Predicting flight delay propagation in the Brazilian aviation network with Heterogeneous Graph Neural Networks

Eduardo Zantut Wittmann[1], Mirela Teixeira Cazzolato[1]

[1]Institute of Mathematics and Computer Science – University of São Paulo (ICMC/USP)
Avenida Trabalhador São-carlense, 400 - Centro CEP: 13566-590 - São Carlos - SP - Brazil

eduardozw@usp.br, mirela@icmc.usp.br

Abstract. *This work takes advantage of a Graph Neural Network (GNN) model to predict flight delay propagation within the Brazilian aviation network. We propose SUBAERO, a method with an architecture that comprises a diverse data ingestion pipeline, integrating weather, airport, flight, and aircraft data via APIs. These datasets are structured in a graph-oriented database (Neo4j) to map complex spatial-temporal dependencies. A Heterogeneous Graph Neural Network (HGT) approach was employed, achieving a Mean Absolute Error (MAE) of 6.7 minutes in delay predictions. This study applies advanced Data Science techniques to optimize urban and regional mobility, providing a timely decision-support tool to mitigate operational bottlenecks and enhance transport infrastructure for social good.*

1. Introduction

The scale of the Commercial Aviation is global, transporting people and cargo worldwide by a huge network of Air Routes and Connection Hubs. According to recent data from the International Air Transport Association (IATA) and the International Civil Aviation Organization (ICAO) [IATA 2025, ICAO 2024], more than 100,000 flights are operated daily. Given this massive scale, the sector nowadays can be characterized mostly by its high dependency between flights, airplanes, and airports. Due to the high optimization of the flight schedules, a single delay in one flight can, sometimes, cause multiple and lasting delays in the flight network [Wu et al. 2024]. This delay and its mitigation can be influenced by multiple factors such as weather, flight schedules, airplane availability and many others.

Correlating and considering the influence of each factor can be a very challenging problem when using usual approaches such as Regression Algorithms, Regular Neural Networks, and other Machine Learning Models, due to the limited capacity of these models to consider the connections between different data points [Rebollo and Balakrishnan 2014, Wu et al. 2024]. For that challenge, an approach such as relating data points using Graphs can be extremely useful and bring significant value to the analysis, as the Graph architecture has the feature of making it easier to correlate seemingly unrelated data points to analyze a behavior. As algorithms evolve, the use of Neural Networks has become more widely used to predict behaviors, and their association with Graphs is a valuable resource, since it combines the great capability of learning by analyzing patterns with the high capacity to associate information in a more appropriate way [Hu et al. 2020, Zanin and Lillo 2013].

Another feature of organizing information into graphs is the native ability to consider different data types within the analysis, making it possible to differentiate aircraft information from weather data in a more natural and clear way. Understanding these distinct kinds of information is essential for the analysis [Hu et al. 2020]. For example, a change in weather impacts a flight schedule differently than an aircraft maintenance delay. Therefore, using Heterogeneous Graph Neural Networks (HGNNs) [Sun and Han 2013] was a logical decision, as this architecture is capable of differentiating connections and determining which are most relevant for the prediction.

To achieve highly accurate predictions, we propose SUBAERO. The methodology integrates the heterogeneous graph architecture with real-time data sources, such as Open Meteo and ANAC (Agência Nacional de Aviação Civil) APIs [1], to comprehensively model the flight network. The results demonstrate the robust predictive capacity of this method, achieving a Mean Absolute Error of 6.7 minutes. This performance highlights the system's potential to provide valuable insights and precise forecasts.

Beyond its technical contribution, this work is motivated by a social challenge specific to the Brazilian context [Aprigliano Fernandes et al. 2021]. As a continental nation with limited high-speed ground transportation, Brazil relies heavily on aviation to connect remote communities to healthcare, employment, and economic opportunity [Gui et al. 2020]. In this setting, cascading flight delays carry an unequal burden: passengers with greater financial resources can more easily absorb disruptions by rebooking flights or arranging alternative transport, while those with tighter budgets face disproportionate consequences.

A reliable predictive tool for delay propagation, therefore, has direct social value, enabling proactive mitigation rather than reactive damage control. Built entirely on openly available data and released as an open tool, SUBAERO allows anyone (*e.g.,* travelers, organizations, and researchers) to access, inspect and build upon it. As an open tool, SUBAERO makes the aerial network more predictable and transparent for those who depend on it.

This paper is organized as follows: Section 2 presents the Theoretical Foundation. Section 3 discusses the related work. Section 4 describes the proposed method SUBAERO and methodology. Section 5 presents and discusses the results obtained from simulating the delay propagation. Finally, Section 6 summarizes the results achieved and provides final considerations.

2. Theoretical Foundation

This section presents the theoretical background related to the SUBAERO methodology. First, the fundamental concepts of Graph Theory and Complex Networks are described, which are important for representing relational aviation data. Next, we discuss the structural characteristics of flight networks, with emphasis on the specific topology and the inherent vulnerability to cascading disruptions. Finally, we motivate the application of Graph Neural Networks to the delay propagation problem, with the adoption of the Heterogeneous Graph Transformer architecture.

[1] ANAC: *https://www.gov.br/anac/pt-br*

2.1. Graph Theory and Complex Networks

Graph theory provides a mathematical foundation to represent complex relational structures in an abstract and efficient manner. Formally, a graph $G = (V, E)$ is defined as a collection of nodes (or vertices), V, and edges (or links), E [Diestel 2017]. The primary strength of a Graph Network is to allow better relationships between data that can correlate with multiple data points simultaneously. This paper uses a Heterogeneous Graph architecture, improving the capability of Graph Networks to correlate distinct data points at the same place as it comprises different types of nodes and edges. This way, it is possible to explicitly represent distinct entities, such as flights, airports, aircraft, and weather conditions, and their unique relational semantics within a single, cohesive network.

2.2. Flight Network Modeling and Structural Limitations

In modern aviation, networks are predominantly organized around hub-and-spoke topologies [Zanin and Lillo 2013, Barabási and Albert 1999]. This configuration utilizes major hub airports to aggregate multiple inbound flights, followed by a concentrated wave of outbound departures within a short timeframe. This strategy allows airlines to provide a vast number of route combinations without scaling the total number of flights proportionally. However, this topology is inherently sensitive to disruptions; a single delay can cascade through the network, preventing passengers from reaching their connections and leading to significant operational inefficiencies and a diminished passenger experience.

Traditional predictive approaches, such as regression and tabular models, inherently ignore the complex sequential dependencies between flights [Rebollo and Balakrishnan 2014, Wu et al. 2024]. They fail to capture the true network structure, often missing the crucial multidimensional interactions between airport operations, aircraft utilization, and localized weather conditions. By utilizing a Heterogeneous Graph architecture, the model natively incorporates these distinct entities and their specific attributes, allowing the algorithm to learn from the exact topology of the flight network.

2.3. Graph Neural Networks for Delay Propagation Prediction

To address the challenge of accurately predicting delays given the multiple relationships between diverse data sources, Graph Neural Networks (GNNs) [Zhou et al. 2020, Kipf and Welling 2017] are applied to the modeled flight network. These models excel at aggregating complex data relationships simultaneously, effectively learning which specific data points or groups have the greatest impact on delay propagation.

This research evaluated multiple GNN architectures, including Graph Attention Networks (GAT) [Velickovic et al. 2018], Temporal Graph Networks (TGN) [Rossi et al. 2020], and Heterogeneous Graph Transformers (HGT) [Hu et al. 2020]. Following preliminary testing on a subset of the dataset, HGT was selected as the primary model. This decision was based on its superior capacity to process heterogeneous networks and its proven ability to capture complex structural patterns, ultimately yielding better predictive performance than the other evaluated GNN architectures.

3. Related Work

The prediction of flight delays and their propagation within aviation networks has been widely studied, traditionally transitioning from isolated statistical models to complex graph-based structures. Existing literature generally splits into approaches that leverage dense operational or meteorological datasets but restrict their scope to isolated flights or single hubs, and studies that model network-wide dynamics but rely on a limited set of data modalities. The methodology proposed in this paper bridges these isolated efforts by combining comprehensive multimodal feature integration with a scalable heterogeneous graph architecture.

In *Flight Delay Prediction Based on Aviation Big Data and Machine Learning* by Gui et al. [Gui et al. 2020], traditional machine learning models and LSTM networks are applied to high-frequency aviation and weather datasets to estimate delay minutes. While data-rich, this methodology analyzes flights independently using tabular and sequential formulations. It restricts its scope to specific routes, ignoring the logistical ripple effects caused by interconnected operations. Our approach directly addresses this limitation by modeling the entire national flight network as a heterogeneous graph, dynamically tracking consecutive aircraft rotations to accurately predict cascading delays rather than treating flights as independent events.

In *A Deep Learning Approach for Flight Delay Prediction through Time-Evolving Graphs* by Cai et al. [Cai et al. 2022], the authors propose a Time-Evolving GCN to capture the spatial-temporal propagation of delays across an airport network. Despite this valid network-centric view, their homogeneous graph relies almost exclusively on operational logs, neglecting external delay triggers like weather conditions and aircraft characteristics. Furthermore, scaling such global graphs often leads to severe memory bottlenecks. We expand upon their foundation by natively integrating multimodal data into a Heterogeneous Graph Transformer (HGT) architecture, utilizing an innovative subgraph partitioning strategy to scale the analysis to millions of flights efficiently.

Finally, in *Predictive Modeling of Flight Delays at an Airport Using Machine Learning Methods* by Hatıpoğlu and Tosun [Hatıpoğlu and Tosun 2024], tree-based models (e.g., LightGBM, XGBoost) are utilized to predict individual arrival delays at a single Turkish airport using local weather and traffic data. While effective locally, this geographically confined, tabular approach is inherently limited, as it cannot foresee cascading delays caused by distant network disruptions. Conversely, our HGT model operates on a national scale, explicitly mapping continuous delay cascades across the Brazilian network and maintaining a stable 6.7-minute MAE even on major regional hubs entirely omitted from the training phase. A comprehensive comparison of these distinct methodological aspects is summarized in Table 1.

4. Methodology

In this work, we propose the SUBAERO[2] method for predicting flight delay propagation in the Brazilian aviation network. This section details the methodology for the data collection, graph modeling, memory-optimized graph construction strategy and the iterative training process using Heterogeneous GNNs.

[2] The complete source code and data pipeline for reproducibility are publicly available at: `https://github.com/zweduardo/SUBAERO`

Table 1. Summary of related work on flight delay prediction.

Work	Data Sources	Scope & Target	Predictive Model
[Gui et al. 2020]	ADS-B, weather, schedules, airport info	Specific routes / Isolated flight delay prediction	Random Forest, LSTM
[Cai et al. 2022]	Flight schedules and historical operations	Airport network / Spatial-temporal delay propagation	Time-Evolving GCN
[Hatıpoğlu and Tosun 2024]	Single-airport operations, local weather	Local hub / Individual arrival delays	LightGBM, XGBoost
SUBAERO (Proposed Method)	Diverse data sources: Flight (ANAC), weather, aircraft rotations, airport topology	National network (129 airports) / Aircraft-driven delay propagation	Heterogeneous Graph Transformer (HGT)

4.1. Data Collection and Feature Engineering

The data used refers to the period from January 2025 to March 2026, representing approximately 1.2 million flights across 129 Brazilian airports. The data was ingested from multiple APIs and then integrated into a single database. The considered sources are:

- **ANAC VRA API[3]:** A public REST API provided by the Brazilian National Civil Aviation Agency (ANAC). It was utilized to retrieve scheduled and actual flight records such as departure and arrival times, continuous delay metrics, aircraft registration, and aerodrome codes.
- **Open-Meteo Archive[4]:** An open-source REST API providing ERA5-based hourly historical weather data. Meteorological conditions, such as temperature, wind speed, precipitation, and cloud cover, were extracted for both the flight's origin and destination airports.
- **OpenSky Network API[5]:** A crowdsourced ADS-B platform used to retrieve information regarding ICAO24 transponder codes and aircraft metadata to assess aircraft age.
- **OurAirports Database[6]:** An open-source global dataset used to enrich the network with structural and geographical airport features, specifically mapping the latitude, longitude, and elevation of each node.

Relevant features, such as arrival delay, were limited to minimize the outliers' influence and also normalized using standard z-score techniques.

4.2. Graph Modeling

Figure 1 shows the graph modeling. In the context of this study, the aviation system is modeled as a Heterogeneous Graph where nodes represent distinct entities: `FLIGHT` , `AIRPORT` , `AIRCRAFT` , and `WEATHER CONDITION` . The edges represent the logistical and operational relationships connecting them, specifically mapped

[3]ANAC VRA API: *https://dados.gov.br/dados/conjuntos-dados/dadosabertos-areas-de-atuacao-voos-e-operacoes-aereas-voo-regular-ativo-vra*

[4]Open Mateo - Free Weather API: *https://open-meteo.com/*

[5]OpenSky Network API: *https://opensky-network.org/data/api*

[6]OurAirports Database: *https://ourairports.com/*

as ORIGIN , DESTINATION , NEXT_ROTATION , and HAS_ORIGIN_WEATHER . Each node has specific properties, presented inside the node boxes. Specifically, flight_departure_delay is the target property that our method SUBAERO aims to predict.

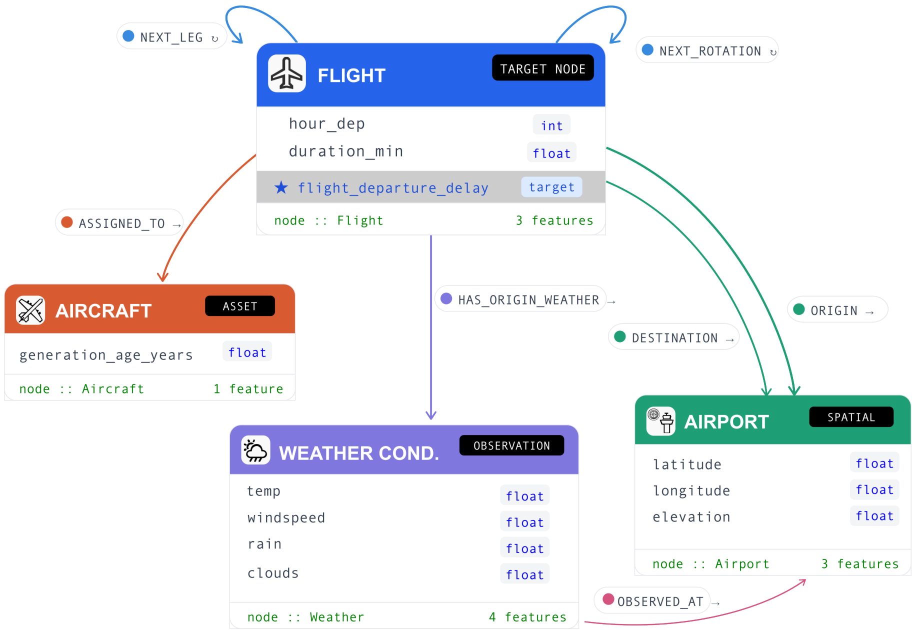

Figure 1. SUBAERO's graph modeling of the flight network, with nodes, edges, and node properties (features) employed.

To better illustrate the proposed graph modeling, 2 presents a real-world data instance extracted from the database, demonstrating how a specific flight and its corresponding entities are interconnected.

4.3. Memory-Optimized Graph Construction

During Training a Graph Neural Network usually maintains all the data allocated in the RAM memory of the Device, but for large Datasets, this degrades the performance or even freezes the Training Iteration. As a form of addressing this issue, a subgraph sampling strategy was developed; this way, the RAM memory only needs to hold some part of the data at a time, resulting in lower RAM usage during training.

The approach used was to split the data into combinations of airports and months, iterating over it for each epoch. Even tough the Data was split, the approach maintained data such as Next aircraft flight, weather nodes, and Other Flight, Airport and Aircraft data, ensuring the Training Data contains the most relationships between data points possible, allowing the model to learn with the data.

4.4. Predictive Modeling with HGT

The Heterogeneous Graph Transformer architecture was selected due to its capabilities of learning given distinct types of nodes and edges, the training of the model was made with the approach detailed previously, and after 5 epochs, the Mean Absolute Error (MAE) converged. To validate the model and affirm the quality of the training phase, cross-validation was performed with hold-out data from airports unseen at the training stage.

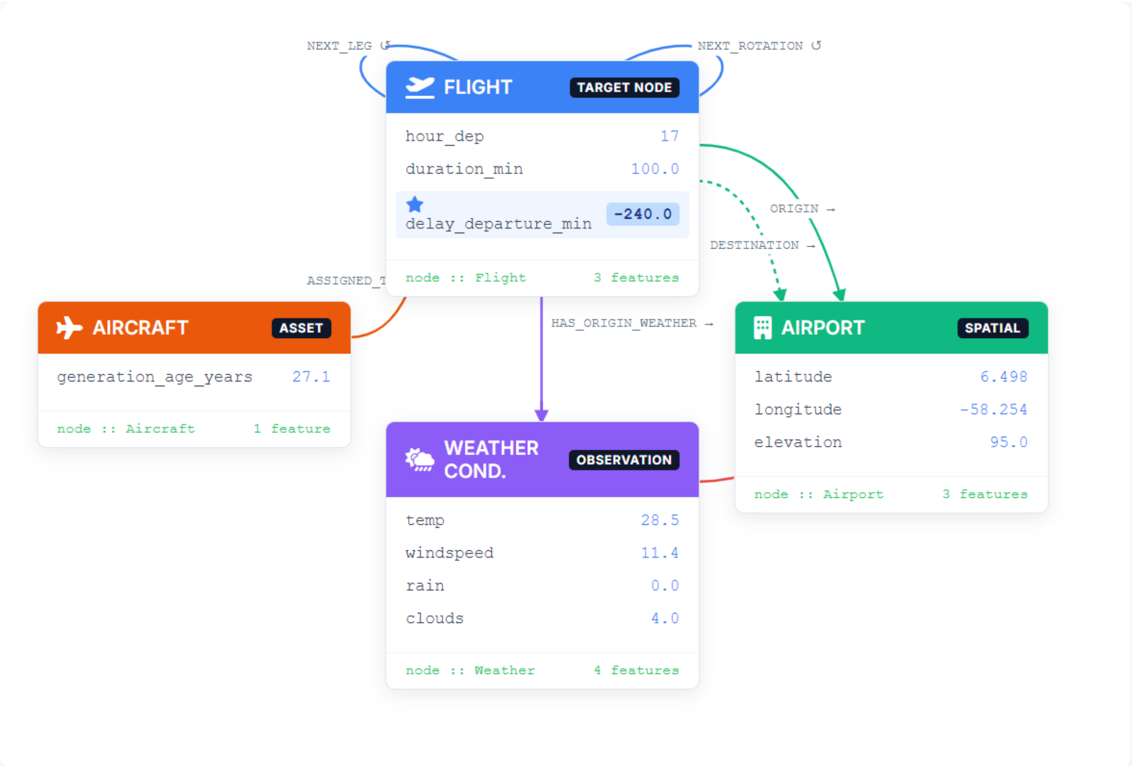

Figure 2. SUBAERO's real-world instance of the heterogeneous graph, illustrating the node properties and relationships for flight.

The model was trained using the Adam optimizer with a learning rate of 1×10^{-3} and a weight decay of 1×10^{-4}. The architecture used consists of two HGT layers with a hidden dimension of 64, 4 attention heads, and a dropout rate of 0.3. For holdout validation, 5 airports were excluded from the training set to check for overfitting afterward.

5. Results and Discussion

Crucially, the performance of the Heterogeneous Graph Transformer (HGT) model was evaluated entirely on real-world historical datasets. Utilizing actual flight operations and meteorological data from open APIs rather than simulated scenarios, the objective was to predict the continuous delay time in minutes for flight arrivals. The model's accuracy in these operational conditions was measured using the Mean Absolute Error (MAE) metric, which assesses the average magnitude of prediction errors.

5.1. Model Performance and Scalability

As described in the methodology, training a global graph with all available data initially resulted in severe memory bottlenecks. By implementing the memory-optimized subgraph construction (iterating over airport-month combinations while preserving *NEXT_ROTATION* edges), the model was successfully scaled to encompass 129 medium and large Brazilian airports, totaling approximately 1.17 million flights.

For this comprehensive dataset, the HGT architecture achieved a highly competitive MAE of 6.7 minutes. This result demonstrates that the model is not only capable of processing large-scale networks with multiple data sources but also adapts at capturing the fine-grained nuances of aircraft-driven delay propagation.

5.2. Generalization and Overfitting Validation

Due to an expressive number of flights with different data sources, it is a big concern that the occurrence of overfitting in the model training. To test for generalization capabilities, an out-of-sample validation was conducted, leaving 5 airports out of the training, one for each Brazil Subregion, and then measuring the Mean Absolute Error (MAE) for each airport. The results can be seen below:

- **GYN (Goiânia):** 5.9 minutes
- **MAO (Manaus):** 6.5 minutes
- **FOR (Fortaleza):** 6.8 minutes
- **POA (Porto Alegre):** 7.6 minutes
- **VCP (Campinas):** 6.8 minutes

For reference, the other evaluated architectures (GAT and TGN) achieved a Mean Absolute Error of 24.76 and 70.02 minutes, respectively, using the same training configuration. The HGT approach reached 6.56 minutes, while the final model, trained with aircraft rotations and validated using the holdout strategy, achieved 6.7 minutes. This result highlights the value of the heterogeneous structure for this task. Taken together with the generalization results, this confirms that the HGT successfully learned structural patterns from the network rather than overfitting to the training data.

The aggregate MAE for airports explicitly seen during training was 6.9 minutes, compared to 6.7 minutes for the entirely unseen airports. This proximity in error margins strongly indicates the absence of overfitting. It confirms that the HGT successfully learned the underlying structural and temporal patterns of delay propagation—such as how weather and consecutive aircraft operations interact—rather than memorizing the behavior of specific geographical nodes.

To further validate the model's generalization capability and rule out overfitting, the learning trajectory was analyzed throughout the training phase. Figure 3 shows that both the training and validation error metrics decrease consistently and converge together over the 50 epochs. The absence of divergence between these curves is a strong indicator that the model could successfully generalize for unseen data and should not be over-fitted.

6. Conclusion

This work presents SUBAERO, an Approach that exploits graphs and graph neural networks to predict, with certain accuracy, the delay propagation of flights on the Brazilian aviation network. By using graph neural networks, it was possible to account for relationships between distinct data sources in order to improve the GNN understanding of the situation, and by using a Heterogeneous Graph Transformer (HGT) architecture, it was possible to strengthen the model's understanding of this diverse data information.

In order to deal with issues of memory limitations when dealing with huge graphs, the dynamic subgraph partitioning strategy was essential not only for the model execution, but also for improving the model results in a significant way. Allowing the system to escalate to almost 130 airports with over a million flights without losing the critical temporal links that dictate delay cascades, achieving a Mean Absolute Error of 6.7 minutes, with strong evidence of a great generalization of the model.

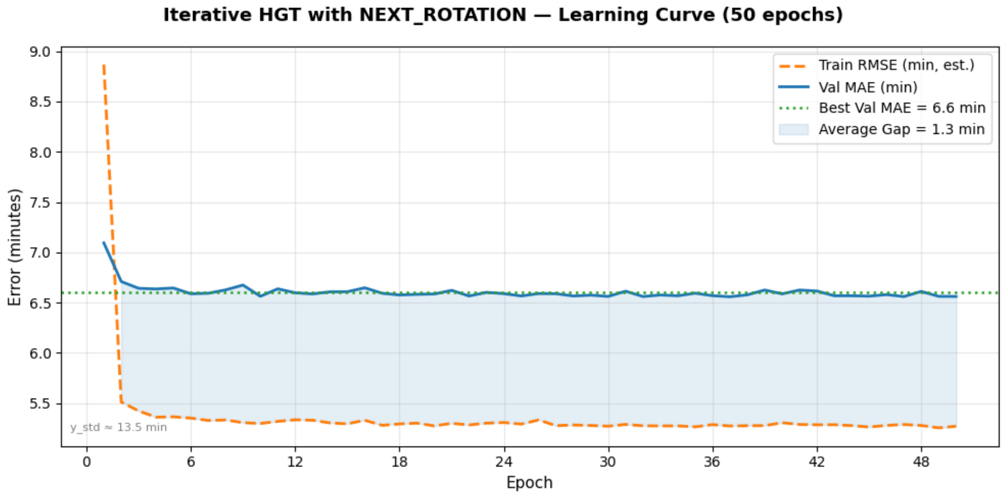

Figure 3. Learning curve over 50 epochs showing the consistent convergence of training and validation metrics.

From a Data Science for Social Good perspective, the practical relevance of this result extends beyond operational efficiency. A prediction accuracy of 6.1 minutes MAE at the national scale is sufficient to support actionable interventions: airlines can consider aircraft reallocation before a delay cascade propagates; ground crews can be repositioned proactively; and passengers can receive early alerts with enough lead time to make alternative arrangements. This is particularly meaningful in the Brazilian context, where aviation connects communities that lack viable ground transport alternatives, and where the social cost of disruption is distributed unequally across the population.

Future work may explore the integration of socioeconomic indicators, such as the Human Development Index (HDI) of municipalities served by each route, to assess the equity dimension of delay propagation more rigorously. Additional directions include expanding the graph structure with critical operational nodes—such as crew scheduling constraints and passenger connection flows to better capture unpredictable operational vagaries. Furthermore, we aim to integrate real-time streaming data pipelines to reduce prediction latency and evaluate the model's transferability to international aviation networks. The source code and data pipeline are publicly available at `https://github.com/zweduardo/SUBAERO` to support reproducibility.

Acknowledgements

This work was supported by the São Paulo Research Foundation (FAPESP – grants #2026/11130-2, #2024/15430-5, #2024/13328-9), the Coordination for Higher Education Personnel Improvement (CAPES – Finance Code 001), the National Research Council (CNPq – grants #401496/2025-2 and #402987/2025-0), and the University of São Paulo (PRPI 1032 - grant #207).

Generative AI usage statement

Grammarly and Gemini tools were used exclusively to support the linguistic revision of the manuscript. The authors take full responsibility for the content presented herein.

References

Aprigliano Fernandes, V., Pacheco, R. R., Fernandes, E., Cabo, M., Ventura, R. V., and Caixeta, R. (2021). Air transportation, economy and causality: Remote towns in brazil's amazon region. *Sustainability*, 13(2). DOI: 10.3390/su13020627.

Barabási, A.-L. and Albert, R. (1999). Emergence of scaling in random networks. *Science*, 286(5439):509–512.

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. DOI: 10.1109/TITS.2021.3103502.

Diestel, R. (2017). *Graph Theory*. Springer Berlin Heidelberg. DOI: 10.1007/978-3-662-53622-3.

Gui, G., Liu, F., Sun, J., Yang, J., Zhou, Z., and Zhao, D. (2020). Flight delay prediction based on aviation big data and machine learning. *IEEE Transactions on Vehicular Technology*, 69(1):140–150. DOI: 10.1109/TVT.2019.2954094.

Hatıpoğlu, I. and Tosun, Ö. (2024). Predictive modeling of flight delays at an airport using machine learning methods. *Applied Sciences*, 14(13):5472. DOI: 10.3390/app14135472.

Hu, Z., Dong, Y., Wang, K., and Sun, Y. (2020). Heterogeneous graph transformer. In *Proceedings of The Web Conference 2020*, WWW '20, page 2704–2710, New York, NY, USA. Association for Computing Machinery. DOI: 10.1145/3366423.3380027.

IATA (2025). IATA Launches 2024 World Air Transport Statistics Report — iata.org. URL: https://www.iata.org/en/pressroom/2025-releases/2025-08-04-01/ [Accessed 23-05-2026].

ICAO (2024). Annual Report 2024 - The World of Air Transport in 2024 - icao.int. URL: https://www.icao.int/about-icao/AnnualReport2024/world-air-transport-2024 [Accessed 23-05-2026].

Kipf, T. N. and Welling, M. (2017). Semi-supervised classification with graph convolutional networks. In *5th International Conference on Learning Representations, ICLR 2017, Toulon, France, April 24-26, 2017, Conference Track Proceedings*.

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

Rossi, E., Chamberlain, B., Frasca, F., Eynard, D., Monti, F., and Bronstein, M. M. (2020). Temporal graph networks for deep learning on dynamic graphs. *CoRR*, abs/2006.10637. URL: https://arxiv.org/abs/2006.10637.

Sun, Y. and Han, J. (2013). Mining heterogeneous information networks: a structural analysis approach. *SIGKDD Explor. Newsl.*, 14(2):20–28. DOI: 10.1145/2481244.2481248.

Velickovic, P., Cucurull, G., Casanova, A., Romero, A., Liò, P., and Bengio, Y. (2018). Graph attention networks. In *6th International Conference on Learning Representations, ICLR 2018, Vancouver, BC, Canada, April 30 - May 3, 2018, Conference Track Proceedings*. URL: https://openreview.net/forum?id=rJXMpikCZ.

Wu, Y., Yang, H., Lin, Y., and Liu, H. (2024). Spatiotemporal Propagation Learning for Network-Wide Flight Delay Prediction. *IEEE Transactions on Knowledge & Data Engineering*, 36(01):386–400. DOI: 10.1109/TKDE.2023.3286690.

Zanin, M. and Lillo, F. (2013). Modelling the air transport with complex networks: A short review. *The European Physical Journal Special Topics*, 215(1):5–21. DOI: 10.1140/epjst/e2013-01711-9.

Zhou, J., Cui, G., Hu, S., Zhang, Z., Yang, C., Liu, Z., Wang, L., Li, C., and Sun, M. (2020). Graph neural networks: A review of methods and applications. *AI Open*, 1:57–81. DOI: 10.1016/j.aiopen.2021.01.001.