categories (i)-(iv).

For each airline, we are interested in obtaining the counts of each interaction category, given that we have already constructed the set of cancellation and $\mathcal{I}$-DRTs from airline's state timeline $\mathcal{H}$. For categories (i) and (ii) where there is an order of precedence between flight delays and cancellation, we require a buffer of at least 1 hour (i.e., one state within $\mathcal{H}$) between the beginnings and endings of the cancellation and $\mathcal{I}$-DRT. We depict this pictorially in Figures 6-14(i) and 6-14(ii). The other two categories, represented in Figures 6-14(iii) and 6-14(iv), have less stringent buffer requirements, as we allow for the possibility of cancellation and $\mathcal{I}$-DRTs starting and/or ending at the same time. We provide the pseudo-code for the algorithm to count each category type in Algorithm 4.

We run Algorithm 4 on four separate state timelines for AA, DL, UA, and WN, and provide the counts of interaction categories per airline in Table 6.3. We see that interaction categories where cancellations lead delay (e.g., proactive cancellations) and delays lead cancellations (e.g., severe delays necessitate flight cancellations) are actually less frequently observed than inclusion-type scenarios (iii) and (iv), regardless

of the specific airline we examine. Furthermore, we note that while interactions (i) and (ii) seem "symmetric" in some sense, i.e., the number of occurrences are approximately the same for all four airlines, there is much more asymmetry between (iii) and (iv), in that interactions where cancellations DRTs being strict subsets of delay DRTs are much more prevalent than the other way around. This imbalance is particularly noticeable for UA, where there were just under 1,000 additional (iv) interaction scenarios compared to (iii) scenarios. For the purposes of benchmarking airline on-time and flight completion performance, it may be of interest to understand the disruptions that precipitate into a (iv) scenario versus a (i) or (ii) scenario. For example, ideally, a cancellation DRT would not be succeeded by a delay DRT, nor be included within a delay DRT, indicating successful mitigation of demand-capacity imbalances post-flight cancellations.

	Category	**AA**	**DL**	**UA**	**WN**
(i)	Cancellations lead delays	1048	850	1082	871
(ii)	Delays lead cancellations	1059	845	1113	880
(iii)	Cancellations include delays	1250	994	1300	1148
(iv)	Delays include cancellations	1809	1540	2268	1820

Table 6.3: Counts for each of the delay-cancellation interaction categories, split by airlines.

Algorithm 4 Counting delay-cancellation interactions

Input: State timeline $\mathcal{H}$; Set of $\mathcal{I}$-DRTs $\mathcal{I}(\mathcal{H})$; Set of cancellation DRTs $\text{canc}(\mathcal{H})$; Delay-cancellation interaction category $\texttt{CAT} \in \{\text{(i), (ii), (iii), (iv)}\}$

Output: Number of interactions $n_{\texttt{CAT}} \in \mathbb{N}_{\geq 0}$

Initialize $n_{\texttt{CAT}} \leftarrow 0$; $\mathbb{1}_{\texttt{CAT}} \leftarrow \text{FALSE}$
for *Each ordered state indexed* $i = 1 : (|\mathcal{H}| - 1)$ *of* $\mathcal{H}$ **do**
    if CAT *start condition at* $i \wedge \mathbb{1}_{\mathit{CAT}} = \text{FALSE}$ **then**
        if CAT *active condition at* $i+1 \vee \mathit{CAT}$ *end condition at* $i+1$ **then**
            $\mathbb{1}_{\texttt{CAT}} \leftarrow \text{TRUE}$
    else if CAT *end condition at* $i \wedge \mathbb{1}_{\mathit{CAT}} = \text{TRUE}$ **then**
        $n_{\texttt{CAT}} \leftarrow n_{\texttt{CAT}} + 1$; $\mathbb{1}_{\texttt{CAT}} \leftarrow \text{FALSE}$
end

6.9 Future work: DRT-based prediction models

The DRT framework leverages a representation of airport network delays as a trajectory object, transitioning between discrete states. The total delay and normalized total variation $\left(\|\mathbf{x}\|_1, \sqrt{\text{TV}(\mathbf{x})}\right)$ capture information regarding both the magnitude and spatial impact of air transportation system disruptions. Thus far in Chapter 6, we had focused on an application of clustering DRTs, where representative DRT clusters can be examined for attributes such as their durations and intensity. We also introduced refinements to the DRT framework, where we prove structural characteristics of, e.g., $\mathcal{S}$-, $\mathcal{D}$-, and $\mathcal{I}$-DRTs.

We now briefly discuss some predictive applications that builds off of the DRT framework. In particular, we focus on two types of questions: Can we predict the *trend* of the airport delays, and can we predict *whether or not* a DRT will begin (or end), given conditions at the current time. The setting is as follows: Suppose that at the current time t, the entire vector of airport delays $\mathbf{x}_t$ is observable, i.e., we have access to the delays $x_{i,t}$ at airport i during current time t, as well as non-schedule-related information (e.g., weather, airport capacities, etc.) Furthermore, we

assume that we have an analogous set of information for the past hours as well. The trend prediction problem addresses whether the total delay will increase or decrease, and similarly for total variation. This can be viewed as a sign classification problem, where we are trying to predict sign $(\|\mathbf{x}_t\|_1 - \|\mathbf{x}_{t+1}\|_1)$ and sign $(\mathrm{TV}(\mathbf{x}_t) - \mathrm{TV}(\mathbf{x}_{t+1}))$.

The second prediction question, termed the "nominal zone prediction" problem for convenience, needs to be framed within the context of a DRT type. For consistency, we will use $\mathcal{I}$-DRTs, which left- or right-terminates in $\{\textcircled{1}\}$ and anchors in $\{\textcircled{2}, \textcircled{3}, \textcircled{4}\}$. If $\mathbf{x}_t$ currently belongs to $\{\textcircled{1}\}$, then we are interested in predicting whether or not a DRT will begin, by predicting if $\mathbf{x}_{t+1}$ will belong to $\{\textcircled{2}, \textcircled{3}, \textcircled{4}\}$. On the other hand, if $\mathbf{x}_t$ currently belongs to $\{\textcircled{2}, \textcircled{3}, \textcircled{4}\}$, we would be interested in predicting if, at the next hour, the system recovers, i.e., if $\mathbf{x}_{t+1}$ will be in $\{\textcircled{1}\}$. To conclude Chapter 6, we will briefly describe a kernel density estimation-based approach for the trend prediction problem, and a time series prediction-based approach, via recurrent neural networks (RNNs), for the nominal zone prediction problem. These prediction problems are part of ongoing work in [156].

6.9.1 Kernel density estimation approaches

The approach we take towards trend predict hinges on constructing a probability density function of the tuple $\left(\|\mathbf{x}\|_1, \sqrt{\mathrm{TV}(\mathbf{x})}\right)$ at the *next* time step, then integrating this density function across the appropriate half-plane to predict the desired quantities sign $(\|\mathbf{x}_t\|_1 - \|\mathbf{x}_{t+1}\|_1)$ and sign $(\mathrm{TV}(\mathbf{x}_t) - \mathrm{TV}(\mathbf{x}_{t+1}))$. To construct the probability density function, we can use a bivariate kernel density estimator with a bivariate Gaussian kernel $\mathcal{K}$ and some 2×2 bandwidth matrix $\mathbf{H}$, where the data for the kernel density estimator comes from the transition behaviors of nearest neighbors in the space of tuples $\left(\|\mathbf{x}\|_1, \sqrt{\mathrm{TV}(\mathbf{x})}\right)$ for our current delay vector $\mathbf{x}_t$. Figure 6-15 shows a depiction of this procedure applied to predicting the trend of total delay. In the case depicted by Figure 6-15, the predicted sign of $\|\mathbf{x}_t\|_1 - \|\mathbf{x}_{t+1}\|_1$ would be negative, indicating an increase in total delay.

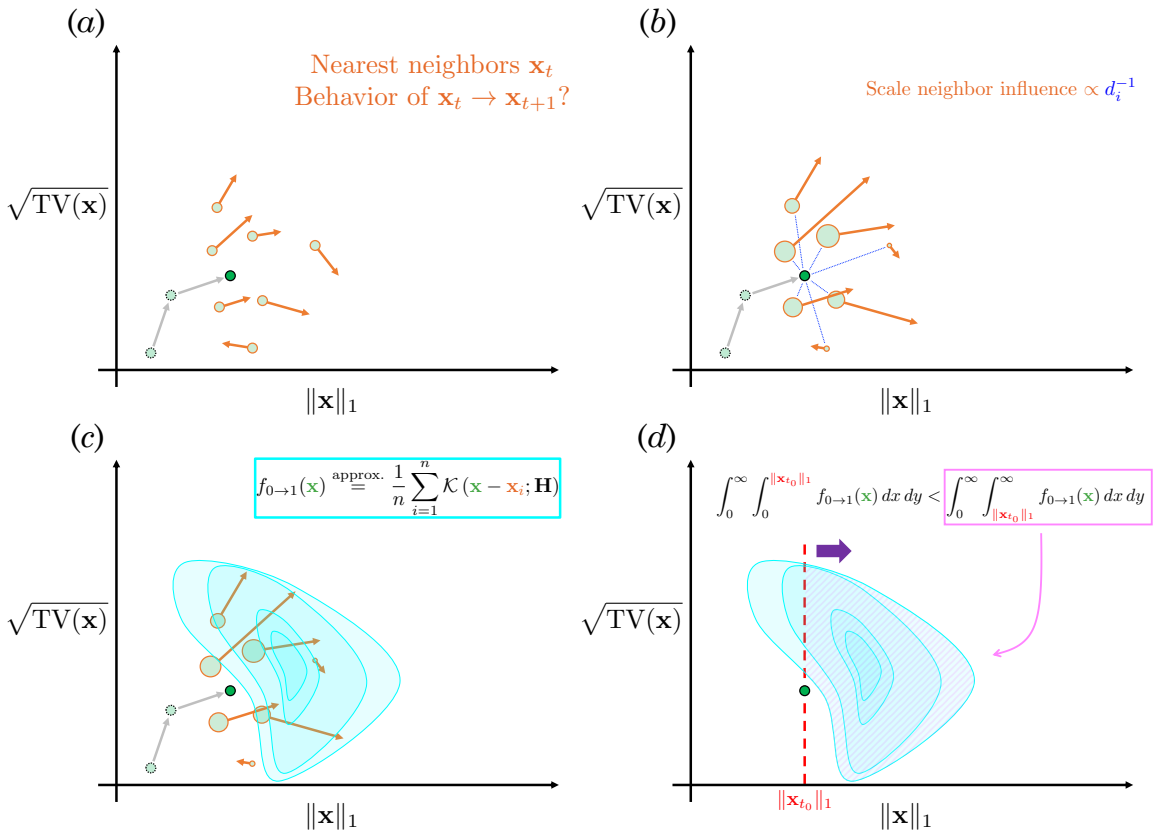


Figure 6-15: Workflow for the kernel density estimation approach for trend prediction on DRTs: (a) Retrieve nearest neighbors; (b) Inversely scale neighbor influence based on distance; (c) Construct density through kernel density estimation; (d) Numerically integrate appropriate regions of the density.

6.9.2 Deep learning approaches via LSTM RNN

For the nominal zone prediction problem, our first attempt was to use the same estimated density from Section 6.9.1, and numerically integrate over region(s) of interest. Note that depending on the prediction direction (e.g., stepping out of or into $\{\textcircled{1}\}$), the region or regions of interest will be either $\{\textcircled{1}\}$ or $\{\textcircled{2},\textcircled{3},\textcircled{4}\}$. However, this approach did not yield an acceptable prediction accuracy, and was computationally intensive. Instead, we will approach this prediction problem through deep learning, specifically using a long short-term memory (LSTM) RNN architecture to predict the tuple $\left(\|\mathbf{x}_{t+1}\|_1, \sqrt{\mathrm{TV}\left(\mathbf{x}_{t+1}\right)}\right)$, given previous time steps along the time series containing sequences of $\left(\|\mathbf{x}\|_1, \sqrt{\mathrm{TV}(\mathbf{x})}\right)$ values. The LSTM RNN is trained to minimize the usual squared ℓ_2 loss between the predicted tuple and the actual tuple, and then this predicted tuple is mapped to the regions $\{\textcircled{1},\textcircled{2},\textcircled{3},\textcircled{4}\}$ independent of the LSTM RNN. In doing so, we actually have two performance metrics to consider: The first is the ℓ_2 loss, which assesses the performance of the LSTM RNN in its prediction of the specific $\left(\|\mathbf{x}\|_1, \sqrt{\mathrm{TV}(\mathbf{x})}\right)$ tuple. The second performance metric is the accuracy of classifying the predicted $\left(\|\mathbf{x}\|_1, \sqrt{\mathrm{TV}(\mathbf{x})}\right)$ tuple into the correct region.

Chapter 7

Towards Hierarchical Traffic Flow Management

7.1 The delay redistribution problem

The key idea formalized in Chapter 4 is the notion that airport delays can be viewed as signals supported on the nodes of a graph, and that adding a perspective of the spatial distribution of these signals is interesting both theoretically (e.g., Section 4.1) as well as operationally in the context of examining disruptions in the air transportation network (e.g., Section 4.5). However, thus far, most of the work presented comes from a post hoc analysis standpoint. In Chapter 7, we ask the question of modeling and targeting "desirable" distributions of delays across a network of airports. We will admittedly overload the term *distribution* by way of representing both geographic or spatial distributions of signals across the network, as well as an actual multivariate probability distribution generating networked delays. Specifically, the ability to now characterize the spatial distribution of graph signals provides a new objective through which a network control problem could be formulated: Is there a better way to *redistribute* delays across my network, given that you cannot arbitrarily remove delay from the system?

We recognize that the notion of delay redistribution may be counterintuitive: Unlike some tangible, physical asset (e.g., packages, containers, etc.), you cannot simply

take the delays experienced by one airport and place it at another airport. However, we argue instead for the following perspective: The network-level redistribution of airport delays reflects an aggregation of microscopic, tactical actions under airport capacity constraints. Examples of such tactical actions include internal airline schedule adjustments or changes in traffic management initiative scopes. Thus, with this new perspective, in Chapter 7 we investigate the problem of designing delay redistribution control policies for an air transportation network under *delay-conserving* constraints. These constraints reflect the fact that incurred delays cannot be easily dissipated in the absence of mechanisms such as flight cancellations. Future work could focus on incorporating flight cancellation mechanisms by way of relaxing the delay conservation constraint.

Our framework offers advantages in terms of its formulation by explicitly handling heterogeneous marginal delay distributions at airports and dependence structures between airport delay distributions arising from network effects. Another advantage is that we provide a flexible and interpretable cost structure for encouraging delay absorption or dissipation at *pre-selected* airports. The resultant approximate control policy dictates required delay reductions or increases at specific airports. We demonstrate our framework on hourly sequences of US air transportation network disruptions between 2008 and 2017 (i.e., the DRTs from Chapter 6), and compare the optimal selective redistribution policies against actual operations. We also provide an estimate of the redistribution cost in delay minutes, resulting in a disruption-specific ranking of least- to most-costly delay-absorbing airports. Our control policies could be used as input constraints into the standard air traffic flow management problem (ATFMP) or multi-airport ground holding problem (MAGHP), ensuring that its solution conforms to network-level redistribution requirements. In turn, the low-level problems (i.e., the ATFMP or MAGHP) play crucial roles of constraining the high-level planner (i.e., the model we will propose in Chapter 7) to a feasible flight schedule. This hierarchical check-and-balance system is precisely the focus of ongoing work in [49], which builds on the results in Chapter 7.

7.1.1 Copula representations for network-wide airport delay distributions

The high-dimensional, model-less setting, along with a sparse set of metrics that are critical to monitor in order to ensure nominal system performance, is precisely found in our aviation setting. Consider the air transportation system, where airports are considered to be nodes in the network, and whose delays we wish to model and control. Interdependencies between different airports due to flight-propagated delays, weather correlations, passenger connectivity, and myriad other factors are complex to model. Consequently, it is difficult to develop an accurate and robust model for delays across all airports within the network. Furthermore, for system operators, key performance requirements are aggregate measures, and may not be defined at specific airports. Examples of such performance targets include the total delay (i.e., the sum of all airport delays), the spatial distribution of delays across the network, and the variance of airport delays. When the system is disrupted, the system operator may want to drive the system to some desired configuration that satisfy aggregate performance targets, with secondary considerations of delays at individual airports. In addition, there may be specific time intervals within which any intervention must occur.

Our goal is to formalize the notion of coming up with ways of redistributing airport delays that is *realistic* given historical observations of delays in the system. Hence, a crucial factor is the ability to generate realistic network delay state observations. The way we chose to do so is by drawing samples from some underlying probability distribution that describes the delay at each airport within the network. Two factors complicate this task: The marginal delay distribution at each airport may differ, and there could be a variety of dependence relations (linear, non-linear, time-varying, etc.) between the delays at different airports. The former encapsulates the fact that different airports have significantly different operating characteristics (e.g., runway capacity, airspace structure, typical weather patterns, etc.), whereas the latter is the result of the networked nature of the system (e.g., tail-propagated delays, shared

airspace constraints, traffic management initiatives, etc.).

To motivate our usage of a particular statistical construct known as a *copula*, we first demonstrate that simply drawing from a simple joint distribution with a simple dependence structure does not sufficiently represent the operational realities of airport delays. The straightforward process of fitting network-wide airport delays to a multivariate Gaussian distribution with some (sample) Pearson correlation-based dependence structure is not representative, given the observation that the marginal distributions for airport arrival and departure delays appear to follow a log-normal or Gamma distribution [12]. This is not entirely surprising, given that sub-process time durations, such as taxi times and terminal area transit times, are not Gaussian either [154, 11]. Furthermore, even in terms of distributional properties, Gaussian distributions are real-valued and symmetric, whereas operationally-relevant airport delay distributions are non-negative, and positive- or right-skewed (e.g., [161] and related work done in Section 5.5).

To simultaneously overcome both of these obstacles, we would like a new way to structure probability distributions that allows us to isolate the dependence structure of a general multivariate probability distribution from the individual marginal distributions. This separation would then allow for separate estimations of the marginal distributions and the dependence structure. This is precisely the advantage of using copulas [199]:

Definition 13 (N-dimensional copula) *An N-dimensional copula, denoted by $C : [0,1]^N \to [0,1]$, is a cumulative distribution function $C(\mathbf{u}) = C(u_1, \ldots, u_N)$ describing a random variable $\mathbf{U}$ in a N-dimensional unit hypercube $[0,1]^N$. $C(\mathbf{u})$ is a copula if, and only if, it satisfies the following properties:*

(i) $C(\mathbf{u})$ is non-decreasing in each component u_i.

(ii) The i^{th} marginal distribution of $C(\mathbf{u})$, obtained by setting $u_j = 1, \forall j \neq i$, is equal to the cumulative distribution function corresponding to a standard uniform random variable.

(iii) For scalars $a_i \leq b_i$, we have that

$$\mathbb{P}\left(\bigcap_{i=1}^{N} a_i \leq U_i \leq b_i\right) = C(b_1, \ldots, b_N) - C(a_1, \ldots, a_N).$$

Copulas disentangle the dependence structure from individual marginals by representing the dependence completely via the form of the copula function C, whereas individual marginals are represented separately as standard univariates u_1 through u_N via a probability integral transform. Furthermore, *any* multivariate cumulative distribution function can be represented by a copula, and this representation is unique as long as the multivariate distribution is continuous. This is given by Sklar's Theorem:

Theorem 2 (Sklar's Theorem [250]) *Consider a N-dimensional cumulative distribution function $F_{\mathbf{X}}$ with marginals $F_{X_1}, \ldots, F_{X_N}$. Then, there exists a copula C such that*

$$F_{\mathbf{X}}(x_1, \ldots, x_N) = C\left(F_{X_1}(x_1), \ldots, F_{X_N}(x_N)\right),$$

for all $x_i \in \mathbb{R} \cup \{-\infty, \infty\}$ and $i = 1, \ldots, N$. Furthermore, if F_{X_i} is continuous for all $i = 1, \ldots, N$, then the copula C is unique.

Sklar's Theorem allows us to first estimate the marginal distributions, and then attempt to fit dependence structures based on the choice of the copula family. We exploit this property by estimating empirical cumulative distribution functions for each airport based on historical data, and then separately fitting a copula-based dependence structure via maximum likelihood estimation. We assume that the marginal distributions and the copula cumulative distribution function are continuous.

In other words, we can now estimate individual marginal airport delay distributions (either parametrically or non-parametrically, although we chose the non-parametric route since misspecification of marginals can lead to unstable results [139]), then separately estimate the copula C through maximum likelihood-type routines [31]. Finally, it is important to note that a copula contains no temporal information, whereas airport delay distributions are by nature highly non-stationary [125].

While there are time-varying *copula processes* that could be used to model such systems [176], we make a deterministic assumption of periodicity and simply estimate a different copula model corresponding to different hours of the day.

7.2 An overview of our approach

Our approximate projection-based control approach for networks is generalizable, and not specific to the air transportation use case. Specifically, our proposed general solution consists of two main components:

(1) Generate a set of feasible high-dimensional states; and

(2) Map a desired trajectory satisfying performance targets in a lower-dimensional projection back to the high-dimensional state space.

Figure 7-1 presents a flowchart depicting our methodological framework. The first step uses a limited number of observations and draws from a fitted multivariate Gaussian copula to represent the set of feasible high-dimensional system states. As mentioned previously, given a complicated multivariate distribution with intertwined dependence structures, copulas provide a way of relating the multivariate joint distributions to univariate marginal distributions. The second step of our framework enforces conformance to lower-dimensional performance targets, and maps them back to the full system state via an optimization subroutine that identifies a corresponding high-dimensional state from the feasible set generated in the first step.

Our framework can accommodate any reasonable lower-dimensional projection space for the desired performance metric. Our approach is also flexible in terms of the objective function and cost structure for the optimization subroutine, which identifies a high-dimensional state that satisfies low-dimensional performance targets.

7.2.1 Elaborating on our approach to delay redistribution

We elaborate on the notion of delay redistribution that we first introduced at the beginning of Section 7.1. As we pointed out previously, while notions of redistributing

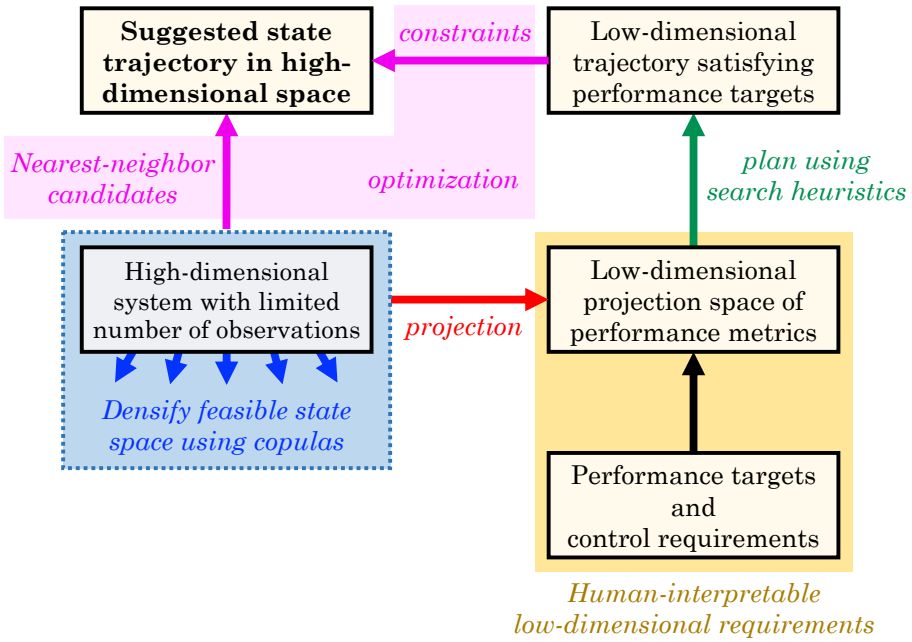


Figure 7-1: Flowchart of the proposed approach. *Reprinted from [155].* © *2020 IEEE*

quantities across a network are not uncommon, in the context of airport delays, it may be counterintuitive given the fact that delays are *accrued quantities* based on differences in scheduled versus actual arrival (or departure) times. This stands in contrast to physical quantities that can be easily thought of as "redistributable." We emphasize that when we talk loosely about redistributing airport delays at a network-level, it really is an aggregation of microscopic, tactical actions made by airlines or air navigation service providers in response to some demand-capacity imbalance. The notion of redistribution is more natural in the context of network-supported signals: We will make this perspective concrete with the following toy example:

Example: Let $f_{A\to C}$ and $f_{B\to C}$ be two same-duration flights from airports A and B, respectively, to destination airport C. Both $f_{A\to C}$ and $f_{B\to C}$ are scheduled to depart at 1600Z, and arrive at C at 1800Z. Suppose airport C has a capacity constraint of one arriving aircraft at 1800Z. In one scenario, flight $f_{A\to C}$ may be assigned an Expected Departure Clearance Time (EDCT) of 1630Z, and arrive at C at 1830Z (we assume no time was made up en route); in another scenario, flight $f_{B\to C}$ may be assigned an EDCT of 1630Z, and arrive at 1830Z. Whether one scenario happens

over another scenario depends on various microscopic factors such as airline schedule adjustments and/or traffic management initiative scopes. However, when observing the time series of delays at airports A, B, and C, the two scenarios look identical, modulo a *redistribution* of delay between airports A and B at 1600Z. We note that one could alternatively separate the accounting of arrival versus departure delays at an airport, and use a "multi-layered network" approach. Furthermore, this setup assumes that delays are recorded at the scheduled arrival and departure times, which is indeed the standardization used in the ASPM database maintained by the FAA [85]. This example is illustrated in Figure 7-2.

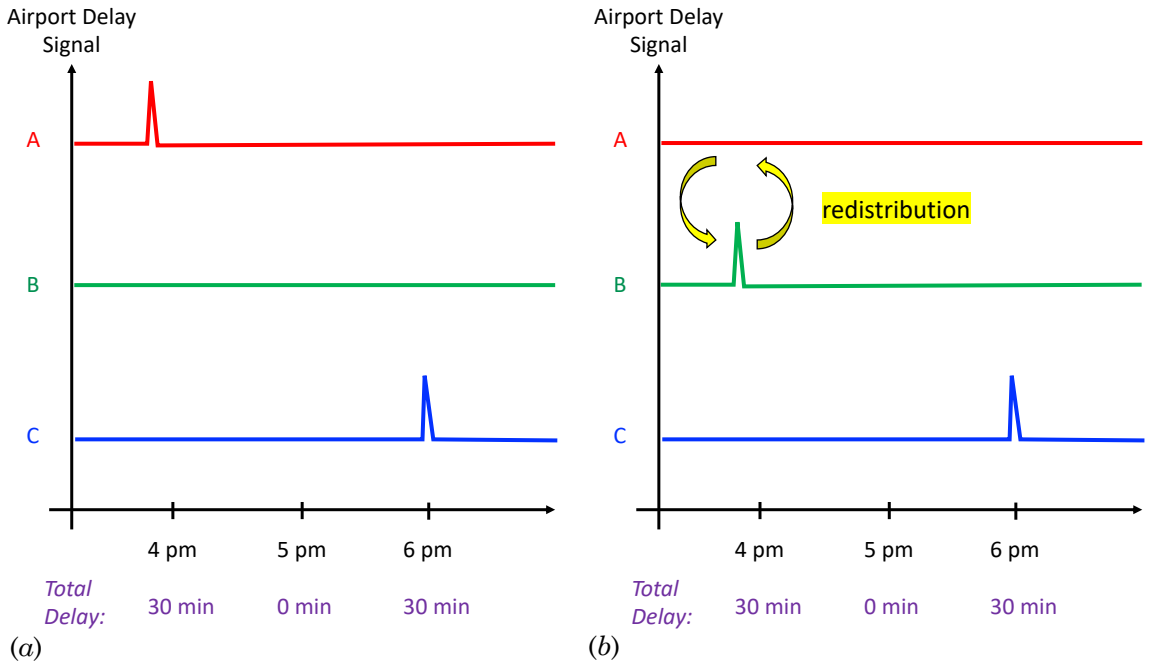


Figure 7-2: A simple delay redistribution scenario, based on whether flight $f_{A\to C}$ (a) or $f_{B\to C}$ (b) was assigned a new, later EDCT.

At the level of airport delay distributions, the observed delay signals reflect the operational inefficiencies incurred after carrying out a specific set of schedules (i.e., the demand-side) against constraints such as traffic management initiatives (i.e., the capacity-side). The classic traffic flow management problem (TFMP), regardless of whether it is implemented at the level of individual flights (Lagrangian approach, e.g., [27, 14]) or aggregate traffic flows (Eulerian approach, e.g., [181, 254, 257]), is

controlling traffic at a microscopic scope, whereas the macroscopic view sets desired performances at the "terminus" point where all delays are accounted for: At the airport. We note that one motivation for focusing on airports within the US NAS stems from the fact that airport capacity is a critical bottleneck in the US NAS [173].

How such a macroscopic perspective can fit into the overall air traffic management structure is by providing a closed feedback loop between the observed delay state and future schedule adjustments. This idea is illustrated in Figure 7-3: Currently, flow planning takes into account a *lagged* perspective of the system state via historical traffic data (e.g., through the Traffic Flow Management System, or TFMS). Given that state-of-the-art flight-level TFMP implementations can be quite fast, even for large numbers of traffic [14], one could consider a near-real time feedback between the current system delay state with desired delay redistribution requirements, and re-running the TFMP. The airport-level delay constraints fed back to the TFMP can be thought of as upper bound on delay management performance, as the magnitude of delay dissipation and absorption shown in Section 7.7 is most likely unattainable given operational constraints.

Even though it is unreasonable to assume that any TFMP can find a feasible solution that achieves the upper bound on delay redistribution actions, this feedback is still viable in the sense that small adjustments and changes in scheduling can have significant impact in terms of on-time performance [125, 214]. This raises an interesting question: If large, airport-level delay dissipation and absorption constraints are added to standard TFMP solutions, *how close* can the solution get to the upper bound without losing feasibility. In other words, previous research has shown one direction of influence (i.e., small schedule perturbations result in large changes in delays), does this imply that high-level delay redistribution policies at the level of airports are potentially easy to translate to microscopic schedule changes, displacements, and adjustments. While the work described in Chapter 7 does not explicitly implement the feedback cycle depicted in Figure 7-3 (a brief discussion regarding how these airport delay constraints could fit into a TFMP model is given in Section 7.8.3), it provides the mechanism that generates the performance upper bounds to be fed

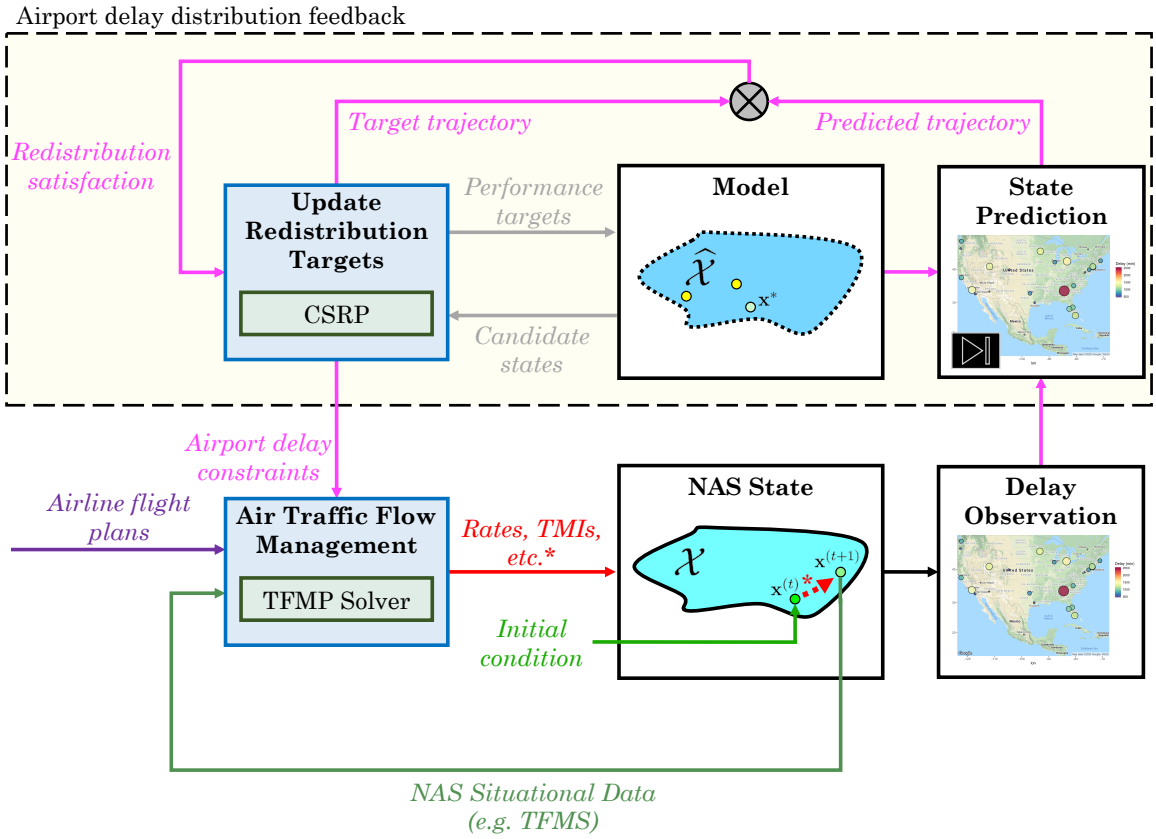


Figure 7-3: Model predictive control-like feedback cycle for incorporating high-level airport delay redistribution targets into the standard air traffic management process. In Chapter 7, we build up to the Conservative Selective Redistribution Problem (CSRP) formulation, the model, and constructing a target redistribution trajectory. Future work (e.g., [49]) will focus on implementing delay constraints into standard TFMP or multi-airport ground holding formulations, as well as the trajectory prediction component.

back as constraints in a TFMP formulation.

7.3 State space description and copula estimation

Given some (networked) system (e.g., a network of airports), we abstract the N signal-generating members (e.g., the airports with airport delays) of the system as vertices $i \in V$, and encode relationships between members i and j as undirected, possibly weighted, edges $(i,j) = (j,i) \in E \subseteq V \times V$. We collect the vertices and edges into a graph $G = (V, E)$, and denote a vertex i-supported signal by $x_i \in \mathbb{R}$. We denote observation l of the vector of signals generated from this system, i.e., the state of the system, by $\mathbf{x}^{(l)} \in \mathbb{R}^{N\times 1}$, with $\mathbf{x}^{(l)} = \left(x_1^{(l)}, \ldots, x_N^{(l)}\right)^\top$. If $\mathbf{x}^{(l)}$ was observed historically, then $\mathbf{x}^{(l)} \in \mathcal{X} \subseteq \mathbb{R}^N$. We will call $\mathcal{X}$ as the state space of the system. We assume we have no other information on $\mathcal{X}$, apart from the historical observations.

We can use a kernel density estimator to construct continuous marginal distributions, then transform them to standard univariates via the probability integral transform. We can find the copula density $c(\mathbf{u})$ by taking the appropriate partial derivatives of $C(\mathbf{u})$:

$$c(u_1, \ldots, u_N) = \frac{\partial^N}{\partial u_1 \cdots \partial u_N} C(u_1, \ldots, u_N). \tag{7.1}$$

Then, from Sklar's Theorem, we can rewrite the unknown probability density function $f_{\mathbf{X}}$ as a product of the copula density in (7.1) and the marginal density functions f_{X_i}, as shown in (7.2). We then estimate $f_{\mathbf{X}}$ written via the copula density through maximum likelihood estimation using the set of historical M historical observations $\mathfrak{O}_M = \left\{\mathbf{x}^{(l)}\right\}_{l=1}^{l=M} \subset \mathcal{X}$.

$$
\begin{aligned}
f_{\mathbf{X}}(\mathbf{x}) &= \frac{\partial^N}{\partial x_1 \cdots \partial x_N} F_{\mathbf{X}}(x_1, \ldots, x_N) \\
&= \frac{\partial^N}{\partial u_1 \cdots \partial u_N} C\left(\underbrace{F_{X_1}(x_1)}_{u_1}, \ldots, \underbrace{F_{X_N}(x_N)}_{u_N}\right) \times \frac{\partial F_{X_1}(x_1)}{\partial x_1} \cdots \frac{\partial F_{X_N}(x_N)}{\partial x_N} \\
&= c(\mathbf{u}) \prod_{i=1}^{N} f_{X_i}(x_i). \qquad (7.2)
\end{aligned}
$$

Let $\ell\left(\boldsymbol{\Theta}; \mathfrak{O}_M\right)$ denote the log-likelihood of (7.2) with respect to the copula parameter, $\boldsymbol{\Theta}$, and historical state observations, $\mathfrak{O}_M$. We have that

$$
\ell\left(\boldsymbol{\Theta}; \mathfrak{O}_M\right) = \sum_{k=1}^{M} \ln c\left(F_{X_1}\left(x_1^{(k)}\right), \ldots, F_{X_N}\left(x_N^{(k)}\right); \boldsymbol{\Theta}\right) + \sum_{k=1}^{M} \sum_{i=1}^{N} \ln f_{X_i}\left(x_i^{(k)}\right). \qquad (7.3)
$$

Note that the log-likelihood is only over the copula parameter $\boldsymbol{\Theta}$. Since we have no information regarding parameterizations of marginal densities or cumulative distribution functions, we utilize the *canonical maximum likelihood* (CML), where empirical marginal distributions are first estimated based on $\mathfrak{O}_M$, and these empirical marginal distributions are used to transform observations in $\mathfrak{O}_M$ to uniform variates via the probability integral transform. Under CML estimation, (7.3) reduces to

$$
\widehat{\boldsymbol{\Theta}} = \underset{\boldsymbol{\Theta} \in \mathcal{M}}{\operatorname{argmax}} \sum_{k=1}^{M} \ln c\left(\widehat{u_1}^{(k)}, \ldots, \widehat{u_N}^{(k)}; \boldsymbol{\Theta}\right), \qquad (7.4)
$$

where $\mathcal{M}$ is the space of copula parameters, and $\widehat{u_i}^{(k)} = \widehat{F}_{X_i}\left(x_i^{(k)}\right)$ is computed as the probability integral transform with an empirical estimate $\widehat{F}_{X_i}$ of F_{X_i} at each vertex i, given by the kernel density estimator in (7.5) with bandwidth h and the standard normal density function $\phi(t)$ as the smoothing kernel. Therefore, we have that

$$
\widehat{F}_{X_i}(x_i) = \frac{1}{M} \sum_{k=1}^{M} \left\{ \int_{-\infty}^{\frac{x_i - x_i^{(k)}}{h}} \phi(t)\, dt \right\}. \qquad (7.5)
$$

Now that we are able to consider a wide range of empirical marginal distributions,

we impose a dependence structure by choosing a family of copulas over which we can carry out the CML estimation in (7.4). There are a variety of parametric copula families for bivariate distributions [199]; for multivariate distributions, the most flexible is the multivariate Gaussian copula, along with other possibilities such as t-copulas and vine copulas [145]. We use CML to fit a multivariate Gaussian copula $C(\mathbf{u};\boldsymbol{\rho})$ with the cumulative distribution function and density $c(\mathbf{u};\boldsymbol{\rho})$:

$$C(\mathbf{u};\boldsymbol{\rho}) = \Phi_{\boldsymbol{\rho}}\left(\Phi^{-1}(u_1),\ldots,\Phi^{-1}(u_N)\right), \tag{7.6}$$

$$c(\mathbf{u};\boldsymbol{\rho}) = \det(\boldsymbol{\rho})^{-\frac{1}{2}} \exp\left(-\frac{1}{2}\boldsymbol{\Xi}^{\top}\left(\boldsymbol{\rho}^{-1} - I_{N\times N}\right)\boldsymbol{\Xi}\right) \tag{7.7}$$

where $\Phi_{\boldsymbol{\rho}}$ is a standardized multivariate normal distribution with correlation matrix $\boldsymbol{\rho} \in \mathbb{S}_{\succeq 0}^{N\times N}$ having unit diagonals, Φ^{-1} is the inverse cumulative distribution function for a standard normal distribution, $\boldsymbol{\Xi} = \left(\Phi^{-1}(u_1),\ldots,\Phi^{-1}(u_N)\right)^{\top}$, and $I_{N\times N}$ is the $N \times N$ identity matrix. Given our choice of the multivariate Gaussian copula, we have that $\boldsymbol{\Theta} \triangleq \boldsymbol{\rho}$ and $\mathcal{M} \triangleq \mathbb{S}_{\succeq 0}^{N\times N}$, and the CML estimation problem in (7.4) becomes

$$\widehat{\boldsymbol{\rho}} = \underset{\boldsymbol{\rho}\in\mathbb{S}_{\succeq 0}^{N\times N}}{\operatorname{argmax}} \left\{-\frac{M}{2}\ln\det(\boldsymbol{\rho}) - \frac{1}{2}\sum_{k=1}^{M}\widehat{\boldsymbol{\Xi}}^{(k)\top}\widetilde{\boldsymbol{\rho}}\,\widehat{\boldsymbol{\Xi}}^{(k)}\right\}, \tag{7.8}$$

over valid correlation matrices, with $\widetilde{\boldsymbol{\rho}} = \boldsymbol{\rho}^{-1} - I_{N\times N}$ and

$$\widehat{\boldsymbol{\Xi}}^{(k)} = \left(\Phi^{-1}\left(\widehat{u_1}^{(k)}\right),\ldots,\Phi^{-1}\left(\widehat{u_N}^{(k)}\right)\right)^{\top}.$$

After obtaining $\widehat{\boldsymbol{\rho}}$ from (7.8), we draw $\widetilde{M}$ samples residing in $[0,1]^N$ from $C(\mathbf{u};\widehat{\boldsymbol{\rho}})$, where $\widetilde{M} \gg M$. We will refer to these samples $\left\{\mathbf{u}^{(1)},\ldots,\mathbf{u}^{(\widetilde{M})}\right\}$ as *simulated observations* drawn from the fitted multivariate Gaussian copula $C(\mathbf{u};\widehat{\boldsymbol{\rho}})$. We transform these simulated observations back into conformance with the original scale of $\mathfrak{O}_M$ (e.g., airport delay-minutes for our application) via the inverse probability integral transform, through the inverse of the empirical marginal distributions found via (7.5). With a slight overload on notation, we denote these transformed simulated observations as *copula-simulated state observations*, and define the set $\widehat{\mathcal{X}} \coloneqq \left\{\mathbf{x}^{(1)},\ldots,\mathbf{x}^{(\widetilde{M})}\right\}$

as the *approximate* space of feasible states. To avoid confusion, we will denote the l^{th} historical state observations in $\mathfrak{O}_M$ as $\mathbf{x}^{(l)}_{\mathfrak{O}_M}$, in contrast to $\mathbf{x}^{(l)}$, the l^{th} copula-simulated state observation from $\widehat{\mathcal{X}}$.

7.4 Projection-based approximate control

Recall that the intuition behind our projection-based control is to drive the system signals from some currently observed state $\mathbf{x}^{(0)}_{\mathfrak{O}_M} \in \mathcal{X}$ to a (sequence) of desired state(s), dictating only performance targets in some lower-dimensional space that captures key, aggregate, system characteristics. For our choice of the lower-dimensional space of metrics, we chose a $\mathbb{R}^2$-projection of $\mathcal{X} \cup \widehat{\mathcal{X}}$ parameterized by the 1-norm of the state vector $\|\mathbf{x}\|_1$ and its *total variation* (TV) with respect to the graphical system abstraction G (e.g., G is the graph with airports as nodes and some measure of dependence as weighted edges). We use the same definition for TV as we did in Chapter 4; specifically, Definition 4 based off of the combinatorial graph Laplacian.

The choice of $\|\mathbf{x}\|_1$ is motivated by the fact that in positive signal-generating systems, this metric captures the total magnitude of signals across the entire system (e.g., the total number of bikes in a bike-share network, or total delay in an airport network). The choice of $\text{TV}(\mathbf{x}) = \mathbf{x}^\top \mathcal{L}\mathbf{x}$ reflects the fact that TV can be interpreted as a measure of signal smoothness, and can be used for outlier detection in graph signals (Chapter 4). It is worth noting that our projection-based control framework is agnostic to the specific choice of metrics: Any reasonable set of low-dimensional metrics that captures the important system performance characteristics may be used.

We define the projection $\text{proj}_{\mathbb{R}^2} : \mathbb{R}^{N\times 1} \to \mathbb{R}^{2\times 1}$ that maps $\mathbf{x} \in \mathcal{X} \cup \widehat{\mathcal{X}}$ to $\left(\|\mathbf{x}\|_1, \sqrt{\text{TV}(\mathbf{x})}\right)$. Note that the square root on TV ensures comparable dimensions between the 1-norm and TV, which is a quadratic form in $\mathbf{x}$. This transformation is permissible, as TV is a positive quantity in our setting, and the square root is one-to-one on the non-negative half-plane. The use of $\sqrt{\text{TV}(\mathbf{x})}$ simplifies the task of defining geometric constraints in $\text{im}\left(\text{proj}_{\mathbb{R}^2} \mathcal{X} \cup \widehat{\mathcal{X}}\right)$. Let $\mathbf{x}^{(0)}_{\mathfrak{O}_M}$ be the initial state of the system that has been observed, i.e., we have full knowledge of all signals $x_i^{(0)}$. We

can compute its projection in our low-dimensional space as

$$\text{proj}_{\mathbb{R}^2}\left(\mathbf{x}^{(0)}_{\mathfrak{O}_M}\right) = \left(\left\|\mathbf{x}^{(0)}_{\mathfrak{O}_M}\right\|_1, \sqrt{\text{TV}\left(\mathbf{x}^{(0)}_{\mathfrak{O}_M}\right)}\right). \tag{7.9}$$

Recall that our goal is to drive the system from $\mathbf{x}^{(0)}_{\mathfrak{O}_M}$ to some unknown terminal state $\mathbf{x}^{(T)}$ via unknown intermediate states $\mathbf{x}^{(t)}$ in discrete time steps $t = 1, \ldots, T-1$, by only constraining the $\mathbb{R}^2$-projected system metrics $\text{proj}_{\mathbb{R}^2}\left(\mathbf{x}^{(1)}\right), \ldots, \text{proj}_{\mathbb{R}^2}\left(\mathbf{x}^{(T)}\right)$ to preset performance targets. In other words, we do not specify any entries $x_i^{(t)}$ in $\mathbf{x}^{(t)}, \forall t = 1, \ldots, T$, and instead consider all *candidate* copula-simulated state observations $\mathbf{x}^{(t)}_{\text{candidate}} \in \widehat{\mathcal{X}}$ that satisfy some performance target in the $\mathbb{R}^2$-projected space, i.e., some constraint on $\text{proj}_{\mathbb{R}^2}\left(\mathbf{x}^{(t)}_{\text{candidate}}\right)$ and $\text{proj}_{\mathbb{R}^2}\left(\mathbf{x}^{(t)}\right)$.

Let us consider an example of a geometric constraint on $\text{im}\left(\text{proj}_{\mathbb{R}^2}\mathcal{X} \cup \widehat{\mathcal{X}}\right)$ that could be used to ensure a type of conformance to performance targets in the $\mathbb{R}^2$-projected space. We first select performance target *anchors* $\mathbf{a}(t) \in \text{im}\left(\text{proj}_{\mathbb{R}^2}\widehat{\mathcal{X}}\right), \forall t = 1, \ldots, T$. We assume that, via the copula-based state space estimation described earlier in Section 7.3, that the selection of the number of copula-simulated state observations $\widetilde{M}$ was large enough such that $\text{im}\left(\text{proj}_{\mathbb{R}^2}\widehat{\mathcal{X}}\right)$ is dense around anchors $\mathbf{a}(t)$. Formally, let $\mathcal{B}\left((x_1, x_2); \varepsilon\right)$ be a ball in $\mathbb{R}^2$ with radius ε centered at (x_1, x_2). Then, for a small positive radius $\varepsilon > 0$, there exists a $\widetilde{M}_\varepsilon = \left|\widehat{\mathcal{X}}\right|$ such that

$$\mathcal{B}\left(\mathbf{a}(t); \varepsilon\right) \cap \text{im}\left(\text{proj}_{\mathbb{R}^2}\widehat{\mathcal{X}}\right) \neq \{\varnothing\}. \tag{7.10}$$

The anchors $\mathbf{a}(t)$ are the prescribed system performance targets $\text{proj}_{\mathbb{R}^2}\left(\mathbf{x}^{(t)}\right)$ that candidate copula-simulated states $\mathbf{x}^{(t)}_{\text{candidate}} \in \widehat{\mathcal{X}}$ must adhere to at each corresponding time step t. We then solve the optimization problem in (7.11), where the geometric constraint encodes our requirement that the candidate copula-simulated states do not deviate more than $\delta_t > 0$ in Euclidean norm from anchor $\mathbf{a}(t)$ at time step t.

$$
\begin{aligned}
\mathbf{x}_*^{(t)}: \quad & \underset{\mathbf{x}^{(t)} \in \widehat{\mathcal{X}}}{\operatorname{argmin}} && \left\| \mathbf{x}^{(t)} - \mathbf{x}_*^{(t-1)} \right\|_2 \\
& \text{s. t.} \quad \delta_t && \geq \left\| \mathbf{a}(t) - \operatorname{proj}_{\mathbb{R}^2} \left(\mathbf{x}^{(t)} \right) \right\|_2 \\
& \mathbf{x}_*^{(0)} && = \mathbf{x}_{\mathfrak{O}_M}^{(0)} \\
& \forall t && = 1, \ldots, T.
\end{aligned}
\tag{7.11}
$$

The objective function reflects the fact that $\operatorname{proj}_{\mathbb{R}^2}$ is surjective, i.e., multiple candidate copula-simulated states could feasibly satisfy the geometric constraint. Although we do not know the underlying state dynamics from t to $t+1$, assuming we select a reasonable time step, we prefer small changes at each vertex signal $x_i^{(t)}$. These assumptions are reasonable in real applications: Reflecting back on our bike-share network example, if we assume that the duration between time steps is 15 minutes, we should not expect a large variation in the number of bikes at a particular station. The same assumptions can be constructed for airport delays, given an appropriate time interval. The objective function in (7.11) retrieves the current state that requires a *minimal-energy* evolution from the fully-known previous state $\mathbf{x}_*^{(t-1)}$. Note that $\mathbf{x}_*^{(t-1)}$ is fully known because it is either the solution to the preceding optimization, or the known initial condition $\mathbf{x}_*^{(0)} = \mathbf{x}_{\mathfrak{O}_M}^{(0)}$.

We define $\mathbb{T}_* \coloneqq \left\{ \mathbf{x}_*^{(0)} = \mathbf{x}_{\mathfrak{O}_M}^{(0)}, \mathbf{x}_*^{(1)}, \ldots, \mathbf{x}_*^{(T)} \right\}$ as the extrapolated system state trajectory starting at $\mathbf{x}_*^{(0)} = \mathbf{x}_{\mathfrak{O}_M}^{(0)}$, obtained from solving (7.11) at each time step $t = 1, \ldots, T$. The control policy $\boldsymbol{\pi}_*^{(t)}$ at time $t = 0, \ldots, T-1$ is simply $\boldsymbol{\pi}_*^{(t)} = \mathbf{x}_*^{(t+1)} - \mathbf{x}_*^{(t)}$, with the initial condition $\mathbf{x}_*^{(0)} = \mathbf{x}_{\mathfrak{O}_M}^{(0)}$. We note that for the setup in (7.11), it is possible to define a feasible set on the space of control actions by observing that

$$
\boldsymbol{\pi}_{\min} \preceq_{\mathbb{R}} \boldsymbol{\pi}_*^{(t)} \preceq_{\mathbb{R}} \boldsymbol{\pi}_{\max}, \tag{7.12}
$$

where $\preceq_{\mathbb{R}}$ denotes the element-wise inequality, $\boldsymbol{\pi}_{\min} = (\pi_{\min,1}, \ldots, \pi_{\min,N})^\top$, and $\boldsymbol{\pi}_{\max} = (\pi_{\max,1}, \ldots, \pi_{\max,N})^\top$ with

$$
\begin{aligned}
\pi_{\min,i} &= \inf \left\{ x_i^{(\tau)} - x_i^{(\sigma)} \,\middle|\, x_i^{(\tau)}, x_i^{(\sigma)} \in \mathcal{X} \cup \widehat{\mathcal{X}} \right\}, \\
\pi_{\max,i} &= \sup \left\{ x_i^{(\tau)} - x_i^{(\sigma)} \,\middle|\, x_i^{(\tau)}, x_i^{(\sigma)} \in \mathcal{X} \cup \widehat{\mathcal{X}} \right\}.
\end{aligned}
$$

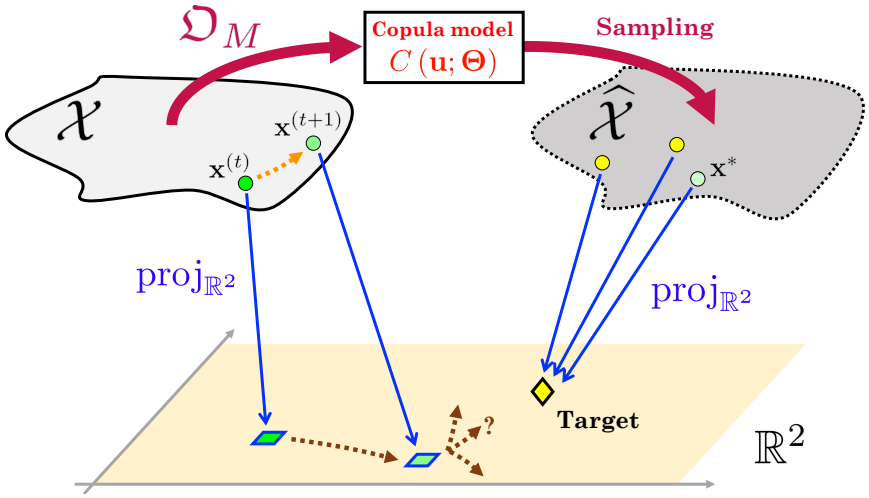


Figure 7-4: Pictorial representation of our methodology, specifically the copula-based state space model (Section 7.3) and the projection-based approximate control framework (Section 7.4). *Reprinted from [155]. © 2020 IEEE*

7.5 Methodology discussion

7.5.1 Interpretation

Our framework provides a control policy for a dynamical system whose model is not explicitly known. Our two-step approach of first generating an approximate feasible state space along with the subsequent optimization mapping between low-dimensional performance targets and high-dimensional states is a proxy for some dynamics $\mathbf{x}^{(t)} = h\left(\mathbf{x}^{(t-1)}\right)$, where h is state-dependent. The optimization procedure in (7.11) can be thought of as a pseudo-inverse projection $\text{proj}_{\mathbb{R}^2}^{-1} : \mathbb{R}^{2\times 1} \to \widehat{\mathcal{X}}$ which selects a unique high-dimensional system state (by way of the objective function in (7.11) being convex; we prove this fact for an extended version of (7.11) in Section 7.6, but it is straightforward to see that (7.11) is convex given that its objective function is an Euclidean norm, and the constraints are all convex as well) from the copula-approximated state space that satisfy performance target constraints.

7.5.2 Generalizability

Furthermore, many points along the workflow of this framework (see Figure 7-1) can be modified to accommodate a range of general applications. Specifically, we highlight five aspects of versatility provided by our framework: (1) The choice and selection of a copula family; (2) the form of the objective function; (3) the cost structure; (4) the constraints enforcing low-dimensional performance targets; and (5) the choice of projection metrics. In terms of (1), although we chose to use the family of multivariate Gaussian copulas, there have been interesting results on copula goodness-of-fit testing [86] and avoiding copula family misspecification [151]. While copula family selection is outside the scope of this work, our framework only requires that a copula is available to be sampled from, regardless of its specific family.

With regards to (2)-(5), we will see that the *Conservative Selective Redistribution Problem* (CSRP) for airport delays (7.13) that we formulate in Section 7.6 was directly extended from (7.11). Suppose, instead, that we want to prescribe control actions over a bike-share network, where certain vertices (bike stations) are rewarded for maintaining a steady inventory of bikes during the bike redistribution process. We could easily adapt a version of this problem from the airport delay CSRP in (7.13) by changing the sign on the third term of the objective function. Furthermore, we could assign priorities to different vertices by switching to a non-binary cost vector in (7.14). Now suppose that the bike-share system operator has access to bike depots located around the city, effectively allowing for a certain number of sink- and source-vertices in the network. This new degree of freedom can be captured by relaxing the conservation constraints in (7.13), i.e., by choosing a more lenient buffer δ. Lastly, our choice of $\text{proj}_{\mathbb{R}^2}$ was made based on our preference of two aggregate system performance measures. The system operator could alternatively choose a modified $\widetilde{\text{proj}}_{\mathbb{R}^2}$ that captures other aggregate performance measures. Some canonical examples would include $\mathbb{E}\left[\mathbf{x}^{(t)}\right]$, $\text{Var}\left[\mathbf{x}^{(t)}\right]$, and $\left\|\mathbf{x}^{(t)}\right\|_\infty$.

7.5.3 Technical remarks and caveats

When choosing the low-dimensional $\mathbb{R}^2$-projected state trajectory that satisfies performance targets, we use a greedy approach at each time step in (7.11). We could have alternatively constructed a set of trajectory options, and then performed a breadth- or depth-first search to select one with the lowest cost. Furthermore, even though the projection operator $\text{proj}_{\mathbb{R}^2}$ we use is smooth and continuous, these conditions are sufficient, but may not be necessary, conditions. An interesting question for future research would be to characterize necessary and sufficient conditions on $\text{proj}_{\mathbb{R}^2}$ such that small perturbations in the low-dimensional projection space do not translate to large perturbations in $\mathcal{X} \cup \widehat{\mathcal{X}}$. Specifically, one way to improve the projection-based framework may be to find a projection operator that preserves distances (within bounds), e.g., the Johnson-Lindenstrauss lemma [6, 135], between two points post-projection while retaining useful network performance interpretations.

Theorem 3 (Johnson-Lindenstrauss (Dasgupta-Gupta)) *For any $0 < \epsilon < 1$ and any integer n, let k be a positive integer such that*

$$k \geq \frac{4 \ln n}{\epsilon^2/2 - \epsilon^3/3}.$$

Then for any set V of n points in $\mathbb{R}^d$, there is a map $f : \mathbb{R}^d \to \mathbb{R}^k$ such that for all $\mathbf{u}, \mathbf{v} \in V$,

$$(1 - \epsilon) \left\| \mathbf{u} - \mathbf{v} \right\|^2 \leq \left\| f(\mathbf{u}) - f(\mathbf{v}) \right\|^2 \leq (1 + \epsilon) \left\| \mathbf{u} - \mathbf{v} \right\|^2.$$

Furthermore, this map can be found in randomized polynomial time.

Proof. See [59]. □

It is important to bear in mind that copulas cannot provide more information than what is encoded in the historical observations. Furthermore, copulas do not capture any time-varying information; they are merely a particular form of a multivariate probability distribution. They should not be interpreted as stochastic processes. In our CSRP formulation and US NAS case study in Section 7.6 and 7.7, we take

into account time-varying dependence structures by calibrating different copulas for different hours of the day.

7.6 The Conservative Selective Redistribution Problem (CSRP)

We now turn our attention to adapting the general framework from Sections 7.3 and 7.4 to the airport delay redistribution problem. Recall that in our projection-based network control framework, we need to select an optimal candidate from the copula-generated state space $\widehat{\mathcal{X}}$. We now provide the CSRP formulation, wherein the optimal solution to the CSRP at each time step provides the optimal candidate. We solve the CSRP sequentially, starting at $t = 1$ and with the initial condition anchored at the first state in trajectory $\mathbb{T}$. The formulation of the CSRP is as follows:

$$\begin{aligned}
\underset{\mathbf{x}^{(t)} \in \widehat{\mathcal{X}}}{\arg\min} \quad & \left\{ \left\| \mathbf{x}^{(t)} - \mathbf{x}_*^{(t-1)} \right\|_2 + \lambda \tfrac{\mathbb{1}^\top \mathbf{c}}{N} \left\| \mathbf{x}^{(t)} - \mathbf{x}_{\mathfrak{O}_M}^{(t)} \right\|_2 + (1-\lambda) \mathbf{c}^\top \left(\mathbf{x}^{(t)} - \mathbf{x}_{\mathfrak{O}_M}^{(t)} \right) \right\} \\
\text{s. t.} \quad & \left\| \mathbf{x}^{(t)} \right\|_1 \geq \left\| \mathbf{x}_{\mathfrak{O}_M}^{(t)} \right\|_1 - \delta, \\
& \left\| \mathbf{x}^{(t)} \right\|_1 \leq \left\| \mathbf{x}_{\mathfrak{O}_M}^{(t)} \right\|_1 + \delta, \\
& \mathbf{x}_*^{(0)} = \mathbf{x}_{\mathfrak{O}_M}^{(0)}, \\
& \lambda \in [0, 1], \\
& \forall t = 1, \ldots, T.
\end{aligned} \tag{7.13}$$

There are three penalty terms in the objective function in (7.13), with the first term $\left\| \mathbf{x}^{(t)} - \mathbf{x}_*^{(t-1)} \right\|_2$ assigning a base cost that enforces *smooth* transitions from one state to another. In reality, for a sensibly chosen time interval, the delays at a particular airport should not vary drastically: The first term in the objective function penalizes large changes in the optimal candidate state compared to the previous optimal solution at time step $t-1$. We refer to this first term as a base cost because it is always incurred; however, it may be possible to multiply this term by some discount factor γ^t where $\gamma \in (0, 1]$, since the assumption of smooth transitions may

be less important given uncertainty about time steps in the far future.

We will refer to λ as a *redistribution workload parameter*, as it controls the importance between adherence to historically-observed airport delay states (given by the second term) versus redistributing delays away from select airports (given by the third term). In one extreme regime, if $\lambda = 1$, then no effort is made to redistribute delays, and the CSRP attempts to *match* $\mathbb{T}$. On the other hand, if $\lambda = 0$, then full effort is made to redistribute delays, with no penalty on not adhering to historically observed airport delay states. Note that since $\lambda \in [0, 1]$, we must ensure that the second and the third term are of comparable scales, given that they both have units of delay minutes. The magnitude of the second term is generally larger than the third term, assuming that the cost vector $\mathbf{c}$ is unit and sparse. We ensure this by using the following form for the cost vector $\mathbf{c}$: Let $S \subset V$ be the set of airports at which delay dissipation is encouraged, and ensure that $|S| < |V|$. Then, construct $\mathbf{c}$ such that:

$$\mathbf{c} = [c_i] = \begin{cases} 1 & \text{if } i \in S \subset V, \\ 0 & \text{otherwise.} \end{cases} \tag{7.14}$$

We add a constant re-scaling factor to the second term of the form $\mathbb{1}^\top \mathbf{c} N^{-1} \in [0, 1]$ to account for the fact that the third term is only evaluated over certain airports, whereas the second term is evaluated over the entire state vector. Note that if $|S| < |V|$, we must have that $\mathbb{1}^\top \mathbf{c} N^{-1} \in [0, 1)$.

The delay conservation constraints in the CSRP formulation is straightforward, requiring that all optimal state candidates have total delays that lie within a small δ-band of the historically-observed total delay. When running our case studies in Section 7.7, we pick a small-enough δ-tolerance such that the CSRP remains feasible (i.e., there is a non-empty set of candidate states) while ensuring that the delay conservation constraints are enforced in essence (i.e., pick δ such that the total delay at an airport factoring in the δ-tolerance does not vary significantly from the actual total delay, across timescale t). The CSRP formulation will provide a global optimal solution, given that it is convex:

Proposition 8 (CSR problem convexity) *The CSRP given in* (7.13) *is a convex optimization problem.*

Proof. See Appendix A.9. □

7.6.1 Calculating delay absorption costs

One question that could be answered through a minor adjustment of the CSRP formulation in (7.13) relates to the *cost* associated with delay absorption at a particular airport within the network. While there are many ways of measuring this cost, we will work with a mechanistic definition of cost as given in Definition 14:

Definition 14 (Delay absorption cost at airport i) *Set $\lambda = 0$ to encourage maximum redistribution, and re-formulate the cost vector $\mathbf{c}$ in the CSRP* (7.13) *to the absorption cost vector $\mathbf{c}_{\text{absp}}$ as follows:*

$$\mathbf{c}_{\text{absp}} = [c_j] = \begin{cases} -1 & \text{if } j = i, \\ 0 & \text{otherwise.} \end{cases}$$

Then, the delay absorption cost at airport i specific to some trajectory $\mathbb{T}$ is defined as the sum of the costs associated with each solution of the modified CSRP with $\mathbf{c}_{\text{absp}}$, at each time step $t = 1, \ldots, T$ across $\mathbb{T}$.

By way of the new absorption cost vector as redefined in Definition 14, it provokes the CSRP into attempting to redistribute delays to airport i at each time step across $\mathbb{T}$. This is because the objective value is reduced by a solution where the proposed delay is higher at airport i during time t compared to its historical value. The airport with the minimized delay absorption cost is the "easiest" airport to redistribute delays to, given the network structure (i.e., the marginal delay distributions at each airport and the dependence structure linking them together). Note that the units on this cost measure is in delay minutes, and the cost measure can be negative. By cycling through all airports i individually and computing their delay absorption cost using the modified CSRP with the absorption cost vector from Definition 14, we obtain a

ranked list of least- to most-costly delay-absorbing airports, *specific to a disruption sequence (i.e., trajectory, or DRT)* $\mathbb{T}$. In Section 7.7, we compute this cost measure across all 30 US airports specific to a particular disruption, as well as an *average* delay absorption costs across clusters (i.e., the DRT clusters from Section 6.4) of disruption sequences.

7.7 Deploying the CSRP: US NAS case study

7.7.1 Data description and processing

We use hourly airport delay data obtained from FAA ASPM [85], and train the hour-specific copula models on airport delay data from 2008-2017. We represent our network via a complete, undirected graph $G = (V, E)$. In terms of the airport network we construct, we pick the $|V| = 30$ busiest US airports in terms of passenger enplanements in 2017. The undirected edges are weighted using the sample Pearson correlation coefficient based on the 2008-2017 airport delay data. Note that this can be taken to be a non-negative system, as the hourly delays at any airport for the entire duration of the 10-year data set are non-negative. Finally, since 3,653 data observations were used to construct the hour-specific copula models (one hour-specific observation per day, across 2008-2017), we select a much larger number of samples $N_{\widehat{\mathcal{X}}} = 100,000$ to construct $\widehat{\mathcal{X}}$. Recall that the need for hour-specific copulas is due to the fact that an individual copula contains no temporal information: Since airport delays are strongly dependent on the hour-of-day (e.g., because of heavy traffic during the morning and afternoon periods), we construct 24 different approximate state spaces $\widehat{\mathcal{X}}_0, \ldots, \widehat{\mathcal{X}}_\tau, \ldots, \widehat{\mathcal{X}}_{23}$, each from a copula model based on historical observations of $\mathbf{x}^{(t)}$ belonging to hour τ. Similarly, we have 24 different graph Laplacians $\mathcal{L}_0, \ldots, \mathcal{L}_\tau, \ldots, \mathcal{L}_{23}$, corresponding to the different hourly airport delay correlation networks.

7.7.2 New York City redistribution example

We demonstrate our redistribution framework on a particular DRT spanning 11 hours, from 1 pm EDT on May 22, 2014 through midnight on May 23, 2014. Recall from Section 6.1 that a DRT is a chronologically-ordered set of network performance metric, specifically, total delay and total variation, that captures the evolution of the magnitude and spatial distribution of airport delays. Intuitively, DRTs project the state of the system in a qualitatively interpretable manner, via a "phase portrait" representation. This particular 11-hour DRT was extremely disruptive to the US NAS, due to convective weather moving off-shore in the ZBW ARTCC, along with emerging pop-up thunderstorms and convective activity near the ZOB and ZNY ARTCC border. Along with various airspace flow programs (AFPs) around ZNY, there were a total of 63 ground stop (GS)-related advisories and 19 ground delay program (GDP)-related advisories issued on May 22, 2014 [84].

Starting with the airport delay state at 1 p.m. EDT, we solve the CSRP in (7.13) for $t = 1$ (2 p.m. EDT) through $t = 11$ (midnight EDT), taking care to use the correct hour-specific copula model. On a technical note, for the delay conservation constraint buffer δ in (7.13), we select a δ-tolerance small enough so that we have a non-empty set of candidate solutions, while still enforcing the delay conservation constraints in practice. Specifically, we set δ such that

$$\delta = 0.01 \times \max \left\{ \left\| \mathbf{x}_{\mathbb{T}_\diamond}^{(t)} \right\|_1 \, \middle| \, \mathbf{x}_{\mathbb{T}_\diamond}^{(t)} \in \mathbb{T}_\diamond \right\}. \tag{7.15}$$

By selecting δ according to (7.15), we set the tolerance to 1% of the maximum total delay observed across any time step within a given DRT $\mathbb{T}_\diamond$. This also means that the buffer will dynamically adjust with respect to the DRT that the CSRP in being deployed on. In practice, for this CSRP deployment example, we have that $\delta \approx 19.6$ minutes: This is a very small amount of delay relative to the total delay at any hour, and indicates that the delay conservation constraints in the CSRP will be enforced.

We configure our cost vector $\mathbf{c}$ in accordance with (7.14) using

$$S = \{\text{EWR, JFK, LGA}\} \subset V.$$

In other words, we configure the third term in (7.13) to reward control policies that shift delay away from the three New York City (NYC) airports. Finally, we repeat the entire sequence of optimizations (i.e., re-run the CSRP) for differing $\lambda = \{0, 0.5, 1\}$. Recall that $\lambda = 0$ represents maximal efforts to shift delay away from the NYC airports with no penalties on not adhering to the original DRT, whereas $\lambda = 1$ represents maximal adherence to the DRT $\mathbb{T}_\diamond$ with no preference on redistributing delay away from the NYC airports. Decreasing λ is a proxy for increasing the workload of air traffic flow managers, since it requires greater efforts towards delay redistribution.

We plot the original delay signals at EWR, LGA, and JFK, along with the CSRP solutions for various redistribution workload parameters in Figure 7-5. We see that for the purple-colored delay signals ($\lambda = 0$), there is a significant dissipation of delays across all three NYC airports for the duration of the DRT. On the other hand, for the orange-colored delay signals ($\lambda = 1$), the delay signals track the baseline (i.e., historic) delay signals quite closely. This verifies the role of the λ redistribution workload parameter, and shows that the solution to the sequentially-applied CSRP across the DRT successfully dissipates delays at the select subset consisting of NYC airports.

Since the delay conservation constraints are upheld, the natural question is to ask where the dissipated delays are reabsorbed in the system. To this end, we leverage the definition of the delay absorption cost (Definition 14), and compute the airport-specific delay absorption costs across this specific DRT. These delay absorption costs can be seen in Figure 7-6. We see that ATL, DFW, and IAH are the least-costly airports to absorb the additional delays incurred from dissipation at NYC, and this is further confirmed in Figure 7-7 where we plot the delay signals at ATL, DFW, and IAH. Interestingly, there seems to be some temporal differences in terms of when the absorption occurs, with ATL and DFW having significant peaks at slightly offset times, whereas IAH maintains a more modest, but longer-duration absorption. This

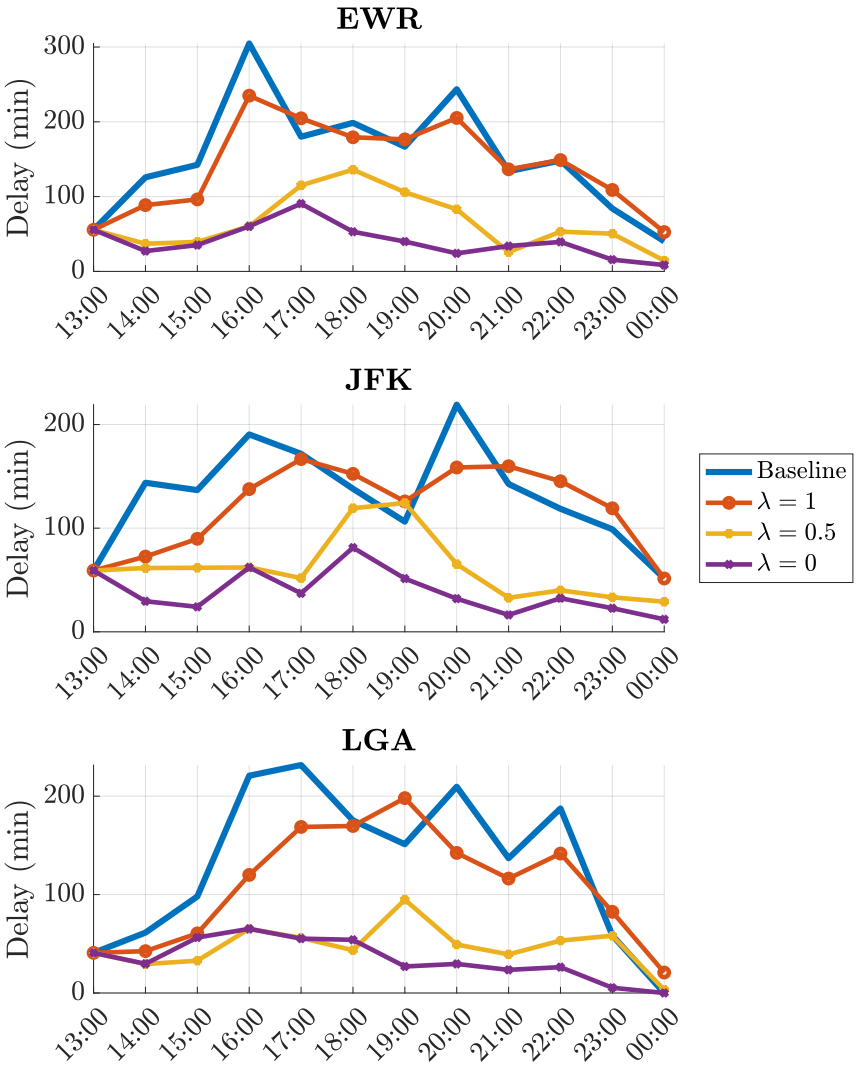


Figure 7-5: Delay state signals at EWR, LGA, and JFK across three separate solutions of the CSRP for static redistribution workload parameter λ values.

potentially has implications when one considers the possibility of *dynamic* redistribution workload parameters, specifically in terms of redistribution equity. We discuss the issue of distribution equity in more detail in Section 7.8.

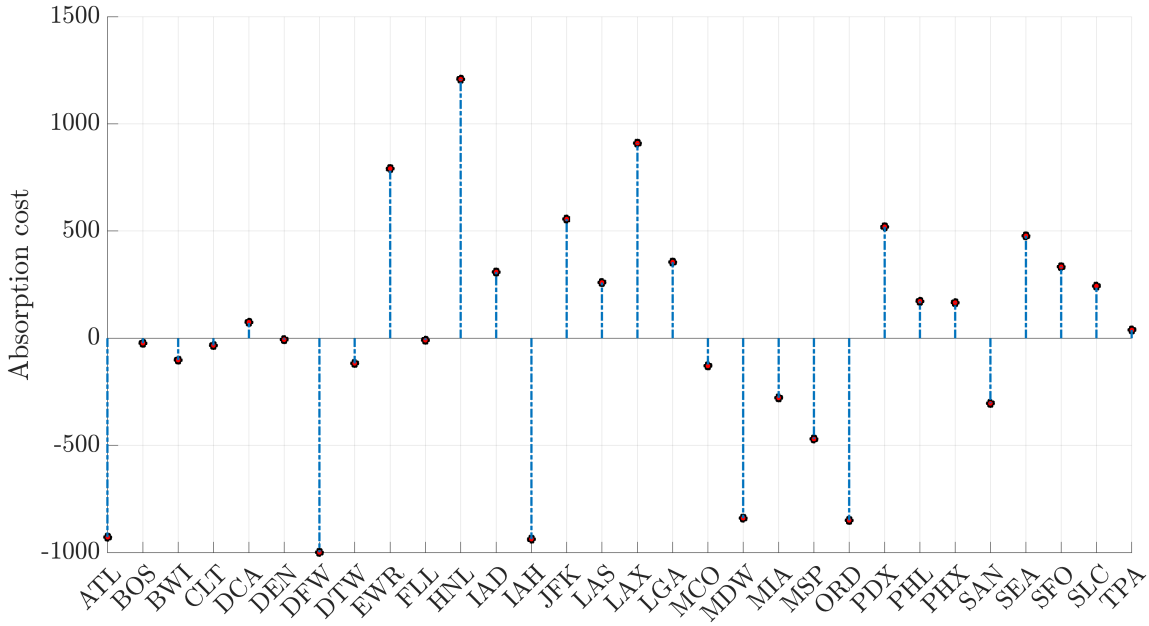


Figure 7-6: Delay absorption costs associated with each airport in the network for the NYC case study DRT.

While common least-costly delay absorption airports are highly connected hubs in more central or US East Coast locations, we see that airports on the West Coast and HNL are particularly difficult (i.e., costly) to redistribute delays to. This is likely due to the fact that, in terms of correlation-based dependence, there are two significant cliques of high correlations among the US Core 30 airports: East Coast versus West Coast (see Figure 4-10 and accompanying discussions in Chapter 4). In other words, the US NAS is somewhat decoupled between the two coasts, which most likely reflects the operational realities of transcontinental flights being less likely to be impacted by scope-based traffic management initiatives. Thus, while redistribution within one clique might be easy, redistributing across cliques is much more costly. Figure 7-8 reflects this: In all CSRP solutions, regardless of λ, the delay signals at these expensive (in terms of absorbing delays) airports do not vary much from baseline values.

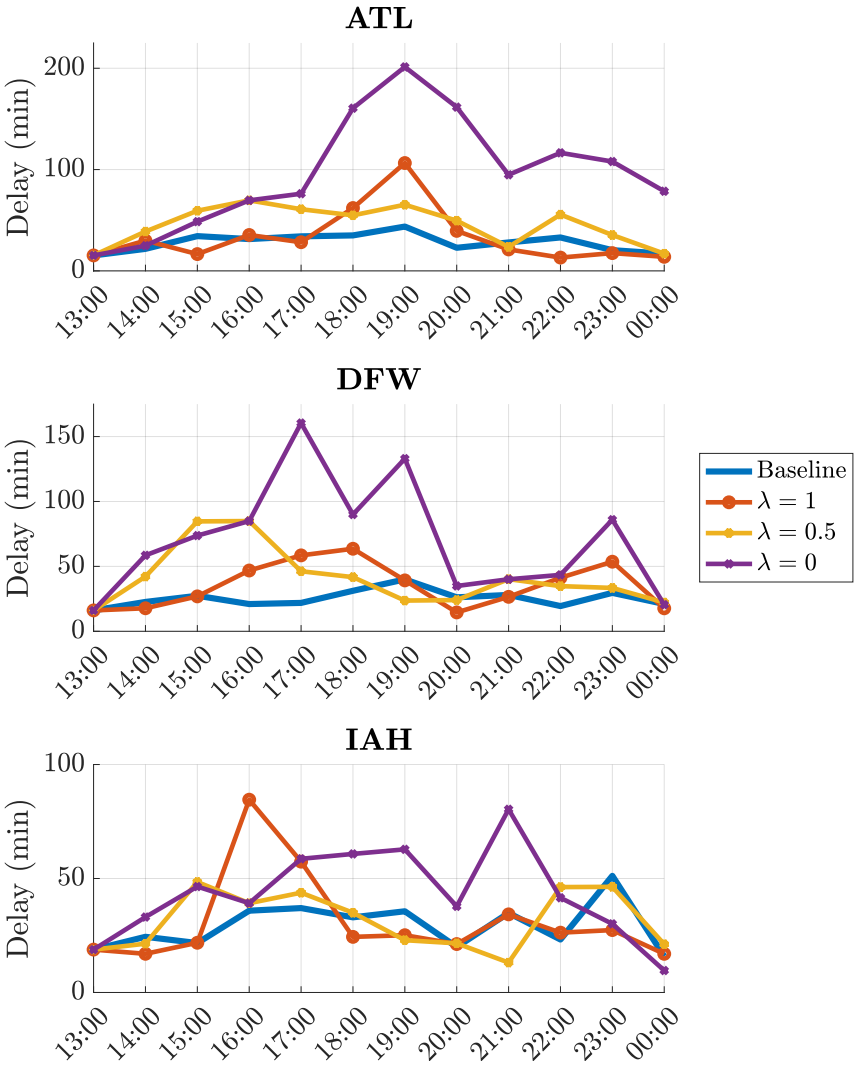


Figure 7-7: Delay state signals at the three most-absorbent airports (i.e., airports with the least expensive delay absorption cost) in terms of delays (ATL, DFW, IAH), specific to the NYC case study DRT, across three separate solutions of the CSRP for static redistribution workload parameter λ values.

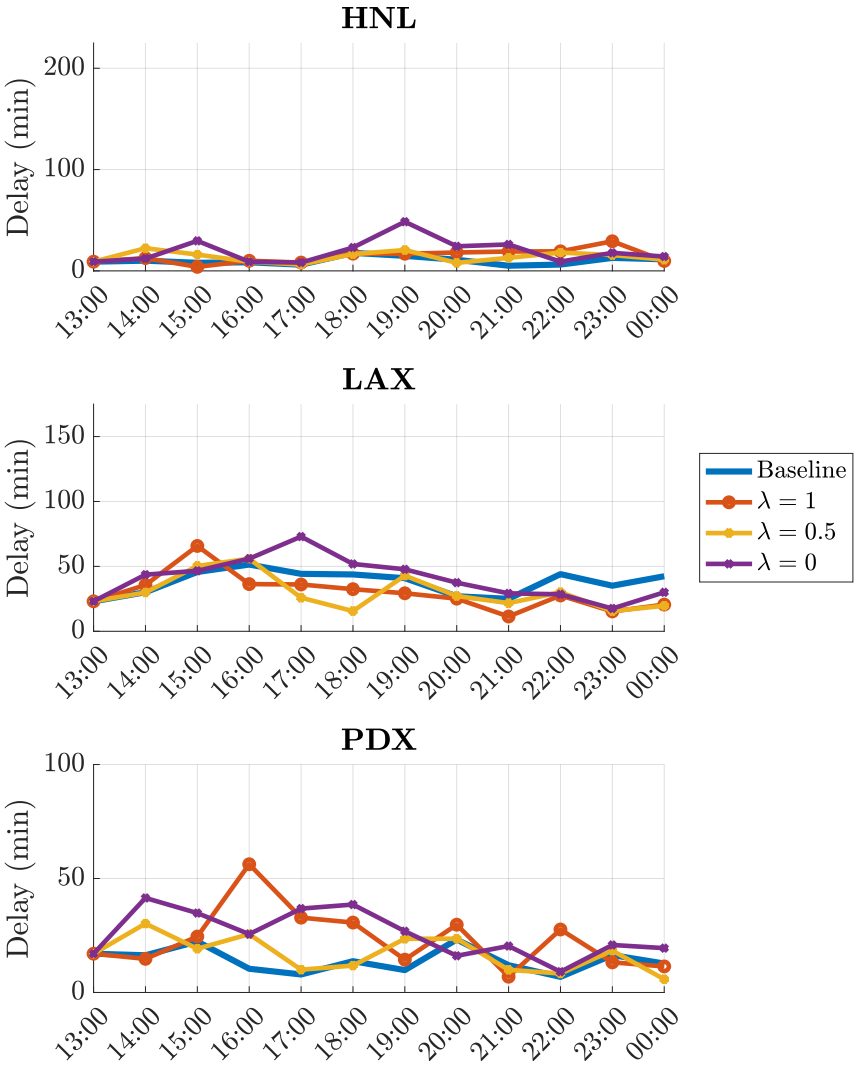


Figure 7-8: Delay state signals at the three least-absorbent airports (i.e., airports with the most expensive delay absorption cost) in terms of delays (HNL, LAX, PDX), specific to the NYC case study DRT, across three separate solutions of the CSRP for static redistribution workload parameter λ values.

7.7.3 Average delay absorption costs across DRT clusters

The delay absorption costs in Figure 7-6 are specific to the NYC DRT case study; here, we examine *average* delay absorption costs across specific DRT clusters from Section 6.1. Recall that these DRT clusters were retrieved based off of features related to operational and DRT-specific geometric characteristics, and was a major focus in Chapter 6. In particular, we examine the average delay absorption costs for three clusters of DRTs: (1) Operational day-long DRTs with spatially perturbed disruptions (`OpsDay_Dis`-type DRTs), (2) operational day-long DRTs with spatially perturbed recoveries (`OpsDay_Rec`-type DRTs), and (3) DRTs that span multiple days (`MultiDay`-type DRTs). The first and second types describe DRTs that last for a significant portion of an operational day within the NAS – specifically, the duration of these DRTs average between 15 to 18 hours. The difference between the first and second group is that the former is characterized by spatial delay distributions that are unexpected during the disruption phase, whereas the latter is characterized by unexpected spatial delay distributions during the recovery phase. The average costs across these two types of DRTs are shown in Figures 7-9 and 7-10, respectively.

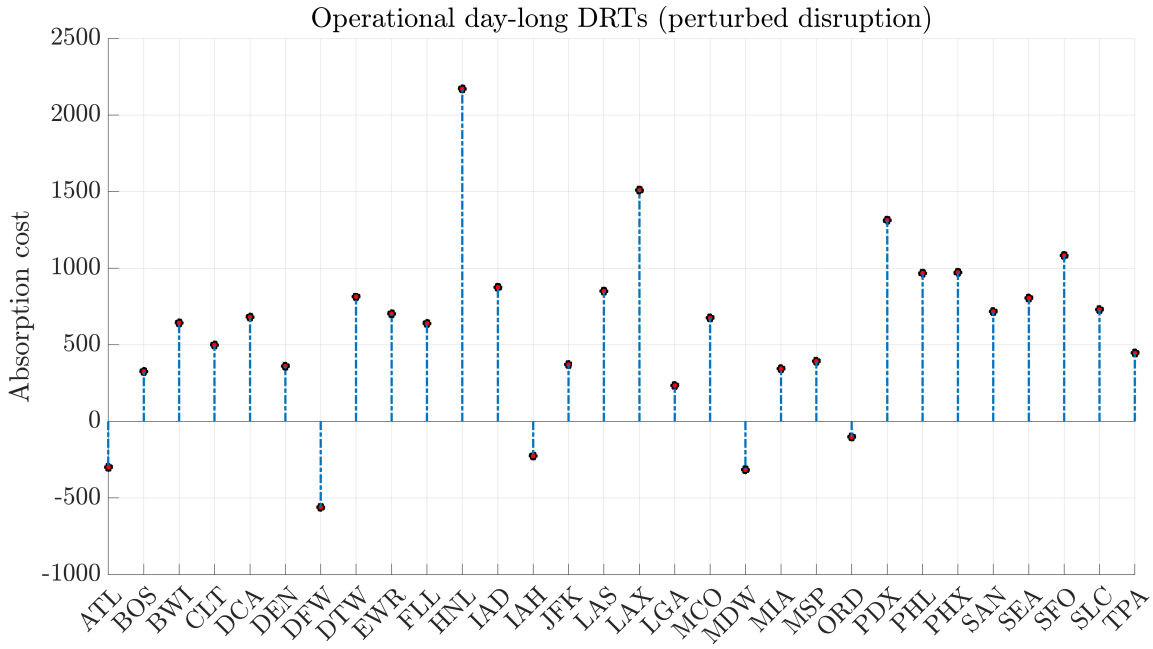


Figure 7-9: Average delay absorption costs associated with each airport in the network for all `OpsDay_Dis`-type DRTs.

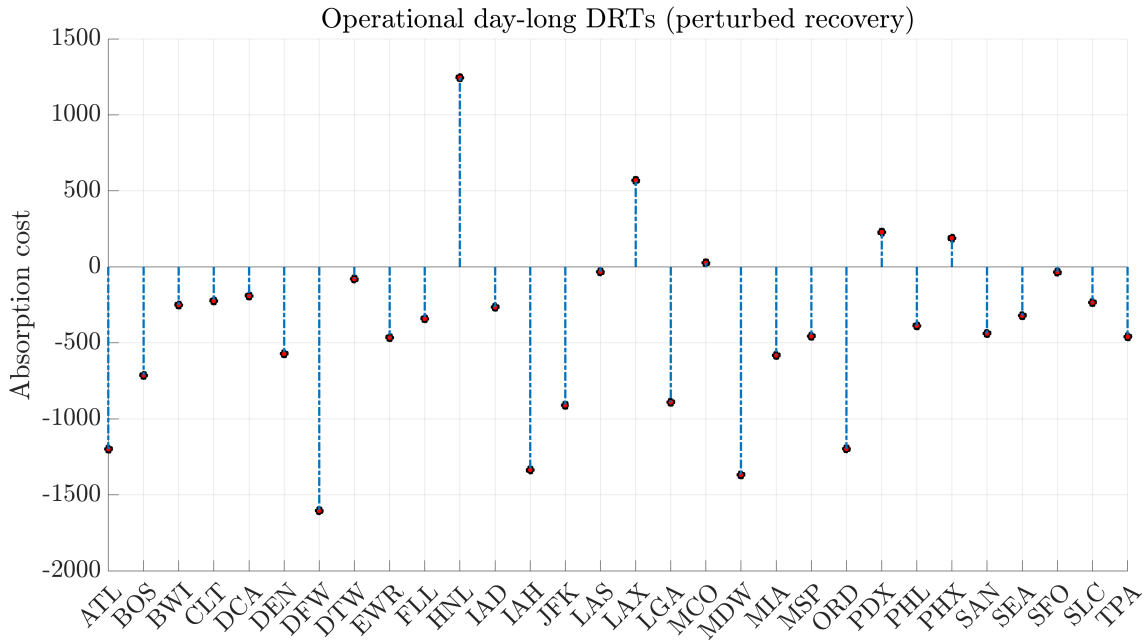


Figure 7-10: Average delay absorption costs associated with each airport in the network for all `OpsDay_Rec`-type DRTs.

We see an interesting juxtaposition in terms of the absorption costs: In Figure 7-9, most airports have positive values for absorption costs, and in Figure 7-10 most have negative values. This indicates that, across these two DRT types, spatially perturbed disruptions translate to the CSRP finding higher-cost redistribution control policies compared to spatially perturbed recoveries. In other words, it is more difficult to find redistribution policies when the spatial variance in delay signals is higher during the disruption phase than during recovery phases. Moving forward, this has the implication that if an ongoing DRT in its disruption phase has lower spatial variance, it may be worthwhile to implement redistribution policies immediately rather than waiting to act during the eventual recovery. On the other hand, if the spatial variance is already high during the ongoing disruption phase, it may be more worthwhile to wait for the recovery phase.

If we look now at the third cluster of DRTs whose average delay absorption costs are given in Figure 7-11, we first note that the duration of these DRTs spans multiple days: These DRTs represent prolonged disruption and recovery events, indicating that the system did not recover even during overnight hours, when there is typically

enough slack in the system to absorb any excess delay. Overall, we note that the costs for `MultiDay`-type DRTs are more similar to spatially perturbed recoveries than spatially perturbed disruptions.

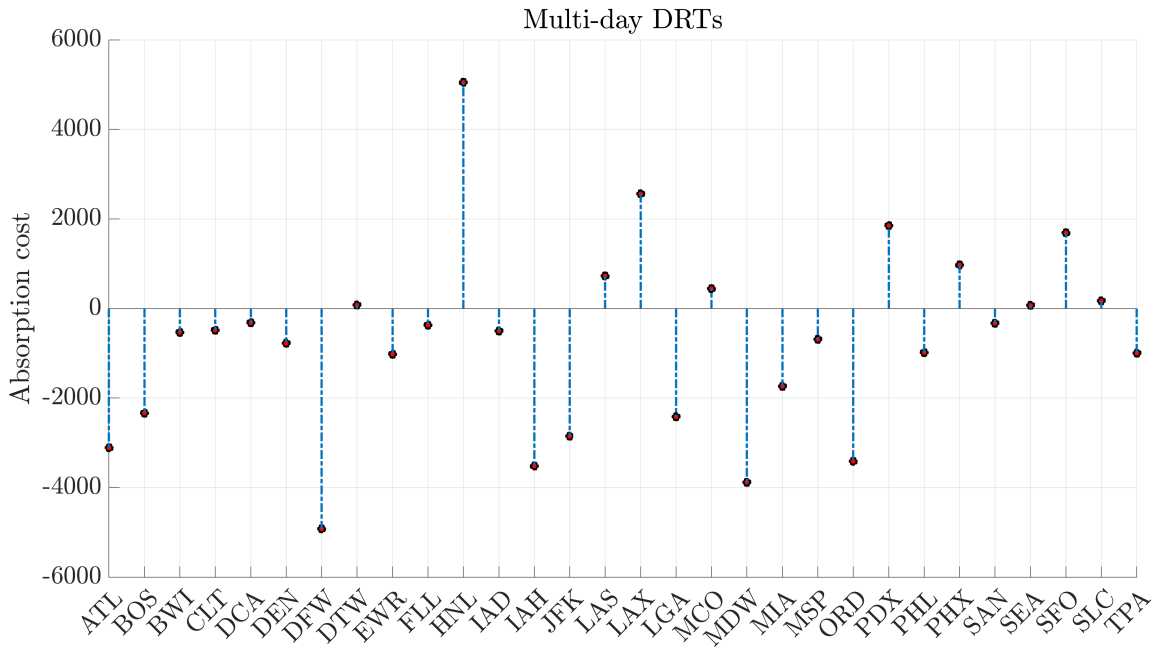


Figure 7-11: Average delay absorption costs associated with each airport in the network for all `MultiDay`-type DRTs.

We do point out the caveat that the standard deviations on these average absorption costs can be quite high, indicating that even within clusters of similar DRTs, the variance in terms of disruptions and recoveries can be quite large. An interesting point of further analysis would be to relate disruption categories back to absorption costs. We know from Section 6.6 that different disruption categories as well as seasonal trends (i.e., month-of-year) are composed of differing fractions of DRT types. Integrating this observation with absorption costs could lead to better strategies in terms of issuing traffic management initiatives during certain disruption events or commonly-occurring, seasonally-entrenched weather patterns. Specifically, these traffic management initiatives would seek to redistribute delay, if the first priority of total dissipation from the system is not immediately achievable (e.g., without massive cancellations).

7.8 Discussion: US NAS delay redistribution results

7.8.1 Delay redistribution mechanisms

Recall that network-level observations of delay redistribution scenarios result from aggregations of multiple scenarios similar to the example described in Section 7.2 happening simultaneously or near-simultaneously across the NAS. Ultimately, airline schedule adjustments are responses to initiations, adjustments, and cancellations of traffic management initiatives set by the FAA during the Collaborative Decision Making (CDM) process. This implies that one of the main mechanisms to induce certain redistribution patterns may be through adjustments of traffic management initiatives such as ground delay programs (GDPs). Taking GDPs as the prototypical control mechanism, one parameter with far-reaching consequences is the *scope* of the GDP, indicating what airports (generally at an ARTCC level) are impacted by a specific GDP [83]. In fact, other initiatives such as ground stops can even be issued by airlines themselves, directed towards their own fleet, with additional nuances such as differentiating between mainline versus regional carrier ground stops.

In terms of redistribution, if we examine the most-absorbent airports from the NYC case study (ATL, DFW, and IAH), any redistribution mechanism result in increased delays at ATL, DFW, and IAH while simultaneously reducing delays at the three NYC airports. Besides changing the impact scope of traffic management initiatives, one could also view airline cancellations as a control action, counting it as a departure delay penalty at the origin airport. In this way, the redistribution action would prioritize NYC-bound flights at ATL, DFW, and IAH for proactive cancellation in order to reduce arrival delays accrued at NYC airports. This might be particularly relevant for network legacy carriers with strong hub presences, but would also raises an interesting case for airlines that practice more point-to-point or circuit-based routings. For the latter class of airline network structures, suppose that an airline (e.g., Southwest Airlines) schedules the same tail to perform LAX-

ATL, followed by ATL-LGA. In this case, redistribution would be straightforwardly assessed at the LAX-ATL leg, resulting in a shift of delays away from LGA and onto LAX and ATL instead.

7.8.2 Redistribution equity

Even though the least-costly subset of airports in terms of delay absorption costs will differ between each DRT (indeed, the order will differ even at each time step t within one run of the CSRP), it is clear from Section 7.7 that certain airports seem to be less costly on average across a wide range of disruptions. While one could argue that since these airports (ATL, DFW, IAH, MDW, and ORD) are serviced by a large range of network legacy carriers, low cost carriers, and ultra low cost carriers, there is, in some sense, an equitable share of delay absorption burdens across airlines. In other words, if the solution to the CSRP typically prefers to redistribute delays to this subset of airports, it may be equitable in that most carriers would be affected, and no single carrier would be impacted more severely than others.

However, this may not hold true if the redistribution workload parameter λ was allowed to dynamically vary throughout the course of a disruption, something that might be allowed in a more realistic implementation of the CSRP. For example, it may not be worth attempting to drastically redistribute delays at the beginning of a disruption, but rather redistribute strategically when the recovery phase is predicted to begin soon (such predictions could follow from the ongoing work in Chapter 6 in terms of predictions on DRTs). These temporally-varying strategies have been hinted at by the fact that different DRT clusters exhibit different average redistribution costs (Section 7.7). In this case, it may no longer be true that delay absorption costs are inflicted equally among all airports, and by extension, carriers. One possible solution is to add a penalty term that scales with the total variation to the CSRP, encouraging solutions that exhibit a "smoothing" of delays across the network, rather than an accumulation at a particular, low-absorption cost airport. Such a term could be written as $K(\mathbf{c}, N)\sqrt{\mathbf{x}^{(t)\top}\boldsymbol{\mathcal{L}}\mathbf{x}^{(t)}}$ and added directly into the objective function. Specifically, $\sqrt{\mathbf{x}^{(t)\top}\boldsymbol{\mathcal{L}}\mathbf{x}^{(t)}}$ penalizes the total variation, with a square root to match

the units of delay minutes of the objective value, and $K(\mathbf{c}, N)$ is some scaling constant similar to $\mathbb{1}^\top \mathbf{c} N$ that ensures comparability between all objective function terms. In summary, such an equity-encouraging CSRP formulation could resemble (7.16):

$$\begin{aligned}
\underset{\mathbf{x}^{(t)} \in \widehat{\mathcal{X}}}{\arg\min} \quad & \left\{ \left\| \mathbf{x}^{(t)} - \mathbf{x}_*^{(t-1)} \right\|_2 + \mathbf{c}^\top \left(\mathbf{x}^{(t)} - \mathbf{x}_{\mathfrak{O}_M}^{(t)} \right) + K(\mathbf{c}, N) \sqrt{\mathbf{x}^{(t)\top} \mathcal{L} \mathbf{x}^{(t)}} \right\} \\
\text{s. t.} \quad & \left\| \mathbf{x}^{(t)} \right\|_1 \geq \left\| \mathbf{x}_{\mathfrak{O}_M}^{(t)} \right\|_1 - \delta, \\
& \left\| \mathbf{x}^{(t)} \right\|_1 \leq \left\| \mathbf{x}_{\mathfrak{O}_M}^{(t)} \right\|_1 + \delta, \\
& \left\| \mathbf{x}^{(t)} - \mathbf{x}_{\mathfrak{O}_M}^{(t)} \right\|_2 \leq \kappa(t), \\
& \mathbf{x}_*^{(0)} = \mathbf{x}_{\mathfrak{O}_M}^{(0)}, \\
& \forall t = 1, \ldots, T.
\end{aligned} \tag{7.16}$$

In particular, to reduce overloading the objective function, one could pull the historical adherence term out as a constraint $\left\| \mathbf{x}^{(t)} - \mathbf{x}_{\mathfrak{O}_M}^{(t)} \right\|_2 \leq \kappa(t)$, where small values of $\kappa(t) > 0$ indicate less flexibility for redistribution at time t, and setting $\kappa(t) = M$ for some large positive M indicate complete redistribution freedom at time t. Note that $\kappa(t)$ takes the place of the redistribution workload parameter λ in this case. An analogous version of a temporally-varying λ could be formulated as well.

7.8.3 Potential microscopic implementation

To complete the feedback cycle depicted in Figure 7-3, the desired delay dissipation and absorption will have to be incorporated as linear constraints on total airport delays within whatever TFMP implementation is used for setting and possibly adjusting traffic management initiatives. If we consider the deterministic or stochastic TFMP framework from [14], the decision regarding whether or not to run a particularly flight is made based on some aggregate "utility" of running that flight. Part of the utility calculation includes a penalty on the arrival delay, which is assessed at the destination airport for a particular flight. Using notation from [14], where $\Delta(f, t)$ denotes the total delay cost of flight f arriving at its destination at time t, let $\Delta_{\max}(a, t)$ denote the maximum allowable delay at airport a at time t according to

the solution of the CSRP $\mathbf{x}_*^{(t)} \in \widehat{\mathcal{X}}$ at time t. Specifically, since [14] specifies delays as an incurred cost directly related to the duration of the delay, there will be some scalar conversion between the value at airport a in $\mathbf{x}_*^{(t)}$ to match the units on $\Delta(f,t)$. Let $\mathcal{F}_a$ be the set of flights arriving at airport a at time t; then, the following airport-level delay constraint could be inserted into the Integer Master Program in [14], for each redistribution-controlled airport a in the network:

$$\sum_{f \in \mathcal{F}_a} \Delta(f,t) \leq \Delta_{\max}(a,t), \qquad \forall t. \tag{7.17}$$

There will have to be adjustments in terms of the time windows t, as the CSRP currently is implemented hourly. Note that the constraint in (7.17) will only need to be over airports a that the CSRP explicitly tries to dissipate delays at, i.e., the subset of pre-selected airports, embodied by non-zero terms in the cost vector $\mathbf{c}$ of the CSRP. The TFMP (e.g., [14]) implicitly assesses any additional delay penalties on flights that need to be incurred in order to satisfy (7.17). Given the *selective* nature of the CSRP, we would not need to modify the TFMP at every airport in the network. Another consideration might be to explicitly encourage delay absorption at particularly airports in the TFMP formulation, potentially guided by the disruption-specific results from Section 7.7.

As a part of ongoing work in [49], in lieu of the TFMP, we examine integrating the CSRP together with the multi-airport ground holding problem, or MAGHP. In [49], we propose a two-stage hierarchical control strategy for rescheduling aircraft (i.e., assigning delays) after network disruptions. In particular, we propose a *high-level planner* based off of a modified version of the CSRP (7.13). The high-level planner proposes a *reference plan* based on user preferences encoded in the modified CSRP. This reference plan accounts for complex objectives such as ensuring a "smooth" redistribution of delays across airports (quantified by the total variation). The low-level controller then solves the MAGHP, augmented to track the reference plan. The solution to the augmented MAGHP yields a revised flight schedule with lower total variation than the original MAGHP, while still satisfying operational constraints.

7.9 Delay redistribution: Summarizing remarks and lessons learned

The observed delays across a network of airports can be thought of as the end result of imbalances in scheduled demand and system capacity. The observed delays also reflect the effects of both strategic and tactical actions that influence both the demand side (schedule adjustments, cancellations, etc.) as well as the capacity side (switching airport runway configurations, traffic management initiatives, etc.). To this end, we investigate the potential to *redistribute* delays at the level of airports, in order to levy more strategic constraints on airport-side capacities (i.e., the feedback cycle depicted in Figure 7-3). Since we are working at the level of delay distributions, we use the approximate projection-based controls framework summarized in Sections 7.3 and 7.4 and detailed in [155].

A key component of this framework is the usage of copulas to generate more realistic airport network delay distributions, as well as the ability to produce interpretable control actions (i.e., how much delay to absorb or dissipate at each airport) and delay absorption cost measures. Specifically, we demonstrated the ability of copulas to provide an approximation for the feasible state space of a complex system that has intricate dependence structures, as well as different distributions of sub-system behavior. In particular, copulas provide a way to separate estimating marginal distributions and estimating dependence structures. We fit historical state observations to the family of multivariate Gaussian copulas, and constructed an approximate state space through copula sampling and the inverse probability integral transform. We then proposed a control framework for large-scale networks that sets performance targets in a lower-dimensional projection of the full state space, and provides future system states by selecting from candidate states in the copula-approximated state space. We demonstrated the applicability of our proposed methodology through a case study of flight delays in the US airport network, as well as its effectiveness in selectively redistributing airport delays away from a particular subset of airports.

In terms of future work related to Chapter 7, there are two distinct categories:

Improvements to the model, and applications-oriented extensions. In terms of the former, we currently use a static copula for each operational hour, whereas it may be beneficial to use time-varying copulas and copula processes [176]. The main benefit would be the ability to estimate some non-linear, differentiable system dynamics from the copula process. With an actual functional estimate of the system dynamics, it would then be possible to use setups such as discrete, finite horizon, constrained, iterated linear quadratic regulators and provide closed-form optimal controllers for the network redistribution problem [166, 60]. Another question of ongoing interest relates to the sample complexity of our copula-approximated state space.

On application-side extensions, the main question of interest would be to actually feed the redistribution constraints into a standard TFMP implementation, and characterize the impacts on the resultant TFMP solution. As we have previously mentioned, the redistribution scenarios shown Chapter 7 could be considered as an upper bound in terms of redistribution performance, as the various TFMP-based constraints would most likely prevent this upper bound from ever being reached. The addition of "sinks" to absorb delay (i.e., cancellations) may have to be incorporated into the CSRP in order to produce airport-level delay constraints that will not result in infeasibilities when coupled with a TFMP implementation. Finally, we hope to model other networked systems through this redistribution framework, particularly systems where redistribution of signals have more natural interpretations (e.g., bikeshare networks with signals being the number of bikes at a station).

Chapter 8

Concluding Remarks

8.1 Summary of thesis work

In the first half of the thesis with Chapter 4 and Chapter 5, the central theme is casting airport network delays as a signal processing problem. The airports are abstracted as nodes in a graph, and the relationship between airport delays are encoded into the edges of the graph. We want to emphasize that there are many ways to examine the relationship between one airport and another: One could weigh the edges simply by using the geographic distance between airports, or by the number of passengers that flow between an origin-destination airport pair, or the number of direct flights, etc. Since the phenomena of interest in this thesis are the on-time performance signals – flight delays and cancellations – generated by mismatches in demand and capacity at airports, we chose to use statistical measures such as airport delay correlations as the edge weights. With this choice, the edge weights reflect an aggregate measure of networked dependence between airports, driven by myriad factors such as localized weather, shared flights, and external traffic management actions.

With this setup in mind, we first motivate the methods developed in Chapter 4 by noting that current delay performance metrics capture only the magnitude of incurred flight delays at airports. The graph-supported signal setup allows us to augment this with another metric which takes into consideration a much more nuanced quantity that is affected by the underlying network: We show that it is also important to

consider the *spatial distribution* of delays across a network of airports. In Chapter 4, we formalize notions of, and develop tools to identify, graph signal outliers. Specifically, we show how the *total variation* metric for graph signals can help identify signal observations with an unexpected distribution across the nodes. Recall that total variation considers node-adjacent signals, and scales signal differences based on the edge weights. Recall also that we distinguish between different types of graph signal outliers, ranging from outliers in scale (OIS) to weak and strong outliers in distribution (OID); the former seeks to identify outliers based on the signal magnitude, whereas the latter uses total variation to identify outliers based on the spatial distribution. We then proceed to analyze and interpret spatial delay patterns across the US 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: For example, nor'easters tend to trigger very unexpected spatial delay distributions at a system-wide level, but thunderstorms and airport outages are far more spatially unexpected at the scales of airline sub-networks. We conduct some further comparison studies between US airline sub-networks and the US system-wide network, as well as between airport networks of different countries (e.g., US and China).

Chapter 5 is complementary to Chapter 4, in that we attempt to identify specific airports whose delays contribute significantly to certain unexpected spatial delay distributions. To achieve this, we use tools from graph signal processing and graph Fourier decomposition: Given an observation of delays across an airport network, we can decompose this network delay observation as linear combinations of commonly-occurring delay modes. Specifically, these delay modes are quite interpretable, and consist of relative delay differences between airports. With this approach, we are able to find critical subsets of airports that should be monitored when it comes to unexpected spatial delay distributions. In particular, we find that the delay distributions at a small set of US East Coast airports tend to be indicative of whether the system-wide delay distribution is expected or unexpected. In contrast, such an equivalent set of airports in China is spread out over a large portion of the airport

network in China. Another key contribution of Chapter 5 is a comprehensive outlier detection and interpretation framework: This framework elucidates the close connections between Chapter 4 and Chapter 6, and ties these two chapters together. We demonstrate this framework by analyzing specific outlier days in China and the US.

While Chapter 4 and Chapter 5 can be considered the methodological parts of this thesis, Chapter 6 presents the first application stemming from this graph signals perspective of airport delays. In Chapter 6, we leverage the notions developed in the first half of the thesis to comprehensively define the start, progression, and end of disruption and recovery cycles in a network of airports. This disruption-recovery cycle definition not only considers the magnitude of delays, but also their spatial distribution, relation to historical delay patterns, and temporal trends. We then identify a set of *disruption-recovery trajectories* using operational data, and develop appropriate features in order to cluster them into representative groups: These clusters represent commonly-occurring disruption-recovery pathways in the air transportation network. Through interpreting the features of these representative disruption-recovery pathways, we are able to extract metrics of interest such as the length of a typical disruption as well as spatial- and magnitude-related characteristics. Towards the end of Chapter 6, we refine the disruption-recovery trajectory framework, and prove structural properties about these objects. We also provide operational interpretations for these structural properties of interest.

Similar to Chapter 6, another application enabled by the framework developed in the first half of the thesis is an approximate control scheme for networked signals. We spend Chapter 7 focused on this, and develop a method to identify a sequence of future states and associated control actions for difficult-to-model, high-dimensional systems. The proposed control actions are based on performance targets in a lower-dimensional space, parameterized by metrics of interest such as total delay and total variation. We then discuss controlling NAS delays, and detail characteristics of the network control scheme via a case study of disruptions focused on the three major New York City airports. In doing so, we show that this control scheme is able to successfully provide potential delay state trajectories that follow pre-specified delay

redistribution and conservation constraints. Finally, we show how a new measure of delay absorption costs can be quantified using this approach. In our setup, the delay absorption costs are readily interpretable: An airport with a low absorption cost is more likely to see increases in delays post-redistribution of delays, and vice versa.

8.1.1 Data-driven network management and new entrants

We dedicate the remainder of this section to some remarks on how air traffic flow management and the aviation system as a whole is moving into a new, data-rich, and expanding environment. At its core, the air transportation system is a canonical example of a capacity-constrained, societal-scale network. The complexity of such a system hinders analysis, but also generates rich data primed for a range of models that provide actionable insights. However, as discussed partially in studies such as [163], there are key obstacles revolving around aviation research data availability and organization that need to be overcome. In particular, using some of the results of this thesis as examples: The inventories of outlier days, DRTs, as well as, e.g., representative airport delay, cancellation, and GDP network clusters from [101], should ideally be centrally cleaned and stored, then made available for analyses by any relevant stakeholders (airlines, ANSPs, etc.). This will also encourage the continued evaluation and incorporation of new performance metrics, especially in light of recent efforts by the FAA in centralizing flight data collection and distribution, i.e., the System Wide Information Management (SWIM) initiative [79]. The eventual goal for traffic and delay management in the NAS should be more flexible, data-driven, and well-targeted control actions, while considering the burden on the human factors side of the system. A desirable equilibrium would balance new, complex feedback-based and data-driven traffic flow management procedures with limiting increasing workloads on traffic flow managers and various hierarchies (en route, terminal, tower, etc.) of air traffic controllers.

Evolving from its core of commercial aviation, the air transportation system is not only expanding in scale, but also preparing to welcome new entrants such as unmanned aerial vehicles (UAVs) and to incorporate new paradigms such as urban

air mobility (UAM) and Advanced Air Mobility (AAM). This evolution has already begun, and there is a critical need to continue ensuring safety and efficiency. In the future, there may be multiple new modalities involved in bringing passengers from their home to the airport: These multiple mobility networks all generate their own set of output signals, giving insights into the health and performance of the system, if stored and analyzed properly. Some of the methods in this thesis could be extended to examining such multi-modal networks, and investigate their interaction dynamics during disruptions. Similar to the need to take into account human factors and controller workload, there is a more policy-oriented side of the story here as well: We need to ensure that new modes of transport enabled by UAM and AAM do not cannibalize funding and support towards current public transportation networks. Even though thoughtful engineering and resilience-building strategies may provide an efficient, seamless, and robust multi-modal transportation network, it may worsen existing issues in transportation access and equity.

8.2 Future directions of interest

We conclude Chapter 8 with some areas of interest for future work, stemming from various portions of Chapter 4 through Chapter 7. Other future work topics may be more tangentially related, but share common themes such as an underlying graphical model, or investigating behaviors of signals in networks.

Outlier prediction using graph convolution filters. Previously, we discussed the prediction problem from the perspective of the TV-TD state space, which provides a compressed representation of airport network delays, as both TD and TV are aggregate network quantities. Specifically, supposing that we know the delays at each airport at the current hour (i.e., we also can compute the current observed TD and TV values, and evaluate whether or not this hour is an OIS, weak OID, or strong OID), can we predict whether or not the system at some future point in time will attain a TD and TV value that would result in a classification as an outlier? Along

with the two prediction methods we briefly remarked on in the conclusion of Chapter 6 which is the subject of ongoing work in [156], it would also be interesting to explore adaptive classification filters, as these filters naturally encode attributes such as the graph eigenvector modes and characteristics of the graph frequency domain.

Controlling delay severity and spatial allocation. One of the most interesting applications of the framework from Chapter 4 and Chapter 5 is the potential for coming up with traffic management procedures that take into account both the severity as well as the spatial allocation of airport delays. A natural question that such a framework would answer would be the trade-offs between accumulating delays at one node versus allowing delays to propagate to other nodes. As an example, an ANSP may want to "quarantine" delays as best as it can, whereas airports and airlines might lean towards a more equitable distribution of delays across the system. We are addressing one facet of this question as it relates to modifying the multi-airport ground holding problem in [49].

Top-down control and management of disruptions in air transportation systems. The observed performance signals across a mobility network (e.g., delays, traffic velocities, canceled trips) are oftentimes the end product of some imbalance in system capacity and demand. For example, the low-level process in air traffic flow management (ATFM) involves solving the traffic flow management problem (TFMP) given airline schedules and system capacity forecasts [27]. Downstream in the ATFM process, the high-level system state is the resultant airport network delay distribution. An interesting and challenging network control question asks whether high-level delay distribution targets could translate to control actions at the traffic assignment (i.e., TFMP) level, potentially leading to a more optimal schedule that conforms to an interpretable delay target. In a first step towards this top-down network control perspective, we explore generating greedy-optimal delay redistribution targets via copula-based bootstrapping in Chapter 7. Another set of tools worth investigating is the use of balanced and unbalanced optimal transport [208] to construct

globally-optimal delay redistribution policies, as well as the integration of this high-level redistribution policy with the low-level TFMP. Such a top-down network control perspective may also be transferable to other mobility settings. In particular, it would be interesting to explore its applicability to bike-sharing systems and UAV networks.

Interplay between graphs and graph-supported signals. For graph-supported signals, one measure of interest is its total variation with respect to the underlying graph. Recall from Chapter 4 that this measure is a quadratic form in the signal, modulated by a matrix containing information regarding the underlying graph (e.g., the graph Laplacian). In the technical discussions in Chapter 4, we note nontrivial dependencies between moments of the total variation and the underlying graph. It would be interesting to explore the relationship between the generated graph signals, underlying graph structure, and distributions of random quadratic forms that are often functions of certain orthogonal polynomials [178]. This thread translates naturally to eigenvalue distributions (e.g., Laguerre ensembles for positive definite random matrices) and other results from random matrix theory (e.g., Marčenko-Pastur law and Wishart matrices), whose intersections have already been explored in the context of networks [210]. Airport delays, like many other signals generated from networked infrastructures, can be viewed as non-stationary stochastic processes, influenced by myriad underlying factors such as airline schedules and weather. Each agent (e.g., airport) has its own marginal signal (e.g., delay) distribution, and their interactions (e.g., back-and-forth flights) lead to nonlinear dependence structures in the joint network signal distribution. In Chapter 7, we consider the use of copulas and canonical maximum likelihood estimation to model such joint network distributions, but bypass the issue of non-stationarity by estimating different copulas for different time periods. An extension of interest would be to use copula processes [275] to work towards a continuous-time model for these non-stationary joint network distributions.

Appendix A

Proofs

A.1 Proof for Proposition 1

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

$$\begin{aligned}
\mathrm{TV}\left(\mathbf{x}^{(k)}\right) &= \left(\mathbf{x}^{(k)}\right)^\top \mathfrak{L}\mathbf{x}^{(k)} \\
&= \left(\mathbf{x}^{(k)}\right)^\top \mathfrak{L}\left(\sum_{i=1}^{N} \alpha_i^{(k)} v_i\right) \\
&= \left(\mathbf{x}^{(k)}\right)^\top \left(\sum_{i=1}^{N} \alpha_i^{(k)} \mathfrak{L} v_i\right) \\
&= \left(\mathbf{x}^{(k)}\right)^\top \left(\sum_{i=1}^{N} \alpha_i^{(k)} \lambda_i v_i\right) \\
&= \left(\sum_{j=1}^{N} \alpha_j^{(k)} v_j\right)^\top \left(\sum_{i=1}^{N} \alpha_i^{(k)} \lambda_i v_i\right) \\
&= \sum_{i,j} \alpha_i^{(k)} \alpha_j^{(k)} \lambda_i v_j^\top v_i \\
&= \sum_{i=1}^{N} \left(\alpha_i^{(k)}\right)^2 \lambda_i.
\end{aligned} \tag{A.1}$$

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

A.2 Proof for Proposition 3

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

A.3 Proof for Proposition 4

Proof. Since $r_{ij|\mathcal{O}_M}$ is a random variable dependent only on M previous observations, and X_i, X_j are currently unobserved random variables, the expectation operator factorizes over the expression for the TV (from (4.31)):

$$\begin{aligned}\mathbb{E}[\text{TV}(\mathbf{X})] &= \frac{1}{2}\sum_{i\neq j}\left\{\mathbb{E}\left[r^+_{ij|\mathcal{O}_M}\right]\,\mathbb{E}\left[(X_i - X_j)^2\right]\right\} \\ &= \frac{1}{2}\sum_{i\neq j}\left\{\mathbb{E}\left[r^+_{ij|\mathcal{O}_M}\right]\left((\mu_i - \mu_j)^2 + \sigma_i^2 + \sigma_j^2 - 2\rho^+_{ij}\sigma_i\sigma_j\right)\right\}.\end{aligned} \tag{A.2}$$

Since the bias of $r^+_{ij|\mathcal{O}_M}$ is given in (4.30), for any M, there exists a $\gamma_{ij} > 0$ that is a function of M (and $\lim_{M\to\infty}\gamma_{ij} = 0$) such that:

$$\left|\mathbb{E}\left[r^+_{ij|\mathcal{O}_M}\right] - \rho^+_{ij}\right| < \gamma_{ij} \iff \rho^+_{ij} - \gamma_{ij} < \mathbb{E}[r^+_{ij|\mathcal{O}_M}] < \rho^+_{ij} + \gamma_{ij}. \tag{A.3}$$

We can use the fact that $\nu_{ij} < \rho^+_{ij} < \varepsilon_{ij}$ in order to rewrite the bounds of (A.3):

$$\max\{0, \nu_{ij} - \gamma_{ij}\} \leq \mathbb{E}\left[r^+_{ij|\mathcal{O}_M}\right] < \varepsilon_{ij} + \gamma_{ij}. \tag{A.4}$$

The maximum operator is included since $\nu_{ij} - \gamma_{ij}$ can be negative, but we know that the expectation of a non-negative random variable $\text{TV}(\mathbf{X})$ is bounded below by 0. We aim to use (A.4) in conjunction with the bounds on ρ^+_{ij} to bound (A.2). We focus first on deriving the upper bound by evaluating (A.2) for the largest-possible contributions from the positive terms, and the smallest-possible deductions from the

negative term. This yields:

$$\mathbb{E}[\mathrm{TV}(\mathbf{X})] = \frac{1}{2}\sum_{i\neq j}\Big\{\underbrace{\mathbb{E}\left[r^{+}_{ij|\mathcal{O}_M}\right]}_{<\varepsilon_{ij}+\gamma_{ij}}\left((\mu_i-\mu_j)^2+\sigma_i^2+\sigma_j^2\right) - 2\underbrace{\mathbb{E}\left[r^{+}_{ij|\mathcal{O}_M}\right]\rho^{+}_{ij}}_{<\nu_{ij}\max\{0,\nu_{ij}-\gamma_{ij}\}}\sigma_i\sigma_j\Big\}. \tag{A.5}$$

Therefore, this gives the upper bound $\mathbb{E}[\mathrm{TV}(\mathbf{X})] < \delta_2$, where

$$\delta_2 = \frac{1}{2}\sum_{i\neq j}\left\{\tilde{\varepsilon}_{ij}\left((\mu_i-\mu_j)^2+\sigma_i^2+\sigma_j^2\right) - 2\tilde{\nu}^{+}_{ij}\nu_{ij}\sigma_i\sigma_j\right\} \tag{A.6}$$

along with rewriting $\tilde{\varepsilon}_{ij} = \varepsilon_{ij}+\gamma_{ij}$ and $\tilde{\nu}^{+}_{ij} = \max\{0, \nu_{ij}-\gamma_{ij}\}$.

To get the lower bound, we evaluate (A.2) for the smallest-possible contribution from the positive terms and the largest-possible contribution in terms of magnitude from the negative terms. This gives $\mathbb{E}[\mathrm{TV}(\mathbf{X})] > \delta_1$, where

$$\delta_1 = \frac{1}{2}\sum_{i\neq j}\left\{\tilde{\nu}^{+}_{ij}\left((\mu_i-\mu_j)^2+\sigma_i^2+\sigma_j^2\right) - 2\,\tilde{\varepsilon}_{ij}\varepsilon_{ij}\sigma_i\sigma_j\right\} \tag{A.7}$$

Note that δ_1 and δ_2 are functions of the bounds on ρ^{+}_{ij} and M, so we have:

$$\lim_{\substack{\nu_{ij}\to\rho^{+}_{ij}\\ \varepsilon_{ij}\to\rho^{+}_{ij}}}(\delta_1) = \lim_{\substack{\nu_{ij}\to\rho^{+}_{ij}\\ \varepsilon_{ij}\to\rho^{+}_{ij}}}(\delta_2) = \mathbb{E}[\mathrm{TV}(\mathbf{X})]. \tag{A.8}$$

□

A.4 Proof for Proposition 5

Proof. The idea behind the proof is similar to that of Proposition 4. We expand $\mathrm{Var}[\mathrm{TV}(\mathbf{X})]$ as done in (4.10) and (4.11). Proposition 2 can then be used to obtain the appropriate higher-order moments, including $\mathrm{Var}[\mathrm{TV}(\mathbf{X})]$ as a scalar quantity that depends on $\mathbb{E}\left[r^{+}_{ij|\mathcal{O}_M}\right]$ and ρ^{+}_{ij}. Finally, these two terms can be used to obtain the desired bounds on $\mathrm{Var}[\mathrm{TV}(\mathbf{X})]$. □

A.5 Proof for Theorem 3

Proof. Select an arbitrary regime $\mathcal{S}$ trajectory $\tau \in \mathcal{S}$; we would like to show that there exists an extension $\tau' =_{\triangle} \text{Ext}(\tau)$ such that $\tau' \in \mathcal{I}$. Since $\tau \in \mathcal{S}$, we know that τ is limited to the following forms:

(i) $\{\textcircled{1}\} \to \mathcal{P}\left(\textcircled{2}, \textcircled{4}\right) \to \{\textcircled{1}\}$,

(ii) $\{\textcircled{1}\} \to \mathcal{P}\left(\textcircled{2}, \textcircled{4}\right) \to \{\textcircled{3}\}$,

(iii) $\{\textcircled{3}\} \to \mathcal{P}\left(\textcircled{2}, \textcircled{4}\right) \to \{\textcircled{1}\}$,

(iv) $\{\textcircled{3}\} \to \mathcal{P}\left(\textcircled{2}, \textcircled{4}\right) \to \{\textcircled{3}\}$.

We first examine case (i) where τ is of the form $\{\textcircled{1}\} \to \mathcal{P}\left(\textcircled{2}, \textcircled{4}\right) \to \{\textcircled{1}\}$. Since the left and right terminal states are already $\textcircled{1}$ and the anchor states are either $\textcircled{2}$ or $\textcircled{4}$, we define $\tau' =_{\triangle} \tau$, and observe that $\tau' \in \mathcal{I}$.

We now examine case (ii) where τ is of the form $\{\textcircled{1}\} \to \mathcal{P}\left(\textcircled{2}, \textcircled{4}\right) \to \{\textcircled{3}\}$. Let $\tau' \supset_{\triangle} \tau$, where

$$\tau' : \quad \underbrace{\{\textcircled{1}\} \to \mathcal{P}\left(\textcircled{2}, \textcircled{4}\right) \to \{\textcircled{3}\}}_{\tau} \to \overbrace{\mathcal{P}\left(\textcircled{2}, \textcircled{3}, \textcircled{4}\right) \to \{\textcircled{1}\}}^{\text{Ext}(\tau) \setminus_{\triangle} \tau}. \tag{A.9}$$

Note that (A.9) can be rewritten as

$$\tau' : \quad \{\textcircled{1}\} \to \mathcal{P}\left(\textcircled{2}, \textcircled{3}, \textcircled{4}\right) \to \{\textcircled{1}\},$$

which clearly shows that $\tau' \in \mathcal{I}$.

We now examine case (iii) where τ is of the form $\{\textcircled{3}\} \to \mathcal{P}\left(\textcircled{2}, \textcircled{4}\right) \to \{\textcircled{1}\}$. Let $\tau' \supset_{\triangle} \tau$, where

$$\tau' : \quad \overbrace{\{\textcircled{1}\} \to \mathcal{P}\left(\textcircled{2}, \textcircled{3}, \textcircled{4}\right)}^{\text{Ext}(\tau) \setminus_{\triangle} \tau} \to \underbrace{\{\textcircled{3}\} \to \mathcal{P}\left(\textcircled{2}, \textcircled{4}\right) \to \{\textcircled{1}\}}_{\tau}. \tag{A.10}$$

Note that (A.10) can also be rewritten as

$$\tau' : \quad \{\textcircled{1}\} \to \mathcal{P}\left(\textcircled{2}, \textcircled{3}, \textcircled{4}\right) \to \{\textcircled{1}\},$$

which clearly shows that $\tau' \in \mathcal{I}$.

Finally, we examine case (iv) where τ is of the form $\{\textcircled{3}\} \to \mathcal{P}\left(\textcircled{2}, \textcircled{4}\right) \to \{\textcircled{3}\}$. We first construct extension $\tau'_\ell \supset_\triangle \tau$, where

$$\tau'_\ell : \quad \underbrace{\{\textcircled{3}\} \to \mathcal{P}\left(\textcircled{2}, \textcircled{4}\right) \to \{\textcircled{3}\}}_{\tau} \to \overbrace{\mathcal{P}\left(\textcircled{2}, \textcircled{3}, \textcircled{4}\right) \to \{\textcircled{1}\}}^{\mathrm{Ext}(\tau) \setminus_\triangle \tau}.$$

Now, we construct $\tau' \supset_\triangle \tau'_\ell$, where

$$\begin{aligned} \tau' : \quad & \overbrace{\{\textcircled{1}\} \to \mathcal{P}\left(\textcircled{2}, \textcircled{3}, \textcircled{4}\right)}^{\mathrm{Ext}\left(\tau'_\ell\right) \setminus_\triangle \tau'_\ell} \to \\ & \underbrace{\{\textcircled{3}\} \to \mathcal{P}\left(\textcircled{2}, \textcircled{4}\right) \to \{\textcircled{3}\} \to \mathcal{P}\left(\textcircled{2}, \textcircled{3}, \textcircled{4}\right) \to \{\textcircled{1}\}}_{\tau'_\ell}. \end{aligned} \tag{A.11}$$

Note that (A.11) can also be rewritten as

$$\tau' : \quad \{\textcircled{1}\} \to \mathcal{P}\left(\textcircled{2}, \textcircled{3}, \textcircled{4}\right) \to \{\textcircled{1}\},$$

which clearly shows that $\tau' \in \mathcal{I}$. □

A.6 Proof for Theorem 4

Proof. Let τ be a trajectory in regime $\mathcal{I}(\mathcal{H})$, and count the number of occurrences of $\textcircled{3}$ in τ, and record this count as k. Since $\textcircled{1}$ and $\textcircled{3}$ are both terminal states for any trajectory belonging to regime $\mathcal{S}$, the number of terminal states in τ with respect to regime $\mathcal{S}$ is $k + 2$, i.e., k plus the left and right terminal states in τ, which by definition of $\tau \in \mathcal{I}(\mathcal{H})$ must both be $\textcircled{1}$. If we were to count the number of (possibly empty) sub-lists of τ, where the sub-lists has left and right terminal states as either

$\textcircled{1}$ or $\textcircled{3}$, there must be $k+2-1=k+1$ such (possibly empty) sub-lists. Observe that empty sub-lists all must be of the following forms:

$$\begin{aligned} \{\textcircled{1}\} &\rightarrow \{\textcircled{3}\}, \\ \{\textcircled{3}\} &\rightarrow \{\textcircled{1}\}, \\ \text{or } \{\textcircled{3}\} &\rightarrow \{\textcircled{3}\}. \end{aligned}$$

Thus, count the number of occurrences of such empty sub-lists, and record this count as m. Now observe that non-empty sub-lists all must be of the following forms:

$$\begin{aligned} \{\textcircled{1}\} &\rightarrow \mathcal{P}\left(\textcircled{2},\textcircled{4}\right) \rightarrow \{\textcircled{3}\}, \\ \{\textcircled{3}\} &\rightarrow \mathcal{P}\left(\textcircled{2},\textcircled{4}\right) \rightarrow \{\textcircled{1}\}, \\ \text{or } \{\textcircled{3}\} &\rightarrow \mathcal{P}\left(\textcircled{2},\textcircled{4}\right) \rightarrow \{\textcircled{3}\}. \end{aligned} \tag{A.12}$$

Note that all forms in (A.12) can be equivalently rewritten as

$$\{\textcircled{1},\textcircled{3}\} \rightarrow \mathcal{P}\left(\textcircled{2},\textcircled{4}\right) \rightarrow \{\textcircled{1},\textcircled{3}\},$$

which is clearly a generic trajectory in regime $\mathcal{S}$ by way of Definition 5. Thus, the number of disjoint trajectories in $\mathcal{S}(\mathcal{H})$ that $\tau \in \mathcal{I}(\mathcal{H})$ decomposes to, or equivalently, the number of non-empty sub-lists, is given by $k+1-m$. □

A.7 Proof for Theorem 5

Proof. We first introduce and prove a lemma that will be used in the proof of Theorem 5:

Lemma 1 (Non-negativity of counts) *Let $\tau \in \mathcal{I}(\mathcal{H})$, and examine the supersets of decomposed trajectories $\mathcal{S}(\tau)$ and $\mathcal{D}(\tau)$. In accordance with Theorem 4 and Corollary 2, record the cardinalities $|\mathcal{S}(\tau)| = k_{\mathcal{S}}+1-m_{\mathcal{S}}$ and $|\mathcal{D}(\tau)| = k_{\mathcal{D}}+1-m_{\mathcal{D}}$. We have that*

$$\begin{aligned} k_{\mathcal{S}}+1-m_{\mathcal{S}} &\geq 0, \\ k_{\mathcal{D}}+1-m_{\mathcal{D}} &\geq 0. \end{aligned}$$

Proof of Lemma 1. We first examine the $\mathcal{S}$ case, and the $\mathcal{D}$ case will follow analogously. Note first that since $k_{\mathcal{S}}$ and $m_{\mathcal{S}}$ are counts, they are necessarily non-negative. Now, we have that $k_{\mathcal{S}}$ increases by 1 for every ③ state in τ, and with every unit increase in $k_{\mathcal{S}}$, $m_{\mathcal{S}}$ (recall from Theorem 4 that $m_{\mathcal{S}}$ counts the number of occurrences of pairs $\{①\} \to \{③\}$, $\{③\} \to \{①\}$, and $\{③\} \to \{③\}$) can increase by 0 (see Figure A-1, first row), or 1 (see Figure A-1, first row), or 2. However, observe that $m_{\mathcal{S}}$ can increase by 2 for an unit increase in $k_{\mathcal{S}}$ at most once. In fact, observe that only in the following situation (illustrated in Figure A-2), where τ is

$$\tau: \quad \{①\} \to \mathcal{P}\left(③\right) \to \{①\},$$

is the worst-case upper bound $k_{\mathcal{S}} + 1 \geq m_{\mathcal{S}}$ required. For the most general $\tau \in \mathcal{I}(\mathcal{H})$, where

$$\tau: \quad \{①\} \to \mathcal{P}\left(②, ③, ④\right) \to \{①\},$$

the bound $k_{\mathcal{S}} \geq m_{\mathcal{S}}$ holds. The same argument can be used to show that $k_{\mathcal{D}} + 1 - m_{\mathcal{D}} \geq 0$, by trivially relabeling ② $\mapsto$ ③ and ③ $\mapsto$ ②. □

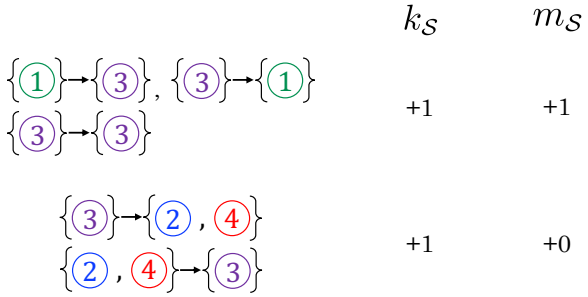


Figure A-1: State pairs in τ and the relationship between $k_{\mathcal{S}}$ and $m_{\mathcal{S}}$.

We now resume the proof for Theorem 5. Let $\tau \in \mathcal{I}(\mathcal{H})$, we will first derive the follow inequality:

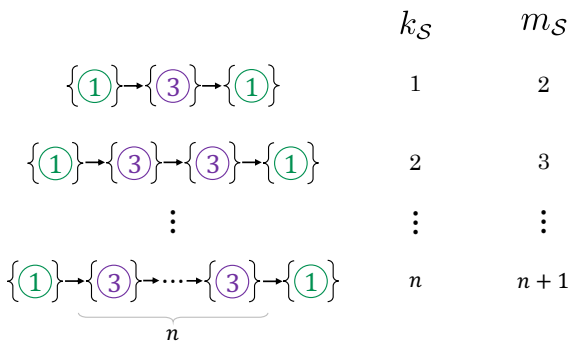


Figure A-2: Trajectory τ under the worst-case bound $k_{\mathcal{S}} + 1 \geq m_{\mathcal{S}}$.

$$\begin{aligned} &|\mathcal{S}(\tau)| + |\mathcal{D}(\tau)| \geq 1 \\ \Leftrightarrow\ &(k_{\mathcal{S}} + 1 - m_{\mathcal{S}}) + (k_{\mathcal{D}} + 1 - m_{\mathcal{D}}) \geq 1. \end{aligned} \tag{A.13}$$

In order to show that the inequality in (A.13) is true, we note that there are four cases with respect to $k_{\mathcal{S}}$ and $k_{\mathcal{D}}$, keeping in mind that $k_{\mathcal{S}} \in \mathbb{N}_{\geq 0}$ and $k_{\mathcal{D}} \in \mathbb{N}_{\geq 0}$:

(i) $k_{\mathcal{S}} = k_{\mathcal{D}} = 0$,

(ii) $k_{\mathcal{S}} \geq 1$ and $k_{\mathcal{D}} \geq 1$,

(iii) $k_{\mathcal{S}} \geq 1$ and $k_{\mathcal{D}} = 0$,

(iv) $k_{\mathcal{S}} = 0$ and $k_{\mathcal{D}} \geq 1$.

Since τ is a valid regime $\mathcal{I}$ trajectory, in the first case where $k_{\mathcal{S}} = k_{\mathcal{D}} = 0$, this is equivalent to constraining τ to be of the form

$$\tau: \quad \{(1)\} \to \mathcal{P}\left((4)\right) \to \{(1)\}, \tag{A.14}$$

since no anchor states can be (2) or (3). Observe that when τ is of the form given in (A.14), we must have that $m_{\mathcal{S}} = m_{\mathcal{D}} = 0$ since pairs $\{(1)\} \to \{(3)\}$, $\{(3)\} \to \{(1)\}$, and $\{(3)\} \to \{(3)\}$ or $\{(1)\} \to \{(2)\}$, $\{(2)\} \to \{(1)\}$, and $\{(2)\} \to \{(2)\}$ cannot occur. Thus, the left-hand side of (A.13) simplifies down to:

$$(k_{\mathcal{S}} + 1 - m_{\mathcal{S}}) + (k_{\mathcal{D}} + 1 - m_{\mathcal{D}}) = 2 \geq 1.$$

For case (ii), observe that since $k_{\mathcal{S}} \geq 1$ and $k_{\mathcal{D}} \geq 1$, τ cannot be of the form

$$\tau : \quad \{\textcircled{1}\} \to \mathcal{P}\left(\textcircled{2}\right) \to \{\textcircled{1}\}$$

or

$$\tau : \quad \{\textcircled{1}\} \to \mathcal{P}\left(\textcircled{3}\right) \to \{\textcircled{1}\}$$

since there is at least one occurrence of a $\textcircled{2}$ anchor state and a $\textcircled{3}$ anchor state. Thus, from Lemma 1, we have that $k_{\mathcal{S}} - m_{\mathcal{S}} \geq 0$ and $k_{\mathcal{D}} - m_{\mathcal{D}} \geq 0$. Rearranging the left-hand side of (A.13) and applying the inequalities $k_{\mathcal{S}} - m_{\mathcal{S}} \geq 0$ and $k_{\mathcal{D}} - m_{\mathcal{D}} \geq 0$, we see that

$$\underbrace{(k_{\mathcal{S}} - m_{\mathcal{S}})}_{\geq 0} + \underbrace{(k_{\mathcal{D}} - m_{\mathcal{D}})}_{\geq 0} + 2 \geq 1.$$

For the last two cases (iii) and (iv), note that they are symmetric cases. Without loss of generality, we set and examine the case where $k_{\mathcal{S}} \geq 1$ and $k_{\mathcal{D}} = 0$. Since $k_{\mathcal{D}} = 0$, the pairs $\{\textcircled{1}\} \to \{\textcircled{2}\}$, $\{\textcircled{2}\} \to \{\textcircled{1}\}$, and $\{\textcircled{2}\} \to \{\textcircled{2}\}$ cannot occur, so we have that $m_{\mathcal{D}} = 0$. From Lemma 1, we know that $k_{\mathcal{S}} + 1 - m_{\mathcal{S}} \geq 0$. Thus, we see that in this case, the left-hand side of (A.13) becomes:

$$\underbrace{(k_{\mathcal{S}} + 1 - m_{\mathcal{S}})}_{\geq 0} + 1 \geq 1.$$

Analogously, if we had set $k_{\mathcal{S}} = 0$ and $k_{\mathcal{D}} \geq 1$, we would have had

$$1 + \underbrace{(k_{\mathcal{D}} + 1 - m_{\mathcal{D}})}_{\geq 0} \geq 1.$$

Hence, we have that $\forall \tau \in \mathcal{I}(\mathcal{H})$, the decomposition cardinalities satisfy $|\mathcal{S}(\tau)| + |\mathcal{D}(\tau)| \geq 1$. Taking the sum over all such trajectories $\tau \in \mathcal{I}(\mathcal{H})$, we have that

$$
\begin{aligned}
|\mathcal{S}(\tau)| + |\mathcal{D}(\tau)| &\geq 1 \\
\sum_{\tau \in \mathcal{I}(\mathcal{H})} \Big\{ |\mathcal{S}(\tau)| + |\mathcal{D}(\tau)| \Big\} &\geq \sum_{\tau \in \mathcal{I}(\mathcal{H})} 1 \\
\sum_{\tau \in \mathcal{I}(\mathcal{H})} |\mathcal{S}(\tau)| + \sum_{\tau \in \mathcal{I}(\mathcal{H})} |\mathcal{D}(\tau)| &\geq |\mathcal{I}(\mathcal{H})| \,.
\end{aligned}
\tag{A.15}
$$

Note that the left-hand side of the last expression in (A.15) records a sum over decompositions of all trajectories $\tau \in \mathcal{I}(\mathcal{H})$. In order to rewrite the inequality in terms of the cardinality of $\mathcal{S}(\mathcal{H})$ and $\mathcal{D}(\mathcal{H})$, note that for any $\tau_{\mathcal{S}} \in \mathcal{S}(\tau)$, since $\tau \in \mathcal{I}(\mathcal{H})$, this implies that $\tau_{\mathcal{S}} \in \mathcal{S}(\mathcal{H})$. Thus, we have the containment $\mathcal{S}(\tau) \subseteq \mathcal{S}(\mathcal{H}), \forall \tau \in \mathcal{I}(\mathcal{H})$. Taking a union over all $\tau \in \mathcal{I}(\mathcal{H})$ preserves the containment, i.e.,

$$
\bigcup_{\tau \in \mathcal{I}(\mathcal{H})} \mathcal{S}(\tau) \subseteq \mathcal{S}(\mathcal{H}).
$$

Since $\mathcal{S}(\tau)$ and $\mathcal{S}(\mathcal{H})$ are supersets of trajectories, taking the cardinality of both sides gives

$$
\begin{aligned}
\left| \bigcup_{\tau \in \mathcal{I}(\mathcal{H})} \mathcal{S}(\tau) \right| &\leq |\mathcal{S}(\mathcal{H})| \\
\Leftrightarrow \sum_{\tau \in \mathcal{I}(\mathcal{H})} |\mathcal{S}(\tau)| &\leq |\mathcal{S}(\mathcal{H})| \,.
\end{aligned}
\tag{A.16}
$$

Note that the same argument applied to regime $\mathcal{D}$ decompositions give the equivalent of (A.16), but for regime $\mathcal{D}$, i.e., $\sum_{\tau \in \mathcal{I}(\mathcal{H})} |\mathcal{D}(\tau)| \leq |\mathcal{D}(\mathcal{H})|$. Thus, we have the desired upper bound (6.2) via

$$
|\mathcal{I}(\mathcal{H})| \leq \sum_{\tau \in \mathcal{I}(\mathcal{H})} |\mathcal{S}(\tau)| + \sum_{\tau \in \mathcal{I}(\mathcal{H})} |\mathcal{D}(\tau)| \leq |\mathcal{S}(\mathcal{H})| + |\mathcal{D}(\mathcal{H})| \,.
$$

□

A.8 Proof for Theorem 6

Proof. Let $\tau \in \mathcal{I}(\mathcal{H})$, similar to the process of the proof for Theorem 5, we first show an equivalent inequality to (6.3) for a single trajectory τ, then use an aggregation argument across all $\tau \in \mathcal{I}(\mathcal{H})$. We first show that the following statement is valid:

$$|\tau|_{\triangle} \leq \sum_{\tau_{\mathcal{S}} \in \mathcal{S}(\tau)} |\tau_{\mathcal{S}}|_{\triangle} + \sum_{\tau_{\mathcal{D}} \in \mathcal{D}(\tau)} |\tau_{\mathcal{D}}|_{\triangle}\,, \tag{A.17}$$

for any $\tau \in \mathcal{I}(\mathcal{H})$ and its decompositions $\mathcal{S}(\tau)$ and $\mathcal{D}(\tau)$. Recall that in the proof for Theorem 5, we showed that $|\mathcal{S}(\tau)| + |\mathcal{D}(\tau)| \geq 1$ for any $\tau \in \mathcal{I}(\mathcal{H})$, implying that $|\mathcal{S}(\tau)| \in \mathbb{N}_{\geq 0}$ and $|\mathcal{D}(\tau)| \in \mathbb{N}_{\geq 0}$ cannot both be 0, i.e., there must be at least one decomposition trajectory. Furthermore, since $|\cdot|_{\triangle} : \tau \to \mathbb{N}_{\geq 0}$ gives the length (or the number of states) belonging to a trajectory, we must have that

$$\sum_{\tau_{\mathcal{S}} \in \mathcal{S}(\tau)} |\tau_{\mathcal{S}}|_{\triangle} \geq 0, \qquad \sum_{\tau_{\mathcal{D}} \in \mathcal{D}(\tau)} |\tau_{\mathcal{D}}|_{\triangle} \geq 0.$$

We first examine the proposed upper bound in (A.17), and observe that it consists of two summation terms summing over trajectories in $\mathcal{S}(\tau)$ and $\mathcal{D}(\tau)$, where at least one of $\mathcal{S}(\tau)$ or $\mathcal{D}(\tau)$ must be non-empty. Consider first the case where $|\mathcal{S}(\tau)| = 0$, forcing $|\mathcal{D}(\tau)| \geq 1$. Note that $|\mathcal{S}(\tau)| = 0$ implies

$$\begin{aligned} &\left(\bigcup_{\tau_{\mathcal{S}} \in \mathcal{S}(\tau)} \right)_{\triangle} \tau_{\mathcal{S}} =_{\triangle} \{\varnothing\} \\ \Leftrightarrow &\sum_{\tau_{\mathcal{S}} \in \mathcal{S}(\tau)} |\tau_{\mathcal{S}}|_{\triangle} = 0, \end{aligned}$$

where $\{\varnothing\}$ denotes an empty list. On the other hand, we have that

$$\begin{aligned} &\left(\bigcup_{\tau_{\mathcal{D}} \in \mathcal{D}(\tau)} \right)_{\triangle} \tau_{\mathcal{D}} =_{\triangle} \tau \\ \Leftrightarrow &\sum_{\tau_{\mathcal{D}} \in \mathcal{D}(\tau)} |\tau_{\mathcal{D}}|_{\triangle} = |\tau|_{\triangle}\,, \end{aligned}$$

and so in this case, (A.17) holds in equality, since

$$\max\left\{\sum_{\tau_{\mathcal{S}}\in\mathcal{S}(\tau)}|\tau_{\mathcal{S}}|_{\triangle}\,,\sum_{\tau_{\mathcal{D}}\in\mathcal{D}(\tau)}|\tau_{\mathcal{D}}|_{\triangle}\right\}=\max\left\{0,|\tau|_{\triangle}\right\}=|\tau|_{\triangle}$$

and $\sum_{\tau_{\mathcal{S}}\in\mathcal{S}(\tau)}|\tau_{\mathcal{S}}|_{\triangle}+\sum_{\tau_{\mathcal{D}}\in\mathcal{D}(\tau)}|\tau_{\mathcal{D}}|_{\triangle}=0+|\tau|_{\triangle}=|\tau|_{\triangle}$. By symmetry, (A.17) holds in equality as well if we had set $|\mathcal{D}(\tau)|=0$, hereby forcing $|\mathcal{S}(\tau)|\geq 1$.

Consider the final possibility that both $|\mathcal{S}(\tau)|\geq 1$ and $|\mathcal{D}(\tau)|\geq 1$. In this case, we have that

$$\begin{aligned}&\left(\bigcup_{\tau_{\mathcal{S}}\in\mathcal{S}(\tau)}\right)_{\triangle}\tau_{\mathcal{S}}\cup_{\triangle}\left(\bigcup_{\tau_{\mathcal{D}}\in\mathcal{D}(\tau)}\right)_{\triangle}\tau_{\mathcal{D}}=_{\triangle}\tau\\ \Leftrightarrow&\sum_{\tau_{\mathcal{S}}\in\mathcal{S}(\tau)}|\tau_{\mathcal{S}}|_{\triangle}+\sum_{\tau_{\mathcal{D}}\in\mathcal{D}(\tau)}|\tau_{\mathcal{D}}|_{\triangle}-C=|\tau|_{\triangle}\,.\end{aligned}$$

Where the integer constant $C\geq 1$ counts overlapping terminal states shared by trajectories in $\mathcal{S}(\tau)$ and $\mathcal{D}(\tau)$. The upper bound now becomes loose, since if both $\mathcal{S}(\tau)$ and $\mathcal{D}(\tau)$ are non-empty, there must be at least one overlapping terminal state, counted by C. Combining all three cases together (i.e., $|\mathcal{S}(\tau)|=0$ and $|\mathcal{D}(\tau)|\geq 1$; $|\mathcal{S}(\tau)|\geq 1$ and $|\mathcal{D}(\tau)|=0$; $|\mathcal{S}(\tau)|\geq 1$ and $|\mathcal{D}(\tau)|\geq 1$), we arrive at the relationship in (A.17).

We can now begin the aggregation argument to show that (A.17) implies (6.3). We have that

$$\begin{aligned}|\tau|_{\triangle}&\leq\sum_{\tau_{\mathcal{S}}\in\mathcal{S}(\tau)}|\tau_{\mathcal{S}}|_{\triangle}+\sum_{\tau_{\mathcal{D}}\in\mathcal{D}(\tau)}|\tau_{\mathcal{D}}|_{\triangle}\\ \sum_{\tau\in\mathcal{I}(\mathcal{H})}|\tau|_{\triangle}&\leq\sum_{\tau\in\mathcal{I}(\mathcal{H})}\sum_{\tau_{\mathcal{S}}\in\mathcal{S}(\tau)}|\tau_{\mathcal{S}}|_{\triangle}+\sum_{\tau\in\mathcal{I}(\mathcal{H})}\sum_{\tau_{\mathcal{D}}\in\mathcal{D}(\tau)}|\tau_{\mathcal{D}}|_{\triangle}\,.\end{aligned}\tag{A.18}$$

Recall from the proof of Theorem 5 that

$$\bigcup_{\tau\in\mathcal{I}(\mathcal{H})}\mathcal{S}(\tau)\subseteq\mathcal{S}(\mathcal{H}),\qquad\bigcup_{\tau\in\mathcal{I}(\mathcal{H})}\mathcal{D}(\tau)\subseteq\mathcal{D}(\mathcal{H}).$$

Hence, we have that

$$\begin{aligned}\sum_{\tau\in\mathcal{I}(\mathcal{H})}|\tau|_{\triangle} &\leq \sum_{\tau\in\mathcal{I}(\mathcal{H})}\sum_{\tau_{\mathcal{S}}\in\mathcal{S}(\tau)}|\tau_{\mathcal{S}}|_{\triangle}+\sum_{\tau\in\mathcal{I}(\mathcal{H})}\sum_{\tau_{\mathcal{D}}\in\mathcal{D}(\tau)}|\tau_{\mathcal{D}}|_{\triangle}\\ &\leq \sum_{\tau\in\mathcal{S}(\mathcal{H})}|\tau|_{\triangle}+\sum_{\tau\in\mathcal{D}(\mathcal{H})}|\tau|_{\triangle}-\sum_{\tau\in\mathcal{S}(\mathcal{H})\cap\mathcal{D}(\mathcal{H})}|\tau|_{\triangle} \qquad \text{(A.19)}\\ &= \sum_{\tau\in\mathcal{S}(\mathcal{H})\cup\mathcal{D}(\mathcal{H})}|\tau|_{\triangle}\,.\end{aligned}$$

Note that we must adjust for double-counting, since it is possible that $\mathcal{S}(\mathcal{H})\cap\mathcal{D}(\mathcal{H})$ is non-empty – for example, the trajectory $\{\textcircled{1}\}\to\mathcal{P}\left(\textcircled{4}\right)\to\{\textcircled{1}\}$ is both a regime $\mathcal{S}$ and a regime $\mathcal{D}$ trajectory. □

A.9 Proof for Proposition 8

Proof. We first show that the cost function $\mathcal{J}\left(\mathbf{x}^{(t)}\right)$ is convex, where

$$\begin{aligned}&\mathcal{J}:\mathbb{R}^N\to\mathbb{R}\\ &\quad\mathbf{x}^{(t)}\mapsto\left\|\mathbf{x}^{(t)}-\mathbf{x}_1\right\|_2+\widetilde{\lambda}\left\|\mathbf{x}^{(t)}-\mathbf{x}_2\right\|_2+(1-\lambda)\mathbf{c}^{\top}\left(\mathbf{x}^{(t)}-\mathbf{x}_2\right).\end{aligned} \qquad \text{(A.20)}$$

Note that $\mathbf{x}_1$ and $\mathbf{x}_2$ are constant vectors as they were either a known initial condition, the optimal solution from the previous time step, or historically observed in $\mathfrak{D}_M$. Since all p-norms with $p\geq 1$ are convex, and we have that $\widetilde{\lambda}\geq 0$ since $\widetilde{\lambda}=\lambda\mathbb{1}^{\top}\mathbf{c}N^{-1}\geq 0$, the first two terms in (A.20) are convex. The third term is an affine function $\mathbf{c}^{\top}\left(\mathbf{x}^{(t)}-\mathbf{x}_2\right)$ scaled by a non-negative quantity $1-\lambda$ given that $\lambda\in[0,1]$. Thus, the third term is also convex, and we have that $\mathcal{J}\left(\mathbf{x}^{(t)}\right)$ is convex as it is the sum of three convex functions.

We can easily verify the convexity of the feasible region defined by the conservation constraints in (7.13) as any closed interval $[a,b]\subseteq\mathbb{R}$ is convex, and such a (non-measure zero) interval is exactly defined by $\left[\left\|\mathbf{x}^{(t)}_{\mathfrak{D}_M}\right\|_1-\delta,\left\|\mathbf{x}^{(t)}_{\mathfrak{D}_M}\right\|_1+\delta\right]\subseteq\mathbb{R}$ given a non-zero buffer $\delta>0$. □

Appendix B

Supplementary Figures

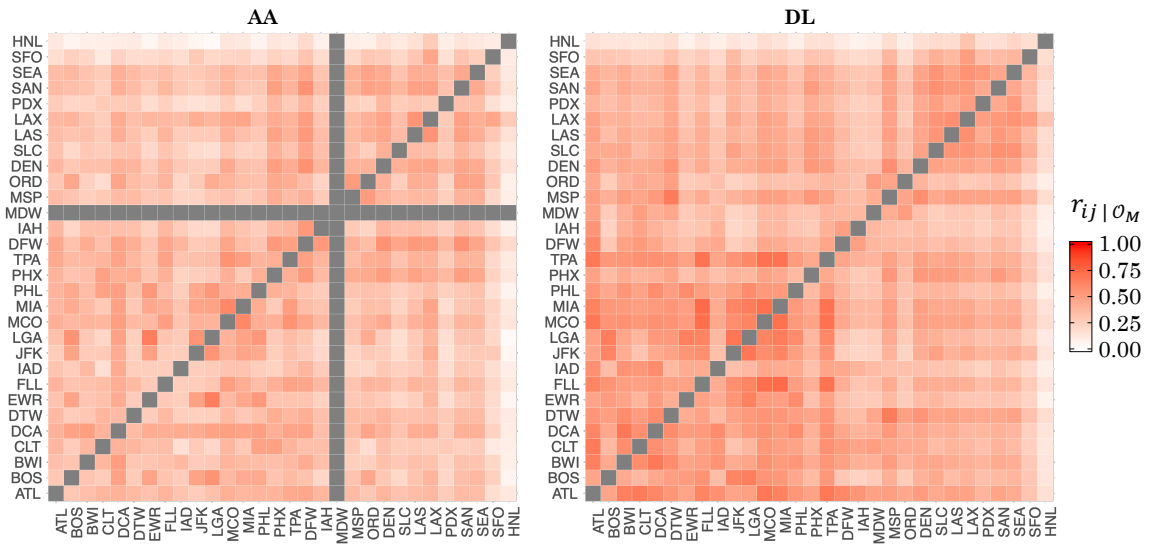


Figure B-1: Heat maps of the delay correlations between the top 30 airports for AA and DL. *Reprinted from [158] (Supplementary Material).*

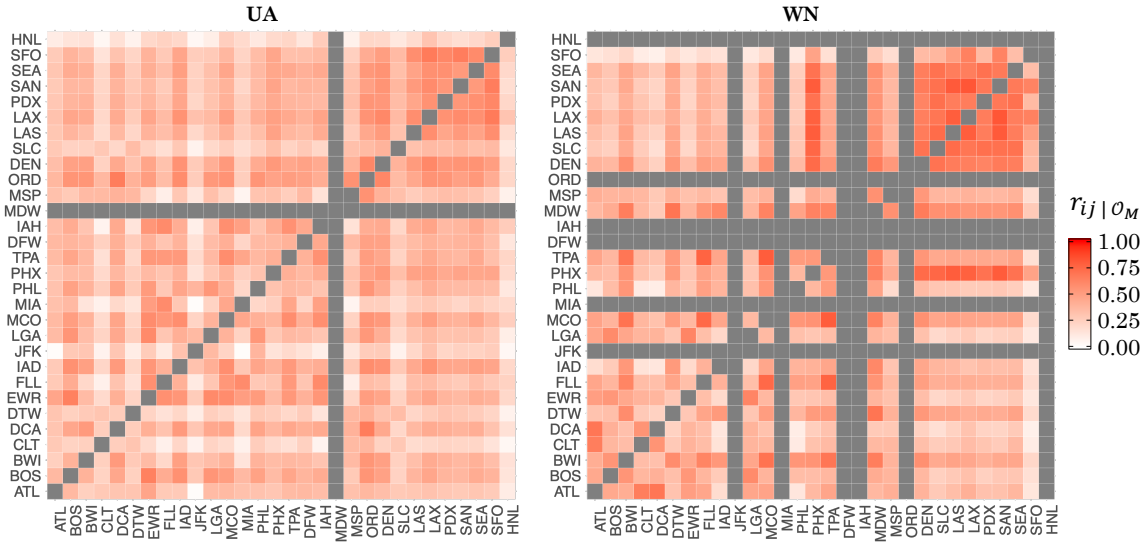


Figure B-2: Heat maps of the delay correlations between the top 30 airports for UA and WN. *Reprinted from [158] (Supplementary Material).*

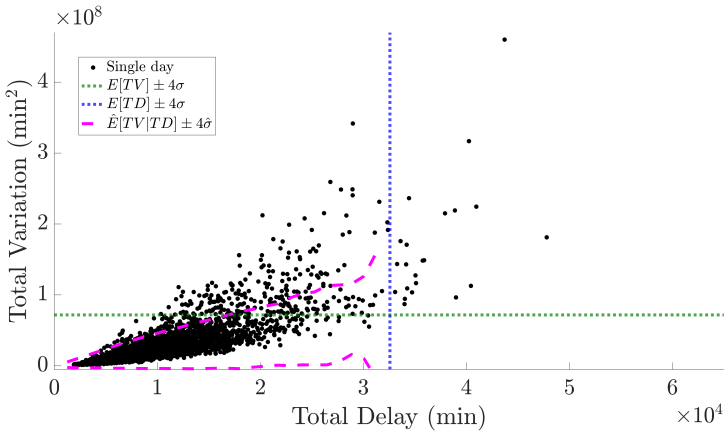


Figure B-3: AA sub-network: TV versus TD for all days in 2008-2017 with level $k = 4$ weak and strong outlier bounds demarcated. *Reprinted from [158] (Supplementary Material).*

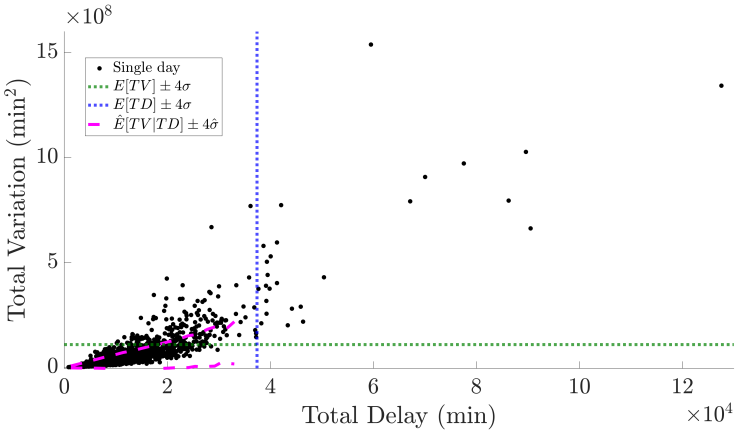


Figure B-4: DL sub-network: TV versus TD for all days in 2008-2017 with level $k = 4$ weak and strong outlier bounds demarcated. *Reprinted from [158] (Supplementary Material).*

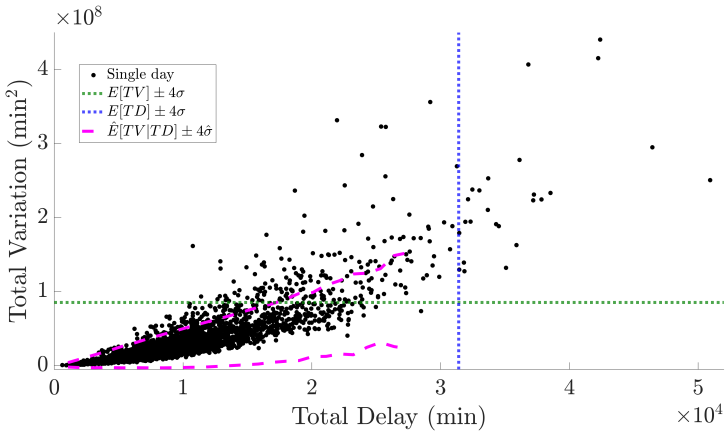


Figure B-5: UA sub-network: TV versus TD for all days in 2008-2017 with level $k = 4$ weak and strong outlier bounds demarcated. *Reprinted from [158] (Supplementary Material).*

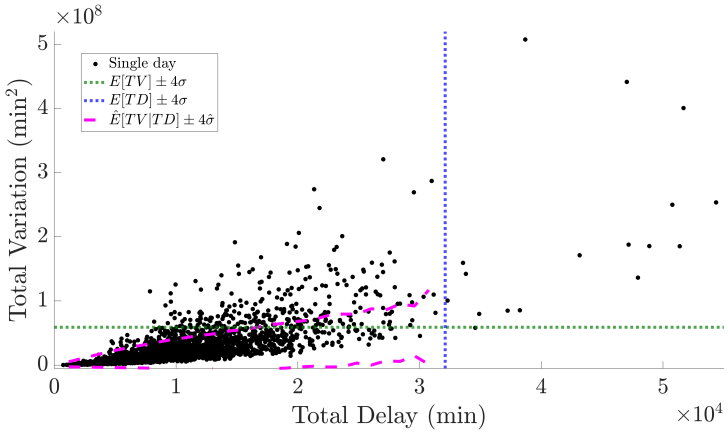


Figure B-6: WN sub-network: TV versus TD for all days in 2008-2017 with level $k = 4$ weak and strong outlier bounds demarcated. *Reprinted from [158] (Supplementary Material).*

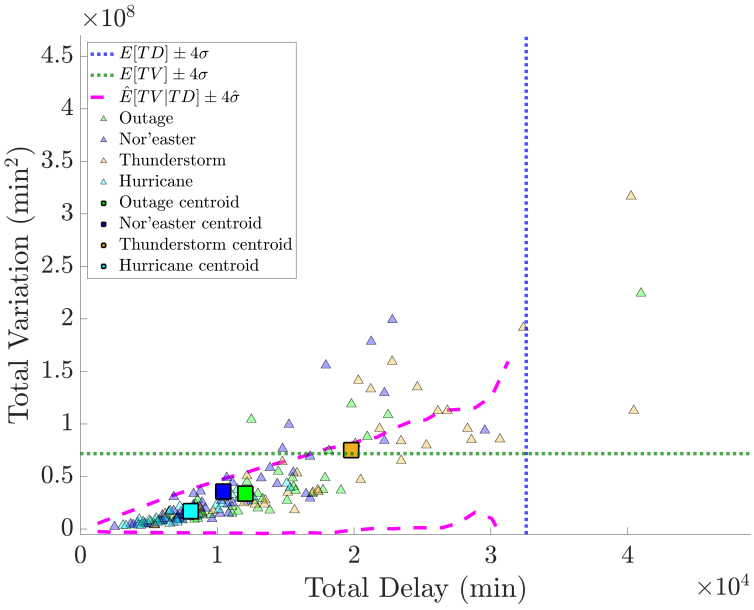


Figure B-7: TV versus total delay plot for American Airlines (AA) during 2008-2017 with specific disruptions and their average values annotated. *Reprinted from [158] (Supplementary Material).*

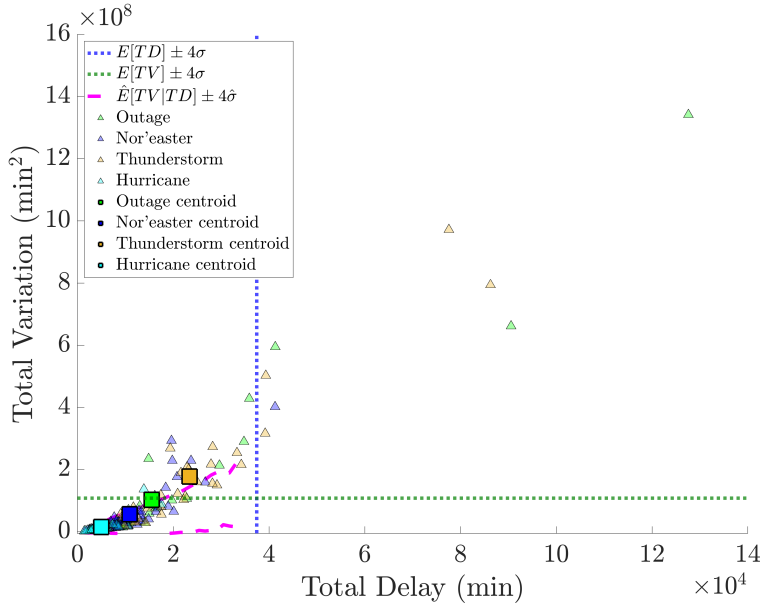


Figure B-8: TV versus total delay plot for Delta Air Lines (DL) during 2008-2017 with specific disruptions and their average values annotated. *Reprinted from [158] (Supplementary Material).*

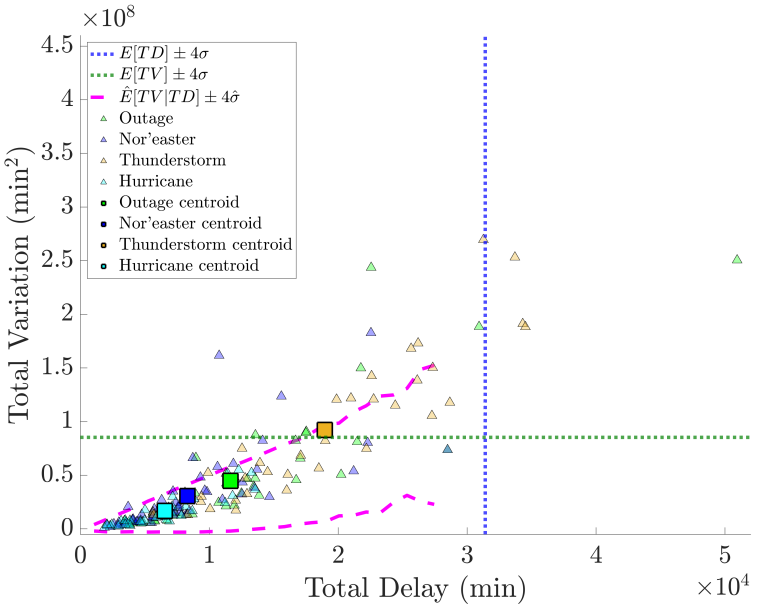


Figure B-9: TV versus total delay plot for United Airlines (UA) during 2008-2017 with specific disruptions and their average values annotated. *Reprinted from [158] (Supplementary Material).*

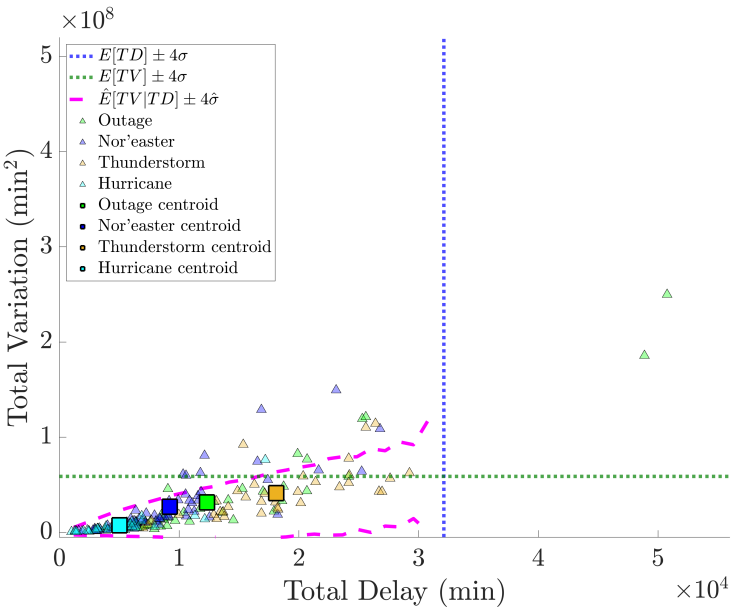


Figure B-10: TV versus total delay plot for Southwest Airlines (WN) during 2008-2017 with specific disruptions and their average values annotated. *Reprinted from [158] (Supplementary Material).*

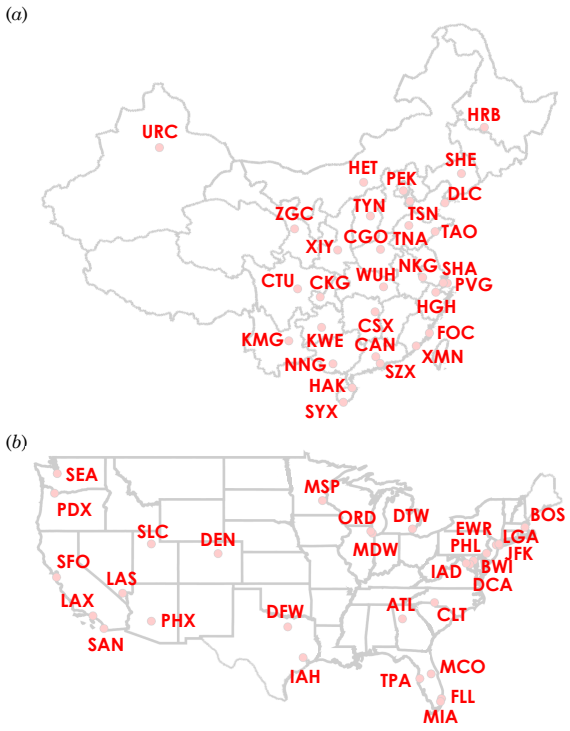


Figure B-11: Geographic locations of the airports (IATA code given) within our graph of China (a) and the US (b). Note that HNL is not shown in (b) for simplicity. *Reprinted from [161]. © 2020 IEEE*

Appendix C

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 C.1: IATA three-letter code and corresponding full airport name; the ARTCC that each airport is located within is also listed.

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

Table C.2: List of 178 disruption days used in the system-wide and airline-specific analysis. *Reprinted from [158] (Supplementary Material).*

Day-type	**Dates**							
	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

Day-type	Dates							
	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
0,0,0,0,1	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	
	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		
0,0,1,0,0	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			

Day-type	Dates							
0,1,0,0,0	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			
	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			
	7/6/13	10/31/14	2/19/16	3/31/17	5/1/17	7/24/17	10/19/17	

Day-type	Dates							
0,0,1,1,0	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 C.4: Inventory of days belonging to the top 9 most frequently occurring day-type tuples (excluding the no-outlier case). *Reprinted from [158] (Supplementary Material).*

Airline	Hubs (AA, DL, UA), Operating Bases and Focus Cities (WN)
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 C.3: 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. *Reprinted from [158] (Supplementary Material).*

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 C.5: Inventory of days belonging to the day-type tuples not captured in Table C.4. *Reprinted from [158] (Supplementary Material).*

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

Table C.6: IATA three-letter code and corresponding full airport name of the airports within our graph of China and the US. *Reprinted from [161].*

Bibliography

[1] Mohamed Abdel-Aty, Chris Lee, Yuqiong Bai, Xin Li, and Martin Michalak. Detecting periodic patterns of arrival delay. *Journal of Air Transport Management*, 13(6):355–361, 2007.

[2] Khaled F. Abdelghany, Sharmila S. Shah, Sidhartha Raina, and Ahmed F. Abdelghany. A model for projecting flight delays during irregular operation conditions. *Journal of Air Transport Management*, 10(6):385–394, 2004.

[3] Cesar O. Aguilar and Bahman Gharesifard. On almost equitable partitions and network controllability. In *2016 American Control Conference (ACC)*, pages 179–184, 2016.

[4] Shervin AhmadBeygi, Amy Cohn, and Marcial Lapp. Decreasing airline delay propagation by re-allocating scheduled slack. *IIE Transactions*, 42(7):478–489, 2010.

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

[6] Nir Ailon and Bernard Chazelle. The fast Johnson–Lindenstrauss transform and approximate nearest neighbors. *SIAM Journal on Computing*, 39(1):302–322, 2009.

[7] Airlines for America. A4A Presentation: Industry Review and Outlook. `https://www.airlines.org/dataset/a4a-presentation-industry-review-and-outlook/`, 2019.

[8] Giovanni Andreatta and Lorenzo Brunetta. Multiairport ground holding problem: A computational evaluation of exact algorithms. *Operations Research*, 46(1):57–64, 1998.

[9] Fatemeh Asadi and Arthur Richards. Ad hoc distributed model predictive control of air traffic management. *IFAC-PapersOnLine*, 48(25):68 – 73, 2015. 16th IFAC Workshop on Control Applications of Optimization CAO'2015.

[10] Samet Ayhan, Johnathan Pesce, Paul Comitz, David Sweet, Steve Bliesner, and Gary Gerberick. Predictive analytics with aviation big data. In *2013 Integrated*

Communications, Navigation and Surveillance Conference (ICNS), pages 1–13, 2013.

[11] Sandeep Badrinath, Max Z. Li, and Hamsa Balakrishnan. Integrated surface-airspace model of airport departures. *Journal of Guidance, Control, and Dynamics*, pages 1–15, Jan 2019.

[12] Xin Bai. Coordination, matchmaking, and resource allocation for large scale distributed systems. Master's thesis, University of Central Florida , 2006.

[13] Poornima Balakrishna, Rajesh Ganesan, and Lance Sherry. Accuracy of reinforcement learning algorithms for predicting aircraft taxi-out times: A case-study of tampa bay departures. *Transportation Research Part C: Emerging Technologies*, 18(6):950–962, 2010.

[14] Hamsa Balakrishnan and Bala G. Chandran. Optimal large-scale air traffic flow management. Technical report, Massachusetts Institute of Technology, 2014.

[15] Michael Ball, Cynthia Barnhart, Martin Dresner, Mark Hansen, Kevin Neels, Amedeo Odoni, Everett Peterson, Lance Sherry, Antonio Trani, and Bo Zou. Total delay impact study, 2010.

[16] Michael Ball, Cynthia Barnhart, George Nemhauser, and Amedeo Odoni. Air transportation: Irregular operations and control. *Handbooks in operations research and management science*, 14:1–67, 2007.

[17] Segio Barbarossa, Stefania Sardellitti, and Alfonso Farina. On sparse controllability of graph signals. In *2016 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pages 4104–4108, March 2016.

[18] Cynthia Barnhart, Dimitris Bertsimas, Constantine Caramanis, and Douglas Fearing. Equitable and efficient coordination in traffic flow management. *Transportation Science*, 46(2):262–280, 2012.

[19] Juan Andrés Bazerque and Pablo Monzón. Control of networked systems in the graph-frequency domain. In *2017 51st Asilomar Conference on Signals, Systems, and Computers*, pages 1444–1448, Oct 2017.

[20] Janet R. Bednarek. *Airports, Cities, and the Jet Age: US Airports Since 1945*, pages 1–9. Springer International Publishing, Cham, 2016.

[21] Hamid Behjat, Nora Leonardi, Leif Sörnmo, and Dimitri Van De Ville. Anatomically-adapted graph wavelets for improved group-level fMRI activation mapping. *NeuroImage*, 123:185 – 199, 2015.

[22] Seddik Belkoura, José Maria Peña, and Massimiliano Zanin. Generation and recovery of airborne delays in air transport. *Transportation Research Part C: Emerging Technologies*, 69:436 – 450, 2016.

[23] Peter Belobaba, Amedeo Odoni, and Cynthia Barnhart. *The Global Airline Industry*. Wiley, Cham, 2015.

[24] Michel Bergmann, Laurent Cordier, and Jean-Pierre Brancher. Optimal rotary control of the cylinder wake using proper orthogonal decomposition reduced-order model. *Physics of fluids*, 17(9):097101, 2005.

[25] Dimitri P. Bertsekas. *Dynamic Programming and Optimal Control*. Athena Scientific, 2nd edition, 2000.

[26] Dimitris Bertsimas and Shubham Gupta. Fairness and collaboration in network air traffic flow management: An optimization approach. *Transportation Science*, 50(1):57–76, 2016.

[27] Dimitris Bertsimas, Guglielmo Lulli, and Amedeo Odoni. An integer optimization approach to large-scale air traffic flow management. *Operations Research*, 59(1):211–227, 2011.

[28] Dimitris Bertsimas and Sarah Stock Patterson. The air traffic flow management problem with enroute capacities. *Operations Research*, 46(3):406–422, 1998.

[29] Peter J. Bickel and Elizaveta Levina. Covariance regularization by thresholding. *Ann. Statist.*, 36(6):2577–2604, 12 2008.

[30] Peter J. Bickel and Elizaveta Levina. Regularized estimation of large covariance matrices. *Ann. Statist.*, 36(1):199–227, 02 2008.

[31] Eric Bouyé, Valdo Durrleman, Ashkan Nikeghbali, Gaël Riboulet, and Thierry Roncalli. Copulas for finance: A reading guide and some applications. *SSRN Electronic Journal*, 03 2000.

[32] Jan K. Brueckner and Chrystyane Abreu. Airline fuel usage and carbon emissions: Determining factors. *Journal of Air Transport Management*, 62:10–17, 2017.

[33] G. Brys, M. Hubert, and A. Struyf. A robust measure of skewness. *Journal of Computational and Graphical Statistics*, 13(4):996–1017, 2004.

[34] Bureau of Transportation Statistics. Air Carrier Financial Reports (Form 41 Financial Data), 2019.

[35] Bureau of Transportation Statistics. Airline On-Time Performance and Causes of Flight Delays. `https://www.bts.gov/topics/airlines-and-airports/airline-time-performance-and-causes-flight-delays`, 2019.

[36] Bureau of Transportation Statistics. Airline On-Time Statistics and Delay Causes. `https://www.transtats.bts.gov/OT_Delay/OT_DelayCause1.asp`, 2019.

[37] Bureau of Transportation Statistics. Glossary. https://www.transtats.bts.gov/glossary.asp, 2019.

[38] Bureau of Transportation Statistics. Understanding the Reporting of Causes of Flight Delays and Cancellations. https://www.bts.gov/topics/airlines-and-airports/understanding-reporting-causes-flight-delays-and-cancellations, 2019.

[39] Ronald W. Butler. *Saddlepoint Approximations with Applications*. Cambridge University Press, 2007.

[40] Gurkaran Buxi and Mark Hansen. Generating day-of-operation probabilistic capacity scenarios from weather forecasts. *Transportation Research Part C: Emerging Technologies*, 33:153 – 166, 2013.

[41] Sandrine Carlier, Ivan de Lépinay, Jean-Claude Hustache, and Frank Jelinek. Environmental impact of air traffic flow management delays. *7th Air Traffic Management Research and Development Seminar*, 2007.

[42] Moncef Chaabouni. Least-cost ground holding strategies with departure and arrival delay uncertainties. Master's thesis, Massachusetts Institute of Technology, 2 1999.

[43] Shuwei Chen, Yanjun Wang, Minghua Hu, Ying Zhou, Daniel Delahaye, and Siyuan Lin. Community detection of chinese airport delay correlation network. In *2020 International Conference on Artificial Intelligence and Data Analytics for Air Transportation (AIDA-AT)*, pages 1–8, 2020.

[44] Siheng Chen, Rohan Varma, Aarti Singh, and Jelena Kovačević. Representations of piecewise smooth signals on graphs. In *2016 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pages 6370–6374, March 2016.

[45] Wen-Hua Chen and Xiao-Bing Hu. Receding horizon control for airport capacity management. *IEEE Transactions on Control Systems Technology*, 15(6):1131–1136, 2007.

[46] Zhenhua Chen, Yuxuan Wang, and Lei Zhou. Predicting weather-induced delays of high-speed rail and aviation in china. *Transport Policy*, 101:1–13, 2021.

[47] Gene Cheung, Enrico Magli, Yuichi Tanaka, and Michael K. Ng. Graph spectral image processing. *Proceedings of the IEEE*, 106(5):907–930, May 2018.

[48] Christopher Chin, Karthik Gopalakrishnan, Maxim Egorov, Antony Evans, and Hamsa Balakrishnan. Efficiency and fairness in unmanned air traffic flow management. *IEEE Transactions on Intelligent Transportation Systems*, pages 1–13, 2021.

[49] Christopher Chin, Max Z. Li, Karthik Gopalakrishnan, and Hamsa Balakrishnan. Airport ground holding with hierarchical control objectives, in progress.

[50] Sun Choi, Young Jin Kim, Simon Briceno, and Dimitri Mavris. Prediction of weather-induced airline delays based on machine learning algorithms. In *2016 IEEE/AIAA 35th Digital Avionics Systems Conference (DASC)*, pages 1–6, Sep. 2016.

[51] Fan R. K. Chung. *Spectral Graph Theory*. American Mathematical Society, 1997.

[52] Civil Aviation Administration of China. Statistical Communique of the Civil Aviation Administration of China on the Operations in the Civil Airports (2018). `https://http://www.caac.gov.cn/XXGK/XXGK/TJSJ/201903/t20190305_194972.html`. Accessed: 2019-10-27.

[53] Kevin L. Clark, Udit Bhatia, Evan A. Kodra, and Auroop R. Ganguly. Resilience of the US national airspace system airport network. *IEEE Transactions on Intelligent Transportation Systems*, 19(12):3785–3794, 2018.

[54] Jens Clausen, Allan Larsen, Jesper Larsen, and Natalia J. Rezanova. Disruption management in the airline industry—concepts, models and methods. *Computers & Operations Research*, 37(5):809–821, 2010.

[55] Gerald N. Cook and Jeremy Goodwin. Airline networks: A comparison of hub-and-spoke and point-to-point systems. *Journal of Aviation Aerospace Education and Research*, 17, 2008.

[56] Jonathan Cox and Mykel J. Kochenderfer. Probabilistic airport acceptance rate prediction. In *AIAA Modeling and Simulation Technologies Conference*, 2016.

[57] Mark Crovella and Eric Kolaczyk. Graph wavelets for spatial traffic analysis. In *IEEE INFOCOM 2003. Twenty-second Annual Joint Conference of the IEEE Computer and Communications Societies*, volume 3, pages 1848–1857 vol.3, March 2003.

[58] H. E. Daniels. Saddlepoint approximations in statistics. *Ann. Math. Statist.*, 25(4):631–650, 12 1954.

[59] Sanjoy Dasgupta and Anupam Gupta. An elementary proof of a theorem of Johnson and Lindenstrauss. *Random Struct. Algorithms*, 22(1):60–65, January 2003.

[60] Sarah Dean, Stephen Tu, Nikolai Matni, and Benjamin Recht. Safely learning to control the constrained linear quadratic regulator. In *2019 American Control Conference (ACC)*, pages 5582–5588, 2019.

[61] Tony Diana. Can machines learn how to forecast taxi-out time? a comparison of predictive models applied to the case of seattle/tacoma international airport. *Transportation Research Part E: Logistics and Transportation Review*, 119:149–164, 2018.

[62] Xiaowen Dong, Antonio Ortega, Pascal Frossard, and Pierre Vandergheynst. Inference of mobility patterns via spectral graph wavelets. In *2013 IEEE International Conference on Acoustics, Speech and Signal Processing*, pages 3118–3122, May 2013.

[63] Xiaoxia Dong and Megan S. Ryerson. Increasing civil aviation capacity in China requires harmonizing the physical and human components of capacity: A review and investigation. *Transportation Research Interdisciplinary Perspectives*, 1, 2019.

[64] Elisabeth Drayer and Tirza Routtenberg. Detection of false data injection attacks in smart grids based on graph signal processing. *IEEE Systems Journal*, pages 1–11, 2019.

[65] Michael Drew and Kapil Sheth. *A Frequency Analysis Approach for Categorizing Air Traffic Behavior*, pages 1–11. 14th AIAA ATIO Conference, Jun 2014.

[66] Michael Drew and Kapil Sheth. *A Wavelet Analysis Approach for Categorizing Air Traffic Behavior*, pages 1–14. AIAA ATIO Conference, Jun 2015.

[67] Wen-Bo Du, Xing-Lian Zhou, Oriol Lordan, Zhen Wang, Chen Zhao, and Yan-Bo Zhu. Analysis of the Chinese Airline Network as multi-layer networks. *Transportation Research Part E: Logistics and Transportation Review*, 89:108 – 116, 2016.

[68] Wenbo Du, Boyuan Liang, Gang Yan, Oriol Lordan, and Xianbin Cao. Identifying vital edges in Chinese air route network via memetic algorithm. *Chinese Journal of Aeronautics*, 30(1):330 – 336, 2017.

[69] William Eberle and Lawrence Holder. Discovering structural anomalies in graph-based data. In *Seventh IEEE International Conference on Data Mining Workshops (ICDMW 2007)*, pages 393–398, Oct 2007.

[70] Hilmi E. Egilmez and Antonio Ortega. Spectral anomaly detection using graph-based filtering for wireless sensor networks. In *2014 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pages 1085–1089, May 2014.

[71] Carsten Eickhoff, Arjen P. de Vries, and Kevyn Collins-Thompson. Copulas for information retrieval. In *SIGIR '13*, page 663–672, 2013.

[72] Paul Embrechts, Alexander J. McNeil, and Daniel Straumann. *Correlation and Dependence in Risk Management: Properties and Pitfalls*, page 176–223. Cambridge University Press, 2002.

[73] Ehsan Esmaeilzadeh and Seyedmirsajad Mokhtarimousavi. Machine learning approach for flight departure delay prediction and analysis. *Transportation Research Record*, 2674(8):145–159, 2020.

[74] Alexander Estes, Michael O. Ball, and David J. Lovell. Predicting performance of ground delay programs. In *12th USA/Europe Air Traffic Management Research and Development Seminar, ATM 2017*, January 2017.

[75] Antony D. Evans, Paul Lee, and Banavar Sridhar. Predicting the operational acceptance of airborne flight reroute requests using data mining. *Transportation Research Part C: Emerging Technologies*, 96:270–289, 2018.

[76] C. Evertse and H.G. Visser. Real-time airport surface movement planning: Minimizing aircraft emissions. *Transportation Research Part C: Emerging Technologies*, 79:224–241, 2017.

[77] FAA Office of Environment and Energy. Aviation emissions, impacts, and mitigation: A primer. Technical report, Federal Aviation Administration, 2019.

[78] Federal Aviation Administration. Traffic Management Initiatives. `http://tfmlearning.faa.gov/Publications/atpubs/FAC/1706.html#1706.html.1`, 2011.

[79] Federal Aviation Administration. SWIM Program Overview. `https://www.faa.gov/air_traffic/technology/swim/overview/`, 2018.

[80] Federal Aviation Administration. Aeronautical Information Manual. `https://www.faa.gov/air_traffic/publications/atpubs/aim_html/index.html`, 2019.

[81] Federal Aviation Administration. Community Involvement in Airport Planning. `https://www.faa.gov/documentLibrary/media/Advisory_Circular/150-5050-4A-Community-Involvement.pdf`, 2019.

[82] Federal Aviation Administration. Questions About Flight Delays. `https://www.fly.faa.gov/FAQ/faq.html`, 2019.

[83] Federal Aviation Administration (FAA). Traffic Flow Management in the National Airspace System, 2009.

[84] Federal Aviation Administration (FAA). Air Traffic Control System Command Center Advisories Database, 2020.

[85] Federal Aviation Administration (FAA). Aviation System Performance Metrics (ASPM) website, accessed 2018. http://aspm.faa.gov/.

[86] Jean-David Fermanian. Goodness-of-fit tests for copulas. *Journal of Multivariate Analysis*, 95(1):119 – 152, 2005.

[87] Peter Filzmoser. *A multivariate outlier detection method*. TU Wien, 2004.

[88] Marco Fiorucci, Francesco Pelosin, and Marcello Pelillo. Separating structure from noise in large graphs using the regularity lemma. *Pattern Recognition*, 98:107070, 2020.

[89] Ronald A. Fisher. Frequency distribution of the values of the correlation coefficient in samples from an indefinitely large population. *Biometrika*, 10(4):507–521, 1915.

[90] Mark I. Freidlin and Alexander D. Wentzell. Diffusion processes on graphs and the averaging principle. *The Annals of Probability*, 21(4):2215–2245, 1993.

[91] Jerome Friedman, Trevor Hastie, and Robert Tibshirani. Sparse inverse covariance estimation with the graphical lasso. *Biostatistics*, 9(3):432–441, Jul 2008.

[92] Pedro Galeano, Daniel Peña, and Ruey S. Tsay. Outlier detection in multivariate time series by projection pursuit. *Journal of the American Statistical Association*, 101(474):654–669, 2006.

[93] Fernando Gama, Elvin Isufi, Alejandro Ribeiro, and Geert Leus. Controllability of bandlimited graph processes over random time varying graphs. *IEEE Transactions on Signal Processing*, 67(24):6440–6454, 2019.

[94] Global Eagle (masFlight). MASFLIGHT. `https://www.globaleagle.com/products/operations-solutions/masflight`, 2019.

[95] Karthik Gopalakrishnan and Hamsa Balakrishnan. A comparative analysis of models for predicting delays in air traffic networks. In *Twelfth USA/Europe Air Traffic Management Research and Development Seminar (ATM2017)*, June 2017.

[96] Karthik Gopalakrishnan and Hamsa Balakrishnan. Control and optimization of air traffic networks. *Annual Review of Control, Robotics, and Autonomous Systems*, 4(1):397–424, 2021.

[97] Karthik Gopalakrishnan, Hamsa Balakrishnan, and Richard Jordan. Clusters and Communities in Air Traffic Delay Networks. In *American Control Conference*, July 2016.

[98] Karthik Gopalakrishnan, Hamsa Balakrishnan, and Richard Jordan. Deconstructing Delay Dynamics: An air traffic network example. In *International Conference on Research in Air Transportation (ICRAT)*, 2016.

[99] Karthik Gopalakrishnan, Hamsa Balakrishnan, and Richard Jordan. Stability of networked systems with switching topologies. In *Decision and Control (CDC), 2016 IEEE 55th Conference on*, pages 2601–2608. IEEE, 2016.

[100] Karthik Gopalakrishnan, Max Z. Li, and Hamsa Balakrishnan. Identification of outliers in graph signals. In *58th IEEE Conference on Decision and Control*, December 2019.

[101] Karthik Gopalakrishnan, Max Z. Li, and Hamsa Balakrishnan. Network-centric benchmarking of operational performance in aviation. *Transportation Research Part C*, 2021.

[102] Sreeta Gorripaty, Yi Liu, Mark Hansen, and Alexey Pozdnukhov. Identifying similar days for air traffic management. *Journal of Air Transport Management*, 65:144 – 155, 2017.

[103] Shon R. Grabbe, Banavar Sridhar, and Avijit Mukherjee. Similar days in the NAS: an airport perspective. In *2013 Aviation Technology, Integration, and Operations Conference*, page 4222, 2013.

[104] Brandon Graver, Kevin Zhang, and Dan Rutherford. CO2 emissions from commercial aviation, 2018. Technical report, The International Council on Clean Transportation, 2019.

[105] Vincent Gripon, Antonio Ortega, and Benjamin Girault. An inside look at deep neural networks using graph signal processing. In *2018 Information Theory and Applications Workshop (ITA)*, pages 1–9, Feb 2018.

[106] R. Guimerà, S. Mossa, A. Turtschi, and L. A. N. Amaral. The worldwide air transportation network: Anomalous centrality, community structure, and cities' global roles. *Proceedings of the National Academy of Sciences*, 102(22):7794–7799, 2005.

[107] Ismail Gultepe and Wayne F. Feltz. Aviation meteorology: Observations and models. introduction. *Pure and Applied Geophysics*, 176(5):1863–1867, May 2019.

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

[109] Ali Haghani and Min-Ching Chen. Optimizing gate assignments at airport terminals. *Transportation Research Part A: Policy and Practice*, 32(6):437 – 454, 1998.

[110] Lu Hao, Megan S. Ryerson, Lei Kang, and Mark Hansen. Estimating fuel burn impacts of taxi-out delay with implications for gate-hold benefits. *Transportation Research Part C: Emerging Technologies*, 80:454 – 466, 2017.

[111] Kanghang He, Lina Stankovic, Jing Liao, and Vladimir Stankovic. Non-intrusive load disaggregation using graph signal processing. *IEEE Transactions on Smart Grid*, 9(3):1739–1747, May 2018.

[112] Floris Herrema, Richard Curran, Hendrikus Visser, Denis Huet, and Régis Lacote. Taxi-out time prediction model at Charles de Gaulle Airport. *Journal of Aerospace Information Systems*, 15(3):120–130, 2018.

[113] Robert Hoffman and Michael O. Ball. A comparison of formulations for the single-airport ground-holding problem with banking constraints. *Operations Research*, 48(4):578–590, 2000.

[114] Chenhui Hu, Lin Cheng, Jorge Sepulcre, Keith A. Johnson, Georges E. Fakhri, Yue M. Lu, and Quanzheng Li. A spectral graph regression model for learning brain connectivity of alzheimer's disease. *PLOS ONE*, 10(5):1–24, 05 2015.

[115] Weiyu Huang, Thomas A. W. Bolton, John D. Medaglia, Danielle S. Bassett, Alejandro Ribeiro, and Dimitri Van De Ville. Graph signal processing of human brain imaging data. In *2018 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pages 980–984, April 2018.

[116] Weiyu Huang, Leah Goldsberry, Nicholas F. Wymbs, Scott T. Grafton, Danielle S. Bassett, and Alejandro Ribeiro. Graph frequency analysis of brain signals. *IEEE Journal of Selected Topics in Signal Processing*, 10(7):1189–1203, Oct 2016.

[117] M. Hubert and E. Vandervieren. An adjusted boxplot for skewed distributions. *Computational Statistics & Data Analysis*, 52(12):5186 – 5201, 2008.

[118] International Air Transport Association. IATA Annual Review 2019. `https://www.iata.org/publications/Documents/iata-annual-review-2019.pdf`, 2019.

[119] International Air Transport Association (IATA). The Importance of Air Transport to the People's Republic of China. `https://www.iata.org/en/iata-repository/publications/economic-reports/china--value-of-aviation/`, 2018.

[120] International Air Transport Association (IATA). IATA Statistics. `https://www.iata.org/en/publications/annual-review/`, 2019. Accessed: 2020-05-14.

[121] International Civil Aviation Organization. Presentation of 2018 Air Transport Statistical Results. `https://www.icao.int/annual-report-2018/Documents/Annual.Report.2018_Air%20Transport%20Statistics.pdf`, 2018.

[122] Takayuki Ishizaki, Risong Ku, and Jun-ichi Imura. Clustered model reduction of networked dissipative systems. In *2016 American Control Conference (ACC)*, pages 3662–3667, 2016.

[123] Leon Isserlis. On a formula for the product-moment coefficient of any order of a normal frequency distribution in any number of variables. *Biometrika*, 12(1/2):134–139, 1918.

[124] Elvin Isufi, Paolo Banelli, Paolo Di Lorenzo, and Geert Leus. Observing and tracking bandlimited graph processes, 2017.

[125] Alexandre Jacquillat and Amedeo R. Odoni. An integrated scheduling and operations approach to airport congestion mitigation. *Operations Research*, 63(6):1390–1410, 2015.

[126] Rishee K. Jain, Jose M. F. Moura, and Constantine E. Kontokosta. Big data + big cities: Graph signals of urban air pollution [exploratory sp]. *IEEE Signal Processing Magazine*, 31(5):130–136, Sep. 2014.

[127] Milan Janić. Modelling the resilience, friability and costs of an air transport network affected by a large-scale disruptive event. *Transportation Research Part A: Policy and Practice*, 71:1 – 16, 2015.

[128] Ahmad I. Z. Jarrah, Gang Yu, Nirup Krishnamurthy, and Ananda Rakshit. A decision support framework for airline flight cancellations and delays. *Transportation Science*, 27(3):266–280, 1993.

[129] John R. Jensen. Analysis of the impact of collaborative ground delay programs in air traffic control. Master's thesis, Massachusetts Institute of Technology, 1999.

[130] Alexander Jung, Peter Berger, Gabor Hannak, and Gerald Matz. Scalable graph signal recovery for big data over networks. In *2016 IEEE 17th International Workshop on Signal Processing Advances in Wireless Communications (SPAWC)*, pages 1–6, 2016.

[131] Alexander Jung and Nguyen Tran. Localized linear regression in networked data. *CoRR*, abs/1903.11178, 2019.

[132] Nabin Kafle and Bo Zou. Modeling flight delay propagation: A new analytical-econometric approach. *Transportation Research Part B: Methodological*, 93:520 – 542, 2016.

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

[134] Megumi Kaneko, Gene Cheung, Weng-Tai Su, and Chia-Wen Lin. Graph-based joint signal/power restoration for energy harvesting wireless sensor networks. In *GLOBECOM 2017 - 2017 IEEE Global Communications Conference*, pages 1–6, Dec 2017.

[135] Zohar S. Karnin, Yuval Rabani, and Amir Shpilka. Explicit dimension reduction and its applications. *SIAM J. Comput.*, 41(1):219–249, February 2012.

[136] Harshad Khadilkar and Hamsa Balakrishnan. Network congestion control of airport surface operations. *Journal of Guidance, Control, and Dynamics*, 37(3):933–940, Mar 2014.

[137] Mohammad Khalil, Sondipon Adhikari, and Abhijit Sarkar. Linear system identification using proper orthogonal decomposition. *Mechanical Systems and Signal Processing*, 21(8):3123–3145, 2007.

[138] Amy Kim and Mark Hansen. Deconstructing delay: A non-parametric approach to analyzing delay changes in single server queuing systems. *Transportation Research Part B: Methodological*, 58:119 – 133, 2013.

[139] Gunky Kim, Mervyn J. Silvapulle, and Paramsothy Silvapulle. Comparison of semiparametric and parametric methods for estimating copulas. *Computational Statistics and Data Analysis*, 51(6):2836 – 2850, 2007.

[140] Myeonghyeon Kim, Yuri Choi, and Ki Han Song. Identification model development for proactive response on irregular operations (IROPs). *Journal of Air Transport Management*, 75:1 – 8, 2019.

[141] Seyun Kim, Jiseon Lee, Soohwan Oh, and Yoonjin Yoon. Assessment of the volcanic hazard of Mt. Paektu explosion to international air traffic using South Korean airspace. *Natural Hazards*, 96(2):647–667, Mar 2019.

[142] Seyun Kim and Yoonjin Yoon. On node criticality of the Northeast Asian air route network. *Journal of Air Transport Management*, 80:101693, 2019.

[143] Young Jin Kim, Sun Choi, Simon Briceno, and Dimitri Mavris. A deep learning approach to flight delay prediction. In *2016 IEEE/AIAA 35th Digital Avionics Systems Conference (DASC)*, pages 1–6, Sep. 2016.

[144] Erik Kole, Kees Koedijk, and Marno Verbeek. Selecting copulas for risk management. *Journal of Banking & Finance*, 31(8):2405 – 2423, 2007.

[145] Alexander Kreuzer, Luciana Dalla Valle, and Claudia Czado. Bayesian multivariate nonlinear state space copula models. *arXiv preprint arXiv:1911.00448*, 2019.

[146] Kenneth D. Kuhn. Ground delay program planning: Delay, equity, and computational complexity. *Transportation Research Part C: Emerging Technologies*, 35:193–203, 2013.

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

[148] Miguel Lambelho, Mihaela Mitici, Simon Pickup, and Alan Marsden. Assessing strategic flight schedules at an airport using machine learning-based flight delay and cancellation predictions. *Journal of Air Transport Management*, 82:101737, 2020.

[149] Steven Landau, Geoffrey D. Gosling, Kenneth Small, and Thomas Adler. Measuring air carrier passengers' values of time by trip component. *Transportation Research Record*, 2569(1):24–31, 2016.

[150] Dan Larsen and Michael Robinson. A non-parametric discrete choice model for airport acceptance rate prediction. In *AIAA Aviation 2019 Forum*, June 2019.

[151] Bo Li and Marc G. Genton. Nonparametric identification of copula structures. *Journal of the American Statistical Association*, 108(502):666–675, 2013.

[152] Lishuai Li, Santanu Das, R. John Hansman, Rafael Palacios, and Ashok N. Srivastava. Analysis of flight data using clustering techniques for detecting abnormal operations. *Journal of Aerospace Information Systems*, 12(9):587–598, 2015.

[153] Lishuai Li, R. John Hansman, Rafael Palacios, and Roy Welsch. Anomaly detection via a gaussian mixture model for flight operation and safety monitoring. *Transportation Research Part C: Emerging Technologies*, 64:45 – 57, 2016.

[154] Max Z. Li. Towards more resilient and predictable skies: Data-driven characterization and modeling of the terminal departure and arrival airspace. Master's thesis, University of Pennsylvania, 2018.

[155] Max Z. Li, Karthik Gopalakrishnan, and Hamsa Balakrishnan. Approximate projection-based control of networks. In *59th IEEE Conference on Decision and Control*, December 2020.

[156] Max Z. Li, Karthik Gopalakrishnan, Hamsa Balakrishnan, Xiyitao Zhu, Aritro Nandi, and Lavanya Marla. Identification and prediction of disruptions in airline networks, in progress.

[157] Max Z. Li, Karthik Gopalakrishnan, Kristyn Pantoja, and Hamsa Balakrishnan. A spectral approach towards analyzing airport performance and disruptions. *13th Air Traffic Management Research and Development Seminar*, 2019.

[158] Max Z. Li, Karthik Gopalakrishnan, Kristyn Pantoja, and Hamsa Balakrishnan. Graph signal processing techniques for analyzing aviation disruptions. *Transportation Science*, 55(3):553–573, 2021.

[159] Max Z. Li, Karthik Gopalakrishnan, Sang Shin, Darsh Jalan, Aritro Nandi, Hamsa Balakrishnan, and Lavanya Marla. Dynamics of disruption and recovery in air transportation networks. *9^{th} International Conference on Research in Air Transportation*, 2020.

[160] Max Z. Li, Karthik Gopalakrishnan, and Yanjun Wang. Spatial Delay Distributions in Chinese Airline Networks. In *AIAA AVIATION Forum*, June 2020.

[161] Max Z. Li, Karthik Gopalakrishnan, Yanjun Wang, and Hamsa Balakrishnan. Outlier Analysis of Airport Delay Distributions in US and China. In *IEEE Artificial Intelligence and Data Analytics in Air Transportation Conference*, February 2020.

[162] Max Z. Li and Megan S. Ryerson. A data-driven approach to modeling high-density terminal areas: A scenario analysis of the new Beijing, China airspace. *Chinese Journal of Aeronautics*, 30(2):538 – 553, 2017.

[163] Max Z. Li and Megan S. Ryerson. Reviewing the DATAS of aviation research data: Diversity, Availability, Tractability, Applicability, and Sources. *Journal of Air Transport Management*, 75:111 – 130, 2019.

[164] Max Z. Li, Megan S. Ryerson, and Hamsa Balakrishnan. Topological data analysis for aviation applications. *Transportation Research Part E: Logistics and Transportation Review*, 128:149 – 174, 2019.

[165] Max Z. Li, Daniel Y. Suh, and Megan S. Ryerson. Visualizing aviation impacts: Modeling current and future flight trajectories with publicly available flight data. *Transportation Research Part D: Transport and Environment*, 63:769 – 785, 2018.

[166] Weiwei Li and Emanuel Todorov. Iterative linear quadratic regulator design for nonlinear biological movement systems. In *Proceedings of the 1st International Conference on Informatics in Control, Automation and Robotics, (ICINCO 2004)*, volume 1, pages 222–229, 01 2004.

[167] Man Liang, Daniel Delahaye, and Pierre Maréchal. Integrated sequencing and merging aircraft to parallel runways with automated conflict resolution and advanced avionics capabilities. *Transportation Research Part C: Emerging Technologies*, 85:268 – 291, 2017.

[168] Pei-chen Barry Liu, Mark Hansen, and Avijit Mukherjee. Scenario-based air traffic flow management: From theory to practice. *Transportation Research Part B: Methodological*, 42(7):685 – 702, 2008.

[169] Yang-Yu Liu, Jean-Jacques Slotine, and Albert-László Barabási. Controllability of complex networks. *Nature*, 473(7346):167–173, 2011.

[170] Yi Liu and Mark Hansen. Incorporating predictability into cost optimization for ground delay programs. *Transportation Science*, 50(1):132–149, 2016.

[171] Lennart Ljung. System identification. *Wiley Encyclopedia of Electrical and Electronics Engineering*, pages 1–19, 1999.

[172] Anna Lombardi and Michael Hörnquist. Controllability analysis of networks. *Phys. Rev. E*, 75:056110, May 2007.

[173] Dou Long, Shahab Hasan, Virginia L. Stouffer, Kris Ramamoorthy, Husni R. Idris, B. David Ballard, and Gregory Carr. Analytical identification and ranking of choke points in the national airspace system. In *15th AIAA Aviation Technology, Integration, and Operations Conference*, 2015.

[174] Oriol Lordan, Jose M. Sallan, Pep Simo, and David Gonzalez-Prieto. Robustness of the air transport network. *Transportation Research Part E: Logistics and Transportation Review*, 68:155 – 163, 2014.

[175] Bengi Manley and Lance Sherry. The impact of Ground Delay Program (GDP) rationing rules on passenger and airline equity. In *International Conference on Research in Air Transportation (ICRAT)*, 2008.

[176] Hans Manner and Olga Reznikova. A survey on time-varying copulas: Specification, simulations, and application. *Econometric Reviews*, 31(6):654–687, 2012.

[177] Aude Marzuoli, Emmanuel Boidot, Pablo Colomar, Mathieu Guerpillon, Eric Feron, Alexandre Bayen, and Mark Hansen. Improving disruption management with multimodal collaborative decision-making: A case study of the asiana crash and lessons learned. *IEEE Transactions on Intelligent Transportation Systems*, 17(10):2699–2717, Oct 2016.

[178] A. M. Mathai and Serge B. Provost. *Quadratic Forms in Random Variables*. Marcel Dekker, Inc., 1992.

[179] Rahul Mazumder and Trevor Hastie. The graphical lasso: New insights and alternatives, 2011.

[180] Scott McCartney. The story behind your canceled flight. `https://www.wsj.com/articles/the-story-behind-your-canceled-flight-1426093364`, 2015.

[181] P. K. Menon, G. D. Sweriduk, T. Lam, G. M. Diaz, and Karl D. Bilimoria. Computer-aided Eulerian air traffic flow modeling and predictive control. *Journal of Guidance, Control, and Dynamics*, 29(1):12–19, 2006.

[182] MIT Global Airline Industry Program. Airline Data Project: Aircraft and Related. `http://web.mit.edu/airlinedata/www/Aircraft&Related.html`, 2019.

[183] Iman Mohammadian, Babak Abbasi, Ahmad Abareshi, and Mark Goh. Antecedents of flight delays in the australian domestic aviation market. *Transportation Research Interdisciplinary Perspectives*, 1:100007, 2019.

[184] Dhanya Menoth Mohan, Muhammad Tayyab Asif, Nikola Mitrovic, Justin Dauwels, and Patrick Jaillet. Wavelets on graphs with application to transportation networks. In *17th International IEEE Conference on Intelligent Transportation Systems (ITSC)*, pages 1707–1712, Oct 2014.

[185] Nima Monshizadeh, Harry L. Trentelman, and M. Kanat Camlibel. Projection-based model reduction of multi-agent systems using graph partitions. *IEEE Transactions on Control of Network Systems*, 1(2):145–154, 2014.

[186] Nima Monshizadeh, Harry L. Trentelman, and M. Kanat Camlibel. Stability and synchronization preserving model reduction of multi-agent systems. *Systems & Control Letters*, 62(1):1–10, 2013.

[187] Nima Monshizadeh and Arjan van der Schaft. Structure-preserving model reduction of physical network systems by clustering. In *53rd IEEE Conference on Decision and Control*, pages 4434–4440, 2014.

[188] Avijit Mukherjee, Shon Grabbe, and Banavar Sridhar. Classification of Days using Weather Impacted Traffic in the National Airspace System. In *AIAA Aviation Technology, Integration and Operations Conference*, 2013.

[189] Biswanath Mukherjee, M. Farhan Habib, and Ferhat Dikbiyik. Network adaptability from disaster disruptions and cascading failures. *IEEE Communications Magazine*, 52(5):230–238, 2014.

[190] Mayara Condé Rocha Murça and McWillian de Oliveira. A data-driven probabilistic trajectory model for predicting and simulating terminal airspace operations. In *2020 AIAA/IEEE 39th Digital Avionics Systems Conference (DASC)*, pages 1–7, Oct 2020.

[191] Mayara Condé Rocha Murça and R. John Hansman. Identification, characterization, and prediction of traffic flow patterns in multi-airport systems. *IEEE Transactions on Intelligent Transportation Systems*, 20(5):1683–1696, May 2019.

[192] Mayara Condé Rocha Murça and R. John Hansman. Predicting and planning airport acceptance rates in metroplex systems for improved traffic flow management decision support. *Transportation Research Part C: Emerging Technologies*, 97:301–323, 2018.

[193] Mathilde Ménoret, Nicolas Farrugia, Bastien Pasdeloup, and Vincent Gripon. Evaluating graph signal processing for neuroimaging through classification and dimensionality reduction. In *2017 IEEE Global Conference on Signal and Information Processing (GlobalSIP)*, pages 618–622, 2017.

[194] Alex H. Nakahara and Tom G. Reynolds. Estimating current and future system-wide benefits of airport surface congestion management. *10th Air Traffic Management Research and Development Seminar*, 2013.

[195] National Oceanic and Atmospheric Administration. Atlantic Hurricane Seasons, 2018.

[196] National Oceanic and Atmospheric Administration. Saffir-Simpson Hurricane Wind Scale, 2018.

[197] National Oceanic and Atmospheric Administration. Radar Data Map, 2019.

[198] National Weather Service. What is a Nor'Easter?, 2019.

[199] Roger B. Nelson. *An Introduction to Copulas*, pages 7–49. Springer, New York, NY, 2006.

[200] Hohsuk Noh, Anouar El Ghouch, and Taoufik Bouezmarni. Copula-based regression estimation and inference. *Journal of the American Statistical Association*, 108(502):676–688, 2013.

[201] Jenaro Nosedal, Miquel A. Piera, Sergio Ruiz, and Alvaro Nosedal. An efficient algorithm for smoothing airspace congestion by fine-tuning take-off times. *Transportation Research Part C: Emerging Technologies*, 44:171 – 184, 2014.

[202] OAG. OAG Schedule Analyzer. `https://www.oag.com/schedules-analyser`, 2019.

[203] Ingram Olkin and John W. Pratt. Unbiased estimation of certain correlation coefficients. *The Annals of Mathematical Statistics*, 29(1):201–211, 1958.

[204] Operation Monitoring Center, the Civil Aviation Administration of China. Civil flights operation performance report. Technical report, Civil Aviation Administration of China, 2018.

[205] Erika J. Palin, Adam A. Scaife, Emily Wallace, Edward C. D. Pope, Alberto Arribas, and Anca Brookshaw. Skillful seasonal forecasts of winter disruption to the UK transport system. *Journal of Applied Meteorology and Climatology*, 55(2):325–344, 2016.

[206] Jiahao Pang, Gene Cheung, Antonio Ortega, and Oscar C. Au. Optimal graph laplacian regularization for natural image denoising. In *2015 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pages 2294–2298, April 2015.

[207] Benjamin Peherstorfer and Karen Willcox. Dynamic data-driven reduced-order models. *Computer Methods in Applied Mechanics and Engineering*, 291:21–41, 2015.

[208] Gabriel Peyré and Marco Cuturi. *Computational Optimal Transport: With Applications to Data Science*. Now Foundations and Trends, 2019.

[209] Jean Pouget-Abadie and Thibaut Horel. Inferring graphs from cascades: A sparse recovery framework. In Francis Bach and David Blei, editors, *Proceedings of the 32nd International Conference on Machine Learning*, volume 37 of *Proceedings of Machine Learning Research*, pages 977–986, Lille, France, 07–09 Jul 2015. PMLR.

[210] Victor M. Preciado and M. Amin Rahimian. Moment-based spectral analysis of random graphs with given expected degrees. *IEEE Transactions on Network Science and Engineering*, 4(4):215–228, 2017. https://doi.org/10.1109/TNSE.2017.2712064.

[211] Hugo M. Proença, Ruben Klijn, Thomas Bäck, and Matthijs van Leeuwen. Identifying flight delay patterns using diverse subgroup discovery. In *2018 IEEE Symposium Series on Computational Intelligence (SSCI)*, pages 60–67, Nov 2018.

[212] Nikolas Pyrgiotis. Public policy model of delays in a large network of major airports. *Transportation Research Record*, 2206(1):69–83, 2011.

[213] Nikolas Pyrgiotis, Kerry M. Malone, and Amedeo Odoni. Modelling delay propagation within an airport network. *Transportation Research Part C: Emerging Technologies*, 27:60 – 75, 2013. Selected papers from the Seventh Triennial Symposium on Transportation Analysis (TRISTAN VII).

[214] Nikolas Pyrgiotis and Amedeo Odoni. On the impact of scheduling limits: A case study at newark liberty international airport. *Transportation Science*, 50(1):150–165, 2016.

[215] J.V. Pérez–Rodríguez, J.M. Pérez–Sánchez, and E. Gómez–Déniz. Modelling the asymmetric probabilistic delay of aircraft arrival. *Journal of Air Transport Management*, 62:90 – 98, 2017.

[216] Shuo Qin, Jianhong Mou, Saran Chen, and Xin Lu. Modeling and optimizing the delay propagation in Chinese aviation networks. *Chaos: An Interdisciplinary Journal of Nonlinear Science*, 29(8):081101, 2019.

[217] Juan Jose Rebollo and Hamsa Balakrishnan. Characterization and prediction of air traffic delays. *Transportation Research Part C: Emerging Technologies*, 44:231 – 241, 2014.

[218] Kexin (May) Ren, Amy M. Kim, and Kenneth Kuhn. Exploration of the evolution of airport ground delay programs. *Transportation Research Record*, 2672(23):71–81, 2018.

[219] Pan Ren and Lishuai Li. Characterizing air traffic networks via large-scale aircraft tracking data: A comparison between China and the US networks. *Journal of Air Transport Management*, 67:181 – 196, 2018.

[220] Octavio Richetta and Amedeo R. Odoni. Solving optimally the static ground-holding policy problem in air traffic control. *Transportation Science*, 27(3):228–238, 1993.

[221] David M. Rocke and David L. Woodruff. Identification of outliers in multivariate data. *Journal of the American Statistical Association*, 91(435):1047–1061, 1996.

[222] S. Roy and Mengran Xue. Vulnerability metrics for the airspace system. In *Twelfth USA/Europe Air Traffic Management Research and Development Seminar (ATM2017)*, 2017.

[223] Franziska Ruh. Munich airport's third runway and stakeholder communications. *Journal of European Management and Public Affairs Studies*, 2(1), 2014.

[224] Liu Rui, Hossein Nejati, and Ngai-Man Cheung. Dimensionality reduction of brain imaging data using graph signal processing. In *2016 IEEE International Conference on Image Processing (ICIP)*, pages 1329–1333, 2016.

[225] Megan S. Ryerson. Diversion ahead: Modeling the factors driving diversion airport choice. *Journal of Infrastructure Systems*, 24(1):04017039, 2018.

[226] Megan S. Ryerson and Andrew M. Churchill. Aircraft rerouting due to abrupt facility outages: Case study of the 2011 Great Tōhoku Earthquake, Japan. *Transportation Research Record*, 2336(1):27–35, 2013.

[227] Megan S. Ryerson and Hyun Kim. Integrating airline operational practices into passenger airline hub definition. *Journal of Transport Geography*, 31:84 – 93, 2013.

[228] M Samimy, M Debiasi, E Caraballo, A Serrani, X Yuan, J Little, and JH Myatt. Feedback control of subsonic cavity flows using reduced-order models. *Journal of Fluid Mechanics*, 579:315–346, 2007.

[229] H. Sandberg and R. M. Murray. Model reduction of interconnected linear systems. *Optimal Control Applications and Methods*, 30(3):225–245, 2009.

[230] Aliaksei Sandryhaila and Jose M. F. Moura. Discrete signal processing on graphs. *IEEE Transactions on Signal Processing*, 61(7):1644–1656, April 2013.

[231] Aliaksei Sandryhaila and Jose M. F. Moura. 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, Sept 2014.

[232] Aliaksei Sandryhaila and Jose M. F. Moura. Discrete signal processing on graphs: Frequency analysis. *IEEE Transactions on Signal Processing*, 62(12):3042–3054, June 2014.

[233] Joseph Sarkis and Srinivas Talluri. Performance based clustering for benchmarking of US airports. *Transportation Research Part A: Policy and Practice*, 38(5):329 – 346, 2004.

[234] David Schaar and Lance Sherry. Analysis of airport stakeholders. In *2010 Integrated Communications, Navigation, and Surveillance Conference Proceedings*, pages J4–1–J4–17, May 2010.

[235] Michael T. Schaub. *Unraveling complex networks under the prism of dynamical processes: Relations between structure and dynamics*. PhD thesis, Imperial College London, 7 2014.

[236] Michael T. Schaub, Austin R. Benson, Paul Horn, Gabor Lippner, and Ali Jadbabaie. Random walks on simplicial complexes and the normalized hodge 1-laplacian. *SIAM Review*, 62(2):353–391, 2020.

[237] Michael T. Schaub and Santiago Segarra. Flow smoothing and denoising: Graph signal processing in the edge-space. In *2018 IEEE Global Conference on Signal and Information Processing (GlobalSIP)*, pages 735–739, 2018.

[238] Ioannis D. Schizas. Online data dimensionality reduction and reconstruction using graph filtering. *IEEE Transactions on Signal Processing*, 68:3871–3886, 2020.

[239] Michael Schultz, Stefan Reitmann, and Sameer Alam. Classification of weather impacts on airport operations. In *2019 Winter Simulation Conference (WSC)*, pages 500–511, Dec 2019.

[240] Thomas Schürmann and Ingo Hoffmann. On biased correlation estimation. *arXiv preprint arXiv:1707.09037*, 2017.

[241] Chze Eng Seah, Alinda Aligawesa, and Inseok Hwang. Algorithm for conformance monitoring in air traffic control. *Journal of Guidance, Control, and Dynamics*, 33(2):500–509, 2010.

[242] Shashi Shekhar, Chang-Tien Lu, and Pusheng Zhang. Detecting graph-based spatial outliers: Algorithms and applications (a summary of results). In *Proceedings of the Seventh ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, KDD '01, pages 371–376, New York, NY, USA, 2001. ACM.

[243] Yanning Shen, Panagiotis A. Traganitis, and Georgios B. Giannakis. Nonlinear dimensionality reduction on graphs. In *2017 IEEE 7th International Workshop on Computational Advances in Multi-Sensor Adaptive Processing (CAMSAP)*, pages 1–5, 2017.

[244] David I. Shuman, Sunil K. Narang, Pascal Frossard, Antonio Ortega, and Pierre Vandergheynst. 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, May 2013.

[245] David I. Shuman, Benjamin Ricaud, and Pierre Vandergheynst. Vertex-frequency analysis on graphs. *Applied and Computational Harmonic Analysis*, 40(2):260 – 291, 2016.

[246] Ioannis Simaiakis and Hamsa Balakrishnan. Impact of congestion on taxi times, fuel burn, and emissions at major airports. *Transportation Research Record*, 2184(1):22–30, 2010.

[247] Ioannis Simaiakis and Hamsa Balakrishnan. A queuing model of the airport departure process. *Transportation Science*, 50(1):94–109, 2016.

[248] Ioannis Simaiakis, Harshad Khadilkar, Hamsa Balakrishnan, Tom G. Reynolds, and R. John Hansman. Demonstration of reduced airport congestion through pushback rate control. *Transportation Research Part A: Policy and Practice*, 66:251 – 267, 2014.

[249] Jonas Sjöberg, Qinghua Zhang, Lennart Ljung, Albert Benveniste, Bernard Deylon, Pierre-Yves Glorennec, Håkan Hjalmarsson, and Anatoli Juditsky. *Nonlinear black-box modeling in system identification: a unified overview*. Linköping University, 1995.

[250] Abe Sklar. Fonctions de répartition à n dimensions et leurs marges. *Publ. Inst. Statist. Univ. Paris*, 8:229–231, 1959.

[251] Michael Stanley Smith and Worapree Maneesoonthorn. Inversion copulas from nonlinear state space models with an application to inflation forecasting. *International Journal of Forecasting*, 34(3):389–407, 2018.

[252] Michael F. Squires, Jay H. Lawrimore, Richard R. Heim, David A. Robinson, Mathieu R. Gerbush, and Thomas W. Estilow. The regional snowfall index. *Bulletin of the American Meteorological Society*, 95(12):1835–1848, 2014.

[253] Banavar Sridhar and P.K. Menon. Comparison of linear dynamic models for air traffic flow management. *IFAC Proceedings Volumes*, 38(1):13 – 18, 2005. 16th IFAC World Congress.

[254] Banavar Sridhar, Tarun Soni, Kapil Sheth, and Gano Chatterji. Aggregate flow model for air-traffic management. *Journal of Guidance, Control, and Dynamics*, 29(4):992–997, 2006.

[255] Alice Sternberg, Diego Carvalho, Leonardo Murta, Jorge Soares, and Eduardo Ogasawara. An analysis of Brazilian flight delays based on frequent patterns. *Transportation Research Part E: Logistics and Transportation Review*, 95:282–298, 2016.

[256] Daniel Suh and Megan S. Ryerson. A large neighborhood search heuristic to establish an optimal ad-hoc hubbing strategy in the wake of a large-scale airport outage. *Journal of Air Transport Management*, 65:156 – 165, 2017.

[257] Dengfeng Sun and Alexandre M. Bayen. Multicommodity Eulerian-Lagrangian large-capacity cell transmission model for en route traffic. *Journal of Guidance, Control, and Dynamics*, 31(3):616–628, 2008.

[258] Miao Sun, Elvin Isufi, Natasja M. S. de Groot, and Richard C. Hendriks. A graph signal processing framework for atrial activity extraction. In *2019 27th European Signal Processing Conference (EUSIPCO)*, pages 1–5, Sep. 2019.

[259] Noboru Takeichi, Ryosuke Kaida, Akihide Shimomura, and Takahiro Yamauchi. Prediction of delay due to air traffic control by machine learning. In *AIAA Modeling and Simulation Technologies Conference*, January 2017.

[260] Georg Theis, Thomas Adler, John-Paul Clarke, and Moshe Ben-Akiva. Risk aversion to short connections in airline itinerary choice. *Transportation Research Record*, 1951(1):28–36, 2006.

[261] Balasubramanian Thiagarajan, Lakshminarasimhan Srinivasan, Aditya Vikram Sharma, Dinesh Sreekanthan, and Vineeth Vijayaraghavan. A machine learning approach for prediction of on-time performance of flights. In *2017 IEEE/AIAA 36th Digital Avionics Systems Conference (DASC)*, pages 1–6, 2017.

[262] Ruey S. Tsay, Daniel Peña, and Alan E. Pankratz. Outliers in multivariate time series. *Biometrika*, 87(4):789–804, 12 2000.

[263] Visual Capitalist. WATS World Air Transport Statistics 2019. `https://www.iata.org/contentassets/a686ff624550453e8bf0c9b3f7f0ab26/wats-2019-mediakit.pdf`, 2019.

[264] Visual Capitalist. USA Airport Delay Statistics. `https://www.visualcapitalist.com/visualizing-the-u-s-airports-with-the-worst-flight-delays/`, 2020. Accessed: 2020-05-14.

[265] Lechen Wang, Xuechun Li, and Jianfeng Mao. Integrating ARIMA and Bidirectional LSTM to Predict ETA in Multi-Airport Systems. In *2020 Integrated Communications Navigation and Surveillance Conference (ICNS)*, Sep. 2020.

[266] Shitong Wang and Vikrant Vaze. Modeling probability distributions of primary delays in the national air transportation system. *Transportation Research Record*, 2569(1):42–52, 2016.

[267] Xiaohan Wang, Pengfei Liu, and Yuantao Gu. Local-set-based graph signal reconstruction. *IEEE Transactions on Signal Processing*, 63(9):2432–2444, 2015.

[268] Yanjun Wang, Jianming Zhan, Xinhua Xu, Lishuai Li, Ping Chen, and Mark Hansen. Measuring the resilience of an airport network. *Chinese Journal of Aeronautics*, 2019.

[269] Yanjun Wang, Ying Zhou, Mark Hansen, and Christopher Chin. Scheduled block time setting and on-time performance of U.S. and Chinese airlines—A comparative analysis. *Transportation Research Part A: Policy and Practice*, 130:825 – 843, 2019.

[270] Yao Wang. Prediction of weather impacted airport capacity using RUC-2 forecast. In *2012 IEEE/AIAA 31st Digital Avionics Systems Conference (DASC)*, pages 3C3–1–3C3–12, Oct 2012.

[271] Zhengyi Wang, Man Liang, and Daniel Delahaye. A hybrid machine learning model for short-term estimated time of arrival prediction in terminal manoeuvring area. *Transportation Research Part C: Emerging Technologies*, 95:280 – 294, 2018.

[272] Prashan Wanigasekara. *Latent state space models for prediction*. PhD thesis, Massachusetts Institute of Technology, 2016.

[273] Duncan J. Watts. A simple model of global cascades on random networks. *Proceedings of the National Academy of Sciences*, 99(9):5766–5771, 2002.

[274] Jessica Weinkle, Chris Landsea, Douglas Collins, Rade Musulin, Ryan P. Crompton, Philip J. Klotzbach, and Roger Pielke. Normalized hurricane damage in the continental United States 1900-2017. *Nature Sustainability*, 1(12):808–813, 2018.

[275] Andrew G. Wilson and Zoubin Ghahramani. Copula processes. In *Advances in Neural Information Processing Systems 23*, pages 2460–2468, 2010.

[276] Michael D. Wittman and William S. Swelbar. Capacity discipline and the consolidation of airport connectivity in the United States. *Transportation Research Record*, 2449(1):72–78, 2014.

[277] Weiwei Wu, Haoyu Zhang, Tao Feng, and Frank Witlox. A network modelling approach to flight delay propagation: Some empirical evidence from China. *Sustainability*, 11(16), 2019.

[278] Jing Xiong. *Revealed preference of airlines' behavior under air traffic management initiatives*. PhD thesis, University of California, Berkeley, 2010.

[279] Yan Xu, Xavier Prats, and Daniel Delahaye. Synchronised demand-capacity balancing in collaborative air traffic flow management. *Transportation Research Part C: Emerging Technologies*, 114:359–376, 2020.

[280] Shuicheng Yan, Dong Xu, Benyu Zhang, Hong-jiang Zhang, Qiang Yang, and Stephen Lin. Graph embedding and extensions: A general framework for dimensionality reduction. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 29(1):40–51, 2007.

[281] Yang Yang, Dong Li, and Xiang (Robert) Li. Public transport connectivity and intercity tourist flows. *Journal of Travel Research*, 58(1):25–41, 2019.

[282] David Yanofsky. Tech glitches keep plaguing US airlines. This dashboard kept track of them all., 2015.

[283] Maryam Farshchian Yazdi, Seyed Reza Kamel, Seyyed Javad Mahdavi Chabok, and Maryam Kheirabadi. Flight delay prediction based on deep learning and Levenberg-Marquart algorithm. *Journal of Big Data*, 7(1):106, Nov 2020.

[284] Bin Yu, Zhen Guo, Sobhan Asian, Huaizhu Wang, and Gang Chen. Flight delay prediction for commercial air transport: A deep learning approach. *Transportation Research Part E: Logistics and Transportation Review*, 125:203 – 221, 2019.

[285] Zhengzhong Yuan, Chen Zhao, Zengru Di, Wen-Xu Wang, and Ying-Cheng Lai. Exact controllability of complex networks. *Nature Communications*, 4(1):2447, 2013.

[286] Massimiliano Zanin and Fabrizio Lillo. Modelling the air transport with complex networks: A short review. *The European Physical Journal Special Topics*, 215(1):5–21, Jan 2013.

[287] Jun Zhang, Xian-Bin Cao, Wen-Bo Du, and Kai-Quan Cai. Evolution of Chinese airport network. *Physica A: Statistical Mechanics and its Applications*, 389(18):3922 – 3931, 2010.

[288] Mingyuan Zhang, Boyuan Liang, Sheng Wang, Matjaž Perc, Wenbo Du, and Xianbin Cao. Analysis of flight conflicts in the Chinese air route network. *Chaos, Solitons & Fractals*, 112:97 – 102, 2018.

[289] Mingyuan Zhang, Xuting Zhou, Yu Zhang, Lijun Sun, Ming Dun, Wenbo Du, and Xianbin Cao. Propagation index on airport delays. *Transportation Research Record*, 2673(8):536–543, 2019.

[290] Guodong Zhu, Chris Matthews, Peng Wei, Matt Lorch, and Subhashish Chakravarty. En route flight time prediction under convective weather events. In *2018 Aviation Technology, Integration, and Operations Conference*, June 2018.

[291] Álvaro Rodríguez-Sanz, Fernando Gómez Comendador, Rosa Arnaldo Valdés, Javier Pérez-Castán, Rocío Barragán Montes, and Sergio Cámara Serrano. Assessment of airport arrival congestion and delay: Prediction and reliability. *Transportation Research Part C: Emerging Technologies*, 98:255–283, 2019.

[292] Álvaro Rodríguez-Sanz, Fernando Gómez Comendador, Rosa Arnaldo Valdés, and Javier A. Pérez-Castán. Characterization and prediction of the airport operational saturation. *Journal of Air Transport Management*, 69:147–172, 2018.