論文 / 著書情報
Article / Book Information

題目(和文)	
Title(English)	A Multifaceted Data-driven Analysis of Flight Departure Delays in the Japanese Domestic Aviation Market
著者(和文)	SADEEKSOUMIK NAFIS
Author(English)	Soumik Sadeek
出典(和文)	学位:博士(学術), 学位授与機関:東京科学大学, 報告番号:甲第512号, 授与年月日:2025年9月22日, 学位の種別:課程博士, 審査員:花岡 伸也,山下 幸彦,髙田 潤一,室町 泰徳,髙山 雄貴,杉下 佳辰
Citation(English)	Degree:Doctor (Academic), Conferring organization: Institute of Science Tokyo, Report number:甲第512号, Conferred date:2025/9/22, Degree Type:Course doctor, Examiner:,,,,,
学位種別(和文)	博士論文
Type(English)	Doctoral Thesis

A Multifaceted Data-driven Analysis of Flight Departure Delays in the Japanese Domestic Aviation Market

A Dissertation
Submitted to the Department of Transdisciplinary Science and Engineering
in Partial Fulfilment of the Requirements of the Degree of
Doctor of Philosophy

Soumik Nafis Sadeek

Graduate Major in Global Engineering for Development, Environment and Society
Department of Transdisciplinary Science and Engineering
School of Environment and Society
Institute of Science Tokyo

September 2025

ABSTRACT

Flight departure delays pose significant operational and economic challenges in the aviation industry. In Japan's domestic market where two full-service carriers: All Nippon Airways (ANA) and Japan Airlines (JAL) dominate (more than 70% domestic flights), these challenges are compounded by dense networks, tightly coordinated schedules, and the need for punctuality. This thesis presents a multifaceted, data-driven analysis of departure delays by integrating network science, machine learning, conformal prediction, and topological data analysis (TDA). Using flight data from the Official Aviation Guide (OAG) between 2018 and 2021, the study focuses on the temporal, structural, and predictive dimensions of delay propagation at major Japanese airports.

The dataset includes all ANA and JAL domestic flights across 36 and 47 airports, respectively including eight key hub airports - Haneda (HND), Narita (NRT), Itami (ITM), Kansai (KIX), Chubu Centrair (NGO), Fukuoka (FUK), New Chitose (CTS), and Naha (OKA). Each day's data is used to build a dynamic delay network where airports are nodes and flights with delays of at least one minute form the edges. These networks serve as the foundation for the three chapters.

Chapter 3 investigates the association between network properties and departure delay duration. Using Prais–Winsten regression with fixed effects and panel-corrected standard errors, the study models delay duration accounting for serial correlation and heteroskedasticity in the panel data. Four network metrics are examined: in-degree, betweenness centrality (BC), eigenvector centrality (EC), and transitivity. Results show that in-degree centrality is consistently associated with longer delays suggesting that airports receiving more delayed flights are more vulnerable to accumulating further delays. BC has mixed effects-hubs may absorb delays more efficiently while high-BC spoke airports can act as bridges for delay propagation. EC generally correlates with shorter delays likely due to stronger integration within well-performing subnetworks. Transitivity exhibits differing effects across carriers: it often reduces delays for ANA but may exacerbate clustered delays for JAL. In 2020, despite fewer flights, some airports experienced longer delays highlighting how reduced resources and procedural constraints can still induce inefficiency. The findings suggest that network metrics can support proactive delay management to implement trajectory-based air traffic operations.

Chapter 4 addresses the need for uncertainty quantification in delay forecasting. While most studies focus on point predictions, this chapter introduces conformal prediction (CP) to produce prediction intervals for average daily departure delays. Three adaptive CP methods are tested: Adaptive Conformal Inference (ACI), Aggregate Adaptive Conformal Inference (AgACI) and Fully Adaptive Conformal Inference (FACI). These models are built using features such as lagged delay, network metrics, seasonal variables,

and state-of-emergency indicators. Random Forest algorithm is used for feature importance consistently identifying lagged delay as the most important predictor underscoring the temporal dependence of delay patterns. At high coverage levels (80%), all models struggle to widen prediction intervals fast enough during delay surges leading to under-coverage. AgACI performs best at moderate coverage levels (50%) by aggregating learning rates and adjusting interval widths adaptively. FACI achieves a balance between precision and responsiveness through expert reweighting based on pinball loss. These findings allow stakeholders to choose forecasting tools that match operational needs: ANA's strategy may prioritize wider intervals for reliability while JAL's narrower intervals emphasize precision. Thus, CP offers a flexible, interpretable, and adaptive framework for delay forecasting under uncertainty.

Chapter 5 introduces Topological Data Analysis (TDA) to capture higher-order patterns of delay propagation not visible through conventional network metrics. The study constructs daily undirected, weighted delay networks and applies Vietoris-Rips complexes to analyze two types of topological features: H_0 (connected components) and H_1 (delay loops). Before COVID-19, ANA's networks exhibited more persistent and broader delay loops whereas JAL's were more fragmented. During the pandemic, delay connectivity collapsed for both carriers as service was suspended leading to sharp drops in connected components and feedback loops. By 2021, particularly in winter, high-persistence loops (lasting $\geq$20 minutes) reemerged especially for ANA. These loops were more triangular in ANA's case while JAL featured more complex four- and five-airport loops. Airports such as HND, OKA, NGO, FUK, and NRT consistently appeared in high-persistence loops identifying them as critical nodes for delay control. Persistence distributions also showed seasonal effects, with longer-lived loops more common in winter. Therefore, TDA provides a powerful lens to detect persistent delay structures, offering operational insights for preemptive buffer allocation and network-wide interventions.

In conclusion, this thesis provides a multidimensional perspective on departure delays in Japan's domestic aviation market. Chapter 3 highlights how network positioning influences delay duration. Chapter 4 introduces adaptive methods for quantifying forecast uncertainty using conformal prediction. Chapter 5 reveals hidden loop/loop structures using topological tools. Together, these contributions offer practical strategies for airlines and air traffic managers to control delays, allocate resources efficiently, and respond dynamically to operational disruptions. The novel integration of CP and TDA into aviation analytics also opens new pathways for delay management in the air transport systems.

ACKNOWLEDGEMENTS

The very first round of deep gratitude goes to my supervisor, Professor Dr. Shinya Hanaoka, for his excellent style of supervision. Especially, Sensei's availability for discussions (even in short notice!) about various thesis directions made life easier when things had gone wrong in an unpleasant way. His ability to pick up even the most microscopic errors and to provide insightful suggestions has significantly improved this work. His most valuable contribution goes beyond the thesis itself: the way he advised me think critically about a research topic will remain with me in the rest of my life. In addition, his flexibility in approving my topic selection and in the modeling methods themselves was invaluable. I abruptly (though planned!) changed my master's topic from supply chain to air transport operation and management during my doctoral studies, and he supported that transition. Likewise, when I chose to apply topological data analysis - a difficult concept and very new in the engineering field - Sensei not only supported my decision but also took the time to understand what I actually wanted to do with it using the big data of aviation delay. So, thank you, Sensei!

Another important thing I should mention is that Hanaoka Sensei is still curious to learn new things. His continuous support and knowledge sharing during my PhD journey, along with his ability to see things and translate them into possible policymaking perspectives or macroscopic viewpoints, and then cluster them together, was the most amazing quality I tried to mimic (but have not achieved yet!).

Then, a big thanks goes to my co-supervisor, Assistant Professor Dr. Kashin Sugishita. His reading seminar on network science in 2021 actually shifted my mindset from purely statistical thinking toward a network-based perspective. He always supported my ideas and took the time to understand them. During paper revisions, Sugishita Sensei gave feedback (often tough!), often in minute detail, that greatly enhanced the quality of the thesis contents. I still have a long way to go to fully grasp classical network science and higher-order network science. Thank you, Sensei, for your continuous support.

I am deeply grateful to my dissertation review committee members - Prof. Yukihiko Yamashita, Prof. Jun-ichi Takada, Prof. Yasunori Muromachi, and Prof. Yuki Takayama - for their invaluable comments, feedback, questions, and constructive criticism. Their guidance has significantly improved the quality of this dissertation and helped make some of its complex concepts more accessible to readers of both this work and my related academic publications.

The Japanese Government (Monbukagakusho, or the Ministry of Education, Culture, Sports, Science and Technology: MEXT) provided me with the opportunity to pursue this research, first supporting my master's program in 2020 and later extending the support to the doctoral program in 2022 at the Institute

of Science Tokyo (formerly, Tokyo Institute of Technology) through enrollment in the Integrated Doctoral Education Program at IGP-A. Dr. Reza and Dr. Hiraide made my life easier at the beginning with their valuable academic advice, encouraging words, and by facilitating my smooth adaptation to the lab environment.

Dr. Sunkyung Choi at Gunma University has always been a source of encouragement. Our discussions on topics such as mobility, disaster management, emergency evacuation, gender equality in evacuation shelters, big data, and text mining were the source of many brainstorming sessions.

Continuing to the three years of PhD journey would not have been possible without the lab life ensured by my friends at Hanaoka Lab. Dr. Liu, Dr. Tagawa, Fahim (Rakuten), Gutt, Tokuda, and Ms. Hattori deserve special mention in this regard. Apart from the lab members, Nihal, Prattoyee, Shahriar, and Nusrat ensured that I have a social life and were all there during my challenging times living in Tokyo without family, especially our mingle at Ma-Mai restaurant.

My wife, Afrida Bari Asha, deserves special mention for two reasons. This thesis could not have been completed if she had not encouraged and insisted me that I must not quit my PhD midway during periods of depression and frustration. Additionally, her continuous support to both of our families in Bangladesh during my absence was invaluable.

My gratitude also extends to my parents, Mr. Mahbubul Alam and Mrs. Ummul Imam Sokaina, and my in-laws, Mr. Emdadul Bari and Mrs. Halima Bari. Lastly, but not the least, my younger brother, Amiya Afwan Jarrah, who has been beside me since 2022, has stood by me through my highs and lows and has always been like a guardian throughout this entire period in Japan. I can’t thank them enough!

DEDICATION

This dissertation is dedicated to my late grandfather, **Md. Shamsul Alam** (1940–2023). I miss him deeply, as he was always the most enthusiastic person to hear about my graduate studies at Tokyo Tech on the MEXT scholarship. If he were present, he would have been the happiest person.

Table of Contents

List of Tables

List of Figures

List of Acronyms

ACI	Adaptive Conformal Inference
ACRP	Airport Cooperative Research Program
AgACI	Aggregate Adaptive Conformal Inference
AIAA	American Institute of Aeronautics and Astronautics
ANA	All Nippon Airways
ANN	Artificial Neural Network
ANOVA	Analysis of Variance
AR	Autoregressive (model)
ATFM	Air Traffic Flow Management
ATM	Air Traffic Management
ATMC	Air Traffic Management Center
BC	Betweenness Centrality (network metric)
BOA	Bernstein Online Aggregation
CAAC	Civil Aviation Administration of China
CANSO	Civil Air Navigation Services Organization
CARATS	Collaborative Actions for Renovation of Air Traffic Systems (Japan)
CC	Closeness Centrality (network metric)
CNN	Convolutional Neural Network
CP	Conformal Prediction (uncertainty quantification method)
DBSCAN	Density-Based Spatial Clustering of Applications with Noise (clustering algorithm)
DCNN	Deep Convolutional Neural Network
DEA	Data Envelopment Analysis
EC	Eigenvector Centrality (network metric)
FAA	Federal Aviation Administration
FACI	Full Adaptive Conformal Inference
FGLS	Feasible Generalized Least Squares
GCN	Graph Convolutional Network
GDP	Ground Delay Programs
GLS	Generalized Least Squares
IATA	International Air Transport Association
ICAO	International Civil Aviation Organization

i.i.d	Independent and Identically Distributed
JAL	Japan Airlines
LFR	Loop Formation Rate
LSTM	Long Short-Term Memory
MAE	Mean Absolute Error
MLIT	Ministry of Land, Infrastructure, Transport and Tourism (Japan)
NASEM	National Academies of Sciences, Engineering, and Medicine (U.S.)
OAG	Official Airline Guide
PBTO	Proportion of Birth Time Overlap
PCSE	Panel-Corrected Standard Errors
PW	Prais-Winsten
RF	Random Forest
RMSE	Root Mean Square Error
SD	Standard Deviation
SE	Standard Error
SOE	State of Emergency (COVID-19 context in Japan)
SVM	Support Vector Machine
TDA	Topological Data Analysis
VIF	Variance Inflation Factor (multicollinearity diagnostic)
WHO	World Health Organization

Glossary

All Nippon Airways	Japan's largest full-service airline, operating hub-and-spoke domestic networks
Airborne Holding	A delay management procedure where aircraft enter a holding pattern at a designated fix in the airspace.
Autoregressive model	A time-series model where the present value depends on its past values.
Air Traffic Flow Management	The process of balancing air traffic demand with available capacity at airports and airspace sectors. Delay management strategies like Ground Delay Programs (GDP) are part of ATFM
Bernstein Online Aggregation	An online learning algorithm that adaptively combines predictors using Bernstein inequalities, ensuring efficient sequential updates and valid uncertainty quantification. Used in modern conformal prediction frameworks.
Boundary Matrix	A matrix that encodes the relationships between simplices in a simplicial complex by specifying how higher-dimensional simplices are built from lower-dimensional ones.
Collaborative Actions for Renovation of Air Traffic System	A Japanese government-led modernization plan. Aims to improve efficiency and delay resilience in Japan's airspace system.
Conformal Prediction	A statistical framework for producing valid prediction intervals without distributional assumptions
Delay propagation	The process by which a delay at one airport spreads through the network, either within the same aircraft's daily schedule or to connected flights at other airports
Departure Delay	Departure delay is defined as the difference between a flight's actual departure time and its scheduled departure time.
Exchangeability	A statistical assumption weaker than *i.i.d.*, meaning that the probability distribution of data is invariant to the order of observations. Conformal prediction methods rely on exchangeability to provide valid uncertainty intervals.

Functoriality	A property from category theory stating that homology is a functor, meaning it preserves structure under mappings between spaces. Without functoriality, persistent homology would not consistently track features across thresholds, and persistence diagrams or barcodes would not be reliable.
Gate Departure Delay	Delay that occurs before pushback, usually caused by late boarding, aircraft turnaround issues, missing crew, or maintenance checks. Part of departure delays.
Ground Delay Program	FAA based traffic management process when aircraft are held on the ground in order to manage capacity and demand at a specific location, by assigning arrival slots.
H_0 Features	Connected Components: Represent isolated airport clusters before merging with the other airports
H_1 Features	Loops: Represent delay loops circulating among multiple airports.
Hub-and-Spoke Network	A network structure in which flights are routed through central airports (hubs) that also serve as transfer points, connecting to a series of secondary or smaller airports (spokes).
Independent and Identically Distributed Variables (i.i.d)	A standard statistical assumption that each data point is drawn independently from the same probability distribution.
Japan Airlines	Japan's second-largest airline. Together with ANA, it accounts for over 70% of domestic market share.
Kronecker delta (δ_{ij})	A mathematical function equal to 1 when indices are equal (i = j) and 0 otherwise. Used in Prais-Winsten regression error covariance structure
Loop Formation Rate	A TDA-derived metric quantifying how frequently delay loops are formed within a given period. Used to capture loop-based delay propagation (for directed network) and(or) delay co-occurrence (for undirected network).
Mean Absolute Error	Evaluation metric measuring average magnitude of forecast errors.
Miles-in-Trail	A type of traffic management initiative where air traffic controllers require a minimum number of nautical miles

	between aircraft flying the same route or into the same sector.
Minutes-in-Trail	Similar to Miles-in-Trail, but separation is measured in time (minutes) instead of distance. Aircrafts are required to cross a fix, waypoint, or sector boundary at a specified time interval apart.
Persistent Homology	A method in TDA that tracks how topological features (clusters, loops, voids) emerge and persist across scales in data.
Proportion of Birth Time Overlap	TDA metric measuring the overlap in timing when loops first appear across networks. Higher PBTO value indicates systemic vulnerabilities inside the network.
Panel-Corrected Standard Errors	Robust error estimators for panel data regressions, accounting for contemporaneous correlation across panels.
Persistence	The duration between birth time and death time of a feature (H_0 or H_1). High persistence indicates structural significance of delay features.
Persistence diagram / barcode	Visualizations plotting birth-death lifespans of topological features.
Prais-Winsten regression	An econometric technique correcting for first-order serial correlation in panel regression models.
Prediction interval coverage	The probability that a prediction interval contains the true outcome. Targeted through CP to ensure reliability of delay forecasts.
Pinball loss	A loss function for quantile regression; used in evaluating CP intervals.
Root Mean Square Error	Forecast accuracy metric emphasizing large errors.
State of Emergency	COVID-19 restrictions declared in Japan during 2020-2021
Topological Data Analysis - TDA	Mathematical framework for studying the *shape* of data using algebraic topology. Here, TDA is applied to detect delay loops and persistence across airline networks.
Vietoris-Rips complex	A simplicial complex construction from pairwise distances, central to building persistence diagrams in TDA.
Variance Inflation Factor	Statistic used to detect multicollinearity in regression models.

Chapter 1 Introduction

1.1 Background and Significance

Flight delays are a pervasive operational challenge in the aviation industry, imposing substantial costs and causing negatively the reliability of air transport networks. They are formally defined as deviations between actual and scheduled takeoff or landing times, and they serve as key performance indicators reflecting constraints in airport capacity and inefficiencies in operational practices. Globally, roughly one in four commercial flights departs or arrives late, leading to significant disruptions for passengers and substantial financial losses for airlines. For example, in the United States, flight delays in 2019 were estimated to have cost airlines and passengers approximately $33-40 billion in combined expenses and lost productivity. Similarly, in China, the majority of delays over 60% are not attributable to exogenous causes such as weather, but rather to cascading propagation effects that transform localized disruptions into system-wide problems. One estimate calculated that such delays in China induced indirect economic losses approaching $50 billion USD in a single year that underscores the scale of the issue and the need for effective mitigation strategies (Sadeek et al., 2025; Yang et al., 2023; ACRP, 2014; Ye et al., 2020).

Empirical evidence from multiple regions demonstrates that delays are both widespread and costly. Data indicate that 24% of European flights experience delays exceeding five minutes, while in Japan, over 69,000 flights in 2022 were delayed by more than 15 minutes (Statista, 2023). Even in markets renowned for punctuality, such as Japan, delays remain a significant operational concern. Beyond their direct costs, delays reduce network resilience and create vulnerabilities to cascading disruptions. In modern air transport systems, aircraft are scheduled across multiple daily segments, and a delay on one flight often precipitates a sequence of subsequent delays within the airline's schedule, a process described as vertical propagation. In parallel, airports act as convergence nodes for arriving and departing flights. When congestion arises at a major hub, it can trigger horizontal propagation that disrupts unrelated flights queued for departure. This dual mechanism means that dense flight schedules and near-capacity airports have limited buffer to absorb deviations, rendering them particularly susceptible to chain reactions in which localized perturbations escalate into systemic congestion.

The Japanese domestic market provides a salient region for analyzing these dynamics. As the world's fifth-largest domestic air travel market, Japan is characterized by high concentration among two major full-service carriers namely: All Nippon Airways (ANA) and Japan Airlines (JAL) which collectively account for over 70% of domestic capacity (Ng et al., 2022). Both airlines maintain extensive networks of routes spanning more than 50 airports, anchored by several major hubs including Tokyo Haneda, Osaka

Itami, and Sapporo New Chitose. While these networks achieve high operational efficiency in normal conditions, their tight schedules and dense flight increase susceptibility to delay amplification particularly at slot-constrained airports such as Haneda. For instance, in 2019, ANA's on-time arrival rate was approximately 86%, while JAL's was 84%. However, by 2021, during the COVID-19 pandemic, both carriers recorded on-time arrival rates exceeding 93%, illustrating that reduced traffic volumes temporarily alleviated congestion and improved punctuality (Cirium, 2019; 2021). Nevertheless, the pandemic also revealed that certain airports continued to experience severe delays despite lower flight activity, as resource constraints, public health measures, and procedural adaptations introduced new inefficiencies (Sadeek et al., 2025). Comparative data (Figure 1.1) indicate that both departure and arrival delays were markedly higher pre-COVID, with significant reductions observed in the pandemic period. For example, pre-COVID departure delays affected 47% of flights, compared to 23% post-COVID.

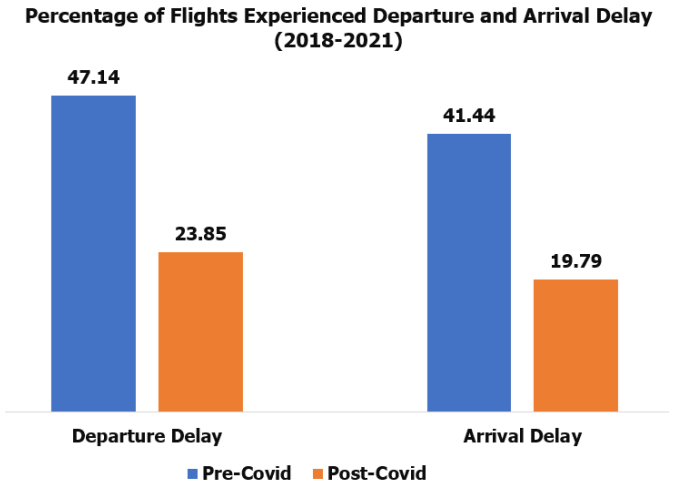


Figure 1.1: Percentage of flights experienced departure and arrival delay from 2018-2021 (Made by the author)

Understanding how delays accumulate requires recognizing their multi-causal origins as shown in Figure 1.2. At a system level, delays emerge from the interplay of four principal domains: air traffic control and management performance (encompassing airspace design and investment in navigation infrastructure), capacity and airport infrastructure constraints, airline scheduling practices and operational preferences, and weather phenomena such as low visibility and high winds. These drivers interact dynamically, often compounding one another over time. From an operational standpoint, delay can be decomposed into distinct phases: gate delay prior to departure, taxi-out delay while queuing for takeoff, airborne delay during flight, and taxi-in delay upon landing. Variability in any of these phases contributes to the cumulative block delay measured at arrival. This holistic perspective underscores that effective mitigation

requires interventions targeting multiple stages of the operational timeline rather than isolated improvements.

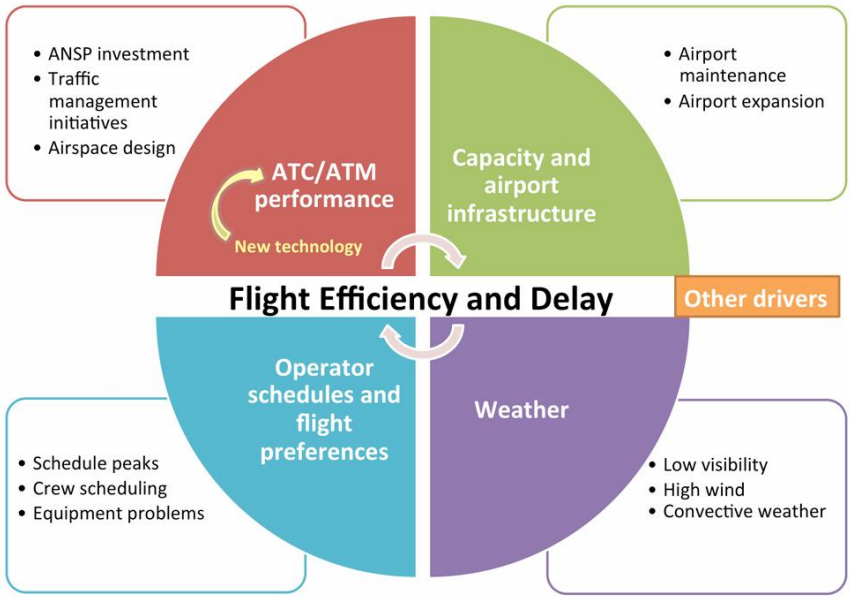


Figure 1.2: System interdependencies that influence flight delay (CANSO, 2020)

A detailed decomposition of operational variability clarifies how delays accumulate across each phase of flights. Scheduled departure is followed by the taxi-out phase, airborne phase, and taxi-in phase before arrival at the gate. Deviations in any segment such as - gate delay, taxi-out delay, airborne delay, or taxi-in delay - accumulate to produce the block delay, which defines overall lateness (Figure 1.2). This schematic makes explicit that effective mitigation strategies must address delay contributors holistically across the entire operational timeline that can start with departure delay at the gate. Focusing on departure delays is analytically advantageous because they are more likely to propagate as domino effects across airports, provide clearer attribution to the origin of disruptions, and offer predictive value for subsequent arrival delays (Zheng et al., 2021)

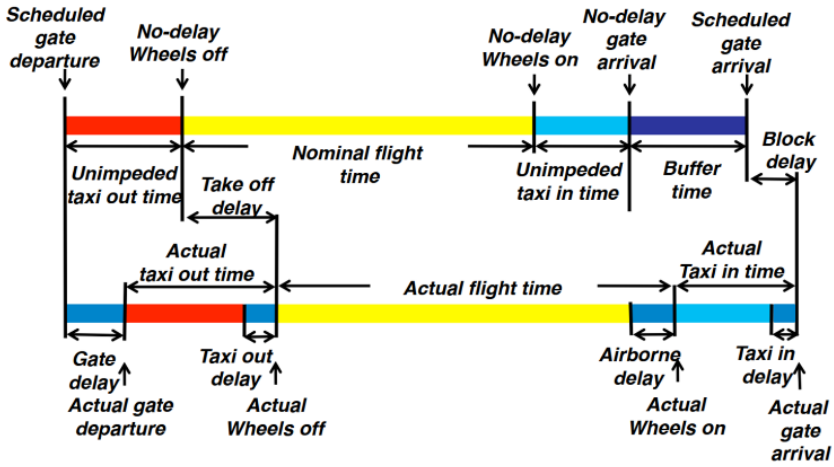


Figure 1.3: Operational variability and flight delays (Balakrishnan, 2015)

Conceptually, flights can be understood as components of an interconnected network that links airports through both scheduled services (the flight network) and the pathways along which delays propagate (the delay network). Traditional tabular data structures often fail to capture these relational dependencies. By contrast, network-based representations enable analysts to identify critical nodes with disproportionate influence, uncover recurrent patterns of delay transmission, and query system-wide interdependencies. This approach is particularly valuable in complex, high-utilization settings like Japan's domestic aviation sector, where local disruptions can spread rapidly across hubs and affect a large share of total traffic (ACRP, 2014; Bechberger and Perryman, 2020).

Motivated by these considerations, the present study aims to develop a data-driven framework capable of characterizing, forecasting, and diagnosing vulnerabilities in Japan's domestic departure-delay network. This framework integrates econometric modeling of network positions, conformal prediction methods to quantify uncertainty, and topological data analysis to detect persistent delay loops and critical airport sets. Through this multi-method approach, the research seeks to study how delay propagation emerges from systemic interdependencies and to inform strategies for improving resilience and operational reliability as air traffic continues to recover to pre-pandemic levels.

1.2 Problem Statement and Research Gaps

Ensuring on-time performance in a complex network like Japan's domestic air system involves more than addressing isolated flights – it requires system-wide insight into how delays originate, propagate, and can be predicted. Traditional studies of flight delays often focus on individual flight leg attributes (e.g. weather, aircraft, or crew connections) or treat airports as independent units. While such approaches can

identify immediate causes (like a thunderstorm or mechanical issue), they struggle to capture the cascading nature of delays across an interconnected schedule. One major gap is the limited integration of network science into understanding flight delays. Modern air transport is inherently a networked system: airports are nodes and flights are links forming a web through which disruptions can flow. Complex network research has demonstrated, for example, that the degree (number of connections) of an airport correlates with its tendency to absorb or propagate delays. High-degree hub airports are often more prone to being affected by upstream delays simply due to the volume of flights interacting there. Similarly, network simulation studies indicate that even a small initial delay can spread and escalate in a scale-free network, reflecting the low critical threshold for contagion in air traffic systems (Yao and Zhang, 2023). These insights beg the question: Can network metrics quantitatively explain why certain airports or routes experience worse delays than others or how the delay differs on an average for each of the airports? And if so, could those metrics be used to predict or anticipate delay propagation? This thesis addresses these questions by explicitly linking graph-theoretic properties of the domestic flight network to delay duration.

Another gap lies in predictive analytics for delays, particularly in forecasting not just point estimates but the uncertainty around those estimates. Airline and airport stakeholders increasingly use machine learning models to forecast delays or on-time performance, leveraging big data on past flights, weather, and operations. Techniques ranging from regression trees and random forests to deep learning have shown promise in predicting whether a given flight will be delayed or estimating daily delay averages. However, most of these predictive models yield a single predicted value (or classification) without an indication of confidence or risk. In practice, knowing that - average departure delay at Haneda is likely to be 5 minutes tomorrow - is far less actionable than knowing it will be 5 minutes ± some range, or that there's, for example, a 10% chance delays will exceed a critical threshold. Uncertainty quantification in delay prediction is thus crucial for risk-informed decision making that the airlines could adjust schedules or standby resources if a high delay day is likely while airports could allocate extra ground staff when the risk of extreme delays is elevated. The literature on probabilistic forecasting and prediction intervals in transportation is still emerging. Some recent work has applied Bayesian deep learning approaches (e.g. Monte Carlo dropout with LSTMs) to produce prediction intervals for daily airport delays (Vandal et al., 2018) demonstrating that it is feasible to generate well-calibrated uncertainty estimates alongside point forecasts. Nonetheless, these methods can be complex and may rely on distributional assumptions or heavy computational training. A newer framework known as conformal prediction (CP) offers an attractive alternative for uncertainty quantification: CP can take any sophisticated predictive model and combine it with a procedure to output prediction intervals with a guaranteed coverage probability without strong parametric assumptions (Susmann et al., 2023). Originally developed for i.i.d. data, CP has

recently been extended to time-series and streaming data through adaptive algorithms. The application of conformal prediction to flight delays remains largely unexplored which represents a clear research gap. This thesis assume that introducing adaptive conformal methods to delay forecasting can fill the void by providing robust and interpretable prediction intervals that help operational planners measure uncertainty in delay outcomes.

A final gap addressed by this research concerns the detection of higher-order patterns in delay propagation that are not visible to traditional network metrics. Conventional network analysis focuses on direct pairwise connections. However, complex propagation phenomena such as: feedback loops where a group of airports exchange delays cyclically can avoid standard analyses. For instance, consider a scenario where delays circulate in a triangle: Airport A's delay affects B, B's affects C, and C's comes back to affect A. Such a closed loop of delays could sustain itself or recur periodically (e.g., under certain weather patterns or seasonal demand surges) without ever standing out in pairwise delay correlations. Topological Data Analysis (TDA) provides a novel toolkit to uncover these kinds of structures by treating the network as a geometric object and searching for homological features like loops (loops) and connected components. TDA, and specifically persistent homology, measures the shape of data across multiple scales (Myers et al., 2023). In a network context, 0-dimensional homology (H_0) corresponds to connected components, while 1-dimensional homology (H_1) corresponds to loops or loops in the network's connectivity. These loops can be crucial: an evolving air traffic network might remain fully connected as one component (no airport is isolated), yet it could contain multiple simultaneous delay loops indicating circulating delays among subsets of airports. By applying persistent homology to sequences of delay networks (e.g., building a simplicial complex where airports connected by delay form higher-dimensional simplices), we can detect when such loops appear and how long they persist (in terms of delay duration). Prior research in transportation analytics suggests that TDA can reveal meaningful patterns missed by classical analysis. For example, in aviation, TDA is still in a nascent stage, with initial studies demonstrating its use on airport surface movement data and other high-dimensional operational datasets (Li et al., 2019). No known study to date has applied TDA to airborne delay propagation. Thus, there is a significant opportunity to contribute new knowledge by exploring how persistent topological features - like long-lived delay loops or multi-airport delay clusters - manifest in an air transportation network and what operational insights they provide.

In summary, the key research gaps motivating this dissertation are: (1) a need for network-level understanding of delay propagation, linking network position and delay severity; (2) a need for uncertainty-aware forecasting of delays, going beyond point predictions to prediction intervals that inform

risk management; and (3) a need for methods to detect higher-order propagation patterns (like loops) that are invisible to conventional metrics. Addressing these gaps is essential for developing a holistic strategy to manage flight delays in a large, complex network such as Japan's domestic air system.

1.3 Research Objectives

To bridge the gaps identified above, this dissertation takes a multi-faceted, data-driven approach that integrates concepts from network science, machine learning, and topological data analysis. The overarching objective is to develop a holistic understanding of departure delays in Japan's domestic airline network, encompassing how delays propagate, how they can be predicted (with quantified uncertainty), and what hidden structural patterns underlie the delay dynamics. The specific research objectives are:

a. **Objective 1:** This objective examines how an airport's network position, captured by in-degree, betweenness, eigenvector centrality, and transitivity, influences the accumulation and propagation of departure delays. It tests their association with delay outcomes. It reveals whether certain centrality measures amplify or mitigate delays, for instance, whether high in-degree increases vulnerability to inbound disruption or whether high betweenness facilitates buffering. The analysis also explores how these relationships evolved during COVID-19 shocks, providing evidence that network-centric features are dynamic and stress-responsive.

b. **Objective 2:** This objective aims to produce prediction intervals rather than point estimates for average daily departure delays. We integrate adaptive conformal prediction methods (ACI, AgACI, FACI) with Random Forest models to generate intervals that reflect uncertainty under both stable and volatile conditions. Each conformal approach is compared in terms of interval sharpness and coverage reliability, highlighting trade-offs for operational decision-making. This objective showcases how uncertainty-aware tools enable risk-based planning and enhances forecast reliability to airlines and air traffic managers.

c. **Objective 3:** The third objective applies Topological Data Analysis (TDA) to detect persistent loops and vulnerable points in the delay network. This approach identifies not only connected components (H_0) but also delay loops (H_1) that reveal multi-airport cyclical delay co-occurrence often hidden in traditional metrics. We quantify which airport sets consistently form persistent loops, indicating systemic vulnerabilities requiring proactive interventions. This objective demonstrates how higher-dimensional analytics can highlight critical delay configurations and imply preemptive scheduling strategies to disrupt sustained loops.

1.4 Data and Study Scope

To achieve the above objectives, we utilize a comprehensive flight dataset from the Official Airline Guide (OAG), covering all domestic flights by ANA and JAL from 2018 through 2021. This dataset includes flight schedules and actual departure times, enabling the calculation of departure delays (actual departure time minus scheduled departure time) for every flight leg. In total, the data spans 36 airports served by the ANA Group and 47 airports by the JAL Group with overlapping coverage on all the major airports in Japan. Focusing on ANA and JAL allows us to capture the majority of domestic air traffic (about three-quarters of the market) while maintaining a consistent full-service carrier context. The eight hub airports highlighted earlier (HND, NRT, ITM, KIX, NGO, FUK, CTS, OKA) feature heavily in the analysis as they account for a large proportion of flights and often serve as origin or destination for multi-leg journeys. Each year of the dataset brings its own characteristics: 2018-2019 represent a baseline of normal, high-volume operations; 2020 reflects the drastic pandemic-induced contraction (with certain months seeing traffic reductions over 50% and the imposition of state-of-the-emergency schedules); and 2021 shows a partial recovery with ongoing pandemic effects and seasonal irregularities. By examining this full timeline, the study captures both steady-state behaviors and stress-test scenarios for delay propagation.

1.5 Structure of the Dissertation

This dissertation is organized into six chapters. Following this introduction, Chapter 2 provides an overview of relevant research in flight delay analysis, including classical studies on delay causation and propagation, network-based analyses, machine learning applications for delay prediction, and prior uses of conformal inference and topological analysis in transportation contexts. Figure 1.4 represents the basic framework of this study

Chapter 3 corresponds to Objective 1. In this chapter, we present the construction of daily delay networks and the regression analysis linking network metrics to delay durations. We detail the Prais-Winsten panel regression methodology and discuss the results for each metric (in-degree, betweenness, eigenvector centrality, transitivity), including variations by airline and by year. The chapter interprets the findings with reference to network theory and operational characteristics – for example, explaining why in-degree centrality emerges as a significant factor and how the differing transitivity effects for ANA vs. JAL might relate to their network design philosophies. The chapter concludes by suggesting how airlines or airports could incorporate network analytics into delay mitigation strategies, such as identifying high-risk inbound flows in real time.

Chapter 4 covers Objective 2. Here, we describe the predictive modeling framework including data preprocessing, feature engineering (e.g., creation of lag features and binary indicators for special events), and the configuration of the Random Forest model. We then explain the conformal prediction techniques (ACI, AgACI, FACI), with a focus on their implementation and tuning in an online forecasting setting. Results are presented for different coverage probability targets, comparing the size of prediction intervals and their empirical coverage across various conditions. We use visualization of prediction interval trajectories to illustrate how the adaptive methods react to sudden changes in delay levels (for instance, showing a plot of actual vs. predicted delays with bands that expand or contract over time). The chapter provides a discussion on choosing conformity scores (absolute error vs. quantile losses) and how this affects the interval behavior. We highlight that the conformal approach is model-agnostic - while we used Random Forest here, any predictive model (including neural networks or hybrid simulations) could be integrated in the same way making it flexible the aviation analytics.

Chapter 5 corresponds to Objective 3. This chapter delves into the application of persistent homology to the delay network data. We provide a primer on how the Vietoris-Rips complex is constructed from a weighted network and how to interpret the resulting persistence barcodes for H_0 and H_1. We here quantify the number of connected components over time relating a spike in components to the pandemic flight cuts. A significant portion of the chapter is dedicated to interpreting what specific loops mean in real terms. We identify representative examples of delay loops and trace how on those days the delays circulated. By doing so, we connect the abstract topological features back to concrete events (for instance: seasonal issues). The chapter demonstrates how TDA can act as an early warning mechanism: a persistence diagram that shows an unusually long loop forming might signal to operators that a set of airports are in a delay feedback situation.

Finally, Chapter 6 summarizes the key findings of the dissertation, reflecting on how the three analytical approach collectively contribute to an approach to managing flight delays. The chapter also acknowledges the limitations of the study to ensure a balanced interpretation of the results. We then outline directions for future research, inspired by this work. The thesis concludes with the broader vision that combining network-aware analysis, probabilistic forecasting, and topological pattern recognition can significantly enhance the robustness of air transportation.

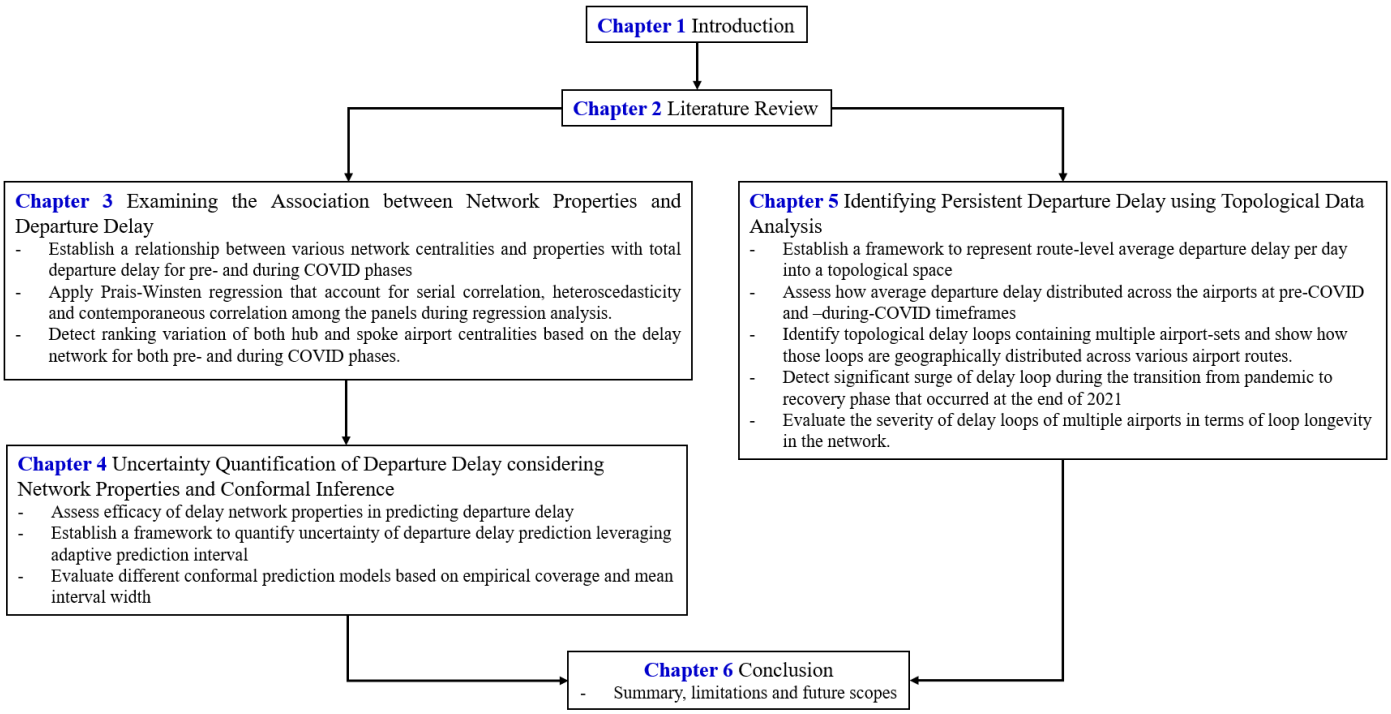


Figure 1.4: Dissertation framework of this study

Chapter 2 Literature Review

This chapter synthesizes the extensive body of literature on flight delay prediction, propagation analysis, and operational optimization. It situates the dissertation within a research landscape spanning traditional statistical approaches, machine learning, network science, causal inference, uncertainty quantification, and topological data analysis. The below Figure 4 is produced based on the detailed aviation delay review papers by Ribeiro et al. (2025), Wandelt et al. (2025) and Carvalho et al. (2020).

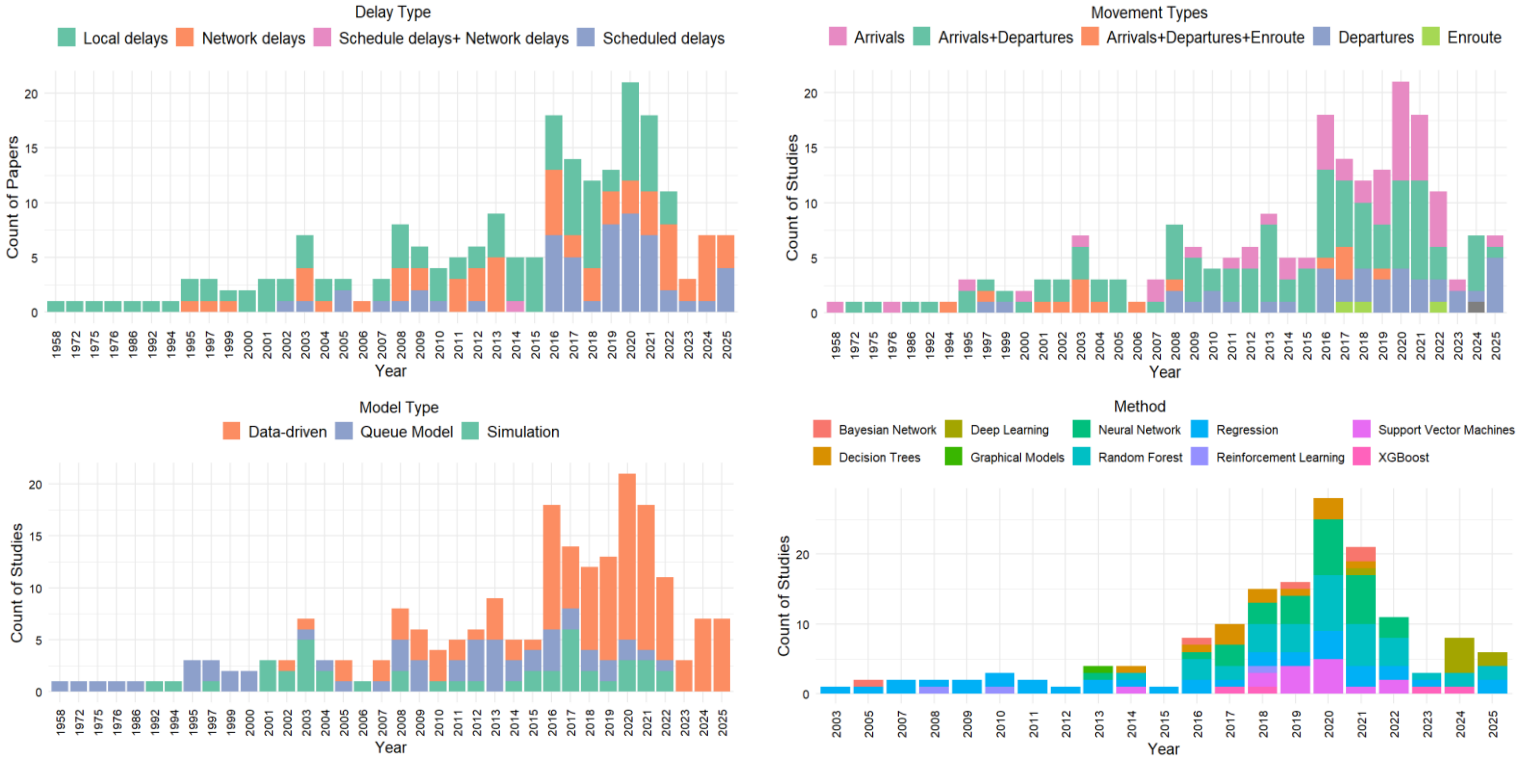


Figure 2.1: Trends of various flight delay studies along with the delay types, movement types and methodology used for data-driven delay studies

In Figure 2.1 (top left panel), it illustrates the distribution of studies by the category of delay they address such as: local delays, network delays, and scheduled delays. Early research, particularly before the 1990s, focused predominantly on local delays, such as individual gate or runway queues. Around 2000, there is an increase in the investigation of network delays reflecting a shift toward understanding systemic propagation phenomena across interconnected airports. Interest in scheduled delays, differences between planned and actual block times, rose sharply after 2005, peaking around 2015-2020 as airlines and regulators prioritized on-time performance metrics for operational benchmarking.

Model-types (top right panel), depict the types of models employed in delay analysis over time. Simulation models and queueing models dominated earlier decades reflecting an operations research orientation and an emphasis on modeling airport-level processes. However, from approximately 2010 onward, there is a rapid expansion of data-driven models which by 2020 had surpassed traditional approaches in frequency. This trend coincides with the proliferation of richer operational data, advances in computational infrastructure, and the adoption of predictive analytics within airline and air traffic management contexts.

The bottom-left panel in Figure 2.1 categorizes studies by the types of flight movements analyzed - arrivals, departures, enroute operations, and there combinations. While research initially emphasized either arrivals or departures in isolation, more recent studies increasingly adopt an integrated perspective that accounts for the interaction of arrivals, departures, and enroute phases. Notably, the volume of studies examining combined movement types accelerated substantially after 2010 indicating growing recognition of the interdependencies among different operational phases and their role in delay propagation. Lastly, the bottom-right panel highlights the range of analytical methods applied especially in the last two decades. Regression models were historically predominant and remain widely used due to their interpretability and established role in econometric analysis. However, since approximately 2010, there has been significant diversification, with growing adoption of machine learning techniques such as Random Forests, Neural Networks, Support Vector Machines, and XGBoost. Among these, Random Forests have seen particularly rapid uptake, reflecting their robustness to nonlinearity and their suitability for high-dimensional operational data. Additionally, emerging applications of reinforcement learning and Bayesian networks suggest increasing experimentation with advanced AI-driven approaches. Below we discuss how various flight delay studies those adopted data-driven approaches used various methods in their studies, mostly focusing on departure delay.

2.1 Machine Learning and Deep Learning Approaches to Delay Prediction

The prediction of flight delays has undergone considerable methodological evolution. Early research applied regression and ensemble methods. Esmaeilzadeh and Mokhtarimousavi (2020) compared Random Forests (RF), XGBoost, and artificial neural networks (ANN) on U.S. BTS data demonstrating XGBoost's superior predictive performance and identifying weather, departure time, and carrier type as key predictors. Kilic and Sallan (2023) evaluated multiple models on U.S. airport data reporting that gradient boosting machines consistently outperformed alternatives, particularly after correcting class imbalance.

Several studies advanced hybrid ensemble methods. Nivitha et al. (2023) combined XGBoost, RF, ANN, and a voting regressor, demonstrating that model fusion improved accuracy. Dai (2024) proposed a hybrid

model integrating ANOVA and FSFS feature selection, DBSCAN clustering, and Coyote Optimization–weighted RF, achieving 97.2% accuracy and reducing training time by over 39%. Mokhtarimousavi and Mehrabi (2023) developed a framework combining random-parameter logit and SVM models optimized by artificial bee colony algorithms, highlighting the relevance of carrier type and temporal factors.

Deep learning has emerged as a principal strategy for delay prediction. Yu et al. (2019) introduced a CNN–LSTM hybrid outperforming conventional machine learning techniques on large datasets. Qu et al. (2020) fused meteorological data in CNN-based architectures (DCNN and SE-DenseNet), attaining accuracies of 92–93%. Mamdouh et al. (2023) designed a BiLSTM with attention mechanisms, illustrating that incorporating arrival delays alongside departure delays enhances predictive accuracy. Bisandu and Moulitsas (2024a, 2024b) explored BiLSTM classifiers and a DeepONet optimized by a gradient-mayfly algorithm, demonstrating improvements in classification and regression tasks under uncertainty.

Graph neural networks have been increasingly applied to exploit spatiotemporal correlations. Bao et al. (2021) developed a graph-to-sequence (G2S) model integrating GCNs and attention mechanisms for multi-step prediction, outperforming LSTM and T-GCN baselines. Cai et al. (2022) demonstrated that dynamic GCNs captured evolving temporal dependencies more effectively than static architectures. Zheng et al. (2024) proposed a Graph Multi-Attention Network for learning selective spatiotemporal dependencies, while Sun et al. (2024) introduced a causality-guided GCN leveraging transfer entropy to capture influence pathways. Wu et al. (2024) used deep spatiotemporal propagation learning, integrating congestion dynamics to enhance accuracy.

Transformer architectures have further extended the modeling landscape. Liu et al. (2024) applied a Temporal Fusion Transformer to forecast 15-minute arrival delays with interpretability. Wang et al. (Book Chapter) implemented multi-airport delay prediction using transformers, capturing inter-airport dependencies.

Recent comparative analyses (Wandelt et al., 2025) have emphasized reproducibility challenges and the need for region-specific validation frameworks. Other works (Anguita and Olariaga, 2024; Khan et al., 2021; Kim and Park, 2024) confirmed that combining ensemble methods and deep learning consistently improves predictive performance, while Dalmau et al. (2023) underscored the utility of probabilistic modeling via quantile regression.

2.2 Network Science and Delay Propagation

Complex network theory has become integral to analyzing delay propagation. Wandelt et al. (2025) conducted a scoping review categorizing network types - route, airline, passenger-centric, trajectory-

based and identified deficiencies in standardization and real-time modeling. Zhixing et al. (2021) offered a taxonomy integrating multilayer networks and machine learning perspectives. Empirical studies have explored network metrics and propagation typologies. Sugishita et al. (2024) demonstrated that hubs disproportionately contribute to propagation in Japanese domestic aviation observing seasonal and spatial asymmetries. Tang et al. (2023) introduced the concepts of *delay width* and *delay strength* applying multilayer network decomposition. Li et al. (2024) constructed aviation, traffic, and delay propagation networks revealing that smaller airports are particularly vulnerable to systemic delays. Nguyen and Lin (2023) developed minimal-path reliability algorithms, quantifying resilience under budget constraints. Delay causality networks have facilitated the analysis of directional dependencies. Du et al. (2018) and Chen et al. (2024) employed nonlinear Granger causality via multilayer perceptrons to detect causal pathways, identifying large hubs as key sources of propagation. Feng et al. (2024) integrated Granger and kernel-based tests within a Systematic Path Isolation framework, capturing static and dynamic propagation. Zeng et al. (2022) applied PC-MCI causal discovery to filter spurious correlations, highlighting the role of small airports under extreme events. Kang et al. (2023) developed an attention-weighted recurrent GCN uncovering bidirectional propagation in North China.

Network-based metrics have also been linked to operational decision support. Rebollo and Balakrishnan (2014), Gopalakrishnan and Balakrishnan (2017), and Wang et al. (2020) analyzed delay networks across the U.S. and China identifying universal propagation classes and multifractal characteristics. Other studies such as: Wang et al. (2022) and Wang et al. (2023) demonstrated that weighted directed graphs and time-delay stability algorithms enable fine-grained identification of temporal dependencies.

Recent analyses (Sadeek et al., 2025) explored associations between network centrality measures and delay durations underscoring the value of centrality metrics as predictive covariates.

2.3 Uncertainty Quantification in Delay Forecasting

Probabilistic modeling of delays has attracted significant interest. Dalmau et al. (2023) demonstrated that quantile regression can produce actionable interval forecasts several days in advance. Bayesian neural networks and Monte Carlo dropout (Gibbs and Candès, 2021) have been adopted for uncertainty estimation, though these methods face computational and independence limitations.

Adaptive Conformal Inference (ACI), Aggregated ACI (AgACI), and Fully Adaptive Conformal Inference (FACI) provide theoretically grounded, distribution-free prediction intervals robust to non-i.i.d. data streams. Despite their potential, these approaches remain unexplored in flight delay prediction.

Several studies have also embedded predictive uncertainty within operational decision frameworks. Mamdouh et al. (2024) employed DeepONet with gradient-mayfly optimization to manage high-

uncertainty regimes, while Birolini and Jacquillat (2023) combined quantile regression with predictive optimization. However, the broader integration of conformal prediction and network-informed forecasts remains an open area.

2.4 Operational Decision Support and Optimization

The integration of predictive analytics into optimization models has advanced significantly. Riveiro et al. (2025) proposed hybrid ML-queuing models for Airport Capacity Management classifying delays and enabling prescriptive slot allocation. Xu et al. (2024) co-optimized ground handling processes with predictive forecasts, demonstrating gains in punctuality and efficiency. Pang et al. (2024) coupled machine learning predictions with mixed-integer programming, reducing landing time by 17%.

Birolini and Jacquillat (2023) introduced a two-step model combining primary delay reconstruction via quantile regression and delay-aware routing optimization, yielding 3–5% cost reductions. Liu et al. (2024) applied reinforcement learning for adaptive GDP management, although environmental assumptions limited transferability. Hardell et al. (2025) assessed point merge procedures, revealing simultaneous benefits for delay reduction and emissions mitigation. Other studies (Chandra and Verma, 2025; Woo and Moon, 2021) highlighted the relevance of stochastic programming and conflict-aware departure management. Collectively, these contributions underscore the need for integrated predictive-prescriptive systems while emphasizing the challenges of real-time deployment.

2.5 Topological Data Analysis of Delay Networks

Topological Data Analysis (TDA) offers a framework to capture higher-order structures beyond dyadic motifs. Edelsbrunner and Harer (2010) formalized persistent homology to identify multiscale connectivity and loops. Li et al. (2019) demonstrated TDA's utility in uncovering persistent surface inefficiencies in airports. Rajput et al. (2023) and Hickok et al. (2024) applied persistent homology to transportation and polling networks, highlighting its capacity to detect robust patterns.

In aviation delay research, no prior study has applied Vietoris-Rips complexes or persistent homology to model airborne delay propagation. This is notable given evidence that motif-based analyses (Sugishita et al., 2024) reveal repeatable delay loops. TDA offers a systematic alternative that tracks feature persistence across thresholds distinguishing structural loops from transient correlations.

This approach holds promise for uncovering nested and overlapping loops of delay propagation that remain invisible to traditional graph measures or fixed-size motifs.

2.6 Synthesis and Research Gaps

Despite methodological advances across machine learning, network science, causal inference, and optimization, several gaps persist:

- Quantitative links between network properties and delay duration remain underexplored in a regression framework.
- No systematic application of conformal prediction has been undertaken to produce distribution-free, adaptive prediction intervals.
- Persistent homology has yet to be applied to detect stable multi-airport delay loops.

This dissertation addresses these gaps by integrating Prais-Winsten regression of network metrics on delay durations, conformal prediction to quantify forecast uncertainty, and topological data analysis to characterize persistent delay loops. Together, these contributions provide a comprehensive, uncertainty-aware framework for delay prediction and mitigation in Japanese domestic aviation.

Chapter 3 Examining the Association between Network Properties and Departure Delay Duration in Japanese Domestic Aviation

3.1 Introduction

With the increasing demand for air traffic, a surge in air traffic delays, defined as instances when the actual take-off or landing time of a flight is later than the scheduled time, has been observed worldwide (Yang et al., 2023; Sugishita et al., 2024). Such delays not only signal that the airport is operating at or near its practical capacity limit as the observed delays reflect an imbalance between flight demand and the airport's throughput capability but also indicates the lack of airline punctuality, passenger dissatisfaction, economic costs, environmental emissions, and loss of passengers' time. The estimation and prediction of delays are universally recognized as complex tasks (NASEM ACRP, 2014). In the United States, 20% of arriving and departing flights were delayed in 2018 (Bureau of Transportation Statistics, 2020). Similarly, in China, prior to the COVID-19 pandemic, the departure punctuality rate at domestic airports was approximately 70%, with an average departure delay of nearly 28 min (Yang et al., 2023). In Japan, approximately 9% of domestic flights were delayed in 2022, marking a 5.4% increase from the previous year, with 69,280 flights experiencing delays exceeding 15 min. These delays were primarily caused by prior flight delays or cancelations (Ye et al., 2020).

Departure delay is an operational classification of flight delays. It can result from a combination of factors, including airline and airport operations, air traffic control, adverse weather, and human-related issues (NASEM ACRP, 2014; Statista, 2023; CAAC, 2021; Choi et al., 2016). Departure delay is frequently characterized as the primary causal factor in delay propagation. When airlines operate on tightly coordinated schedules, flights naturally become interdependent. Therefore, a departure delay in one flight can trigger a cascading effect on subsequent flights and connections. Even a single delayed flight can cause a domino effect that can spread delays from one airport to another. This domino effect often disrupts the entire flight network of a particular airline as well as the operations of interconnected airports (NASEM ACRP, 2014; Choi et al., 2016; Wu, 2010). Consequently, departure delays often serve as a networked phenomenon among airports. Therefore, departure delays may exhibit network properties that warrant further investigation.

Departure delay duration is a critical parameter in air traffic management and is characterized differently by various stakeholders. For instance, the Federal Aviation Administration prioritizes departure delays exceeding 15 min to align with airport and airline expectations and enhance passenger satisfaction (NASEM ACRP, 2014). In contrast, airports approach departure delays from an operational perspective,

focusing on throughput and the average delay duration to manage airport capacity, often treating delays exceeding 10 min as problematic (NASEM ACRP, 2014). Airlines measure departure delay durations at the gate within 5 to 15 min of the scheduled time and are more concerned with the maximum delay than the average delay (NASEM ACRP, 2014; Choi et al., 2016). Thus, the departure delay duration is a vital parameter requiring investigation in the context of network representation.

Traditional databases for departure delays typically record information, such as origin and destination airports, scheduled and actual departure times, and delay durations (NASEM ACRP, 2014). However, representing departure delays as a network, where airports serve as nodes and delays form the edges, offers a more insightful perspective. This approach facilitates the identification of critical airports within a network that requires particular attention - an analysis that is less efficient with column-oriented databases. Using a network database enables queries such as identifying the most influential airports in terms of delays, identifying airports with similar or dissimilar delay profiles, and detecting patterns of delay occurrences across multiple airports (Bechberger and Perryman, 2020).

Despite the extensive research on flight departure delays, limited attention has been paid to exploring the network properties of these delays and their association with delay durations. This study aims to address this research gap by identifying the statistically significant network properties of departure delays and analyzing their relationship with delay durations. We employ a network representation of departure delay data from Japanese domestic airports served by All Nippon Airways (ANA) and Japan Airlines (JAL) between 2018 and 2021. Initially, we calculate the network properties for each airport connected by departure delays on a daily basis, resulting in a dataset containing the daily network properties and the total departure delay duration for each airport. Subsequently, using a Prais–Winsten regression in a panel data framework, we analyze the relationship between the network properties (independent variables) and departure delay durations (dependent variable). To achieve this objective, we examine four network properties: in-degree centrality, betweenness centrality, eigenvector centrality, and transitivity. The representation of network properties as panel data has been utilized in various transportation studies. For example, Kopsidas et al. (2023) analyzed the association of network centralities with passenger flow in metro systems using regression models. Shi et al. (2024) investigated the spatial distribution of commercial facilities using network centralities, while Wang et al. (2022) applied this methodology to analyze taxi travel patterns.

3.2 Study Area and Data

In this study, data on Japanese domestic flights from 2018 to 2021 were analyzed using the Official Aviation Guide (OAG) historical flight data (OAG, 2023). The dataset contains detailed information on each flight, including flight numbers, departure and arrival airports, terminals, gate numbers, scheduled,

estimated, and actual departure and arrival times, operating airlines, runway times, and instances of flight cancelations. From this dataset, departure delay information was extracted to construct departure delay networks. The departure delay was calculated as the difference between the actual departure time at the gate and the scheduled departure time. The analysis focused on two full-service carriers in Japan: ANA and JAL.

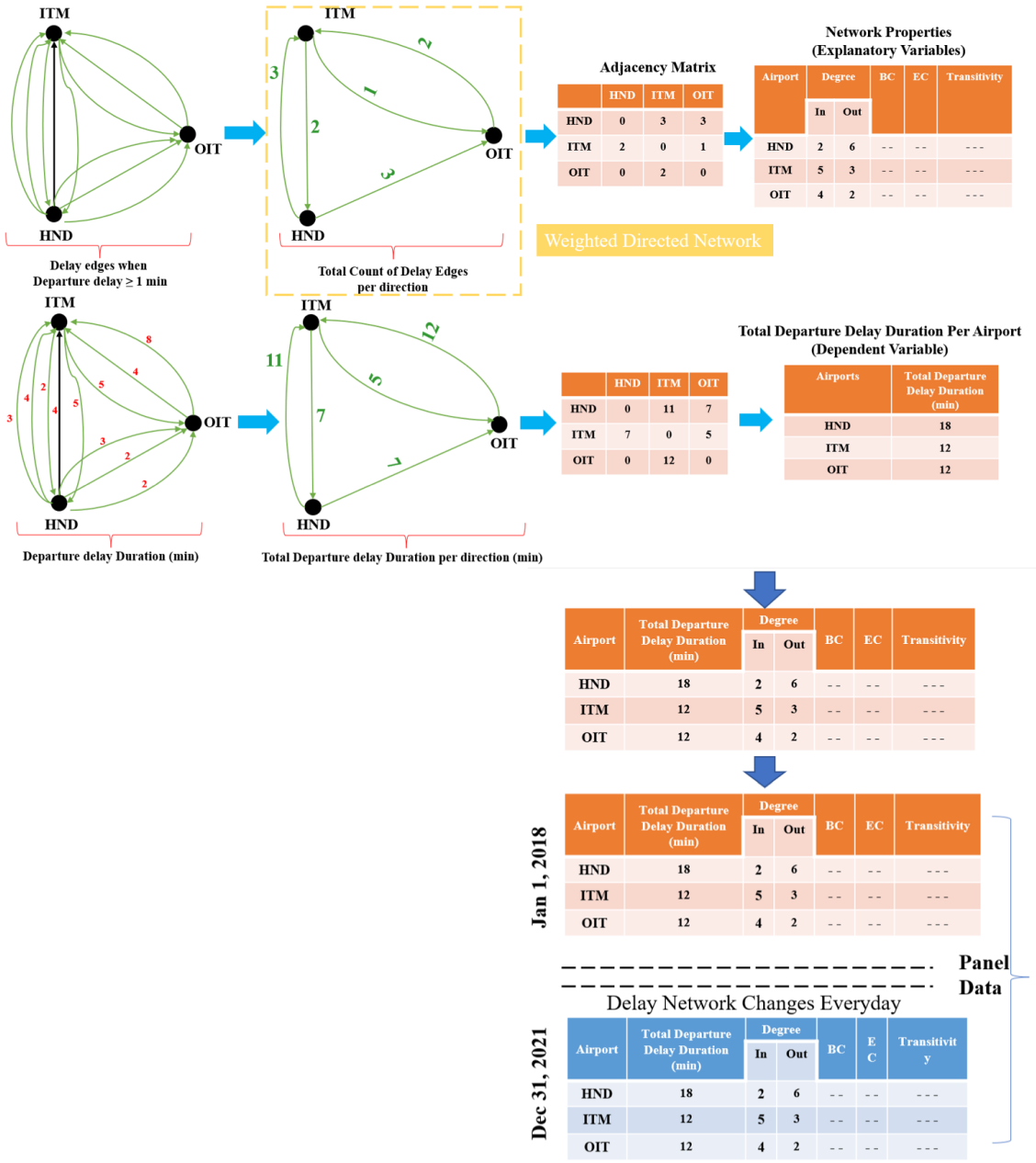


Figure 3.1: Steps of data pre-processing to prepare panel format data from the network-based data followed in Chapter 3

Daily snapshots of the delay networks were created to represent the departure delays of domestic Japanese flights from 2018 to 2021. In these networks, nodes represent airports, and edges denote flights with departure delays of 1 min or more. Focusing only on the delay value would help to understand the delay propagation across the airports in terms of regression point of view through various network metrices that helps to understand how readily it imports, amplifies or dissipates delay shocks and it would also help to understand how airport's structural position in the network relates to total departure delay duration. Figure 3.1 represent the steps of data pre-processing to prepare data from the network-centric database of departure delay.

Four network properties were analyzed, consisting of four continuous and two categorical variables. The continuous variables included in-degree centrality, betweenness centrality, eigenvector centrality, and transitivity. The categorical variables were airport type (hub or spoke) and seasonality (quarters 1, 2, 3, and 4). Based on Hanaoka's study (Hanaoka, 2018), eight airports were categorized as hub airports for domestic flights in Japan: New Chitose (CTS), Fukuoka (FUK), Tokyo International (HND), Narita International (NRT), Chubu Centrair International (NGO), Kansai International (KIX), Osaka International (ITM), and Naha (OKA). These hub airports are depicted in Figure 3.2. For classification purposes, airports were assigned a value of hub = 1 if they were among these eight hubs; otherwise, they were categorized as spoke = 0.

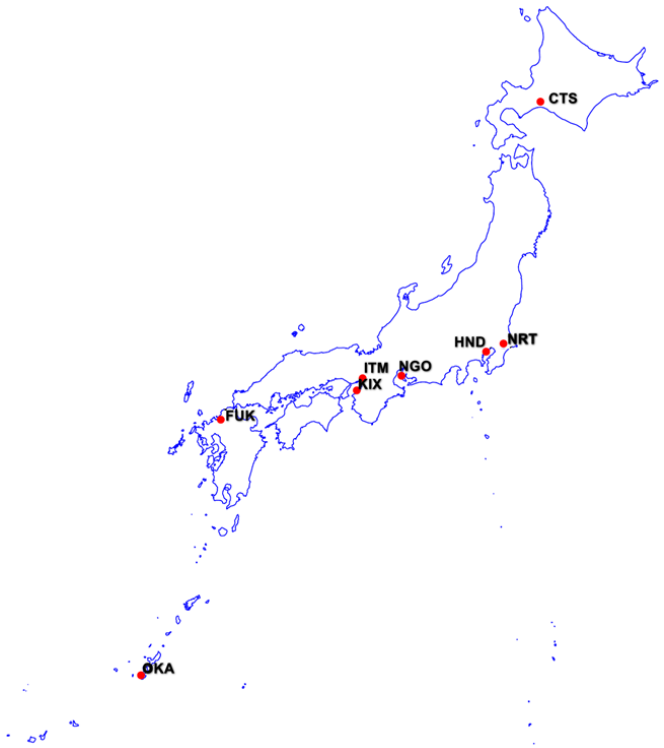
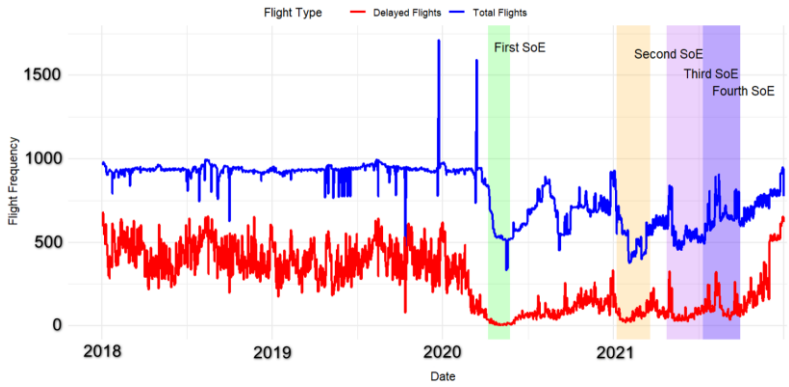
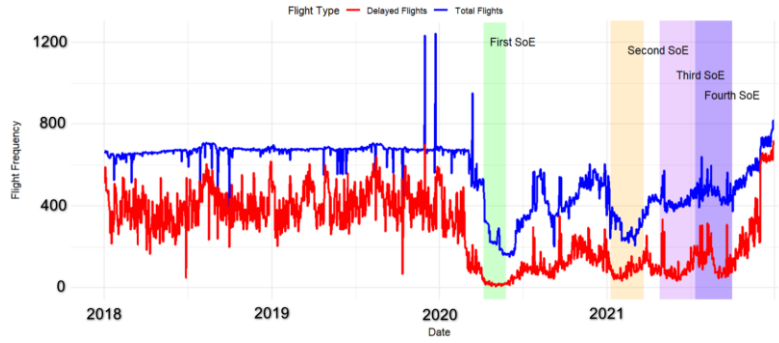


Figure 3.2: Eight hub airports in the Japanese domestic air transport network

Seasonality was divided into four groups: January to March (Q1 = 1), April to June (Q2 = 2), July to September (Q3 = 3), and October to December (Q4 = 4). This study developed a model covering two periods across four individual years, from 2018 to 2021. The years 2018 and 2019 were designated as the pre-COVID-19 phase, while 2020 and 2021 were considered the during-COVID-19 phase. For the pre-COVID-19 period, seasonality was not incorporated; however, for the during-COVID-19 period, seasonality was included because of the multiple rounds of state-of-emergency (SoE) declarations by the Government of Japan. The first SoE lasted from 7 April to 25 May 2020: the second from 8 January to 21 March 2021: and the third from 25 April to 19 June 2021, except for Okinawa, where it was lifted on 30 September 2021. The fourth SoE spanned from 12 July to 30 September 2021 (Okamoto, 2022). Figures 3.3(a) and 3.3(b) illustrate the frequency of domestic flights and number of delayed departure flights per day for ANA and JAL, respectively, along with the durations of the four SoEs during 2020–2021. The data indicate that both flight frequency and the frequency of delayed departures declined significantly during the first and second SoEs. However, following the lifting of the fourth SoE, flight frequency and delayed departures began to recover, mirroring trends observed during the pre-COVID-19 period.



(a) ANA



(b) JAL

Figure 3.3: Variation in the frequency of total flights and departure-delayed flights in domestic regions in Japan from 2018 to 2021

For each airport and day, the total departure delay duration is calculated as the sum of the departure delay durations for all flights departing from that airport on that day. For instance, if three ANA flights depart from CTS to HND, NRT, and FUK on a given day, and two of these flights experience delays of 1 min or more, two edges are added from CTS to the respective destination airports. This information is used to derive network data, such as the out-degree centrality of CTS and the in-degree centrality of the destination airports. Hypothetically, if the departure delay durations for CTS-HND and CTS-NRT are 5 min and 10 min, respectively, the total departure delay duration for that day would be 15 min. This total departure delay duration is treated as the dependent variable, whereas the other network data are considered independent variables, as detailed in Section 5.1. Not all flights departing from a particular airport experience delays each day. This variability implies that the characteristics of the departure delay network can vary daily. Understanding the associations between changes in departure delay networks and the total daily departure delay duration across airports is of critical importance.

3.3 Methodology

This study employs a panel data regression model to analyze the association between network properties and the departure delay duration. Network adjacency metrics are utilized to calculate the network properties of each airport for each day, which are subsequently represented as panel data to fit the regression model. Figure 3.4 outlines a flow-chart of the detailed usage of network properties and statistical methods.

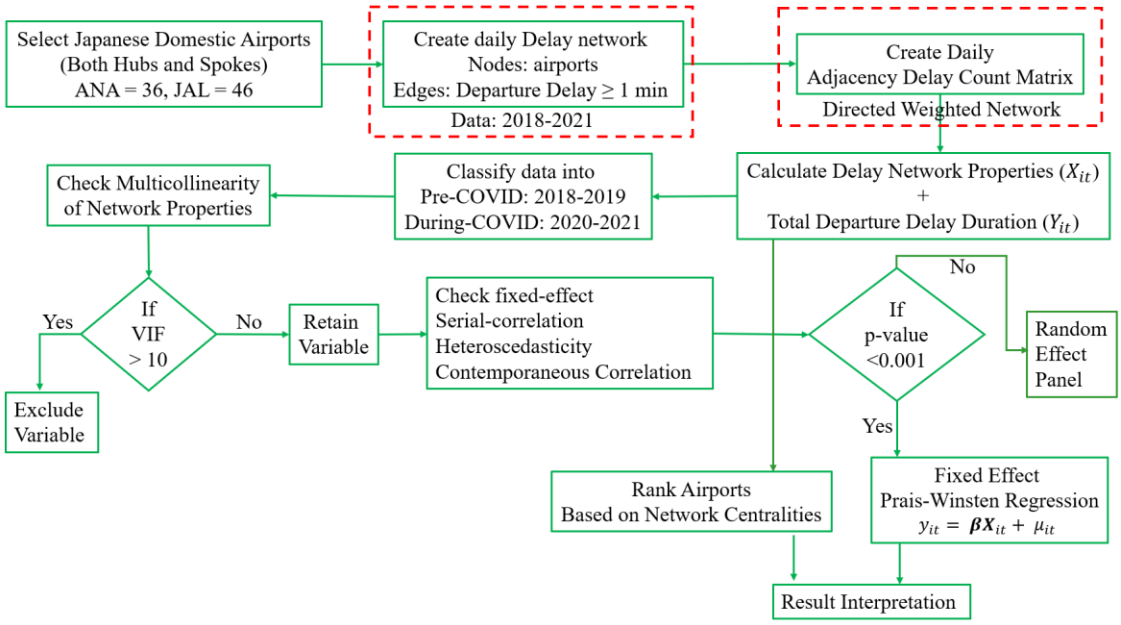


Figure 3.4: Methodological flow-chart followed in Chapter 3

3.3.1 Network Properties

In this study, departure delay data between airports over a four-year period is analyzed. Airports are represented as nodes, while departure delays are represented as edges. Based on this data representation, four network properties of the delay network, as previously described, were calculated. For instance, in-degree centrality represents the number of delayed flights received by an airport from other airports. In the context of this study, in-degree centrality measures the total count of flights that departed late from their respective originating airports. It is important to note that this metric differs from arrival delay. A high in-degree centrality value indicates that an airport in the delay network receives a substantial number of delayed departure flights. Betweenness centrality (BC) typically measures the frequency with which a node lies on the shortest path between two other nodes (Newman, 2018). In this study, BC is examined to evaluate how an airport experiencing a departure delay is connected to other airports, either as a hub or a spoke (An, 2023). This metric reflects the extent to which an airport serves as a bridge between different airport types experiencing departure delays. An increase in BC indicates that certain airports are critical nodes situated on major paths connecting numerous origin-destination pairs. Consequently, delays at such airports are more likely to propagate across a wide range of routes. From the perspective of delays, airports with higher BC values generate more delayed departure links to both hub and spoke airports, thus playing a key role in the spread of delays. Eigenvector centrality (EC) measures the degree to which an airport is connected to other well-connected airports. In the context of departure delays, EC evaluates whether an airport experiencing delays is also connected to other airports that exhibit frequent delays. A higher EC value suggests that delays are more prevalent at airports connected to other delay-prone airports. Finally, transitivity quantifies the extent to which a node (in this case, a departure-delayed airport) tends to cluster or form triads within the delay network. This metric reflects the likelihood that two nodes connected to a common node are also directly connected to each other, forming a triangle. High transitivity indicates that the network contains communities or groups of nodes that are densely interconnected. From the perspective of delays, an airport with high transitivity is likely connected to other airports that also experience frequent departure delays, and thus form a triad among themselves. This could be important because when airports are tightly connected like this, delays can spread more easily within these groups. Some studies have incorporated global network measures alongside local ones for econometric analyses (Rainone, 2020; Menzel, 2020). Table 3.1 provides a detailed mathematical explanation of the network properties.

Table 3.1: Mathematical expression and explanation of network properties

Network property	Equation	Explanation
In-degree Centrality	$C_{in}(v) = \sum_{u \in V} w_{uv}$	If w_{uv} represents delay minutes from u to v, this captures the magnitude of delays, not just the count of delayed connections.
Betweenness Centrality (BC)	$C_{BC}(v) = \sum_{s \neq v \neq t} \frac{\sigma_{st}(v)}{\sigma_{st}}$	How often a node v appears on the shortest paths between pairs of nodes *s* and *t*. Here, σ_{st} is the number of weighted shortest paths from node s to node t, and $\sigma_{st}(v)$ is the number of those weighted paths passing through *v*.
Eigenvector Centrality (EC)	$C_{EC}(v) = \frac{1}{\lambda} \sum_{u \in V} w_{vu} C_{EC}(u)$	It refers to that airports connected to other delay-heavy and central airports inherit higher importance. Here, λ is the largest eigenvalue of the weighted adjacency matrix *W*, and w_{vu} are the edge weights (delay strength).
Transitivity	$C_T = \frac{3 \times Number\ of\ triangles}{Number\ of\ connected\ triples\ of\ nodes}$	Here, connected triples refer to nodes (i.e., airports) with edges (departure delay edges with delay counts and delay duration) running to two other nodes (i.e., airports). The factor *3* accounts for a single triangle as triad.

3.3.2 Panel Data Model using Prais-Winsten Regression

To identify the significant delay network properties associated with departure delay duration, once the network properties for each airport on each day were calculated, one-year panel data were utilized to develop four models for the years 2018 through 2021 for ANA and JAL. For each day, the total departure delay duration at each airport was aggregated, thereby forming the panel dataset. This study employs unbalanced panel data due to differences in the time observations across individuals, with a maximum time observation of 365 days and a maximum of 46 airports. This structure is also referred to as long-panel data, as the time dimension exceeds the individual dimension (i.e., 365 days > 46 airports).

Traditionally, panel data modeling involves either fixed-effects or random-effects models. Fixed-effects models facilitate the examination of temporal changes within panels (i.e., airports), while random-effects models allow for analysis between panels. After performing the Hausman test (Greene, 2012), the fixed-effects model was determined to be more appropriate for this study.

Network properties are often highly correlated with one another (Cimenler et al., 2014), which in-creases the likelihood of serial correlation, cross-sectional dependence, and heteroscedasticity. To test for cross-sectional dependence or contemporaneous correlation, the Breusch-Pagan Lagrange multiplier test and the Pasaran cross-sectional dependence test were employed (Breusch and Pagan, 1980; Pesaran, 2004). Serial correlation was assessed using the Lagrange multiplier test (Baltagi et al., 2012). The results indicated the presence of cross-sectional dependence, heteroscedasticity, and serial correlations in the dataset. In light of these findings, a fixed-effects Prais-Winsten regression model with correlated panels corrected standard errors (PCSE) was employed. This approach accounts for regression coefficients with robust and reliable standard errors (Hoechle, 2007; Prais and Winsten, 1954; Beck and Katz, 1995). The PCSE coefficients were estimated by calculating standard errors and variance-covariance estimates, where the Prais-Winsten regression method assumes disturbances to be heteroscedastic and contemporaneously correlated across panels. Given the unbalanced nature of the panel data, parameters were estimated with first-order autocorrelation by specifying panel-specific autocorrelation weights based on T_i. Unbalanced panels often contain missing data; to address this issue, pairwise estimation of the inter-panel covariance matrix of the disturbances was used. This method ensures that all available observations common to each panel contribute to the computation of the covariance matrix. To the best of our knowledge, the modeling framework employed in this study represents a novel approach for estimating parameters and standard errors in a robust and reliable manner for this type of data representation in the analysis of delay data in the air transport domain.

In this study, the dependent variable is the departure delay duration for each airport, while the independent variables include the network properties of in-degree, betweenness, eigenvector, and transitivity for each airport per day. Dummy variables for airport type and seasonality are also included as independent variables.

This model can be written as:

$$y_{it} = \boldsymbol{\beta} \boldsymbol{X}_{it} + \mu_{it} \tag{3.1}$$

where y_{it} represents the departure delay duration for airport i at time t, $\boldsymbol{X}_{it}$ represents the network properties derived for airport i at time t, i represents the individual airports (connected in the departure delay network) (panels), and $t = 1, 2, \dots, , T_i$; T_i = number of periods in panel i, and μ_{it} depicts the

disturbances autocorrelated along t or contemporaneously correlated across i. As autocorrelation was specified in our model, we used Prais-Winsten estimates with the Generalized Least Squares (GLS) method (Hoechle, 2007; Beck and Katz, 1995) using *STATA 15* with the *xtpcse* library (Hoechle, 2007).

Note that the serial correlation is computed as the average of daily residual auto-correlations as:

$$\rho = \frac{\rho_1 + \rho_2 + \ldots\ldots + \rho_n}{n} \tag{3.2}$$

The PCSE covariance matrix used the Kronecker product of the panel covariance and an identity matrix.

$$Var(\beta) = (\boldsymbol{X}'\boldsymbol{X})^{-1}\, \boldsymbol{X}'\boldsymbol{\tau}\boldsymbol{X}\, (\boldsymbol{X}'\boldsymbol{X})^{-1} \tag{3.3}$$

where, $\tau = \boldsymbol{\Sigma}_{n\times n} \otimes \boldsymbol{I}_{n\times n}$ where, $\boldsymbol{\Sigma}$ is the n by n panel-by-panel covariance matrix of the disturbances

$\boldsymbol{\Sigma}$ is estimated as $\widehat{\Sigma_{ij}} = \frac{\epsilon_i \epsilon_j}{T_{ij}}$

where, ϵ_i and ϵ_j are the residuals for panels i and j, respectively that can be matched by period, and

T_{ij} is the number of residuals between panel i and j that is matched by time period. Here, the $\boldsymbol{\Sigma}_{n\times n} \otimes \boldsymbol{I}_{n\times n}$ is the simplest structure that captures cross-sectional dependence (airports correlated at each date) while maintaining serial independence (no unwanted correlation across days) once it is already handled the $AR(1)$ time process in the Prais–Winsten transformation.

In our case, as not all airports in all days experienced departure delay, so not all airports each day produced delay edges, so this study used unbalanced panel data. Therefore, here the matching was done with case-wise basis i.e., only those residuals from period that are common to all panels.

This GLS-based approach let isolate the within-airport effect. For example: how a one unit-change in in-degree centrality on a given day adds β minutes of delay beyond each airport's baseline.

3.4 Results and Discussion

3.4.1 Summary of Network Properties

Table 3.2 presents the fundamental properties of the ANA and JAL flight departure delay networks for the period from 2018 to 2021. The number of nodes represents the airports connected by flights operated by each airline, while the number of edges corresponds to the directed edges, indicating the number of departure-delayed flights (≥ 1 min) between these nodes or airports. The mean path length serves as a measure of network cohesiveness, with smaller values suggesting that nodes are more closely connected.

Table 3.2: Departure delay network properties of ANA and JAL

Properties		2018	2019	2020	2021
Nodes	ANA	36	36	36	35
	JAL	46	45	44	46
Edges	ANA	188,279	166,926	61,060	68,006
	JAL	161,258	169,138	64,202	74,236
Mean Path Length	ANA	1.90	1.92	1.92	1.88
	JAL	2.21	2.17	2.14	2.09

Table 3.2 reveals minimal variation in the number of nodes between the pre-COVID (2018–2019) and during-COVID (2020–2021) periods, with at most one or two nodes being omitted for either airline. This consistency suggests that most flights continued to experience departure delays despite the COVID-19 outbreak. Similarly, relatively minor variations were observed in the mean path length, which ranged from 1.90 for ANA to 2.21 for JAL. This indicates that the network structure remained largely stable, as flights continued to operate with departure delays during the COVID-19 period without significant changes in network cohesiveness when compared to the pre-COVID period. However, the number of edges decreased notably during the COVID-19 period in the Japanese domestic flight network. This reduction likely reflects a decline in the number of flights between airports due to the surge in COVID-19 cases and subsequent government interventions, as illustrated in Figure 3.3.

Table 3.3 provides the basic descriptive statistics of the network properties and total departure delay duration for ANA and JAL from 2018 to 2021. These statistics include the means and standard deviations for all airports during the study period. For ANA, the mean number of incoming flights that departed late (in-degree centrality) was approximately 12, with a standard deviation of 20.67, indicating substantial variation among airports. The mean BC was 30.09, and the EC was 0.23, both exhibiting relatively large standard deviations. The mean transitivity was 0.62, with a standard deviation of 0.39. The mean value of total departure delay duration was 142.29 min, with a high standard deviation of 489.38, signifying considerable variability in delay durations across connected airports. For JAL, the mean in-degree centrality was 9.28, with a standard deviation of 19.55, which was higher than ANA's variation. The mean BC was 44.65, exceeding that of ANA, while the mean EC was 0.18. The mean value of total departure delay duration was 103.73 min, which was lower than ANA's, although its standard deviation was comparable to that of ANA.

Table 3.3: Descriptive statistics of network properties

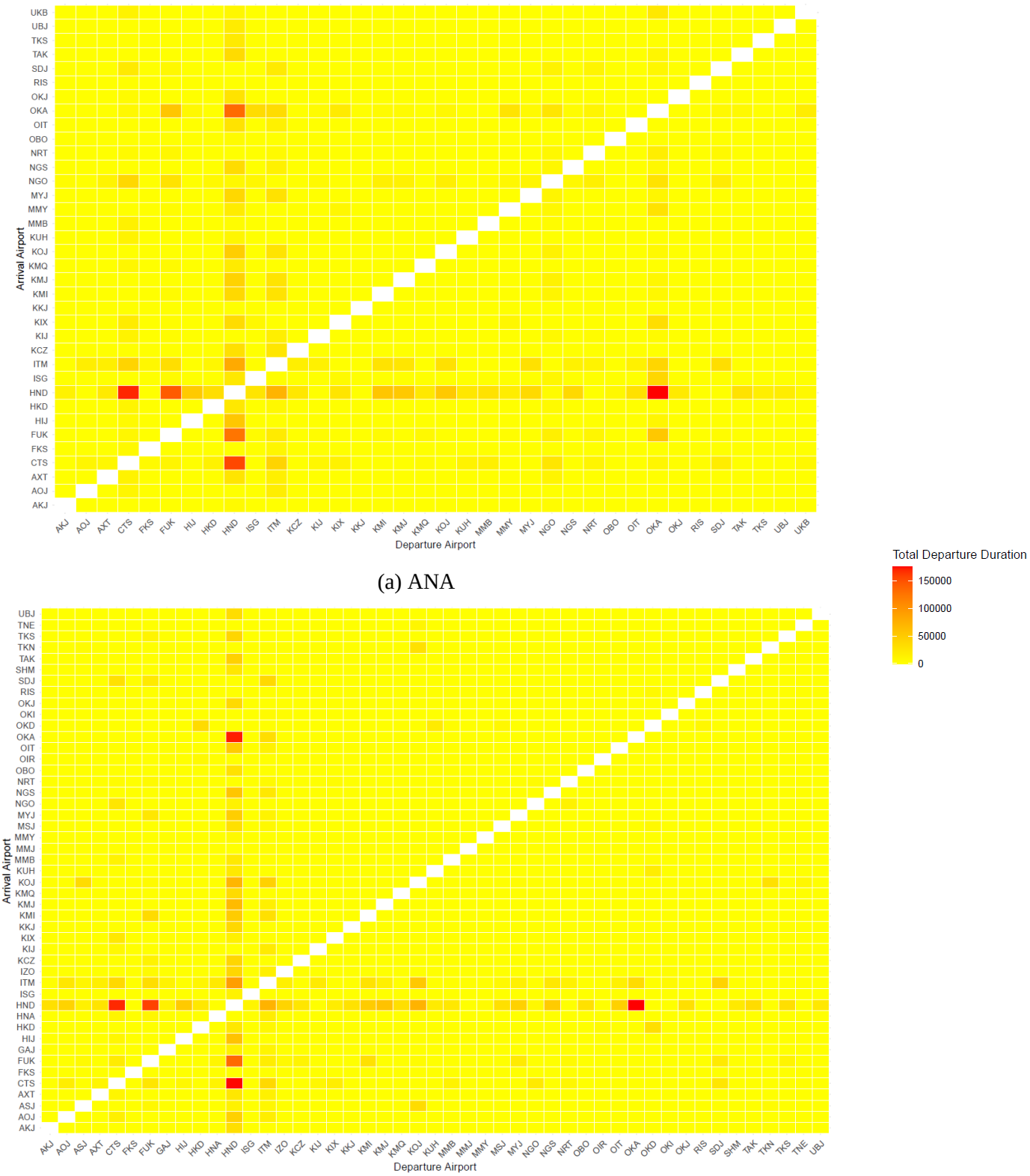
Variables	ANA		JAL	
	Mean	Standard Deviation	Mean	Standard Deviation
In-degree Centrality	12.30	20.67	9.28	19.55
Betweenness Centrality (BC)	30.09	98.53	44.65	174.12
Eigenvector Centrality (EC)	0.23	0.24	0.18	0.21
Transitivity	0.62	0.39	0.45	0.45
Total Departure Delay Duration (min)	142.29	489.38	103.73	482.94

Table 3.4 provides the Pearson correlation coefficients for in-degree, betweenness, eigenvector, and transitivity. The results demonstrate that in-degree exhibits a strong correlation with betweenness (> 0.85) and a moderate correlation with eigenvector (> 0.70). Furthermore, eigenvector and betweenness display a moderate inter-correlation (≥ 0.70). Notably, transitivity is negatively correlated with all other variables, a result that merits particular attention. An analysis of these correlation values reveals the potential for multicollinearity among the network properties, particularly between in-degree and betweenness. To assess the severity of multicollinearity, the Variance Inflation Factor (VIF) was calculated for these variables. The results confirmed that the VIF values were below 10, indicating acceptable levels of collinearity, which aligns with standard practice (James et al., 2021). Specifically, the VIF values for ANA and JAL were 4.69 and 6.51 for in-degree, 5.02 and 6.13 for betweenness, 2.25 and 2.28 for eigenvector, and 1.15 and 1.10 for transitivity, respectively.

Table 3.4: Pearson correlation coefficients between network centralities for ANA and JAL

	ANA/JAL			
	In-degree	Betweenness	Eigenvector	Transitivity
In-degree	1.00	0.88/0.91	0.71/0.74	-0.25/-0.09
Betweenness		1.00	0.73/0.70	-0.33/-0.19
Eigenvector			1.00	-0.29/-0.01
Transitivity				1.00

Accordingly, despite the relatively high correlation between in-degree and betweenness observed in Table 3.4, both variables were retained in the model for two principal reasons. First, the VIF values were within



(a) ANA

(b) JAL

Figure 3.5: Adjacency matrix of Total Departure Delay Duration (in minutes) of each airport from 2018 to 2021

the acceptable threshold of 10. Second, both network properties play essential roles in analyzing their relationship with departure delay duration within the delay network.

Figures 3.5(a) and 3.5(b) display adjacency matrix tables representing the total departure delay durations (in minutes) for flights departing from various airports between 2018 and 2021. In these figures, both axes correspond to the IATA codes of the airports, with further details on the full names of these airports provided in Appendix Table A1. For ANA flights, the HND-CTS and OKA-HND routes recorded the highest departure delay durations in both directions during this period. Similarly, for JAL flights, the routes HND-OKA, HND-CTS, and HND-FUK exhibited the highest departure delay durations. A review of the adjacency matrices highlights that departure delays were most prevalent on flights originating from HND, ITM, and OKA, affecting a considerable number of connecting airports for both ANA and JAL.

3.4.2 Model Estimation

3.4.2.1 Pre-COVID Phase

The Table 3.5 summarizes single variable Prais-Winsten regressions of total departure delay duration on network-derived variables for two Japanese carriers (ANA and JAL) during the pre-COVID period (2018–2019). In the Prais-Winsten model, the first-order serial correlation (ρ) is estimated rather than assumed to be zero. Each independent variable is entered separately, so the R^2 values reflect the explanatory power of individual network measures. For ANA, for in-degree (incoming connections), a one-unit increase in the number of incoming delay links was associated with longer departure delays. The coefficient was 0.32 in 2018 and 0.20 in 2019, both statistically significant. The associated intercepts were close to zero (- 0.003 in 2018 and 0.013 in 2019). This variable explained a substantial portion of the variation in delay duration ($R^2 = 0.58$ in 2018 and 0.34 in 2019) and exhibited moderate AR(1) autocorrelation ($\rho = 0.22 - 0.37$). In case of BC, airports that sit on many shortest paths tended to exhibit longer delays: the coefficient was 0.25 in 2018 and 0.14 in 2019. The intercepts were -0.010 (2018) and 0.040 (2019), and the AR(1) parameter for this model was relatively high ($\rho \approx 0.37$ for both years). BC accounted for about one-third of the delay variation in 2018 ($R^2 = 0.32$) but just 3% in 2019 ($R^2 = 0.03$). For EC, airports connected to other well delay-connected airports also faced more delay. The EC coefficients were 0.16 (2018) and 0.10 (2019), with intercepts 0.001 and 0.01. his variable produced the highest R^2 among the network measures (0.46 in 2018 and 0.23 in 2019), and ρ increased from 0.22 to 0.40 across the two years. Lastly, for transitivity, greater clustering in the delay network slightly reduced total delay. The coefficient was -0.03 in 2018 and -0.02 in 2019 (intercepts -0.003 and 0.027). This measure showed low explanatory power ($R^2 \approx 0.28$ in 2018 and 0.03 in 2019) and modest serial correlation ($\rho \approx 0.32 - 0.37$).

Table 3.5: Estimation result for single variable regression at the phase of pre-COVID (2018 and 2019)

Airline	Variables	2018				2019			
ANA	In-degree	0.32				0.20			
	BC		0.25				0.14		
	EC			0.16				0.10	
	Transitivity				-0.03				-0.02
	Constant	-0.003	0.013	-0.010	0.040	0.001	0.01	-0.003	0.027
	ρ	0.22	0.37	0.37	0.37	0.22	0.40	0.32	0.37
	R-squared	0.58	0.34	0.32	0.03	0.46	0.23	0.28	0.03
	Prob > chi2	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
	n	11,710				11,531			
	Number of airport	36				36			
	Number of time (range)	18-365				17-356			
JAL	In-degree	0.32				0.28			
	BC		0.23				0.17		
	EC			0.15				0.11	
	Transitivity				-0.003				-0.004
	Constant	-0.002	0.009	-0.010	0.017	-0.002	0.005	-0.007	0.012
	ρ	0.24	0.41	0.42	0.59	0.18	0.37	0.42	0.56
	R-squared	0.63	0.38	0.31	0.01	0.65	0.40	0.30	0.01
	Prob > chi2	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
	n	14,242				14,291			
	Number of airport	46				45			
	Number of time (range)	1-364				1-358			

NB: All the variables are statistically significant at 5% significance level

For JAL, in-degree centrality, JAL's regression produced coefficients of 0.32 (2018) and 0.28 (2019), again indicating that airports receiving delays from more partners had longer total delay durations. he intercepts were roughly –0.002 in 2018 and 0.009 in 2019, and the AR(1) coefficient increased from $\rho = 0.24$ to 0.41. R^2 values show strong explanatory power (0.63 in 2018 and 0.38 in 2019). For BC, the coefficients were 0.23 (2018) and 0.17 (2019); intercepts were - 0.010 and 0.017. This variable explained 31% of delay variation in 2018 but just 1% in 2019 ($R^2 = 0.01$), and ρ rose from 0.42 to 0.59. EC remained a stronger predictor; coefficients were 0.15 in 2018 and 0.11 in 2019 with intercepts around −0.002 and 0.005. R^2 values were high (0.65 in 2018 and 0.40 in 2019), and the ρ values were 0.18 and 0.37. For transitivity – as with ANA, higher clustering slightly decreased delay duration. The coefficient was - 0.003 in 2018 and - 0.004 in 2019; intercepts were - 0.007 and 0.012. The explanatory power was low ($R^2 \approx 0.30$ and 0.01), and ρ was moderate (0.42 and 0.56).

Overall, the regressions show that network position strongly influences flight-delay propagation. For both airlines, in-degree, BC and EC have positive and highly significant coefficients, implying that airports

with more incoming delay links or greater centrality experience longer cumulative departure delays. Transitivity is negative and small, suggesting that densely clustered delay networks can slightly mitigate delays. The AR(1) coefficients (ρ) indicate moderate serial correlation, and the high R^2 values for in-degree and EC highlight their importance in explaining total departure delay durations.

For multi-variable Prais-Winsten Regression estimates shown in Table 3.6 presents the estimation results of the Prais-Winsten regression for the pre-COVID phase, aiming to identify the association between significant network properties and departure delay duration for both ANA and JAL. Two models, one for each airline, are estimated for the years 2018 and 2019. The coefficients for in-degree centrality were found to be 0.410 and 0.515 for ANA and JAL, respectively, at the 1% significance level in 2018, and 0.379 and 0.402 for ANA and JAL, respectively, at the 1% significance level in 2019. These results strongly suggest that the observed relationship is not due to random chance. These estimates can be interpreted as follows: for a given airport in the delay network, on a day when the in-degree centrality is one unit higher than its typical value for that airport, the departure delay duration for each flight may be expected to increase by 0.410 min (ANA, 2018), 0.379 min (ANA, 2019), 0.515 min (JAL, 2018), or 0.402 min (JAL, 2019). This suggests that before the COVID-19 out-break, the greater the number of arriving flights that departed late, the more likely it was that departure delays would occur at the receiving airports, thereby increasing the daily departure delay duration at those airports. When considering only hub airports, the departure delay duration remained positively associated with in-degree centrality for ANA ($\beta = 0.111$ in 2018) and JAL ($\beta = 0.109$ in 2018; $\beta = 0.191$ in 2019), all at the 1% significance level. This finding potentially indicates that when hub airports receive a higher number of late-departing flights, the delay duration at these receiving hub airports tends to increase. Figures 3.6 and 3.7 present the rankings of all four network properties of each airport based on departure delays for ANA and JAL, respectively, for each year from 2018 to 2021. It was observed that HND, ITM, CTS, and FUK ranked in the top five for in-degree centrality in both 2018 and 2019, which reflects the dominance of hub airports in receiving late-departing flights. This dominance aligns with Budd and Ison (2017) since hub airports typically handle a significant number of international and domestic flights under normal conditions. Our results align with those of Wu et al. (2010), who noted that delayed flights at their origin airports may lead to increased turnaround times at the arrival gate of the receiving airports, which could further contribute to departure delays for subsequent flights. As our delay network includes both hub and spoke airports, some spoke airports with higher in-degree centrality rankings, such as KMI, MYJ, KOJ, and KMJ (see Figures 3.6 and 3.7), were found in the top 10 list for both ANA and JAL. This highlights the influence of certain spoke airports within the delay network and suggests a possible increase in departure delay durations at these airports.

Table 3.6: Estimation result for multiple variable regression at the phase of pre-COVID (2018 and 2019)

Variables	ANA		JAL	
	2018	**2019**	**2018**	**2019**
In-degree	0.410***(0.031)	0.379***(0.016)	0.515***(0.021)	0.402***(0.017)
In-degree*Hub	0.111***(0.033)	-0.048***(0.018)	0.109***(0.029)	0.192***(0.033)
Betweenness	0.251 (0.547)	0.103 (0.145)	-0.028***(0.009)	-0.005 (0.006)
Betweenness *Hub	-0.365 (0.546)	-0.210 (0.144)	-0.143***(0.018)	-0.115***(0.016)
Eigenvector	-0.046***(0.009)	-0.057***(0.004)	-0.057***(0.005)	-0.029***(0.003)
Eigenvector *Hub	-0.011 (0.011)	0.048***(0.006)	-0.023**(0.009)	-0.049***(0.008)
Transitivity	-0.0002 (0.0004)	-0.001***(0.0002)	0.0007**(0.0002)	-0.0001 (0.0001)
Transitivity*Hub	-0.007***(0.002)	-0.004***(0.001)	0.005***(0.001)	0.006***(0.0006)
Constants	-0.0008* (0.0005)	0.0019***(0.0003)	-0.001***(0.0002)	-0.001***(0.0001)
Number of Observations	11,710	11,531	14,242	14,291
Number of airport entities	36	36	46	45
Number of time entities (range)	18-365	17-356	1-364	1-358
R-squared	0.61	0.50	0.67	0.68
Prob > chi2	0.000	0.000	0.000	0.000

(In the parenthesis standard errors are reported in parentheses. The star marks represent the statistical significance level at '***' 1% '**' 5% '*' 10%)

In 2019, for ANA, we observed a different sign in the interaction effect of hub airports with in-degree centrality ($\beta = -0.048$, at a 1% significance level). This outcome is plausible because of the presence of outliers in the 2019 data, which may alter the magnitude and direction of the regression coefficients (Aggarwal, 2017). Figure 3.3(a) illustrates that an outlier in the frequency of delayed flights (denoted in red) occurred in the last quarter of 2019. In our study, we did not exclude outliers because we utilized specialized data derived from a network database and considered the entire delay network for each day to calculate the network properties for each airport. Therefore, removing outliers could have hindered the proper establishment of relationships and may have resulted in the omission of important information.

An increase in BC suggests that certain airports function as potentially critical hubs, which may spread departure delays to other airports. In other words, airports with higher BC values are more likely to lie on the shortest path between other airports and are more likely to serve as bridges for airports that are less directly connected in the network. Our findings indicate that BC is positively associated with departure

delay duration for ANA, with coefficients of $\beta = 0.251$ in 2018 and $\beta = 0.103$ in 2019. However, neither of these results is statistically significant. In contrast, BC shows a negative association with departure delay duration for JAL, with $\beta = -0.028$ in 2018 at the 1% significance level and $\beta = -0.005$ in 2019, although the latter lacks statistical significance. This result suggests that, in the delay network, along with hub airports, some spoke airports may also serve as critical airports. For example, for ANA, spoke airports such as SDJ, KOJ, and KMI were ranked among the top 10 in BC (Figure 3.6). For JAL, airports including HKD, KOJ, OKD, KUH, and MSJ were ranked in the top 10, along with hub airports (Figure 3.7). These spoke airports are more likely to be critical within the network when departure delays are a concern. If this is the case, there could be a chance that the departure delay duration is more likely to be longer for ANA, whereas it is more likely to be shorter for JAL. In such instances, one potential measure for ANA flights could involve increasing turnaround time to provide a greater buffer against delays at critical airports.

When focusing exclusively on hub airports, both airlines exhibited a negative association between BC and departure delay duration in 2018 and 2019. However, statistical significance was observed only for JAL, with coefficients of $\beta = -0.143$ in 2018 and $\beta = -0.115$ in 2019, both at the 1% significance level. These results suggest that the relationship between BC and departure delay duration differs between ANA and JAL when considering the entire delay network, but similar patterns emerge for both airlines when focusing exclusively on hub airports. When considering only hub airports, both airlines were more likely to experience shorter delays with higher BC values. This suggests that for JAL, connecting flights departing from high-BC airports (in terms of delay) are more likely to encounter shorter delays, particularly if those airports are categorized as hubs. In contrast, spoke airports that act as higher-BC airports (in terms of delay) for JAL are more likely to experience longer departure delays. Figure 3.7 illustrates that some spoke airports, namely HKD, KOJ, OKD, KUH, and MSJ, ranked 3rd, 4th, 6th, 8th, and 9th in the BC ranking list. This indicates that these airports are high-BC airports in terms of departure delays, making them more prone to longer delays. Zhou et al. (2022) and Zanin et al. (2017) found that in China, 40%–50% of smaller airports are more prone to experience delays in a flight network and engage in delay propagation, whereas, in the USA, Li et al. (2024) reported that smaller airports are more likely to be sensitive to delay propagation.

Hub airports are typically prone to frequent departure flights, and it is plausible that the number of delayed departures may also be higher at these airports. However, the observed negative association suggests that the departure delay duration is more likely to be shorter for connecting airports with higher BC. This implication is particularly relevant for hub airports in Japan because, for both ANA and JAL,

some domestic airports serve as transit points for both airlines. For example, for JAL, the connecting airports are HND, NRT, and OKA (JAL., 2024), and for ANA, the airports are HND, NRT, ITM, KIX, OKA, and FUK (ANA., 2014). These airports are already established as hub airports in terms of flight operations and connections, and they are well equipped to handle domestic transit flights, ensuring minimal departure delays.

In particular, although in-degree centrality and BC exhibit relatively high positive correlations with each other, Figures 3.6 and 3.7 reveal that there are airports where the BC ranking exceeds that of in-degree centrality, and vice versa. For example, in Figure 3.7 (for JAL flights), the spoke airports HKD, OKD, KUH, and MSJ were ranked 14th, 18th, 26th, and 30th for in-degree centrality in 2018, but ranked 3rd, 6th, 8th, and 9th for BC in the same year. Similar patterns were observed in 2019 for JAL at these airports. For ANA, a similar trend was observed for SDJ, which was ranked 11th and 12th in 2019 and 2018 for in-degree centrality, respectively, but ranked 9th for BC in both years. Therefore, it is not uncommon for the coefficients of BC to show the opposite sign compared to in-degree centrality. These two network centralities capture distinct aspects of the role of spoke airports within the network. Even if these airports experience fewer delayed flights, as indicated by their in-degree centrality rankings, their BC rankings suggest that they have the potential to become critical in spreading delays to other airports.

In the case of EC, both ANA and JAL show that an increase in EC is associated with a decrease in departure delay duration, with results that are statistically significant at the 1% level. Specifically, for ANA, the coefficients are $\beta = -0.046$ in 2018 and $\beta = -0.057$ in 2019, while for JAL, the coefficients are $\beta = -0.057$ in 2018 and $\beta = -0.029$ in 2019. This trend persists when focusing exclusively on hub airports, where higher EC is associated with shorter departure delays at the 1% significance level. For ANA, the co-efficient is $\beta = -0.011$ in 2018, while for JAL, the coefficients are $\beta = -0.023$ in 2018 and $\beta = -0.049$ in 2019. These findings suggest that, during the pre-COVID phase, higher EC at hub airports was more likely to be linked to reduced delay durations across the delay network. When an airport with a higher number of departure delay links is connected to another airport that also has a higher number of departure delay links, the departure delay duration tends to decrease as EC increases. A higher EC in a delay network may indicate the propagation of delays across connected airports (Kang et al., 2023). This occurs because when one higher-degree node is connected to another high-degree nodes, information can be disseminated more rapidly across the network. In the context of our study, the increased EC in the delay network generally leads to a reduction in delay duration on a daily basis. One possible explanation for this is that some airports exhibit higher EC values in the delay network each day, indicating that they are more likely to have a greater number of delay links connected

to other airports with similarly high delay links. Consequently, these delay links tend to produce shorter delay durations. Consequently, the total departure delay duration per day is typically shorter than what would be expected for that airport. Therefore, it is plausible that airports in Japan are more likely to produce shorter delay propagation across the delay network; however, the frequency of these shortened delays could be higher because of higher EC. Jia et al. (2022) reported that, in China, larger airlines are more likely to propagate delays slowly but cover a broader area in aviation networks, although they did not report its connection with the duration of the delay. The interaction term between the EC and the hub indicates that hub airports are less prone to longer daily departure delays than spoke airports because of the negative association. This further implies that spoke airports with higher EC are more prone to longer departure delays, which aligns with the BC result that spoke airports may also act as critical airports that generate a significant number of departure-delayed flights with longer durations. For both ANA and JAL, spoke airports, namely KMI, HIJ, KOJ, MYJ, and KMJ, were ranked in the top ten lists of EC rankings (Figure 3.6 and 3.7), with the potential to increase departure delay duration. Hub airports in Japan are more likely to have higher EC in terms of delay, such as HND, OKA, ITM, FUK, and CTS (as indicated in Figures 3.6 and 3.7), and are connected to other well-connected airports. Here, a potential inconsistency may arise in the implications of in-degree centrality and EC concerning their relationship with delay duration. The positive association between in-degree centrality and delay duration suggests that delays may accumulate at the receiving hub airport since delayed flights arrive from both hub and spoke airports. Consequently, the receiving hubs exhibit a higher number of delay connections. Conversely, in the case of EC, well-connected airports with a higher number of delay connections tend to be linked with other similarly well-connected airports. However, not all airports exhibit high EC values. Fewer airports are strongly interconnected in terms of delay propagation. Consequently, the negative association between EC and delay duration could refer to shorter departure delay durations among these few airports, facilitated by better infrastructure, more efficient ground operations, and well-established procedures for handling irregularities. These capabilities enable faster recovery from delays. Although some airports may be critical in terms of EC, they generally experience shorter departure delay durations. Conversely, the spoke airports mentioned above have higher EC in terms of delay because of a lack of resources, like the hub airports, and thus, are more likely to produce longer departure delay durations. It is important to note that, for ANA in 2019, the interaction between hub airports and EC showed a positive association with delay duration ($\beta = 0.048$) at the 1% significance level. As previously mentioned, the presence of an outlier in 2019 may have altered the sign of the coefficient, thereby affecting the interpretation of the results.

In the case of transitivity, both airlines in 2019 exhibited a negative association between departure delay duration and transitivity, suggesting a tendency to experience shorter delay durations. Specifically, for ANA, the coefficient was $\beta = -0.001$ at the 1% significance level, while for JAL, the coefficient was $\beta = -0.0001$, though the latter was not statistically significant. This relationship also holds for ANA in 2018. However, a positive association was observed for JAL in 2018. The transitivity property behaves differently for JAL when focusing on hub airports. For ANA, the association was negative, with $\beta = -0.007$ in 2018 and $\beta = -0.004$ in 2019, both at the 1% significance level. In contrast, for JAL, the association was positive, with $\beta = 0.005$ in 2018 and $\beta = 0.006$ in 2019, both statistically significant at the 1% level. For JAL, these results suggest hub airports are more likely to form triads with two other airports that frequently experience departure delays and are more likely to encounter longer delay durations.

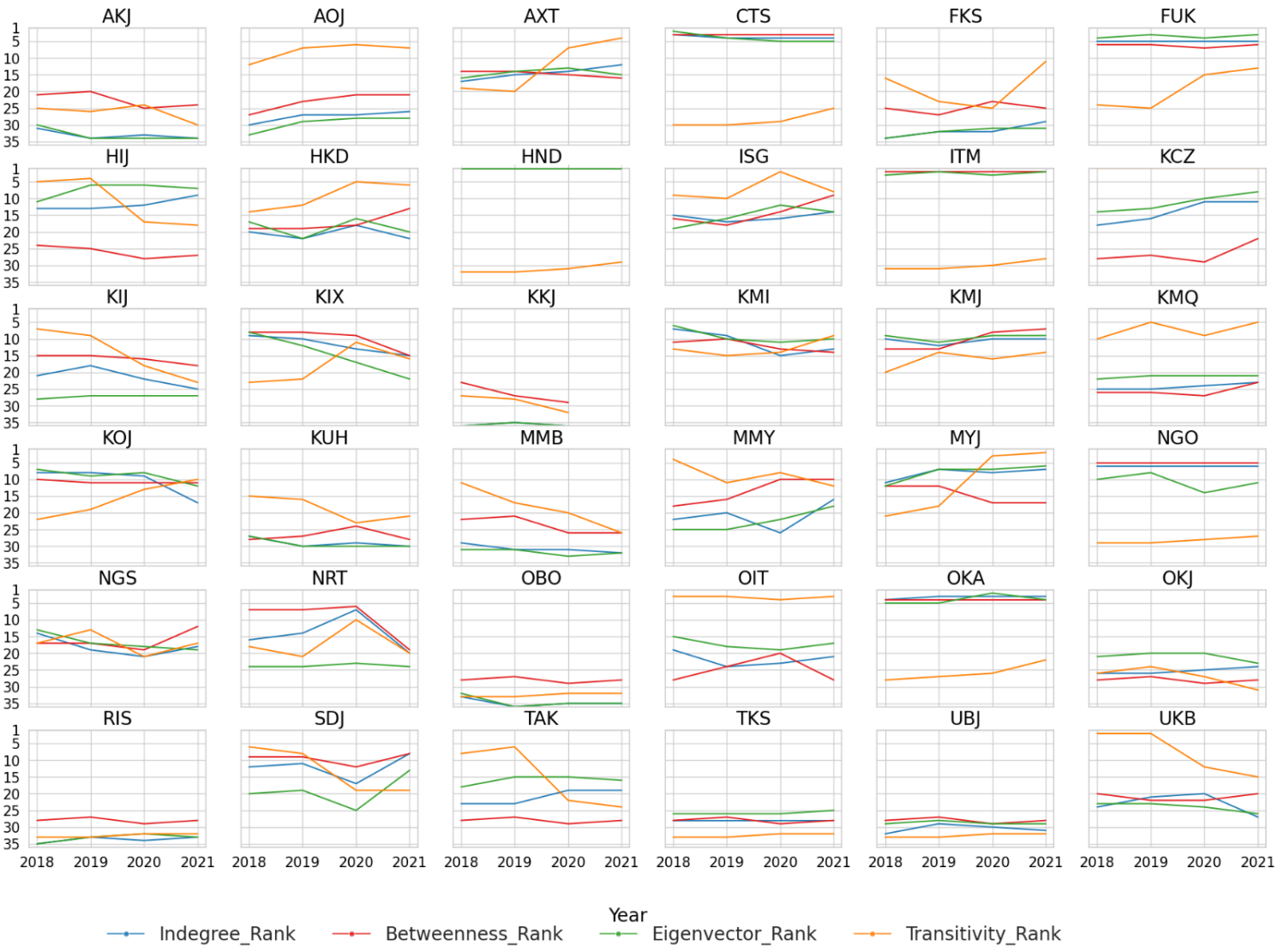


Figure 3.6: Ranking of departure delay network properties for airports connecting ANA flights from 2018 to 2021

In other words, higher transitivity is associated with an increased likelihood of delays within airport triads. This implies that if an airport in a tightly connected group (via delay edges) experiences departure delays, these delays are more likely to propagate within the group, thereby creating localized triads of delayed flights. Consequently, departure delay duration tends to increase among interconnected airports. For example, for ANA, airports such as KCZ, UKB, OIT, HIJ, TAK, KMQ, KIJ, and ISG ranked in the top 10 transitivity rankings in both 2018 and 2019, indicating that these airports are more likely to have higher transitivity, meaning that they are more likely to form departure delay triads. Similarly, for JAL, spoke airports such as KMI, MYJ, SDJ, AOJ, NGS, and KMJ were found in the top 10 transitivity ranking lists in both pre-COVID years. This result aligns with that of Lonzius and Lange (2024), who indicated that routing strength, which focuses on frequent flights between certain airports, is associated with longer delays. Although arrival delays were examined, the airport-cluster analogy coincides with our results. Li et al. (2021) found that among smaller airports, delay clustering is prominent in the US aviation network. In addition, Zhou et al. (2022) reported that the flight network in China tends to create dense delay propagation clusters. Given both the positive and negative associations between transitivity and departure delay duration, it may be important to monitor certain airports on a daily basis where departure delay links are produced at a higher rate within themselves. In this context, a shorter departure delay duration should not be overlooked, as in certain situations, this shorter duration may shift to a longer duration owing to various unobserved factors.

3.4.2.2 During-COVID Phase

In terms of single-variable regression model (Table 3.7), for ANA, in 2020, in-degree has the largest coefficient (0.57) and explains about 41% of the variation in delay duration, indicating that airports with many incoming delay connections suffer longer departure delays. BC is also positively associated with delay duration (0.32) but with less explanatory power ($R^2 = 0.25$). EC has a small positive effect (0.07; $R^2 = 0.08$). Transitivity is slightly negative (–0.01) with negligible explanatory power ($R^2 \approx 0.01$), suggesting that increased clustering of an airport's neighbours may marginally reduce delays. Intercepts range from 0.000 to 0.025 and ρ values from 0.19 to 0.39 indicate moderate serial correlation. All models have p-values <0.001. For 2021, in-degree remains significant but with a smaller coefficient (0.37) and $R^2 = 0.34$. BC's effect declines to 0.07 ($R^2 = 0.03$). EC (0.01; $R^2 = 0.01$) and transitivity (~0.0001; $R^2 = 0.01$) show negligible effects. Intercept estimates range from –0.003 to 0.007. High ρ values (0.75–0.87) imply strong autocorrelation in delays.

Table 3.7: Estimation result for single variable regression at the phase of during-COVID (2020 and 2021)

Airline	Variables	2020				2021			
ANA	In-degree	0.57				0.37			
	BC		0.32				0.07		
	EC			0.07				0.01	
	Transitivity				-0.01				0.0001
	Constant	0.000	0.012	0.005	0.025	-0.003	0.005	0.005	0.007
	ρ	0.19	0.30	0.33	0.39	0.75	0.87	0.84	0.78
	R-squared	0.41	0.25	0.08	0.01	0.34	0.03	0.01	0.01
	Prob > chi2	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
	n	7,389				7,828			
	Number of airport	36				35			
	Number of time (range)	3-365				18-365			
JAL	In-degree	0.50				0.47			
	BC		0.27				0.11		
	EC			0.07				0.02	
	Transitivity				-0.001				0.001
	Constant	-0.003	0.006	-0.001	0.013	-0.004	0.003	0.002	0.004
	ρ	0.15	0.24	0.28	0.40	0.69	0.83	0.78	0.81
	R-squared	0.63	0.44	0.14	0.01	0.49	0.07	0.01	0.01
	Prob > chi2	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
	n	9,671				11,229			
	Number of airport	44				46			
	Number of time (range)	22-363				21-364			

NB: All the variables are statistically significant at 5% significance level (Grey shading represents insignificant variable)

In case of JAL (Table 3.7), the results include, for 2020, in-degree has a positive coefficient (0.50) with a high R^2 of 0.63, making it the strongest predictor. BC (0.27; $R^2 = 0.44$) also shows a substantial association. EC (0.07) has modest explanatory power ($R^2 = 0.14$). Transitivity has a slight negative coefficient (–0.001) and near-zero R^2, again suggesting that clustering may modestly mitigate delays. Constants range from –0.003 to 0.013 and ρ values (0.15–0.40) indicate mild serial correlation. All $p\text{-}values$ are < 0.001. Additionally, for 2021, in-degree continues to be important (coefficient 0.47; $R^2 = 0.49$). BC's effect reduces (0.11; $R^2 = 0.07$), while EC (0.02) and transitivity (0.001) remain tiny with R^2 around 0.01. Intercepts range from –0.004 to 0.004, and ρ values (0.69–0.81) point to strong serial correlation. Models are significant at $p < 0.001$.

Across both airlines and years, in-degree centrality consistently exhibits the largest positive coefficients and explains a substantial portion of the variance in delay durations. This suggests that airports that receive delays from many other airports experience longer departure delays, highlighting the importance of inbound connections in delay propagation. BC also contributes positively but to a lesser extent; airports acting as bridges within the network tend to accumulate delays. EC and transitivity have minor or

negligible effects, indicating that the broader influence of connected neighbours and local clustering play limited roles in determining delay duration. The large, significant ρ estimates confirm that serial correlation is present and justify the use of the Prais–Winsten estimator

For multi-variable Prais-Winsten regression, Table 3.8 presents the estimation results of network properties during the COVID-19 phase, with the only distinction from the pre-COVID analysis being the inclusion of seasonality as a categorical variable (i.e., January to March as Q1, April to June as Q2, July to September as Q3, and October to December as Q4). This categorization allowed for an understanding of the differences in departure delay networks resulting from the four rounds of SoEs declared from 2020 to 2021 during the COVID-19 pandemic in Japan. From Table 3.8, it is evident that during the initial phase of the COVID-19 outbreak in 2020, in-degree centrality was positively associated with departure delay duration for both ANA ($\beta = 0.629$) and JAL ($\beta = 0.662$) at the 1% significance level, indicating an increase in departure delay duration. When focusing on hub airports, both airlines exhibited a tendency toward longer departure delays (significant positive association) in 2020 and 2021; however, statistical significance at the 1% level was only observed in 2021, with $\beta = 0.162$ for ANA and $\beta = 0.192$ for JAL. The results suggest that during the COVID-19 pandemic, flights departing late from origin airports were more likely to cause longer departure delays at destination airports. Figure 3.3 illustrates the significant decrease in the number of departure delayed flights for both ANA and JAL during the first SoE. These results indicate that, although the number of delayed flights decreased during these time periods, the departure delay duration tended to increase.

Table 3.8: Estimation Result at the phase of during-COVID (2020 and 2021)

Variables	**ANA**		**JAL**	
	2020	**2021**	**2020**	**2021**
In-degree	0.629***(0.056)	-0.094 (0.086)	0.662***(0.042)	-0.102 (0.078)
In-degree*Hub	-0.039 (0.049)	0.162***(0.016)	0.052 (0.032)	0.192***(0.020)
In-degree*Q2	-0.399* (0.238)	0.054 (0.115)	-0.324* (0.181)	0.043 (0.109)
In-degree*Q3	-0.129 (0.126)	0.080 (0.108)	-0.046 (0.104)	0.114 (0.099)
In-degree*Q4	-0.173* (0.094)	0.935***(0.095)	-0.421***(0.089)	1.03***(0.090)
Betweenness	-0.129* (0.075)	0.013 (0.027)	-0.016 (0.022)	0.065***(0.024)
Betweenness *Hub	0.081 (0.074)	-0.005 (0.018)	-0.064***(0.019)	-0.064***(0.013)
Betweenness *Q2	0.141 (0.087)	-0.003 (0.03)	0.146 (0.094)	-0.011 (0.038)
Betweenness *Q3	0.156***(0.045)	-0.007 (0.032)	0.113***(0.041)	-0.027 (0.034)
Betweenness *Q4	0.228***(0.042)	-0.185***(0.031)	0.156***(0.037)	-0.237***(0.033)
Eigenvector	-0.017* (0.011)	0.009 (0.007)	-0.043***(0.008)	0.010* (0.006)

Eigenvector *Hub	0.029***(0.007)	-0.012***(0.003)	-0.004 (0.005)	-0.013***(0.003)
Eigenvector *Q2	-0.012 (0.414)	-0.005 (0.011)	0.041***(0.011)	-0.003 (0.008)
Eigenvector *Q3	-0.035**(0.015)	-0.011 (0.012)	-0.00007 (0.012)	-0.010 (0.009)
Eigenvector *Q4	-0.050***(0.016)	-0.114***(0.013)	0.028**(0.013)	-0.112***(0.010)
Transitivity	-0.002 (0.002)	0.0007 (0.0007)	-0.002***(0.0007)	0.0008 (0.0005)
Transitivity*Hub	-0.001 (0.003)	-0.002* (0.002)	0.004***(0.002)	-0.0004 (0.641)
Transitivity*Q2	0.0007 (0.003)	0.0006 (0.001)	0.001 (0.002)	-0.0001 (0.0007)
Transitivity*Q3	0.005**(0.002)	0.002 (0.001)	0.004***(0.001)	-0.0003 (0.0006)
Transitivity*Q4	0.007***(0.002)	-0.002**(0.001)	0.005***(0.001)	-0.002**(0.0007)
Constants	0.007***(0.001)	0.0006 (0.0004)	0.0009**(0.0004)	0.0002 (0.0002)
Number of Observations	7,389	7,828	9,671	11,229
Number of airport entities	36	35	44	46
Number of time entities (range)	3-365	18-365	22-363	21-364
R-squared	0.43	0.73	0.68	0.79
Prob > chi2	0.000	0.000	0.000	0.000

(In the parenthesis standard errors are reported in parentheses. The star marks represent the statistical significance level at '***' 1% '**' 5% '*' 10%)

Figures 3.6 and 3.7 confirm that the hub airports were ranked among the top for in-degree centrality, specifically HND, ITM, OKA, CTS, and FUK. During the COVID-19 pandemic in the United States, flights arriving late at hub airports were more likely to experience increased delays, with gate delays increasing substantially compared to the pre-COVID phase, potentially owing to labor shortages that may have restricted airlines from achieving on-time departures (Yimga, 2021; Xu et al., 2024). Additionally, the turnaround time during the COVID-19 period was reported to have increased at terminals, which may have contributed to longer ground times, further exacerbating departure delay durations at receiving airports despite a decrease in the number of flights and passengers (Schultz et al., 2020).

When accounting for seasonality, both ANA and JAL exhibit similar associations in 2020 and 2021, with consistent coefficient signs across each year. In 2020, statistically significant results were found only for Q2 and Q4, where a negative relationship was observed between in-degree centrality and departure delay duration. For ANA, the coefficients were -0.399 in Q2 and 0.173 in Q4, whereas for JAL, the coefficients were -0.324 in Q2 and -0.421 in Q4. In 2021, the relationship shifted for both airlines, with all quarters from Q2 to Q3 showing positive associations with delay duration, indicating longer delays compared to Q1. However, statistical significance at the 1% level was only observed in Q4, with $\beta = 0.935$ for ANA

and $\beta = 1.03$ for JAL. At the onset of the COVID-19 pandemic, the reduction in flight numbers and departure delayed flights may have contributed to shorter departure delays for both airlines. However, in 2021, as operations began to recover following the fourth SoE, longer delays emerged within the expanded delay network connections, a pattern resembling the pre-COVID phase. This outcome, however, was only evident when considering seasonality variation.

For both ANA and JAL, BC was negatively associated with departure delay duration, indicating shorter departure delays in 2020, with $\beta = -0.129$ for ANA and $\beta = -0.016$ for JAL. However, only JAL's estimate was statistically significant at the 10% level. In 2021, this trend reversed, with BC being positively associated with longer de-lay durations for both airlines ($\beta = 0.013$ for ANA and $\beta = 0.065$ for JAL), although only JAL's result was statistically significant at the 1% level. When focusing solely on hub airports, flights operated by JAL showed a negative association with departure delay duration in 2020 ($\beta = -0.064$ at the 1% significance level) and 2021 ($\beta = -0.064$ at the 1% significance level), although the standard error differed between the two years. For ANA, the association was not statistically significant. These findings suggest that during the COVID-19 phase, hub airports with higher BC are more likely to experience shorter departure delays, while spoke airports with higher BC tend to experience longer delays. Figures 3.6 and 3.7 show that KOJ, HKD, OKD, KUH, MSJ, and IZO ranked in the top ten for BC in both years. This indicates that, during the COVID-19 phase, spoke airports were more likely to act as critical nodes in terms of departure delays, making them more vulnerable to longer delay durations. These airports may have lacked the capacity and personnel to enforce stringent COVID-19 measures, such as temperature screening, physical distancing, and restricted terminal access (WHO, 2020; Narita Airport, 2020). Consequently, these airports may have become more susceptible to delays in connecting with other airports, contributing to longer delay durations. Notably, during the first SoE in Q2 2020, both ANA and JAL experienced an increase in delay duration associated with higher BC. A similar pattern was observed in Q3 and Q4 of 2020. During this period, the total number of flights and departure delayed flights decreased (Figure 3.3). In 2021, BC exhibited a negative association with departure delay duration across Q2, Q3, and Q4 for both airlines. Notably, the Q4 results were statistically significant at the 1%, suggesting that higher BC contributed to shorter delays for both ANA and JAL flights. Consequently, both ANA and JAL displayed an opposite relationship between 2020 and 2021 when seasonality was considered, with the initial phase of COVID-19 tend to increase delay duration, while late

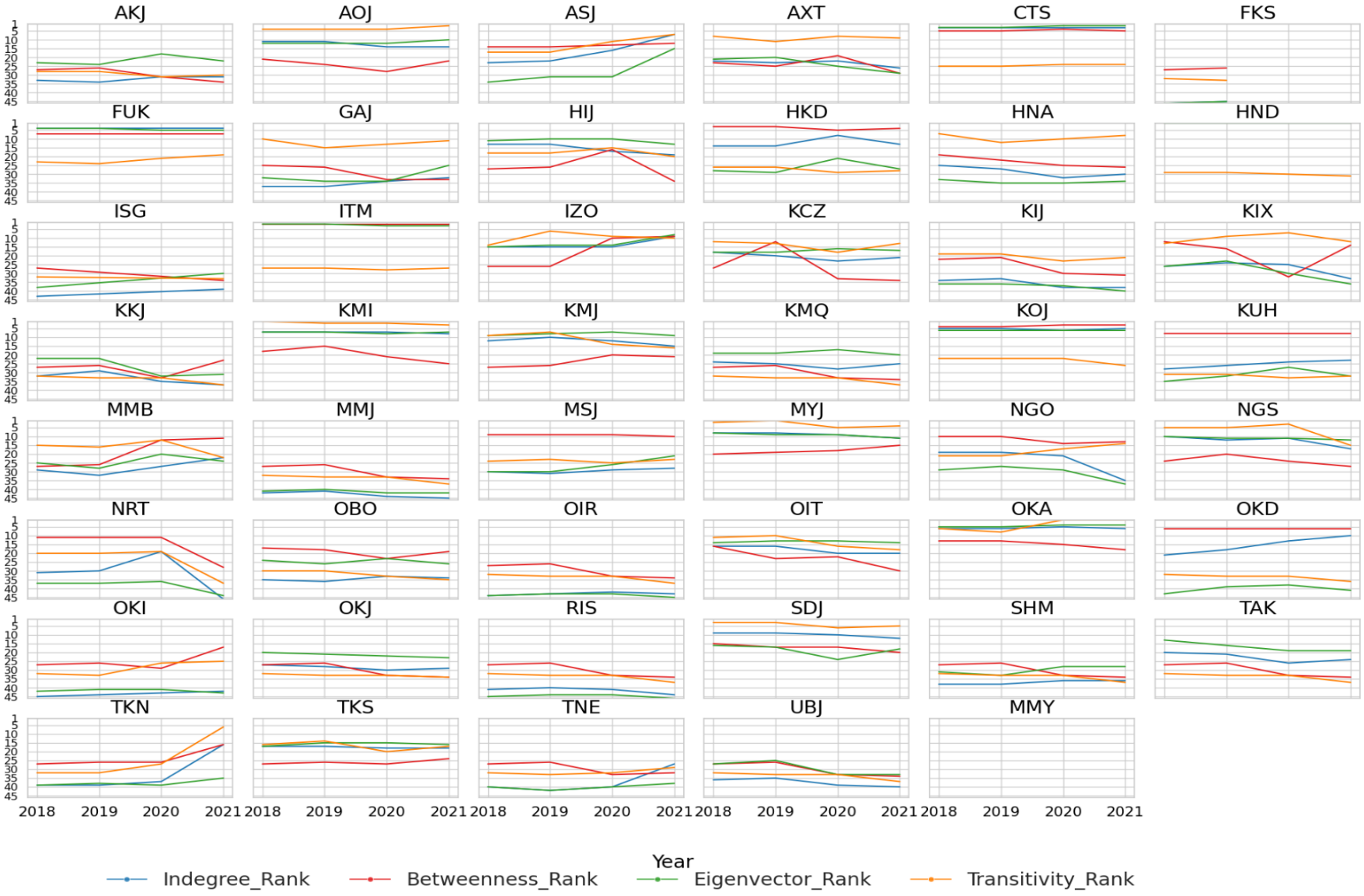


Figure 3.7: Ranking of departure delay network properties for airports connecting JAL flights from 2018 to 2021

years showed a tendency for decreased delay duration. The three rounds of SoEs in 2021 may have influenced these shifts in the relationship, compared to 2020, when only one SoE occurred.

EC was generally negatively associated with departure delay duration in 2020, indicating a reduction in delay duration, with coefficients of $\beta = -0.017$ for ANA and $\beta = -0.043$ for JAL. A similar pattern was observed when focusing solely on hub airports in 2021, with coefficients of $\beta = -0.012$ for ANA and $\beta = -0.013$ for JAL, reflecting a consistent negative relationship. However, in 2021, the opposite trend emerged, showing a positive association between EC and departure delay duration. For ANA, the coefficient was $\beta = 0.009$, though it lacked statistical significance, while for JAL, the coefficient was $\beta = 0.010$, with significance at the 10% level. During Q2 of 2020, which coincided with the first SoE, and Q4 of the same year, JAL experienced increased delay durations. In 2021, the coefficients for both airlines remained consistent across all quarters, indicating a persistent negative association between EC and departure delay duration throughout each quarter and SoE. The interaction with seasonality indicates that as flight and delay networks gradually recovered in 2021, airports began exhibiting delay-based EC patterns resembling the pre-COVID times, characterized by frequent but shorter-duration delay propagation across the network.

In the case of transitivity, delayed flights operated by both ANA and JAL exhibited shorter departure delay durations in 2020, but experienced longer delays in 2021. A similar trend to that observed in 2020 was evident in 2021 when focusing solely on hub airports, with $\beta = -0.002$ at the 10% significance level for ANA and $\beta = -0.0004$, which lacked statistical significance, for JAL. Regarding seasonality, transitivity contributed to longer departure delay durations in Q2, Q3, and Q4 of 2020. This suggests that following the declaration of the first SoE on April 7, 2020, clusters of airports emerged and circulated delays among themselves, thereby leading to increased delay durations. However, in Q4 of 2021, both ANA and JAL demonstrated a negative association between transitivity and departure delay, indicating that higher transitivity during that period corresponded to shorter delay durations.

From the preceding discussion, an important pattern during the COVID-19 phase was identified: the association between network properties and departure delay duration differed in 2021 compared to 2020 and pre-COVID. During the initial phase of the COVID-19 pandemic in 2020, the number of domestic and delayed flights decreased significantly (Figure 3.3), particularly following the declaration of the first SoE. After the end of the first SoE, the frequency of delayed flights increased slightly and remained moderately perturbed until the fourth SoE in 2021. Despite the declaration of three consecutive SoEs in 2021, the frequency of delayed departure flights remained higher than that in 2020. This could explain the changes in the observed relationship patterns. Additionally, during 2021, a recovery trend was noted in the total number of flights and delayed departure flights. Consequently, the coefficient signs may have

also shift-ed compared with 2020. In 2021, we observed variations in the coefficient signs of seasonal interactions with network properties. This outcome is expected because network recovery following a disruption may lead to perturbed characteristics during the entire network reformation process. Furthermore, it is possible to obtain both intuitive and counterintuitive results (Gao et al., 2016; Moustsinas and Guo, 2020) that align with our findings regarding the regression coefficients of all network properties.

To test the coefficient stability across different years from 2018 to 2021, we tested Wald test for coefficient stability. We tested the joint null hypothesis: $All\ interaction\ terms\ (year \times network\ properties) = 0$; we tested whether the effect of in-degree, betweenness, eigenvector and transitivity stay constant across the years from 2018 to 2021. For ANA, the Wald test for coefficient stability across years yields that the Wald chi-square test statistics is $973.74\ (p - value < 0.001)$ and for JAL, the test statistics is $737.17\ (p - value < 0.001)$ which indicates that the effects of network properties on total departure delay duration vary over time. So, it is implied that network-centric predictors do not hold temporal stability; rather it requires year-specific model. In addition, when we compare the coefficient stability between ANA and JAL over time, we get Wald test statistics as $298.95\ (p - value < 0.001)$ which implies that ANA and JAL exhibit structurally different delay network characteristics when associated with departure delay duration. These findings emphasize the need for airline-specific delay management strategies. Airlines should not rely on historical averages and generic models to predict or manage delay risks. Instead, they should develop adaptive models that account for the temporal evolution of their network structure, operational behavior and external disruptions.

Moreover, the single-variable coefficients vary little across years. In-degree, BC and EC remain positive and significant in both pre- and during-COVID periods, while transitivity stays small and negative. This stability occurs because each model isolates a direct relationship and avoids the confounding influence of other network measures. By examining one predictor at a time, the models are not affected by correlations among predictors, so the estimated slope captures a consistent average association between the network property and delay duration. In contrast, the multivariable regressions are noticeably unstable. Coefficients change magnitude or sign across years and between airlines, and standard errors increase. The probable reasons could be - the networks experienced structural changes. In 2019, an outlier (a surge in delayed flights) appears to have flipped the interaction effect of $hub \times in\text{-}degree$ for ANA, and during COVID-19 the relationship between network properties and delays changed as flight schedules and delay patterns were drastically altered. Multi-variable models are more sensitive to such outliers and temporal changes because they estimate multiple interrelated effects simultaneously. The multivariable

models include hub interactions and seasonal dummies to capture differences between hub and spoke airports and between quarters. These terms can alter the coefficient on a network measure; for example, BC becomes negative when interacted with hub status, implying shorter delays at high-BC hubs but longer delays at high-BC spoke airports. Such heterogeneity increases the complexity of the model and contributes to instability across years. Lastly, the Wald test suggests that the relationship between network position and delay duration is not temporally stable; it evolves with changes in traffic volume, scheduling and external shocks such as COVID-19. Single-variable models mask much of this heterogeneity, whereas multi-variable models, by isolating partial effects, reveal these shifts and thus appear less stable.

We also checked results the results for 1 minute of threshold with 15-min threshold of departure delay that usually airports take as a disruption indicator. The same regression analysis was carried out. The results are reported in Appendix B (Table B1-B3). The results are identical to that of 1-minute threshold limit in terms of coefficient instability across the years.

3.5 Chapter Conclusion

This study was conducted to examine the relationship between network properties and departure delay duration in Japanese domestic aviation, focusing on two full-service carriers, JAL and ANA, from 2018 to 2021, considering both the pre-COVID-19 and during-COVID-19 phases. The analysis examines various network properties that characterize the daily departure delay network of connecting airports and investigates the relationship between these properties and the departure delay duration using a Prais–Winsten regression model with fixed effects, accounting for serial correlation, heteroskedasticity, and contemporaneous correlation.

The key findings from the analysis indicate that, during the pre-COVID-19 phase, higher in-degree centrality was associated with longer departure delays for both ANA and JAL, suggesting that airports receiving more delayed flights tended to experience longer delays. During the COVID-19 phase, for hub airports, a positive association between in-degree centrality and delay duration was observed for both airlines. This association changed to shorter durations for both ANA and JAL in 2021. This result indicates not only compounding delays at the receiving airports but also the longer departure delay durations that may translate into network-wide cascading delays if airline fleets are tightly scheduled at the receiving airports. In contrast, increased betweenness centrality generally corresponded to reduced delay durations at hub airports during the pre-COVID-19 phase. However, certain spoke airports, such as SDJ, KOJ, and KMI for ANA and HKD and KOJ, OKD, KUH, and KMI for JAL, may have acted as critical airports in terms of delays and were more likely to experience longer departure delays. During the COVID-19 phase, betweenness centrality was negatively associated with delays in 2020 but positively associated with delays for both ANA and JAL in 2021. Certain spoke airports, such as KOJ, HKD, OKD,

KUH, MSJ, and IZO, acted as critical airports that tended to experience longer delays. During the pre-COVID-19 phase, eigenvector centrality was associated with shorter departure delays for both airlines, indicating that airports connected to highly central airports experienced frequent but shorter delay durations and were more likely to propagate delays with shorter durations among connected airports. In the during-COVID-19 phase, eigenvector centrality yielded results similar to those observed during the pre-COVID-19 phase. Transitivity effects vary by airline and whether one focuses on hubs or the entire network. ANA often exhibited a negative association (more transitive triads linked to shorter delays), whereas JAL at times showed a positive association (higher transitivity associated with longer delays). Focusing only on hubs yielded contrasting signs, airline-specific network structures. In 2020, higher transitivity mostly coincided with shorter delays, potentially owing to overall fewer flights and a partial network collapse. By 2021, higher transitivity often reflected localized clusters of delayed flights, lengthening delays in those triads. Several quarters (especially Q3 and Q4 of 2020) showed that transitive clusters could prolong delays owing to localized COVID-related disruptions.

This comparative result shows that the key network properties generally hold consistent explanatory power for delay duration under normal (pre-COVID-19) conditions. During COVID-19, the relationships became more unpredictable, driven by external factors such as SoEs and reduced flights. These patterns during the pandemic period underscore how extreme disruptions can affect delay networks and their relationship with delay duration. One might assume that fewer flights during COVID-19 inherently reduce overall delays. Our results challenge this assumption by demonstrating that smaller network connections can sometimes experience proportionally longer delays. Regulators should be cautious in applying uniform delay-related policies across different airlines. As structural characteristics and responses differ by carrier and over-time, policies aiming to manage and mitigate systemic delays will be more effective if they incorporate airline-level and time-varying features.

From the above discussion, our study of how network properties relate to flight departure delay duration shows that, in some cases, higher values of properties are linked to longer delays, and, in some cases, higher values seem to be connected with shorter delays. Understanding these associations can significantly assist Japan's Air Traffic Management Center (ATMC) in reducing the risk of network-wide disruptions through collaborative decision-making with other airports (Nakagawa, 2012). In Japan, Fukuoka ATMC oversees domestic air traffic flows, such as flow control and route management, in coordination with three Area Control Centers located in Tokyo, Kobe, and Fukuoka (Nakagawa, 2012; Japan Aviation Hub, 2024). These flow-control measures include ground delay programs, minutes-in-trail, ground stops (for pre-flight), miles-in-trail, speed adjustments, and airborne holding (for in-flight conditions) (ICAO WP Agenda 32, 2022). From our findings, if the in-degree centrality of delayed flights

at the origin airports increases, this typically signals a buildup of delays at the receiving airports. When the number of delayed flights from either the hub or spoke airports is high, the ATMC can recommend speed reductions or increase miles-in-trail on inbound flights to mitigate the overall delay durations at the destination. In terms of betweenness centrality, even airports with lower in-degree (often spokes) can still spread long-lasting delays when their betweenness is high and positively linked to departure delay duration. By tracking it, ATMC can track airports that are most likely to amplify or transfer delays to other airports. Even if these airports do not have a higher number of delayed flights (lower in-degree), their bridging role makes them crucial in controlling the overall delay propagation. Meanwhile, when eigenvector centrality shows a positive link with delay duration, it flags the risk of extensive delays spread across highly interconnected airports with lengthy durations. Here, real-time monitoring and strategic adjustments - such as modifying flight schedules, holding patterns, or route assignments - could help mitigate the downstream impact of delays. Conversely, instances in which a negative association exists may reveal opportunities to reduce the delay impacts since they imply a shorter duration of delay propagation among fewer but highly interconnected airports. Understanding these different effects enables the ATMC to design targeted measures to prevent network configurations from worsening delays and to take advantage of those that help reduce them. Finally, a positive link between transitivity and delay duration means that groups of airports may fall into a loop in which delays reinforce each other over a longer duration. In these cases, ATMC may consider some targeted actions, such as adjusting flight spacing that could help break the loop and reduce disruptions. Conversely, a negative relationship suggests that some airport clusters are better at dispersing delays with shorter delay durations, offering practices that could be adopted in more vulnerable areas of the network where the delay durations are high among the airport clusters. These results can be applied in parallel with Japan's ATMC. We recommend that delay-related network properties be monitored collectively rather than in isolation and further emphasize the correlation of these properties with each delayed flight's duration. We believe that these network-based properties of delay networks help shift to trajectory-based air traffic management operations from the traditional airspace-based operation that the Japan Civil Aviation Bureau has envisioned to establish CARATS: Collaborative Actions for Renovation of Air Traffic Systems in Japan (MLIT, 2024; Hirata, 2013). This enabled a clearer assessment of the number of flights affected and the overall severity of delays, allowing for more effective intervention measures.

This study has certain limitations along with various opportunities for future re-search. One key limitation is that it establishes only a regression-based association between network properties and departure delay duration rather than identifying a causal relationship. As observed in the findings discussed earlier, the association between network properties and delay duration often exhibited inconsistencies across different years and among the various network metrics. This finding suggests the dynamic nature of this

relationship. Furthermore, the potential influence of confounding factors may have contributed to the observed inconsistencies in the associations between delay network metrics. To address these limitations, future research should focus on establishing a causal relationship between network properties and delay duration by incorporating causal inference-based methodologies (e.g., graphical causal models and do-calculus) to account for confounding effects and temporal variations. In addition, this study does not consider the entire departure delay network connecting all JAL and ANA flights as a single network for analysis. Low-cost carriers are also popular in Japan, and a future comparison of the departure delay network properties between full-service carriers and low-cost carriers would be valuable. Weather data that could have provided further insight were not included in this study. Future research could explore the analysis of arrival delays and their connection to departure delays to determine how they both contribute to the overall network, offering a more realistic understanding of delay network characteristics and their eventual propagation. This study is designed for daily delayed networks and durations. The proposed framework can be applied to investigate the relationship at more granular timestamps, such as 1 h, 4 h, and 6 h delay networks. Moreover, the current study framework should be tested and transferred to other delay networks in other regions to establish a generalization of the outcomes that may eventually lead to a causal explanation of this relationship. In particular, in the USA and UK, where centralized air traffic management may incur substantial delays (Campanelli et al., 2016), this study can help assess the vulnerability of the delay network by leveraging both the delay network properties and the corresponding delay duration. In addition, comprehensive and quality-controlled flight delay data beyond 2021 are unavailable during the analysis period. Therefore, we limited our objective to examining the relationship between normal operations (pre-COVID-19) and operations during a major external shock (COVID-19). This allowed us to highlight how key network properties behaved under stable conditions versus extreme disruptions. A future study investigating the entire post-pandemic period from 2022 onward could enhance our understanding of long-term delay network resilience. While this study focused on positive delay occurrences (≥ 1 minute) to better capture disruption propagation and avoid zero-inflation, future research could expand the framework to incorporate zero-delay observations. Including days with no aggregated delay would allow for analyzing the probability of experiencing a delay (incidence) and the magnitude of delay (severity) simultaneously. This extension could provide a comprehensive understanding of on-time performance, schedule robustness, and the conditions under which network centrality influences not only delay severity but also the likelihood of a delay occurring.

Chapter 4 Uncertainty Quantification of Departure Delay considering Network Properties and Conformal Prediction Framework

4.1 Introduction

Flight delays create significant challenges for the aviation industry due to monetary loss and complexity of various interconnected factors within aviation systems (Pyrgiotis et al., 2013). Delays further lead to airport capacity-demand imbalance, consumer dissatisfaction (e.g., wasted passenger time), economic losses, and often increased environmental emissions. Therefore, predicting and managing delays are widely recognized as a highly complex task due to numerous uncertain factors (ACRP, 2014).

The global air traffic has significantly grown at an annual rate of 3.2% from 2010 to 2019 (Airports Council International, 2019). Although this growth was severely disrupted by the COVID-19 pandemic, air travel demand has rapidly rebounded which later highlighted the existing constraints in airport and airspace capacities. These constraints often lead to bottlenecks, especially during peak periods and unexpected disruptions (Bieger et al., 2007; Sun et al., 2023). Airport Capacity Management plays a critical role in addressing these issues by controlling demand and supply through various planning stages. These stages include long-term infrastructure planning, strategic flight scheduling through slot allocation (months in advance), preparation of flight plans (weeks ahead), adjustments based on day-ahead information, and real-time tactical decisions to manage immediate operations (Ribeiro et al., 2025).

Delays remain difficult to predict due to inherent uncertainties stemming from unpredictable weather conditions, aircraft maintenance issues, airport and airspace congestion, and cascading effects within interconnected flight networks (Birolini and Jacquillat, 2023). Adverse weather conditions, such as heavy rain, thunderstorms, and low visibility affect both airport and airline operations. Furthermore, airport infrastructure is typically limited and often operating near full capacity, increasing delays when disruptions occur (Lee and Balakrishnan, 2012). In particular, passengers are often sensitive to delays, whether during departure or arrival (Cook et al., 2009).

To anticipate and predict delays effectively, various approaches have been developed. For example: simulation models (Odoni, 1997, Ribeiro et al., 2025), queueing models (Shone et al., 2021), and machine learning models - either independently or combined with other techniques (Carvalho et al., 2021). Schedule delays are particularly critical in airline planning. By predicting these delays accurately with some level of confidence, both airlines and airports can proactively incorporate buffer times, reducing the potential cascade severity at the downstream side (Birolini and Jacquillat, 2023).

Recent studies emphasize the interconnected nature of aviation systems by using network-based methods like graph theory and complex network analysis (Wandelt et al., 2025a). Unlike traditional methods that focus on individual airports or flights, network-based analyses explore how delays propagate through interconnected airports and schedules. The aviation network, consisting of nodes (airports) and edges (flight/ delayed flight connections), highlights critical elements and interactions that significantly influence delay propagation (Wandelt et al., 2025a; Sano and Berton, 2021; Tang et al., 2021; Sugishita et al., 2024). For instance, in the United States, delays frequently accumulate throughout the day that emphasizes their temporal dependency (Wandelt et al., 2025b).

Departure delays, specifically, are crucial because they often trigger cascades of delays throughout the flight network. Departure delays result from multiple interacting factors, including airline operational strategies, airport management, air traffic control, weather disruptions, and human errors (Wandelt et al., 2025a). The interconnected scheduling of flights increases the likelihood that a delay in one flight will cascade to subsequent flights and across airports.

Typically, when forecasting delays, the common practice in both industry and academia is to predict delay as a point estimate. However, a key concern arises regarding the uncertainty associated with these estimates. Accurate demand forecasting and delay estimation significantly influence the accuracy of delay predictions. Often, delays need to be projected six months in advance to allow for operational adjustments, administrative preparations, and timely communication to policymakers and consumers about anticipated delays (Zografos et al., 2013). However, as previously mentioned, there is no standard practice for quantifying the uncertainty around these predicted delay estimates. This gap highlights the importance of conformal prediction (CP).

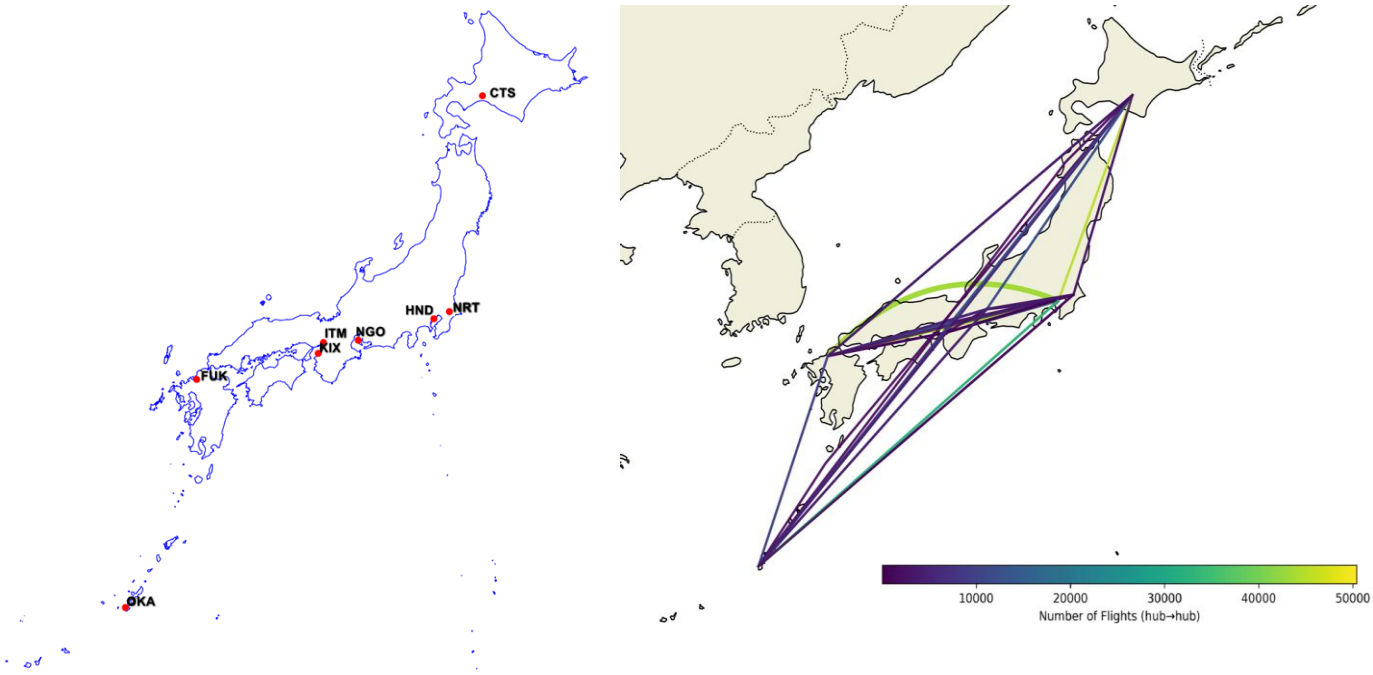
CP aims to produce prediction intervals for the delay of new flights, ensuring that with a high probability (e.g., 90%), the actual delay will fall within this interval. Although it might seem ideal for prediction intervals to cover all possible delays, such intervals would be overly broad and not practically useful. Instead, CP focuses on creating intervals that are as narrow as possible while still capturing the true delay with the specified probability. Furthermore, the width of the interval should adapt based on prediction difficulty: narrower intervals for easily predictable delays and wider intervals for more unpredictable cases (Tibshirani, 2024).

Given the complexity inherent in aviation delay networks, we address the gap related to uncertainty quantification by employing CP models specifically designed to forecast average daily departure delays at various airports. Typically, airports consider departure delays from an operational perspective which focuses on the average delay duration to manage airport capacity (ACRP, 2014). The model integrates

multiple features, including network centralities (degree, betweenness, eigenvector, and closeness), global measures (e.g., transitivity), temporal factors (month, day, quarter), and historical departure delay data. Understanding which network-centric features and other temporal variables drive departure delays is crucial because it unveils how an airport's connectivity and position within the route network amplify or mitigate delay propagation. By pinpointing the important predictors, we can build more accurate, data-driven models that not only forecast delays but also guide targeted operational interventions to improve on-time performance. Therefore, the primary objectives of our study are: (i) to explore variation of network measures in a time-series manner and to identify key network-centric features influencing average departure delays at major airports in Japan, and (ii) to develop and compare adaptive CP models to quantify prediction uncertainty at the airport-level average daily departure delays. To the best of author's knowledge, CP framework has not been applied and tested in the domain of air traffic management, specifically, in the field of delay prediction.

4.2 Study Area and Data Collection

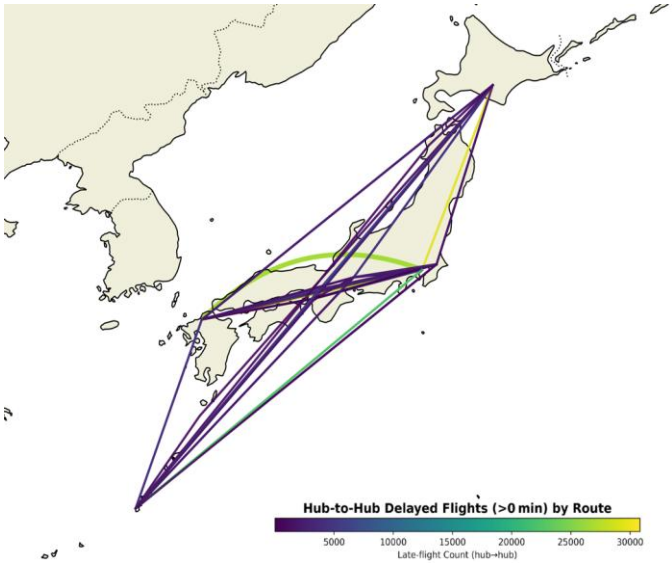
In this study, data on Japanese domestic flights from 2018 to 2021 were analyzed using the Official Aviation Guide (OAG) historical flight data (OAG, 2021). The dataset contains detailed information on each flight, including departure and arrival airports, scheduled, estimated, and actual departure and arrival times, and operating airlines. From this dataset, departure delay information was extracted to construct departure delay networks. Daily snapshots of the departure delay networks were created to represent the departure delays of domestic Japanese flights from 2018 to 2021. In these networks, nodes represent airports, and edges denote flights with departure delays of 1 min or more. The departure delay was calculated as the difference between the actual departure time at the gate and the scheduled departure time. The analysis focused on two full-service carriers in Japan: All Nippon Airways (ANA) and Japan Airlines (JAL). For this study, we took eight major hub airports of Japan. Based on the study of Hanaoka (2018), eight airports were categorized as hub airports for domestic flights in Japan: New Chitose (CTS), Fukuoka (FUK), Tokyo International (HND), Narita International (NRT), Chubu Centrair International (NGO), Kansai International (KIX), Osaka International (ITM), and Naha (OKA). Figure 4.1(a) represents the spatial map of the eight hub airports in Japan and Figure 4.1(b) represents the total number of domestic flights departed from these hub airports (hub-to-hub scenario) from 2018-2021 combining ANA and JAL.



(a) (b)

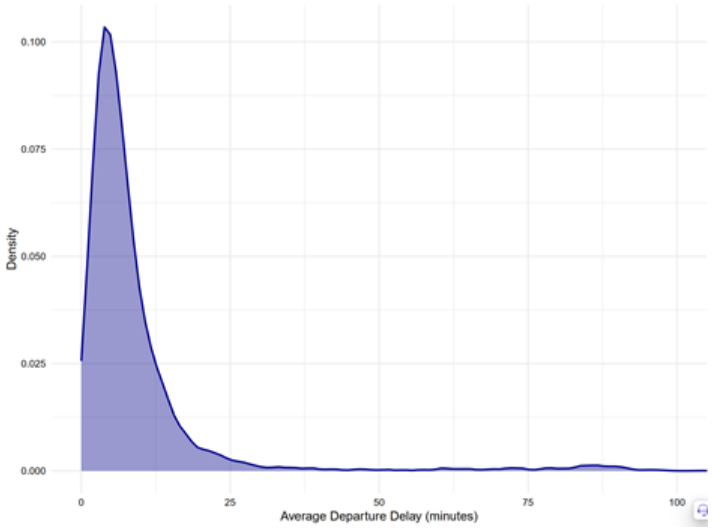
Figure 4.1: (a) Eight hub airports in Japanese domestic aviation, (b) Number of departed flights (hub-to-hub) from 2018-2021

Figure 4.2(a) represents the total number of delayed flights departed in hub-to-hub condition, and Figure 4.2(b) represents the heatmap of this count of hub-hub delayed flights combining ANA and JAL. In this study, we took average departure delay per day for each of the airports in this timeframe. For each of the eight airport and day, average departure delay is calculated as the mean of the departure delay for all flights of ANA and JAL departing from that airport on that day (Figure 4.2: c-d). This average departure delay, then, is treated as the dependent variable, whereas the other network data and categorical data are considered as independent variables. Not all flights departing from a particular airport experience delays each day. This variability implies that the various derived properties (average delay, network properties, etc) of departure delay network can vary daily. The CP model will be developed to predict this average delay, and to quantify the uncertainty at various level of statistical confidence.

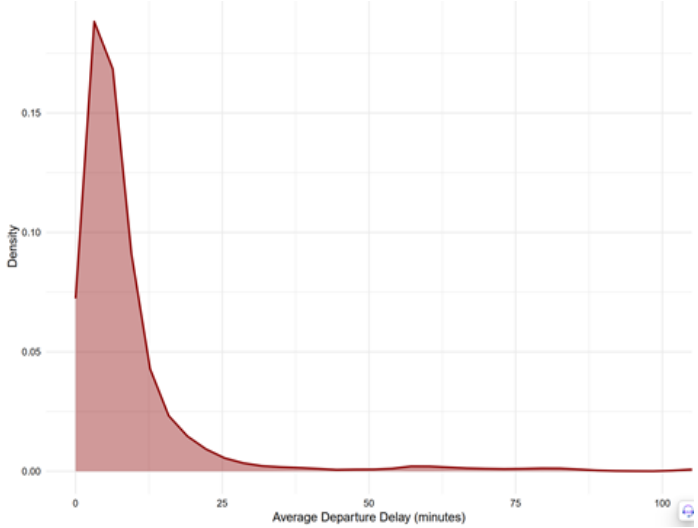


(a)

(b)



(c)



(d)

Figure 4.2: (a) Departure delayed flights (hub-to-hub), (b) heatmap of total late departed flights from 2018-202, (c) distribution of average departure delay for ANA, and (d) distribution of average departure delay for JAL

To quantify the uncertainty of delay prediction, a total twelve explanatory variables are considered. It includes seven network properties - *in-degree centrality*, *out-degree centrality*, *in-closeness centrality*, *out-closeness centrality*, *betweenness centrality*, *eigenvector centrality*, and *transitivity* capturing the topological properties of the airport network. Detailed definitions and interpretations of these metrics are provided in Section 4.1, with summary statistics discussed in Section 5.1. In addition, To capture the temporal and seasonal effects, four temporal factors were included - *quarter of the year* (Q1 to Q4) which is of January to March (Q1 = 1), April to June (Q2 = 2), July to September (Q3 = 3), and October to

December (Q4 = 4); *day of the week* (Monday to Sunday), *month* (January to December) and a binary indicator for *state-of-emergency (SoE)* periods declared by the Japanese government between 2020 and 2021 to consider SoE effect on the prediction model where SoEs from April 7 to May 25, 2020; January 8 to March 21, 2021; April 25 to June 19, 2021 (except for Okinawa), and July 12 to September 30, 2021 were coded as 1 and other days were coded as 0 (Sadeek et a., 2025; Okamoto, 2022). Also, to capture the temporal dependence, *lagged average departure delay* per day is also used for each of the airports to incorporate information about past observations which can be highly predictive of future delay. Time-series based aviation delay data may have autocorrelation meaning todays' delay may be incurred or related to the previous day delay. So, including lagged values allow model to learn temporal relationships.

Detailed data pre-processing path is described in Figure 4.3.

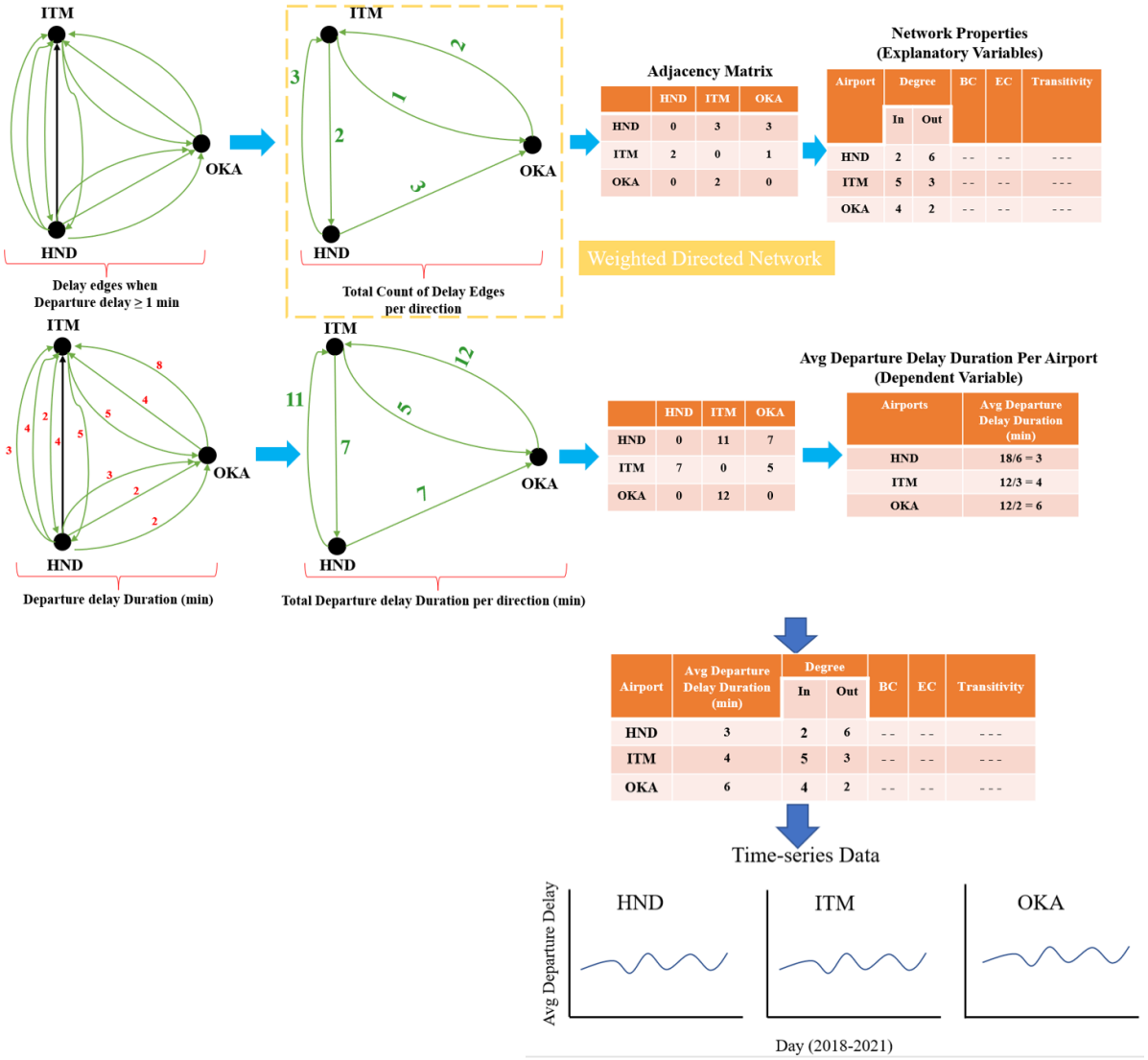


Figure 4.3: Data pre-processing flow chart followed in Chapter 4

4.3 Methodology

Figure 4.4 represents the methodological flow-chart that is followed in this study to predict average departure delay for each of the 8 airports and the association uncertainty quantifications.

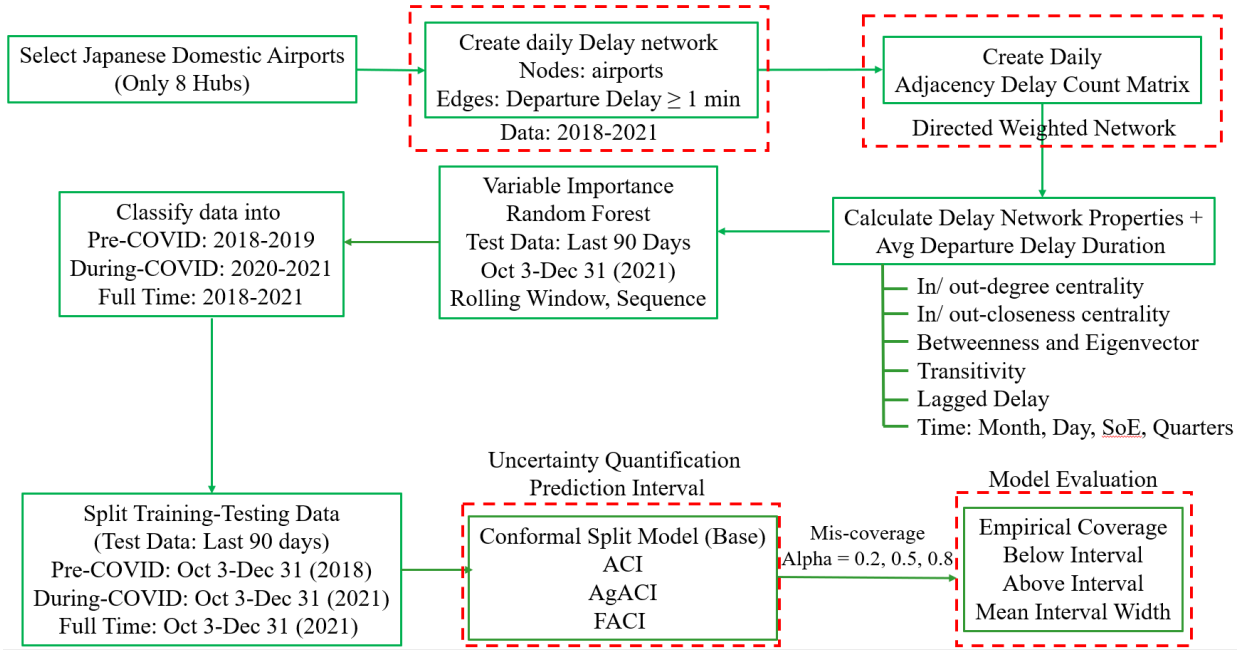


Figure 4.4: Methodological flow-chart followed in Chapter 4

4.3.1 Network properties

In our analysis, we use degree (in and out), closeness (in and out), betweenness, eigenvector centrality and transitivity; for formal definitions of these metrics, see Newman (2018). The degree of an airport denotes the number of delayed links it shares with other airports. In this study, in-degree centrality counts all flights that departed late from their origin airport, while out-degree counts the downstream or destination airports affected by those delays. Betweenness centrality (BC) measures how often an airport lies on the shortest path between two others, highlighting its role as a bridge in the network (Newman, 2018). We use BC to determine whether a delayed airport functions more like a hub or a spoke (An, 2023); airports with higher BC are more capable of influencing and spreading delays across the network (Wandelt et al., 2025b). Closeness centrality (CC) captures how quickly an airport can spread or receive delays: out-closeness measures the speed of propagating delays to all other airports along directed routes, and in-closeness measures the speed of receiving delays from elsewhere (Wandelt et al., 2025b; Li et al., 2024). Airports with higher CC both transmit and absorb delays more rapidly. Eigenvector centrality (EC) reflects how strongly an airport connects to other highly connected, delay-prone airports; a higher EC suggests that delays cluster at well-connected nodes (Güner et al., 2024; Sadeek et al., 2025). Finally, transitivity quantifies the tendency of airports linked to the other delayed airport to also link to each other, forming triangles in the delay network. From a delay perspective, high transitivity indicates that an airport

is part of a tight cluster of frequently delayed airports. Transitivity is a network-based measures which is different from node-based measures like various centrality metrices. In the air transport network, various network-based measures have been used e.g., community, density, network efficiency, average path length, etc. (Wandelt et al., 2025c). However, transitivity measures how much the whole airport delay network is made up of small, tightly connected groups. A high transitivity means that many airports form closed triangles across the network, not just locally. This could be important because when airports are tightly connected like this, delays can spread more easily within these groups. Some studies have incorporated global network measures alongside local ones for prediction tasks (An, 2023) and econometric analyses (Rainone, 2020; Menzel, 2020).

4.3.2 Random Forest Algorithm

Random Forest (RF) is an ensemble learning method that uses boosting and bagging, where multiple decision trees are constructed and merged to evaluate and rank important explanatory variables. RF reduces variance by avoiding the overfitting problem. In this study, the purpose of applying RF was to obtain the most important features (i.e., network properties, seasonal trend driven categories and historical delay data) that have significant impact on the average departure delay of each airports. For this analysis, we focused on permutation-based measurement criteria of variable selection. In this approach, the out-of-bag sample is passed to each tree, and prediction accuracy is recorded. Then the values for each variable are randomly permuted, and the accuracy is recalculated. The decrease in accuracy is averaged over all the trees for each predictor. Then, the variables with the largest average decrease in accuracy are returned as important. For theoretical and practical implementation of RF algorithms, please refer to Breiman (2001) and Boehmke and Greenwell (2020), respectively.

4.3.3 Conformal Inference Framework

Conformal inference, also known as Conformal prediction (CP) provides a framework for quantifying uncertainty in predictions from any machine learning algorithms. It was developed by min-1990s by Vladimir Vovk, Alexandr Gammerman and Vladimir Vapnik where CP converts point predictions into prediction intervals with finite-sample coverage guarantees without relying on distributional assumptions (Tibshirani, 2024). This work was introduced by Vovk (2005) which is still a definitive guideline for CP concepts. The property of CP is that, for a new observation (X_{n+1}, Y_{n+1}), it produces $C_n(X_{n+1})$ satisfying:

$$P\big(Y_{n+1} \in C_n(X_{n+1})\big) \geq 1 - \alpha \tag{4.1}$$

where, $\alpha \in (0,1)$ is the target miscoverage rate. CP achieves finite-sample coverage by using non-conformity scores, computed from residuals of prediction accuracy and comparing ranks and quantiles within the data.

In our analysis, we have integrated the prediction results of RF with CP framework. We first trained the data $\{(X_t, Y_t)\}_{t=1}^{n}$, where response variable Y is the average departure delay and predictors include the network metrices and other temporal factors, fit a RF model $\widehat{f_{RF}}(X)$ obtaining predictions of $forecast = \widehat{f_{RF}}(X_t)$. Then we compute the non-conformity scores using absolute residuals as $|y_t - forecast|$. Then the predictions intervals and update are conducted using Equation 4.2 and 4.3 below when applying into adaptive conformal framework. The adaptive conformal inference framework extends the CP framework to online, streaming data scenarios where data arrives sequentially and potentially non-exchangeably. It then dynamically adjusts the width of prediction intervals based on past performance ensuring intervals remain accurate as must as possible.

4.3.3.1 Conformal Split Model

As indicated above, conformal prediction aims to construct prediction bands for new data points that guarantee finite-sample coverage mode than 1-α without strong assumptions about underlying data distribution. However, full CP models are computationally heavy because it retains the model on every augmented dataset. This split conformal framework overcomes this by splitting the dataset into two parts – training set: used to fit a regression model, and calibration set: used to assess residual errors and determines a quantile threshold for prediction intervals (Lei et al., 2018). This approach maintains valid predictive inference with much lower computational cost making it more practical for time-series data prediction. This split conformal framework can be adapted for time-series predictions. Any regression based or machine learning models can be used in the training step. Then, past observed time-ordered data serves as calibration points to estimate residual quantiles reflecting temporal error distribution. For a future time-stamp, t+1, the interval is constructed using past residual quantiles capturing uncertainty in future predictions that leverages concept of distribution-free prediction intervals that adapt to the empirical variability of residuals without any assumptions.

4.3.3.2 Adaptive Conformal Inference (ACI)

ACI is an online algorithm designed to adaptively adjust the width of prediction intervals. In a prediction problem, we want an interval that covers the true outcome (for example, the actual departure delay per day of a particular airport for a particular airline's flight network) with high probability. However, if the interval is too narrow, it might miss the true value and if it is too wide, it loses precision. ACI automatically updates the width based on past performance. For a detailed theoretical information, please refer to Gibbs and Candès (2021). Here, we give the basic steps of this method as follows:

Step 1: Initialize with a value of θ_t which determines the initial width of the prediction interval. A learning rate, $\gamma > 0$ is chosen to control how quickly the algorithm adjusts.

Step 2: For each time step t, a prediction interval $C_t(v_t)$ is produced. This interval is

$$C_t(\theta_t) = (forecast - \theta_t, forecast + \theta_t) \quad (4.2)$$

The observed true outcome is y_t (here, actual average departure delay).

Step 3: Define an error indicator err_t by checking whether the true outcome y_t falls outside the interval as $err_t = I[y_t \notin C_t(\theta_t)]$. Here, the indicator is 1 if $y_t$is outside the prediction interval and is 0 if it is inside the interval covered.

Step 4: The key update is:

$$\theta_{t+1} = \theta_t + \gamma(err_t - (1 - \alpha)) \quad (4.3)$$

Here, α is related to the target miscoverage level (for example, if for 80% coverage, $\alpha = 0.2$) If an error occurs ($err_t = 1$), then the update becomes $\theta_t + \gamma\alpha$. It, therefore, increases θ_t and thus widens the next prediction interval. If the prediction is successful (i.e., $err_t = 0$), then $\theta_t - \gamma(1 - \alpha)$. The interval becomes narrower and to be more precise.

This algorithm guarantees that the error in coverage (the difference between the desired coverage and the actual coverage) is bounded. Since this is the first application of this algorithm in this specific area, there are no established benchmarks or prior studies to guide the selection of γ. Therefore, we chose $\gamma = 0.01$ through preliminary experimentation suggested in Susmann et al. (2023). the cumulative regret (or suboptimality) grows sub-linearly. A larger learning rate means faster adaptation which means coverage error decays more quickly but can lead to oscillations in the width. A smaller learning rate yields smoother but slower adaptation. In practice, learning rate is tuned to balance responsiveness and stability.

4.3.3.3 Aggregate Adaptive Conformal Inference (AgACI)

While using ACI, one key tuning parameter was the learning rate γ. Different choices of γ can lead to prediction intervals that are too jumpy or that do not adapt quickly enough. AgACI addresses this by running several instances of the ACI algorithm in parallel - each with a different candidate learning rate—and then aggregating their lower and upper prediction bounds separately. In other words, rather than having to pick one single learning rate, AgACI combines multiple candidate intervals. For a detailed theoretical information, please refer to Zaffran et al. (2022). Here, we give the basic steps of this method as follows

Step 1: Choose a set of candidate learning rates $\{\gamma_k\}_{1 \le k \le K}$, starting at θ_1. All candidates start with the same initial parameter θ_1(which influences the initial interval width). Two online aggregation methods (in this case, the Bernstein Online Aggregation, BOA) are initialized: one for the lower bound and one for the

upper bound. Each aggregation algorithm is set to target a quantile: (i) the lower bound is set to target the $\frac{1-\alpha}{2}$ quantile, and (ii) the upper bound target the $1-\frac{1-\alpha}{2}$ quantile.

Step 2: For each candidate *k*, and ACI instance A_k is run. Each such instance produces its own prediction interval $[l_t^k, u_t^k]$ at time *t*. Recall that a basic ACI update uses equation 1 where the error indicator is

$$err_t = I[y_t \notin C_t(\theta_t)] \tag{4.4}$$

Where, y_t is the observed outcome and $C_t(\theta_t)$ the current prediction interval.

Step 3: At each time t, the algorithm retrieves all candidate lower bounds $\{l_t^k\}_{k=1}^K$ and all candidate upper bounds $\{u_t^k\}_{k=1}^K$. Two separate BOA aggregation algorithms combine the candidate bounds: (i) aggregated lower bound is computed as:

$$\widetilde{l_t} = BOA_l(\{l_t^k\}_{k=1}^K) \tag{4.5}$$

and (ii) aggregated upper bound is computed as

$$\widetilde{u_t} = BOA_l(\{u_t^k\}_{k=1}^K) \tag{4.6}$$

These aggregators assign weights to each candidate based on their past performance (using quantile loss) so that better-performing candidates have more influence. The final output at time t is the aggregated interval $[\widetilde{l_t}, \widetilde{u_t}]$.

Step 4: After the output of the interval, the algorithm observes the true outcome y_t (e.g., the actual average departure delay). Each candidate ACI algorithm A_k is updated with y_t using its own update rule. Additionally, the BOA aggregators update their weights based on the performance (i.e., how close the candidate bounds were to covering y_t.

4.3.3.4 Fully Adaptive Conformal Inference (FACI)

FACI is designed to address the challenge of choosing an optimal learning rate when adapting prediction intervals. Like ACI and AgACI, FACI works in an online fashion—but here it fully adapts by running several copies of ACI, each with a different learning rate. Their outputs are then aggregated using an exponential reweighting scheme based on the pinball loss. This method allows us to combine the candidate estimates into a single interval while enabling theoretical performance guarantees. For a detailed theoretical information, please refer to Gibbs and Candès (2024). Here, we give the basic steps of this method as follows

Step 1: Starting value θ_1(e.g., determining the initial interval width). The candidate learning rates as $\{\gamma_k\}_{1\leq k\leq K}$ where each candidate expert runs its own ACI update. Some additional parameters such as: σ and η are tuning parameters for the reweighting scheme. For each candidate *k*, an ACI instance $A_k$is initialize with $\gamma = \gamma_k$ and starting value θ_1.

Step 2: At each time step *t*, each candidate expert *k* has a weight w_t^k. These are normalized into probabilities as:

$$p_t^k = \frac{w_t^k}{\sum_{i-1}^{K} w_t^i} \tag{4.7}$$

The overall prediction parameter is computed as a weighted average of the candidate parameter as:

$$\theta_t = \sum_{k=1}^{K} \theta_{t,k} p_t^k \tag{4.8}$$

This θ_t is then used to form the prediction interval $C_t(\theta_t)$. For example: the interval is constructed as $C_t(\theta_t) = [forecast - \theta_t, forecast + \theta_t]$, then θ_t determines how wide the interval is. After the output of the interval, the algorithm observes the true outcome y_t. It then computes a residual r_t (which in many conformal methods reflects the error) and evaluates the pinball loss $L_\alpha(\theta_{t,k}, r_t)$ for each candidate expert. The pinball loss is defined piecewise by:

$$\text{if } r_t > \theta_{t,k},: L_\alpha(\theta_{t,k}, r_t) = (1-\alpha)(r_t - \theta_{t,k}) \tag{4.9}$$

$$\text{if } r_t \leq \theta_{t,k},: L_\alpha(\theta_{t,k}, r_t) = \alpha(\theta_{t,k} - r_t) \tag{4.10}$$

Each expert's weight is updated according to its loss:

$$\overline{w_t^k} = w_t^k \exp(\eta L_\alpha(\theta_{t,k}, r_t)) \tag{4.11}$$

Then with $W_t = \sum_{i=1}^{K} \overline{w_t^i}$, the weights are smoothed for the next round:

$$w_{t+1}^k = (1-\alpha)\overline{w_t^k} + \frac{\sigma W_t}{K} \tag{4.12}$$

The overall error indicator is defined in equation 4.3, and each candidate ACI expert updates its own parameter with a modification in equation 1 as

$$\theta_{t+1,k} = \theta_{t,k} + \gamma_k(err_t - (1-\alpha)) \tag{4.13}$$

therefore, if the prediction interval missed the true value ($err_t = 1$), he candidate increases its interval width; if it covered the outcome ($err_t = 0$), it narrows the interval.

By reweighting experts according to the pinball loss, FACI adapts over time, and theoretical bounds (in terms of strongly-adaptive regret) can be derived. Unlike AgACI that aggregates lower and upper bounds separately, FACI aggregates the candidate estimates directly, which simplifies theoretical analysis and facilitates the derivation of coverage error bounds.

4.3.4 Performance metrics of conformal prediction framework

The effectiveness of CP methods integrated with RF is assessed by the following performance metrices.

i. Mean Interval Width: It measures the interval precision as:

$$Mean\ Interval\ Width\ (T) = \frac{1}{T}\sum_{t=1}^{T}(u_t - l_t) \tag{4.14}$$

ii. Empirical Coverage: It assesses interval accuracy as:

$$Empirical\ Coverage\ (T) = \frac{1}{T}\sum_{t=1}^{T}\mathbf{1}[y_t \in C_t(\theta_t)] \tag{4.15}$$

Coverage error that indicates deviation from nominal coverage $(1-\alpha)$ is defined as:

$$Coverage\ Error(T) = Empirical\ Coverage\ (T) - (1-\alpha) \tag{4.16}$$

4.3.5 Applicability of conformal inference framework in non-i.i.d setting of times-series data

While classical CP framework provides coverage guarantees under the assumption of exchangeability, this assumption is violated in temporally structured data, e.g., flight departure delay in our case which are subjected to autocorrelation, seasonal effects, network-centric dynamics, etc. However, recent studies have addressed these limitations and enabled distribution-free uncertainty quantification under non independent and identically distributed (i.i.d) and non-stationary data cases. In this case, our three applied models – ACI, AgACI and FACI are theoretically built on non-i.i.d data. For ACI, Gibbs and Candès (2021) proposed no constraints on the data distribution and obtains correct coverage frequency without any assumptions on the distribution. ACI adaptively updates the conformal quantile level using recursive rule and the distributional shift is handled by online adaptation. For ACI, Gibbs and Candès (2021) provided a theoretical as well as empirical justification under non-stationary and temporally autocorrelated conditions where modeling the shift of the distribution is done using hidden markov models. For AgACI, Zaffran et al. (2022) proposed this CP framework with a guarantee of long-term coverage for time-series non-exchangeable data with incorporating weighting to past residuals to construct prediction intervals that adapts when there is a change in the data distributions. It handles time series data by applying exponentially decaying weights to the residuals, emphasizing on the recent data. Lastly, Gibbs and Candès (2024) proposed FACI with fully online setting. FACI adjusts the conformal quantile dynamically based on recent miscoverage history using feedback mechanism. This directly

addresses the problem of calibration drift under distributional shift or in case of autocorrelated data. Therefore, CP framework can be applied to time-series data like departure delay in our case but only with careful adaptation that we have considered by applying these three CP frameworks in adaptive setting. Adaptive CP can be used by weighting recent samples and updating prediction intervals online. These techniques relax the i.i.d. requirement and it maintains valid coverage.

4.4 Results and Discussion

4.4.1 Summary of the network properties

To show how the network-centric measures and average delay patterns vary over time, especially before and after the COVID-19 disruption, Table 4.1 presents key network metrics and average departure delays (minutes) for ANA and JAL from 2018 to 2021. Both ANA and JAL experienced their average departure delays plunge in 2020 before rebounding sharply in 2021, reflecting the sharp drop and gradual recovery in flight volumes during the COVID-19 period. Network connectivity followed the same pattern: in- and out-degrees roughly halved from 2019 to 2020 as routes were cut back, then climbed again steadily in 2021. Closeness centrality, which had been relatively stable, became more variable in 2020 (larger SD), signaling fragmented route structures and uneven access, before settling back toward pre-pandemic levels. BC spiked and grew more volatile for JAL than for ANA, while ANA's BC stayed more constant but dipped in 2020. EC remained consistently higher for ANA, suggesting its network retained a stronger core of well-connected airports, and transitivity edged down for both carriers during 2020-21 as triangle-forming connectivity weakened. Overall, these shifts in network metrics track closely with the fall and rise of average delays, underlining how structural changes in route networks during the pandemic drove and were driven by delay dynamics. Also, the time-series plot comparing ANA and JAL of average departure delay and various network metrices are given in Appendix Figure B1.

Table 4.1: Descriptive statistics of average delay and network-centric variables from 2018 to 2021 for ANA and JAL

Variable	Year	ANA				JAL			
		Mean	**SD**	**Min**	**Max**	**Mean**	**SD**	**Min**	**Max**
Average	2018	9.22	11.51	0	498.00	8.64	8.59	0	96.00
Departure	2019	9.31	8.45	0	206.00	9.19	9.56	0	221.00
Delay	2020	6.09	8.05	0	116.00	5.80	7.55	0	150.00
(in min)	2021	12.22	23.99	0	304.00	13.85	43.11	0	1623.00
In-/out-	2018	42.43/47.99	36.89/45.99	0/1	189/199	33.36/37.07	41.41/48.02	0/1	187/199
degree	2019	40.14/42.22	35.64/34.22	0/1	187/244	35.64/39.36	44/50.95	0/1	254/308

	2020	17.05/19.83	19.44/24.31	0/1	181/139	16.91/19.59	23.45/29.83	0/1	182/186
	2021	18.61/20.71	21.75/25.72	0/1	170/171	19.09/21.01	28.44/32.13	0/1	205/205
	2018	0.020/0.019	0.004/0.003	0/0	0.037/0.029	0.015/0.014	0.004/0.003	0/0	0.091/0.027
In-/out-	2019	0.020/0.020	0.004/0.003	0/0	0.036/0.031	0.014/0.014	0.004/0.003	0/0	0.091/0.050
closeness	2020	0.043/0.048	0.092/0.133	0/0	1/1	0.033/0.032	0.065/0.098	0/0	1/1
	2021	0.032/0.030	0.054/0.076	0/0	1/1	0.021/0.026	0.043/0.092	0/0	1/1
	2018	126.03	198.04	0	786.60	192.03	371.14	0	1290.38
BC	2019	123.88	173.79	0	654.88	196.67	377.87	0	1335.81
	2020	86.93	129.96	0	774.34	143.09	282.53	0	1308.65
	2021	90.03	128.21	0	628.95	179.71	330.49	0	1449.22
	2018	0.459	0.276	0	1	0.395	0.309	0	1
EC	2019	0.452	0.279	0	1	0.394	0.305	0	1
	2020	0.563	0.323	0	1	0.511	0.320	0	1
	2021	0.565	0.313	0	1	0.535	0.314	0	1
	2018	0.429	0.256	0	1	0.485	0.345	0	1
Transitivity	2019	0.419	0.245	0	1	0.499	0.348	0	1
	2020	0.344	0.275	0	1	0.415	0.379	0	1
	2021	0.340	0.282	0	1	0.380	0.371	0	1

4.4.2 Identifying important features

For feature selection and validation of predicted results from Random Forest algorithm, we combined the entire span from January 1, 2018 to December 31, 2021 (1,461 observations). The final 90 days of the full record (October 3, 2021–December 31, 2021) were reserved for testing. To emulate genuine out-of-sample forecasting, the final 90 days of this period were held out as a test window. At each phases, for each day, t in that window, we fit a new Random Forest model using only the data from January 1, 2018 up through day $t-1$, and we used that model to predict the delay on day t. Because each forecast is trained on strictly earlier observations, every one-day-ahead prediction is genuinely out-of-sample. For this study, we avoided k-fold cross-validation as the data itself is time-series in nature; therefore, shuffling or randomly partitioning the observations would violate the temporal ordering that is essential for forecasting. So, the rolling-window procedure is use which is itself a form of time-series cross-validation. At each step, we trained on all data up to time $t-1$ and test on time t. By holding out the final 90 days and re-fitting on every preceding day, we both preserved chronological integrity and evaluate out-of-sample performance.

Although we originally examined twelve predictors, seven network centralities (in-degree, out-degree, in-closeness, out-closeness, betweenness, eigenvector, transitivity), four-time factors (quarter, day of week, month, SOE), and the one-day lag of delay, in the results we consistently selected just six features for each fit: These variables are always drawn from the training set of days 1 through $t-1$. Based on RF importance result, we selected lagged delay, in-degree, out-degree, betweenness, month, and SOE to model departure-delay uncertainty for ANA (Table 4.3) and JAL (Table 4.4). Although SOE proved important only for JAL at NRT, we included it to capture states of emergency occurred in 2020-2021.

In Table 4.2, for RF - permutation-based feature importance shows lagged delay as the top predictor across eight Japanese airports in both ANA and JAL networks. At CTS, ANA's lagged delay importance exceeds in-degree and month, whereas JAL balances lagged delay relatively closely with in-degree and out-degree. At HND, lagged delay dominates, supplemented by ANA's out-degree and in-degree and JAL's in-degree. Connectivity metrics are vital: at FUK, ANA prioritizes out-degree and lagged delay while JAL distributes importance among in-degree, out-degree, out-closeness, and betweenness. At OKA, ANA emphasizes in-degree and betweenness, while JAL weights out-degree, lagged delay, and eigenvector centrality. Seasonal month appears consistently, especially for ANA: at KIX, month, lagged delay, and out-closeness; JAL at KIX relies on lagged delay and transitivity. At NGO, ANA considers month, lagged delay, and connectivity, while JAL highlights eigenvector centrality. NRT shows uniformly low importance for both carriers, suggesting stable operations. Overall, lagged delays remain pivotal; ANA leans more on historical delays and connectivity, whereas JAL distributes importance across centrality metrics, lagged delays, and temporal factors.

Table 4.2: Top 5 Important Features for ANA and JAL using RF

Airport	ANA	JAL
CTS	Lagged delay = 91	In-degree = 66.5
	In-degree = 26.9	Out-degree = 65
	Month = 19.6	Lagged delay = 42.9
	Out-degree = 15.8	Betweenness = 24
	Betweenness = 11.9	Month = 23.5
FUK	Out-degree = 336	In-degree = 33.8
	Lagged delay = 272	Out-degree = 21.9
	In-degree = 90.6	Lagged delay = 19.8
	In-closeness = 54	Out-closeness = 11
	Betweenness = 34.5	Betweenness = 7.67

	Lagged delay = 99.5	Lagged delay = 70.4
	Out-degree = 28	In-degree = 39.8
HND	In-degree = 27.1	Out-degree = 17.6
	Month = 25.5	Month = 7.04
	Betweenness = 7.12	Betweenness = 5.96
	In-degree = 45.4	Lagged delay = 32.4
	Out-degree = 28.2	In-degree = 28.5
ITM	Lagged delay = 23.5	Out-degree = 21.9
	Transitivity = 19.5	Transitivity = 12.5
	Betweenness = 15.8	Month = 8.68
	Lagged delay = 68.4	Lagged delay = 797
	Month = 51.5	Transitivity = 653
KIX	Out-closeness = 32.3	In-degree = 44.8
	Out-degree = 26.8	Days = 40.9
	Betweenness = 16.8	Month = 34.1
	Lagged delay = 80.7	Eigenvector = 29.9
	In-degree = 30.4	Month = 23
NGO	Out-degree = 27.1	Lagged delay = 20.4
	Month = 25.5	In-degree = 13.3
	Betweenness = 15.5	In-closeness = 11.3
	Transitivity = 2.91	Transitivity = 0.75
	Quarter = 2.91	SOE = 0.75
NRT	Month = 2.91	Quarter = 0.75
	Lagged delay = 2.91	Month = 0.75
	Eigenvector = 2.91	Lagged delay = 0.75
	In-degree = 121	Out-degree = 50.9
	Betweenness = 103	Lagged delay = 49.6
OKA	Lagged delay = 65	In-degree = 41.1
	Out-degree = 64.4	Eigenvector = 40.3
	Transitivity = 36.9	Month = 31.9

Based on RF importance result, we selected lagged delay, in-degree, out-degree, betweenness, month, and SOE for ANA and JAL. Although SOE is found important only for JAL at NRT, we included it to capture states of emergency occurred in 2020-2021. The Figure 4.5 presents an evaluation of predictive model

performance for forecasting average departure delays at Japan's eight major hub airports operated by ANA and JAL. M1: Lagged Delay Only act as baseline model using only previous day delays as predictors. M2: Lagged + Month + SOE adds seasonal and pandemic context variables. M3: Lagged + Month + SOE + Network Metrics extends M2 by including network position measures with in-degree, out-degree, betweenness centrality.

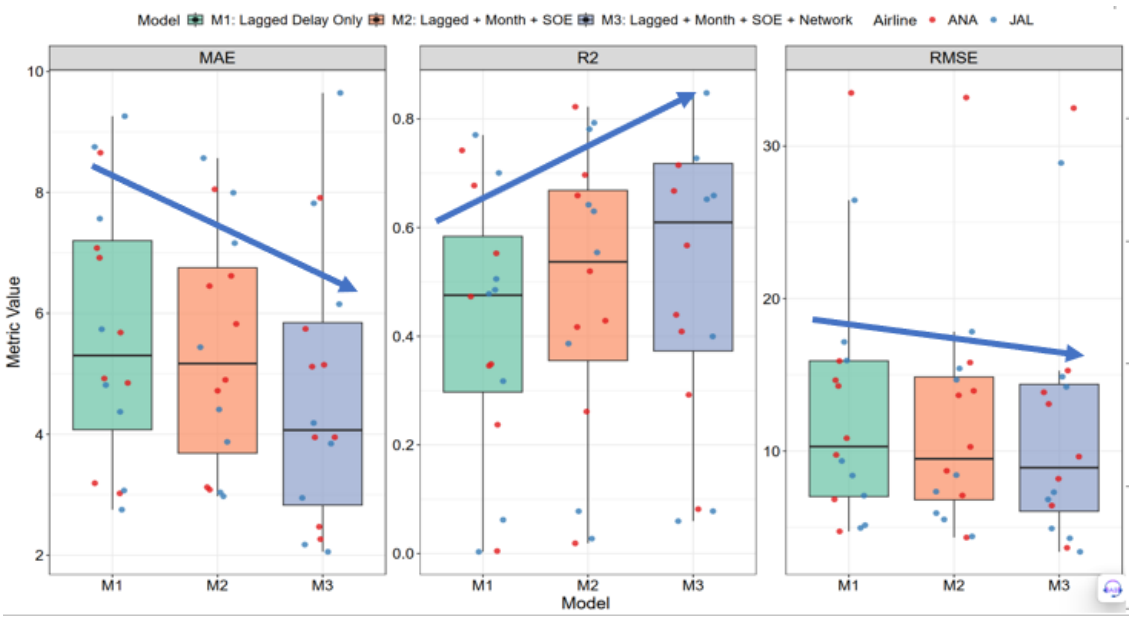


Figure 4.5: Predictive model performance for forecasting average departure delays

The three boxplots visualize comparative performance across three evaluation criteria. Mean Absolute Error (MAE) reflects the average magnitude of prediction errors. Lower values indicate better performance. Results show a decreasing trend in MAE from M1 to M3, demonstrating that incorporating seasonal and network variables progressively improves prediction accuracy. Coefficient of Determination (R^2) captures the proportion of variance explained by the model. Higher values signify better explanatory power. It's values increase consistently across models, with M3 showing the highest R^2, indicating that the most comprehensive model achieves the strongest fit. Root Mean Square Error (RMSE) quantifies the square root of the average squared errors penalizing large deviations more heavily than MAE. Lower RMSE denotes superior model performance. Similar to MAE, RMSE declines across models further confirming the added predictive value of including network metrics and seasonal context.

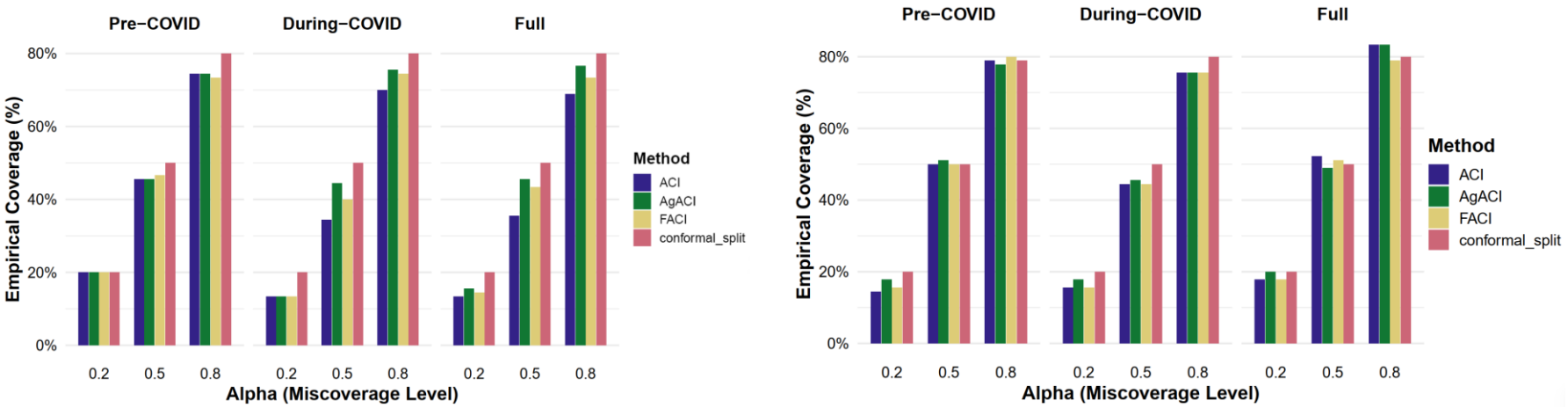
4.4.3 Uncertainty analysis of delay prediction for ANA and JAL

Based on RF importance result, we selected lagged delay, in-degree, out-degree, betweenness, month, and SOE to model departure-delay uncertainty for ANA (Figure 4.6) and JAL (Figure 4.8). Although SOE proved important only for JAL at NRT, we included it to capture states of emergency occurred in 2020-

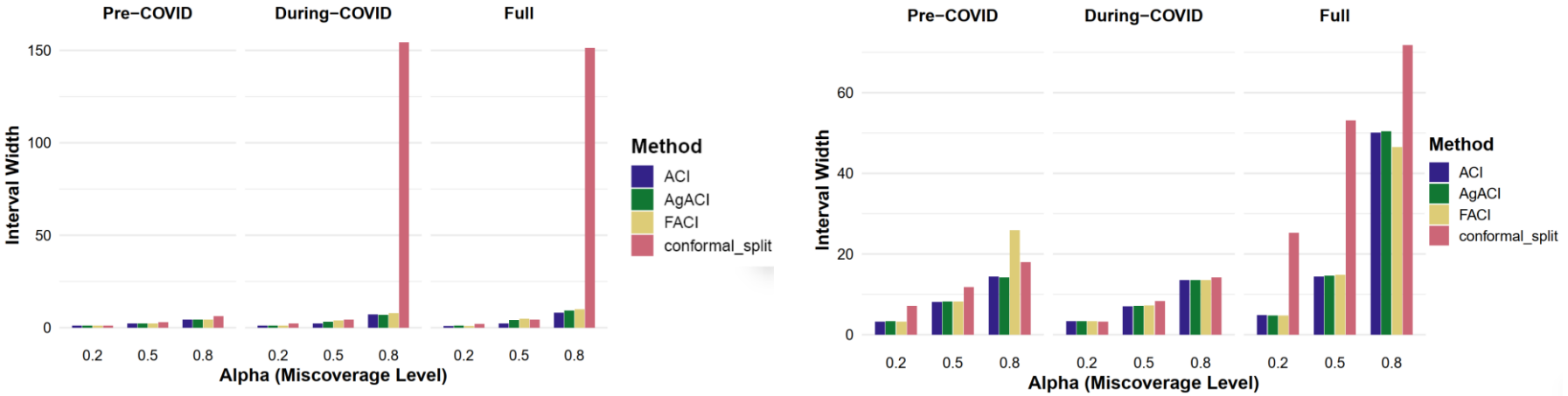
2021. We applied ACI, AgACI, and FACI at $\alpha = 0.2, 0.5$, and 0.8, and report empirical coverage, mis-coverage patterns (below vs. above intervals), and mean interval width. The results are reported in Figure 4.6 and Figure 4.8 below for only two airports HND and NRT for ANA and JAL, respectively. The rest of the results of other 6 airports are shown in the Appendix (Figure D1 and Figure D2 for ANA and JAL, respectively). The phases are categorized into three time-frames: pre-COVID (stable region) - from January 1, 2018 through December 31, 2019, yielding 730 observations. To emulate genuine out-of-sample forecasting, the final 90 days of this period (October 3, 2019-December 31, 2019) were held out as a test data. We then repeated the same procedure for the COVID period (unstable region), using data from January 1, 2020 through December 31, 2021 (731 days). Again, the last 90 days (October 3, 2021-December 31, 2021) served as the out-of-sample test set. Finally, for a comprehensive view, we combined the entire span from January 1, 2018 to December 31, 2021 (1,461 observations). The final 90 days of the full record (October 3, 2021-December 31, 2021) were again reserved for testing.

Figure 4.6 compares three adaptive conformal regression methods - Adaptive Conformal Inference (ACI), Aggregated-ACI (AgACI) and Fully Adaptive Conformal Inference (FACI) - against the standard inductive conformal split baseline. Each method produces prediction intervals for ANA's average departure delay duration (in minutes) at Haneda (HND) and Narita (NRT). The panels depict empirical coverage, mean interval width, and the proportions of observed delay durations falling below or above their prediction intervals. Results are shown for three mis-coverage levels ($\alpha = 0.2, 0.5$ and 0.8) and for three time segments: pre-COVID, during-COVID and the full sample. Adaptive conformal inference frameworks adjust interval widths online to preserve coverage under distribution shift. AgACI combines multiple learners with different learning rates, while FACI aggregates multiple experts; both extend ACI but FACI tends to be more conservative at high-coverage levels.

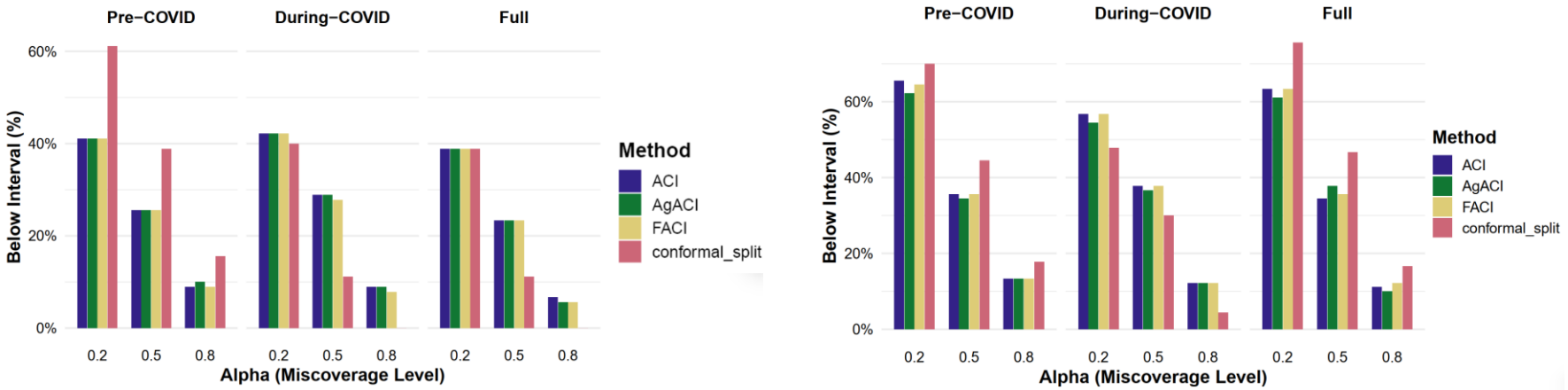
For ANA, in terms of empirical coverage (Figure 4.6-a), at HND, for ACI, AgACI and FACI the empirical coverage increases roughly linearly with α, reaching ~20 % at $\alpha = 0.2$, ~45 % at $\alpha = 0.5$ and ~70–75 % at $\alpha = 0.8$. The conformal-split baseline has slightly higher coverage, especially at high α: at α = 0.8 it achieves ≈80 % (close to the nominal $1-\alpha$ target) whereas the adaptive methods under-cover by ~5–10 %. During-COVID the differences shrink - AgACI's coverage at $\alpha = 0.5$ is about the same as the baseline, while ACI and FACI slightly under-cover. Across all time periods, however, none of the methods reaches the 80 % coverage expected at $\alpha = 0.2$, suggesting that the delay distribution has heavy tails that are difficult to cover. When $\alpha = 0.2$, corresponding to an ambitious 80% coverage goal, ACI's adaptive update rule θ_{t+1} fails to widen its intervals fast enough after surprise delay spikes at around day 60 of the test phase (Figure 4.7) for COVID-phase and full-time time-frames. The



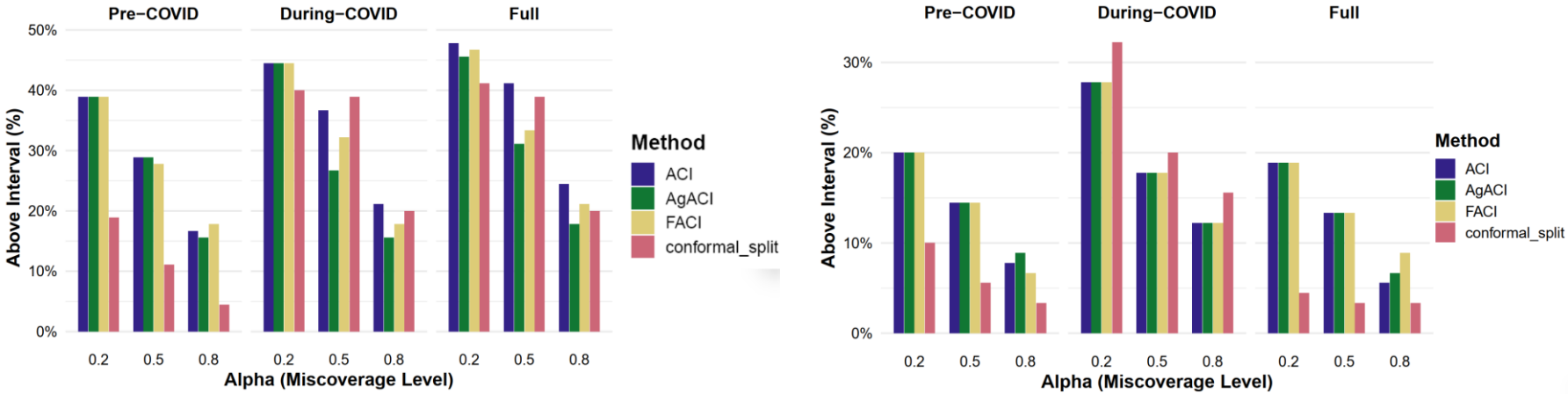
(a) Empirical Coverage



(b) Mean Interval Width



(c) Below Interval



(d) Mean Interval Width

HND NRT

Figure 4.6: Comparison Summary Statistics of Adaptive Conformal Inference for ANA

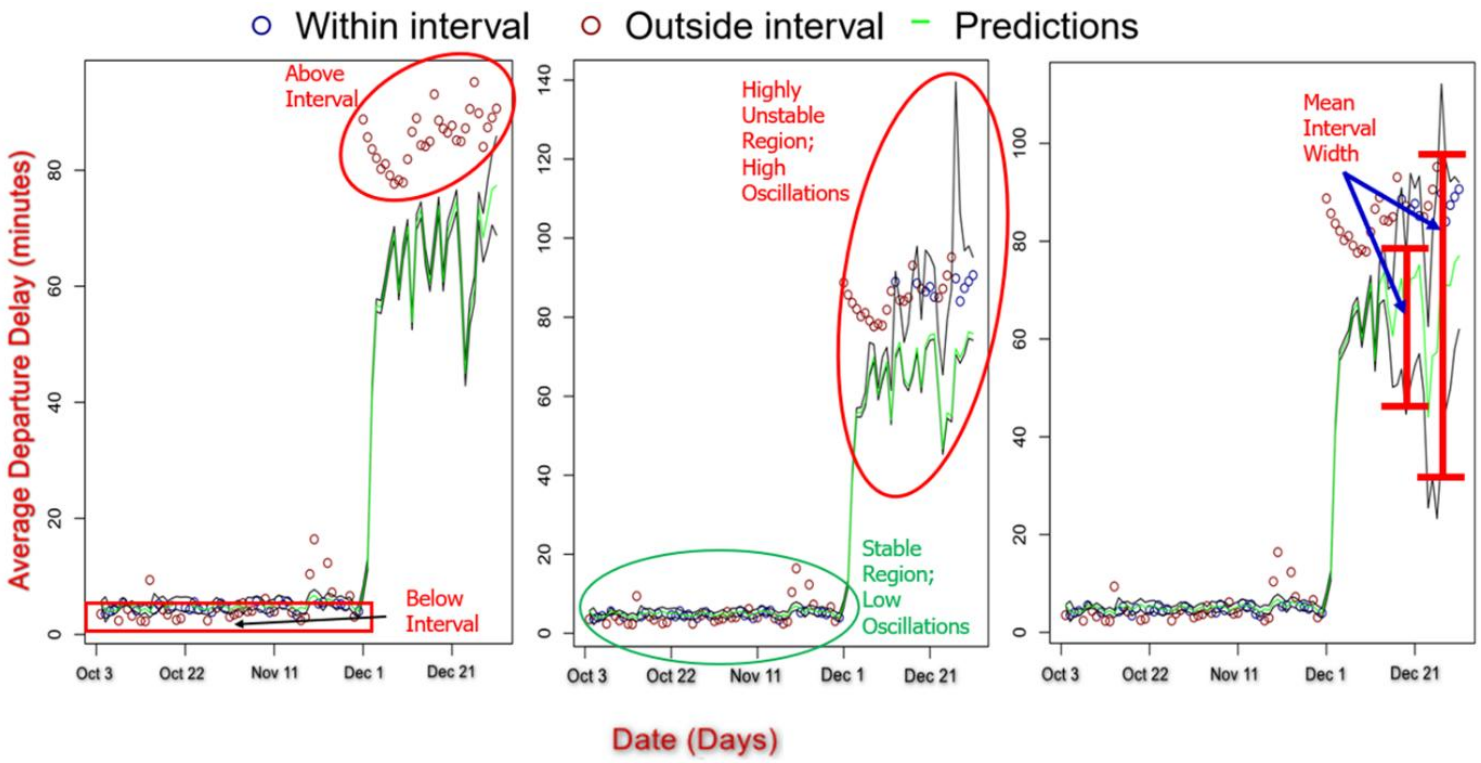


Figure 4.7: Example of testing data (October 3 to December 31) of adaptive conformal inference for full-time frame case

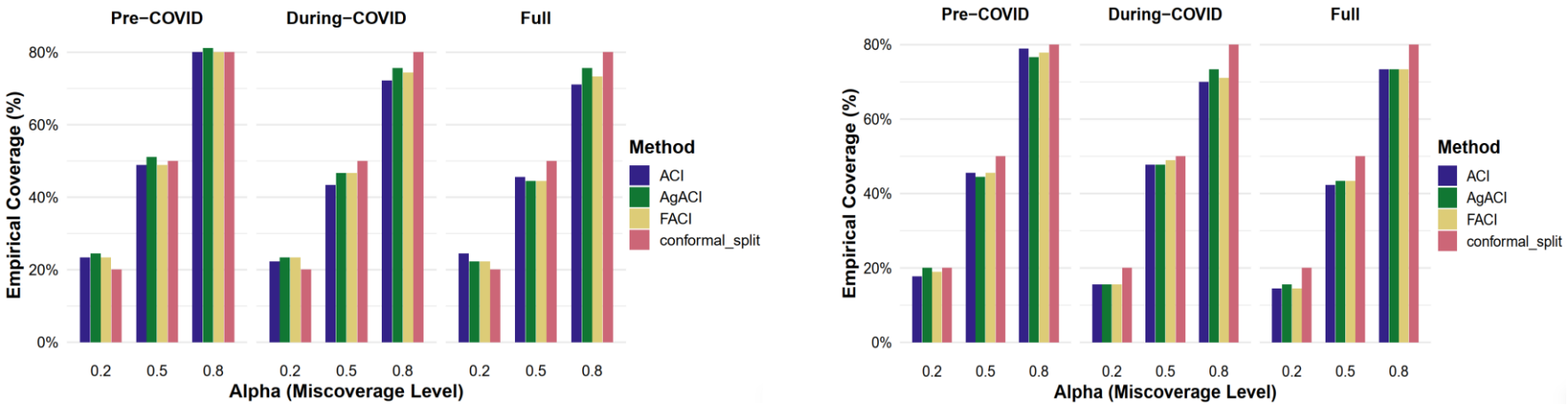
result is narrow mean widths and lower empirical coverage. AgACI's ensemble of learning rates achieves only a marginal gain by averaging bounds across candidate of γ, while FACI's expert-reweighting adds smoothness at the cost of similar under-coverage around the sudden spikes day 60 around the December 1-2 in 2021. These outcomes confirm that, under stringent coverage demands, neither a single-rate update nor more sophisticated aggregation can overcome the inertia of θ when errors dramatically exceed the allowance $(1-\alpha)$. At NRT, the adaptive methods perform better relative to the baseline. At $\alpha = 0.5$ and $\alpha = 0.8$ in the pre-COVID and full periods, ACI and AgACI achieve coverage very close to the baseline and occasionally slightly exceed it. At $\alpha = 0.2$ all methods under-cover (~15–20 %), but the gap between the adaptive and baseline methods is small. Thus, for NRT the adaptive approaches yield coverage comparable to the conformal-split baseline except at the lowest miscoverage level.

In case of, mean interval width (Figure 4.6-b), at, HND, the advantage of the adaptive methods is most pronounced in interval width. For $\alpha = 0.8$, the conformal-split intervals become extremely wide (>150 minutes in the pre-COVID and full periods). In contrast, ACI, AgACI and FACI keep the mean width below about 10 minutes (and below 5 minutes at $\alpha = 0.2$ and 0.5), delivering much tighter prediction intervals. The price for the baseline's high coverage is therefore severe over-conservatism. At, NRT: At $\alpha = 0.8$ in the full period, the baseline interval width explodes to more than 60 minutes, whereas the adaptive methods maintain widths around 15 minutes. Pre-COVID, FACI has wider intervals than ACI or AgACI (≈25 minutes at $\alpha = 0.8$) due to its fairness constraints, but even in this case the adaptive intervals are comparable to or narrower than the baseline.

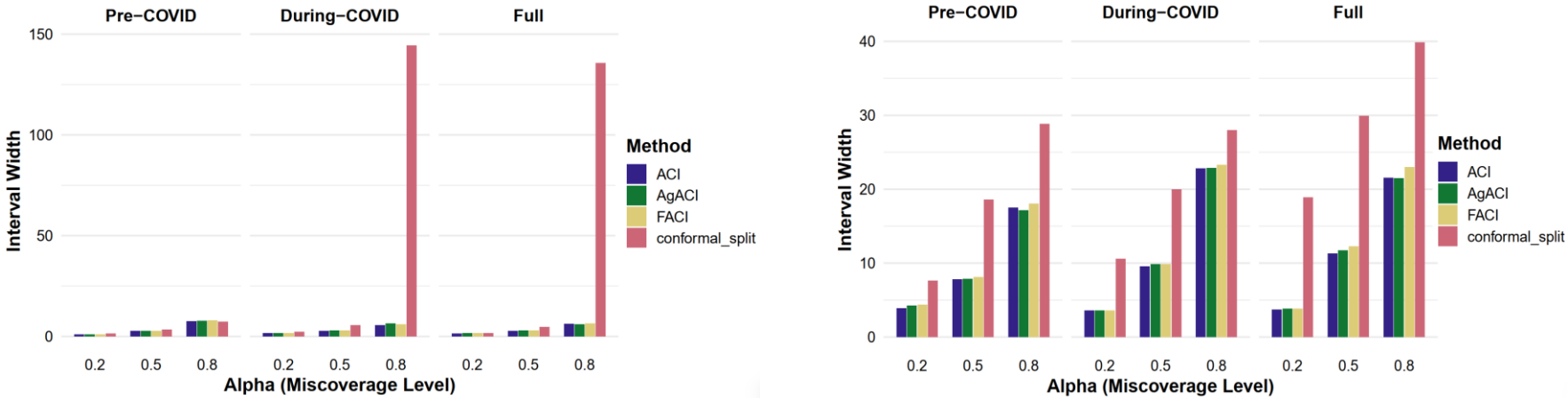
Below interval in Figure 4.6-d, for HND, the conformal-split method produces a higher fraction of true delays lying below the interval at every α (e.g., about 60 % at $\alpha = 0.2$ pre-COVID vs. ≈40 % for the adaptive methods). During-COVID the differences narrow, but the baseline still has more under-predictions at higher α. At NRT the pattern is similar: the baseline has a larger fraction of below-interval events during the pre-COVID and full periods (≈65-70 % at $\alpha = 0.2$ compared with 55-60 % for ACI/AgACI/FACI). Whereas the above interval (Figure 4.6-b), adaptive methods generally have higher above-interval percentages (i.e., their intervals are lower) than the baseline. At HND and NRT, ACI and AgACI have ~20-40 % of points above the interval at $\alpha = 0.2$, whereas the baseline has only ~10-20 %. For larger α the differences shrink, but the baseline remains slightly more conservative.

For ANA, the conformal-split baseline usually attains slightly higher empirical coverage, especially for high miscoverage levels, but at the cost of very wide intervals (particularly at HND and at $\alpha = 0.8$). Adaptive methods sacrifice a small amount of coverage for much narrower intervals. NRT data are easier to predict; all methods achieve closer to nominal coverage at $\alpha = 0.5$ and 0.8. At HND, heavy-tailed

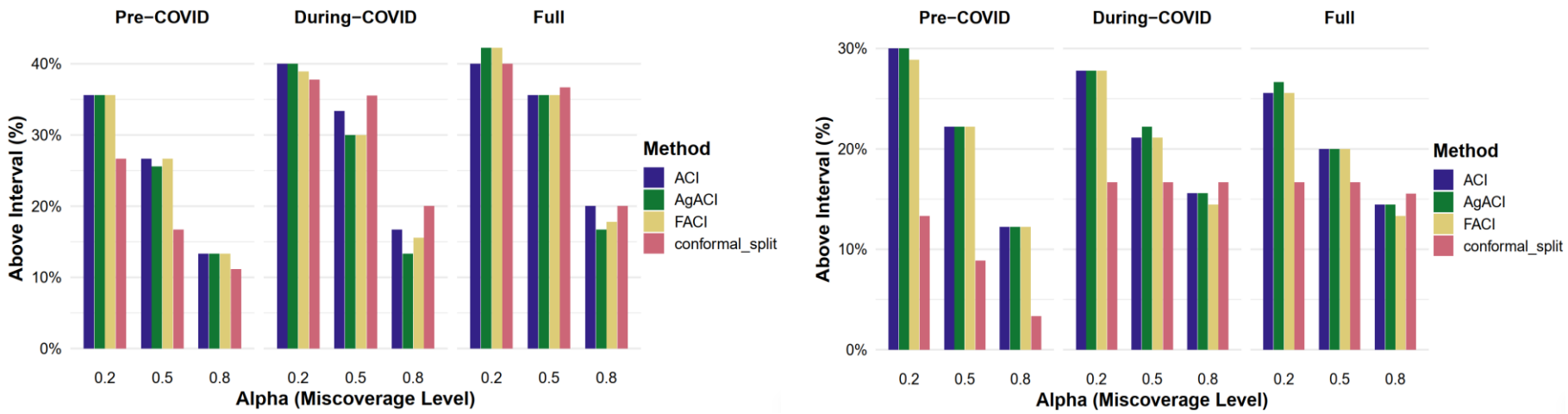
delays cause all methods to under-cover at $\alpha = 0.2$ and to require very wide intervals for the conformal-split baseline. Among the adaptive methods, AgACI often provides the best coverage (close to the baseline) while still maintaining narrow intervals. FACI sometimes has slightly wider intervals due to fairness constraints (e.g., pre-COVID at NRT), but also achieves more balanced miscoverage across above- and below-interval errors.



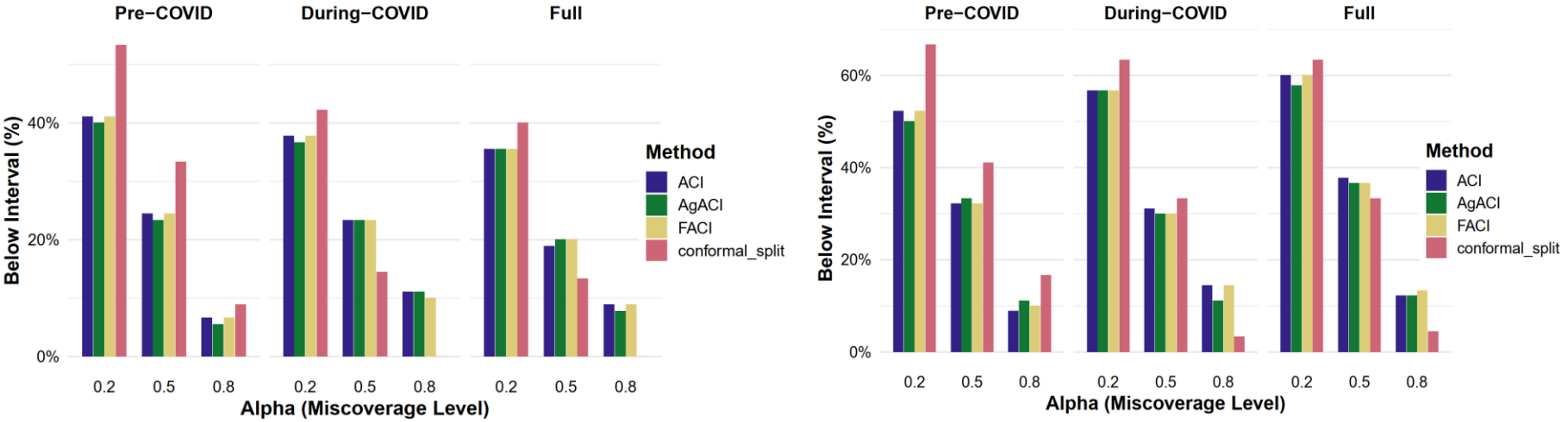
(a) Empirical Coverage



(b) Mean Interval Width



(c) Above Interval



(d) Below Interval

HND NRT

Figure 4.8: Comparison Summary Statistics of Adaptive Conformal Inference for JAL

For JAL in Figure 4.8-a, for empirical coverage, across both airports HND and NRT, the empirical coverage increases with α. For $\alpha = 0.2$ the desired coverage is only 20%, and the intervals from all models cover roughly 20 - 25 % of the departures; for $\alpha = 0.5$ the coverage is around 45-50 %; and for $\alpha = 0.8$ the coverage is about 80 %. On HND, the adaptive methods (especially FACI) achieve slightly higher coverage at $\alpha = 0.5$ and $\alpha = 0.8$ than the split-conformal baseline, while the baseline sometimes under-covers at $\alpha = 0.2$. On NRT the coverage is very similar across methods; FACI is marginally higher at $\alpha = 0.5$ and $\alpha = 0.8$, whereas the baseline achieves comparable coverage at $\alpha = 0.8$ but is slightly lower at $\alpha = 0.2$. Coverage patterns are consistent across time frames - Pre-COVID, During-COVID and Full - suggesting that all adaptive methods maintain their long-run coverage guarantees even when the distribution of delays shifts.

A key benefit of the adaptive methods for JAL is much narrower prediction intervals (Figure 4.8-b). At HND, the split-conformal baseline produces extremely wide intervals when $\alpha = 0.8$ (around 150 units wide during both pre- and during-COVID), reflecting a conservative response to distribution shifts. ACI, AgACI and FACI keep the widths around 4-5 units, achieving near-nominal coverage without inflating uncertainty. For $\alpha = 0.2$ and $\alpha = 0.5$, all adaptive methods yield small widths (≈1-3 units) compared with the baseline. On NRT the widths grow gradually with α: at $\alpha = 0.2$ the intervals are narrow (~3-4 units), at $\alpha = 0.5$ they widen to about 8-12 units, and at $\alpha = 0.8$ they reach 20-30 units. The baseline's intervals are consistently wider - especially at $\alpha = 0.8$ and during-COVID while AgACI often produces the narrowest intervals. These patterns highlight the ability of adaptive algorithms to adjust interval widths efficiently while maintaining coverage

The above interval (Figure 4.8-c) metric measures the proportion of departures with delays exceeding the upper bound. For both airports this proportion decreases as α increases: when $\alpha = 0.2$ (target coverage 20 %) about 20-30 % of departures fall above the interval, indicating that low-coverage intervals often underestimate delays. At $\alpha = 0.5$ the proportion drops to ~15-25 %, and at $\alpha = 0.8$ it falls further to 10-20 %. The split-conformal baseline has consistently lower above-interval percentages because its intervals are wider, particularly at HND where the baseline produces almost no above-interval violations for $\alpha = 0.8$. Among adaptive methods, ACI and AgACI sometimes have slightly higher above-interval percentages (reflecting narrower intervals) but remain close to the expected mis-coverage.

The "below interval" metric is generally larger than the above-interval metric, indicating that the models tend to over-estimate departure delays (actual delays fall below the lower bound). At $\alpha = 0.2$ roughly 40-50 % of HND observations and more than 50% of NRT observations lie below the lower bound, implying that short-coverage intervals overshoot the true delay. As α increases to 0.5, below-interval percentages fall to about 30-40 %, and at $\alpha = 0.8$ they drop further to 10-20 %. During-COVID the below-interval

percentages are slightly higher, especially at NRT where over-estimation exceeds 60% for $\alpha = 0.2$. The split-conformal baseline often has the highest below-interval rates because its wide, symmetric intervals allocate a large portion of mis-coverage to the lower tail. Among adaptive methods, AgACI and FACI generally show lower below-interval percentages than ACI, reflecting better calibration of the interval centre.

Overall, for JAL, the adaptive conformal regression models (ACI, AgACI and FACI) maintain empirical coverage close to the target mis-coverage while producing much narrower prediction intervals than the standard split-conformal approach. FACI's exponential re-weighting of multiple ACI experts and AgACI's aggregation of different learning rates help these methods achieve slightly higher coverage with modest width increases. The split-conformal baseline remains valid but often resorts to overly wide intervals especially under distribution shifts such as COVID-19 leading to low above-interval and high below-interval percentages.

4.5 Discussion

The adaptive algorithms behave differently at HND versus NRT. At HND the distribution of ANA's departure delays is heavy-tailed; none of the methods achieved the nominal 80% coverage for $\alpha = 0.2$ and the standard conformal-split baseline over-conservatively widened intervals beyond 150 minutes. ACI, AgACI and FACI all generated much tighter intervals (mostly under 10 minutes) but under-covered by 5-10 percentage points at high coverage levels ($\alpha = 0.8$). In contrast, at NRT the delay distribution is more predictable; the adaptive methods achieved coverage comparable to the split baseline for $\alpha \geq 0.5$ and their interval widths remained modest (≈15 minutes at $\alpha = 0.8$). For ANA, AgACI often provided the closest coverage to the baseline while maintaining narrow intervals, whereas FACI occasionally generated wider intervals because its fairness constraints spread mis-coverage more evenly between above- and below-interval violations. For JAL, the adaptive methods maintained near-nominal coverage for all α values and produced much narrower intervals than split-conformal, particularly at HND where FACI and AgACI delivered 4-5-minute intervals for $\alpha = 0.8$. FACI's re-weighting of experts allowed it to approximate the baseline coverage at $\alpha = 0.5$ and 0.8 without AgACI's extreme interval inflation, making it a good middle ground. A persistent observation is the imbalance between below- and above-interval errors. Split-conformal assigns much of the mis-coverage to the lower tail; at HND over 60% of true delays lie below the interval for $\alpha = 0.2$, while adaptive methods reduce this to roughly 40%, shifting mis-coverage toward above-interval violations. FACI, because of its fairness-aware re-weighting, often achieves a better balance between over- and under-estimation than either ACI or AgACI. In operational terms this means ACI and AgACI may underestimate delays more often, risking

unpreparedness for prolonged disruptions, whereas the baseline and FACI may overestimate delays, encouraging conservative buffers.

Regulators and airport operators should encourage airlines to use adaptive conformal methods with fairness constraints. Heavy-tailed delay distributions at major hubs like HND require models that widen quickly during disruptions but also retract when conditions normalized. FACI's balanced mis-coverage reduces systemic bias in delay predictions; such fairness is essential for equitable slot allocation and avoiding persistent over- or under-estimation for particular flights. Performance benchmarks could be based on the high on-time rates achieved by ANA and JAL. According to FAA GDP manual, the U.S. FAA's GDP assigns arrival slots and equitably distributes delays while dynamically updating schedules using the Flight Schedule Monitor. Japan could adopt similar collaborative decision-making frameworks to manage congestion at Haneda and Narita. A dynamic GDP-like system would assign delays based on real-time predictive intervals and allow user options for swapping slots, reducing airborne holding and ensuring that all carriers, including ANA and JAL, share delay burdens fairly.

Policymakers should incentivize investments in predictive maintenance, demand forecasting and adaptive conformal inference to improve resilience. Cross-carrier data sharing through integrated platforms would enhance these models, enabling rapid adjustments when demand shifts. Some airport specific strategies can be done - like - Because NRT delays are easier to predict, leaner methods like ACI or moderate aggregation may be suitable, freeing resources for other operational improvements. At HND, where heavy-tailed delays persist, AgACI or FACI should be standard, and regulators might require additional buffers or incentives for airlines to maintain resilience during peak volatility. Policy should reflect local conditions rather than imposing uniform requirements across airports. As regulators adopt adaptive models, they should ensure that delay-prediction errors do not disproportionately affect certain flights. Equitable delay assignment, as emphasized in the FAA's GDP policy, should be mirrored in Japanese air-traffic management.

4.6 Chapter Conclusion

In this paper, we set out to identify the key network-centric features that drive day-to-day average departure delays at Japan's eight major airports, and to develop and empirically compare three adaptive conformal prediction schemes: ACI, AgACI, and FACI, to quantify uncertainty around these average-delay forecasts. To our knowledge, this is the first application of conformal inference in the air-traffic management domain, and our results demonstrate both the promise and the practical trade-offs of these methods.

Our random forest analyses consistently show that lagged delay is the most important predictor of next-day average departure delay across all airports and for both ANA and JAL carriers followed by some temporal predictor and importantly network predictors. This underscores the strong temporal correlation in delay patterns: yesterday's congestions almost invariably carry over into today's operations. Beyond this, network centrality metrics, especially, in- and out-degree and betweenness centrality regularly rank among the top five predictors demonstrating that an airport's position and connectivity in the route network significantly shape its delay exposure. Seasonal factors (month) also contribute meaningfully suggesting the importance of temporal and seasonal dependence on the forecasting departure delays. While lagged delays remain the most influential single predictor, adding structural and temporal context consistently improves model accuracy and explanatory power as evidenced by lower MAE and RMSE and higher R^2. These findings underscore the value of combining temporal persistence, network position, and seasonal factors in predicting departure delays.

In this work, we have brought conformal prediction (CP) into the realm of flight-delay forecasting, quantifying uncertainty around average daily departure delay estimates. We have uncovered patterns of under- and over-coverage that are tightly linked to each airport's characteristic delay volatility. ACI's single learning-rate update delivers the narrowest bands but systematically under-covers during sudden delay surges, making it well suited to stable, low-variance contexts. AgACI's ensemble of learning rates closely tracks nominal coverage even under extreme spikes, albeit with substantially wider intervals, a conservatism that is highly valuable for strategic planning and disruption management. FACI's exponential re-weighting strikes a middle ground, adapting almost as swiftly as AgACI while maintaining much tighter bands, and thus offers an attractive balance between precision and reliability for potential real-time decision support. In comparing ANA and JAL at both Haneda and Narita, we find that JAL's conformal intervals are roughly 30-40 % narrower than ANA's, yielding more precise but slightly less reliable forecasts, while ANA's wider bands provide stronger coverage at the expense of precision. Coverage quantifies how often true delays fall inside the predicted range, mis-coverage patterns reveal whether we tend to under- or over-estimate delays, and mean interval width captures the precision versus safety trade-off. In practice, airline and airport managers can choose JAL-style tight intervals to maximize gate and crew utilization when delay variability is moderate, or ANA-style conservative bands to guard against extreme surges in high-volatility environments. This demonstrates the practical benefit of adaptive conformal inference: it tailors uncertainty quantification to each airport's delay profile, enabling more informed, context-sensitive operational decisions.

Our findings echo and extend the results of Ribeiro et al. (2025) in demonstrating that data-driven interval methods by outperforming static and deterministic approaches where in their case, gradient boosting versus linear regression, and in our study, adaptive versus single-rate updates, by delivering prediction bands that both adapt to the data and reflect the true uncertainty inherent in delay dynamics. As Ribeiro et al. (2025) illustrated for airport capacity management, meaningful uncertainty quantification is critical to assessing operational robustness beyond average-delay metrics alone. Similarly, our CP intervals expand most when delay variability is highest (e.g., around day 60 surges), and contract again as the system stabilizes, thereby offering a nuanced, context-sensitive gauge of risk that simple quantile regressions or fixed bands cannot provide.

This result of framework can help airlines, airports and air traffic managers where they can choose their CP tool according to each airport's volatility profile and their tolerance for interval width versus mis-coverage risk. In routine, low-volatility feeder-hub settings, ACI's tight bands can maximize resource efficiency and throughput. At high-volatility trunk hubs like Haneda, especially during exogenous shocks like SoEs, AgACI's robust coverage ensures downstream delays are anticipated, preventing cascading disruptions. FACI emerges as an all-purpose solution, widening intervals promptly when needed but retracting sooner during calm periods. By embedding these adaptive intervals into prescriptive analytics workflows with capacity planning, operators can shift from deterministic planning to a more resilient, uncertainty-aware regime, ultimately enhancing both punctuality and robustness in today's dynamic air-transport networks.

Future works hold various exciting possibilities. For example: along with network metrices, other factors, for instance: weather patterns, land-side airport-centric variables may reduce the uncertainty of the delay forecast. This particular framework has only been tested in Japanese domestic market. Future scopes hold opportunity to transfer it for other domestic air transport systems in other countries as well as international flight network cases. From methodological perspective, we have more options to hold future works. For instance: numerous experiments can be done with a variety of machine-learning and deep-learning techniques to evaluate which methods produce the narrowest prediction intervals for delay estimates. Adjustment and tuning of the conformal update learning rate can be experimented to identify the optimal γ that balances rapid adaptation and interval stability specifically for flight-delay forecasting. Also, this uncertainty-quantification framework can be extended beyond departure delays, applying it equally to arrival delays, total journey delays, or other performance metrics.

Chapter 5 Identifying Persistent Departure Delay in Japanese Aviation Network using Topological Data Analysis

5.1 Introduction

Managing departure delays remains a major challenge in air transportation. Even short gate-departure delays can propagate across tightly scheduled operations. This leads to significant disruptions in airline networks and among interconnected airports. Departure delays arise from various factors. These include such as operational inefficiencies, airport congestion, air-traffic control instructions, weather events, and human limitations (ACRP, 2014; CAAC, 2021). Due to the interdependency of flights through aircraft rotations, crew links, and connecting passengers, a delay in one segment can trigger subsequent delays across space and time (Wu, 2010; Choi et al., 2016).

Arrival delays, while relevant, often reflect merged effects from routing changes, airborne buffers, and en-route disturbances. This makes it harder to identify where the delay originated. By focusing specifically on departure delays, we can more accurately diagnose network vulnerabilities. Additionally, departure delay has been shown to predict arrival delay (Zheng et al., 2021). This justifies its use as the core metric in this study.

Various studies have used graph-theoretic and network-science methods to understand delay propagation. These studies model airports as nodes and delay or flight connections as edges. Such work has uncovered influential hubs and community-level delay structures (Wandelt et al., 2025a; Sano & Berton, 2021; Tang et al., 2021; Sugishita et al., 2024). Spatio-temporal models further improve prediction accuracy by modeling how delays evolve across time and network topology structure (Rebollo & Balakrishnan, 2014; Gopalakrishnan & Balakrishnan, 2017; Zheng et al., 2024; Wu et al., 2024).

However, most classical network analyses focus on pairwise relationships. These methods are limited to dyadic interactions. They do not fully capture higher-order delay propagation structures. For example, if three or more airports experience correlated delays in a cyclic pattern, traditional metrics cannot identify this loop explicitly. Such feedback loops may play a major role in delay escalation.

To investigate dyads, some researchers have proposed methods like delay causality networks and motif analysis (Du et al., 2018; Sugishita et al., 2024). These uncover structures that propagate delays directionally and non-randomly. However, such methods are still limited by the need for thresholding or causal estimation. Topological Data Analysis (TDA), especially persistent homology offers a powerful alternative. TDA captures clusters and loops from data by constructing a growing simplicial complex

(Aktas et al., 2019; Li et al., 2019). In our case, we define a network where airports are connected when their average delay falls below a threshold. We then build triangles when all three pairwise links meet the same criterion, and extend to higher-dimensional simplices.

This construction allows tracking of 0-dimensional components (connected clusters) and 1-dimensional holes (loops). Persistence measures how long these features survive as the threshold increases. Longer-lived features are more likely to reflect stable and significant delay structures (Cuerno et al., 2025). Persistent homology avoids arbitrary thresholding and provides a multi-scale view of delay propagation. TDA has been applied in other transportation and network domains. For example, Rajput et al. (2023) used TDA to measure flood-induced disruptions in road networks using travel time data. Hickok et al. (2024) applied persistent homology to evaluate polling site accessibility and detect underserved regions. These studies demonstrate that TDA excels in identifying persistent spatial structures and feedback loops.

This study applies persistent homology to four years of Japanese domestic flight delay data (2018–2021). We focus on two major carriers: All Nippon Airways (ANA) and Japan Airlines (JAL). Together, these carriers represent over 70% of Japan's domestic market (Ng et al., 2022). Despite operational similarities, their delay dynamics may differ. Previous work by Sadeek et al. (2025) shows that their delay network metrics are not always aligned. In this study, our objectives are three-fold: i) to compare clustering and loop structures in ANA and JAL delay networks; ii) to study how homological features and their persistence change during normal, pandemic, and recovery periods; iii) to identify persistent delay loops and the specific airport sets that repeatedly participate in them.

Evidence supports the existence of repeatable delay motifs. Sugishita et al. (2024) found recurrent delay propagation patterns in Japan's domestic network. These varied by season and airline. While their method was based on causality networks, it implies that persistent topological loops may be present and detectable. Our method offers a formal another way to track and quantify those loops. While motif-based analyses can uncover frequently occurring subgraph patterns, they are inherently limited to small, fixed-size structures typically involving three or four nodes. These motifs capture local connectivity but cannot fully account for nested or overlapping loops. Also, they do not provide a multiscale view of their persistence across varying thresholds . In contrast, TDA, using simplicial complexes, captures both local and global geometric features, such as nested clusters and higher-order loops, regardless of their size or embedding. Moreover, persistent homology quantifies not only the presence of these features but also their lifespan across a filtration, offering a robustness criterion that motif counting lacks. This means TDA can distinguish between transient, noisy loops and truly structural feedback loops in the delay network. Unlike motifs, which are static and threshold-dependent, persistent homology tracks the

evolution of loops over a continuum of network states, revealing deeper, more stable patterns of delay interdependency.

Li et al. (2019) used TDA to reveal inefficiencies on U.S. airport surfaces. Their persistent features aligned with real traffic constraints. These results support the use of persistent homology in aviation settings. Delay propagation, especially in cyclical structures, requires tools that go beyond traditional metrics. Additional work by Chen et al. (2020) shows that delay communities do not always match physical networks. They used correlation networks and found geographically clustered delay behaviors. In TDA terms, these appear as persistent clusters. If inter-regional loops exist, PH persistent homology can detect them as 1-dimensional loops. Güvercin et al. (2021) demonstrated that network-aware clustering improves delay prediction. Their model used betweenness centrality and community membership to enhance forecast accuracy. This supports our approach: persistent loops offer an alternative, data-driven way to uncover such critical structures. It reveals the topological backbone of delay propagation and complements existing modeling efforts. Overall, this study contributes a multi-scale, topological framework for delay analysis. It highlights persistent delay loops that inform scheduling, resilience, and mitigation strategies in air traffic operations.

5.2 Study Area and Data

In this study, data on Japanese domestic flights from 2018 to 2021 were analyzed using the Official Aviation Guide (OAG) historical flight data (OAG, 2021). The dataset contains detailed information on each flight, including departure and arrival airports, scheduled, estimated, and actual departure times, and operating airlines. The analysis focused on two full-service carriers in Japan: ANA and JAL. From this dataset, departure delay information was extracted to construct departure delay networks. Daily snapshots of the departure delay networks were created to represent the departure delays of domestic Japanese flights from 2018 to 2021.

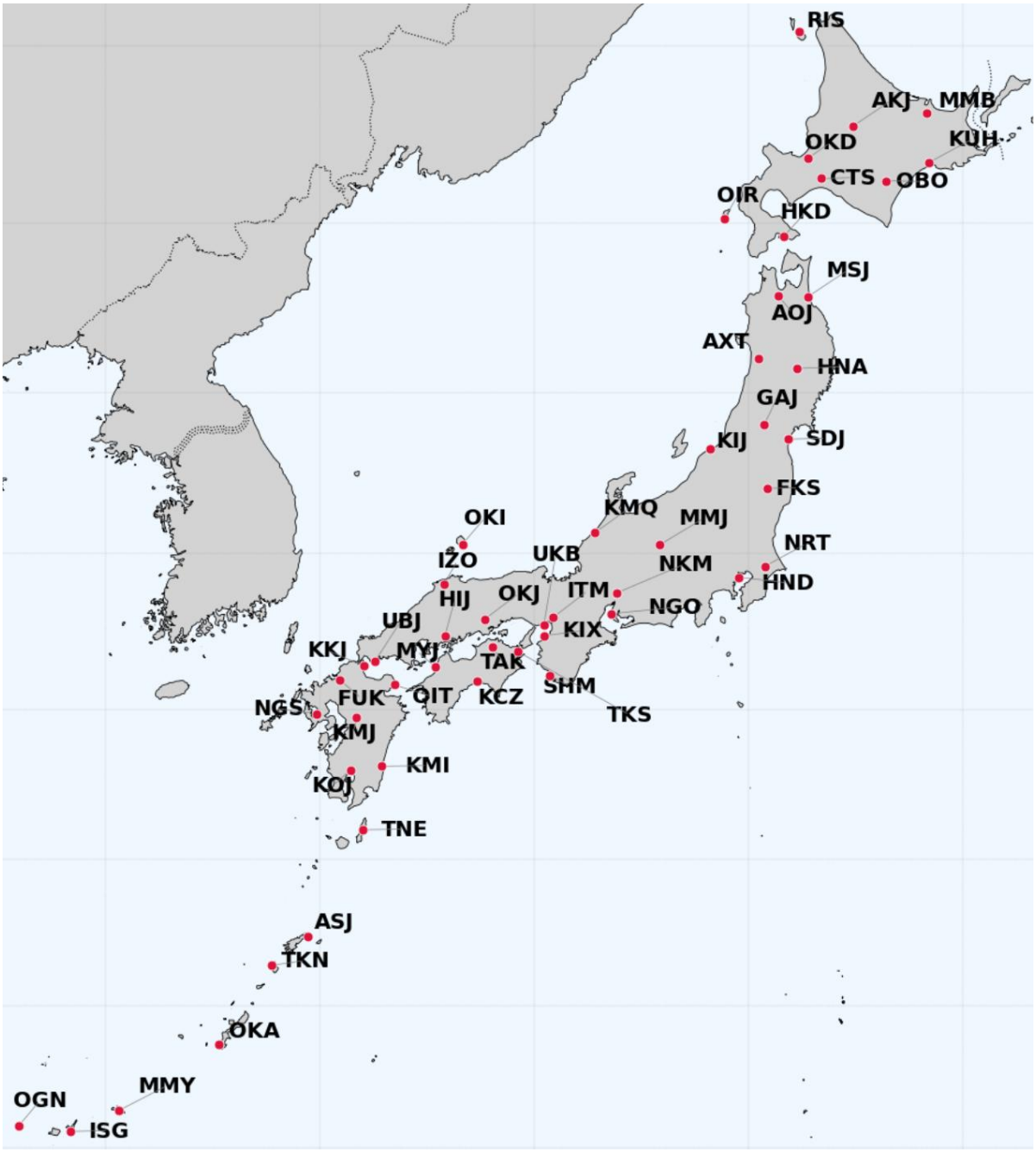


Figure 5.1: Major domestic airports in Japan for ANA and JAL flight operation

Figure 5.1 represents the domestic airports considered in this analysis where ANA and JAL operate their flights. For ANA, 36 airports and for JAL, 47 airports are considered initially while extracting domestic flight data. These airports are selected from the major airports based on the airport development law determined by the Ministry of Land, Infrastructure, Transport and Tourism (MLIT, 2025).

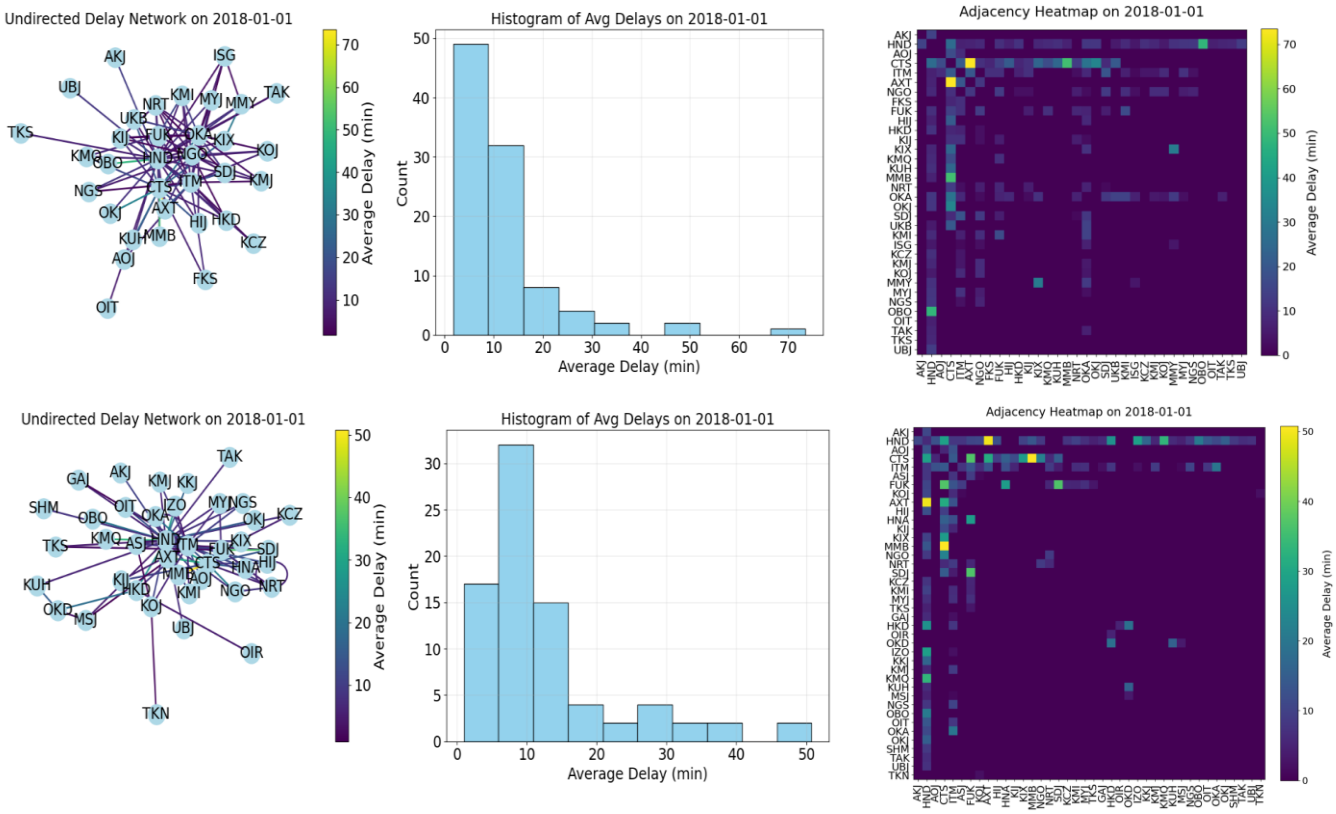


Figure 5.2: Average departure Delay network (undirected), histogram of delay and adjacency matrix of average delay of delayed flight operation for airport pairs for ANA and JAL

In our delay networks, each airport is represented by a node and an edge is drawn only between pairs of airports that experienced at least one departure delay of one minute or more. For every such pair, we computed the average departure delay over all delayed flights on that route during the day. Figure 5.2 (for 2018-01-01) displays, for both ANA and JAL, the corresponding delay-network graph, a histogram of the delay distribution, and a heatmap of these daily mean delays (the adjacency matrix). For our topological data analysis, each day's delay network requires a cutoff or threshold to reveal its underlying shapes and holes. We, therefore, chose the day's maximum average departure delay across all airport pairs as that cutoff. In Figure 5.2, for example, ANA's largest mean delay is about 70 minutes on the CTS-AXT route, and JAL's is roughly 50 minutes on the HND-AXT route; accordingly, we set the threshold for ANA at 70 minutes and for JAL at 50 minutes. By using the maximum value each day, we ensure that every possible loop or void in the delay network is exposed, giving us a complete picture of its topological structure. Also, we compared seasonal delay network variation in terms of TDA. In this analysis, the season indicates: Summer from April to October and Winter from November to March according to Japan Schedule Coordination and Japan Aeronautic Association (2024). Figure 5.3 represents a flow-chart about how the undirected weighted delay network data is converted into topological information using

Persistent Homology Concepts. The flow-chart represents an example with four airport nodes and the corresponding average departure delay per routes converting into adjacency matrix. Then with Vietoris-Rips complex implementation, we track the evolution of various topological features like H_0 and H_1 and get the birth time, death time and persistence of each of the features.

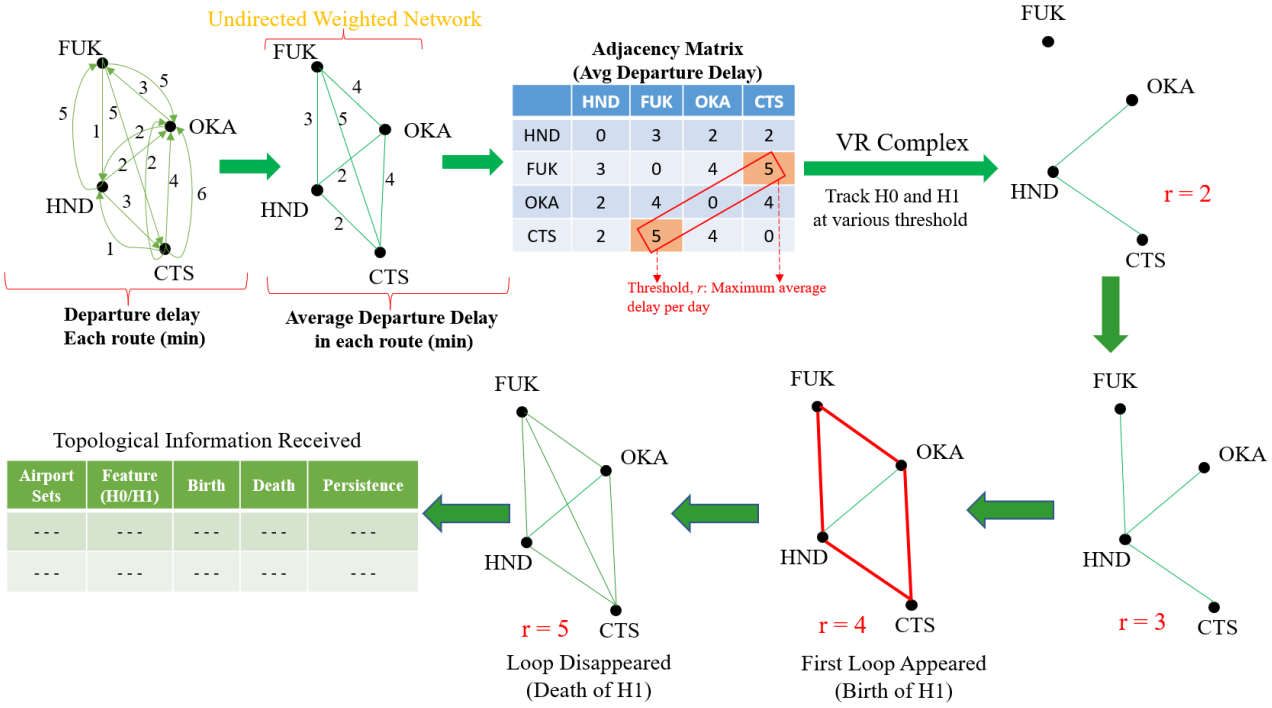


Figure 5.3: Flow-chart describing extracting topological information derived from the delay network

5.3 Modeling Method

Traditional network measurement techniques in a disrupted network can miss deeper shifts in its overall shape. By investigating higher-order structures e.g., how groups of nodes connect beyond simple pairs, we uncover more of the network's evolving topology during disruptions like Covid-19 pandemic (Dey et al., 2019). To reveal these hidden dynamics in terms of departure delay of flight operation, we analyze daily Japanese domestic flight departure delays (January 2018-December 2021) by building undirected, edge-weighted graphs, forming a Vietoris-Rips (VR) simplicial complexes (Carlsson, 2009) – one of the most popular TDA filtration methods tracking the evolution of various topological features using Betti numbers across different delay thresholds, and computing persistent homology (Torres and Ginestra, 2020). Betti numbers are core topological invariants that describe these multi-node relationships when we model a network as a simplicial complex (Bianconi, 2021). Mathematically we can represent this concept as following:

Let, $V = \{v_1, v_1, \ldots, v_n\}$ be the set of *n* domestic airports in Japan. For each day, *t* and each pair (v_i, v_j), we compute the average departure delay

$$w_{ij} = \frac{1}{N_{ij}(t)} \sum_{k=1}^{N_{ij}(t)} \Delta_k(t) \tag{5.1}$$

where, $N_{ij}(t)$ is the number of flights from v_i *to* v_j on day *t* and $\Delta_k(t)$ is the departure delay (in minutes) of flight *k*.

For any delay threshold, $r \leq 1$, define the graph as

$$G(t; r) = \big(V, E(t; \varepsilon)\big), where, E(t; \varepsilon) = \{\{ v_i, v_j \}: w_{ij} \leq r\} \tag{5.2}$$

In particular, $G_1(t) = G(t; 1)$ captures all non-zero delays.

After than we convert each graphs into a simplicial complex by taking cliques. Given as undirected graph *G = (V, E)*, its VR complex, *VR(G)* is the simplicial complex whose simplices are all complete subgraphs of *G*. Concretely, for each day *t* and each threshold *r*, set

$$A_t(r) = VR(G(t; r)) \tag{5.3}$$

As *r* increases from 1 to the day's maximum average delay, these complexes for a filtration

$$A_t(r_o) \subset A_t(r_1) \subset A_t(r_2) \ldots\ldots \subset A_t(r_m), r_o < r_1 < r_2 \ldots < r_m \tag{5.4}$$

where, $A_t(r_k) = VR(G(t; r_k)), G(t; r_k) = \{\{ i, j \}: w_{ij} \leq r_k \}$

here, $r_o = 0$ (all airports isolated) and $r_m = max\{w_{ij}(t)\}$ (fully-filled complex).

For any simplicial complex *A*, its d^{th} homology group $H_d(A)$ has dimension

$$\beta_d (A) = dim\, H_d(A) \tag{5.5}$$

called d^{th} betti number where β_o = number of connected components and β_1 = number of independent loops and so on.

In easy language, The VR-filtration produces a sequence of simplicial complexes that represent how the shape of our data evolves as we gradually increase the scale parameter, r. Because our dataset has only a finite number of points, the Vietoris–Rips complex does not change infinitely often in practice: there are only finitely many distinct configurations as r grows. Formally, although the filtration is defined over all positive real values of r, we only need to consider a finite collection of nested simplicial complexes:

$$A_0 \subset A_1 \subset \ldots \subset A_t$$

(5.6)

where, $A_0 = \emptyset$ is the empty complex (i.e., no points or simplices), and A_t is the fully connected complex, containing all possible simplices (i.e., edges, triangles, etc)

Our goal is not only to count how many holes (connected components, loops, voids, etc.) each complex has at each step but also to track when each hole appears and disappears across the filtration. To do this systematically, we rely on homology, which assigns to every simplicial complex a sequence of vector spaces that record these holes in different dimensions. Specifically:

$$H_0(A), H_1(A), H_2(A), \cdots, \tag{5.7}$$

where, $H_0(A)$ measures connected components (0-dimensional holes), $H_1(A)$ measures loops (1-dimensional holes), $H_2(A)$ measures void (2-dimnesional holes), and so on. The dimension of each homology vector space is called Betti number in the dimension:

$$\dim(H_d(A)) = \beta_d \tag{5.8}$$

It means β_0 counts the number of connected components, β_1 counts the number of loops, β_2 counts the number of enclosed voids, etc. Importantly, each homology space has a basis of loops representing those holes.

A key property that makes homology useful for persistent homology is called functoriality. In simple terms, this property ensures that if we have two simplicial complexes where one is contained inside the other as $A \subset A'$ then in every dimension d, there is a linear map:

$$T : H_d(A) \to H_d(A') \tag{5.9}$$

This linear map lets us track how loops (holes) evolve as we move from a smaller complex A to a larger one A'. Specifically, if a loop $c \in H_d(A)$ maps to zero $(T(c) = 0)$, the hole has died in A'. If $T(c) \neq 0$, the hole persists. Now if we consider the full filtration in equation (), applying homology in dimension, d, we obtain a sequence of linear transformations:

$$H_d(A_0) \xrightarrow{T_0} H_d(A_1) \xrightarrow{T_1} H_d(A_1) \cdots \xrightarrow{T_{t-2}} H_d(A_{t-1}) \xrightarrow{T_{t-1}} H_d(A_t) \tag{5.10}$$

This sequence allows us to record exactly when a hole appears (“birth”) and disappears (“death”).

A loop c is born at index i if it does not come from the previous complex as if:

$$c \notin \mathrm{Im}\, age(T_{i-1})$$

(5.11)

Eqn (5.11) represents intuitively that the feature was not present before step i.

The loop dies at index j if when mapped forward, it becomes zero:

$$T_j(c) = 0 \tag{5.12}$$

Eqn (5.12) means the feature is no longer present in A_{j+1}

The whole calculation of this feature birth and death is conducted through boundary matrix. The boundary matrix keeps track of which simplices are faces of other simplices. If D_d is the boundary matrix mapping d-simplices to $(d-1)$ simplices, then

$$\beta_d(A) = |\Sigma_d(A) - rank(D_d) - rank(D_{d+1})| \tag{5.13}$$

where, $\Sigma_d(A)$ is the set of *d-simplices*.

An example and matrix formulation is given in Appendix E1. Note that D_d encode how everything fits together in a boundary matrix. Then we reduce that matrix to figure out which splices created features and which simplices kill them. Then we track the birth and death as below.

A homology class γ births at r_A and dies at r_l for the first time. The we record its birth-death pair as:

$$(r_{birth}, r_{death}) = (r_k, r_l) \tag{5.14}$$

If γ never dies, we write (r_k, ∞). The collection of these pairs in dimension d,

$$P_d(t) = \{ r_{birth}, r_{death} \} \tag{5.15}$$

which is the *persistence diagram*. An equivalent *"barcode"* representation draws each $[r_{birth}, r_{death})$ as a horizontal bar.

The algorithm of this whole process is as follows (adapted from Elchesen, 2025):

Once the filtration has been prepared and ordered by simplices, the next step is to build what is called the boundary matrix. This matrix is similar in spirit to the boundary matrices D_1, D_2, and so on, which are used to compute Betti numbers. However, in this context, all simplices from all dimensions are combined into a single matrix rather than treated separately. Formally, suppose we have a filtration consisting of a

nested sequence of simplicial complexes as of Equation 5.6 where each simplex is indexed as $\sigma_1, \sigma_1, \cdots, \sigma_t$. The boundary matrix D is then defined so that it has one row and one column corresponding to each simplex in this list. For any pair of simplices σ_i and σ_j, the matrix $D(\sigma_i, \sigma_j)$ is set to the following rule:

$$D(\sigma_i, \sigma_j) = \begin{cases} 1 & if\ \sigma_i\ is\ a\ face\ of\ \sigma_j \\ 0 & otherwise \end{cases} \tag{5.16}$$

This construction creates a square matrix encoding which simplices are faces of which others across the entire filtration. The main purpose of the boundary matrix is to serve as the input to the persistence algorithm, which processes it to extract the so-called birth-death pairs. Importantly, all calculations are carried out using $mod - 2$ arithmetic, meaning that whenever two identical rows or columns are added, they cancel each other out (since 1 + 1 = 0).

For each column j of the matrix, corresponding to simplex σ_j we define the function $low(j)$ as the index of the lowest row containing a 1 in that column. This position is often called pivot. If the column is all zeros, then $low(j)$ is set to zero.

The persistence algorithm proceeds column by column, moving from left to right. At each column j, the algorithm checks all earlier columns $i < j$. If there is an earlier column where $low(i) = low(j)$, this means two simplices share the same pivot row, so we add column i to column j $(mod - 2)$ to eliminate that pivot. This addition changes the position of the pivot in column j effectively "raising" it higher in the column. After this operation, there might still be other columns to the left that share the new pivot, so we repeat the process until no such columns remain or until the column becomes all zeros.

Once all columns have been processed in this way, the boundary matrix is said to be reduced. At this point, we can read off the birth and death of topological features directly from the reduced matrix. Specifically, if a column j zero, this indicates that the corresponding simplex created a new feature (such as a connected component or a loop) that has not yet been killed by any other simplex. Conversely, if column j is nonzero, then the pivot row $i = low(j)$ shows that the feature born when σ_i was added is killed exactly when σ_j is added. Therefore, every nonzero column encodes a birth-death pair $(low(j), j)$. This pair means the feature was born when simplex $\sigma_{low(j)}$ appeared and died when simplex σ_j appeared.

It is also possible for a feature to be born but never to die anywhere in the filtration. In that case, we record this by pairing the birth index b with infinity, writing (b, ∞). To finalize the persistence diagram, we build two lists: B, which will hold indices of simplices that gave birth to features, and P, which will store the resulting birth-death pairs. We then loop through the reduced boundary matrix from left to right.

Whenever we encounter a zero column i, we add its index to B. When we find a nonzero column j, we look at the pivot $low(j)$. We then record the pair $(low(j), j)$ in P and remove $low(j)$ from B because its feature has been matched with a death. After processing all columns, any indices that remain in B correspond to features that were born but never died. For each such index i, we add (i, ∞) to P. The set P is then the complete persistence diagram summarizing when topological feature in the filtration.

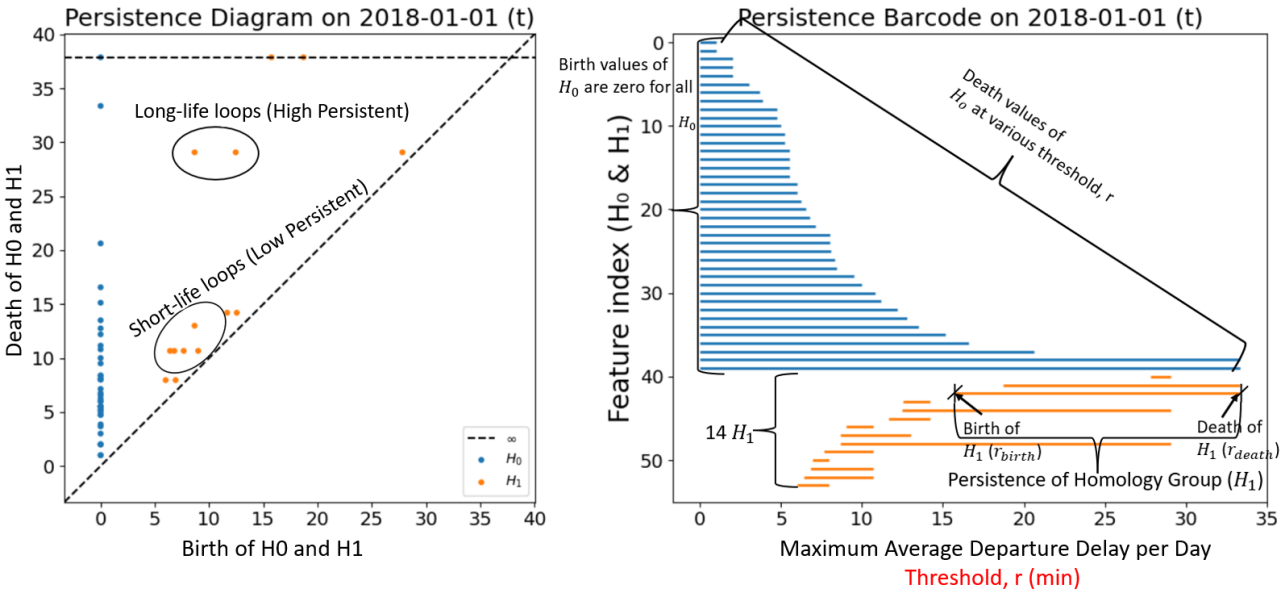


Figure 5.4: An example of topological threshold, r and its corresponding persistent diagram (left) and persistent barcode (right) for JAL on 2018-01-01 (t). Here, Ho refers to the connected component (0D) and H1 refers to the loops (1D) of the simplicial complex.

In our analysis, two H_d that is, H_0(connected components) and H_1 (loops) are the two homology groups of interest (See Figure 5.4). H_0 death values mark the delay threshold at which this airport component merges with another via a delay-based edge. It describes how initially disconnected airports gradually connect into components as the delay threshold increases. The death value itself reflects how isolated or dissimilar the airports' delay profile is from others. It represents how distributed the departure delay for ANA/ JAL comprising of there connected airports. If the distribution is spreaded then the airports experience larger delays and it distribution shrinks then those airports experience smaller duration of delays. In contrast, H_1 features are described by their birth and death thresholds: birth is when a multi-airport loop first appears, and death is when a higher-order simplex fills it in. Together, these values reveal robust loop of delay among airport groups. The birth of H_1 represents when multiple airport sets appear in the loop and a delay loop continues to exist among those airports for some time. That continuation is represented by Persistence value. High persistence (death - birth), especially in H_1, indicates structurally significant delay-propagation patterns that endure across many thresholds, while

low-persistence features usually reflect transient or noisy effects. Performing this analysis every day over four years produces a multiscale topological portrait of Japan's domestic departure-delay network, allowing us to trace seasonal dynamics and to quantify how COVID-19 reshaped both network connectivity (H_0) and loops (H_1). Figure 5.5 represents the work-flow of the VR method explained above.

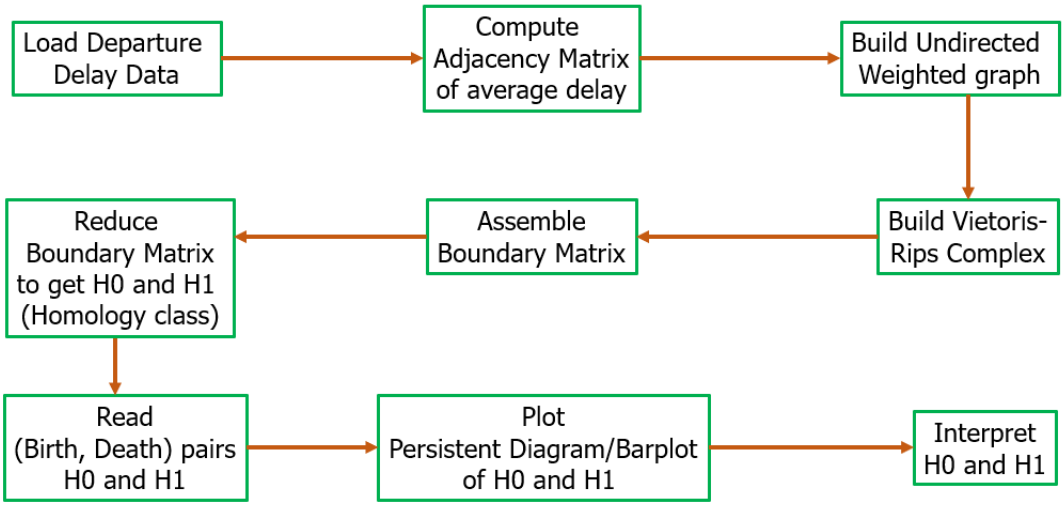


Figure 5.5: Methodology flow chart of the Vietoris-Rips (VR) complex and its resulting parameters

5.4. Results and Discussions

In this section, we delve into seasonal comparisons and airport-specific loop behaviors and provide a general understanding of delay dynamics as captured through topological features. These features, specifically, the birth, death, and persistence of topological loops serve as the building blocks for analyzing structural feedback in delay propagation. Figure 5.6 shows a flow chart of how various results have been derived from the information of extracted H_0 and H_1. This flow chart helps to understand the flow of various results interpreted from Section 5.4.1 to 5.4.7.

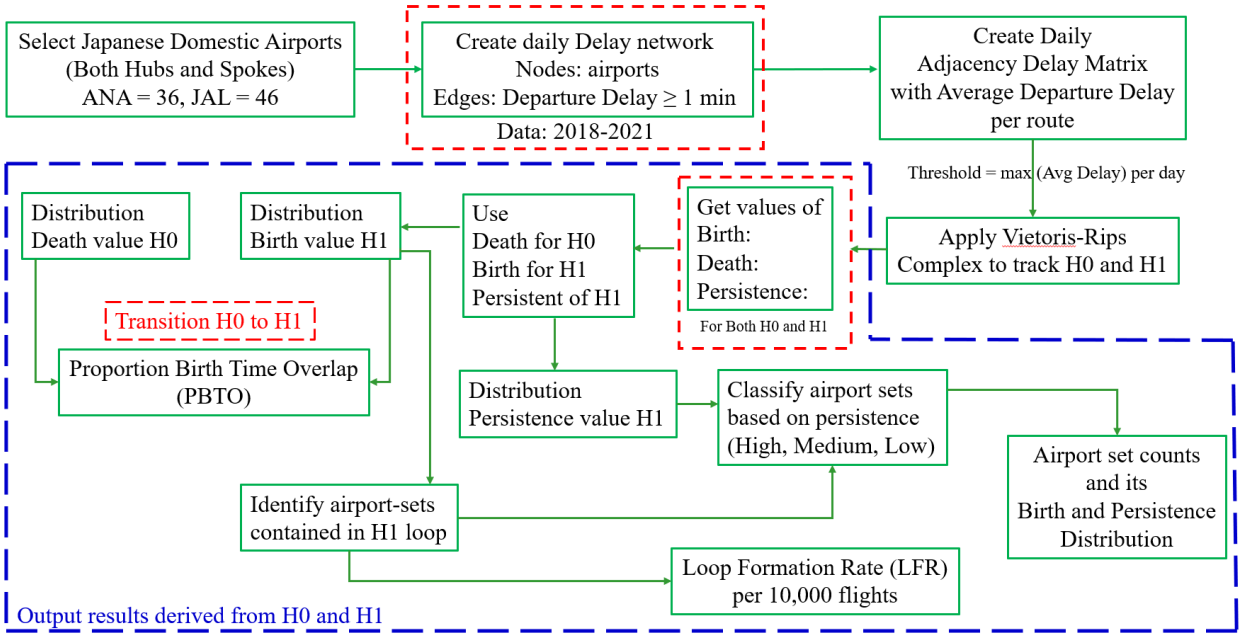


Figure 5.6: Flow-chart of the various output results derived from H_0 and H_1 of the delay network

5.4.1 Distribution of death value of H_0

The histograms of H_0 death-value in Figure 5.7 show how long groups of airports remain separate before becoming connected to the rest of the network as the delay threshold increases. It reflects how spread out the average departure delays are between airports during the summer and winter seasons from 2018-2021. In most seasons and years, about 80% of these separate airport groups connect to the wider network once the delay threshold reaches 5-15 minutes, while only a few remain unconnected until delays exceed 100 minutes. Intuitively, it suggests that for both ANA and JAL, most airport pairs become connected in the delay network when their average departure delays fall in the range of about 5 to 15 minutes, while only a few airport routes exhibit much longer delays, often exceeding 100 minutes. Before the pandemic, peaks in both airlines' distributions occurred at 5-10 minutes, with winter distributions showing faster reconnection, especially for JAL. In 2020, during widespread flight reductions, these times shortened to 3-8 minutes. By 2021, the dominant 5-10 minute peak returned, but winter (2021) introduced a secondary mode between 30-60 minutes, likely due to partial reroutes and weather-induced diversions. ANA's airport delay components consistently persist longer than JAL's under extreme conditions, suggesting variations in delay structure.

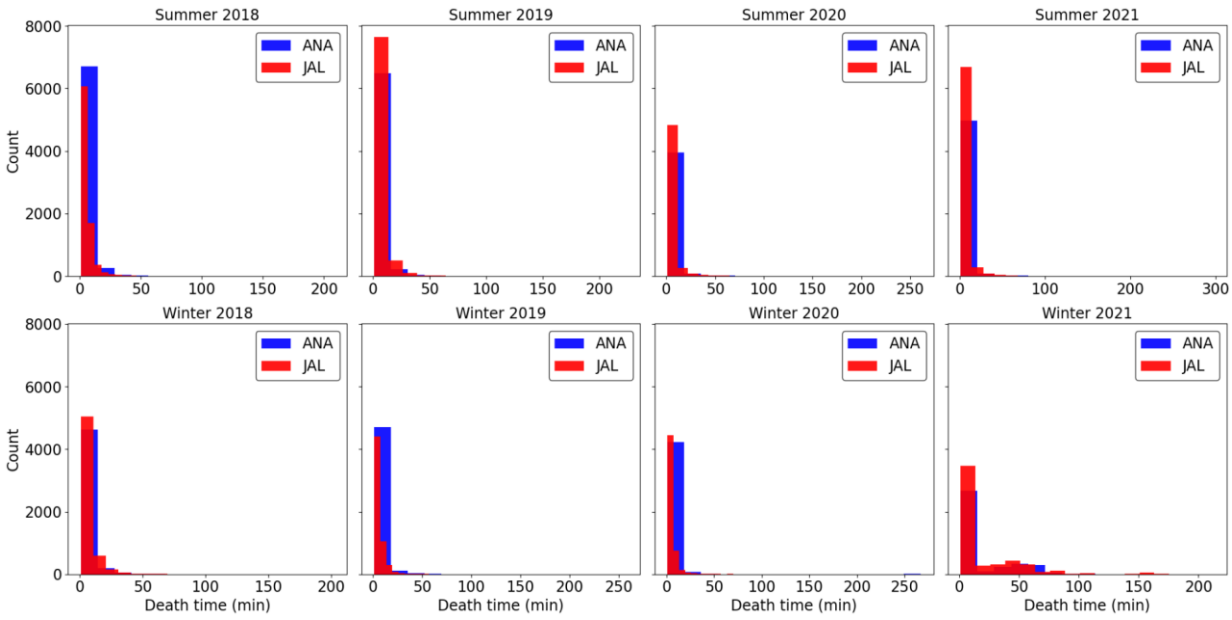


Figure 5.7: Distribution plot of death values of H_0 for ANA and JAL for each season

5.4.2 Distribution of birth value of H_1

In this section, we examine the conditions under which delay loops (H_1 features) emerge. Fig. 5.8 displays the birth-time histograms across seasons and years. Most loops form at low thresholds under 10 to 15 minutes. JAL exhibited a longer tail meaning longer delay duration before COVID-19, with loops forming as late as approximately 80 minutes in 2018. ANA, in contrast, rarely exceeded 30 to 40 minutes. It indicates that delay loop formed earlier for ANA than JAL at the pre-pandemic phase. During the pandemic year 2020, loops formed at much lower thresholds (2-5 minutes) due to reduced flight volumes. In 2021, while early loop formation remained common, ANA experienced longer birth times, with some loops forming near 100 minutes. This trend implies that ANA's delay loops are more sensitive to severe delays during pandemic. These conditions, where loops form at higher thresholds, reflect an environment where moderate delays are insufficient to trigger network-wide loops requiring instead a critical accumulation of disruption.

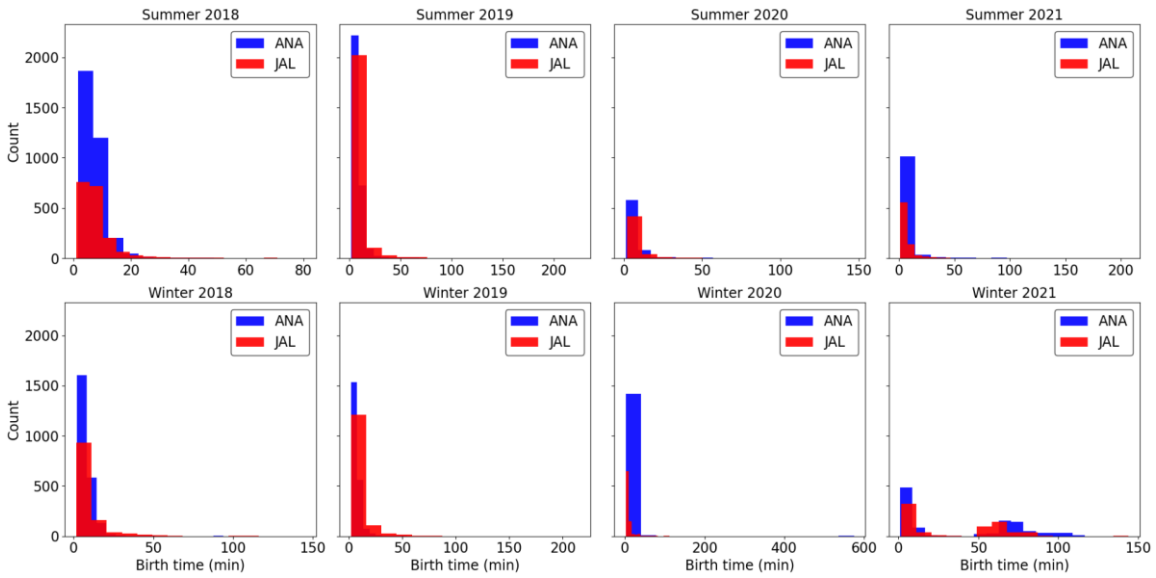


Figure 5.8: Distribution plot of birth values of H_1 for ANA and JAL for each season

5.4.3 Distribution of persistence of H_1

Following the examination of loop emergence, we turn to loop longevity (Figure 5.9). Across all years, over 80% of loops vanish within ten minutes. ANA's loops prior to COVID-19 peaked at 5-8 minutes of persistence but exhibited longer tails, occasionally lasting up to 50 minutes. JAL's loops were more concentrated around 4-6 minutes. Winter seasons extended loop persistence durations, particularly for ANA possibly due to weather and peak traffic. The 2020 collapse in operations caused loops to shrink in lifespan under 5 minutes in summer and 3-5 minutes in winter. Recovery in 2021 introduced longer persistence durations and mid-range loops (20-30 minutes), especially in ANA. These persistent winter loops point to structural vulnerabilities that require early intervention. Greater persistence indicates more difficult delay loops to break, implying operational thresholds beyond which disruptions become more severe. Throughout, ANA's loops consistently show longer tails (more long-lived loops), while JAL's collapse and recover more sharply. Summers always yield shorter-lived loops than winters, underscoring the sustaining effect of winter weather and holiday traffic. Longer persistence may imply to be an alert point to react within this persistent time, otherwise, the delay could be worse (i.e., escalated to higher-order betti number and homology group). Persistence could imply how much time we have from the moment a bad delay-loop appears. Longer persistence could indicate tough delay loops to break (i.e., to control or to reduce delay propagation).

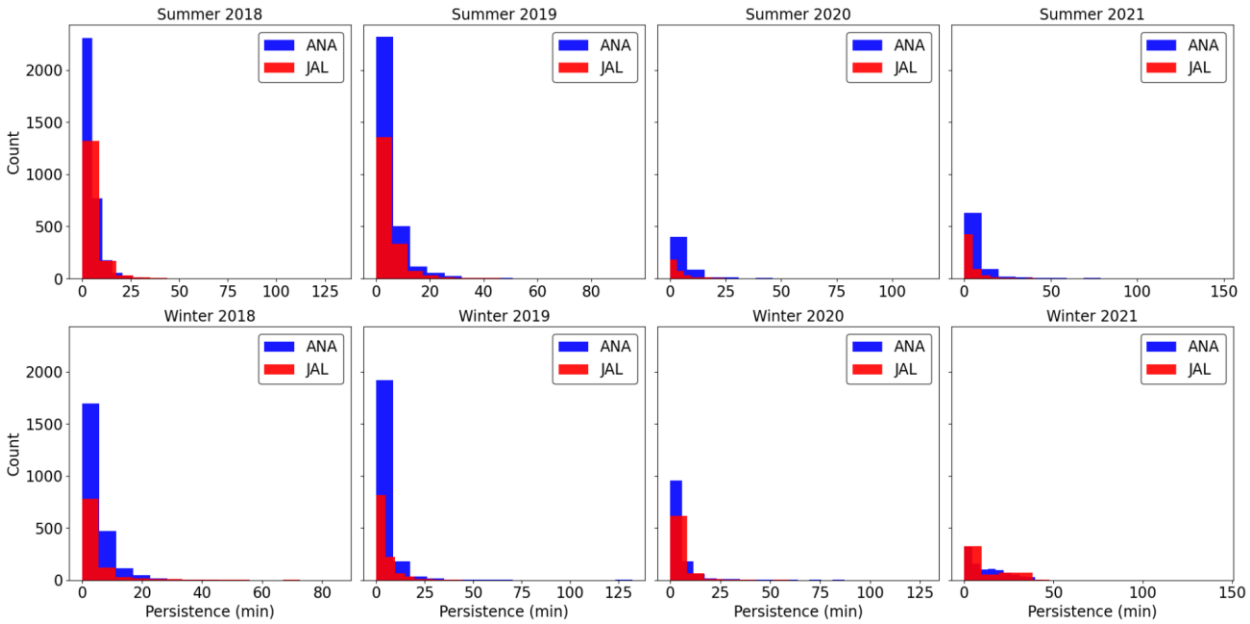


Figure 5.9: Distribution plot of persistence of H_1 for ANA and JAL for each season

5.4.4 Transition from H_0 to H_1

To quantitatively assess the temporal transition from network fragmentation to cyclic structure formation, we proposed and computed the Proportion Birth Time Overlap (PBTO) for each day in the dataset. PBTO is designed to capture the extent to which cyclic delay associations (i.e., H_1 loops in persistent homology) emerge after the underlying delay network becomes structurally unified (i.e., after most H_0 components have merged into a single connected component). PBTO was calculated as the proportion of H_1 loops whose birth time occurred after network had become nearly or fully connected. A higher PBTO value indicates that a larger fraction of cyclic delay structures (H_1 loops) emerge only after the entire network is structurally unified, suggesting a transition from localized to system-wide delay propagation. Conversely, lower PBTO values imply that delay loops tend to form while the network is still fragmented. A conceptual flow-chart of PBTO concept is given in Figure 5.10. A detailed calculation of PBTO is given in Appendix F.

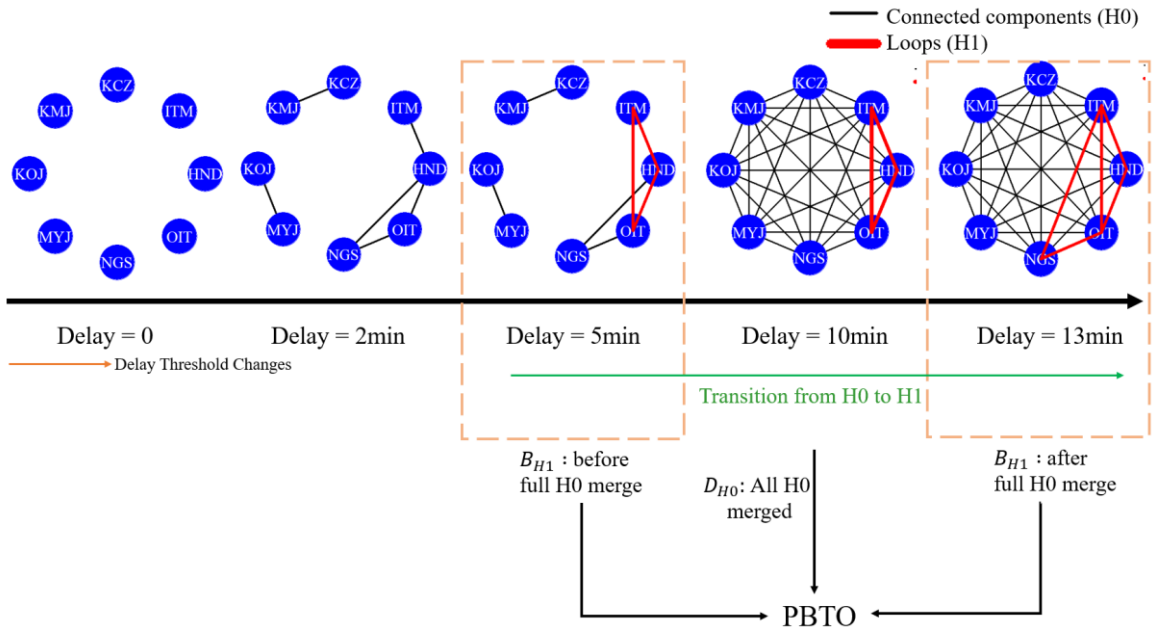
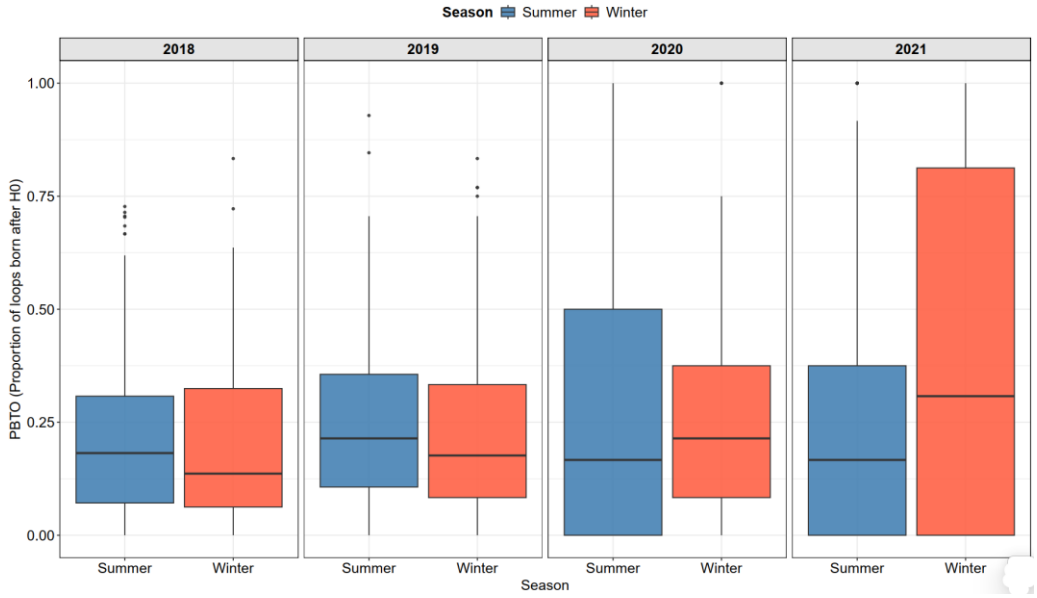


Figure 5.10: Conceptual diagram of PBTO calculation

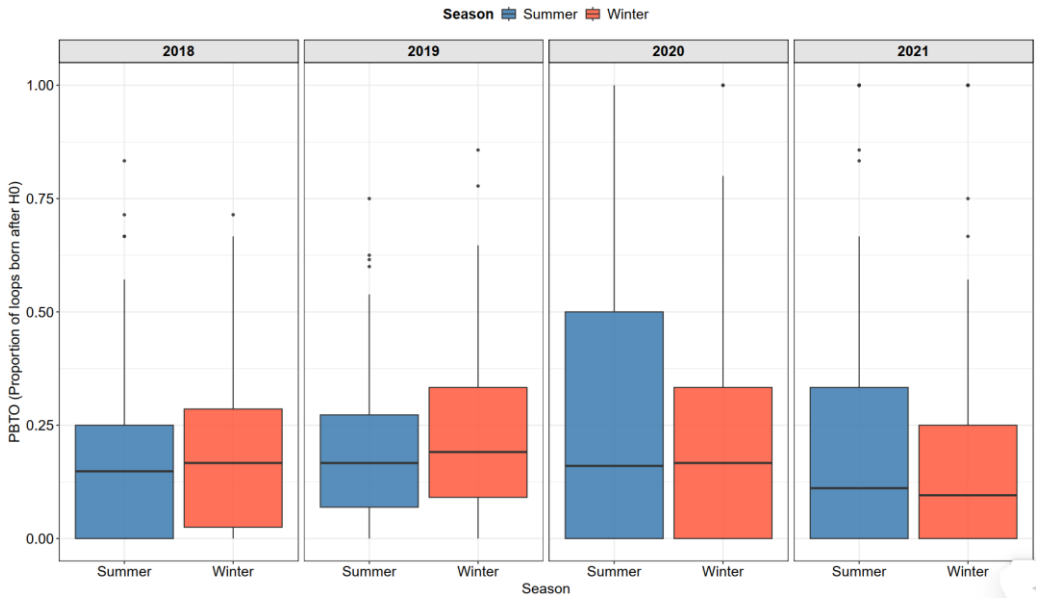
In Fig. 5.10, we observe that a loop is formed between HND-OIT-ITM at 5 min threshold which has born before the 10 min threshold which is maximum value at which all the airports are connected through delay links and continued for 8 minutes until 13 min when another delay loop NGS-OIT-HND-ITM which appears after that maximum threshold limit of 10 min.

Figure 5.11 (a and b) represents the PBTO distribution of each of the Summer and Winter seasons per year from 2018-2021. For ANA (in Figure 11-a), the PBTO patterns reveal notable trends in the evolution of delay co-occurrence loops within the Japanese domestic airport network. In the pre-pandemic years of 2018 and 2019, both Summer and Winter exhibit relatively low and stable PBTO values, with medians below 0.25 and narrow interquartile ranges. This indicates that, under normal operational conditions during pre-pandemic years, most loop-based delay associations tended to form before the network achieved full connectivity, reflecting more localized or fragmented delay loops rather than widespread systemic loop. A marked shift emerges during the COVID-19 pandemic period. In 2020, the Summer season demonstrates an increase in both the median and variability of PBTO, with several extreme outliers, suggesting that pandemic-related disruptions and the first SoE held in April – summer season, led to greater temporal volatility and an increased tendency for delay loops to emerge after the network became structurally unified. In contrast, Winter 2020 remains similar to pre-pandemic patterns, with lower and more stable PBTO values when there was no SoEs were declared in 2020. The most dramatic change occurs in Winter 2021, where the PBTO median rises sharply and the distribution widens considerably, reaching values close to 1.0 for some days. This pattern implies that, during this period, a

significant fraction of delay loops formed only after the entire network had become connected, pointing to a transition from localized to more network-wide delay propagation. From Nov-Dec, 2021 a surge of flights were observed, so as a surge of departure delayed flights have been observed. In Summer 2021, PBTO values return to the lower and more stable levels observed before the pandemic.



(a)



(b)

Figure 5.11: Plots representing Proportion of Birth Time Overlap for (a) ANA and (b) JAL

In summary, seasonal differences are relatively minor in the pre-pandemic years but become much more pronounced in 2020 and 2021, with Winter 2021 standing out as a period of exceptionally high system-wide delay feedback. Overall, these results suggest that while loop-based delay co-occurrence was primarily a localized phenomenon in normal years, the pandemic and subsequent recovery phases - particularly in Winter 2021 - were more associated with a higher prevalence of network-wide delay feedback loops, reflecting increased vulnerability of the network to disruptions.

In addition, for JAL (in Figure 3.11-b), the figure displays annual and seasonal variations in the Proportion Birth Time Overlap (PBTO), revealing consistent and interesting trends across 2018-2021. In both pre-pandemic years (2018 and 2019), PBTO values for Summer and Winter remain relatively low, with medians close to 0.2 and a majority of observations clustered toward the lower end of the scale. This indicates that most H_1 tended to emerge before the airport delay network became fully unified, reflecting a predominantly localized pattern of delay propagation under normal conditions. Seasonal differences in these years are minor, with Winter showing slightly higher medians and outlier values. A more pronounced divergence appears from 2020 onward. In 2020, Summer exhibits a marked increase in the spread of PBTO values, with a higher median and a broader interquartile range, including several days where PBTO approaches 1. This suggests a temporary shift during the pandemic's disruptive phase, where delay loops more frequently formed after the network became structurally unified, pointing toward increased system-wide propagation of delays. In contrast, Winter 2020 remains more consistent with previous years, though its median is slightly elevated. In 2021, both Summer and Winter return to a more stable PBTO distribution. However, the medians and IQRs for both seasons decrease slightly compared to 2020, and the extreme values become less common, implying a partial return to pre-pandemic dynamics. Notably, the distributions of Summer and Winter in 2021 are almost identical, with Winter showing a slight reduction in the upper range compared to Summer. This suggests that the exceptional system-wide feedback effects observed during the initial pandemic year were temporary and that, by 2021, the network's delay propagation characteristics had largely normalized across both seasons. Overall, these results show that loop-based delay co-occurrence comprising of JAL's route was primarily a localized phenomenon in normal years, with limited seasonal influence. Pandemic-related disruptions temporarily increased the prevalence of systemic delay loops - particularly in the summer of 2020 - before returning to more stable, pre-pandemic patterns in 2021.

5.4.5 Comparison of persistence categories for H_1

To further dissect loop characteristics, we classify loop persistence into Low (0-10 minutes), Medium (11-20 minutes), and High (>20 minutes) categories as summarized in Table 5.1. Across both carriers, summer operations are characterized by overwhelmingly low-persistence loops, with more than 89% of

airport sets in the *Low* category across all loop sizes. This suggests that summer delays are largely short-lived, likely due to favorable weather conditions, more predictable operations, and possibly increased operational capacity during peak travel periods. However, a shift occurs in winter, particularly for larger airport sets. ANA sees a moderate increase in high-persistence loops during winter, especially in 4-airport and 3-airport configurations, where high-persistence percentages climb to 12.29% and 6.91% respectively both notably higher than their summer counterparts. For JAL, the contrast is even sharper.

Table 5.1: Percentage (%) of airport sets at each persistent category

Carrier	Season	Airport-sets	High	Medium	Low
ANA	Summer	3-airport	2.32	6.62	91.07
		4-airport	2.58	7.59	89.83
		5-airport	2.91	7.56	89.53
		6+ airport	1.35	17.57	81.08
	Winter	3-airport	6.91	11.38	81.71
		4-airport	12.29	9.97	77.74
		5-airport	9.79	7.69	82.52
		6+ airport	4.41	7.35	88.24
JAL	Summer	3-airport	3.42	7.50	89.08
		4-airport	3.35	7.14	89.51
		5-airport	2.29	17.56	80.15
		6+ airport	4.72	19.81	75.47
	Winter	3-airport	7.98	4.24	87.78
		4-airport	3.62	6.58	89.80
		5-airport	24.43	7.63	67.94
		6+ airport	17.86	20.24	61.90

Its winter operations exhibit a dramatic rise in high-persistence delay loops for large configurations. In particular, 5-airport sets reach an exceptionally high 24.43% in the *High* category during winter, the highest value in the entire table, while 6+ airport sets also show elevated persistence at 17.86%. These figures far exceed ANA's values for the same configurations and season, suggesting that JAL's network in winter is more susceptible to entrenched feedback loops and systemic delay accumulation. The results indicate that persistent delay feedback loops are more likely to emerge under winter conditions and especially within larger airport clusters. For JAL, the problem appears more acute, pointing to potential weaknesses in its wintertime operational resilience, such as fewer buffers, longer turnaround times, or

greater exposure to weather-affected regional routes. By contrast, ANA maintains a more stable persistence distribution across seasons and loop sizes, potentially reflecting more robust delay mitigation strategies or centralized scheduling coordination.

Expanding on these trends, Figure 5.12 shows the density of various airport-set (3-airport, 4-airport, 5-airport, and 6+ airport) counts for ANA and JAL across summer and winter season for each year from 2018 to 2021.

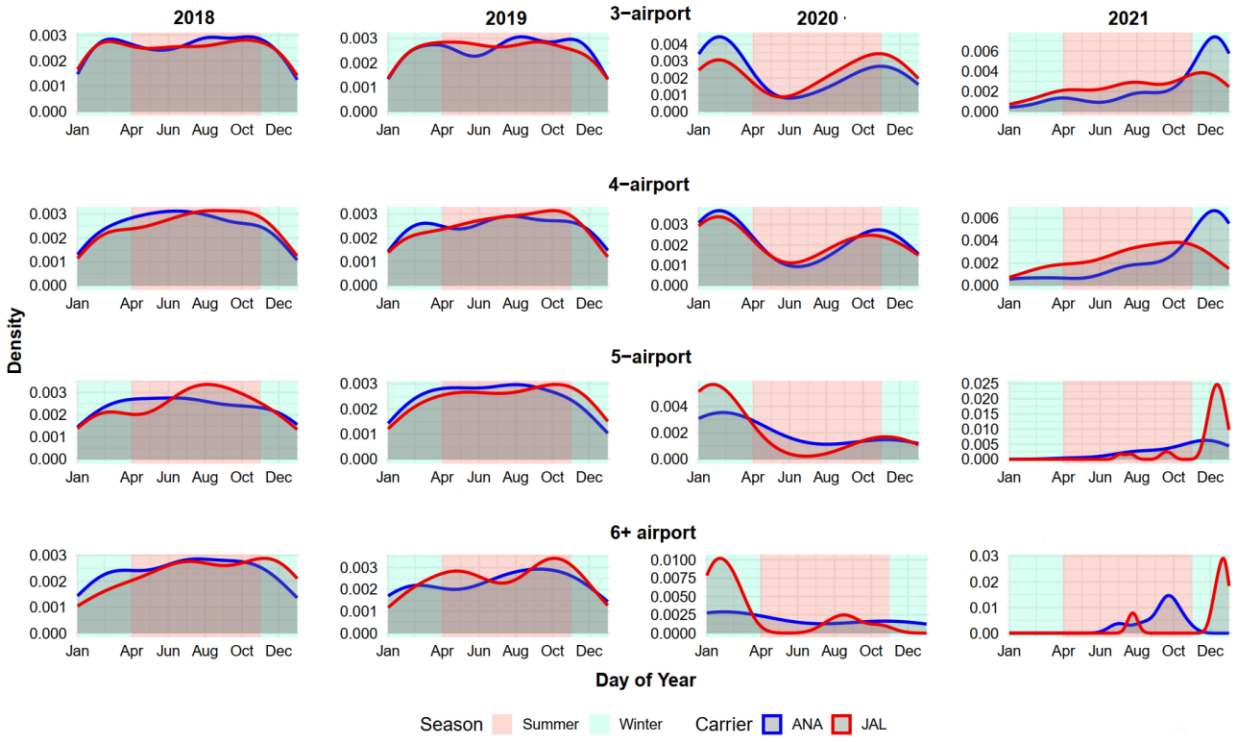


Figure 5.12: Airport-set density plot at each seasons from 2018 to 2021

In the pre-COVID years 2018-19, both carriers show relatively smooth, year-round loop activity. Small loops (3- and 4-airport sets) peak modestly in summer, while larger loops (5- and 6+-airport sets) have a slight uptick in winter but ANA and JAL largely overlap - indicating stable, seasonally muted delay-feedback loops in normal operations. In 2020, summer densities collapse nearly to zero for all loop sizes reflecting the massive scale-back of flights while only winter sees any persistent H1 loops at all, and then only for the smallest 3- and 4-airport loops. By 2021, loop activity rebounds but has fundamentally shifted: winter now dominates across every loop size, and especially for the larger 5- and 6+-airport loops where both carriers and JAL in particular show very sharp end-of-year peaks.

In summary, during pre-COVID, delay loops existed all year, with slight summer bias for small loops and minor winter bias for big loops with ANA and JAL being broadly similar. During the pandemic year

(2020) almost no summer loops; only a muted winter rebound in the smallest loops. During the recovery (2021), strong, winter-only persistence in all loop sizes and especially in large loops (5- and 6+ airports), with JAL's winter peaks most pronounced signaling that post-COVID delay-propagation feedback has concentrated into a few highly interdependent airport clusters during the colder months.

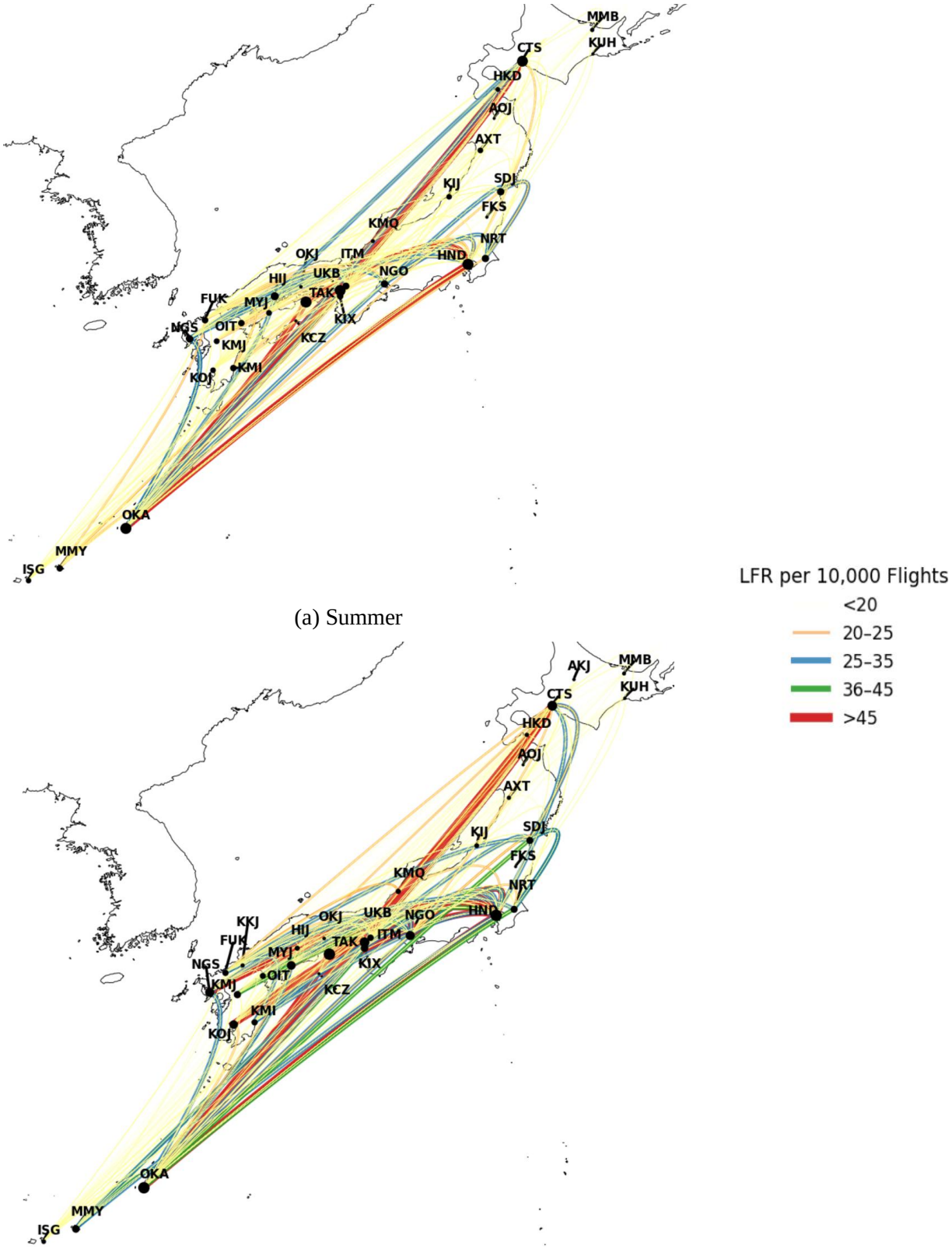
The sharp rise in airport-set loop densities during Winter 2021, especially notable for larger loops (5- and 6+-airport sets) and more observed for JAL compared to ANA can be observed due to in large part to Japan's "Go-To-Travel" campaign which resumed in late 2021 to stimulate domestic travel and economic recovery from the COVID-19 pandemic. the differences between This policy significantly increased passenger volumes in December 2021, particularly around major transportation hubs and popular tourist destinations. Consequently, airports experienced heightened operational pressures which amplified the likelihood of initial delays quickly escalating into network-wide disruptions and cyclic feedback loops. Additionally, the "Go-To-Travel" policy's timing during December aligned directly with the peak winter season. This overlap further increased the vulnerability of JAL's and ANA's networks to systemic delays. Thus, the interplay of sudden policy-induced demand spikes, carrier-specific hub structures, and seasonal weather factors effectively explains ANA and JAL observed in Figure 5.8 for Winter 2021

To quantitatively evaluate how frequently cyclic delay patterns (i.e., persistent homology H_1 loops) occur relative to network traffic, we introduced a metric called Loop Formation Rate (LFR). LFR measures how often delay feedback loops form in the context of actual flight operations, enabling direct comparison across airports, routes, seasons, and operational regimes. Similarly, to assess the prevalence of loop formation relative to total flight volume (not just delayed flights), we used the same loop detection but normalized by the total number of flights (delayed and on-time) between all airport pairs in the loop. The LFR per 10,000 flights was then computed as: $\left(\frac{Loop\ Count}{Total\ Flights\ in\ the\ Loop\ Route}\right) \times 10{,}000$. For example: ('HND', 'OKA', 'TAK') loop was detected 152 times from 2018-2021 in the summer season. Total count flights in this route (HND-OKA, HND-TAK, OKA-TAK - both way) was counted 27,534. So, the loop formation rate per 10,000 flights in this loop was $\left(\frac{152}{27{,}534}\right) \times 10{,}000 = 55$. This formulation is particularly useful for evaluating the structural vulnerability of the network to loop-based delay co-occurrence under different traffic levels. Higher LFR values indicate that particular airport-sets (i.e,, loops) are more prone to repeatedly forming delay loop patterns relative to the volume of traffic, highlighting potential hotspots in the network.

Figure 5.13 represents the geographical distribution of LFR in both Summer and Winter season comprising of the airport routes for ANA. Figure 5.13 (a) shows the Summer patterns. It reveals that the higher LFR values appear along the heavily travelled south-western corridor linking Haneda (HND) or

Osaka (ITM/KIX) and Okinawa (OKA). These loops involve airports in Kanto, Kansai and Kyushu regions, that very year frequently affected by summer typhoons and convective storms. Even after normalising by the large number of flights (these routes are some of ANA's busiest), loops form repeatedly, indicating that weather-related delays and tightly connected flight rotations make this part of the network structurally vulnerable to delay loop formation. In contrast, loops in less busy parts of the network (for example, routes connecting mid-sized regional airports in the Chubu or Tohoku regions) have LFR values below 25. These routes either experience fewer delays or have more schedule slack, so delays dissipate quickly instead of forming persistent loops.

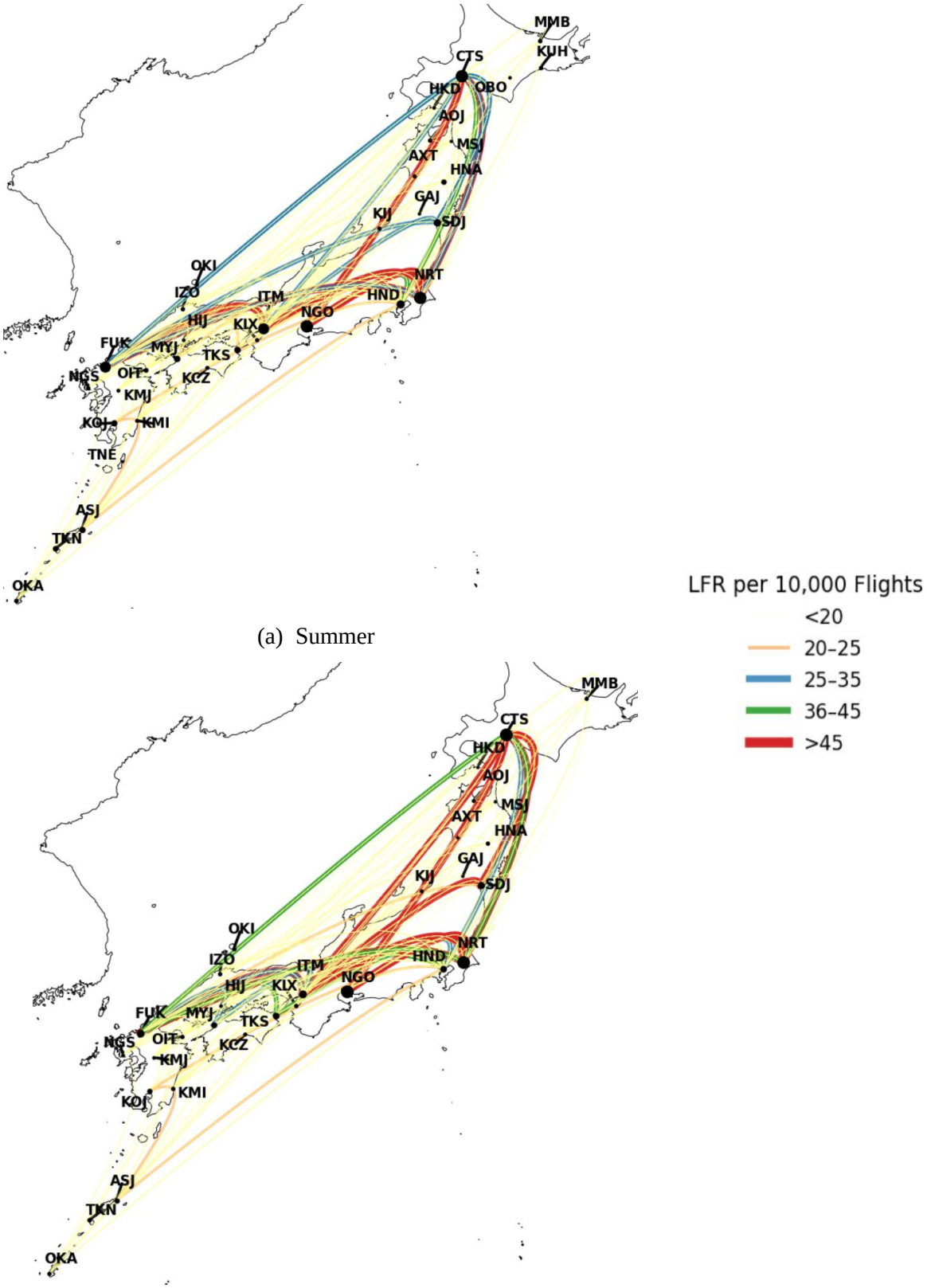
In contrast, the Winter patterns (Figure 13-b) shows the pattern shifts northward. High LFR loops still occur along the HND–KIX–FUK–OKA corridor, but similar or higher LFR values also appear on routes involving Hokkaido and the Sea-of-Japan coast, such as Haneda–Sapporo (CTS). Heavy snowfall and strong winds in the north may be caused by de-icing delays and runway closures, leading to delay loops. Compared with summer, many of the loops involving remote islands in the south (e.g., Ishigaki (ISG) or Miyako (MMY)) show slightly lower LFRs, indicating that typhoon-related disruptions are seasonal.



(a) Summer

(c) Winter

Figure 5.13: Loop Formation Rate per 10,000 flights for ANA for each routes across domestic Japan



(a) Summer

(b) Winter

Figure 5.14: Loop Formation Rate per 10,000 Flights for JAL for each routes across domestic Japan

Figure 5.14 represents the LFR for JAL connecting flights. In the summer network (Figure 5.14-a), most loop edges are light yellow or orange, meaning that fewer than 25 loops occur per 10 000 flights. Many loops connecting southwestern airports (Kyushu and Shikoku) to the major hubs (Haneda (HND), fall into this low-LFR range, suggesting that cyclic delay propagation is uncommon on these routes. A few routes show higher LFR values: a diagonal blue edge (~25–35 per 10,000 flights) connects the southwest with Sapporo (CTS), and several green lines (36-45 per 10,000 flights) cluster around Haneda/Narita and the Tōhoku airports. The red lines (>45 per 10,000 flights) occur between Kanto-Kansai-Kyushu regions. These hotspots indicate that although most summer operations are resilient to loops due to smaller loops in some routes, certain regional routes in northern Japan and some long-distance connections are structurally prone to loop-based delay patterns. In contrast the winter LFR (Figure 5.14-b) shows a very different picture. The number of high-LFR loops increases substantially, and these loops are concentrated in the northern and central part of the network. Many loop edges connecting Haneda/Narita to Sendai (SDJ), and Sapporo (CTS) are red, indicating more than 45 loops per 10,000 flights making them major delay-occurrence hotspots. Also, the loop formation between Kansai region and Hokkaido region substantially increased with red edges indicating again 45 loops per 10,000 flights. A green edge (~36–45 per 10 000 flights) connects the southwest to Sapporo (CTS), suggesting that long-haul flights linking southern Japan with snow-prone Hokkaido are also vulnerable in winter. In contrast, routes within Kyushu and between small southwestern islands remain mostly yellow (<20), implying relatively low loop propensity despite winter conditions. The overall pattern indicates that winter weather and de-icing delays at Tohoku and Hokkaido airports greatly amplify loop formation, whereas the southern network remains robust.

Overall, for ANA, the LFR maps reveal that cyclic delay patterns are not uniformly distributed. A small set of airport groups - largely those connecting Haneda, Osaka, Fukuoka and Okinawa, and (in winter) the Hokkaidō region experience LFRs above 35-45 per 10,000 flights, meaning that loops form once every 200-300 flights. These routes are critical because they carry a large proportion of ANA’s traffic, yet their high LFRs indicate that their schedules offer little resilience to disruptions. Lower-traffic or less-central routes show much lower LFRs, suggesting they either avoid severe weather or have operational slack that prevents loop formation. The seasonal shift in loop hotspots from typhoon-prone southern routes in summer to snow-prone northern routes in winter, underscores the role of weather and travel demand. Because ANA’s network covers the entire country, mitigating delay propagation will likely require tailored strategies: additional buffer time on flights linking major hubs, improved recovery procedures for weather-disrupted regions, and perhaps diversifying aircraft rotations to reduce the coupling that allows loops to form so frequently.

And for JAL, the LFR maps reveal that JAL's network vulnerability to cyclic delays is highly seasonal. In summer, loop formation is relatively rare and localized; only a few northern routes show high loop frequencies. In winter, snow-related disruptions and reduced flight frequencies lead to a dense cluster of high-LFR loops around Kanto, Kansai and Hokkaido region. These loops, normalized by the number of flights, highlight structural weaknesses where delays repeatedly circulate through the same set of airports. Identifying these hotspots allows airlines and regulators to target mitigation strategies such as adding schedule slack, prioritizing de-icing resources, or reconfiguring crew/aircraft rotations on affected routes to prevent small delays from forming persistent feedback loops.

5.4.6 Statistical difference between summer and winter season

Having established a broad understanding of loop dynamics, we now turn to testing whether seasonal differences are statistically significant. We compared the persistent diagrams between ANA and JAL for each of the seasons using persistent landscape two-sample test. It converts the diagram to persistent landscape. It is a piecewise linear function summarizing each diagram in a Banach space. We the compute the landscapes for each diagram. Then we need to treat each landscape as functional data. Then we perform a two-sample test between the set of ANA landscapes and the set of JAL landscapes (Bubenik, 2015; Bubenik and Dlotko, 2017). A two-sample *t-test* on the mean of the first persistence-landscape layer gives $t = 2.575$ and $p - value = 0.010$. It indicates that we reject the null and conclude that ANA's and JAL's delay-network topologies differ significantly in terms of their first-layer persistence landscapes. In practical terms, the "shape" of their birth-and-death patterns for delay-propagation features (condensed into the first landscape) is not just random noise. There is a real, measurable structural difference between the two carriers' networks. Table 5.2 represents the significance test between winter and summer season for ANA and JAL. These results tell us that ANA's higher-order delay-feedback structures are much more sensitive to seasonal and disruption effects than JAL's. ANA shows no meaningful summer-winter shift in the pre-COVID years (2018-19), but strong, highly significant seasonality during the pandemic and recovery. JAL only exhibits a significant summer vs winter difference back in 2018, and no statistically detectable seasonality in the other three years. Before COVID, neither airline's topological delay-patterns changed markedly by season. JAL had a slight seasonality in 2018 but otherwise stayed uniform. ANA's network became strongly seasonal once the pandemic hit. Its' delay-loop dropped in Summer 2020 (likely due to collapsed schedules) and then flipped to be unusually higher in Summer 2021 (when recovery surges may have stressed its network).

Table 5.2: Comparing summer and winter seasonal persistent diagram between ANA and JAL

Year	Airline	t-stat	p-value	Interpretation
2018	ANA	-0.41	0.679	No seasonal difference ($p < 0.05$)
	JAL	-2.83	0.005	Significant
2019	ANA	0.66	0.509	No seasonal difference
	JAL	1.61	0.108	No seasonal difference
2020	ANA	-4.67	0.000	Significant
	JAL	-0.89	0.377	No seasonal difference
2021	ANA	6.66	0.000	Significant
	JAL	0.87	0.383	No seasonal difference

NB: The grey shade represents statistical significance at 1% significance level

5.4.7 Airport-set counts and its birth and persistence distribution

Finally, we offer a network-level breakdown of specific airport-set loop dynamics. The multi-panel in Figure 5.15 summarizes all 1-dimensional topological loops (H_1) identified from ANA's daily departure-delay networks between 2018 and 2021. The top bar chart shows the frequency of each airport-set's appearance as a loop, categorized by persistence duration: Low (≤10 minutes), Moderate (11–20 minutes), and High (>20 minutes). Beneath this, the dot matrix visualizes the specific airports that build each loop. The middle panel presents boxplots of birth times indicating the delay threshold at which each loop forms, and the bottom panel displays how long these loops persist. The most frequently occurring loop is the 3-airport set ['HND', 'OKA', 'TAK'], appearing over 300 times and almost exclusively within the low-persistence category. This suggests a routine, short-lived feedback pattern among Tokyo Haneda, Okinawa, and Takamatsu airports likely resulting from recurring interdependencies in their flight schedules. A smaller subset of loops exhibits moderate persistence, such as the 3-airport set ['HND', 'NGO', 'MYJ']. These loops generally emerge at higher delay thresholds marking them as more structurally impactful and operationally meaningful. They persist longer and are less likely to collapse quickly, which suggests they may act as early-warning signs for more serious network disruptions. Only a few airport-sets enter the high-persistence category. One notable example is ['NGO', 'CTS', 'SDJ', 'KIJ'], which exhibits outlier behavior with persistence occasionally exceeding 20 minutes. These long-lasting loops, while rare, signal severe network entanglements and are critical to identify in real-time for effective delay mitigation.

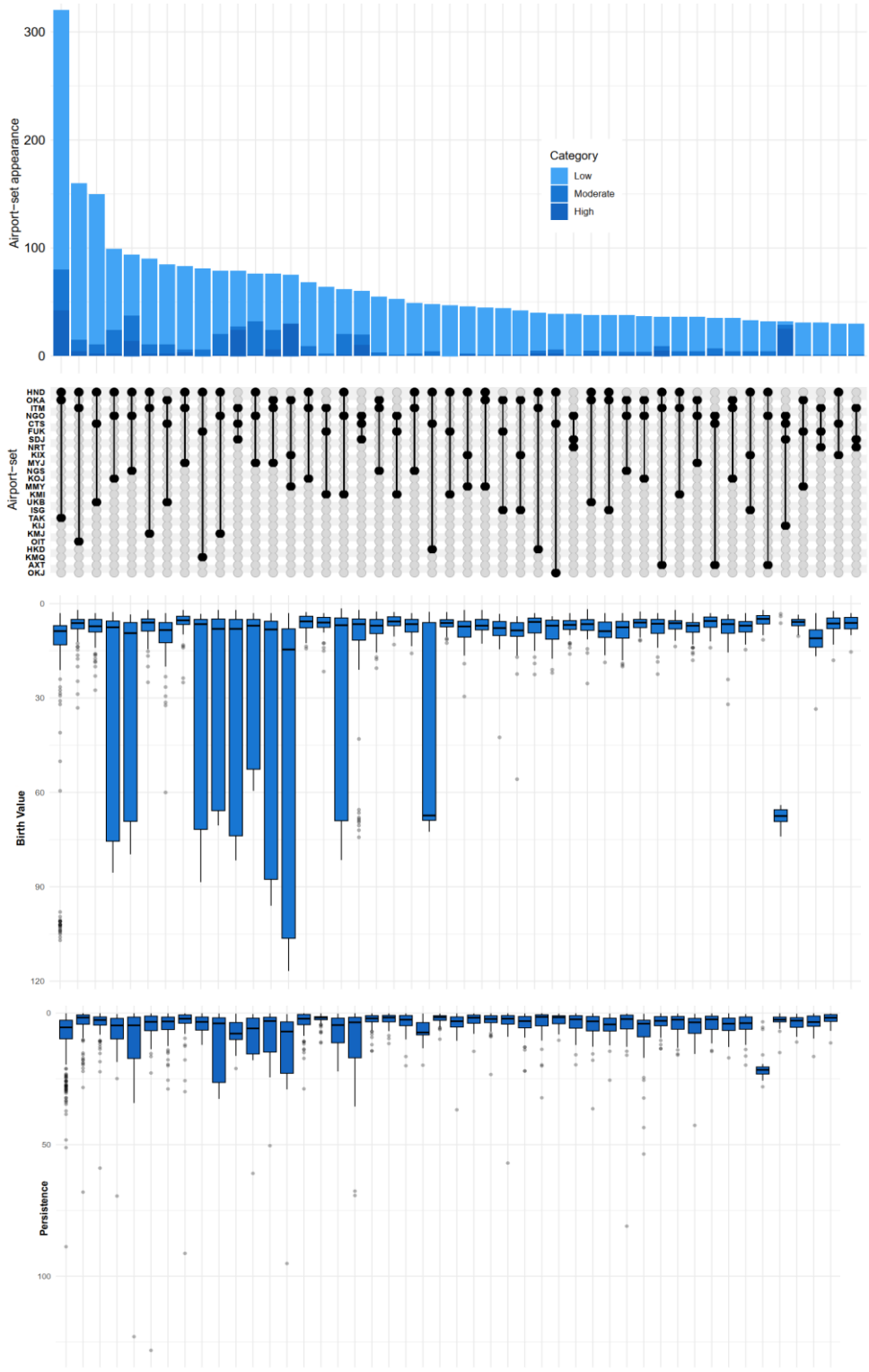


Figure 5.15: ANA airport-set appearances from 2018 to 2021 based its proportion in low, medium and high persistence. Last two panels show the boxplot distribution of the birth time (in minutes) and persistence (in minutes) of each of these airport sets

Across all sets, low-persistence loops tend to form at early delay thresholds (median birth times < 20 minutes), implying that minor disruptions already induce feedback patterns. In contrast, moderate- and high-persistence loops are born under more severe delay conditions (birth times 20-40 minutes), underscoring their potential use as indicators of emerging systemic issues. Persistence boxplots show that frequent loops typically collapse quickly, whereas the more robust loops exhibit greater variability in how long they remain open. Operationally, the most common loops such as HND–FUK–KMI are likely short, requiring minimal intervention. However, moderate-persistence loops like HND–NGO–MYJ demand closer attention when delay thresholds begin to rise, as they can indicate worse delay loops. High-persistence configurations, although infrequent, represent critical points of vulnerability; their identification and monitoring can enable targeted interventions to prevent cascading delays across multiple airports.

In case of JAL, Figure 5.16 summarizes the structure and dynamics of 1-dimensional topological loops (H_1) observed across JAL's daily departure-delay networks from 2018 to 2021. A dominant trend is the recurrence of small 3-airport loops, such as [HND, CTS, NGO] and [HND, FUK, TKS], each appearing over 200 times. These loops tend to emerge early (at delay thresholds below 10 minutes) and collapse quickly, characterizing them as routine operational feedback loops among these two airport loops with high-frequency connections and tightly scheduled operations. While common, they present minimal risk unless compounded by further disruptions.

In contrast, several mid-frequency loops, including [HND, CTS, KIX] demonstrates moderate persistence and appear only once delays become more substantial (birth times typically between 20–30 minutes). These structures serve as early indicators of growing network instability and are operationally meaningful, as their persistence suggests the potential for spreading delays.

Only a few loops fall into the high-persistence category, with sets like [ITM, CTS, SDJ, HNA, KIJ] standing out for occasionally lasting more than 40 minutes. These rare, slow-closing loops signal serious congestion and require immediate intervention, particularly as they represent multi-airport feedback chains that can escalate if left unaddressed. The birth-time distribution across all loops shows a natural progression: frequent, short-lived loops appear early in the delay timeline, while more persistent and structurally critical loops emerge only under higher-delay conditions. The persistence plots reinforce these frequent loops exhibit minimal variability and short durations, while less common, higher-persistence sets show broader distributions and elevated medians.

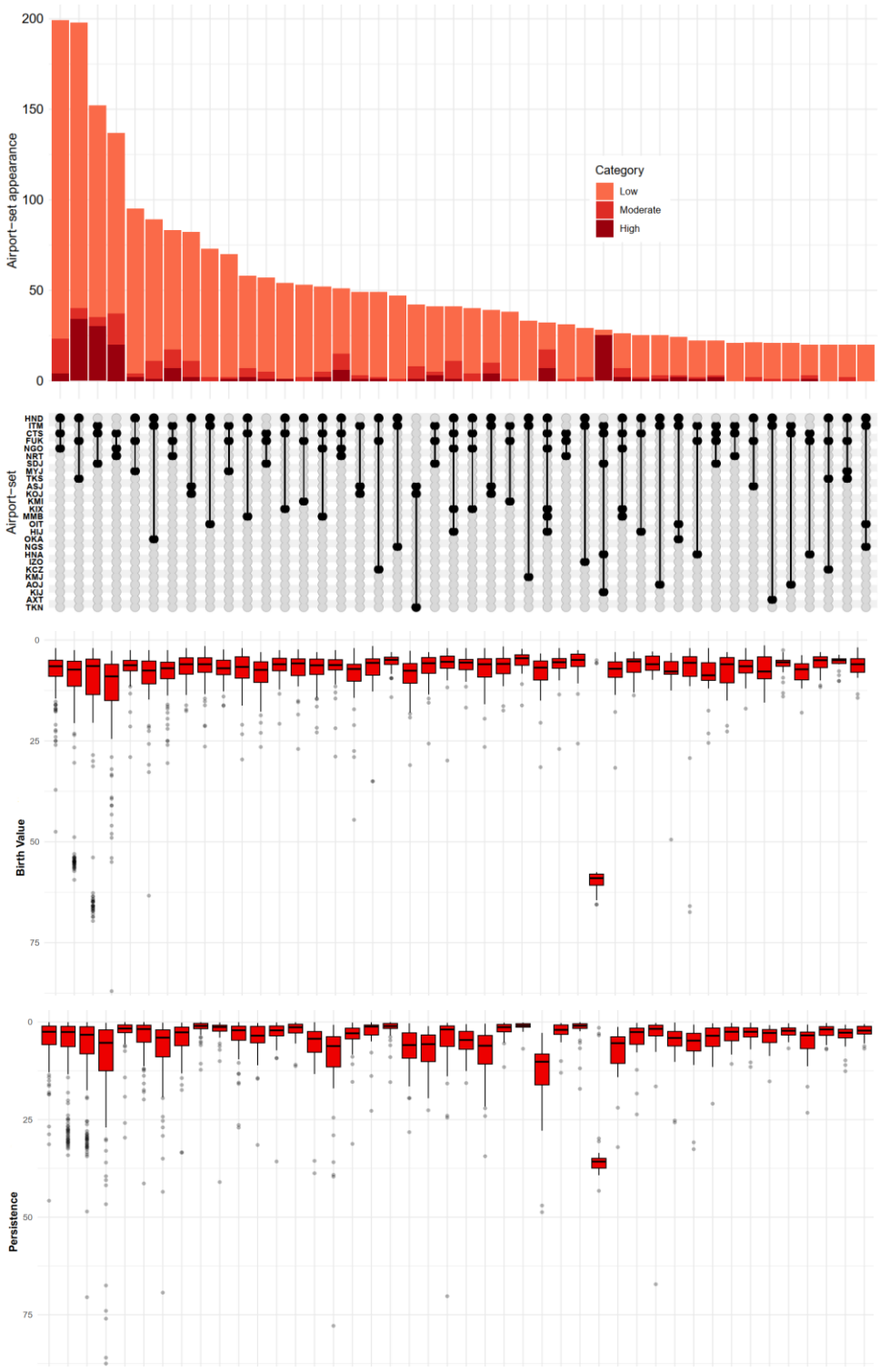


Figure 5.16: JAL airport-set appearances from 2018 to 2021 based its proportion in low, medium and high persistence. Last two panels show the boxplot distribution of the birth time (in minutes) and persistence (in minutes) of each of these airport sets

For JAL's operations, this layered analysis provides a foundation for scalable response strategies. Routine monitoring of frequently recurring short loops can help prevent their trivial disruptions from escalating. Moderate-persistence loops offer timely warning signs when delays begin to stack, allowing for proactive scheduling adjustments. High-persistence loops, although rare, warrant immediate cross-airport mitigation once detected, given their potential to trigger systemic delay disruptions. Overall, the synthesis of frequency, formation threshold, and persistence characteristics offers a strategic toolset for prioritizing interventions and designing buffer capacity across the network.

5.5 Discussion

The analysis began with the death-time distribution of H_0 features (in section 5.4.1), which reflects how quickly isolated delay clusters dissolve. The majority of clusters reconnect within 5 to 15 minutes across both airlines, with longer persistence in winter and under pandemic-induced disruptions. This aligns with findings by Fleurquin et al. (2013), who showed that delay clusters often exhibit heavy-tailed lifespans, with occasional long-lasting events indicating structural weaknesses in the network. Building on this, the birth-time of H_1 loops (in section 5.4.2) captures the delay threshold at which delay loops first emerge. The observed tendency of loops to form at low thresholds (<15 minutes) suggests a fragile network that reacts early to small disruptions mirroring observations by Gopalakrishnan & Balakrishnan (2017), who emphasized how minor delays can quickly translates into tightly coupled networks. Later, PBTO is introduced to capture H_0 to H_1 transition. It captures when H_1 loops tend to appear - only after the delay network has largely unified (high PBTO = system-wide timing) versus while it is still fragmented (low PBTO = localized timing). While ANA's PBTO is low and stable during pre-COVID time, it rises during-COVID and peaks during the recovery in late 2021 whereas JAL shows a Summer rise but largely returns to lower, pre-pandemic like PBTO in 2021. Monitoring PBTO over time can serve as an early warning for systemic vulnerability. When PBTO rises, policy-makers and airlines should consider adding buffer times and decoupling certain rotations to prevent network-wide loops. PH provides a method to identify such transitions objectively.

Persistence distributions (in section 5.4.4) revealed that most loops dissolve within 10 minutes. However, ANA exhibits more long-lived loops, especially in winter suggesting a higher structural inertia in its network. This resilience in delay propagation echoes Prandini et al. (2012), who proposed complexity metrics for flagging high-risk areas in air traffic; persistent loops here serve a similar role, flagging entrenched patterns. The seasonal lengthening of loop persistence, especially in ANA during winter, may also reflect weather-induced congestion and limited recovery capacity a condition noted in operational delay analyses such as by Fleurquin et al. (2013). Importantly, longer persistence may also imply not just severity but predictability. Dmochowski & Skorupski (2017) introduced *air traffic smoothness* as a

structural metric; in this spirit, persistence serves as a delay-smoothness indicator low persistence implies reactive but recoverable loops; high persistence implies self-sustaining, and thus harder-to-manage disruptions.

By categorizing loops into *Low*, *Medium*, and *High* persistence (in section 5.4.5), we identified sharp seasonal asymmetries. Winter conditions dramatically increase the frequency of high-persistence loops, especially in larger airport sets (5- and 6+ airports). JAL is particularly vulnerable here, with 24.43% of 5-airport loops in 2021 winter falling into the *High* category. This supports earlier findings from Fleurquin et al. (2013) and Zhang et al. (2024) that delay propagation is disproportionately concentrated in select subnetworks, which can form large, long-lived clusters under stress. Our results add topological granularity to these observations by showing how those clusters manifest as persistent feedback loops. Notably, the emergence of complex loops in JAL under winter conditions echoes Chen et al. (2020)'s delay community detection, where regional weather-affected clusters intensified delay interdependence. Later, the LFR captures how often those loops occur per unit traffic (loops per 10,000 flights), i.e., their frequency after normalizing for volume. For ANA, high LFR concentrates along HND-(KIX/ITM)-FUK-OKA in Summer (typhoon/convective weather and tightly coupled rotations) and shifts northward in Winter toward Hokkaidō/Sea-of-Japan routes (snow, de-icing). JAL's Summer LFR is generally lower and localized; Winter shows a denser band of higher LFR around Kanto–Kansai–Hokkaidō, highlighting snow-exposed corridors. Persistence distributions, categories and LFR can act as indicators of delay inertia. Persistent loops can be used as flags for high-risk patterns. Airports associated with high persistence should receive priority in delay-mitigation strategies such as rerouting, dynamic scheduling or resource reinforcement. Regulators could require airlines to maintain larger buffers on routes that repeatedly host high-persistence loops.

Airport-set density plots show how loop emergence changes over time, especially across pandemic phases. While pre-COVID networks had balanced loop activity, the pandemic-induced traffic reductions in 2020 eliminated almost all summer loops. In 2021, however, delay loops rebounded but shifted to winter-dominant persistence, particularly in larger loops. This structural transformation supports Du et al. (2018)'s insight that only subsets of airports become delay hubs depending on external conditions. Persistent homology contributes a new dimension by identifying not just hubs, but cyclical motifs that recur during system stress. Unlike causality networks that rely on statistical inference between airport pairs, our method finds persistent, threshold-robust groups of airports forming feedback loops (Wang et al., 2022).

The two-sample t-tests on persistence landscapes validate these structural differences: ANA shows strong seasonality post-COVID, whereas JAL remains statistically consistent except in 2018. This divergence

implies that ANA's delay network became structurally less robust. Our results show that persistent homology detects these structural weaknesses reflected not just in average delays, but in the shape of delay feedback patterns across thresholds. Moreover, this confirms Cook et al. (2015)'s recommendation to analyze ATM as a complex system with emergent behaviors. Persistent loops are precisely such emergent features: they are not defined a priori but arise from the global shape of the delay network. Their seasonal amplification, particularly under abnormal schedules, highlights vulnerabilities invisible to linear or node-level metrics.

Airport-set loop analysis (in section 5.4.6) revealed that some airport combinations such as HND–OKA–TAK forms frequent, low-persistence loops, whereas others (e.g., NGO–CTS–SDJ–KIJ) exhibit high persistence and operational criticality. These airport sets align with internal structural risks, akin to the findings of Chen et al. (2020) and Güvercin et al. (2021), who linked route clusters and centrality to delay propagation strength. These loops, particularly those with high birth thresholds and long persistence, serve as early warnings for severe disruptions analogous to holding patterns in spatial trajectories (Li et al., 2019) or structural complexity zones (Prandini et al., 2012). Identifying and preemptively addressing such loops (e.g., by allocating additional buffers or rerouting flights within these clusters) could prevent delay escalation, a mitigation strategy supported by Dalmau et al. (2022) and Menon & Park (2016).

Persistent loops offer concrete operational insights. From an ATFM perspective, loops indicate where "cherry-picking" delay interventions (Dalmau et al., 2022) would be most impactful e.g., regulating one flight in a loop could unwind the entire loop. From an airline control center viewpoint, knowing that a particular triplet or quartet of airports repeatedly forms a loop can guide dynamic re-planning (Kuhn, 2016) and trajectory buffering.

Persistent homology also enhances predictive modeling: as Güvercin et al. (2021) and Zhang et al. (2024) show, accounting for group-level or layer-level structure boosts delay forecasting. Persistent loops offer a data-driven way to define those groups without prior assumptions. As such, they could serve as real-time features in delay prediction systems, or as overlays in collaborative digital twins (Chen et al., 2024), visually flagging regions of concern.

Regulators and airlines should tailor interventions based on airline-specific vulnerabilities. Persistent homology can guide resource allocation by highlighting which airport combinations require extra buffers or rerouting. Additionally, network-level analysis supports calls for complex network control: research shows that controlling delay networks using network-science tools can reduce delays at relatively low cost (Niu et al., 2021). Integrating PH-derived features into airline control centres and predictive models

(e.g., graph-based machine learning) could improve delay forecasting and support collaborative decision making.

5.6 Chapter Conclusion

This study leveraged Topological Data Analysis (TDA), particularly persistent homology, to uncover structural feedback patterns in Japan's domestic departure-delay networks. By constructing daily Vietoris–Rips complexes where nodes represent airports, edges signify delayed-flight connections, and delay durations define the filtration threshold, we extracted two core dynamic phenomena: (1) the fragmentation of the network into isolated delay clusters (captured via H_0 death times), and (2) the emergence and persistence of cyclical delay feedback loops (H_1 features). These topological features provide a structural lens into how delays accumulate, propagate, and reinforce across multiple interconnected airports.

Comparative analysis between ANA and JAL from 2018 to 2021 revealed clear airline-specific operational patterns. JAL's network exhibited greater fragmentation, with loops that tended to form and collapse abruptly, whereas ANA sustained longer-lasting loops with broader tails especially in winter and during disrupted periods. Seasonal disaggregation confirmed that winter schedules, shaped by adverse weather and holiday surges, give rise to longer-lived clusters and more persistent loops, consistent with findings by Sugishita et al. (2024) on seasonal delay motifs. Across both carriers, over 80% of delay loops dissolved within 10 minutes under normal operations. However, winter loops exhibited rightward-shifted persistence distributions, with both ANA and JAL frequently generating loops that persisted beyond 20-30 minutes.

COVID-19 served as a quasi-natural experiment. During the 2020 operational collapse, delay loops were smaller and short-lived due to reduced flight interdependencies—validating the link between network density and feedback loop formation (Fleurquin et al., 2013). By 2021, loop behavior returned but exhibited new bimodal patterns, particularly in ANA's winter network, with mid-range persistence modes (30–60 minutes) coexisting alongside short-lived loops signaling evolving structural vulnerabilities in the recovery phase.

Crucially, persistent loops consistently centered around key airports such as Tokyo Haneda (HND), Chitose (CTS), Okinawa (OKA), Itami (ITM), Fukuoka (FUK), and Nagoya (NGO). These airport-sets, particularly in three- to five-airport configurations, dominated both frequency and high-persistence rankings, marking them as structurally entrenched delay circuits. This supports earlier research on critical hubs in propagation networks (e.g., Güvercin et al., 2021; Fleurquin et al., 2013) and validates the utility

of persistent homology in highlighting emergent, threshold-robust risk zones beyond what standard network metrics can reveal.

The birth time of a delay loop serves as a critical indicator of when tightly interdependent airport sets begin exhibiting synchronized delay behavior. This metric enables air traffic managers and airline operators to anticipate the onset of systemic delay propagation. When observed delays approach known birth thresholds, targeted interventions such as buffer allocation, gate reassignment, or flow regulation can be implemented preemptively to prevent full-loop formation. Persistence, in contrast, reflects the structural resilience of a delay loop. High-persistence loops signify enduring feedback mechanisms that persist across a range of delay thresholds. These loops are indicative of prolonged vulnerabilities and warrant strategic interventions, including schedule adjustments, turnaround time extensions, or infrastructure enhancement at involved nodes. Recurrent loop patterns of airport-sets involving the same airports across multiple operational periods should be mapped into network-level delay risk profiles. Such patterns offer insight into structural hot spots and inform joint management protocols. Collaborative coordination especially during seasonal peaks can involve synchronized pushback planning, contingency asset sharing, and regional flow harmonization among frequently co-occurring airports.

Operationally, persistence and birth metrics can guide both short-term response and long-term planning. Longer-lived loops may signal feedback loops that require immediate disruption through minimal yet strategic changes, such as rerouting a single flight to break the loop. This aligns with high-leverage, low-cost interventions supported in adaptive ATFM frameworks (Dalmau et al., 2022).

Overall, the use of persistent homology offers a shift from local, pairwise delay analysis to a holistic, topological understanding of delay dynamics. This approach not only supports comparative assessments such as between ANA and JAL or across seasonal phases but also enables integration with predictive analytics and collaborative platforms like digital twin systems (Chen et al., 2024), promoting coordinated, data-driven decision-making across the air traffic ecosystem.

This analysis focused solely on departure delays as edge weights and used daily-aggregated data to construct simplicial complexes. Future studies could enhance the framework by incorporating multiple edge weights such as turnaround times, arrival delays, or passenger connection pressures and define a multivariate filtration function. Additionally, a finer temporal granularity (e.g., hourly snapshots) could offer deeper insights into within-day loop formation and collapse, supporting more tactical delay management and schedule buffering. Exploring multiplex delay networks (as in Zhang et al., 2024) could also uncover inter-airline feedback loops and shared vulnerabilities, further bridging individual carrier insights with systemic airspace resilience.

In conclusion, persistent homology opens new avenues for understanding and managing delay propagation in complex, seasonally fluctuating airline networks. As air transport systems grow in complexity, TDA provides a mathematically grounded, data-driven tool to detect—and ultimately preempt—the feedback patterns that drive operational inefficiency.

Chapter 6 Conclusion

6.1 Summary and Conclusions

This dissertation is set to develop a data-driven understanding of departure delay in Japan's domestic airline network by integrating network science, machine learning, and topological data analysis. Across three major analytic objectives, this work produced novel insights, demonstrated the utility of advanced methodologies, and provided actionable recommendations for airlines, air traffic managers, and policymakers.

Firstly, the study established that network-centric features constitute effective predictors of departure delay that provided explanatory power beyond conventional temporal and operational variables. Through Prais-Winsten regression model in panel data setting with panel corrected standard error on daily delay networks, in-degree centrality consistently emerged as a significant factor: airports receiving more delayed departed flights often experienced longer departure delays. Conversely, Betweenness centrality displayed context-dependent effects: while hub airports sometimes buffered delays effectively, certain spoke airports with high betweenness became critical nodes for delay propagation. Eigenvector centrality revealed that airports embedded within well-connected subnetworks (i.e., highly delayed subnetwork) often experienced shorter delays suggesting the benefits of coordinated operations. Transitivity effects were heterogeneous across airlines and seasons with higher clustering sometimes dissipating delays (as with ANA) and other times amplifying them (as observed with JAL).

Secondly, the study introduced uncertainty-aware forecasting using adaptive conformal prediction models representing the first application of this approach in air traffic delay management. Random Forest confirmed that lagged delays, reflecting temporal dependence, were the most important predictor of next-day average delays, while some network metrices such as: in-/out degree and betweenness ranked consistently among the top predictive features. Also, adding network metrices actually enhances the predictability of departure delay forecasts. Comparative analysis across Adaptive Conformal Inference (ACI), Aggregated Adaptive Conformal Inference (AgACI), and Fully Adaptive Conformal Inference (FACI) methods demonstrated that: ACI produced the narrowest prediction intervals under stable conditions but under-covered during abrupt surges. AgACI delivered robust coverage in volatile periods with more conservative interval widths. FACI achieved a balance between adaptiveness and precision which ensured dynamically adjusting to real-time variability.

These findings illustrate that airlines can use prediction intervals to their operational priorities whether emphasizing efficiency or resilience. Furthermore, the observation that prediction region size and

adaptiveness varied between ANA and JAL underscored the importance of accounting for airline-specific operational practices.

Lastly, Topological Data Analysis (TDA) revealed higher-order structural patterns of delay propagation that conventional pairwise network analysis often lacks in detection. Persistent homology through Vietoris-Rips simplicial complex uncovered recurrent multi-airport loops, particularly among Haneda, Chitose, Naha, Itami, Fukuoka, and Nagoya, that tended to be more prominent in winter and the early pandemic recovery period. These loops provided quantitative indicators (birth and persistence) of systemic vulnerability that enables operators to anticipate where and when loops of delays were likely to emerge. The research also highlighted airline-specific susceptibilities, such as characteristic three-airport loops for ANA and more extensive five- to six-airport loops for JAL.

Collectively, these contributions demonstrate that combining network measures, adaptive uncertainty quantification, and topological techniques provides both predictive insights and diagnostic capabilities for monitoring and managing departure delays. By enumerating the frequency of delay-generating configurations across seasons and quantifying their persistence, the dissertation also offers an empirical basis for proactive operational interventions.

6.2 Limitations

Despite these contributions, the research has several limitations. The study exclusively focused on the delay network rather than integrating the broader flight network of scheduled operations, potentially omitting structural context. In defining delay events, it employed a 1-minute or more threshold instead of the conventional 15-minute benchmark commonly used in industry reporting which may limit comparability. The reliance on daily-aggregated data constrained the granularity of within-day dynamics and lacked the examination of flight-by-flight variability or passenger connections. Weather data and finer-grained operational variables such as: turnaround times and load factors were not incorporated. Furthermore, while associations (correlations) were established, the analysis did not implement causal inference frameworks limiting the ability to draw definitive conclusions about interventions or simulate policy impacts. Finally, given the novelty of combining network metrics, conformal prediction, and TDA in this setting, there were no established benchmark models for comparative validation which underscores exploratory nature of the work.

6.3 Future Studies

Future studies can address these limitations and build on the insights generated here. Some important research directions include the integration of causal inference techniques, such as graphical causal models and do-calculus, to more rigorously account for confounding effects and temporal dependencies.

Expanded data integration incorporating weather, arrival delays, turnaround times, and land-side factors would likely improve both explanatory power and forecast reliability. Exploring multivariate filtration functions in TDA could further enrich the characterization of multidimensional delay dynamics. Sub-daily (e.g., hourly) network analysis would enable understanding of intra-day delay formation and dissipation. Additionally, extending the framework to multiplex and international networks including low-cost carriers and cross-border flights could assess generalizability and refine operational best practices. Methodological innovations such as: experimenting with deep learning models or systematically tuning conformal update rates hold promise for improving real-time adaptation. Finally, developing decision-support systems that embed adaptive prediction intervals into prescriptive analytics workflows can help resource allocation and collaborative traffic management.

This dissertation advances the understanding of departure delay propagation in complex airline networks by combining classical network properties, machine learning uncertainty quantification, and topological data analysis. In doing so, it demonstrates not only how these methods reveal previously hidden structural patterns but also how they can directly inform operational and strategic decision-making. The tools and insights developed here have possibility to contribute to the vision of trajectory-based air traffic management supporting Japan's CARATS initiative and offering transferable approaches for aviation systems worldwide. As air traffic networks become increasingly interconnected and exposed to new disruptions, embracing data-driven and uncertainty-aware strategies can be essential to achieve resilient and efficient operations.

References

Aggarwal, C. C. (2017). Outlier analysis (2nd ed.). Springer.

Ajayi, J., Xu, Y., Li, L., & Wang, K. (2024). Enhancing flight delay predictions using network centrality measures. Information, 15(9), 559. https://doi.org/10.3390/info15090559

Airport Cooperative Research Program (ACRP), National Academies of Sciences, Engineering, and Medicine. (2014). Defining and measuring aircraft delay and airport capacity thresholds. The National Academies Press. https://doi.org/10.17226/22428

All Nippon Airways (ANA). (2024). ANA transit information. https://www.ana.co.jp/en/jp/guide/boarding-procedures/checkin/domestic/transit/

An, W. (2021). A tale of twin dependence: A new multivariate regression model and an FGLS estimator for analyzing outcomes with network dependence. Sociological Methods & Research. https://doi.org/10.1177/00491241211031263

An, W. (2023). A tale of twin dependence: A new multivariate regression model and an FGLS estimator for analyzing outcomes with network dependence. Sociological Methods & Research, 52(4), 1947–1980. https://doi.org/10.1177/00491241211031263

Balakrishnan, H. (2015). Delay propagation in the air transportation network. Lecture notes on Air Transportation Systems and Infrastructure (Module 22), Istanbul Technical University.

Bao, J., Yang, Z., & Zeng, W. (2021). Graph to sequence learning with attention mechanism for network-wide multi-step-ahead flight delay prediction. Transportation Research Part C: Emerging Technologies, 130, 103323. https://doi.org/10.1016/j.trc.2021.05.011

Baltagi, B. H., Feng, Q., & Kao, C. (2012). A Lagrange multiplier test for cross-sectional dependence in a fixed effects panel data model. Journal of Econometrics, 170(1), 164–177. https://doi.org/10.1016/j.jeconom.2012.04.004

Bechberger, D., & Perryman, J. (2020). Graph databases in action. Manning Publications.

Beck, N. K., & Katz, J. N. (1995). What to do (and not to do) with time-series cross section data. American Political Science Review, 89(3), 634–647. https://doi.org/10.2307/2082979

Bianconi, G. (2021). Higher-order networks: An introduction to simplicial complexes. Cambridge University Press.

Birolini, S., & Jacquillat, A. (2023). Day-ahead aircraft routing with data-driven primary delay predictions. European Journal of Operational Research, 310(1), 379–396. https://doi.org/10.1016/j.ejor.2023.02.035

Bisandu, D. B. and Moulitsas, I. (2023). A deep bilstm machine learning method for flight delay prediction classification. The Journal of Aviation/Aerospace Education and Research, 32(2):1–32.

Bisandu, D. B. and Moulitsas, I. (2024). Prediction of flight delay using deep operator network with gradient-mayfly optimisation algorithm. Expert Systems with Applications, 247:123306.

Boehmke, B., & Greenwell, B. (2020). Hands-on machine learning with R. CRC Press.

Breiman, L. (2001). Random forests. Machine Learning, 45(1), 5–32.

Breusch, T. S., & Pagan, A. R. (1980). The Lagrange multiplier test and its applications to model specification in econometrics. The Review of Economic Studies, 47(1), 239–253. https://doi.org/10.2307/2297111

Bubenik, P. (2015). Statistical topological data analysis using persistence landscapes. Journal of Machine Learning Research, 16, 77–102.

Bubenik, P., & Dłotko, P. (2017). A persistence landscapes toolbox for topological statistics. Journal of Symbolic Computation, 78, 91–114. https://doi.org/10.1016/j.jsc.2016.03.009

Budd, L., & Ison, S. (2017). Air transport management: An international perspective. Routledge.

Bureau of Transportation Statistics. (2020). https://www.bts.gov/ (Accessed March 26, 2020)

CAAC. (2021). Civil Aviation Administration of China report No.88.

Cai, K., Li, Y., Fang, Y.-P., and Zhu, Y. (2022). A deep learning approach for flight delay prediction through time-evolving graphs. IEEE Transactions on Intelligent Transportation Systems, 23(8):11397–11407.

CANSO (2020). Recommended kpi for measuring anso operational performance. https://www2023.icao.int/NACC/Documents/Meetings/2018/ASBU18/OD-09-RecommendedKPIforMeasuringANSOOperationalPerformance.pdf. Accessed on: July 5, 2025.

Carlsson, G. (2009). Topology and data. Bulletin of the American Mathematical Society, 46(2), 255–308.

Carvalho, L., Sternberg, A., Maia Goncalves, L., Cruz, A. B., Soares, J. A., Brandão, D., Carvalho, D., & Ogasawara, E. (2021). On the relevance of data science for flight delay research: A systematic review. Transport Reviews, 41(4), 499–528. https://doi.org/10.1080/01441647.2020.1861123

Chandra, A. and Verma, A. (2025). To delay or not to delay? a hybrid relationship between departure delay, en-route conflict probability, and number of conflicts. Journal of the Air Transport Research Society, 4:100053.

Chen, G., Fricke, H., Okhrin, O., and Rosenow, J. (2024). Flight delay propagation inference in air transport networks using the multilayer perceptron. Journal of Air Transport Management, 114:102510.

Chen K, Nadirsha TNM, Lilith N, Alam S, Svensson Å (2024) Tangible digital twin with shared visualization for collaborative air traffic management operations. Transportation Research Part C: Emerging Technologies 161:104546. https://doi.org/10.1016/j.trc.2024.104546

Choi, S., Kim, Y. J., Briceno, S., & Mavris, D. (2016). Prediction of weather-induced airline delays based on machine learning algorithms. In 2016 IEEE/AIAA 35th Digital Avionics Systems Conference (pp. 1–10). IEEE.

Cimenler, O., Reeves, K. A., & Skvoretz, J. (2014). A regression analysis of researchers' social network metrics on their citation performance in a college of engineering. Journal of Informetrics, 8(3), 667–682. https://doi.org/10.1016/j.joi.2014.06.004

Cirium (2019). On-time performance review 2019. https://www.chinatravelnews.com/images/202001/8be7533843f16ebe.pdf. Accessed: July 5, 2025.

Cirium (2021).On-time performance review 2021. https://assets.fta.cirium.com/wp-content/uploads/2023/11/27234144/2021-On-Time-Performance-Report-prmc.pdf.
Accessed: July 5, 2025.

Cook, A., Tanner, G., Williams, V., & Meise, G. (2009). Dynamic cost indexing: Managing airline delay costs. Journal of Air Transport Management, 15(1), 26–35. https://doi.org/10.1016/j.jairtraman.2008.07.001

Cook A, Blom HAP, Lillo F, Mantegna RN, Miccichè S, Rivas D, Vázquez R, Zanin M (2015) Applying complexity science to air traffic management. Journal of Air Transport Management 42:149–158. https://doi.org/10.1016/j.jairtraman.2014.09.011

Dai, M. (2024). A hybrid machine learning-based model for predicting flight delay through aviation big data. Scientific Reports, 14(1):4603.

Dalmau, R., De Falco, P., Spak, M., & Rodriguez Varela, J. D. (2023). Probabilistic pre-tactical arrival and departure flight delay prediction with quantile regression: A case study for Geneva International Airport using operational data. In Proceedings of the Fifteenth USA/Europe Air Traffic Management Research and Development Seminar (ATM2023).

Dey, A. K., Gel, Y. R., & Poor, H. V. (2019). What network motifs tell us about resilience and reliability of complex networks. Proceedings of the National Academy of Sciences, 116(39), 19368–19373. https://doi.org/10.1073/pnas.1819529116

Dmochowski PA, Skorupski J (2017) Air traffic smoothness: a new look at the air traffic flow management. Transportation Research Procedia 28:127–132.

Dong, X., Zhu, X., & Zhang, J. (2024). Departure flight delay prediction due to ground delay program using multilayer perceptron with improved sparrow search algorithm. The Aeronautical Journal, 128, 706–724. https://doi.org/10.1017/aer.2023.83

Du, W.-B., Zhang, M.-Y., Zhang, Y., Cao, X.-B., and Zhang, J. (2018). Delay causality network in air transport systems. Transportation Research Part E: Logistics and Transportation Review, 118:466–476.

Edelsbrunner, H. and Harer, J. (2010). Computational Topology: An Introduction. American Mathematical Society.

Elchesen, A. (2025). Persistence homology: Introduction [Lecture notes: Week 5]. DSCI 475 – Topological Data Analysis. Retrieved from https://www.alexelchesen.com/teaching/tda-spring-25

Esmaeilzadeh, E. and Mokhtarimousavi, S. (2020). Machine learning approach for flight departure delay prediction and analysis. Transportation Research Record: Journal of the Transportation Research Board, 2674(8):1–12.

Feng, D., Hao, B., and Lai, J. (2024). Tracing delay network in air transportation combining causal propagation and complex network. International Journal of Intelligent Networks, 5:63–76.

Fleurquin P, Ramasco JJ, Eguiluz VM (2013) Systemic delay propagation in the US airport network. Scientific Reports 3:1159. https://doi.org/10.1038/srep01159

Gao, J., Barzel, B., & Barabási, A.-L. (2016). Universal resilience patterns in complex networks. Nature, 530(7590), 307–312. https://doi.org/10.1038/nature16948

Gibbs, I., & Candès, E. J. (2021). Adaptive conformal inference under distribution shift. In Advances in Neural Information Processing Systems, 35. https://papers.neurips.cc/paper/2021/file/0d441de75945e5acbc865406fc9a2559-Paper.pdf

Gibbs, I., & Candès, E. J. (2024). Conformal inference for online prediction with arbitrary distribution shifts. Journal of Machine Learning Research, 25, 1–36. http://jmlr.org/papers/v25/22-1218.html

Gopalkrishnan, K., & Balakrishnan, H. (2017). A comparative analysis of models for predicting delays in air traffic networks. In Proceedings of the 12th USA/Europe Air Traffic Management Research and Development Seminar (ATM 2017) (pp. 27–30). Seattle, WA, USA.

Greene, W. H. (2012). Econometric analysis (7th ed.). Prentice Hall.

Güner, S., Antunes, J. J. M., Codal, K. S., & Wanke, P. (2024). Network centrality driven airport efficiency: A weight-restricted network DEA. Journal of Air Transport Management, 116, 102551. https://doi.org/10.1016/j.jairtraman.2024.102551

Güvercin, M., Ferhatosmanoglu, N., & Gedik, B. (2021). Forecasting flight delays using clustered models based on airport networks. IEEE Transactions on Intelligent Transportation Systems, 22(5), 3179–3189. https://doi.org/10.1109/TITS.2020.2986940

H2O.ai. (2024). h2o: R interface for the H2O scalable machine learning platform (Version 3.44.0.3). https://cran.r-project.org/package=h2o

Hanaoka, S. (2018). Low-cost carriers in the Japanese aviation market. In X. Fu & J. Peoples (Eds.), Airline economics in Asia (pp. 9–31). Emerald Publishing. https://doi.org/10.1108/S2212-160920180000007003

Hardell, H., Otero, E., Polishchuk, T., and Smetanová, L. (2025). Optimizing air traffic management through point merge procedures: Minimizing delays and environmental impact in arrival operations. Journal of Air Transport Management, 123:102706.

Hickok, A., Jarman, B., Johnson, M., Luo, J., and Porter, M. A. (2024). Persistent homology for resource coverage: A case study of access to polling sites. SIAM Review, 66(3).

Hoechle, D. (2007). Robust standard errors for panel regressions with cross-sectional dependence. The Stata Journal, 7(3), 281–312. https://doi.org/10.1177/1536867X070070030

James, G., Witten, D., Hastie, T., & Tibshirani, R. (2021). An introduction to statistical learning (2nd ed.). Springer.

Japan Airlines (JAL). (2024). JAL connection information. https://www.jal.co.jp/jp/en/dom/boarding/onto/ (Accessed October 12, 2024)

Jayasinghe, A., Sano, K., & Nishiuchi, H. (2015). Network centrality assessment as an alternative approach to predict vehicular traffic volume: A case of Colombo, Sri Lanka. Journal of the Eastern Asia Society for Transportation Studies, 11, 834–853. https://doi.org/10.11175/easts.11.834

Kang, J., Yang, S., Shan, X., Bao, J., & Yang, Z. (2023). Exploring delay propagation causality in various airport networks with attention-weighted recurrent graph convolution method. Aerospace, 10(5), 453. https://doi.org/10.3390/aerospace10050453

Kato, H., Kimura, Y., Okumura, M., & Tsukai, M. (2008). Analysis of flight delays and cancellations at Sendai airport. Journal of the Japan Society of Civil Engineers, Ser. D3 (Infrastructure Planning and Management), 25, 723–729. https://doi.org/10.2208/journalip.25.723

Khan, W. A., Ma, H.-L., Chung, S.-H., & Wen, X. (2021). Hierarchical integrated machine learning model for predicting flight departure delays and duration in series. Transportation Research Part C: Emerging Technologies, 129, 103225. https://doi.org/10.1016/j.trc.2021.103225

Kiliç, K. and Sallan, J. M. (2023). Study of delay prediction in the us airport network. Aerospace, 10(4):342.

Kim, S., & Park, E. (2024). Prediction of flight departure delays caused by weather conditions adopting data-driven approaches. Journal of Big Data, 11(1), Article 11. https://doi.org/10.1186/s40537-023-00867-5

Kopsidas, A., Douvaras, A., & Kepaptsoglou, K. (2023). Exploring the association between network centralities and passenger flows in metro systems. Applied Network Science, 8(1), 69. https://doi.org/10.1007/s41109-023-00583-2

Li, C., Mao, J., Li, L., Wu, J., Zhang, L., Zhu, J., and Pan, Z. (2024). Flight delay propagation modeling: Data, methods, and future opportunities. Transportation Research Part E: Logistics and Transportation Review, 185:103525.

Li, M. Z., Ryerson, M. S., and Balakrishnan, H. (2019). Topological data analysis for aviation applications. Transportation Research Part E: Logistics and Transportation Review, 128:148–174.

Liu, K., Hu, F., Lin, H., Cheng, X., Chen, J., Song, J., Feng, S., Su, G., and Zhu, C. (2024). Deep reinforcement learning for real-time ground delay program revision and corresponding flight delay assignments. arXiv preprint arXiv:2405.08298.

Lonzius, C., & Lange, A. (2024). Aircraft routing clusters and their impact on airline delays. Journal of Air Transport Management, 114, 102493. https://doi.org/10.1016/j.jairtraman.2023.102493

Luo, D., Cats, O., & van Lint, H. (2020). Can passenger flow distribution be estimated solely based on network properties in public transport systems? Transportation, 47, 2757–2776. https://doi.org/10.1007/s11116-019-09990-w

Mamdouh, M., Ezzat, M., & Hefny, H. (2023). A novel intelligent approach for flight delay prediction. Journal of Big Data, 10, Article 179. https://doi.org/10.1186/s40537-023-00854-w

Mamdouh, M., Ezzat, M., & Hefny, H. (2024). Improving flight delays prediction by developing attention-based bidirectional LSTM network. Expert Systems with Applications, 238, 121747. https://doi.org/10.1016/j.eswa.2023.121747

Menon PK, Park SG (2016) Dynamics and control technologies in air traffic management. Annual Reviews in Control 42:271–284. https://doi.org/10.1016/j.arcontrol.2016.09.012

Menzel, K. (2020). Many player asymptotics for large network formation problems. In B. Graham & Á. de Paula (Eds.), The econometric analysis of network data (pp. 183–223). Academic Press.

Mokhtarimousavi, S. and Mehrabi, A. (2023). Flight delay causality: Machine learning technique in conjunction with random parameter statistical analysis. International Journal of Transportation Science and Technology, 12(1):230–244.

Muros Anguita, J. G. M., & Olariaga, O. D. (2024). Prediction of departure flight delays through predictive tools based on machine learning/deep learning algorithms. The Aeronautical Journal, 128, 111–133. https://doi.org/10.1017/aer.2023.41

Myers, A., Muñoz, D., Khasawneh, F.A., & Munch, E. (2023). Temporal network analysis using zigzag persistence. EPJ Data Science, 12(6). https://doi.org/10.1140/epjds/s13688-023-00379-5

National Academies of Sciences, Engineering, and Medicine. (2014). Defining and measuring aircraft delay and airport capacity thresholds. The National Academies Press. https://doi.org/10.17226/22428

Newman, M. (2018). Networks (2nd ed.). Oxford University Press.

Ng, K. T., Fu, X., Hanaoka, S., and Oum, T. H. (2022). Japanese aviation market performance during the covid-19 pandemic: Analyzing airline yield and competition in the domestic market. Transport Policy, 116:237–247.

Niu, X., Jiang, C., Gao, J., Korniss, G., & Szymanski, B.K. (2021). From data to complex network control of airline flight delays. Scientific Reports, 11, 18715. https://doi.org/10.1038/s41598-021-98112-7

Nivitha, K., Madhavan, V., Megha, V., Dinesh Kumar, S., and Jothi Lakshmi, S. (2023). An ensemble approach for flight delay prediction through spatiotemporal parameters. In Proceedings of the 2023 9th International Conference on Advanced Computing and Communication Systems (ICACCS), pages 1–7.

OAG. (2023). OAG historical flight data. https://www.oag.com/historical-flight-data

Odoni, R. (1997). Existing and required modelling capabilities for evaluating ATM systems and concepts (Final Report No. NAG2-997). International Centre for Air Transportation.

Okamoto, S. (2022). State of emergency and human mobility during the COVID-19 pandemic in Japan. Journal of Transport and Health, 26, 101405. https://doi.org/10.1016/j.jth.2022.101405

Operational considerations for managing COVID-19 cases or outbreak in aviation. (2020). World Health Organization. https://iris.who.int/bitstream/handle/10665/331488/WHO-2019-nCoV-Aviation-2020.1-eng.pdf (Accessed September 30, 2024)

Pang, Y., Zhao, P., Hu, J., and Liu, Y. (2024). Machine learning-enhanced aircraft landing scheduling under uncertainties. Transportation Research Part C: Emerging Technologies, 158:104444.

Pesaran, M. H. (2004). General diagnostic tests for cross section dependence in panels. Cambridge Working Papers in Economics, No. 0435. University of Cambridge.

Prais, S. J., & Winsten, C. B. (1954). Trend estimators and serial correlation. Cowles Commission Discussion Paper No. 383.

Prandini, M., Putta, V., & Hu, J. (2012). Air traffic complexity in future air traffic management systems. Journal of Aerospace Operations, 1(3), 281–292. https://doi.org/10.3233/AOP-2012-0015

Pyrgiotis, N., Malone, K. M., & Odoni, A. (2013). Modelling delay propagation within an airport network. Transportation Research Part C: Emerging Technologies, 27, 60–75. https://doi.org/10.1016/j.trc.2011.05.017

Qu, J., Zhao, T., Ye, M., Li, J., and Liu, C. (2020). Flight delay prediction using deep convolutional neural network based on fusion of meteorological data. Neural Processing Letters, 52(2):1461–1484.

Rainone, E. (2020). Estimating spillover effects with bilateral outcomes. In Á. de Paula, E. Tamer, & M.-C. Voia (Eds.), The econometrics of networks (Vol. 42, pp. 293–314). Emerald Publishing. https://doi.org/10.1108/S0731-905320200000042014

Rajput, A. A., Nayak, S., Dong, S., and Mostafavi, A. (2023). Anatomy of perturbed traffic networks during urban flooding. Sustainable Cities and Society, 97:104693.

Rebollo, J. J., & Balakrishnan, H. (2014). Characterization and prediction of air traffic delays. Transportation Research Part C: Emerging Technologies, 44, 231–241. https://doi.org/10.1016/j.trc.2014.04.007

Ribeiro, N. A., Tay, J., Ng, W., & Birolini, S. (2025). Delay predictive analytics for airport capacity management. Transportation Research Part C: Emerging Technologies, 171, 104947. https://doi.org/10.1016/j.trc.2024.104947

Sadeek, S. N., Hanaoka, S., & Sugishita, K. (2025). Examining the association between network properties and departure delay duration in Japan's domestic aviation. Aerospace, 12(2), 137. https://doi.org/10.3390/aerospace12020137

Schultz, M., Evler, J., Asadi, E., Preis, H., Fricke, H., & Wu, C.-L. (2020). Future aircraft turnaround operations considering post-pandemic requirements. Journal of Air Transport Management, 89, 101886. https://doi.org/10.1016/j.jairtraman.2020.101886

Statista. (2023). Share of domestic flights delayed in Japan. https://www.statista.com/statistics/1280563/japan-share-domestic-flights-delayed/ (Accessed August 29, 2024)

Sugishita, K., Arisawa, K., & Hanaoka, S. (2024). Delay propagation patterns in Japan's domestic air transport network. Transportation Research Interdisciplinary Perspectives, 27, 101235. https://doi.org/10.1016/j.trip.2024.101235

Sun, M., Tian, Y., Wang, X., Huang, X., Li, Q., Li, Z., and Li, J. (2024). Transport causality knowledge-guided gcn for propagated delay prediction in airport delay propagation networks. Expert Systems with Applications, 240:122426.

Sun, X., Wandelt, S., & Zhang, A. (2023). A data-driven analysis of the aviation recovery from the COVID-19 pandemic. Journal of Air Transport Management, 109, 102401.

Susmann, H., Chambaz, A., & Josse, J. (2024). Adaptive conformal inference under distribution shift. Computo. https://computo.sfds.asso.fr/published-202407-susmann-adaptive-conformal/

Tang, Z., Huang, S., and Han, S. (2021). Recent progress about flight delay under complex network. Complexity, 2021:Article ID 5513093, 18 pages.

Tang, Z., Huang, S., Zhu, X., Pan, W., Han, S., and Gong, T. (2023). Research on the multilayer structure of flight delay in China air traffic network. Physica A: Statistical Mechanics and its Applications, 609:128309.

Tay, R. (2017). Correlation, variance inflation and multicollinearity in regression model. Journal of the Eastern Asia Society for Transportation Studies, 12, 2006–2015. https://doi.org/10.11175/easts.12.2006

Tibshirani, R. (2024). Conformal prediction [Lecture notes]. In Advanced Topics in Statistical Learning. University of California, Berkeley. https://www.stat.berkeley.edu/~ryantibs/statlearn-s24/lectures/conformal.pdf

Torres, J. J., & Bianconi, G. (2020). Simplicial complexes: Higher-order spectral dimension and dynamics. Journal of Physics: Complexity, 1(1), 015002. https://doi.org/10.1088/2632-072X/ab82f5

Tu, Y., Ball, M. O., & Jank, W. S. (2008). Estimating flight departure delay distributions: A statistical approach with long-term trend and short-term pattern. Journal of the American Statistical Association, 103(481), 112–125. https://doi.org/10.1198/016214507000000257

Vandal, T., Livingston, M., Piho, C., & Zimmerman, S. (2018). Prediction and uncertainty quantification of daily airport flight delays, Proceedings of The 4th International Conference on Predictive Applications and APIs. Proceedings of Machine Learning Research, 82:45-51.

Vovk, V., Gammerman, A., & Shafer, G. (2005). Algorithmic learning in a random world. Springer.

Wandelt, S., Chen, X., & Sun, X. (2025a). Networks for flight delay analysis: A scoping review and research agenda. IEEE Transactions on Network Science and Engineering, 12(2), 1250–1266. https://doi.org/10.1109/TNSE.2025.3526850

Wandelt, S., Chen, X., & Sun, X. (2025b). Flight delay prediction: A dissecting review of recent studies using machine learning. IEEE Transactions on Intelligent Transportation Systems, 26(4), 4283–4297. https://doi.org/10.1109/TITS.2025.3528536

Wandelt, S., Wang, S., Chen, X., Zheng, C., Chang, S., & Sun, X. (2025c). Network structures in air transportation: A comprehensive review of applications and challenges. Journal of Air Transport Management, 126, 102794.

Wang, L., Tien, A., and Chou, J. (2021). Multi-airport delay prediction with transformers. arXiv preprint arXiv:2111.04494.

Wang, Y., Cao, Y., Zhu, C., et al. (2020). Universal patterns in passenger flight departure delays. Scientific Reports, 10:6890

Wang, D., Gong, Y., Zhang, X., & Lin, Y. (2022). The impact of multiscale urban road network centralities on taxi travel: A case study in Shenzhen. Complexity, Article 1780667. https://doi.org/10.1155/2022/1780667

Wang, Y., Li, M. Z., Gopalkrishnan, K., & Liu, T. (2022). Timescales of delay propagation in airport networks. Transportation Research Part E: Logistics and Transportation Review, 161, 102687. https://doi.org/10.1016/j.tre.2022.102687

Woo, Y.-B. and Moon, I. (2021). Scenario-based stochastic programming for an airline driven flight rescheduling problem underground delay programs. Transportation Research Part E: Logistics and Transportation Review, 150:102360.

Wright, M. N., & Ziegler, A. (2017). ranger: A fast implementation of random forests for high dimensional data in C++ and R. Journal of Statistical Software, 77(1), 1–17. https://doi.org/10.18637/jss.v077.i01

Wu, C.-L. (2010). Airlines operations and delay management: Insights from airline economics, networks and strategic schedule planning. Ashgate Publishing.

Wu, Y., Yang, H., Lin, Y., & Liu, H. (2024). Spatiotemporal propagation learning for network-wide flight delay prediction. IEEE Transactions on Knowledge and Data Engineering, 36(1), 386–400. https://doi.org/10.1109/TKDE.2023.3286690

Xu, J., Dai, L., & Hansen, M. (2024). Flight time and flight traffic before, during and after the pandemic: What has changed. Transportation Research Record, 2678(4), 203–216. https://doi.org/10.1177/03611981231184248

Xu, Y., & Zhao, X. (2024). How does node centrality in a financial network affect asset price prediction? The North American Journal of Economics and Finance, 73, 102163. https://doi.org/10.1016/j.najef.2024.102163

Yang, Z., Chen, Y., Hu, J., Song, Y., & Mao, Y. (2023). Departure delay prediction and analysis based on node sequence data of ground support services for transit flights. Transportation Research Part C: Emerging Technologies, 152, 104217. https://doi.org/10.1016/j.trc.2023.104217

Yao, H. G., & Zhang, H. (2023). Critical and steady-state characteristics of delay propagation in an airport network. PLOS ONE, 18(7): e0288200. https://doi.org/10.1371/journal.pone.0288200

Yimga, J. (2021). The airline on-time performance impacts of the COVID-19 pandemic. Transportation Research Interdisciplinary Perspectives, 10, 100386. https://doi.org/10.1016/j.trip.2021.100386

Yu, B., Guo, Z., Asian, S., Wang, H., & Chen, G. (2019). Flight delay prediction for commercial air transport: A deep learning approach. Transportation Research Part E: Logistics and Transportation Review, 125, 203–221. https://doi.org/10.1016/j.tre.2019.03.007

Zeng, L., Wang, B., Wang, T., and Wang, Z. (2022). Research on delay propagation mechanism of air traffic control system based on causal inference. Transportation Research Part C: Emerging Technologies, 138:103622

Zhang H, Wu W, Jiang Y, Chen X (2024) Flight delay propagation in the multiplex network system of airline networks. Physica A: Statistical Mechanics and its Applications 648:129883. https://doi.org/10.1016/j.physa.2024.129883

Zheng, H., Wang, Z., Zheng, C., Wang, Y., Fan, X., Cong, W., & Hu, M. (2024). A graph multi-attention network for predicting airport delays. Transportation Research Part E: Logistics and Transportation Review, 181, 103375. https://doi.org/10.1016/j.tre.2023.103375

Zheng, Z., Wei, W., and Hu, M. (2021). A comparative analysis of delay propagation on departure and arrival flights for a Chinese case study. Aerospace, 8(8):212.

Zografos, K., Andreatta, G., & Odoni, A. (2013). Modelling and managing airport performance: Theory and practice. John Wiley & Sons.

Appendices

Appendix A: Airports considered for ANA and JAL flights in the domestic Japanese market

Table A1: Considered Airports in Japan with IATA Code

Airport Name	IATA Code	Airport Name	IATA Code
Kansai International Airport	KIX	Yamagata Airport	GAJ
Narita International Airport	NRT	Amami Airport	ASJ
Chubu Centrair International Airport	NGO	Aomori Airport	AOJ
Tokyo International Airport (Haneda Airport)	HND	Fukushima Airport	FKS
Osaka International Airport (Itami Airport)	ITM	Hanamaki Airport	HNA
Akita Airport	AXT	New Ishigaski Airport	ISG
Asahikawa Airport	AKJ	Izumo Airport	IZO
New Chitose Airport	CTS	Kobe Airport	UKB
Fukuoka Airport	FUK	Matsumoto Airport	MMJ
Hakodate Airport	HKD	Memanbetsu Airport	MMB
Hiroshima Airport	HIJ	Miyako Airport	MMY
Kagoshima Airport	KOJ	Nanki-Shirahama Airport	SHM
Kitakyushu Airport	KKJ	Okayama Airport	OKJ
Kochi Airport	KCZ	Oki Airport	OKI
Kushiro Airport	KUH	Okushiri Airport	OIR
Kumamoto Airport	KMJ	Rishiri Airport	RIS
Matsuyama Airport	MYJ	Tanegashima Airport	TNE
Miyazaki Airport	KMI	Tokunoshima Airport	TKN
Nagasaki Airport	NGS	Yonaguni Airport	OGN
Naha Airport	OKA	Komatsu Airport	KMQ
Niigata Airport	KIJ	Misawa Airport	MSJ
Oita Airport	OIT	Nagoya (Komaki) Airport	NKM
Sendai Airport	SDJ	Okadama Airport	OKD
Takamatsu Airport	TAK	Tokushima Airport	TKS
Obihiro Airport	OBO	New Ishigaki Airport	ISG
Yamaguchi-Ube Airport	UBJ	Kitakyushu Airport	KKJ

Appendix B: Results for 15-minute threshold of Prais-Winsten Panel Regression (Chapter 3)

Table B1: Estimation results of single variable PW regression for ANA and JAL

Airline	Variables	2018				2019				2020				2021			
ANA	In-degree	0.35				0.32				0.23				0.89			
	BC		0.17				0.15				0.12				0.66		
	EC			0.03				0.03				0.004				0.01	
	Transitivity				-0.001				0.001				0.003				0.01
	Constant	0.001	0.004	0.002	0.009	0.001	0.003	0.002	0.008	0.002	0.003	0.004	0.005	-0.004	0.011	0.019	0.015
	ρ	0.14	0.12	0.16	0.33	0.16	0.13	0.16	0.35	0.15	0.14	0.19	0.25	0.50	0.74	0.64	0.84
	R-squared	0.58	0.55	0.17	<0.001	0.49	0.42	0.13	<0.001	0.21	0.28	0.01	<0.001	0.93	0.44	<0.001	<0.001
	Prob > chi2	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
	n	6,888				6,662				2,609				3,036			
	Number of airport	36				36				35				35			
	Number of time (range)	5-365				8-352				1-273				3-278			
JAL	In-degree	0.35				0.40				0.33				0.89			
	BC		0.19				0.21				0.17				0.61		
	EC			0.02				0.03				0.01				0.02	
	Transitivity				0.002				0.002				0.002				0.005
	Constant	0.0001	0.003	0.001	0.005	-0.001	0.002	0.0002	0.005	0.0003	0.002	0.002	0.003	-0.004	0.008	0.013	0.012
	ρ	0.17	0.16	0.19	0.41	0.13	0.12	0.12	0.39	0.05	0.10	0.13	0.22	0.66	0.74	0.54	0.81
	R-squared	0.73	0.62	0.17	<0.001	0.78	0.69	0.19	<0.001	0.60	0.61	0.07	<0.001	0.93	0.60	0.01	<0.001
	Prob > chi2	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
	n	6,608				6,999				2,710				4,005			
	Number of airport	44				45				44				46			
	Number of time (range)	3-361				1-357				1-296				5-313			

(NB: Variables are statistically significant at 5% significance level; Grey shading represents the insignificant variable)

Table B2: Estimation result of multi-variable PW regression (Pre-COVID phase) (Chapter 3)

Variables	ANA		JAL	
	2018	**2019**	**2018**	**2019**
In-degree	0.325***(0.018)	0.334***(0.013)	0.336***(0.010)	0.340***(0.010)
In-degree*Hub	-0.098***(0.023)	-0.078***(0.021)	-0.006 (0.018)	0.060**(0.023)
Betweenness	0.019 (0.014)	0.017*(0.007)	0.044***(0.006)	0.068***(0.006)
Betweenness *Hub	0.066***(0.016)	0.049***(0.011)	-0.027*(0.011)	-0.047***(0.013)
Eigenvector	-0.006***(0.001)	-0.005***(0.0006)	-0.002***(0.0003)	-0.002***(0.0005)
Eigenvector *Hub	-0.002 (0.002)	-0.006**(0.002)	-0.0001 (0.001)	-0.007***(0.001)
Transitivity	-0.0005* (0.0003)	0.0004**(0.0002)	0.0007***(0.0001)	0.0005***(0.0001)
Transitivity*Hub	0.004***(0.0008)	0.006***(0.0008)	0.003***(0.0006)	0.002***(0.0006)
Constant	0.002***(0.0002)	0.001***(0.0002)	0.0002*(0.0001)	-0.0001 (0.0001)
Number of Observations	6,888	6,662	6,608	6,999
Number of airport	36	36	44	45
Number of time (range)	5-365	8-352	3-361	1-357
R-squared	0.62	0.52	0.73	0.79
Prob > chi2	0.000	0.000	0.000	0.000

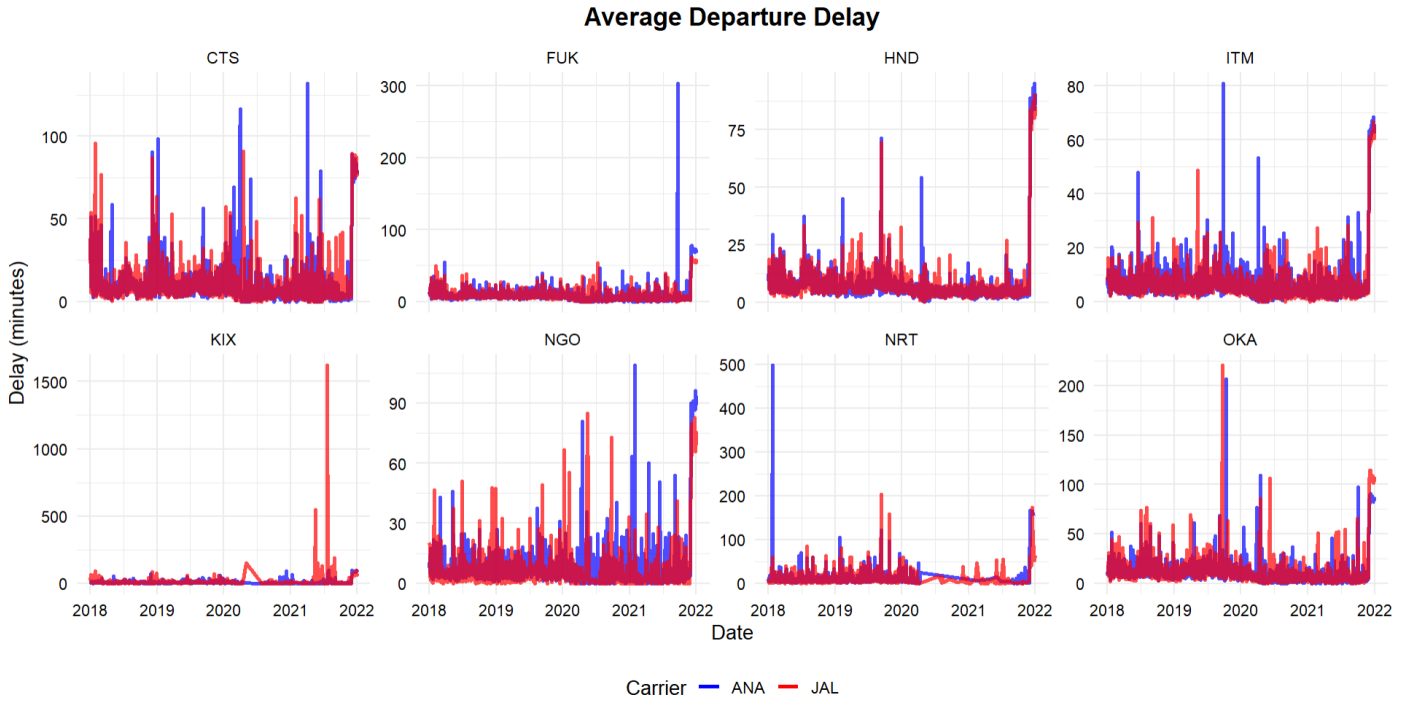
(Standard errors are reported in parentheses. The star marks represent the statistical significance level at '***' 1% '**' 5% '*' 10%)

Table B3: Estimation result of multi-variable PW regression (During-COVID phase) (Chapter 3)

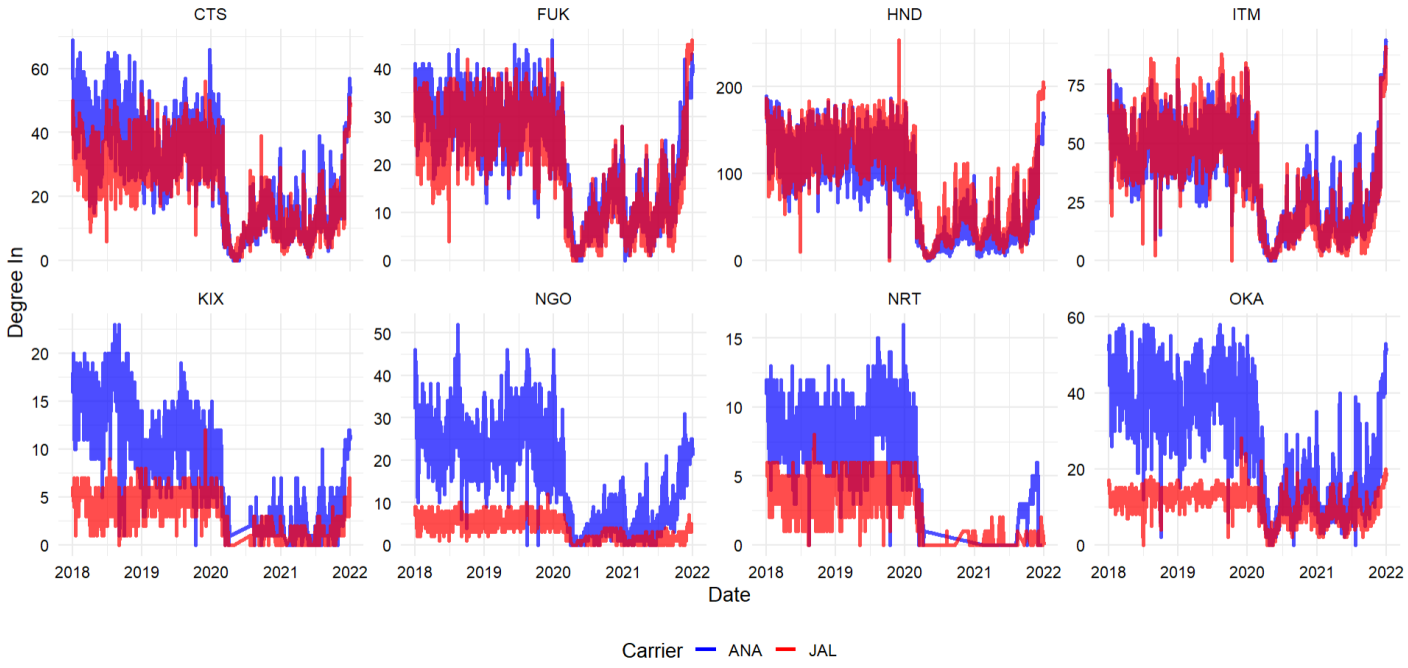
Variables	ANA		JAL	
	2020	**2021**	**2020**	**2021**
In-degree	0.157***(0.037)	-0.116 (0.157)	0.344***(0.024)	-0.033 (0.126)
In-degree*Hub	-0.158***(0.041)	0.218***(0.019)	-0.068*(0.028)	0.197***(0.015)
In-degree*Q2	0.490*(0.247)	-0.041 (0.215)	-0.153 (0.110)	0.018 (0.166)
In-degree*Q3	0.015 (0.096)	0.325*(0.180)	-0.240***(0.041)	0.277*(0.144)
In-degree*Q4	0.160 (0.056)	0.802***(0.157)	-0.227***(0.036)	0.690***(0.127)
Betweenness	0.047*(0.018)	0.113 (0.086)	0.008 (0.515)	0.170 (0.131)
Betweenness *Hub	0.071**(0.021)	-0.020 (0.031)	0.020 (0.015)	0.079***(0.016)
Betweenness *Q2	-0.408 (0.486)	0.031 (0.112)	0.602**(0.228)	-0.124 (0.149)
Betweenness *Q3	0.115*(0.054)	-0.078 (0.416)	0.215***(0.017)	-0.225*(0.135)
Betweenness *Q4	0.086**(0.035)	-0.064 (0.089)	0.130***(0.022)	-0.154 (0.132)
Eigenvector	-0.004**(0.001)	0.004*(0.002)	-0.002**(0.0007)	0.004***(0.001)
Eigenvector *Hub	0.003**(0.001)	-0.005***(0.001)	0.0008 (0.0006)	-0.005***(0.001)
Eigenvector *Q2	-0.002 (0.003)	0.0003 (0.003)	0.003*(0.001)	-0.0005 (0.002)
Eigenvector *Q3	0.0004 (0.002)	-0.004 (0.003)	0.003***(0.0009)	-0.003 (0.002)
Eigenvector *Q4	0.003*(0.002)	-0.017***(0.003)	0.003**(0.008)	-0.015***(0.002)
Transitivity	0.002***(0.0006)	0.004 (0.002)	0.0008*(0.0003)	0.002 (0.001)
Transitivity*Hub	0.0002 (0.001)	-0.004 (0.003)	0.0006 (0.0006)	-0.0007 (0.001)
Transitivity*Q2	-0.004 (0.003)	0.002 (0.004)	0.007 (0.005)	0.0004 (0.018)
Transitivity*Q3	-0.0006 (0.002)	0.00002 (0.003)	-0.0007 (0.011)	-0.001 (0.001)
Transitivity*Q4	0.0002 (0.001)	-0.004 (0.003)	-0.0004 (0.0007)	0.0001 (0.001)
Constant	0.003***(0.0002)	0.001***(0.0004)	0.0008***(0.0001)	0.0002 (0.0002)
Number of Observations	2,609	3,036	2,710	4,005
Number of airport	35	35	44	46
Number of time (range)	1-273	3-278	1-296	5-313
R-squared	0.30	0.95	0.71	0.97
Prob > chi2	0.000	0.000	0.000	0.000

(Standard errors are reported in parentheses. The star marks represent the statistical significance level at '***' 1% '**' 5% '*' 10%)

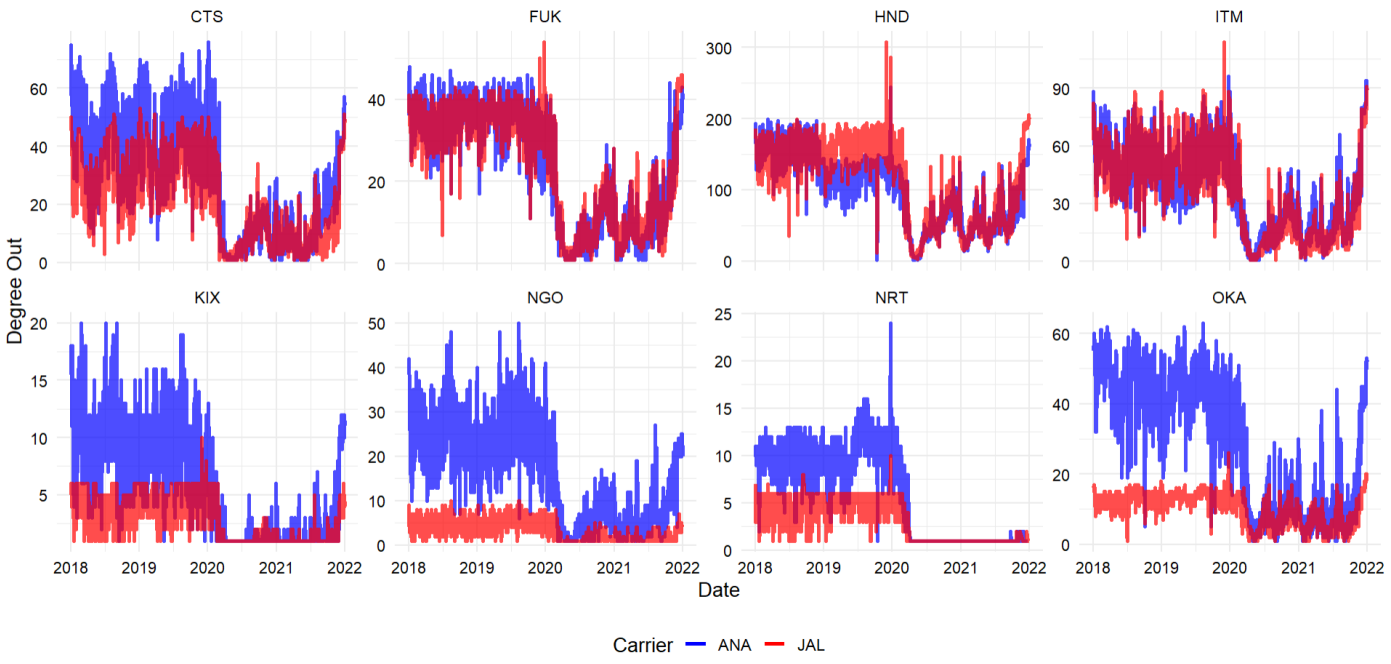
Appendix C: Comparison of time-series plot between ANA and JAL of different network properties and average departure delay (Chapter 4)



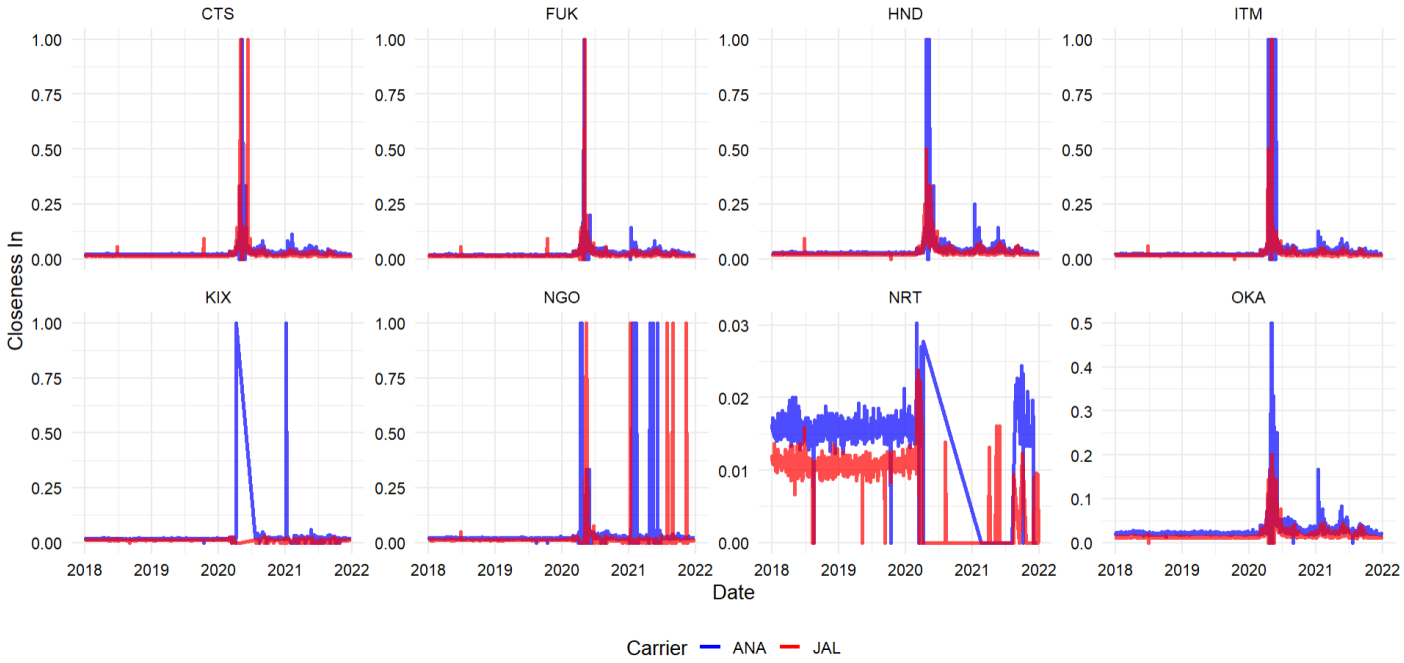
(a)



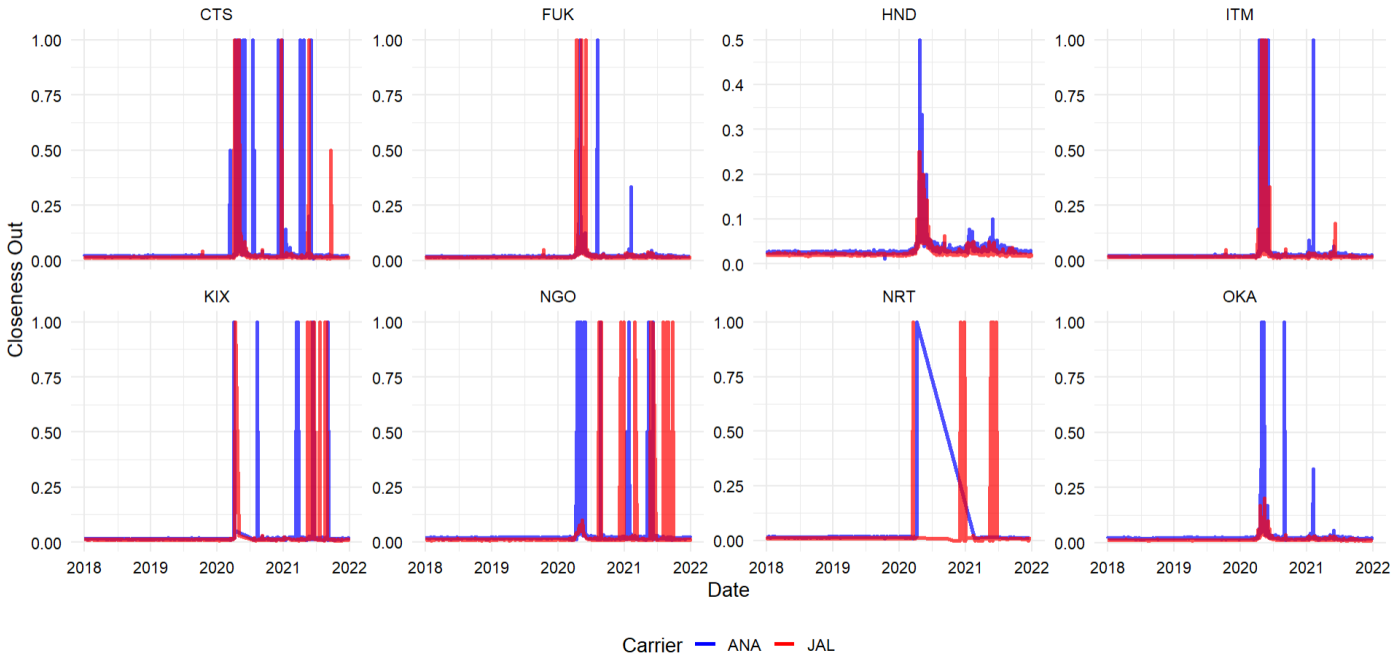
(b)


CTS
FUK
HND
ITM
KIX
NGO
NRT
OKA
Degree Out
Date
Carrier
ANA
JAL

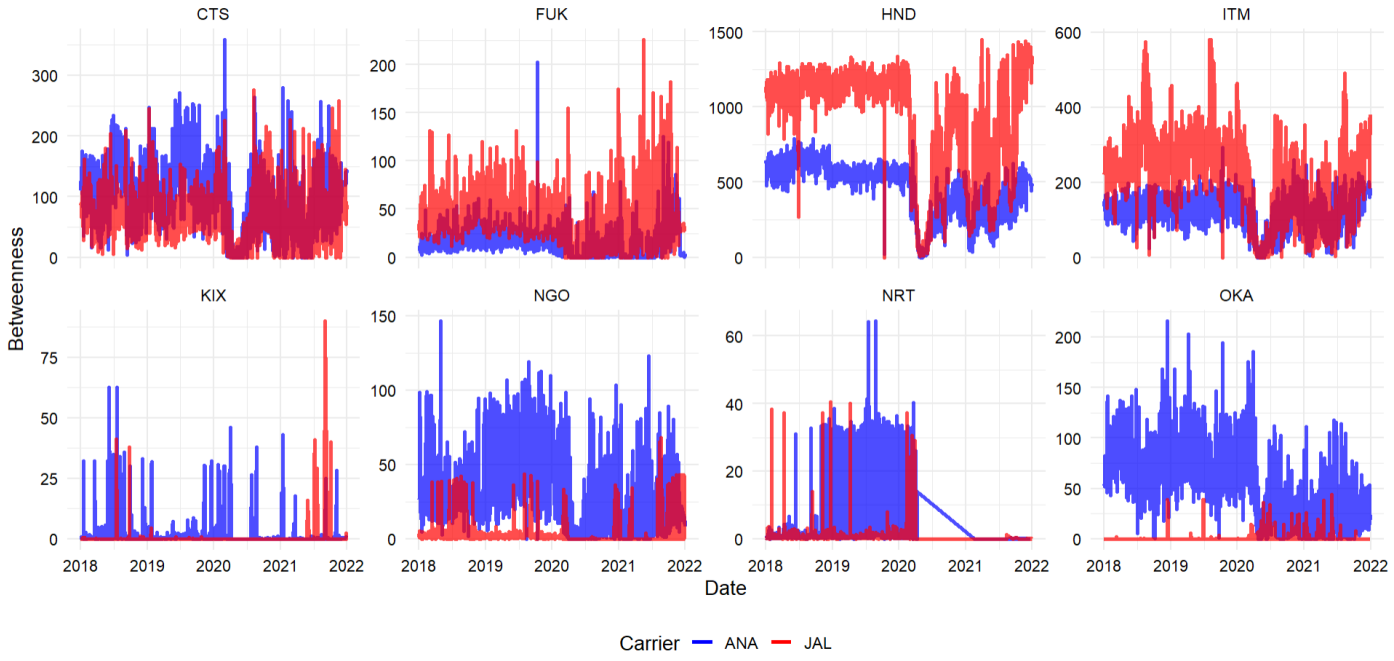
(c)


CTS
FUK
HND
ITM
KIX
NGO
NRT
OKA
Closeness In
Date
Carrier
ANA
JAL

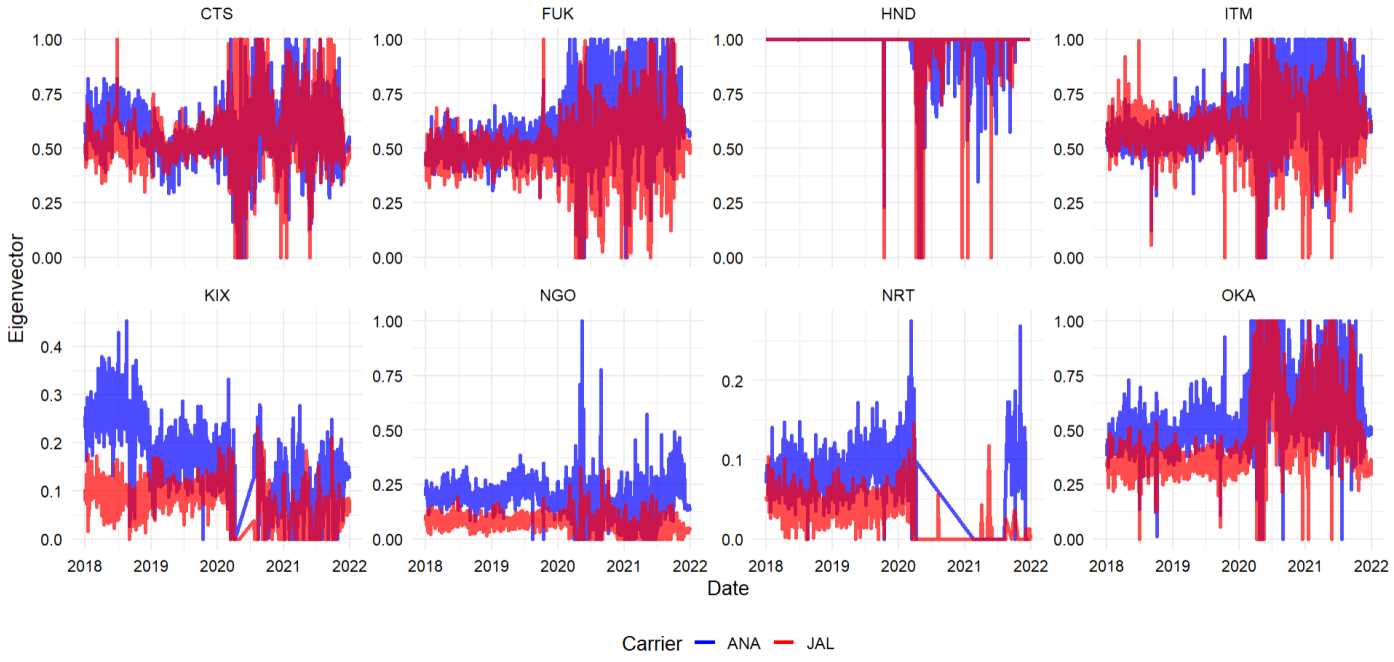
(d)


CTS
FUK
HND
ITM
KIX
NGO
NRT
OKA
Closeness Out
Date
Carrier
ANA
JAL

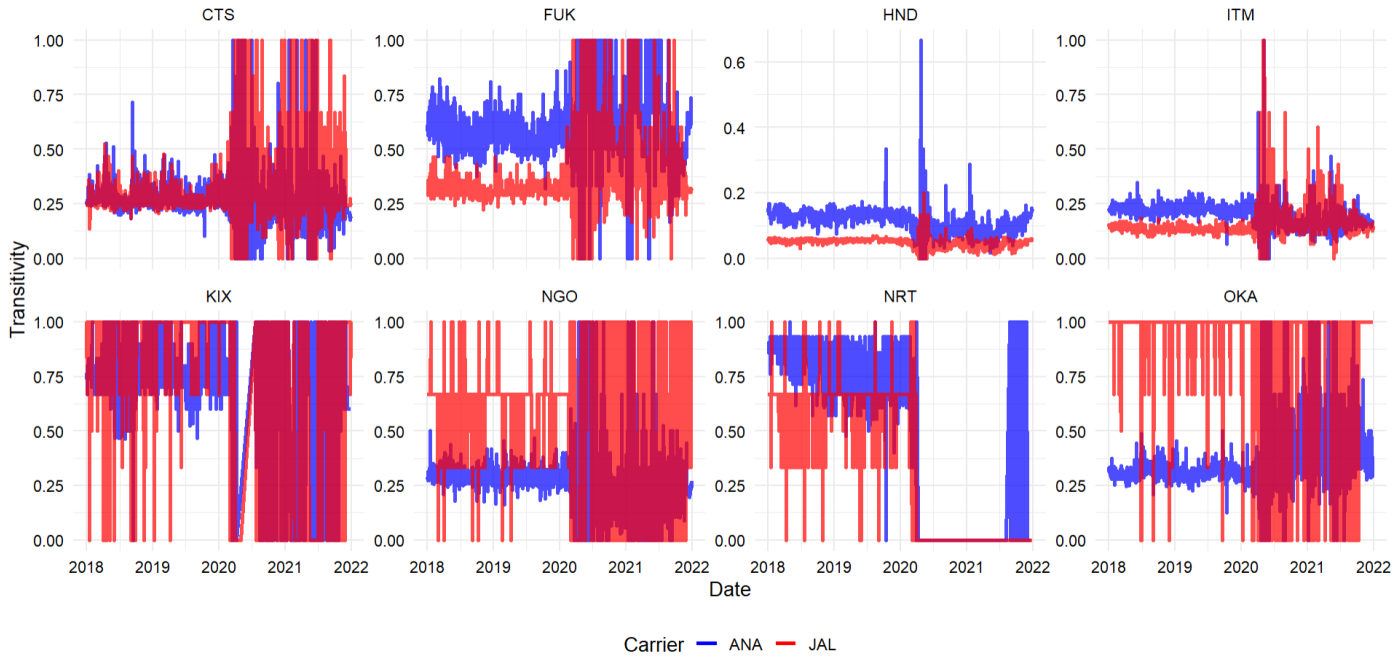
(e)


CTS
FUK
HND
ITM
KIX
NGO
NRT
OKA
Betweenness
Date
Carrier
ANA
JAL

(f)



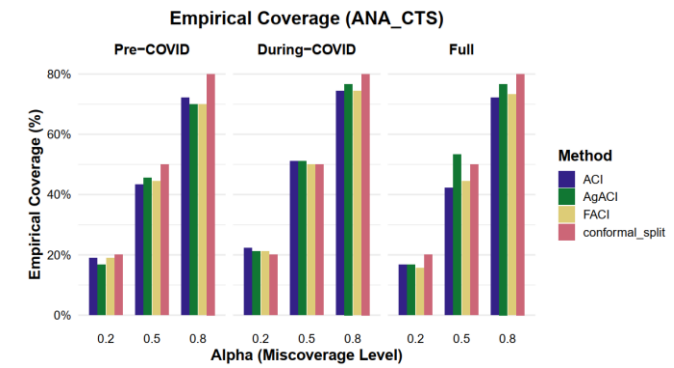
(g)

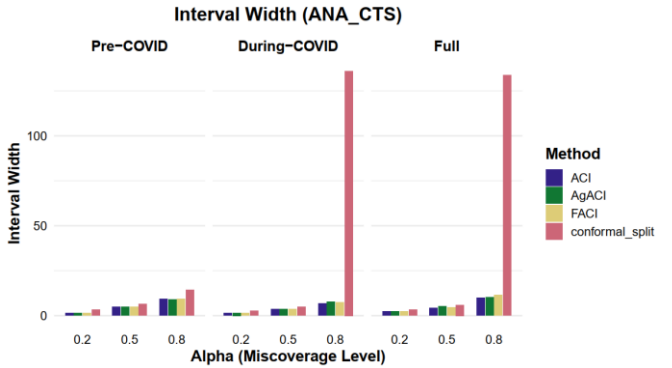


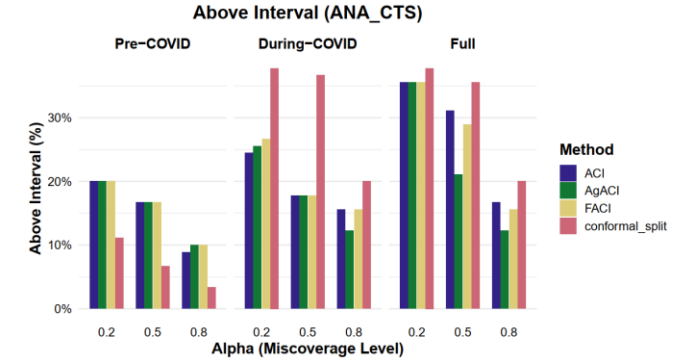
(h)

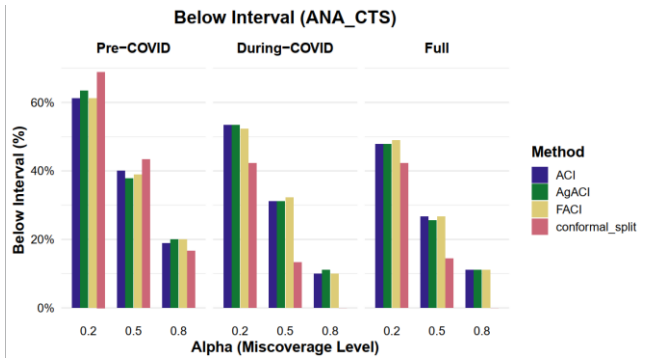
Figure C1: Comparison of time-series plot between ANA and JAL of dependent variable (a) average departure delay, and independent variables (b) in-degree centrality, (c) out-degree centrality, (d) closeness-in centrality, (e) closeness-out centrality, (f) betweenness centrality, (g) eigenvector centrality, and (h) transitivity

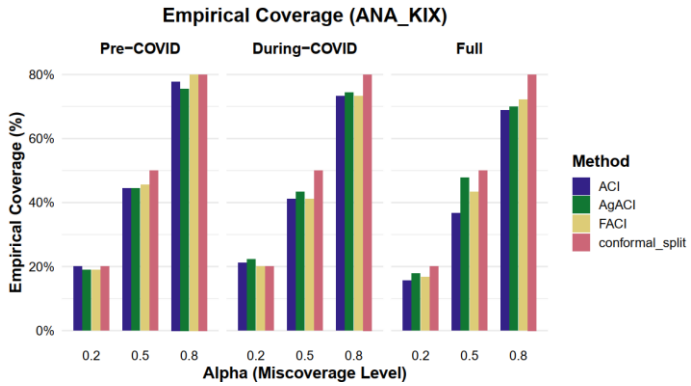
Appendix D: Uncertainty quantification results of conformal split model (base), ACI, AgACI and FACI for average departure delay for rest of the six hub airports (Chapter 4)

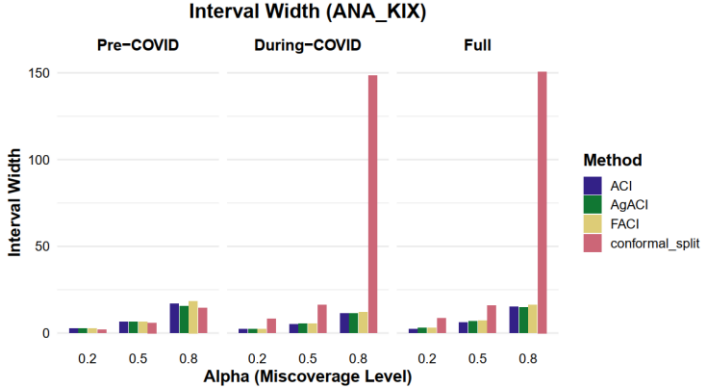


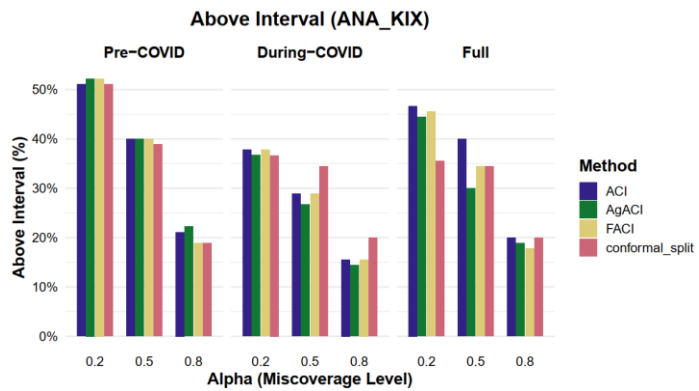


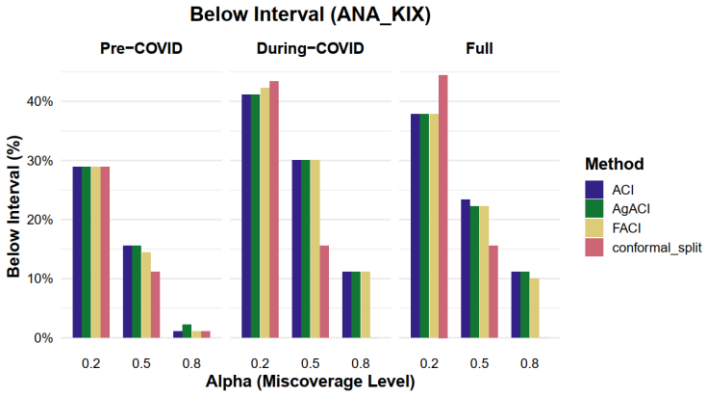


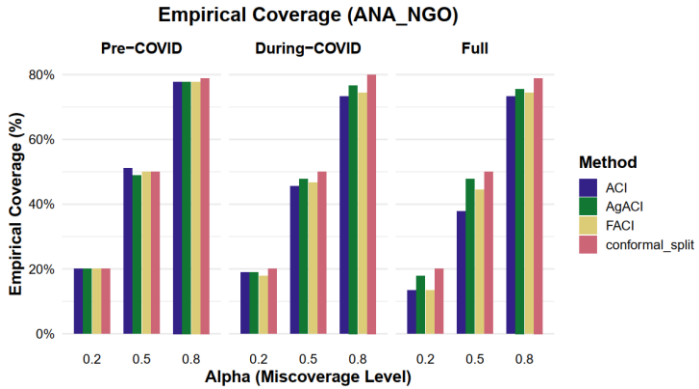


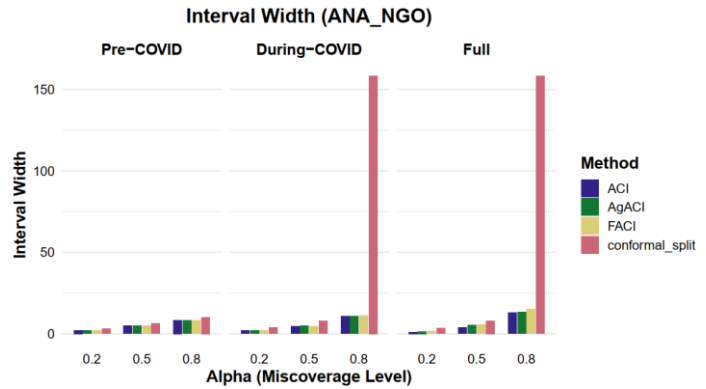


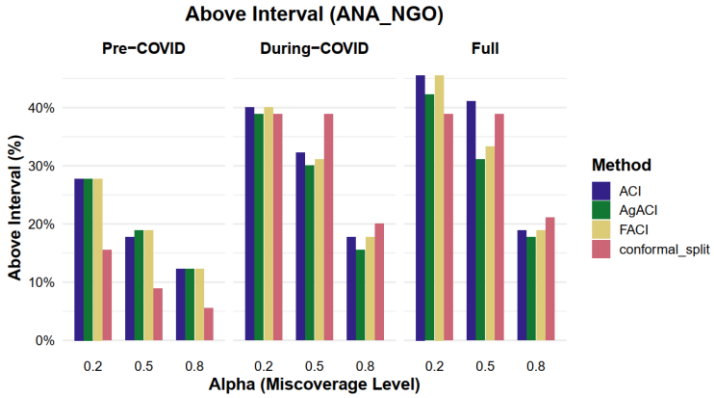


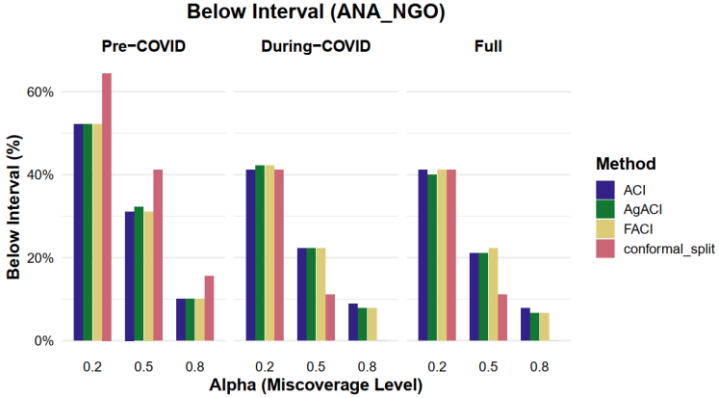

Above Interval (ANA_KIX)
Pre−COVID
During−COVID
Full
Above Interval (%)
50%
40%
30%
20%
10%
0%
0.2
0.5
0.8
Alpha (Miscoverage Level)
Method
ACI
AgACI
FACI
conformal_split

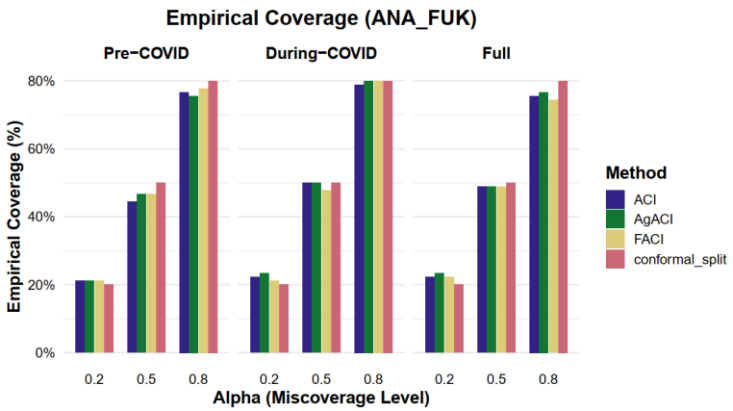

Below Interval (ANA_KIX)
Pre−COVID
During−COVID
Full
Below Interval (%)
40%
30%
20%
10%
0%
0.2
0.5
0.8
Alpha (Miscoverage Level)
Method
ACI
AgACI
FACI
conformal_split

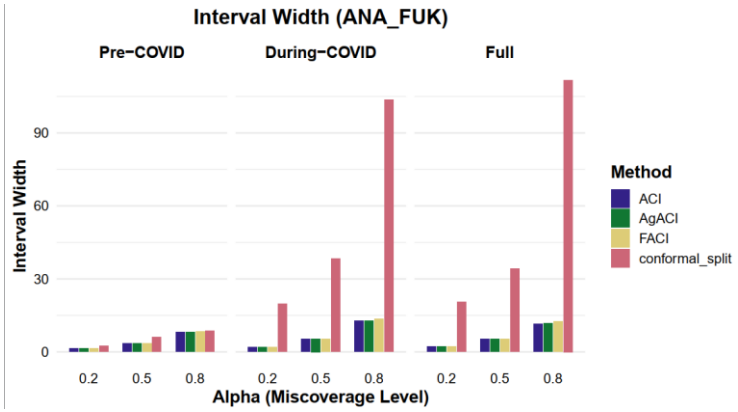

Empirical Coverage (ANA_NGO)
Pre−COVID
During−COVID
Full
Empirical Coverage (%)
80%
60%
40%
20%
0%
0.2
0.5
0.8
Alpha (Miscoverage Level)
Method
ACI
AgACI
FACI
conformal_split

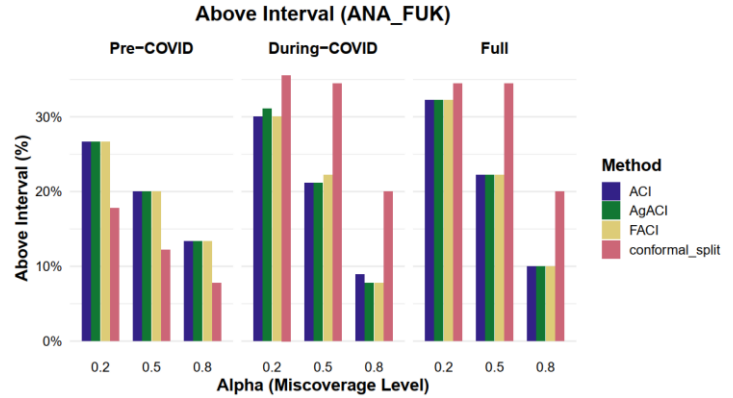

Interval Width (ANA_NGO)
Pre−COVID
During−COVID
Full
Interval Width
150
100
50
0
0.2
0.5
0.8
Alpha (Miscoverage Level)
Method
ACI
AgACI
FACI
conformal_split

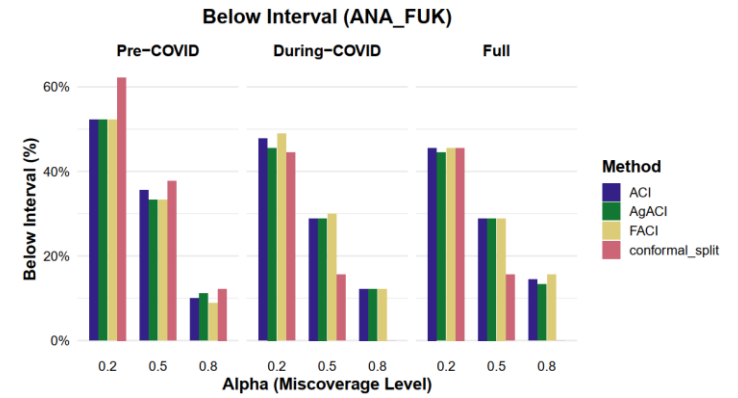

Above Interval (ANA_NGO)
Pre−COVID
During−COVID
Full
Above Interval (%)
40%
30%
20%
10%
0%
0.2
0.5
0.8
Alpha (Miscoverage Level)
Method
ACI
AgACI
FACI
conformal_split

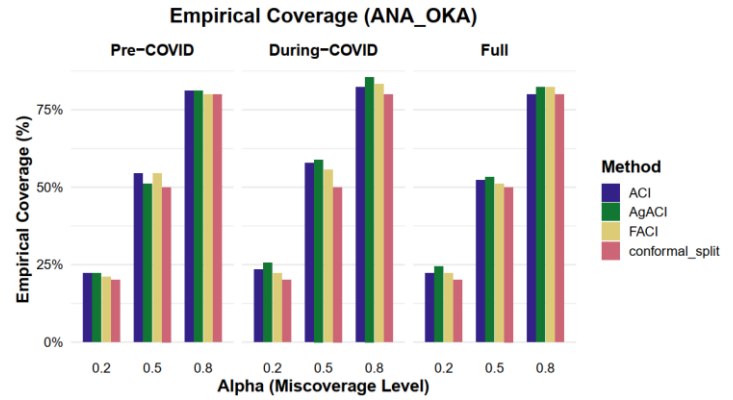

Below Interval (ANA_NGO)
Pre−COVID
During−COVID
Full
Below Interval (%)
60%
40%
20%
0%
0.2
0.5
0.8
Alpha (Miscoverage Level)
Method
ACI
AgACI
FACI
conformal_split

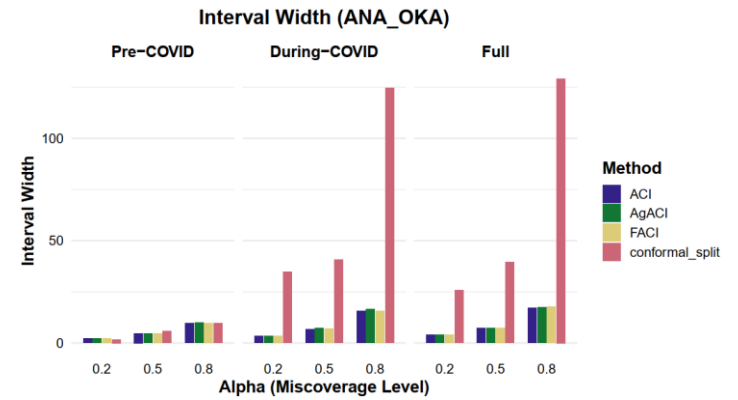

Empirical Coverage (ANA_FUK)
Pre−COVID
During−COVID
Full
Empirical Coverage (%)
80%
60%
40%
20%
0%
0.2
0.5
0.8
Alpha (Miscoverage Level)
Method
ACI
AgACI
FACI
conformal_split


Interval Width (ANA_FUK)
Pre−COVID
During−COVID
Full
Interval Width
90
60
30
0
0.2
0.5
0.8
Alpha (Miscoverage Level)
Method
ACI
AgACI
FACI
conformal_split


Above Interval (ANA_FUK)
Pre−COVID
During−COVID
Full
Above Interval (%)
30%
20%
10%
0%
0.2
0.5
0.8
Alpha (Miscoverage Level)
Method
ACI
AgACI
FACI
conformal_split


Below Interval (ANA_FUK)
Pre−COVID
During−COVID
Full
Below Interval (%)
60%
40%
20%
0%
0.2
0.5
0.8
Alpha (Miscoverage Level)
Method
ACI
AgACI
FACI
conformal_split


Empirical Coverage (ANA_OKA)
Pre−COVID
During−COVID
Full
Empirical Coverage (%)
75%
50%
25%
0%
0.2
0.5
0.8
Alpha (Miscoverage Level)
Method
ACI
AgACI
FACI
conformal_split


Interval Width (ANA_OKA)
Pre−COVID
During−COVID
Full
Interval Width
100
50
0
0.2
0.5
0.8
Alpha (Miscoverage Level)
Method
ACI
AgACI
FACI
conformal_split

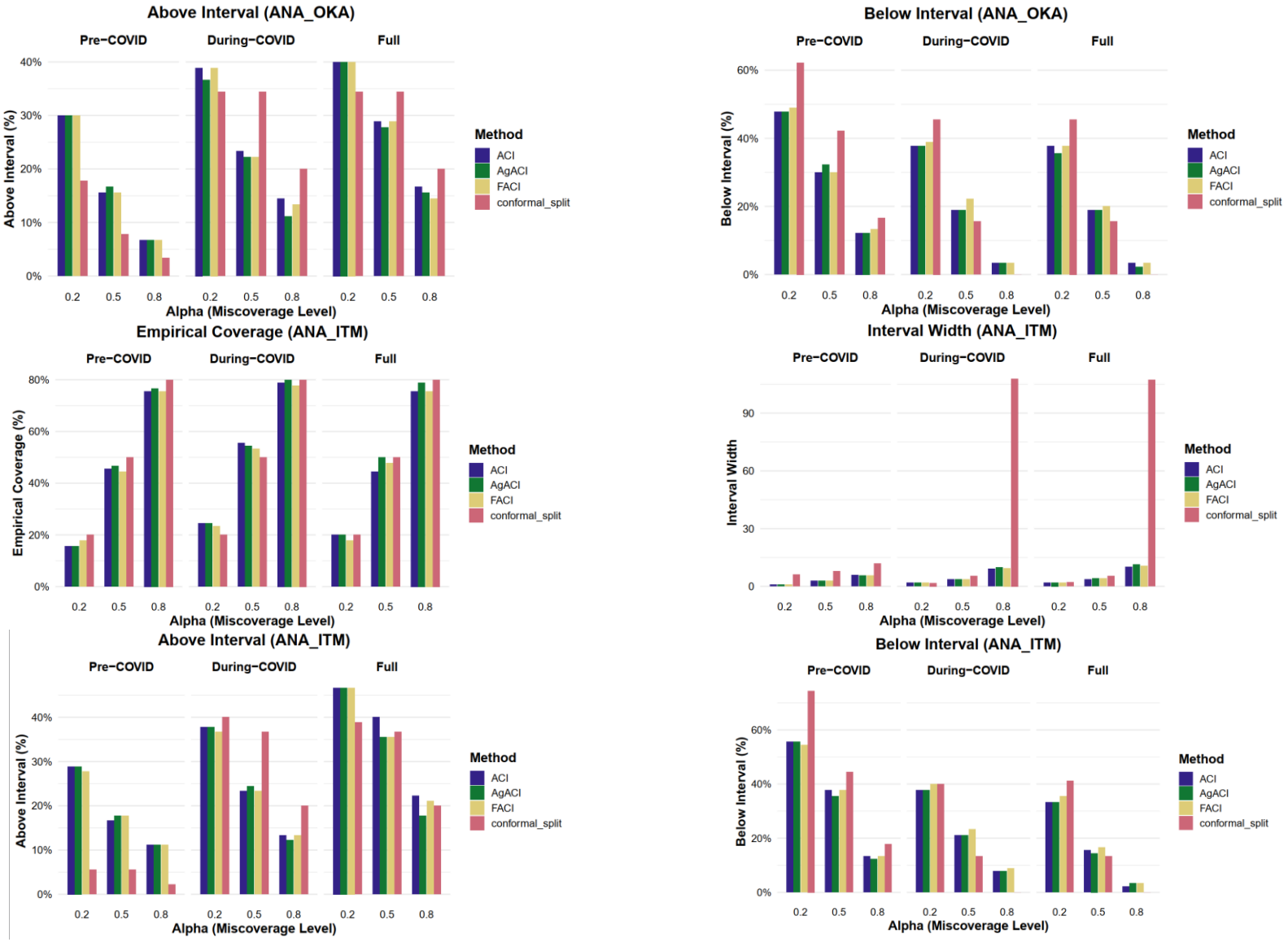
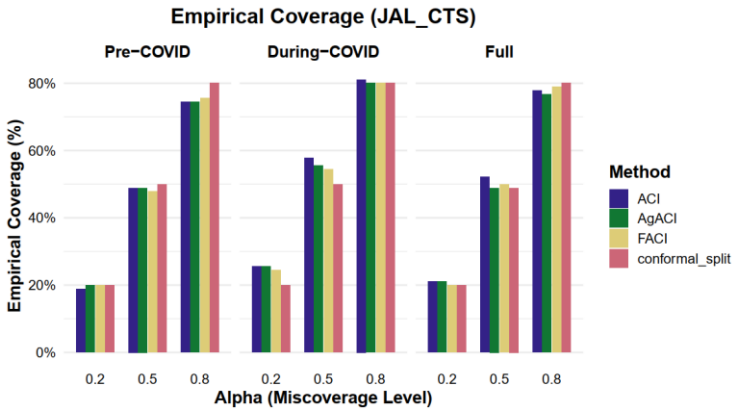
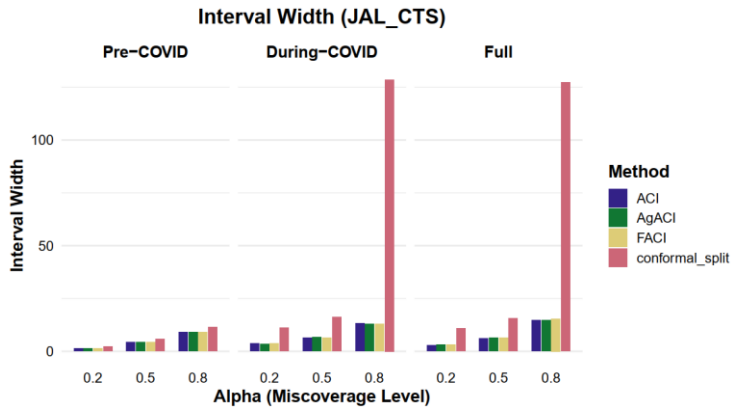
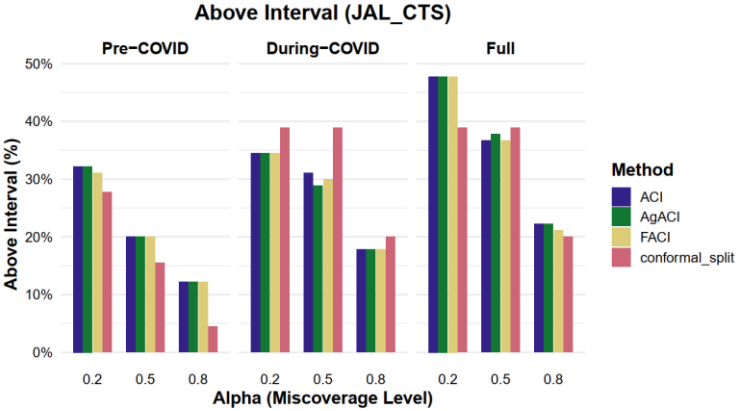
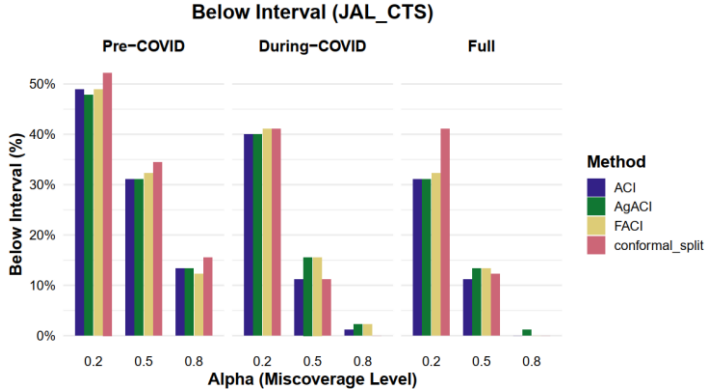


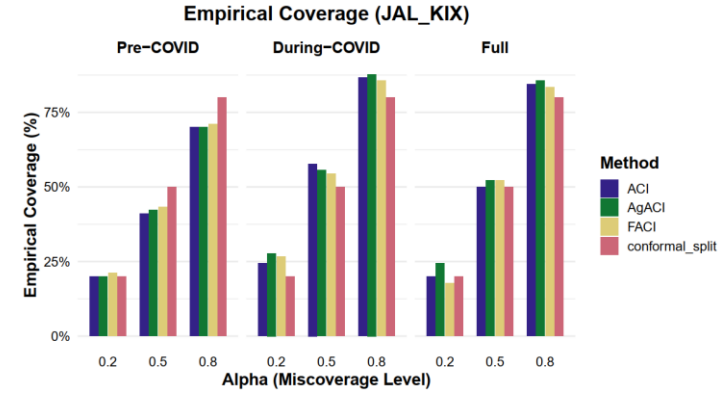
Figure D1: Summary statistics of the three conformal inference frameworks for ANA

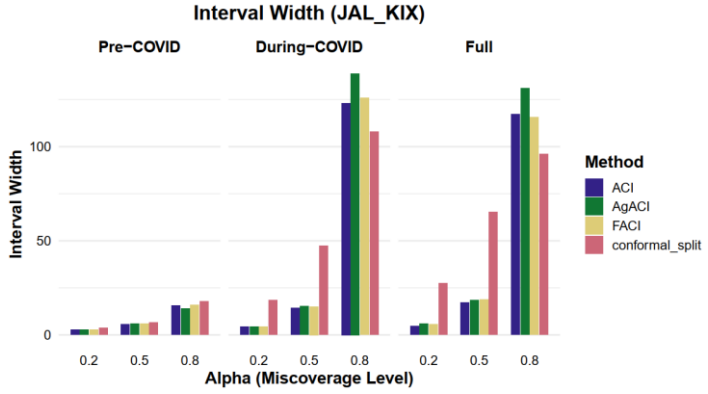

Empirical Coverage (JAL_CTS)
Pre−COVID
During−COVID
Full
Empirical Coverage (%)
80%
60%
40%
20%
0%
0.2
0.5
0.8
Alpha (Miscoverage Level)
Method
ACI
AgACI
FACI
conformal_split

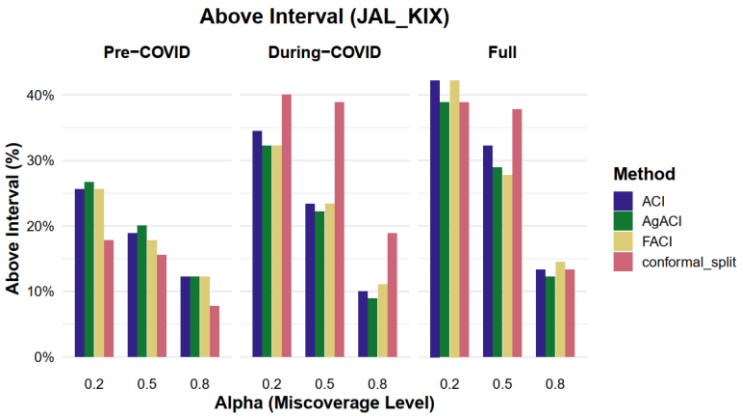

Interval Width (JAL_CTS)
Pre−COVID
During−COVID
Full
Interval Width
100
50
0
0.2
0.5
0.8
Alpha (Miscoverage Level)
Method
ACI
AgACI
FACI
conformal_split

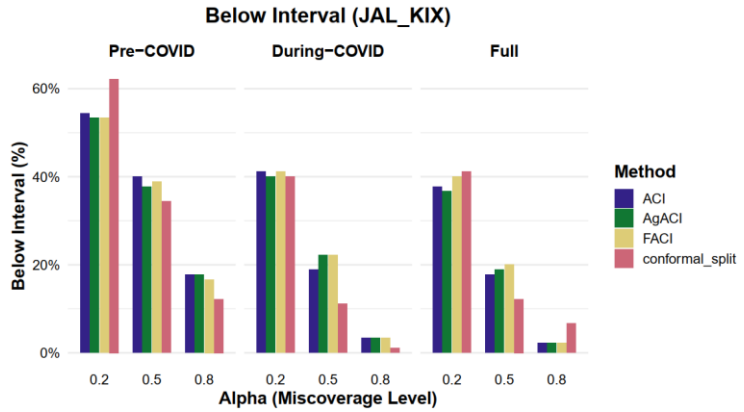

Above Interval (JAL_CTS)
Pre−COVID
During−COVID
Full
Above Interval (%)
50%
40%
30%
20%
10%
0%
0.2
0.5
0.8
Alpha (Miscoverage Level)
Method
ACI
AgACI
FACI
conformal_split

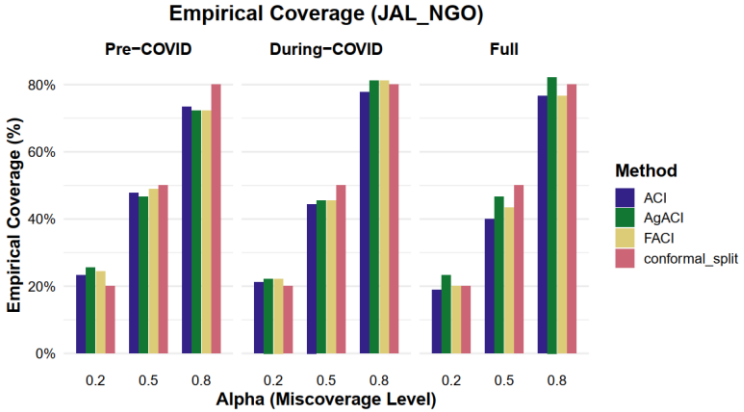

Below Interval (JAL_CTS)
Pre−COVID
During−COVID
Full
Below Interval (%)
50%
40%
30%
20%
10%
0%
0.2
0.5
0.8
Alpha (Miscoverage Level)
Method
ACI
AgACI
FACI
conformal_split

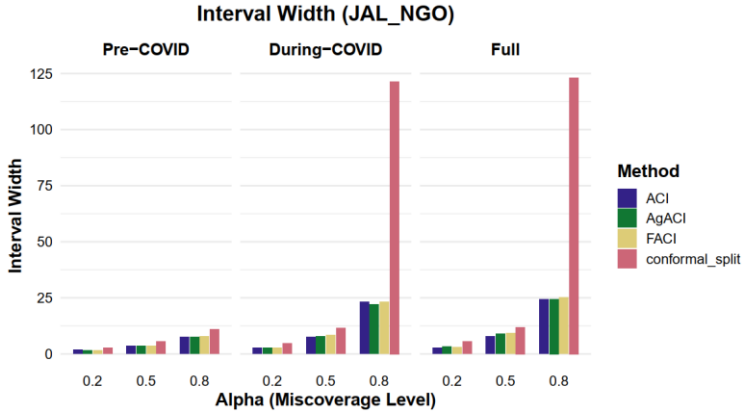

Empirical Coverage (JAL_KIX)
Pre−COVID
During−COVID
Full
Empirical Coverage (%)
75%
50%
25%
0%
0.2
0.5
0.8
Alpha (Miscoverage Level)
Method
ACI
AgACI
FACI
conformal_split

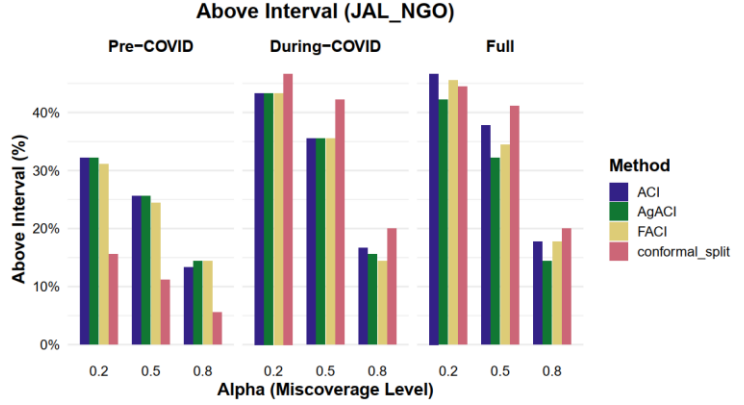

Interval Width (JAL_KIX)
Pre−COVID
During−COVID
Full
Interval Width
100
50
0
0.2
0.5
0.8
Alpha (Miscoverage Level)
Method
ACI
AgACI
FACI
conformal_split

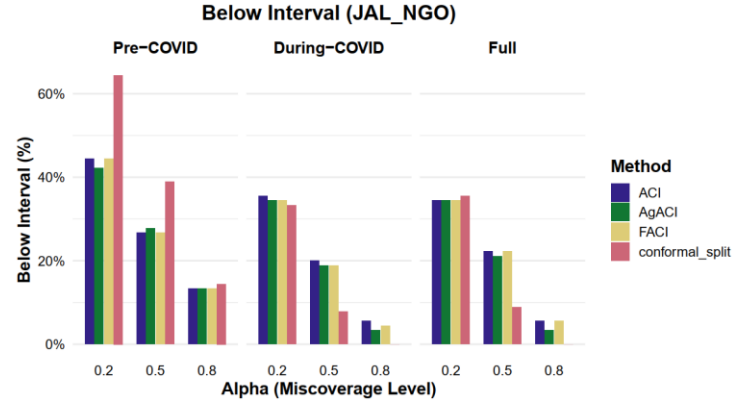

Above Interval (JAL_KIX)
Pre−COVID
During−COVID
Full
Above Interval (%)
40%
30%
20%
10%
0%
0.2
0.5
0.8
Alpha (Miscoverage Level)
Method
ACI
AgACI
FACI
conformal_split

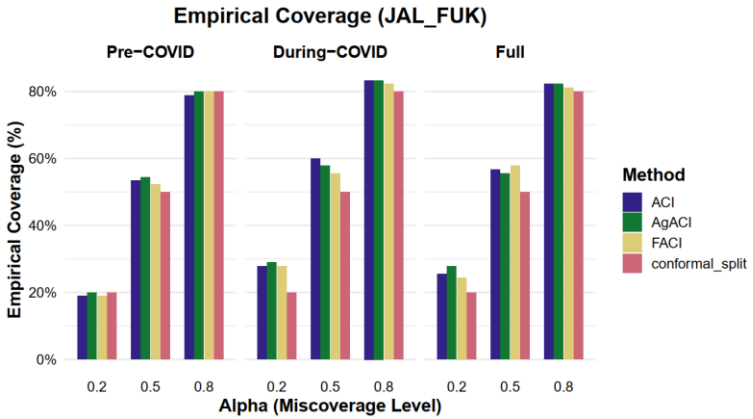

Below Interval (JAL_KIX)
Pre−COVID
During−COVID
Full
Below Interval (%)
60%
40%
20%
0%
0.2
0.5
0.8
Alpha (Miscoverage Level)
Method
ACI
AgACI
FACI
conformal_split

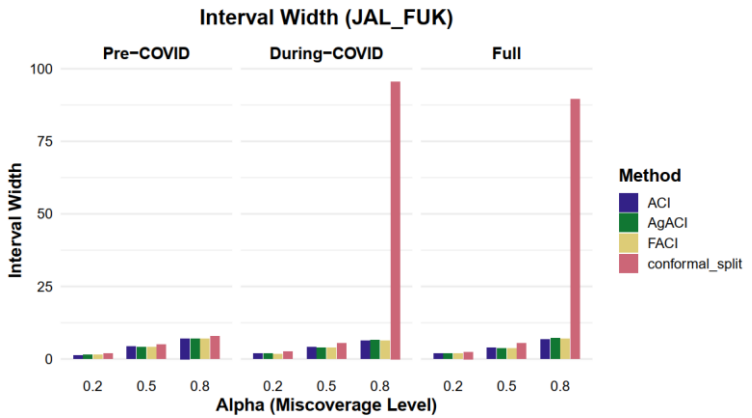

Empirical Coverage (JAL_NGO)
Pre−COVID
During−COVID
Full
Empirical Coverage (%)
80%
60%
40%
20%
0%
0.2
0.5
0.8
Alpha (Miscoverage Level)
Method
ACI
AgACI
FACI
conformal_split

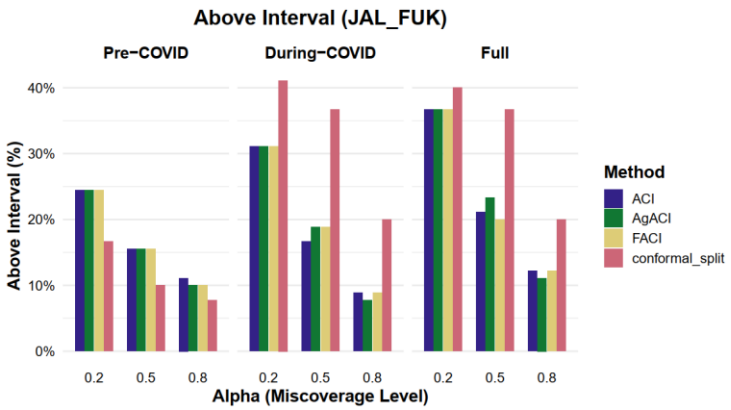

Interval Width (JAL_NGO)
Pre−COVID
During−COVID
Full
Interval Width
125
100
75
50
25
0
0.2
0.5
0.8
Alpha (Miscoverage Level)
Method
ACI
AgACI
FACI
conformal_split

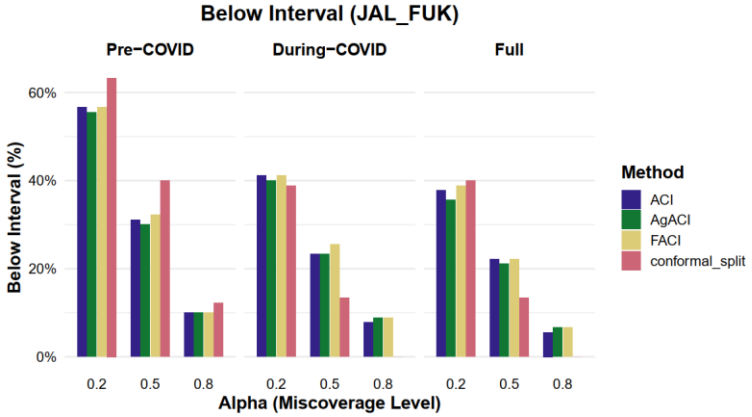

Above Interval (JAL_NGO)
Pre−COVID
During−COVID
Full
Above Interval (%)
40%
30%
20%
10%
0%
0.2
0.5
0.8
Alpha (Miscoverage Level)
Method
ACI
AgACI
FACI
conformal_split

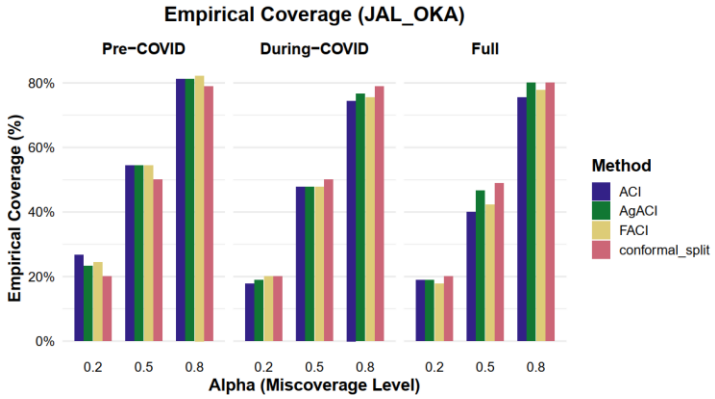

Below Interval (JAL_NGO)
Pre−COVID
During−COVID
Full
Below Interval (%)
60%
40%
20%
0%
0.2
0.5
0.8
Alpha (Miscoverage Level)
Method
ACI
AgACI
FACI
conformal_split

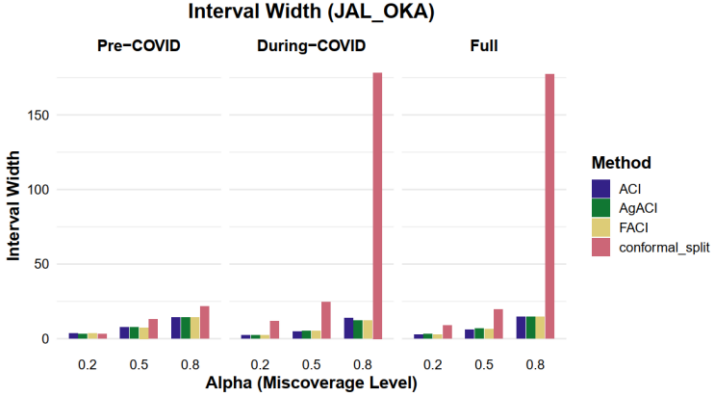

Empirical Coverage (JAL_FUK)
Pre−COVID
During−COVID
Full
Empirical Coverage (%)
80%
60%
40%
20%
0%
0.2
0.5
0.8
Alpha (Miscoverage Level)
Method
ACI
AgACI
FACI
conformal_split


Interval Width (JAL_FUK)
Pre−COVID
During−COVID
Full
Interval Width
100
75
50
25
0
0.2
0.5
0.8
Alpha (Miscoverage Level)
Method
ACI
AgACI
FACI
conformal_split


Above Interval (JAL_FUK)
Pre−COVID
During−COVID
Full
Above Interval (%)
40%
30%
20%
10%
0%
0.2
0.5
0.8
Alpha (Miscoverage Level)
Method
ACI
AgACI
FACI
conformal_split


Below Interval (JAL_FUK)
Pre−COVID
During−COVID
Full
Below Interval (%)
60%
40%
20%
0%
0.2
0.5
0.8
Alpha (Miscoverage Level)
Method
ACI
AgACI
FACI
conformal_split


Empirical Coverage (JAL_OKA)
Pre−COVID
During−COVID
Full
Empirical Coverage (%)
80%
60%
40%
20%
0%
0.2
0.5
0.8
Alpha (Miscoverage Level)
Method
ACI
AgACI
FACI
conformal_split


Interval Width (JAL_OKA)
Pre−COVID
During−COVID
Full
Interval Width
150
100
50
0
0.2
0.5
0.8
Alpha (Miscoverage Level)
Method
ACI
AgACI
FACI
conformal_split

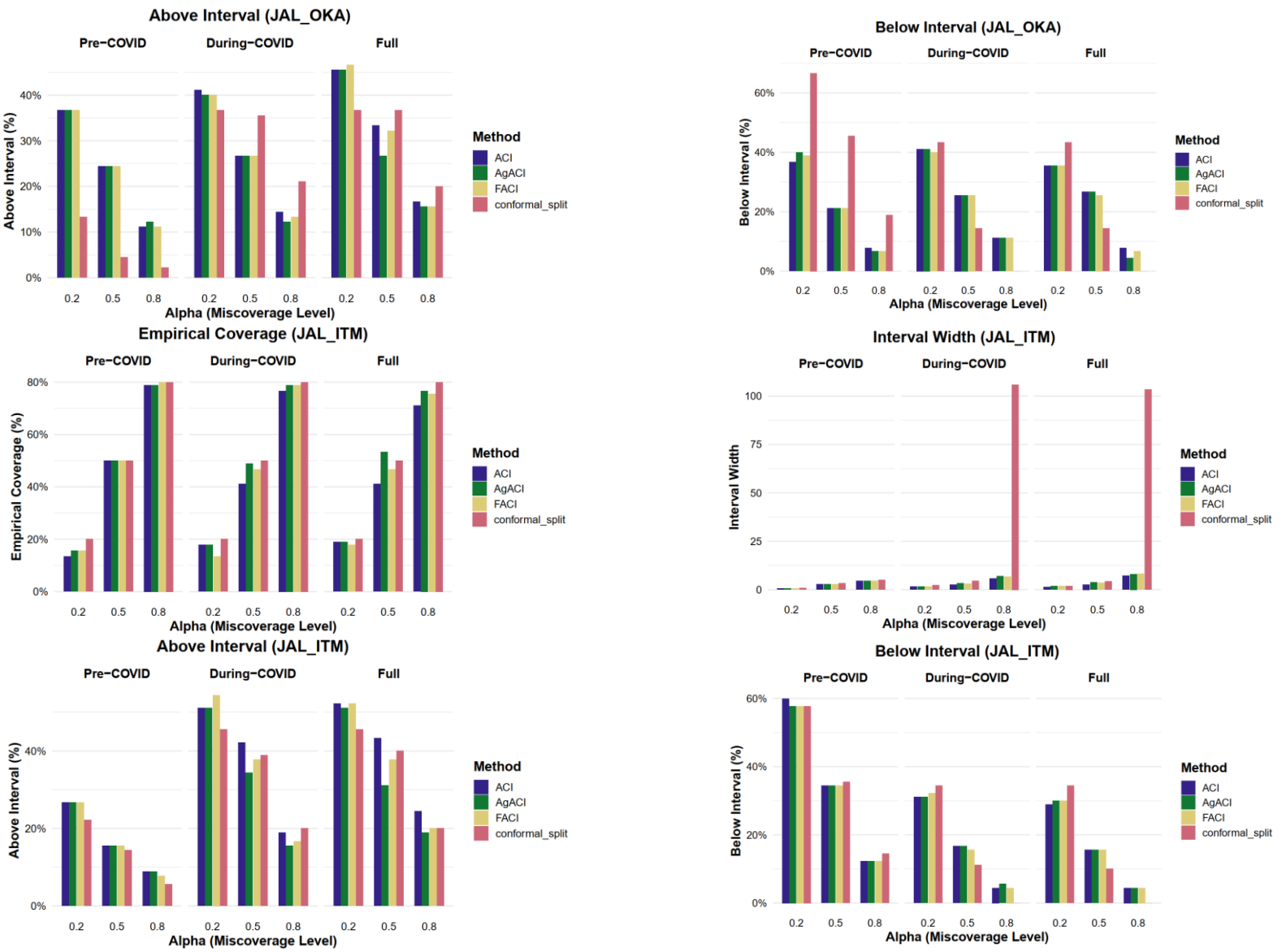


Figure D2: Summary statistics of the three conformal inference frameworks for JAL

Appendix E: Boundary Matrix and Example Calculation of Persistent Homology (Chapter 5)

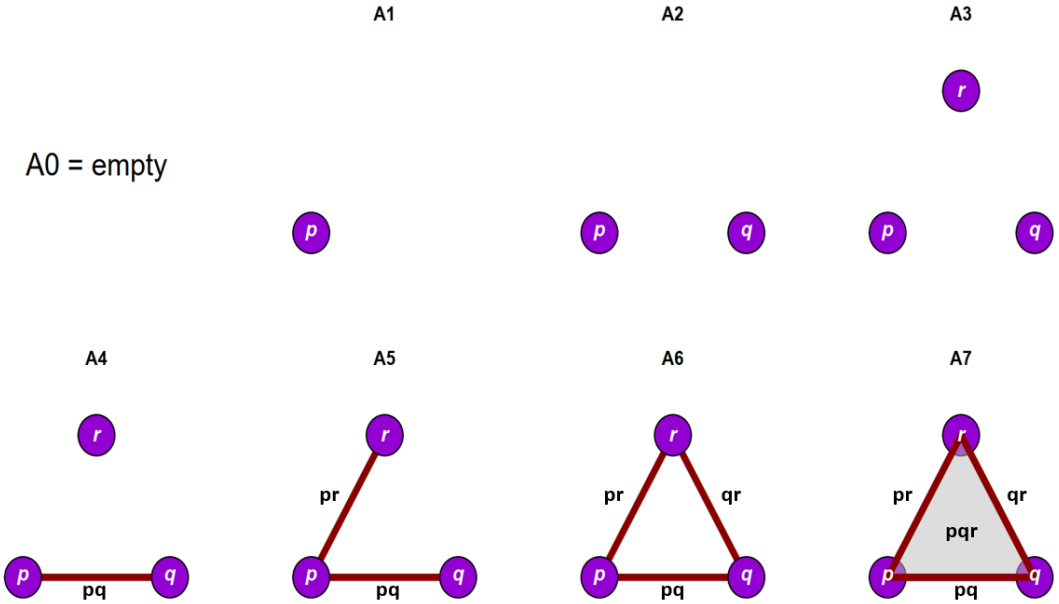


Figure E1: Filtration of a simplicial complex A created by adding one simplex at a time (Adapted example from Elchesen, 2025)

Below matrix represents the Boundary matrix, D of Vietoris-Rips complex.

$$D = \begin{array}{c} \\ p \\ q \\ r \\ pq \\ pr \\ qr \\ pqr \end{array} \begin{bmatrix} 0 & 0 & 0 & 1 & 1 & 0 & 1 \\ 0 & 0 & 0 & 1 & 0 & 1 & 1 \\ 0 & 0 & 0 & 0 & 1 & 1 & 1 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix}$$

D keeps track of which simplices are faces of other simplices. Here, as pqr is a triangle, so, its faces are the three edges pq, pr, qr. So, in the triangle's column, the rows for those edges have a 1. Similarly, I have an edge pq, therefore, so its faces are the points p and q. So, in the edge's column, the rows for p and q have 1. The persistence algorithm takes the growing simplicial complex, encodes it in a boundary matrix, D, reduces that matrix column by column, and extracts birth-death pairs to tell exactly when every topological feature (H_0 and H_1) appeared and disappeared. This example is given in the lecture notes prepared by Elchesen, 2025.

Appendix F: Sample calculation of Proportion of Birth Time Overlap (Chapter 5)

To quantitatively assess the transition from connected components to loop structure formation, we computed the Proportion Birth Time Overlap (PBTO) for each day in the dataset. PBTO is designed to capture the extent to which cyclic delay associations (i.e., H_1 loops in persistent homology) emerge after the underlying delay network becomes structurally unified (i.e., after most H_0 components have merged into a single connected component - 90th percentile). PBTO was calculated as the proportion of H_1 loops whose birth time occurred after network had become 90th percentile of connected. A higher PBTO value indicates that a larger fraction of cyclic delay structures (H_1 loops) emerge only after the entire network is structurally unified suggesting a transition from localized to system-wide (higher-order) delay propagation. Conversely, lower PBTO values imply that delay loops tend to form while the network is still fragmented.

Example: Death time of H_0 Components

Nodes	C1	C2	C3	C4	C5	C6	C7	C8	C9	C10
Death	2.5	4.0	5.5	6.0	6.5	8.0	9.5	10.0	11.5	13.0

If we take 90th percentile death time then the value is 11.5.

Birth time of H_1 loop

Loop	L1	L2	L3	L4	L5	L6
Birth time	9.2	10.8	11.5	12.6	13.8	14.1

PBTO calculation:

Loop	L1	L2	L3	L4	L5	L6
Birth	9.2	10.8	11.5	12.6	13.8	14.1
>11.5	X	X	X	Y	Y	Y

Loops born after the threshold value merge is L4, L5, L6 – total 3. So, PBTO value is 3/6 = 0.5.

It means that 50% of loops were born after the network had become mostly connected. This suggests that half of the loop delay emerged only after delays had spread enough to unify the network structurally. PBTO actually gives a sense of transition from H_0 to H_1. It informs how much of each day's loop activity waits until the network is already connected (after most H_0 components have merged). A high PBTO therefore signals that the appearance of loops is largely a second-phase phenomenon, while a low PBTO means loops are appearing during the fragmentation phase.