Correlation Analysis Between Rail Network Nodal Attributes and Passenger Staying Rate: A Case Study of China's Yangtze River Delta Region

Zhoulanyi Xing[1], Junting LIN[1], enshan CHEN[2]
Affiliation: 1 Tongji University, 2 Delft University of Technology
Email:2410328@tongji.edu.cn

Abstract
The development of railway station areas hinges on their transition from "transfer hubs" to "destinations," where the scale and structure of passenger flows determine how built environments respond to traveler needs. Traditional node-place models often characterize station node attributes using volatile metrics like passenger volume or train frequency. This study selects 49 railway stations in China's most economically developed region, constructing a multilayer L-space rail complex network model from timetable data to investigate how nodal attributes across distinct rail networks influence Passenger Staying Rate and Passenger Structure. Key findings reveal that the Economically-Weighted PageRank (integrating city GDP) in the Integrated Rail Network exhibits significant correlations with all passenger-related metrics. Ultimately, a "Node-Passenger" analytical framework is established, offering new perspectives for railway station area development.

Keywords: Station areas; multi-scale accessibility; multi-centrality analysis; node-place model

1. Introduction

The emergence of high-speed rail (HSR) technology heralds the advent of the "Second Railway Era," further compressing spatiotemporal distances between cities. This acceleration has intensified intercity rail passenger flows, significantly boosting urban economic growth and regional integration. Within this context, as noted by Professor Edwards of the University of Sheffield School of Architecture in The Modern Station: New Approaches to Railway Architecture, "Railway stations stand as integral elements of urban life and culture. Their unique significance stems from serving as interfaces between 'two worlds'—the convergence points of railway systems and urban ecosystems."(Edwards, 2013). The railway system, a network of constant flow, has railway passenger stations as its nodes. In contrast, the urban system is where people live and conduct various activities, functioning as a place. Nodes represent movement and transfer, while places imply stays and activities. As a key public space in cities, the development of railway station areas hinges on transforming stations from mere transfer points to destinations for diverse social interactions. This shift not only saves time for time-sensitive passenger groups like business and commuter travelers but also reduces overall transportation costs and alleviates environmental pollution from urban congestion and car dependency (Carlton, 2009).

Since the launch of its high-speed rail system in 2008, China has witnessed the rapid development of its high-speed rail network. However, there remain significant differences between China's railway passenger structure and that of Europe and Japan. In China, long-distance travelers constitute a larger proportion, whereas the number of commuters and business travelers is relatively smaller. Variations in passenger volumes and the composition of passenger types at different stations place diverse demands on the built environment of station areas. As conduits for transporting various passenger flows, is there a relationship between the configuration of different railway networks and the types of passenger groups they serve? Can this relationship be observed at the meso-scale level of station areas?

To answer these questions, Section 2 provides a review of relevant literature, Section 3 outlines the methodological approach, data sources, and study scope, and Section 4 presents and analyzes the research findings. Finally, Section 5 concludes the study.

2. Literature Review

The classic Node-Place model proposed by Bertolini(1999), has been widely used in station typology research and in analyzing development strategies for station areas—that is, assessing whether a station area should further develop its "place" functions or enhance its "node" functions. This model has been extensively applied around the world, including in the Netherlands, Japan, and Australia(Chorus and Bertolini, 2011; Kamruzzaman et al., 2014; Peek et al., 2006; Zemp et al., 2011) Internationally, transit oriented development (TOD. It is not only used to analyze railway station areas but also stations along urban rail transit and bus rapid transit (BRT) corridors(Amini Pishro et al., 2022; Pezeshknejad et al., 2020)Transit-Oriented Development (TOD.

The model often evaluates the strength of a node based on quantitative indicators, such as the number of train services or passenger volumes. This approach works well in regions where differences in passenger types are minimal. However, in China, long-distance travelers still dominate the railway system overall. Yet, in economically developed regions, the proportion of business and commuting passengers has been increasing, thanks to the rapid expansion of high-speed and intercity rail in recent years. These two groups exhibit distinct behavioral patterns, travel choices, functional needs, and sensitivity to time. Therefore, whether passenger volume alone can accurately reflect the “node value” of a station remains questionable in the Chinese context.

On the other hand, passenger volume and train frequency are values that are easily influenced by the built environment and are subject to fluctuations over time. Changes in the built environment inherently imply potential shifts in passenger flow. For newly constructed stations or those undergoing major redevelopment, it is difficult to accurately predict future passenger volumes or train frequencies, which, to some extent, limits the applicability of the Node-Place model.

However, as the conduit for passenger flows, the structure and characteristics of a transport network influence the ease of movement within it. A station’s node attributes within different types of networks may offer a more stable proxy for the “node value.” Yet, most existing studies on railway complex networks typically treat cities rather than individual railway stations as nodes, making them of limited value when addressing development issues at the meso-scale of station areas.

At the “place” level, traditional Node-Place models often use indicators such as the number and diversity of employees across industries (Zemp et al., 2011)which focuses on transport related issues, blending performance with preconditions at a given site. We argue that a classication system for strategic planning should focus on the demands and conditions of the site within which the railway station must function, i.e. system context. Here, we present such a classication system: a cluster analysis of the 1700 Swiss railway stations relying solely on context factors. The resulting classes vary primarily in density (of land use and transport services, rental prices of various types of buildings, or the number of public facilities (Amini Pishro et al., 2022). However, these indicators are not direct measures of the extent to which a station area serves as a true “destination” for travelers. In recent years, with the advancement of big data technologies, Location-Based Services (LBS) data have made it possible to directly determine whether travelers remain in the station area after alighting, and what activities they engage in, allowing for the inference of passenger types. Such indicators can serve as direct performance metrics for evaluating the degree to which a station area functions as a destination.Therefore, this study employs complex network analysis techniques to quantify the centrality of stations within different network structures and explores the relationship between network centrality and both passenger volume and passenger composition using LBS data. Finally, the study attempts to construct a Node-Place model based on these two dimensions, offering a new perspective for analyzing and understanding station area development.

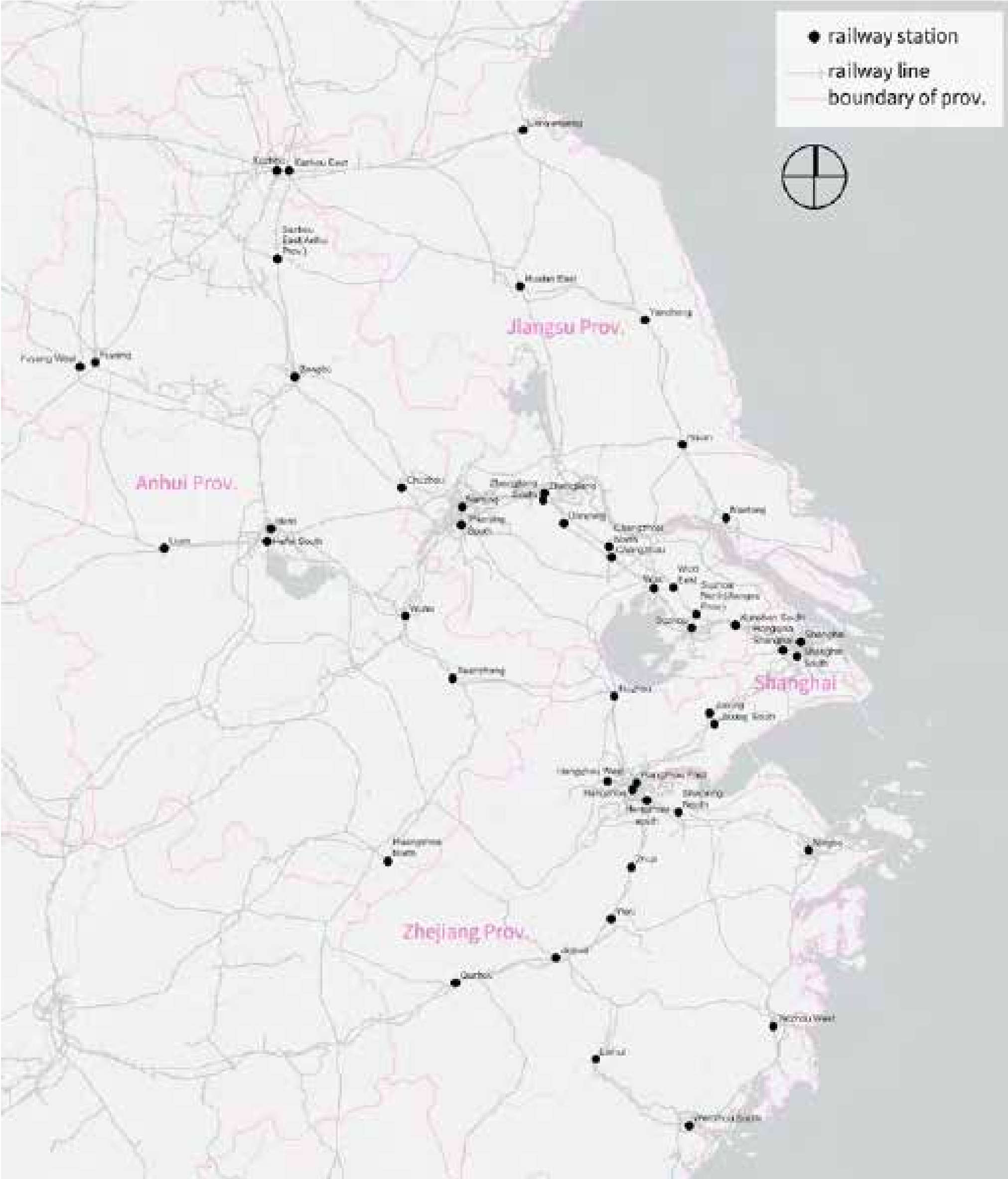

Figure 1. The research cases

3.Methodology

3.1 Samples and the Data Source

This study selects 49 railway stations within the Yangtze River Delta—one of the most economically developed regions in China—as research subjects. These stations are all served by at least two railway lines, with at least one being a high-speed or intercity rail line (Fig. 1). Passenger group data associated with these stations are derived from Sun Juan et al. (2024), who used LBS trajectory data covering 72 million person-trips in the Yangtze River Delta railway station areas in 2021. Individuals who traveled by train and conducted activities within a 5 km radius of a station (excluding the station itself) were defined as “arrival-activity travelers.” Among them, travelers who stayed in office buildings for more than one hour and for longer durations than in leisure venues were defined as “business travelers.” Travelers who spent more than one hour in scenic areas, shopping zones, or entertainment venues—and whose time in these areas exceeded time spent in business districts—were defined as “leisure travelers.”Travelers whose residence and workplace (for more than three months) were located in different cities and who regularly commuted by train were categorized as “commuters.” All remaining travelers were classified as “other travelers,” with travel purposes including cross-city medical visits, family or friend visits, and transport transfers.

The railway network data were obtained from the Chinese Railway Passenger Train Timetable, which includes train numbers, station information, and departure and arrival times. The dataset contains a total of 10,737 passenger trains and 3,212 railway stations. The initial letter of each train number indicates its type, with a total of eight categories

(Wang et al., 2020) (see Table 1). Due to differences in speed and ticket pricing among these train types, their correspondence with different passenger groups also varies.

Business travelers and commuters, who are generally more time-sensitive, mainly correspond to the G, D, C, and S train types. These are grouped in this study as “High-speed/Short-haul trains,” and the corresponding railway network is referred to as the HSR (High-Speed Railway) Network. The remaining train types—Z, T, K, numeric-only trains, and Y—are classified as “Conventional/Long-haul trains,” forming the CLR (Conventional Long-distance Railway) Network, which mainly serves passengers with lower time sensitivity.The combination of both networks is referred to as the Integrated Railway Network, abbreviated as IR.

Table 1. Chinese Passenger Train Classification Framework

Type		Discreption
Z	Non-Stop Express Passenger Train	- Max speed: 160km/h - Limited intermediate stops - Point-to-point service
T	Limited-Stop Express Train	- Max speed: 140km/h - Major city stops only - Medium-long distance focus
K	Fast Passenger Train	- Max speed: 120km/h - Multi-stop configuration - Cost-effective service
-	Conventional Local Train	- Speed: ≤120km/h - Green-body coaches - Frequent stops, low fares
Y	Scenic Tour Train	- Max speed: 120km/h - Routes: Tourist attractions - Themed service
G	High-Speed EMU Train (G-Series)	- Design speed: 250-400km/h - Business/Top-class seats - Long-distance premium service
D	High-Speed EMU Train (D-Series)	- Operating speed: 200-250km/h - Regional corridor service - Medium-distance comfort
C	Intercity EMU Shuttle	- Speed: 200-350km/h - High-frequency operation - Metropolitan area connectivity
S	Metropolitan Commuter Rail	- Speed: 100-160km/h - Urban-suburban transit - Mass transit configuration

3.2 Modeling the Railway Networks

There are various methods for constructing railway complex networks. In this study, the L-space approach is adopted, where stations are represented as nodes, and any two adjacent stations traversed by the same train are considered connected(Lin and Ban, 2013). The train number and travel time between each pair of adjacent stations are recorded, and the average travel time of all trains passing through that segment is used as the time cost of the corresponding edge. This process is conducted separately for the HSR, CLR, and IR networks.

The constructed HSR network contains 1,629 nodes and 9,771 edges nationwide, with an average edge travel time of 30.05 minutes. The CLR network consists of 2,071 nodes and 7,049 edges, with an average edge travel time of 77.48 minutes. The average travel time between CLR stations is 257.58% of that between HSR stations. The combined IR network contains 3,212 nodes and 15,810 edges (Fig. 2).

3.3 Measuring Node Attributes of Stations in the Rail Network

This section proposes a multidimensional node evaluation framework that systematically assesses the structural functions and strategic positions of stations within the railway network. The framework integrates the topological characteristics of different networks with both the time cost between stations and the economic attributes of the cities where the stations are located. The measurement system includes two core indicators: Temporal Closeness Centrality and Economically-Weighted PageRank.

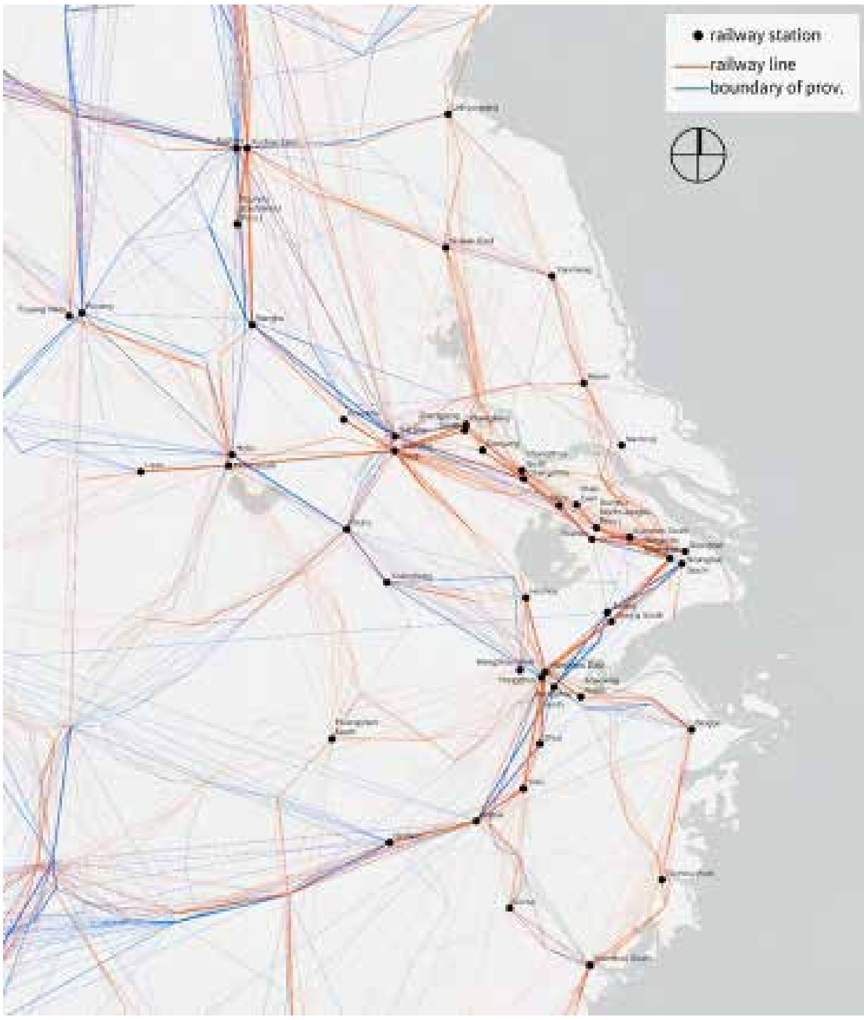

Figure 2. Multi-tiered Rail Complex Network in L-Space Constructed from Train Services (Link width corresponds to service frequency between stations)

Given prior research confirming the symmetric nature of China's railway network (Wang et al., 2020), all indicator calculations in this study are based on an undirected network. By weighting the traditional closeness centrality with time cost between stations, the Temporal Closeness Centrality is defined as follows:

$$Closeness_v = \frac{n-1}{\sum_{u \neq v} T_{uv}} \quad (1)$$

Where $Closeness_v$ represents the Temporal Closeness Centrality of station v; T_{uv} denotes the shortest travel time between stations u and v (assuming u and v are within the same connected component), calculated as the sum of average travel times across all edges in the time-optimal path between them; and n is the number of nodes within the connected component that includes station v. Temporal Closeness Centrality measures the average ease (or difficulty) of reaching all other stations from station v, using travel time as the cost metric across different types of railway networks.By incorporating both the economic context of each city and the connectivity between stations, we define the following:

$$PageRank_v = \alpha \sum_{u \in M(v)} PR_0(u)/L_u + (1-\alpha)/N \quad (2)$$

$$PR_0(u) = GDP/m \quad (3)$$

Where a is the damping factor, set to the default value of 0.85; N is the total number of stations within the connected component; $M(v)$ denotes all neighboring stations of station v; L_u is the number of neighboring nodes of station u; and $PR_0(u)$ is the initial importance of station u, calculated as the GDP of the city where station u is located divided by the number of railway stations (m) in that city. After several iterations, the final importance values for each station can be obtained.

This study calculates both Temporal Closeness Centrality and Economically-Weighted PageRank for each station in the HSR, CLR, and IR networks, and analyzes their correlations with different types of passenger group indicators at each station (Table 2).

Table 2. The Metrics and Specifications

Dimension	Metric	Specification
Node Metrics	Closeness	Temporal Closeness Centrality
	PageRank	Economically-Weighted PageRank
Passenger Metrics	N_Y	number of yearly passengers depart only
	P_S	Proportion of passengers staying in the station area
	P_B	Proportion of business passengers among passengers staying in the station area
	P_C	Proportion of commuters among passengers staying in the station area
	P_L	Proportion of passengers for leisure among passengers staying in station area
	P_O	Proportion of other passengers among passengers staying in station area

4. Results

4.1 Correlation Analysis

Among the 49 station samples, 23 stations only have high - speed rail lines. Thus, samples with only high - speed rail lines will be excluded when calculating CLR - related indicators. After normality testing, C_IR, C_HSR, PR_CR, P_B, P_L, and P_O follow a normal distribution. The Pearson correlation coefficient will be used to calculate the correlation between them. For the rest, the Spearman correlation coefficient will be applied (Table 3).

Table 3. The Correlation Analysis Results

Rail Networks	Metric	N_Y	P_S	P_B	P_C	P_L	P_O
IR	Closeness	0.066	0.028	*-0.092*	-0.111	*-.366***	*0.229*
	PageRank	**.726****	**.639****	**.577****	.456**	.331*	**-.597****
HSR	Closeness	0.087	0.006	*-0.144*	-0.108	*-.301**	*0.253*
	PageRank	**.632****	.490**	**.525****	.477**	0.245	**-.514****
CLR	Closeness	0.091	0.079	-0.015	-0.278	-0.247	0.088
	PageRank	0.350	.406*	*0.198*	-0.058	*0.172*	*-0.207*

Italicized values in the table represent Pearson correlation coefficients;
**. Significant at the 0.01 level (2-tailed);
*. Significant at the 0.05 level (2-tailed).

The correlation calculation results reveal the following characteristics:
(1)PageRank Correlations: Station PageRank on IR and HSR networks correlates with passenger metrics, with stronger overall correlations on IR (except for Proportion of commuters, where HSR PageRank shows a stronger correlation).

(2)CLR Network Irrelevance: Station node properties on the CLR network generally show no significant correlation with passenger metrics, except for the Proportion of passengers staying in the station area, which has a significant positive correlation with CLR PageRank (correlation coefficient: 0.406).

(3)Temporal Closeness Centrality: On IR and HSR networks, Temporal Closeness Centrality only shows a significant negative correlation with the Proportion of passengers for leisure and no correlation with other metrics.
(4)Strongest Correlation: Proportion of passengers staying at the station area has the strongest positive correlation with IR PageRank, followed by HSR PageRank.

(5)Significant Negative Correlation: Proportion of other passengers shows a strong significant negative correlation with PageRank on both IR and HSR networks.

These characteristics indicate that the High-speed and Short-haul Rail Network is positively correlated with the proportion of business and commuter travelers but negatively correlated with other traveler types. As stations become more important in the HSR network, the increase in business and commuter travelers outpaces that of leisure and other travelers. Node importance considering urban GDP better explains station attractiveness than high temporal accessibility, and the Integrated Rail Network, combining HSR and CLR, shows stronger explanatory power.

Despite the Yangtze River Delta being one of China's most HSR-developed regions, incorporating the conventional rail network into railway complex network modeling for station analysis remains valuable.

The PageRank of stations in the conventional long-distance rail network positively impacts passenger station-area staying rates. This might be because the conventional rail system, with a longer development history, has more mature and comprehensive urban construction around high-importance stations, making them more attractive for passenger stays.

4.2 Node-Passenger modal
Drawing on the Node-place model for 'Node-Passenger' analysis: Min-max normalize the PageRank values of stations in the IR network and the Proportion of passengers staying at station areas. Plot the normalized PageRank values on the x-axis and the normalized proportion of passengers on the y-axis, then perform curve fitting(Fig. 3).

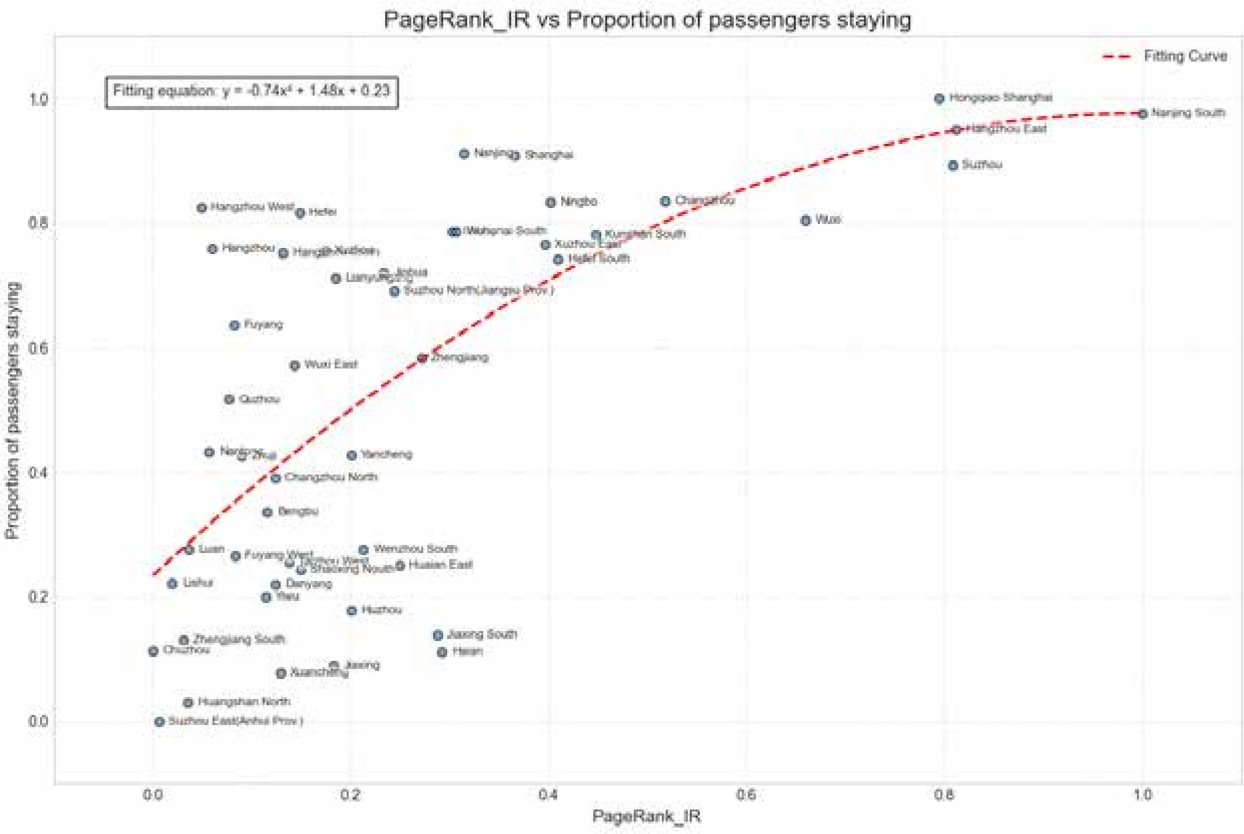

Figure 3 Node-Passenger Modal

The fitted trend line, a parabola, shows that in the Yangtze River Delta, the proportion of passengers staying at station areas increases with the PageRank value of stations in the IR network, but at a declining rate. When PageRank reaches around 63%, growth in passenger proportion halts.

Five stations hitting this threshold are Nanjing South, Hongqiao Shanghai, Hangzhou East, Suzhou, and Shanghai Station, all major stations in the region's most economically advanced cities.

Using the classic Node-Place Theory to explain the Node-Passenger model shows no "unbalanced nodes" in the Yangtze River Delta, but many "unbalanced places" cluster in the upper-left quadrant. These are mostly old stations (e.g., Hangzhou, Hefei, and Xuzhou Stations) with low railway network importance but high urban development around them, making them attractive for passenger stays. Numerous Dependency stations are in the lower-left quadrant. They have low IR network importance, struggle to attract business and commuter travelers, and thus have low passenger retention rates and destination-formation levels.

5. Conclusions

This paper focuses on railway stations' node properties in rail networks. It uses stable and accessible network structure indicators instead of the volatile and hard - to - get passenger flow and train frequency indicators to represent station node properties. It checks their correlation with the Proportion of passengers staying, which directly shows the "destination-formation degree" of stations. Results show that PageRank (considering city GDP) of stations in the combined HSR and CLR networks strongly correlates with passenger activity rates (correlation coefficient: 0.726) and with proportions of business, commuter, and leisure travelers. This proves that stations' importance in HSR networks significantly attracts time-sensitive passenger groups. It also finds an upper limit

(63%) for passenger retention rates in the Yangtze River Delta.

For future research, it's suggested to include more built-environment indicators in the analysis system. This will help form a comprehensive "Node - Passenger - Place" analysis system.

Funding. National Natural Science Foundation of China (52178047).

Reference

Amini Pishro, A., Yang, Q., Zhang, S., Amini Pishro, M., Zhang, Z., Zhao, Y., Postel, V., Huang, D., Li, W., 2022. Node, place, ridership, and time model for rail-transit stations: a case study. Sci Rep 12, 16120. https://doi.org/10.1038/s41598-022-20209-4

Bertolini, L., 1999. Spatial Development Patterns and Public Transport: The Application of an Analytical Model in the Netherlands. Planning Practice and Research 14, 199–210. https://doi.org/10.1080/02697459915724

Carlton, I., 2009. Histories of transit-oriented development: perspectives on the development of the TOD concept.

Chorus, P., Bertolini, L., 2011. An application of the node place model to explore the spatial development dynamics of station areas in Tokyo. Journal of transport and land use 4, 45–48.

Edwards, B., 2013. The modern station: new approaches to railway architecture. Taylor & Francis.

Kamruzzaman, Md., Baker, D., Washington, S., Turrell, G., 2014. Advance transit oriented development typology: case study in Brisbane, Australia. Journal of Transport Geography 34, 54–70. https://doi.org/10.1016/j.jtrangeo.2013.11.002

Lin, J., Ban, Y., 2013. Complex Network Topology of Transportation Systems. Transport Reviews 33, 658–685. https://doi.org/10.1080/01441647.2013.848955

Peek, G.-J., Bertolini, L., De Jonge, H., 2006. Gaining insight in the development potential of station areas: A decade of node-place modelling in The Netherlands. Planning Practice and Research 21, 443–462. https://doi.org/10.1080/02697450701296247

Pezeshknejad, P., Monajem, S., Mozafari, H., 2020. Evaluating sustainability and land use integration of BRT stations via extended node place model, an application on BRT stations of Tehran. Journal of Transport Geography 82, 102626. https://doi.org/10.1016/j.jtrangeo.2019.102626

Juan, S.,Chunhui, G., Zuyuan G., Luofeng, Y., Xingjian, W., Xiaomin S., 2024. Assessment of the Development of High-speed Railway Hub Areas in the Yangtze River Delta Based on Accurate Identification of Passenger Groups: Thresholds, Divergences, and Core Spaces. Urban Planning Forum.

Wang, W., Cai, K., Du, W., Wu, X., Tong, L. (Carol), Zhu, X., Cao, X., 2020. Analysis of the Chinese railway system as a complex network. Chaos, Solitons & Fractals 130, 109408. https://doi.org/10.1016/j.chaos.2019.109408

Zemp, S., Stauffacher, M., Lang, D.J., Scholz, R.W., 2011. Classifying railway stations for strategic transport and land use planning: Context matters! Journal of Transport Geography 19, 670–679. https://doi.org/10.1016/j.jtrangeo.2010.08.008