applied sciences

MDPI

Article

A Multi-Relational Graph Ranking Framework for Identifying Potential Short-Haul Air-Service Links in County-Level Transport Networks

Yu Wang [1], Xisheng Li [1,*], Jiannan Chi [1,2,*], Xin Jiang [1], Yixu Wang [1] and Jiahui Liu [1,2]

[1] Beijing Engineering Research Center of Industrial Spectrum Imaging, School of Automation and Electrical Engineering, University of Science and Technology Beijing, Beijing 100083, China; d202110332@xs.ustb.edu.cn (Y.W.); w_yixu999@163.com (Y.W.)

[2] Shunde Innovation School, University of Science and Technology Beijing, Foshan 528399, China

* Correspondence: lxs@ustb.edu.cn (X.L.); ustbjnc@ustb.edu.cn (J.C.)

Abstract

The development of low-altitude passenger transport and emerging short-haul air services has made inter-county short-distance routes an important subject for regional aviation market research and preliminary planning. However, piloted eVTOL and low-altitude short-haul passenger services remain at an early stage, lacking continuous and stable historical operational data for supervised modelling. Moreover, unobserved county-level routes cannot be simply regarded as having no market potential. This study formulates county-level candidate route identification as a large-scale priority ranking problem under sparse proxy labels and develops a multi-relational graph ranking framework. The framework integrates county attributes, mobility, transport impedance, high-speed rail substitution conditions, and estimated air travel time variables through multiple relational graphs, and employs an MR-GAT ranking model to learn relative priorities among candidate routes using a pairwise ranking objective. Observed civil aviation route labels and flight frequencies are used as primary supervision signals, while airport catchment weak labels are introduced for auxiliary analysis. Experiments conducted over 240,350 county-level candidate ODs show that, across ten random seeds, the MR-GAT model achieves a mean internal Top-1000 recall of 16.33%, compared with 7.55% for XGBoost and 2.00% for the gravity-based heuristic. Paired bootstrap analysis further supports the higher internal top-ranked retrieval performance of MR-GAT over XGBoost. Ten-seed relational ablation shows relatively small differences among individual relation-removal variants, indicating that no single relational graph dominates the ranking performance under the current evaluation setting. Sensitivity analysis further shows that the resulting ranking is highly robust to uniform additions of up to 90 min to the estimated air travel time. Analysis of the resulting candidate set indicates that high-ranking ODs are typically associated with stronger inter-county interactions, higher ground-transport impedance, and limited direct high-speed rail connectivity. The framework therefore provides an analytical screening tool for reducing a large candidate space to a smaller set of links for subsequent market and operational feasibility assessment.

Keywords: low-altitude passenger services; multi-relational graph; graph attention network; transport impedance

Academic Editors: Artur Kierzkowski, Tomasz Nowakowski, Agnieszka A. Tubis, Franciszek Restel, Tomasz Kisiel, Anna Jodejko-Pietruczuk, Mateusz Zając, Viktoria Ivannikova, Michał Stosiak and Andrija Vidović

Received: 29 August 2026
Revised: 17 September 2026
Accepted: 18 September 2026
Published: 19 September 2026

1. Introduction

The development of the low-altitude economy and advanced air mobility (AAM) is reshaping the technological pathways, operational scenarios, and organizational modes of passenger air services. The continuous advancement of emerging aircraft, including electric-powered aircraft, piloted compound-wing aircraft, and electric vertical takeoff and landing (eVTOL) aircraft, provides important technological foundations for transforming low-altitude passenger transport from conceptual visions into practical applications. Meanwhile, relevant pilot programs are promoting the integration of these emerging aircraft into real operational environments. For example, the Electric Vertical Takeoff and Landing and Advanced Air Mobility Integration Pilot Program (eIPP) initiated by the Federal Aviation Administration (FAA) aims to explore safe operation and deployment pathways for eVTOL and other emerging aircraft within the U.S. airspace system through public–private cooperation [1]. Driven by both technological progress and operational validation, future passenger air services may gradually evolve from traditional airport-centered service networks toward more flexible and fine-grained regional point-to-point mobility services. Correspondingly, China has incorporated the low-altitude economy into the development framework of emerging and future industries, providing a new policy context for regional aviation, general aviation, and low-altitude passenger transport applications [2].

From an application perspective, low-altitude passenger transport can be broadly divided into intra-city and inter-city services. Intra-city services mainly focus on airport connections, cross-river or cross-bay commuting, and rapid connections within urban cores, where operations typically rely on dense vertiport deployment, refined airspace management, and integrated urban transport connections. In contrast, inter-city services primarily address short-distance travel between cities and counties. Their applicability depends not only on aircraft operational capabilities but also on existing ground transport efficiency, accessibility to takeoff and landing facilities, regional mobility intensity, and competition with alternative transport modes. Because such travel demands are more spatially dispersed and involve a larger number of potential service links, systematic market screening and regional transport condition assessment are particularly important during the early development stage of inter-city passenger services. From a transport-system planning perspective, this creates a large-scale screening problem in which analytical models are needed to prioritize candidate connections before detailed market, infrastructure, and operational assessments are undertaken. This study focuses on county-level short-haul passenger transport within the inter-city context. At the current early stage of low-altitude passenger transport development, the key research objective is not to directly determine whether an individual route can be launched, but rather to identify county-level links with higher research value and prioritization potential from a large number of potential connections. Therefore, this study investigates the prioritization of potential service links in county-level short-haul passenger transport, providing preliminary support for subsequent market research, route assessment, vertiport planning, and operational feasibility analysis.

Traditional aviation market analysis and route planning studies generally rely on relatively mature operational data and market-related variables. Existing studies commonly use airports or operated routes as analytical units and evaluate aviation demand, route performance, or market potential based on variables such as population size, gross domestic product (GDP), income level, airport throughput, historical passenger volume, flight frequency, spatial distance, and competing transport modes [3,4]. These approaches have strong explanatory capability in mature aviation markets, mainly because the investigated objects are embedded in relatively stable air service networks and have accumulated observable records of passenger demand, flight supply, and operational performance. However, county-level short-haul low-altitude passenger transport remains at an early development stage and lacks continuous, stable, and directly observable operational samples. Therefore, existing prediction frameworks based on historical passenger demand or established route performance cannot be directly applied. Under such conditions, the research focus should shift from determining whether a specific route has clear conditions for operation toward identifying the relative priorities among a large number of potential county-level routes. Accordingly, this study does not attempt to construct ground-truth labels equivalent to future demand. Instead, it formulates the limited availability of operational observations as a learning-to-rank problem under sparse proxy labels.

To address this issue, this study develops a multi-relational graph ranking framework that integrates county-level characteristics, inter-county interactions, and transport constraints into a unified candidate route prioritization process. Within this framework, the MR-GAT ranking model serves as the core scoring model, which combines information from multi-relational networks and candidate route attributes to generate continuous ranking scores for candidate routes. Through this approach, the framework identifies candidate links that deserve priority consideration in route assessment and provides support for preliminary planning and screening in the early stage of low-altitude passenger transport development.

The main contributions of this study are threefold. First, this study formulates county-level short-haul air-service screening as a large-scale priority-ranking problem under limited direct operational observations. Rather than treating unobserved OD pairs as negative samples, the proposed formulation focuses on their relative priorities for subsequent market and planning analysis. Second, this study develops a multi-relational analytical framework that jointly represents spatial proximity, observed inter-county mobility, and transport impedance, and integrates these relational contexts with county and OD attributes through an MR-GAT ranking model. This design enables heterogeneous transport-system information to be incorporated into a unified candidate-link screening process. Third, the framework is evaluated over 240,350 county-level candidate ODs using held-out regular air-service labels, auxiliary airport-service-coverage information, relational-graph ablation experiments, and an independent general-aviation reference set. The resulting rankings provide a practical means of reducing a large candidate space to a smaller set of priority links for subsequent market, infrastructure, and operational feasibility assessment.

The remainder of this paper is organized as follows. Section 2 reviews related studies on airport systems and short-haul aviation planning, door-to-door impedance and air–rail competition, network modelling and graph-based ranking, and mobility data. Section 3 introduces the construction of candidate routes, node attributes, route attributes, the multi-relational county network, and aviation proxy labels. Section 4 presents the MR-GAT ranking model, candidate route scoring method, and pairwise ranking objective. Section 5 reports the empirical results and discusses model performance, component contributions, top-ranked candidate enrichment, spatial patterns, and research limitations. Section 6 concludes the study.

2. Literature Review

County-level short-haul route prioritization involves three interrelated issues: first, whether air services can form effective supply within regional transport systems; second, whether potential links have realistic travel demand foundations and transport compensation value; and third, how to rank a large number of candidate links when direct operational observations are limited. Based on this logic, this section reviews related studies from four perspectives: airport systems and short-haul aviation planning, door-to-door impedance and air–rail competition, mobility and demand-side signals, and network modelling and candidate-link ranking.

2.1. Airport Systems, Regional Air Services and Short-Haul Planning

Research on airport systems emphasizes that aviation service capability depends not only on the existence of airport infrastructure but also on the interaction between airports and regional transport systems. Ground accessibility, route connectivity, service frequency, destination combinations, and competition from surrounding airports all influence actual airport catchment areas and regional aviation accessibility. Existing studies have integrated ground access conditions with airport connectivity to identify airport hinterlands, overlapping service areas, and regional aviation service gaps. Other studies have shown that multi-airport distribution and substitution relationships among airports can reshape aviation accessibility patterns at urban and regional scales [5,6]. These findings suggest that potential aviation service conditions cannot be assessed solely based on administrative boundaries or airport proximity, but should be understood within regional airport systems and ground access networks.

Research on aviation demand forecasting and route planning further indicates that aviation market potential results from the matching between demand conditions and service supply. Variables such as population, GDP, income level, airport throughput, flight frequency, and spatial distance are commonly used to explain aviation demand, route performance, and market potential. However, these variables generally rely on existing airport networks and observable operational records [3,4]. For mature aviation markets, such approaches can utilize historical passenger demand and flight supply information for prediction. However, for county-level short-haul low-altitude passenger transport that has not yet formed stable operational samples, directly applying existing route forecasting frameworks is constrained by the limited availability of observations.

Studies on regional air mobility and low-altitude transport further extend the discussion toward the operational conditions of short-haul aviation. The attractiveness of short-haul air services depends not only on flight distance but also on door-to-door time savings, accessibility to regional airports or vertiports, first- and last-mile connections, aircraft range, airspace capacity, and operational complexity [7–9]. Therefore, short-haul aviation should not be considered a market determined solely by distance, but rather as a travel option jointly shaped by technological capability, operational constraints, service organization, and regional transport conditions.

2.2. Door-to-Door Impedance, Air–Rail Competition and Multimodal Accessibility

The competitiveness of air transport in short- and medium-distance travel depends heavily on the complete travel chain. Passengers typically do not compare only in-vehicle flight time, but also consider departure access, waiting time, main-trip travel, arrival access, service frequency, reliability, and travel costs [10–12]. This issue becomes more prominent at the county scale. Accessibility evaluations conducted at the city scale may overlook differences in internal county-level node distances, access time, and service availability, while actual travel impedance between counties may be jointly determined by road detours, high-speed rail station accessibility, and transfer conditions.

The expansion of high-speed rail networks has changed the applicability boundaries of short- and medium-distance air services. Existing studies show that short-distance air demand is more likely to be affected by substitution in corridors with mature high-speed rail services. However, the relationship between air and rail is not simply a substitution relationship; passenger choices are also influenced by access convenience, reliability, value of time, and service experience [10,13]. Therefore, short- and medium-distance markets should not be regarded as homogeneous markets naturally suitable for aviation.

Short-haul air services are more likely to provide complementary value when ground travel impedance is high, high-speed ground transport alternatives are insufficient, or access efficiency is limited.

Multimodal accessibility research further suggests that the advantages of different transport modes should be evaluated within the complete door-to-door travel chain. High-speed rail accessibility depends not only on the existence of rail lines and stations, but also on station–city distance, intra-city access, travel time, and transfer conditions [11,14]. This perspective has direct implications for county-level short-haul route prioritization. High-speed rail conditions should not be simply represented as whether a station or route exists; instead, they should be transformed into transport constraint information by comparing ground travel time, access conditions, and potential air travel time advantages.

2.3. Mobility Data and Demand-Side Signals

Traditional travel surveys provide strong interpretability at the sample level, but their high collection costs and low update frequency make them difficult to apply for capturing large-scale county-level interactions. In contrast, mobile signaling data, call detail records (CDRs), and other passive location data can provide more frequent and extensive observations of regional mobility patterns, and have been widely used for travel matrix reconstruction and spatial interaction analysis [15,16]. However, the application of such data still faces challenges related to stay-point identification, spatial aggregation, user coverage, sampling bias, and privacy protection. Therefore, the interpretation of mobility-based results requires integration with population size, employment distribution, land use, and spatial proximity conditions.

For county-level short-haul aviation research, the value of mobility data does not lie in directly providing aviation demand labels, but rather in characterizing whether realistic interaction foundations exist between counties. Inter-county mobility may involve various purposes, including business, commuting, family visits, tourism, and other activities. Different mobility purposes may correspond to different willingness to pay, values of travel time, and transport mode preferences. Spatial interaction studies have long demonstrated that regional interactions are jointly influenced by scale and impedance, while distance decay effects vary with travel purposes, transport modes, and socioeconomic conditions [17,18]. Research on regional air mobility has similarly shown that different passenger groups exhibit heterogeneous sensitivities to time savings, fares, and service patterns [8].

Therefore, mobility intensity is more appropriately regarded as a proxy signal of existing inter-county interactions rather than a direct representation of future air passenger demand. If mobility volume is directly interpreted as aviation demand, the model may assign excessive priority to all highly connected links while ignoring the substitution effects of high-speed rail, road transport, and access conditions. Conversely, excluding mobility information entirely may cause candidate-link identification to rely excessively on distance or transport impedance, thereby failing to capture existing regional interaction foundations. Based on this understanding, this study incorporates mobility as demand-side relational information and combines it with spatial proximity and transport impedance for candidate-link ranking.

Mobility patterns also exhibit substantial spatial heterogeneity. The same level of mobility may correspond to different conditions for potential aviation conversion under different transport environments. In regions with mature high-speed rail services and convenient ground access, strong mobility interactions may not require additional air services. In contrast, in regions characterized by significant geographical barriers, severe ground detours, or insufficient high-speed transport coverage, medium-scale mobility interactions may still present relatively high transport compensation value. Therefore, mobility information should be interpreted jointly with ground transport impedance, high-speed rail substitution conditions, and spatial contexts.

2.4. Network-Based Modelling, Graph Learning and Candidate-Link Ranking

Traditional aviation network studies generally represent airports as nodes and routes as edges, and identify important airports or hub nodes through indicators such as degree, betweenness, closeness, eigenvector centrality, and PageRank [19,20]. These methods can explain connectivity strength and intermediary roles within existing aviation networks. However, their analytical objects are usually established airport networks. Some studies have attempted to identify potential routes based on existing airport networks and connectivity structures, but their research scales have mainly focused on airports or city pairs, making it difficult for them to directly address the preliminary screening of a large number of unserved or weakly served links at the county level [21].

County-level short-haul route prioritization differs substantially from traditional aviation network centrality analysis. The former does not focus on the importance of nodes within existing airport networks, but rather on the relative priorities among a large number of potential county-level links. Many potential links have no historical air passenger records and lack explicit negative samples that can be directly used for supervised learning. Therefore, relying solely on existing route topology or traditional centrality measures is insufficient to capture the spatial context, real-world interactions, and transport compensation conditions underlying candidate links.

Graph learning methods provide more flexible tools for representing multi-source relationships. Graph attention networks can adaptively assign weights according to differences among neighboring nodes and are suitable for modelling heterogeneous relationship strengths [22]. In transportation and aviation studies, graph attention, spatial dependency modelling, temporal attention mechanisms, and dynamic graph learning approaches have been applied to characterize complex relationships among nodes and alleviate the limitations of fixed physical adjacency structures [23,24]. This study draws on the relational representation capability of these approaches, but its objective is not operational state prediction; instead, it focuses on candidate service-link ranking.

From the perspective of task formulation, county-level short-haul route prioritization is closer to a learning-to-rank problem. Ranking learning focuses on relative priorities among candidate objects rather than class assignment for individual samples [25,26]. In transportation research, previous studies have constructed heterogeneous graphs using OD flows, paths, and road segments, and transformed the identification of critical road segments into ranking problems. These studies indicate that the importance of transportation objects is determined not only by physical adjacency but also by mobility relationships, path dependence, and functional interactions [27]. This perspective provides important implications for this study: when the objective is to prioritize candidates from a large candidate set rather than determine whether an object belongs to a positive class, ranking learning is more consistent with the underlying problem structure than binary classification.

Aviation demand modelling also faces challenges related to short historical sequences and sparse labels. Previous studies on aviation demand forecasting have indicated that aviation market data often exhibit a "short-but-wide" characteristic, where individual route histories are relatively short while numerous correlated links exist across the network, enabling information sharing through cross-series learning [28]. County-level short-haul route prioritization is positioned earlier than conventional route demand forecasting, as many candidate links lack historical passenger records and the number of observed route labels is limited. Treating unobserved links directly as negative samples may incorrectly classify undeveloped but potentially valuable links as having no value. This issue is conceptually similar to the treatment of unobserved samples in implicit-feedback ranking problems [29]. The main research streams reviewed above and their relevance to this study are summarized in Table 1.

Table 1. Main research streams and their relevance to this study.

Research Stream	Representative Studies	Main Focus	Relevance and Remaining Gap
Airport systems and short-haul planning	[3–9]	Airport catchment areas, regional aviation accessibility, route planning, service supply, and short-haul aviation operational conditions	Indicates that potential aviation services should be analyzed by considering airport systems, ground access, and operational constraints. However, most studies rely on existing airport networks or operational samples.
Door-to-door impedance and air–rail competition	[10–14]	Door-to-door travel time, air–rail competition, multimodal accessibility, and access conditions	Demonstrates that the value of short-haul aviation should be evaluated considering complete travel chains and alternative transport conditions. However, large-scale county-level candidate route prioritization is rarely considered as the research objective.
Mobility data and demand-side signals	[8,15–18]	Spatial interaction, mobility observations, travel connections, and demand-side proxy signals	Provides information for characterizing existing inter-county interaction foundations. However, mobility intensity cannot be directly interpreted as aviation demand and should be jointly explained with transport constraints.
Network modelling, graph learning and ranking	[19–29]	Network structure analysis, graph learning, ranking learning, unobserved sample handling, and cross-link information sharing	Provides methodological foundations for multi-relational representation and candidate-link ranking. However, existing applications mainly focus on established networks, operational prediction, or other transportation objects, with limited attention to county-level short-haul route prioritization.

3. Problem Formulation and Data Construction

This section formulates the county-level short-haul route prioritization problem from the perspectives of dataset construction, node attributes, edge attributes, relational graphs, and proxy labels. County administrative units are defined as the basic spatial nodes, while potential short-haul air-service links between counties are considered as the ranking targets. Based on this formulation, the attribute data, relational structures, and label references required for subsequent modelling are constructed.

3.1. OD Set and Attribute Construction

This study considers county units as the basic spatial nodes. Let N denote the set of county nodes, where $i,j \in N$ represent two county nodes. An OD represents a potential short-haul air-service link between two county nodes. Since this study focuses on the relative priority of OD pairs rather than specific travel directions, flight schedules, flight frequencies, or operational organization schemes, an undirected OD setting is adopted.

For the construction of the OD set, the great-circle distance between county centroids is used as the spatial screening criterion, and a distance window of 200–400 km is defined for empirical analysis. This distance range does not correspond to the fixed design range of any specific aircraft type, nor is it interpreted as a universal operational boundary. Instead, it is used to determine the OD pairs included in the ranking task from a large number of county combinations. Accordingly, the OD set is defined as

$$E^{OD}=\{(i,j)\in N\times N|\, i<j, 200\le d_{ij}^{gc}\le 400\}, \quad (1)$$

where E^{OD} represents the set of OD pairs included in the ranking task. The condition $i<j$ ensures that each undirected OD pair is counted only once, and d_{ij}^{gc} denotes the great-circle distance between the centroids of county i and county j.

For attribute construction, multi-source data are organized into two categories: node attributes and edge attributes. Node attributes describe the characteristics of individual county nodes, while edge attributes characterize the interaction intensity and transport conditions between two county nodes.

For any county node $i\in N$, its node attribute vector is defined as

$$\mathbf{x}_i=[P_i, GDP_i]^{\top}, i\in N, \quad (2)$$

where population size and gross regional product are selected as node attributes to represent the basic scale and economic activity level of each county. Specifically, P_i denotes the population size of county node i, and GDP_i represents the gross regional product of county node i.

For any OD pair $(i,j)\in E^{OD}$, the edge attribute vector is defined as

$$\mathbf{e}_{ij}=\left[M_{ij}, T_{ij}^{road}, D_{ij}^{road}, H_{ij}^{hsr}, T_{ij}^{hsr}, T_{ij}^{air}\right]^{\top}, (i,j)\in E^{OD}, \quad (3)$$

where M_{ij} represents the mobility intensity of OD pair (i,j), which describes the existing interaction between the two counties. T_{ij}^{road} denotes ground travel time, D_{ij}^{road} represents ground travel distance, H_{ij}^{hsr} indicates high-speed rail connectivity status, T_{ij}^{hsr} denotes high-speed rail travel time, and T_{ij}^{air} represents the estimated air travel time. These variables are collected and processed from multiple data sources. To clarify the data sources, processing approaches, and roles of different variables in attribute construction, Table 2 summarizes the data used in this study.

Table 2. Data sources.

Symbol	Data Description	Data Source
P_i	Population size of county node i	China County Statistical Yearbook 2024
GDP_i	Gross regional product of county node i	China County Statistical Yearbook 2024
M_{ij}	Mobility intensity of OD pair (i,j)	Anonymized and aggregated sampled mobile signaling data in 2023
T_{ij}^{road}	Ground travel time of OD pair (i,j)	Obtained through map API queries
D_{ij}^{road}	Ground travel distance of OD pair (i,j)	Obtained through map API queries
H_{ij}^{hsr}	High-speed rail connectivity status of OD pair (i,j)	Compiled based on queries of train schedules in December 2025
T_{ij}^{hsr}	High-speed rail travel time of OD pair (i,j)	Compiled based on queries of train schedules in December 2025
T_{ij}^{air}	Estimated air travel time of OD pair (i,j)	Calculated based on the great-circle distance between counties i and j and the cruising speed specified in this study

https://doi.org/10.3390/app16189299

It should be noted that the travel-time variables in this study are not intended to represent complete door-to-door travel times for different transport modes. Ground travel time is obtained from road-network routing between county-level representative locations, while high-speed rail travel time represents the identified train running time and does not consistently include station access, waiting, transfer, or egress time. Similarly, the estimated air travel time is calculated from great-circle distance and an assumed cruise speed and does not include access, terminal processing, takeoff and landing, or egress processes. Therefore, these variables should be interpreted as mode-specific transport-impedance measures used for large-scale preliminary OD screening rather than as directly comparable door-to-door travel times. The purpose of introducing these variables is to characterize relative transport conditions among candidate ODs, rather than to estimate passenger mode-choice utilities or the actual door-to-door competitiveness of aviation against ground transport. In this study, high-speed rail connectivity refers to direct HSR/EMU services identified under the adopted timetable-matching rule. Therefore, an OD pair coded as having no HSR connectivity only indicates that no direct service was identified; potentially convenient one-transfer rail itineraries are not represented by the current binary variable.

3.2. Multi-Relational County Network

This study constructs three types of relational graphs to characterize different relationship structures among county nodes and provide network context for subsequent graph modelling. The three relational graphs share the same county node set N, but have different edge sets and edge attributes. The multi-relational graph is defined as $G = \{G^S, G^F, G^I\}$ where G^S represents the spatial proximity graph, G^F represents the mobility graph, and G^I represents the transport impedance graph. For any relational type $r \in \{S,F,I\}$, the corresponding relational graph can be represented as

$$G^r = \left(N, A^r, \{\mathbf{a}_{ij}^r : (i,j) \in A^r\}\right), r \in \{S,F,I\}, \tag{4}$$

where N denotes the set of county nodes, A^r represents the edge set of relational graph G^r, and $\mathbf{a}_{ij}^r$ denotes the edge attribute vector of edge (i,j) in relational graph G^r.

The spatial proximity graph G^S is constructed to characterize local spatial associations among county nodes. This graph is established based on county centroid coordinates, with its edge set denoted as A^S. Considering the spatial density differences among counties, an adaptive neighborhood strategy is adopted to determine spatial edges. In densely distributed county areas, neighboring nodes with shorter distances are preferentially retained, while in sparsely distributed areas, isolated nodes are avoided by adding nearest neighbors. Therefore, the spatial proximity graph maintains local spatial relationships while reducing the influence of regional spatial density differences on graph structures. For any spatial edge $(i,j) \in A^S$, its edge attribute is used to describe the degree of spatial proximity between two county nodes:

$$\mathbf{a}_{ij}^S = \left[d_{ij}^{gc}, \omega_{ij}^S\right]^\top, \omega_{ij}^S = \exp\left(-\frac{d_{ij}^{gc}}{\sigma_i}\right), (i,j) \in A^S, \tag{5}$$

where d_{ij}^{gc} represents the great-circle distance between the centroids of county i and county j; ω_{ij}^S denotes the spatial distance decay weight; and σ_i represents the local spatial scale of node i, which is determined by the median distance among the spatial neighbors selected for node i. This weight transforms spatial distance into neighborhood relationship strength, enabling the graph model to distinguish different levels of spatial proximity among neighboring nodes.

The mobility graph G^F is constructed to characterize existing population interaction relationships among county nodes. Based on inter-county mobility data, the edge set A^F

is constructed, and variables including total mobility intensity, mobility intensity by purpose, mobility composition ratios, bidirectional mobility imbalance, and logarithmic mobility weights are incorporated as edge attributes. For mobility records containing directional information, bidirectional mobility flows are aggregated into the same county pair to form mobility relationships consistent with the undirected OD definition. For any mobility edge $(i,j) \in A^F$, its edge attribute vector is defined as

$$\mathbf{a}_{ij}^F = \left[M_{ij}, M_{ij}^{bus}, M_{ij}^{com}, M_{ij}^{lei}, q_{ij}^{bus}, q_{ij}^{com}, q_{ij}^{lei} b_{ij}^F, log(1+M_{ij})\right]^{\top}, (i,j) \in A^F, \tag{6}$$

where M_{ij} represents the total mobility intensity of OD pair (i,j); M_{ij}^{bus}, M_{ij}^{com} and M_{ij}^{lei} represent mobility intensities for different purposes, including business, commuting, and leisure activities, respectively; $q_{ij}^{bus}, q_{ij}^{com}$ and q_{ij}^{lei} denote the proportions of the three mobility purposes within total mobility; b_{ij}^F represents bidirectional mobility imbalance; and $\log(1+M_{ij})$ denotes the logarithmic mobility weight, which is introduced to reduce the influence of extreme mobility values on model training.

The transport impedance graph G^I is constructed to characterize ground transport constraints and high-speed rail substitution conditions among county nodes. Based on ground travel, high-speed rail connectivity, and estimated air travel time information of OD pairs (i,j), the edge set A^I is constructed. Variables including ground travel distance, ground travel time, great-circle distance, high-speed rail connectivity status, high-speed rail travel time, effective ground travel time, relative time indicators, and time-decay weights are incorporated as edge attributes. For any transport impedance edge $(i,j) \in A^I$, its edge attribute vector is defined as

$$\mathbf{a}_{ij}^I = \left[D_{ij}^{road}, T_{ij}^{road}, d_{ij}^{gc}, T_{ij}^{air}, \tilde{T}_{ij}^{surf}, H_{ij}^{\text{hsr}}, T_{ij}^{\text{hsr}}, R_{ij}^{road/air}, R_{ij}^{surf/air}, \omega_{ij}^{road}, \omega_{ij}^{surf}\right]^{\top}, (i,j) \in A^I, \tag{7}$$

where D_{ij}^{road} represents the ground travel distance, T_{ij}^{road} denotes the ground travel time, d_{ij}^{gc} represents the great-circle distance between the centroids of the two counties, T_{ij}^{air} denotes the estimated air travel time, H_{ij}^{hsr} is a binary variable indicating high-speed rail connectivity status, and T_{ij}^{hsr} represents high-speed rail travel time. To simultaneously reflect conventional ground transport and high-speed rail substitution conditions, this study defines the effective ground travel time $\tilde{T}_{ij}^{surf}$. When high-speed rail is available, the smaller value between ground travel time and high-speed rail travel time is selected; otherwise, the ground travel time is used:

$$\tilde{T}_{ij}^{surf} = \begin{cases} min\left(T_{ij}^{\text{road}}, T_{ij}^{\text{hsr}}\right), & H_{ij}^{\text{hsr}} = 1 \\ T_{ij}^{\text{road}}, & H_{ij}^{\text{hsr}} = 0' \end{cases} \tag{8}$$

Based on this definition, two relative time indicators are constructed to characterize the relative relationship between the estimated air travel time and the available ground-transport time measures:

$$R_{ij}^{road/air} = \frac{T_{ij}^{road}}{T_{ij}^{air}}, \qquad R_{ij}^{surf/air} = \frac{\tilde{T}_{ij}^{surf}}{T_{ij}^{air}}, \tag{9}$$

where $R_{ij}^{road/air}$ represents the ratio between conventional ground travel time and estimated air travel time, while $R_{ij}^{surf/air}$ represents the ratio between effective ground travel time considering high-speed rail substitution and estimated air travel time. To further represent the decay effect of transport impedance, this study defines the transport weights as

$$\omega_{ij}^{road} = exp\left(-\frac{T_{ij}^{road}}{\tau_{road}}\right), \qquad \omega_{ij}^{surf} = exp\left(-\frac{\tilde{T}_{ij}^{surf}}{\tau_{surf}}\right), \tag{10}$$

where τ_{road} and τ_{surf} are the time-decay parameters for ground travel time and effective ground travel time, respectively. They control the decreasing rate of edge weights as travel time increases. Both parameters are set as the medians of their corresponding time variables to reduce the influence of extreme travel times on weight calculation. Through these constructions, the transport impedance graph characterizes transport conditions among county nodes from three perspectives: conventional ground transport, high-speed rail substitution, and the relative relationship between ground-transport time measures and estimated air travel time. Together with the spatial proximity graph and mobility graph, the transport impedance graph provides structural inputs for the subsequent MR-GAT ranking model.

3.3. Aviation Proxy Label Construction

This study constructs three types of proxy labels to provide pairwise ranking supervision and independent evaluation references, including route labels, airport catchment weak labels, and external evaluation labels.

Route labels are constructed to represent existing regular air-service links. The original aviation service data use routes as the basic recording unit. Based on the spatial correspondence between airports and counties, this study transforms route records into county-level air-service labels within the OD set. Let $E^{\text{route}} \subseteq E^{\text{OD}}$ denote the set of route labels. If an OD pair corresponds to an observed aviation service record, it is included in E^{route}. The corresponding label is defined as

$$y_{ij}^{\text{route}} = \begin{cases} 1, & (i,j) \in E^{\text{route}} \\ unlabelled, & (i,j) \notin E^{\text{route}} \end{cases}, \tag{11}$$

where y_{ij}^{route} denotes the route label. Unlabelled OD pairs are not assigned confirmed negative class labels. During pairwise training, a subset of unlabelled ODs may be sampled as lower-priority comparison instances relative to observed route-label ODs. This represents a weak ranking preference rather than evidence that the sampled ODs have no market potential. Route records with available service frequency information are further used to characterize the intensity of existing aviation services.

Airport catchment weak labels are introduced to complement the coverage limitations of route labels. Route labels only represent observed aviation service links, while airport catchment weak labels indicate whether both ends of an OD pair satisfy existing airport service coverage conditions. Let $E^{\text{weak}} \subseteq E^{\text{OD}}$ denote the set of airport catchment weak labels. If both counties of an OD pair satisfy the airport catchment coverage condition, the OD pair is included in E^{weak}. The corresponding label is defined as

$$y_{ij}^{\text{weak}} = \begin{cases} 1, & (i,j) \in E^{\text{weak}} \\ unlabelled, & (i,j) \notin E^{\text{weak}} \end{cases}, \tag{12}$$

where y_{ij}^{weak} denotes the airport catchment weak label. This label is only used as auxiliary supervision information to alleviate the sparsity of route labels and does not replace route labels.

External evaluation labels are constructed to provide an independent evaluation reference. Let $E^{\text{ext}} \subseteq E^{\text{OD}}$ denote the set of external evaluation labels. This set is generated by compiling general aviation routes and mapping them to the OD level. It is used to examine whether the final ranking results can capture county-level links observed in another type of aviation activity. The corresponding label is defined as

$$y_{ij}^{\text{ext}} = \begin{cases} 1, & (i,j) \in E^{\text{ext}} \\ unlabelled, & (i,j) \notin E^{\text{ext}} \end{cases}, \tag{13}$$

where y_{ij}^{ext} denotes the external evaluation label. The external evaluation labels are not used to construct model training supervision signals or generate relational graph structures, but only serve as independent evaluation references. The external evaluation set contains 40 positive OD labels constructed from general aviation routes. These labels are evaluated against the same complete ranking of 240,350 candidate ODs and therefore do not constitute a separate balanced classification test set. External Top-K recall represents the proportion of these 40 positive reference ODs retrieved within the corresponding Top-K range.

In summary, this section formulates and organizes the county-level short-haul route prioritization problem based on the OD set, node attributes, edge attributes, relational graphs, and proxy labels. The OD set defines the ranking targets, while node attributes and edge attributes characterize county-level conditions and inter-county transport interactions, respectively. The multi-relational county network further represents spatial proximity, mobility, and transport impedance relationships. The proxy labels provide references for ranking learning and model evaluation.

4. Methodology

This section develops an MR-GAT ranking model for OD priority ranking. Taking county node attributes, OD edge attributes, and multi-relational county networks as inputs, the model transforms county-level characteristics, inter-county interactions, and relational graph structures into relative priority scores of OD pairs through multi-relational graph encoding, OD-level scoring, and pairwise ranking training.

4.1. Overview of the MR-GAT Ranking Model

The objective of the MR-GAT ranking model is to learn a scoring function over the OD set E^{OD}. For any OD pair $(\mathrm{i,j}) \in \mathrm{E}^{\mathrm{OD}}$, the model takes the attributes of the two county nodes, edge attributes, and multi-relational county networks as inputs, and outputs a continuous ranking score:

$$s_{ij} = f_{\Theta}\left(\mathbf{x}_i, \mathbf{x}_j, \mathbf{e}_{ij}, G\right), \qquad (i,j) \in E^{\mathrm{OD}}, \tag{14}$$

where $\mathbf{x}_i$ and $\mathbf{x}_j$ denote the attribute vectors of the two county nodes, $\mathbf{e}_{ij}$ represents the edge attribute vector of OD pair (i,j), and G represents the three relational graphs. f_{Θ} denotes the ranking function with learnable parameters Θ, and s_{ij} is the ranking score generated by the model. A larger s_{ij} indicates that the model assigns a higher priority to OD pair (i,j). This score is only used for relative comparison among OD pairs and should not be interpreted as passenger demand, route opening probability, or operational feasibility.

As illustrated in Figure 1, the MR-GAT ranking model mainly consists of three modules: data input, multi-relational graph encoding, and OD scoring. The first module is the data input module, which includes county node attributes, edge attributes, and three relational graphs: the spatial proximity graph, mobility graph, and transport impedance graph. The second module is the multi-relational graph encoding module. The model updates county node representations on the three relational graphs separately, obtains node representations under different relational contexts, and then integrates these representations through relation fusion to generate unified node representations. The third module is the OD scoring module, where the node representations of the two counties and the edge attributes of the corresponding OD pair are concatenated to construct an OD feature vector, and the scoring network outputs the ranking score s_{ij}. Through this process, the model transforms county-level characteristics, inter-county attribute relationships, and multi-relational network structures into relative priority scores for OD pairs.

https://doi.org/10.3390/app16189299

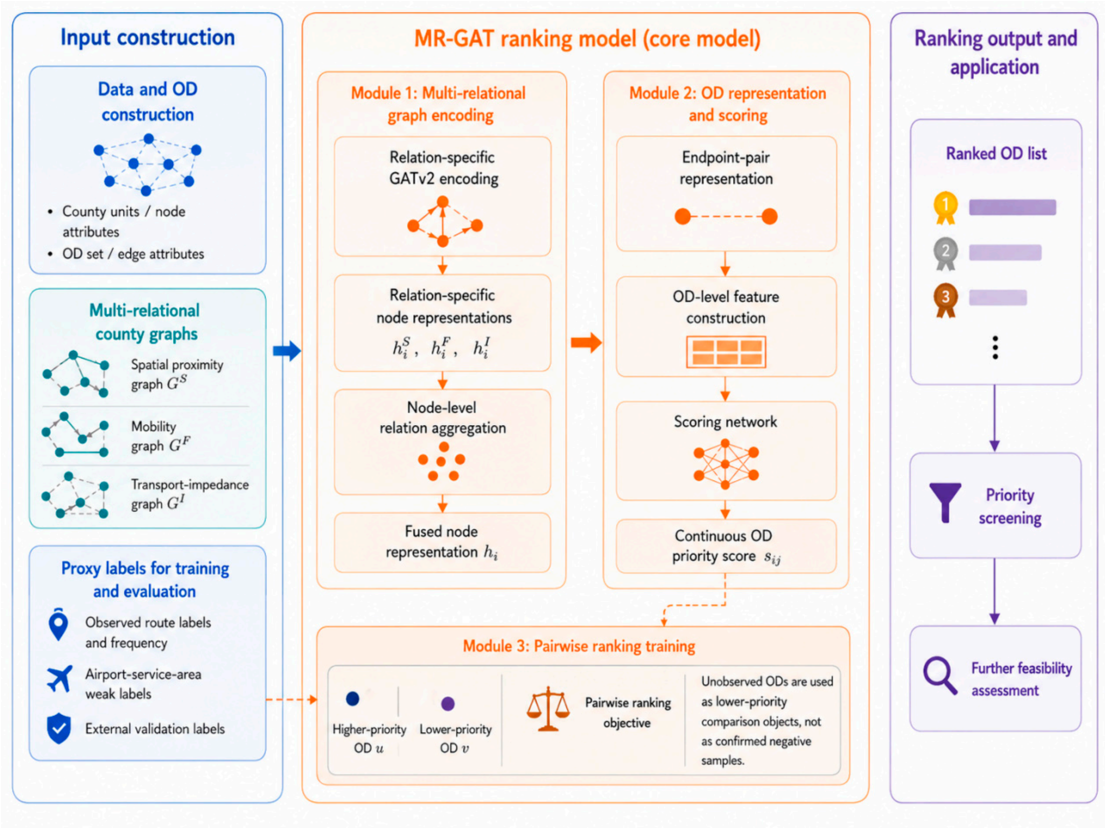

Figure 1. Multi-relational graph ranking framework for county-level OD prioritization.

Before introducing the graph encoding process, it should be clarified that route labels and airport catchment weak labels are used as supervision signals during model training or evaluation, while external evaluation labels are only used for result assessment. This setting avoids directly using label information to generate ranking scores and ensures that model scoring mainly relies on node attributes, edge attributes, and multi-relational graph structures.

4.2. Multi-Relational Graph Encoding

In the MR-GAT ranking model, the multi-relational graph encoding module first transforms the county node attribute vector $\mathbf{x}_i$ into an initial node representation suitable for graph neural network computation through a node attribute mapping function:

$$\mathbf{h}_i^{(0)} = \phi_x(\mathbf{x}_i), \qquad i \in N, \tag{15}$$

where $\mathbf{h}_i^{(0)}$ represents the initial node representation of node i, and $\phi_x(\cdot)$ denotes a learnable node attribute mapping function. The initial node representations of all nodes are organized into a matrix $\mathbf{H}^{(0)}$, which is used as the input of the graph encoding module. For the three relational graphs $\{G^S, G^F, G^I\}$, independent GATv2 encoders are constructed to extract structural information and edge attribute information from different relational graphs. GATv2 is a dynamic graph attention mechanism that alleviates the limitations of the original GAT related to attention ranking and query-node sensitivity [30]. To distinguish the encoding parameters of different relational graphs, the GATv2 encoder corresponding to relation type $r \in \{S, F, I\}$ is denoted as $\Psi_r(\cdot)$. For relational graph G^r, the encoding process is formulated as

$$\mathbf{H}^r = \Psi_r\left(\mathbf{H}^{(0)}, A^r, \mathbf{B}^r\right), \qquad r \in \{S, F, I\}, \tag{16}$$

where $\mathbf{H}^r$ represents the node representation matrix after encoding through relational graph G^r, $\mathbf{H}^{(0)}$ denotes the initial node representation matrix of all nodes, A^r represents the edge set of relational graph G^r, and $\mathbf{B}^r$ represents the edge attribute matrix composed of edge attribute vectors in relational graph G^r. After encoding through the three relational graphs, each node obtains three relation-specific representations, $\mathrm{h}_{\mathrm{i}}^{\mathrm{S}}, \mathrm{h}_{\mathrm{i}}^{\mathrm{F}}, \mathrm{h}_{\mathrm{i}}^{\mathrm{I}}$, corresponding to spatial proximity, mobility, and transport impedance relationships, respectively. Since the three relational graph encoders output node representations with the same dimensionality, these representations can be further fused. However, information sources and dependency patterns may differ across the three relational graphs. Therefore, this study introduces node-level relation attention to calculate the relative contribution of different relational graphs to the same node representation.

Since the information sources and dependency patterns of different nodes may vary across the three relational graphs, the model further adopts node-level relation attention to fuse the three types of node representations. This attention mechanism measures the relative contribution of different relational graphs to the representation of the same node, which differs from the neighborhood aggregation attention used inside the GATv2 encoder. For node i and relation type r, the unnormalized attention score of relational graph G^r is first calculated as

$$u_i^r = \mathbf{q}^\top \tanh(\mathbf{P}_r \mathbf{h}_i^r), \qquad r \in \{S,F,I\}, \tag{17}$$

where $\mathbf{P}_r$ denotes a learnable linear transformation matrix corresponding to relation type r, which maps node representations from different relational graphs into a unified attention space. $\mathbf{q}$ is a learnable attention vector used to compress the transformed node representation into a scalar score. $\tanh(\cdot)$ is the hyperbolic tangent activation function, which introduces nonlinear transformation. $\mathbf{h}_i^r$ represents the node representation of node i in relational graph G^r. The three relation-specific attention scores are then normalized using the softmax function:

$$\beta_i^r = \frac{exp(u_i^r)}{\sum_{r' \in S,F,I} exp(u_i^{r'})}, \qquad r \in \{S,F,I\}, \qquad \sum_{r \in S,F,I} \beta_i^r = 1, \tag{18}$$

where β_i^r represents the weight of relational graph G^r in the fused representation of node i. Since softmax normalization is performed among the three relational graphs, β_i^S, β_i^F and β_i^I reflect the relative dependency of node i on spatial proximity, mobility, and transport impedance information, rather than globally fixed weights. After obtaining the relation weights, the fused node representation is defined as

$$\mathbf{h}_i = \sum_{r \square S,F,I} \beta_i^r \mathbf{h}_i^r, \tag{19}$$

Accordingly, node i obtains three relation-specific representations, $\beta_i^S, \beta_i^F, \beta_i^I$, together with the fused node representation $\mathbf{h}_i$. The former preserve information from different relational sources, while the latter integrates the three relational contexts according to node-specific relation weights and serves as the node input for subsequent OD scoring.

4.3. OD-Level Scoring

After obtaining the relation-specific node representations $\beta_i^S, \beta_i^F, \beta_i^I$ and the fused node representation $\mathbf{h}_i$, this section further combines the endpoint node representations with OD edge attributes to construct the input of the OD scoring module. For any OD pair $(i,j) \in E^{\mathrm{OD}}$, the model first constructs an endpoint pair representation under each relational graph. For relation type $r \in \{S,F,I\}$, it is defined as

$$\mathbf{b}_{ij}^r = [\mathbf{h}_i^r + \mathbf{h}_j^r \| |\mathbf{h}_i^r - \mathbf{h}_j^r| \| \mathbf{h}_i^r \odot \mathbf{h}_j^r], \tag{20}$$

where $\mathbf{h}_i^r$ and $\mathbf{h}_j^r$ represent the node representations of the two county nodes under relational graph G^r, respectively; $\|$ denotes vector concatenation; $\odot$ denotes element-wise multiplication; and $|\mathbf{h}_i^r - \mathbf{h}_j^r|$ represents the element-wise absolute difference between the two node representations. The endpoint pair representation consists of three components: $\mathbf{h}_i^r + \mathbf{h}_j^r$ captures the overall feature level of the two endpoint nodes under the corresponding relational graph; $\left|\mathbf{h}_i^r - \mathbf{h}_j^r\right|$ describes the difference between the two node representations; and $\mathbf{h}_i^r \odot \mathbf{h}_j^r$ captures the interaction relationship between the two endpoint node representations. Similarly, based on the fused node representations $\mathbf{h}_i$ and $\mathbf{h}_j$, an overall endpoint pair representation is constructed as

$$\mathbf{b}_{ij} = \left[\mathbf{h}_i + \mathbf{h}_j \| \left|\mathbf{h}_i - \mathbf{h}_j\right| \| \mathbf{h}_i \odot \mathbf{h}_j\right], \tag{21}$$

Unlike $\mathbf{b}_{ij}^r$, which corresponds to a single relational graph, $\mathbf{b}_{ij}$ is constructed based on the fused node representation after integrating information from the three relational graphs. After obtaining the endpoint pair representations, the OD scoring module further combines OD edge attributes to construct relation-specific OD feature vectors. For relation type $r \in \{S,F,I\}$, the OD representation is defined as

$$\mathbf{p}_{ij}^r = \rho_r\left(\left[\mathbf{b}_{ij}^r \| \mathbf{b}_{ij} \| \mathbf{e}_{ij}\right]\right), \qquad r \in \{S,F,I\}, \tag{22}$$

where $\rho_r(\cdot)$ denotes a learnable mapping function corresponding to relation type r, $\mathbf{e}_{ij}$ represents the edge attribute vector of OD pair (i,j), and $\mathbf{p}_{ij}^r$ represents the OD feature vector under relational graph G^r. Through this process, each OD pair obtains three relation-specific feature representations corresponding to spatial proximity, mobility, and transport impedance relationships. To fuse the three relation-specific OD representations, the OD scoring module further learns OD-level relation weights γ_{ij}^r. Unlike node-level relation weights, OD-level relation weights characterize the relative contribution of different relational graphs to the score of a specific OD pair. For relation type $r \in \{S,F,I\}$, the OD-level relation weight is defined as

$$\gamma_{ij}^r = \frac{exp\left(\eta_r\left(\left[\mathbf{p}_{ij}^r \| \mathbf{e}_{ij}\right]\right)\right)}{\sum_{r' \in S,F,I} exp\left(\eta_{r'}\left(\left[\mathbf{p}_{ij}^{r'} \| \mathbf{e}_{ij}\right]\right)\right)}, \qquad r \in S,F,I, \qquad \sum_{r \in S,F,I} \gamma_{ij}^r = 1, \tag{23}$$

where $\eta_r(\cdot)$ denotes the learnable scoring function corresponding to relation type r. The OD-level relation weight γ_{ij}^r is jointly determined by the relation-specific OD representation $\mathbf{p}_{ij}^r$, the OD edge attribute vector $\mathbf{e}_{ij}$, and the learnable scoring function $\eta_r(\cdot)$. The softmax normalization is performed among the three relational graphs. Therefore, $\gamma_{ij}^S, \gamma_{ij}^F, \gamma_{ij}^I$ represent the relative contributions of spatial proximity, mobility, and transport impedance information to the ranking of the same OD pair, respectively, and their sum equals one. The three relation-specific OD representations are then fused through weighted aggregation to obtain the final OD-level representation:

$$\mathbf{p}_{ij} = \sum_{r \square S,F,I} \gamma_{ij}^r \mathbf{p}_{ij}^r, \tag{24}$$

Finally, the model feeds the fused OD representation together with the OD edge attribute vector into a learnable scoring function $f_s(\cdot)$ to obtain the ranking score:

$$s_{ij} = f_s\left(\left[\mathbf{p}_{ij} \| \mathbf{e}_{ij}\right]\right), \qquad (i,j) \in E^{\mathrm{OD}}, \tag{25}$$

4.4. Pairwise Ranking Objective and Training Protocol

The previous sections defined the OD ranking score s_{ij}. This section further describes the construction of training supervision. Since the objective of this study is to learn relative priorities among OD pairs, proxy labels are not directly used to construct binary classification labels, but are transformed into ordered comparison pairs. Let $u,v \in E^{\mathrm{OD}}$ denote two OD pairs. If the supervision information indicates that u has a higher priority than v, it is denoted as $u \succ v$, and the model is expected to satisfy $s_u > s_v$. For an ordered comparison pair (u,v), this study adopts the pairwise logistic ranking loss:

$$\ell(u,v) = \log(1+exp[-(s_u - s_v)]), \tag{26}$$

where $\ell(u,v)$ represents the ranking loss of comparison pair (u,v). The loss takes the score difference $s_u - s_v$ as input. When s_u is substantially larger than s_v, $\ell(u,v)$ becomes smaller. When $s_u \le s_v$, the loss increases and encourages the model to assign higher scores to higher-priority OD pairs. This formulation is consistent with commonly used pairwise ranking losses in learning-to-rank methods [25].

Based on route labels, flight frequencies, and airport catchment weak labels, this study constructs three types of ordered comparison pair sets. To reduce the risk that held-out positive samples are incorrectly used as lower-priority comparison objects, positive labels in the validation set, internal test set, and external evaluation set are excluded from low-priority OD sampling. The low-priority comparison pool is drawn from candidate ODs that are not used as positive route labels after excluding held-out positive ODs. These ODs are used only as comparison instances for constructing pairwise ranking preferences and are not treated as confirmed negative classes.

The first type is the route-label comparison pair set, denoted as $P^{\mathrm{route}} = \{(u,v){:}u,v \in E^{\mathrm{OD}}, u \succ v\}$. The route-label ranking loss is defined as

$$\mathrm{L}_{\mathrm{route}} = \frac{1}{|P^{\mathrm{route}}|}\sum\nolimits_{(u,v)\in P^{\mathrm{route}}} \mathrm{l}(u,v), \tag{27}$$

where $|P^{\mathrm{route}}|$ represents the number of route-label comparison pairs.

The second type is the frequency comparison pair set, denoted as $P^{\mathrm{freq}} = \{(u,v){:}u,v \in E^{\mathrm{OD}}, u \succ v\}$. Within OD pairs containing route labels, if the service frequency of OD(u) is higher than that of OD(v), a frequency-based ranking relationship $u \succ v$ is constructed. The corresponding frequency ranking loss is defined as

$$\mathrm{L}_{\mathrm{freq}} = \frac{1}{|P^{\mathrm{freq}}|}\sum\nolimits_{(u,v)\in P^{\mathrm{freq}}} \mathrm{l}(u,v), \tag{28}$$

where $|P^{\mathrm{freq}}|$ represents the number of frequency-based comparison pairs. A larger frequency difference provides a stronger ranking constraint during model training.

The third type is the airport catchment weak-label comparison pair set, denoted as $P^{\mathrm{weak}} = \{(u,v){:}u,v \in E^{\mathrm{OD}}, u \succ v\}$. The weak-label ranking loss is defined as

$$\mathrm{L}_{\mathrm{weak}} = \frac{1}{|P^{\mathrm{weak}}|}\sum\nolimits_{(u,v)\in P^{\mathrm{weak}}} \mathrm{l}(u,v), \tag{29}$$

where $|P^{\mathrm{weak}}|$ represents the number of airport catchment weak-label comparison pairs. Since airport catchment weak labels only indicate whether an OD pair satisfies existing airport service coverage conditions and do not provide service intensity information, this study assigns equal weights to weak-label comparison pairs.

By integrating the three ranking constraints, the training objective of the model is defined as

$$\mathrm{L} = \lambda_{\mathrm{route}}\mathrm{L}_{\mathrm{route}} + \lambda_{\mathrm{freq}}\mathrm{L}_{\mathrm{freq}} + \lambda_{\mathrm{weak}}\mathrm{L}_{\mathrm{weak}} + \lambda_{\mathrm{reg}}\|\Theta\|_2^2, \tag{30}$$

where Θ denotes the set of learnable parameters of the MR-GAT ranking model; λ_{route},λ_{freq}, λ_{weak} represent the weights of route-label ranking loss, frequency ranking loss, and weak-label ranking loss, respectively; and λ_{reg} denotes the regularization coefficient. Instead of assigning additional weights to individual comparison pairs, this study applies weighting at the loss-term level to balance different supervision sources. The main model uses route-label ranking loss and frequency ranking loss as the primary supervision signals, while the weak-label ranking loss is not activated in the main model. The model variant incorporating airport catchment weak labels is implemented by assigning a non-zero weight to the weak-label loss term. The weighting strategy adopted here reflects the different semantic roles of the supervision sources. Route-label ranking provides the primary supervision target, while flight-frequency information and, in the corresponding variant, airport-catchment weak labels provide auxiliary ranking constraints. Therefore, fixed loss-term weights are retained to preserve this predefined supervision hierarchy. It should also be distinguished from the fusion of the three relational graphs, whose contributions are adaptively learned through the node-level and OD-level softmax relation-attention mechanisms rather than optimized as separate task losses. Training regularization is further provided by dropout, gradient clipping, early stopping, and AdamW optimization, while stability is assessed through repeated runs with different random seeds. Adaptive gradient- or loss-balancing methods may be investigated in future extensions when a larger number of heterogeneous supervision objectives are jointly optimized.

The MR-GAT ranking model is trained using the AdamW optimizer [31]. During training, route-label comparison pairs and frequency comparison pairs are sampled using mini-batches and jointly contribute to the training loss. For the weak-label variant, airport catchment weak-label comparison pairs are also sampled through mini-batches for loss computation. After training, the optimal model parameters are selected according to ranking performance on the validation set. External evaluation labels are only used for final result assessment and do not participate in parameter updates or model selection. The main training configurations are summarized in Table 3.

Table 3. Training parameter settings of the MR-GAT ranking model.

Item	Setting
Optimizer	AdamW
Learning rate	1×10^{-3}
AdamW weight decay	1×10^{-4}
(L_2)*regularization weight*(λ_{reg})	1×10^{-6}
Maximum epochs	18
Early stopping patience	6 epochs
Graph encoder layers	2
Attention heads	4
Node hidden dimension	64
Graph hidden dimension	128
Graph output dimension	128
OD scoring vector dimension	128
Dropout rate	0.1
Route-pair batch size	512
Frequency-pair batch size	768
Weak-label pair batch size	1024
λ_{route}	1
λ_{freq}	0.05
Maximum(λ_{weak})	0.05
Weak-label warm-up	3 epochs

Weak-label ramp-up	3 epochs
Gradient clipping norm	2

In summary, this section develops an MR-GAT ranking model for OD priority ranking and provides a formal description of the graph encoding process, OD scoring mechanism, and pairwise ranking training strategy. Based on the proposed model framework, the next section further conducts empirical analyses to evaluate the performance of the framework in OD ranking and preliminary candidate screening.

5. Results and Discussion

5.1. Experimental Setup and Evaluation Metrics

This section evaluates the OD ranking performance of the MR-GAT ranking model for county-level short-haul route prioritization. Given the OD set E^{OD}, the model outputs a ranking score s_{ij} for each OD pair $(i,j) \in E^{\mathrm{OD}}$, and generates a complete ranking list by sorting all OD pairs in descending order according to their ranking scores. The candidate set contains 240,350 OD pairs. Therefore, the objective of the experimental evaluation is not to determine whether an individual OD pair should be operated as a route, but to examine whether known positive-label OD pairs can be concentrated at the top positions of the ranking list.

The experiments follow the data partitioning strategy defined in Section 3. The training set is used for parameter estimation, the validation set is used for model selection, the internal test set is used to report the main ranking performance, and the external evaluation set is used for additional validation. Unless otherwise specified, all results are calculated based on the ranking results of the complete OD set rather than sampled subsets. Let $Y \subset E^{\mathrm{OD}}$ denote the set of positive-label OD pairs used for evaluation. Let $\mathrm{Top}K$ represent the top-K OD pairs after sorting according to the ranking scores s_{ij}. The number of positive-label OD pairs retrieved within the top K positions is defined as

$$\mathrm{Hit}_Y(K) = \left|\mathrm{Top}_K \cap Y\right|, \tag{31}$$

where $\mathrm{Hit}_Y(K)$ represents the number of positive-label OD pairs included in the top-K ranking results, and $|\cdot|$ denotes the cardinality of a set. The corresponding top-K recall is defined as

$$\mathrm{Recall}_Y(K) = \frac{|\mathrm{Top}_K \cap Y|}{|Y|}, \tag{32}$$

where $\mathrm{Recall}_Y(K)$ represents the proportion of positive-label OD pairs retrieved within the top-K ranking results relative to the entire evaluation label set Y. This study reports top-K recall values for K = 100, 500, 1000, and 2000. To reduce the dependence on a fixed K relative to the candidate set size, recall values within the top 1% and top 5% ranking ranges are also reported. Let

$$K_\tau = \left\lceil \tau \left|E^{\mathrm{OD}}\right| \right\rceil, \qquad \tau \in \{0.01, 0.05\}, \tag{33}$$

$$\mathrm{Recall}_{Y,\tau} = \mathrm{Recall}_Y(K_\tau), \tag{34}$$

In addition, this study reports the best-rank percentile (BRP) to describe the relative position of the highest-ranked OD pair among the evaluation labels within the complete ranking list. Let $rank_{ij}$ denote the ranking position of OD pair (i,j) in the complete ranking list, where the ranking starts from 1. If $rank_{ij}$ = 1, OD pair (i,j) is ranked first. The best-rank percentile is defined as

$$\mathrm{BRP}_Y = \frac{\min_{(i,j)\in Y} rank_{ij}}{|E^{\mathrm{OD}}|}, \tag{35}$$

A smaller BRP_Y value indicates that at least one OD pair in the evaluation label set Y is ranked closer to the top of the complete ranking list. This metric complements top-K recall and reflects the ability of the model to identify positive-label OD pairs at the very front of the ranking list.

It should be noted that the training, validation, and internal test partitions refer to the partition of observed positive route labels rather than a disjoint partition of the complete candidate OD set. All models generate ranking scores for the same 240,350 candidate OD pairs. Among the observed route labels, 231 are used for constructing training supervision, 50 are held out for model selection, and 49 are reserved exclusively for internal evaluation.

To further characterize the uncertainty associated with the limited number of held-out positive route labels, the primary comparison between MR-GAT and XGBoost is additionally evaluated using paired bootstrap resampling. For each positive OD, the Top-K hit indicator is first aggregated across the ten fixed random seeds for each model. The same held-out positive ODs are then resampled with replacement for 10,000 bootstrap replications, and the difference in recall between MR-GAT and XGBoost is calculated in each replication. The 2.5th and 97.5th percentiles of the bootstrap distribution are reported as the 95% confidence interval. This paired design ensures that both models are compared on exactly the same held-out positive OD samples. The experimental setup is summarized in Table 4.

Table 4. Experimental setup overview.

Item	Setting
County units	2837 county units
Complete OD set	240,350 undirected OD pairs
Training positive route labels	231 route labels
Validation positive route labels	50 route labels
Internal test positive route labels	49 route labels
External evaluation positive labels	40 external evaluation labels
Weak-label training set	4983 airport catchment weak labels
Baseline models	Gravity score, XGBoost
Main model	MR-GAT ranking model
Model variants	MR-GAT with weak labels
Evaluation metrics	$\mathrm{Recall}_Y(K)$, BRP_Y
Random seeds	Primary MR-GAT/XGBoost comparison and relational ablation—10 fixed random seeds: 42, 2025, 3407, 7, 17, 73, 314, 777, 1024, 4096 Auxiliary weak-label and aviation-time sensitivity analyses—42, 2025, 3407

The subsequent results are organized into four levels. First, the overall ranking performance of the MR-GAT ranking model is presented. Second, comparisons with baseline models are conducted. Third, ablation analyses are performed to investigate the effects of weak supervision and relational graph structures. Finally, external evaluation results and the characteristics of top-ranked OD pairs are analyzed.

5.2. Overall Ranking Results of the MR-GAT Model

The experimental procedure is as follows. First, the model is trained and ranking scores are calculated using the same OD set and label partitions. Second, the optimal model parameters are selected according to validation-set performance, and the main ranking performance is evaluated using route labels in the internal test set. Third, external evaluation labels from general aviation routes and airport catchment weak labels are used for supplementary validation. Finally, the top-ranked OD pairs are analyzed in terms of label coverage, transport attributes, and spatial distribution. This procedure ensures that model comparison, component analysis, and top-ranked candidate analysis are conducted based on the same OD set and consistent evaluation criteria.

This section evaluates the overall ranking performance of the MR-GAT ranking model on the complete OD set. The model first estimates parameters using the training set and selects the optimal parameters according to validation-set performance. Then, ranking scores are generated for all 240,350 OD pairs, and a complete ranking list is produced by sorting the scores in descending order. The number of retrieved labels, top-K recall, and best-rank percentile within different top-ranked ranges are calculated to evaluate the model's ability to concentrate positive-label OD pairs at the top of the ranking list.

The ranking performance of the MR-GAT ranking model on the internal test set is presented in Table 5. The model generates a complete ranking list for 240,350 OD pairs and evaluates the ranking results using route labels and airport catchment weak labels as evaluation sets. The corresponding hit counts, top-K recall values, and best-rank percentile values are calculated.

Table 5. Detailed internal-test ranking results of the MR-GAT model under fixed random seed 42.

Metric	Route Labels	Airport Catchment Weak Labels
Number of evaluation labels	49	43
$\mathrm{Hit}_Y(100)$	0	2
$\mathrm{Hit}_Y(500)$	5	6
$\mathrm{Hit}_Y(1000)$	8	9
$\mathrm{Hit}_Y(2000)$	12	16
Top-100 recall	0.00%	4.65%
Top-500 recall	10.20%	13.95%
Top-1000 recall	16.33%	20.93%
Top-2000 recall	24.49%	37.21%
Top-1% recall	24.49%	41.86%
Top-5% recall	51.02%	81.40%
BRP_Y	0.07%	0.00%

For route labels, the fixed seed-42 model retrieves 5 internal test labels within the Top 500 OD pairs and retrieves 8 and 12 labels within the Top 1000 and Top 2000 OD pairs, respectively. The corresponding Top-1000 and Top-2000 recall values are 16.33% and 24.49%. For airport catchment weak labels, the model retrieves 2, 9, and 16 labels within the Top 100, Top 1000, and Top 2000 ranges, respectively. These detailed results are used to characterize one fixed ranking realization rather than the multi-seed average performance reported in the subsequent model comparison.

Overall, the MR-GAT ranking model is able to generate concentrated ranking results at the top of the complete OD set. The following sections further examine the relative performance and underlying mechanisms of these ranking results through baseline comparisons and ablation analyses.

5.3. Baseline Comparison

All ranking methods generate complete ranking lists based on the 240,350 OD pairs constructed in Section 3. To examine whether multi-relational graph structures improve the identification capability of high-priority OD pairs at the top of the ranking list, this study compares three types of methods: the Gravity score based on scale and distance decay, XGBoost based on OD attributes, and the proposed MR-GAT ranking model.

The Gravity score represents a heuristic ranking method based on traditional spatial interaction theory [18]. Its formulation is defined as

$$s_{ij}^{gravity} = \log(1+M_i)+\log\left(1+M_j\right) - \gamma \log\left(d_{ij}^{\mathrm{gc}}\right), \tag{36}$$

where M_i and M_j denote the scale variables of the two counties in the OD pair, d_{ij}^{gc} represents the great-circle distance between the centroids of the two counties, and γ denotes the distance decay coefficient. The Gravity score only considers scale and distance information and does not incorporate mobility information, ground transportation conditions, high-speed rail substitution effects, or multi-relational graph structures.

XGBoost is used as a feature-based machine learning baseline without graph structure propagation information [32]. It generates OD scores using the same supervised sample construction strategy as the training labels and ranks OD pairs according to the generated scores. The purpose is to examine whether node attributes and OD edge attributes alone can achieve effective ranking performance without introducing graph structural information. The ranking score is formulated as

$$s_{ij}^{\mathrm{xgb}} = f_{\mathrm{xgb}}\left(\mathbf{x}_i,\mathbf{x}_j,\mathbf{e}_{ij}\right), \tag{37}$$

where $\mathbf{x}_i$ and $\mathbf{x}_j$ represent the node attribute vectors of the two counties, $\mathbf{e}_{ij}$ denotes the OD attribute vector, and $f_{\mathrm{xgb}}(\cdot)$ represents the scoring function learned by XGBoost. XGBoost and MR-GAT use the same node attributes and OD attributes, but XGBoost does not utilize the graph structural information contained in the spatial proximity graph, mobility graph, and transport impedance graph. To maintain consistency with the ranking task of this study, XGBoost also generates a complete OD ranking list according to the ranking scores and is evaluated using the top-K recall metrics defined in Section 5.1.

Table 6 reports the ranking results of different models on the internal test set and external evaluation set. The internal test set uses route labels as the evaluation target, while the external evaluation set provides additional validation references. To further characterize the concentration of labels within the Top-1000 ranking range, this study introduces the Top-1000 enrichment factor:

$$EF(1000) = \frac{\mathrm{Recall}_Y(1000)}{1000/\left|E^{\mathrm{OD}}\right|}, \tag{38}$$

This metric represents the multiple of the model's label recall within the Top-1000 ranking range relative to the expected recall under random ranking.

Table 6. Comparison results between the proposed model and baseline methods.

Model	Internal Test Set			External Evaluation Set	
	Top-1000 Recall	Top-2000 Recall	Top-1000 Enrichment Factor	Top-1000 Recall	Top-2000 Recall
Gravity score *	2.00%	4.10%	4.9	0.00%	0.00%
XGBoost	7.55% ± 2.73%	12.65% ± 2.32%	18.1 ± 6.6	7.50% ± 2.36%	11.50% ± 3.76%
MR-GAT ranking model	16.33% ± 2.55%	22.65% ± 2.25%	39.2 ± 6.1	10.50% ± 1.97%	12.75% ± 2.49%

* Gravity is reported as a deterministic result, whereas XGBoost and MR-GAT are reported as mean ± standard deviation over ten fixed random seeds.

Across ten fixed random seeds, the MR-GAT ranking model achieves mean internal Top-1000 and Top-2000 recall values of 16.33% ± 2.55% and 22.65% ± 2.25%, respectively, compared with 7.55% ± 2.73% and 12.65% ± 2.32% for XGBoost. The Gravity score achieves corresponding recall values of 2.00% and 4.10%. Considering that the Top-1000 range accounts for only approximately 0.42% of the complete candidate set, these results indicate that MR-GAT concentrates substantially more held-out route labels near the front of the ranking than the two baseline methods.

The Gravity score considers only county scale and distance decay and therefore provides limited discrimination within the predefined 200–400 km candidate window. XGBoost substantially improves upon the Gravity score by incorporating county- and OD-level attributes, indicating that mobility intensity, ground-transport conditions, high-speed rail connectivity, and estimated aviation time contain useful ranking information. MR-GAT further improves the internal ranking performance by incorporating multi-relational network context in addition to these OD-level attributes.

Paired bootstrap analysis further supports the observed internal-test advantage of MR-GAT over XGBoost. The mean Top-1000 recall difference is 8.78 percentage points, with a 95% bootstrap confidence interval of [1.02, 17.35] percentage points. For Top-2000 recall, the mean difference is 10.00 percentage points, with a 95% confidence interval of [1.84, 18.98] percentage points. Both intervals remain above zero. In contrast, the corresponding external-evaluation confidence intervals include zero ([−2.50, 9.00] percentage points for Top-1000 recall and [−5.25, 7.50] percentage points for Top-2000 recall). Therefore, the internal route-label evaluation provides the primary evidence for model comparison, while the external general-aviation evaluation is interpreted as supplementary cross-scenario evidence.

5.4. Component Analysis: Weak Supervision and Relational Graph Ablation

The results in Table 6 show that the MR-GAT ranking model achieves better top-ranked identification performance on internal test route labels than the Gravity score and XGBoost. To further investigate the sources of model performance, this study conducts component analysis from two perspectives: supervision information and relational graph structure.

First, the influence of introducing airport catchment weak labels is examined to evaluate the contribution of weak supervision information to OD ranking. Second, while keeping other settings unchanged, the spatial proximity graph, mobility graph, and transport impedance graph are removed separately to evaluate the marginal contribution of the three relational graphs to ranking performance. All experiments use the same OD set, data partition, and evaluation metrics.

Because the weak-label variant is evaluated under the original three-seed protocol, Table 7 reports both models using seeds 42, 2025, and 3407 and is interpreted as an auxiliary supervision analysis rather than as the primary model comparison.

Table 7. Three-seed comparison of weak-label auxiliary supervision.

Model	Route Label Top-1000 Recall	Route Label Top-2000 Recall	Airport Catchment Weak Labels Top-1000 Recall
MR-GAT ranking model	17.70% ± 2.40%	24.50% ± 0.00%	1.10% ± 0.70%
MR-GAT with weak labels	17.00% ± 1.20%	22.40% ± 0.00%	2.30% ± 1.90%

The weak-label variant improves the top-1000 recall on airport catchment weak labels from 1.10% to 2.30%, but shows a slight decline when evaluated using route labels. This indicates that airport catchment weak labels can enhance the model's identification of existing airport service coverage conditions, but these coverage conditions are not completely consistent with the existing aviation service relationships reflected by route labels. In other words, airport catchment weak labels are more suitable as auxiliary supervision signals and explanatory information for service coverage conditions, rather than as a direct replacement for route labels.

Table 8 further presents the ablation results of the three relational graphs. The full model simultaneously incorporates the spatial proximity graph, mobility graph, and transport impedance graph. The ablation models remove one type of relational graph separately while keeping the training settings, data partition, and evaluation metrics unchanged. This experiment is used to evaluate the marginal contribution of different relational structures to the top-ranked OD performance.

Table 8. Ten-seed relational-graph ablation results on the internal test and external evaluation labels.

Model	Relational Graphs	Route Label Top-1000 Recall *	Route Label Top-2000 Recall	External Evaluation Label Top-1000 Recall	External Evaluation Label Top-2000 Recall
Full MR-GAT	spatial + mobility + transport	16.33% ± 2.55%	22.65% ± 2.25%	10.50% ± 1.97%	12.75% ± 2.49%
w/o spatial	mobility + transport	15.51% ± 1.72%	22.04% ± 1.88%	8.75% ± 1.77%	13.50% ± 2.93%
w/o mobility	spatial + transport	16.12% ± 2.63%	23.06% ± 1.68%	11.25% ± 2.12%	14.50% ± 3.07%
w/o transport	spatial + mobility	15.71% ± 1.68%	22.86% ± 1.61%	10.00% ± 2.04%	12.50% ± 1.67%

* Values are reported as mean ± standard deviation over ten fixed random seeds. Paired bootstrap comparisons are additionally conducted over the same held-out positive ODs. The external evaluation set is treated as supplementary cross-scenario evidence.

Across ten random seeds, the full MR-GAT model achieves the highest mean internal Top-1000 recall of 16.33%. Removing the spatial-proximity, mobility, and transport-impedance relations results in mean Top-1000 recall values of 15.51%, 16.12%, and 15.71%, respectively. However, these differences are relatively small compared with the observed seed-to-seed variability. Paired bootstrap analysis further shows that the 95% confidence intervals for all three internal comparisons between the full model and the relation-removal variants include zero. Therefore, the ablation results do not support a strict statistical ordering of the marginal importance of the individual relational graphs. Rather, they indicate that ranking performance is not dominated by a single relation and that different relational sources may contain partially overlapping contextual information.

The external evaluation shows a similar degree of uncertainty. In particular, removing the mobility relation yields slightly higher mean external recall than the full model. However, the corresponding differences are small and are not statistically distinguishable under the bootstrap analysis. Therefore, this result should not be interpreted as evidence that mobility information adversely affects external generalization. One possible additional explanation is that the external reference set consists of general-aviation routes, whose formation may depend on scenario-specific factors such as tourism-oriented demand, specialized transport tasks, demonstration operations, or local operating conditions and may therefore be less directly aligned with aggregate inter-county mobility than regular civil air-service links. Accordingly, the external evaluation is treated as supplementary evidence of cross-scenario generalization rather than as the primary basis for determining the importance of individual relations.

5.5. Front-End Enrichment and Transport Profiles

Based on the model comparison and component analysis presented above, this study applies the MR-GAT ranking model to rank all 240,350 OD pairs and further investigates the label coverage and transportation characteristics of top-ranked OD pairs. The ranking profiles presented in this section are generated using the fixed seed-42 model run, consistent with the detailed results reported in Table 5, whereas the primary model comparison in Table 6 is based on results aggregated over ten random seeds. Therefore, the following analysis is intended to describe the structural and transportation characteristics of one fixed ranking realization rather than the average ranking across repeated runs.

Table 9 presents the label hits and transportation characteristics across different top-ranked sets. As the ranking range expands from Top 100 to Top 5%, the number of hits for route labels, external evaluation labels, and airport catchment weak labels gradually increases. Within the Top 500 OD pairs, the model retrieves 5 route labels and 4 external evaluation labels. Within the Top 1000 OD pairs, the model retrieves 8 route labels and 5 external evaluation labels. Considering that the Top 1000 ranking range accounts for only 0.42% of the complete OD set, these results indicate that the model achieves a certain degree of label enrichment at the top of the ranking list.

Table 9. Label hits and transportation characteristics across different ranking ranges.

Top Set	Cand. Share	Test Hits	External Hits	Catchment-Label Hits	Avg. GCD/km	Ground h	Air Proxy h	Gap h	HSR Share
Top 100	0.04%	0	1	44	348.88	6.73	1.09	5.64	0
Top 500	0.21%	5	4	185	341.96	6.45	1.07	5.38	0.018
Top 1000	0.42%	8	5	305	330.28	6.22	1.03	5.19	0.03
Top 2000	0.83%	12	7	538	326.07	5.96	1.02	4.94	0.047
Top 1%	1.00%	12	8	618	325.15	5.86	1.02	4.84	0.052
Top 5%	5.00%	25	16	1939	319.45	5.32	1	4.32	0.073

From the perspective of transportation characteristics, the average ground travel times of the Top 100, Top 500, and Top 1000 OD pairs are 6.73 h, 6.45 h, and 6.22 h, respectively, which are substantially higher than the corresponding estimated air travel time values. For the Top 1000 OD pairs, the average numerical difference between ground travel time and estimated air travel time is 5.19 h. This difference should not be interpreted as an actual passenger door-to-door time saving from air transport. The ground and aviation time variables are constructed under different mode-specific measurement conventions and do not consistently include access, waiting, terminal, and other components of the complete travel chain. Rather, the observed difference indicates that high-ranking OD pairs tend to combine relatively high road-network travel impedance with relatively short estimated airborne travel time. As the ranking range expands to the Top 5%, both the average ground travel time and the corresponding time difference decrease. This pattern suggests that the model tends to prioritize OD pairs with relatively unfavorable ground-transport conditions, rather than simply the geographically shortest connections.

To examine whether the simplified aviation-time specification materially affects the ranking results, an additional sensitivity analysis was conducted by uniformly adding 0.5 h, 1.0 h, and 1.5 h to the estimated air travel time, corresponding to progressively more conservative assumptions regarding non-cruise aviation time. Using the three original random seeds (42, 2025, and 3407), the mean internal Top-1000 and Top-2000 recall remained unchanged at 17.69% and 24.49%, respectively, under all three scenarios. The Top-1000 candidate-set overlap with the original ranking remained between 99.63% and 99.73%, while the Spearman rank correlation exceeded 0.99996 in all cases. These results indicate that the resulting ranking is highly insensitive to uniform additions of up to 90 min to the estimated air travel time. This sensitivity analysis does not convert the aviation-time measure into a complete door-to-door travel-time measure; rather, it shows that the main ranking results are robust to substantial uniform perturbations of this simplified time proxy.

The proportion of high-speed rail connectivity further reflects the transportation substitution conditions of top-ranked OD pairs. No high-speed rail-connected OD pairs are identified within the Top 100 set. The proportion of high-speed rail-connected OD pairs is 0.030 within the Top 1000 set and only 0.073 within the Top 5% ranking range. These results indicate that high-ranking OD pairs are more likely to occur in county-level connections with limited direct high-speed rail connectivity or insufficient high-speed ground transportation alternatives. Since the high-speed rail variable in this study is mainly constructed based on train service identification, it does not fully incorporate station access conditions, service frequency, ticket prices, or transfer conditions. Therefore, this result only indicates that top-ranked OD pairs have relatively low direct high-speed rail connectivity and should not be interpreted as a final assessment of comprehensive air–rail competition.

Airport catchment weak labels also show a certain degree of concentration at the top of the ranking list. Among the Top 1000 OD pairs, 305 airport catchment weak labels are retrieved, while 1939 are retrieved within the Top 5% ranking range. This indicates that after ranking based on route labels, mobility relationships, and transport impedance relationships, the model can still identify a subset of OD pairs with existing airport service coverage conditions. This result complements the weak-label analysis in Section 5.4: airport catchment weak labels are not suitable as direct substitutes for route labels, but they can provide auxiliary information for interpreting the service coverage characteristics of top-ranked OD pairs.

Overall, the MR-GAT ranking model places OD pairs with existing interaction foundations, relatively high ground transportation impedance, low direct high-speed rail connectivity, and certain airport accessibility conditions at the top of the ranking list. From a planning perspective, the value of this result is not to directly determine whether a route should be operated, but to reduce more than 240,000 OD pairs into a smaller priority evaluation set. This provides a basis for subsequent assessment incorporating takeoff and landing infrastructure, airspace constraints, aircraft capabilities, operational costs, and local demand conditions.

5.6. Spatial Patterns and Representative ODs

Based on the analysis of label retrieval and transportation characteristics at the top of the ranking list, this section further examines the spatial distribution of high-priority OD pairs. Figure 2 presents the spatial distribution of the Top 1000 OD pairs and the frequency of endpoint provinces. Overall, high-ranking OD pairs are not uniformly distributed across space but instead show clear regional concentrations. These clusters are mainly observed in regions including Xinjiang, Yunnan, Inner Mongolia, areas surrounding the Qinghai–Tibet Plateau, southern Northeast China, the Shandong Peninsula, Guangdong–Guangxi region, and Sichuan.

To illustrate the label coverage of the Top 1000 OD pairs, Figure 2 classifies the ranked OD pairs into four categories: route-label hits, external evaluation label hits, airport catchment weak-label hits, and unmatched OD pairs. The first three categories indicate that the corresponding OD pairs are included in the internal test route labels, external evaluation labels, and airport catchment weak labels, respectively. The unmatched category only indicates that an OD pair is not covered by the current label system and does not imply that it lacks potential value for short-haul air services.

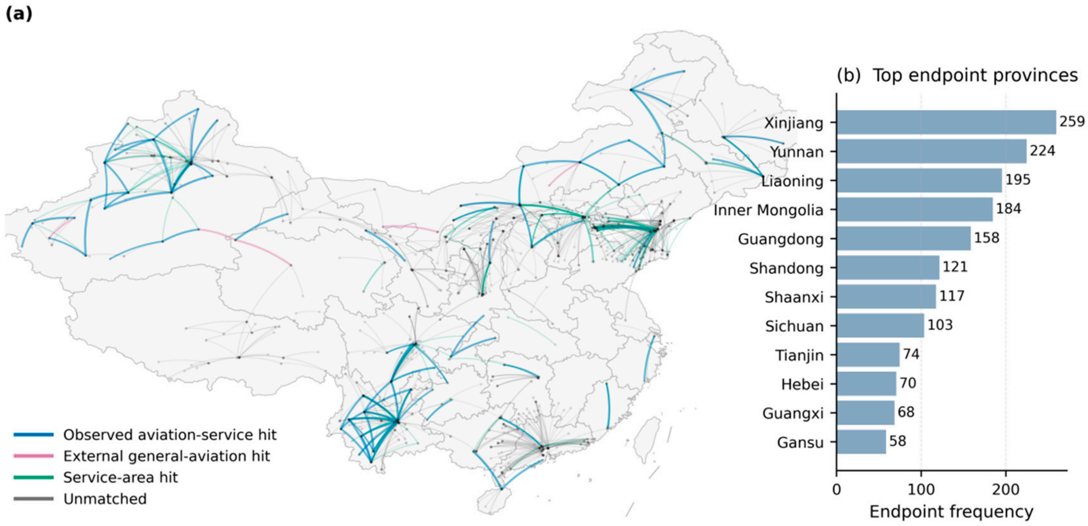

Figure 2. Spatial distribution and endpoint provinces of the Top 1000 ODs: (**a**) spatial distribution of the Top 1000 ODs, classified as observed aviation-service hits, external general-aviation hits, service-area hits, and unmatched ODs; (**b**) provinces with the highest endpoint frequencies among the Top 1000 ODs.

From the perspective of spatial distribution, the Top 1000 OD pairs show concentrations in western regions, border areas, regions surrounding high-altitude plateaus, and areas with intensive regional interactions. Combined with the transportation profile analysis in Section 5.5, high-ranking OD pairs in Xinjiang, Yunnan, Inner Mongolia, and areas surrounding the Qinghai–Tibet Plateau are often associated with longer ground travel times, limited direct high-speed rail connectivity, or substantial ground detour effects. In contrast, high-ranking OD pairs in the Shandong Peninsula, Guangdong–Guangxi region, Sichuan, and southern Northeast China mainly reflect strong regional interaction foundations and cross-regional connectivity demands. The endpoint province frequency shown in Figure 2 further indicates that Xinjiang, Yunnan, Liaoning, Inner Mongolia, Guangdong, Shandong, Shaanxi, and Sichuan have relatively high occurrence frequencies among the Top 1000 OD pairs. This suggests that highly ranked OD pairs are not confined to remote areas with accessibility constraints, but also occur in regions characterized by relatively strong inter-county interaction intensity.

To facilitate interpretation of the practical implications of top-ranked OD pairs, this study further provides a descriptive classification of selected high-priority OD pairs based on label coverage and major transportation characteristics. It should be emphasized that these categories are not directly generated by the model, but are post hoc classifications derived from ranking results, label coverage, and transportation characteristics. Table 10 presents several representative high-priority OD pairs and categorizes them into four groups: route-label-hit type, ground-transport compensation type, airport service coverage type, and enhanced regional interaction type.

Table 10. Representative high-priority OD pairs.

Type [1]	OD	Rank	GCD/km	Ground h	Air Proxy h	HSR	Flow	Label
Observed hit	Aral City–Korla City	2	389.28	7.03	1.22	0	503.5	route + catchment
Observed hit	Hotan City–Aksu City	3	370.04	7.24	1.16	0	572.2	route + catchment
Observed hit	Yining City–Karamay District	8	319.67	6.99	1	0	338.4	route + catchment
Ground-compensation	Golmud City–Amdo County	5	308.94	25.06	0.97	0	2797.3	unmatched [2]
Ground-compensation	Hami City–Subei Mongolian Autonomous County	12	313.6	6.03	0.98	0	1021.4	unmatched
Ground-compensation	Golmud City–Qumalai County	14	249.28	8.95	0.78	0	926.7	unmatched
Catchment-label hits	Kuqa City–Shayibake District of Urumqi City	1	393.89	8.98	1.23	0	620	catchment
Catchment-label hits	Shihezi City–Yining City	4	381.63	6.5	1.19	0	459.5	catchment
Mobility-enhanced	Mengla County–Guandu District of Kunming City	7	397.29	7.45	1.24	0	721.3	unmatched
Mobility-enhanced	Xifeng District of Qingyang City–Chengguan District of Lanzhou City	19	344.99	5.61	1.08	0	580.9	unmatched

[1] The categories in the Type column are descriptive post hoc classifications rather than model-generated classes. [2] Unmatched indicates only that the OD is not covered by the current label system and does not imply the absence of potential short-haul air-service value.

Table 10 shows that the top-ranked OD pairs include not only OD pairs matching route labels, but also OD pairs matching airport catchment weak labels and unmatched OD pairs that are not covered by the current label system. For route-label-hit OD pairs, OD pairs such as Aral City–Korla City, Hotan City–Aksu City, and Yining City–Karamay District are ranked among the top positions, indicating that the model can identify some route-label OD pairs that are not included in the training set.

OD pairs such as Kuqa City–Shayibake District of Urumqi City and Shihezi City–Yining City match airport catchment weak labels. This indicates that these connections have certain airport service coverage conditions; however, this label does not imply that these OD pairs have already formed actual air routes.

For unmatched high-ranking OD pairs, this study does not interpret them as direct route-opening recommendations. Instead, they are regarded as candidate objects that are not covered by the current label system but receive relatively high ranking scores from the model. For example, OD pairs such as Golmud City–Amdo County and Golmud City–Qumalai County exhibit long ground travel times and large differences between ground travel time and estimated aviation travel time, indicating strong ground transportation compensation characteristics. OD pairs such as Mengla County–Guandu District of Kunming City and Xifeng District of Qingyang City–Chengguan District of Lanzhou City show certain population mobility foundations and represent enhanced regional interaction characteristics. These OD pairs still require further evaluation considering takeoff and landing infrastructure, operational organization, local demand conditions, and cost–benefit factors.

Overall, high-priority OD pairs do not simply correspond to county pairs with the largest population scale, shortest distance, or strongest existing airport coverage conditions. Instead, the observed ranking patterns reflect a combination of inter-county mobility, ground-transport impedance, and the difference between ground travel time and the estimated air-time proxy. The spatial distribution and representative OD analysis further indicate that the ranking results may reflect both strong ground transportation compensation demand and existing regional interaction foundations. Therefore, the OD ranking results should be interpreted as preliminary inputs for subsequent market research and operational feasibility assessment rather than direct route-opening decisions.

5.7. Limitations and Future Research

This study has several limitations. First, route labels, airport catchment weak labels, and external evaluation labels cannot directly represent future demand for low-altitude aviation or emerging short-haul air services. They only provide indirect references based on existing civil aviation routes, observed airport service coverage relationships, and selected general aviation routes. Second, the travel-time variables used in this study are simplified, mode-specific representations of transport impedance rather than complete door-to-door travel times. Ground travel time is based on road-network routing between county-level representative locations, while high-speed rail travel time mainly represents train running time and does not consistently incorporate station access, waiting, transfer, and egress processes. The estimated air travel time is calculated from great-circle distance and an assumed cruise speed and therefore does not include access and egress travel, terminal processing, waiting time, security procedures, takeoff and landing processes, turnaround time, or weather effects. The aviation-time variable is consequently more idealized than the observed road and rail time measures. For this reason, the relative time indicators and the differences reported in this study should be interpreted as screening-oriented transport-impedance measures rather than actual passenger door-to-door time savings or a complete assessment of modal competitiveness. A more comprehensive multimodal comparison would require consistent door-to-door generalized travel-time and cost data for all transport modes. Third, this study does not evaluate operational and economic feasibility. Whether an OD pair can support actual air service also depends on aircraft range capability, takeoff and landing facilities, operational support capacity, route organization, and cost–benefit conditions. Fourth, the number of positive labels available for evaluation remains limited. The internal test set contains 49 positive route labels, while the external evaluation set contains 40 positive general-aviation OD labels. Although the ten-seed evaluation and paired bootstrap analysis provide additional characterization of result uncertainty, the statistical evidence remains conditional on the currently available observed aviation relationships. Therefore, the reported performance differences should be interpreted within the scope of the present evaluation sets and should not be generalized to broader future route-demand conditions without additional operational observations.

Future research can be extended in two directions. First, more comprehensive aviation operational data can be incorporated, including flight frequency, passenger volume, ticket prices, load factors, access time, waiting time, service frequency, and travel costs, to improve the explanatory capability of labels and impedance-related variables. Second, the OD ranking results can be further integrated with takeoff and landing location planning, airspace conditions, aircraft configuration, and cost–benefit analysis to establish a complete workflow from preliminary candidate screening to operational feasibility evaluation. Future work may also investigate adaptive gradient- or loss-balancing strategies when additional heterogeneous supervision signals are jointly incorporated.

6. Conclusions

This study develops a multi-relational graph ranking framework for county-level candidate short-haul route screening in the early development stage of low-altitude passenger services. Unlike existing studies focusing on air passenger demand forecasting or individual route opening decisions, this study formulates the research task as OD priority ranking, aiming to identify county-level OD pairs that deserve further investigation in market analysis, route assessment, takeoff and landing site planning, and operational feasibility evaluation. Based on 2837 county units and 240,350 undirected OD pairs within the 200–400 km distance range, this study integrates county node attributes, population mobility, ground transportation impedance, high-speed rail substitution conditions, estimated aviation travel time variables, and proxy aviation labels. Three relational graphs, including the spatial proximity graph, mobility graph, and transport impedance graph, are constructed, and the MR-GAT ranking model is applied to generate OD priority rankings.

The ten-seed experimental results show that the MR-GAT ranking model achieves higher internal top-ranked route-label retrieval performance than the Gravity and XGBoost baselines. Paired bootstrap analysis further supports the internal recall advantage of MR-GAT over XGBoost under the current held-out sample. In contrast, the relational-graph ablation analysis shows relatively small differences among the full and relation-removal configurations, and the corresponding bootstrap confidence intervals do not support a strict statistical ordering of the marginal importance of individual relations. These results suggest that the value of the multi-relational framework lies in integrating multiple sources of contextual information rather than relying on a single dominant relational structure.

The aviation-time sensitivity analysis further shows that the main ranking results remain highly stable when 30–90 min are uniformly added to the estimated air travel time. This robustness does not imply that the variable represents complete door-to-door aviation travel time, but indicates that the ranking is not driven by the specific simplified aviation-time assumption adopted in the baseline setting. The external general-aviation evaluation shows greater uncertainty than the internal route-label evaluation and is therefore interpreted as supplementary cross-scenario evidence. Airport catchment weak labels provide auxiliary information for identifying OD pairs with existing airport service coverage conditions, but they do not consistently improve route-label ranking performance and therefore should not replace observed route labels as the primary supervision source.

Overall, the experimental results support the research objective proposed in the Introduction. First, the proposed framework can generate differentiated priority rankings among a large number of county-level OD pairs and concentrate a portion of held-out route labels near the front of the ranking list. Second, the comparison with non-graph baselines indicates that incorporating multi-relational network context can improve the retrieval of observed regular air-service relationships beyond the use of OD-level attributes alone. Third, the ablation results indicate that the contribution of individual relational graphs should be interpreted cautiously because their marginal effects are small relative to the uncertainty observed in the current evaluation. Therefore, the empirical evidence supports the usefulness of multi-relational integration as a whole rather than a strict importance ordering among the spatial, mobility, and transport-impedance relations.

It should be emphasized that the ranking results only represent the relative priority of OD pairs for subsequent research and evaluation, and do not indicate passenger demand levels, route opening probabilities, or operational feasibility. Therefore, the proposed multi-relational graph ranking framework is more appropriately regarded as a preliminary identification tool within the low-altitude passenger service planning process. It can reduce the candidate search space and determine priority objects for further investigation, while subsequent assessments should incorporate aircraft performance, infrastructure conditions, airspace operation constraints, cost–benefit considerations, and market demand analysis. More broadly, the framework illustrates how multi-source mobility and transport-system data can be organized into an analytical screening tool for planning problems in which the number of potential connections substantially exceeds the number of directly observed services. The statistical evidence reported in this study is also conditional on the limited number of currently observed positive aviation relationships.

Author Contributions: Conceptualization, Y.W. (Yu Wang), X.L. and J.C.; Methodology, Y.W. (Yu Wang) and J.C.; Software, Y.W. (Yu Wang), X.J. and J.L.; Validation, Y.W. (Yu Wang); Data curation, X.J., Y.W. (Yixu Wang) and J.L.; Writing—original draft, Y.W. (Yu Wang); Writing—review & editing, Y.W. (Yu Wang), X.L. and J.C.; Visualization, Y.W. (Yixu Wang); Project administration, X.L. and J.C.; Funding acquisition, X.L. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by the National Natural Science Foundation of China, grant number 62206016; the Guangdong Basic and Applied Basic Research Foundation, grant numbers 2023A1515140086 and 2022A1515140016; the Fundamental Research Funds for the Central Universities, grant number FRF-TP-24-062A; the Project "Vice President of Science and Technology" of Changping District, Beijing, grant number 2025040040042; the Foundation of Engineering Research Center of Intelligence Perception and Autonomous Control, Ministry of Education, China; and the Open Project of the State Key Laboratory of Digital Manufacturing Equipment and Technology, Huazhong University of Science and Technology, grant number DMETKF2021023.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The data presented in this study are available from the corresponding author upon reasonable request. Some of the raw data were obtained from third-party or publicly available data sources and may be subject to their respective terms of use and redistribution restrictions.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Federal Aviation Administration. The eIPP: What You Need to Know. 9 March 2026. Available online: https://www.faa.gov/newsroom/eipp-what-you-need-know (accessed on 29 July 2026).
2. State Council of the People's Republic of China. *Report on the Work of the Government*; State Council of the People's Republic of China: Beijing, China, 2024; Official English text.
3. Li, M.; Birolini, S.; Ribeiro, N.A. Hub competition and demand modeling in long-haul aviation markets: A forward-looking scenario-based game-theoretical approach. *Transp. Res. Part E Logist. Transp. Rev.* **2026**, *211*, 104894. https://doi.org/10.1016/j.tre.2026.104894.
4. Muros Anguita, J.G.; Díaz Olariaga, O. Air cargo transport demand forecasting using ConvLSTM2D, an artificial neural network architecture approach. *Case Stud. Transp. Policy* **2023**, *12*, 101009. https://doi.org/10.1016/j.cstp.2023.101009.
5. Birolini, S.; Avogadro, N.; Malighetti, P.; Paleari, S. Con-Accessibility: Logit-based catchment area modeling for strategic airport system planning. *Transp. Res. Part A Policy Pract.* **2024**, *190*, 104270. https://doi.org/10.1016/j.tra.2024.104270.
6. Sun, F.; Yin, H.; Sun, X.; Wang, X.; Zhang, Y. Resilience analysis of cities' air accessibility under disruptions. *J. Air Transp. Manag.* **2024**, *120*, 102671. https://doi.org/10.1016/j.jairtraman.2024.102671.
7. Fan, Z.; Pei, M.; Chen, P.; Wei, Z.; Zhong, L. Optimizing urban transport: An optimization framework for urban air mobility site selection and route operation under real-world travel demand. *Transp. Res. Part E Logist. Transp. Rev.* **2026**, *209*, 104758. https://doi.org/10.1016/j.tre.2026.104758.
8. Fu, M.; Othman, M.M.; Moeckel, R.; Hornung, M.; Antoniou, C. User preferences for regional air mobility: Insights from Germany. *J. Air Transp. Manag.* **2026**, *132*, 102934. https://doi.org/10.1016/j.jairtraman.2025.102934.
9. Wang, Z.; Delahaye, D.; Farges, J.-L.; Alam, S. Complexity optimal air traffic assignment in multi-layer transport network for urban air mobility operations. *Transp. Res. Part C Emerg. Technol.* **2022**, *142*, 103776. https://doi.org/10.1016/j.trc.2022.103776.
10. Chu, Y.; Ai, Y.; Wu, Z.; Zhang, H. Study on travel choice behavior of medium and long distance intercity passengers in the context of air-rail competition. *Transp. Policy* **2025**, *170*, 163–177. https://doi.org/10.1016/j.tranpol.2025.05.011.
11. Jiang, S.; Cai, C.; Xiao, L. Intercity accessibility and equity: Empirical study of high speed rail and air travel network in China. *Case Stud. Transp. Policy* **2023**, *13*, 100995. https://doi.org/10.1016/j.cstp.2023.100995.
12. Yu, L.; Li, M.; Dai, Z.; Cui, M.; Levinson, D. Access door-to-door: An intercity efficiency and distributional analysis of the costs of travel by plane, train, and automobile. *Transp. Policy* **2026**, *184*, 104188. https://doi.org/10.1016/j.tranpol.2026.104188.

13. Signori, A.; Sun, X.; Birolini, S. City-pair air connection under high-speed railway development: Simulation of air travel demand affected by the new Chinese high-speed rail lines opening in 2030. *Transp. Res. Part A Policy Pract.* **2026**, *207*, 104924. https://doi.org/10.1016/j.tra.2026.104924.
14. Zhang, X.; Loo, B.P.Y.; Wang, L.; Wei, S. Disparities in high-speed rail accessibility in the Yangtze River Delta, China: A door-to-door travel time perspective. *Res. Transp. Bus. Manag.* **2025**, *59*, 101308. https://doi.org/10.1016/j.rtbm.2025.101308.
15. de Rolland, C.; Bayart, C. From zonal aggregated mobile phone data to origin-destination matrices: Revisiting methodologies developed for traffic counts data. *Transp. Res. Part A Policy Pract.* **2025**, *202*, 104715. https://doi.org/10.1016/j.tra.2025.104715.
16. Zhang, X.; Wang, X.; Sun, L.; Miranda-Moreno, L. Reconstruction of OD demand using mobile phone data: A regression model with spatially varying coefficients. *J. Transp. Geogr.* **2026**, *134*, 104677. https://doi.org/10.1016/j.jtrangeo.2026.104677.
17. Verma, R.; Ukkusuri, S.V. What determines travel time and distance decay in spatial interaction and accessibility? *J. Transp. Geogr.* **2025**, *122*, 104061. https://doi.org/10.1016/j.jtrangeo.2024.104061.
18. Wilson, A.G. A family of spatial interaction models, and associated developments. *Environ. Plan. A* **1971**, *3*, 1–32. https://doi.org/10.1068/a030001.
19. Chung, H.M.; Kwon, O.K.; Han, O.S.; Kim, H.-J. Evolving network characteristics of the Asian international aviation market: A weighted network approach. *Transp. Policy* **2020**, *99*, 299–313. https://doi.org/10.1016/j.tranpol.2020.09.002.
20. Sun, X.; Wandelt, S.; O'Kelly, M. From structure to function: A comprehensive benchmark on the estimation of hubs in air transportation through complex network metrics. *J. Transp. Geogr.* **2026**, *131*, 104535. https://doi.org/10.1016/j.jtrangeo.2025.104535.
21. Wong, C.W.H.; Cheung, T.K.Y.; Zhang, A. A connectivity-based methodology for new air route identification. *Transp. Res. Part A Policy Pract.* **2023**, *173*, 103715. https://doi.org/10.1016/j.tra.2023.103715.
22. Veličković, P.; Cucurull, G.; Casanova, A.; Romero, A.; Liò, P.; Bengio, Y. Graph attention networks. In Proceedings of the International Conference on Learning Representations (ICLR), Vancouver, BC, Canada, 30 April–3 May 2018.
23. Tong, Y.; Chen, S.; Yang, Y.; Liu, P.; Tang, J.; Xu, C.; Luan, T. ADIG-Net: Spatiotemporally decoupled dynamic graph learning for efficient traffic forecasting. *Knowl.-Based Syst.* **2026**, *345*, 116133. https://doi.org/10.1016/j.knosys.2026.116133.
24. Zheng, H.; Wang, Z.; Zheng, C.; Wang, Y.; Fan, X.; Cong, W.; Hu, M. A graph multi-attention network for predicting airport delays. *Transp. Res. Part E Logist. Transp. Rev.* **2024**, *181*, 103375. https://doi.org/10.1016/j.tre.2023.103375.
25. Burges, C.J.C.; Shaked, T.; Renshaw, E.; Lazier, A.; Deeds, M.; Hamilton, N.; Hullender, G.N. Learning to rank using gradient descent. In Proceedings of the 22nd International Conference on Machine Learning, Bonn, Germany, 7–11 August 2005; pp. 89–96. https://doi.org/10.1145/1102351.1102363.
26. Liu, T.-Y. Learning to rank for information retrieval. *Found. Trends Inf. Retr.* **2009**, *3*, 225–331. https://doi.org/10.1561/1500000016.
27. Xu, M.; Xiang, J.; Xie, Z.; Meng, X. Learning to rank critical road segments via heterogeneous graphs with origin-destination flow integration. *Inf. Process. Manag.* **2026**, *63*, 104702. https://doi.org/10.1016/j.ipm.2026.104702.
28. Lee, K.K.; Abdelghany, A.F. Beyond route-specific forecasting: An empirical test of two cross-series transfer learning strategies for airline demand with short-data constraints. *J. Air Transp. Manag.* **2026**, *133*, 102981. https://doi.org/10.1016/j.jairtraman.2026.102981.
29. Rendle, S.; Freudenthaler, C.; Gantner, Z.; Schmidt-Thieme, L. BPR: Bayesian personalized ranking from implicit feedback. In Proceedings of the Twenty-Fifth Conference on Uncertainty in Artificial Intelligence, Montreal, QC, Canada, 18–21 June 2009; pp. 452–461.
30. Brody, S.; Alon, U.; Yahav, E. How attentive are graph attention networks? In Proceedings of the International Conference on Learning Representations (ICLR), Virtual, 25–29 April 2022.
31. Loshchilov, I.; Hutter, F. Decoupled weight decay regularization. In Proceedings of the International Conference on Learning Representations (ICLR), New Orleans, LA, USA, 6–9 May 2019.
32. Chen, T.; Guestrin, C. XGBoost: A scalable tree boosting system. In Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, San Francisco, CA, USA, 13–17 August 2016; pp. 785–794. https://doi.org/10.1145/2939672.2939785.