

Machine Learning driven complex network analysis of transport systems

Yuqin Xia^a, Kewei Wang^a, Purin Tanirat^b, Bryan Lee^a, Irene Moulitsas^a, Jun Li^{a,*}

^a Centre for Computational Engineering Sciences, Cranfield University, Cranfield MK43 0AL, UK

^b Space Technology Research Center, Geo-Informatics, and Space Technology Development Agency (GISTDA), Chonburi 20230, Thailand

ARTICLE INFO

Keywords:

Clustering
Community detection
Network resilience
Link prediction
Graph convolutional networks
Multi-layer perceptron
China railway network
Paris multi-modal transport system

ABSTRACT

A complex network is a system of interconnected nodes linked by edges, exhibiting non-trivial structural features such as community structure or scale-free distributions. This study develops a novel and generic Machine Learning-driven framework that integrates Complex Network Theory and Machine Learning methods for a comprehensive and multifaceted analysis of transport systems. Specifically, four key functional development and analysis are undertaken: 1) Network analysis, using complex network indicators to study the static properties of the transport systems; 2) Network clustering, employing K-means and hierarchical clustering methods to identify underlying community structures; 3) Network resilience, examining the networks' dynamic characteristics and structural evolution under escalating node attacks to evaluate their robustness; 4) Link and feature prediction, developing Graph Convolutional Networks (GCNs) and Multi-Layer Perceptron (MLP) models to predict hidden links and features. The proposed framework is subsequently applied to two distinct transport systems, namely, the China railway network and the Paris multi-modal transport system. The complex network analysis reveals distinct complex network features in network scale, density, and efficiency, yet both demonstrate a power-law distribution. The clustering analysis based on various node and edge properties exhibits a pattern of concentric circles, radiating outward from the urban to peripheral cities in China railway network, while a high density of short-distance connections within central Paris and a prevalence of long-distance connections in the outskirts. The network attack simulations show fine resilience of the Parisian multi-modal system and low resilience of the China railway network. For link prediction, an encoder-decoder model based on GCN and multiple MLPs are developed for various scenarios. The results for the China railway network reveal critical interregional links, emphasizing the need to strengthen regional connectivity, such as expanding the high-speed railway between Hainan Island and the mainland, and establishing a major transportation artery running from south to north. In the Paris transport system, this study predicts an interesting link extending from southern Paris eastward toward northern Seine-et-Marne, indicating a demand for a direct connection. For both networks, the hidden links are largely concentrated in more developed areas, likely driven by strong economic and social interaction demands, highlighting the need for more balanced transport network development. Overall, the results of this study align closely with existing literature and official transport development plans. This research contributes to the theoretical development in Complex Network Analysis using Machine Learning and offers valuable insight to improve the two transport systems.

1. Introduction

Complex networks are composed of interconnected elements, graphically represented by nodes (i.e., the elements) and edges (i.e., their interactions) (Newman, 2003; Wang and Chen, 2003; Boccaletti et al., 2006). However, unlike traditional graphs complex networks are distinguished by their large scale and intricate structure, resembling real-world systems. This complexity poses significant challenges for

graph-based analysis using conventional methods. In recent years, researchers have developed various approaches to leverage the interdependent structural features of complex networks for theoretical analysis and prediction across diverse domains (Ait Rai et al., 2023; Gao et al., 2015). For example, in social networks, complex network theory has been applied to analyse information cascades on social media (Gleeson et al., 2021; Huang and Chen, 2015). By analysing the network structures and user behaviors, the theory is able to identify key influencers

* Corresponding author.

E-mail addresses: yuqin.xia.058@cranfield.ac.uk (Y. Xia), kewei.wang.611@cranfield.ac.uk (K. Wang), i.moulitsas@cranfield.ac.uk (I. Moulitsas), jun.li@cranfield.ac.uk (J. Li).

<https://doi.org/10.1016/j.jtrangeo.2025.104270>

Received 3 December 2024; Received in revised form 22 April 2025; Accepted 7 May 2025

Available online 15 May 2025

0966-6923/© 2025 The Authors. Published by Elsevier Ltd. This is an open access article under the CC BY license (<http://creativecommons.org/licenses/by/4.0/>).

and predict information flow. Similarly, in the transport domain (Lin and Ban, 2013; Háznagy et al., 2015; Huang et al., 2018; Rahmani et al., 2023), this perspective is essential for understanding the characteristics and dynamics of transport networks such as spatial patterns, network vulnerabilities, resilience and hidden connections.

In transport networks modeled as complex systems, nodes represent locations such as cities or stations, while edges signify the connections or routes between them. Current Complex Network Analysis (CNA) of transport networks commonly employs metrics such as degree centrality, betweenness centrality and modularity to identify critical stations and routes, as well as to assess transport vulnerabilities and robustness (Lin and Ban, 2013; Háznagy et al., 2015; Jiao et al., 2017). However, transport network research and management often require to model transport dynamics and predict their features such as spatial patterns, network resilience and potential connections, whereas traditional CNA methods often fail to address effectively. With advancements in Machine Learning and Deep Learning, there is a growing trend in applying these approaches to Complex Network Analysis (Haghighyan et al., 2021; Jiang and Luo, 2022; Li et al., 2024; Srinivasan et al., 2019). Nevertheless, limited research has been conducted on the Machine Learning-driven Complex Network Analysis specifically applied to transport systems (Behrooz and Hayeri, 2022). Moreover, although there are comparative analysis of transport systems based on CNA available, these studies often focus on similar transport cases (Háznagy et al., 2015; Berche et al., 2009). There is a lack of comparative analyses across multiple transport scenarios at different scales within distinct economies. More importantly, there is a growing need for incorporating topological properties and dynamic modeling of transport networks to analyse spatial and functional economic processes across regions. Such analyses can inform targeted social and economic planning, policy-making, and development strategies, fostering economic growth and reducing disparities between areas.

Building on these insights, this study integrates various Machine Learning and Deep Learning approaches, such as clustering, Multi-Layer Perceptrons (MLPs) and Graph Convolutional Networks (GCNs) (Zhang et al., 2019) into traditional CNA and develop a novel framework for spatial pattern discovery, network resilience analysis and link prediction (Kumar et al., 2020), to offer a fresh perspective on comprehensive transport network analysis by leveraging both network structure and dynamics. The research encompasses a detailed analysis of network properties, resilience, clustering, hidden edge discovery, and the prediction of new nodes and edges, based on two real-world transport system scenarios.

The remainder of this paper is organized as follows: Section 2 reviews the current state-of-the-art research in Machine Learning, Deep Learning and CNA applied to transport systems. Section 3 outlines the datasets, methodology, and analysis procedures applied in this study. Section 4 presents the CNA, clustering, node and link prediction results and evaluates the effectiveness of the proposed methodology. Section 5 discusses the findings, highlighting strengths and limitations. Finally, Section 6 concludes the study.

2. Literature review

Complex Network Analysis (CNA) (Newman, 2003; Wang and Chen, 2003; Boccaletti et al., 2006) is widely used to explore the topological properties of transport networks such as connectivity and centrality. A comprehensive review of various transport systems such as railway systems, urban public transport networks and aviation networks (Lin and Ban, 2013) shows that the complex network theory consistently discloses shared properties such as small-world effect and scale-free distribution across these systems. The urban public transportation systems of five Hungarian cities are analysed based on global network characteristic and centrality measures to identify similarities and differences in the organization of the public transport and the most sensitive routes and stations (Háznagy et al., 2015). Similarly, the impact of

China's high-speed railway network on city connections is studied using complex network properties including beta index, clustering coefficient, and average path length, showing that the network significantly increased the overall city connectivity (Jiao et al., 2017). The Chinese Air Transport Network (ATNC) is analysed through complex network properties such as degree distribution, average path length and clustering coefficient, revealing its scale-free structure characterized by a few highly connected hubs and less connected airports (Wang et al., 2011). This study also explores how the network's structure correlates with geographical and economic factors. Despite its growing applications, most of the research mainly focuses on the static complex network properties. CNA faces challenges when applied to dynamic network analysis. Issues such as network resilience assessment and link prediction remain difficult to address due to the inherently static nature of traditional CNA property-oriented methods. There is a lack of comprehensive research and solutions in the development and optimization of current transport systems, hindering progress toward efficiency, robustness and future advancement. A framework (Kölker et al., 2025) that integrates and categorizes a range of network analysis measures is introduced for analysis of Europe's air transportation system. However, the study primarily focuses on intra-European network connectivity and its evolution. The applicability of the framework to other transportation systems is not explored. Furthermore, the analysis is based solely on classical CNA technique, overlooking more advanced approaches such as Machine Learning and Deep Learning methods.

Emerging studies are increasingly integrating CNA with Machine Learning, particularly leveraging Deep Learning approaches for graph-based data analysis. For example, community detection in complex networks applies Machine Learning algorithms such as random walk and clustering to identify clusters of nodes (or communities) that are more densely interconnected than the rest (Huang et al., 2018; Oubaalla and Benhlila, 2018; Agrawal and Patel, 2020). The Infomap algorithm based on random walks in networks is used for clustering transport nodes, ensuring connectedness and reducing heterogeneity across clusters (Haghighyan et al., 2021). A method for transport cluster detection, based on K-shortest paths and hierarchical clustering, is proposed to assess whether the public intercity ground transport infrastructures and services can support the integration and development of urban agglomerations in mainland China. An ensemble clustering algorithm based on influential nodes is proposed to enhance community detection CNA (Zheng et al., 2024). The results demonstrate that the method enables more effective and efficient identification of communities. Graph Neural Networks (GNNs) such as GCNs exploit complex networks to learn representations of nodes, edges and overall network structures, and therefore effectively capture the inherent patterns, relationships and dynamics of transport systems (Rahmani et al., 2023; Jiang and Luo, 2022). A complex network-oriented model, SAGE-Het, combining the GraphSAGE and a heterogeneous GNN, is developed to predict train delays in railway networks (Li et al., 2024). This approach captures interactions between heterogeneous nodes, offering improved analysis of complex railway systems. A novel Bayesian framework, called Variational Graph Recurrent Attention Neural Networks based on GNN, is proposed for traffic forecasting. The results show that this model outperforms state-of-the-art approaches (Zhou et al., 2020).

Link prediction (Kumar et al., 2020; Wei and Ling, 2020; Zhang et al., 2018; Yao et al., 2021) is frequently applied in transport networks to predicts the likelihood of potential missing or future links, including similarity-based, probabilistic models, dimensionality reduction-based and other approaches (e.g., learning-based and clustering-based models). Recently Machine Learning-based link prediction methods are developed by incorporating both structural and node features of complex networks. For example, a three-stage Machine Learning-based technique employing Support Vector Machine, MLP, and Random Forest is developed for link fault identification and localization in complex networks based on the traffic flow data (Srinivasan et al., 2019). This method demonstrates high accuracy while significantly reducing fault

localization time. Similarly, a hybrid forecasting framework integrates Machine Learning, link prediction, and CNA to predict port container throughput (Liang et al., 2024). In this framework, CNA is employed for Feature Engineering and Machine Learning methods such as Support Vector Machines and Long Short-Term Memory are applied for link prediction. A spatial GCN model utilizing geographical unit embedding is developed to predict missing flows and evaluate the spatial relation between two regions (Yao et al., 2021). The results on Beijing taxi trip data verify the effectiveness of the model in predicting spatial flows. Nevertheless, despite these advancements, the theoretical development and application of sophisticated Machine Learning including Deep Learning algorithms in integrated complex networks and transport system analysis remains limited (Behrooz and Hayeri, 2022; Kumar et al., 2020).

Moreover, the stability and robustness of complex networks are critical to ensure transport systems functioning properly and effectively coping with potential threats such as natural disasters, malicious attacks and system crashes. Therefore, understanding network resilience and the influential factors is essential for designing more stable, efficient and reliable networks (Gao et al., 2015; Zhang et al., 2018). Researchers have made significant progress in understanding the fundamental features of network resilience, particularly in relation to the small-world phenomenon, scale-free properties, and community structure (Holme et al., 2002; Crucitti et al., 2004; Gao et al., 2011), demonstrating that network topology plays a key role in network resilience. For instance, scale-free networks with power-law degree distributions are more vulnerable to targeted attacks (Holme et al., 2002), while local network performance is more sensitive to cascading failures (Crucitti et al., 2004). A study evaluating the robustness of fourteen major global city transport networks under random and targeted attacks find that public transport networks exhibit varying degrees of robustness depending on the type of network structure and the nature of disruptions (Berche et al., 2009). The study of Europe's air transportation system (Kölker et al., 2025) offers insights into seasonal variations in network robustness, emphasizing the stronger local connectivity among airports and the prominent role of the top 20 airports. Moreover, interdependencies between networks significantly affect overall network resilience (Gao et al., 2011). However, these studies often focus on individual or specific scenarios and lack comparative analyses across multiple scenarios at different scales, such as national and city scales, with multi-modal interdependent networks under distinct economic contexts.

The review above shows that, despite the theoretical advance in Machine Learning and Deep Learning approaches and growing applications in CNA, limited research has been conducted on the Machine Learning-driven CNA of transport systems. There is lack of a comprehensive and generic framework in transport system research that integrates CNA with Machine Learning for a multifaceted analysis of transport systems. Moreover, existing network resilience methods based on CNA may have limitations in the face of changing network topologies and dynamic behaviors. Current studies mainly focus on specific network scenarios and lack comparative analyses at national and city levels. This paper aims to advance the theory and applications of Machine Learning-driven CNA in transport network clustering, resilience and link prediction through two real-world transport scenarios at national and city scales. Firstly, this paper presents a novel and generic Machine Learning-driven framework that integrates Complex Network Theory and Machine Learning methods for a comprehensive analysis of transport systems. Then, traditional CNA indices and Machine Learning clustering methods are applied to examine and analyse the overall network structure and spatial patterns. Thirdly, an encoder-decoder model based on GCN and multiple MLPs are developed for link prediction. This paper also investigates the impacts of network attacks on transport network resilience.

3. Methodology

The study develops a novel and generic Machine Learning-driven complex network analysis of transport systems framework as shown in Fig. 1. It integrates various Machine Learning and Deep Learning algorithms, such as clustering, Multi-Layer Perceptrons (MLPs) and Graph Convolutional Networks (GCNs) into traditional CNA, primarily to support four key functional areas, i.e., complex network characterization, resilience assessment, cluster or community detection, and edge prediction with different scenarios.

In the framework, the transport network datasets are first pre-processed, including data cleaning and transformation, to ensure the data accuracy and consistency. Next, the complex networks are constructed, and their properties such as correlation and network structure metrics (Wang et al., 2011; Freeman, 1977; Kitsak et al., 2010; Zhang et al., 2016) are computed. These properties are then used for clustering, network resilience analysis and link prediction based on Machine Learning and Deep Learning techniques. Finally, the results are evaluated based on designed metrics, comparison and uncertainty analysis (incl. Sensitivity analysis) (Schaller et al., 2025) and refined through an iterative optimisation process before being finalized and presented.

For case studies, this research utilizes two transport network scenarios, i.e., the China railway network and the Paris multi-modal transport network. The scenarios are chosen based on their distinct scales (i.e., national and city scales), multi-modal interdependent characteristics and varying economic contexts. The China railway network data, scraped from online sources (China Academy Of Railway Sciences Corporation Limited, 2022; China State Railway Group Corporation Limited, 2022), includes train timetables, train types and journey location information in Excel format. The Paris transport network data, obtained online in Geo and JSON file formats (Asgari et al., 2016), encompasses multiple transport modalities within Paris, including roads, trains, metro, trams, and inter-layer connections. The datasets are pre-processed to remove noise and subsequently validated through expert inspection and corroboration with academic literature (Jiao et al., 2017; Wang et al., 2020a; Cubukcu, 2021). In the China railway network, routes shared by multiple trains are consolidated into a single edge, characterized by a train frequency property, while stations within the same city are consolidated into a single node (Berche et al., 2009). In the complex network of Paris's multi-modal transport system, nodes are multi-modal stops, and edges represent the routes that connect them. Each node is mapped to its corresponding state based on its geographical location. The key features of both datasets are summarized in Table 1.

The node and edge types are categorized as Road, Train, Metro, Tram, and Combined Transportation (cross-layer), labeled 1 to 5 respectively, while the remaining properties in the table are self-explanatory.

3.1. Network properties and clustering

Literature (Wang and Chen, 2003; Freeman, 1977; Kitsak et al., 2010; Zhang et al., 2016) shows that complex networks exhibit unique static network characteristics such as homogeneity, invariable regularity and randomness. This study analyses typical characteristics of static complex networks, focusing on their nodes, edges and network structures, as summarized in Table 2.

Clustering methods group data points based on their similarities, providing insight into the structures and patterns within the data. This study applies K-means and hierarchical clustering algorithms (Greene et al., 2008) to analyse both datasets based on node and edge features. The results are then compared to evaluate the accuracy and consistency of the clustering methods. K-means method groups data points based on the nearest mean, while hierarchical clustering organizes data points into a tree-like structure based on pairwise distances or similarities, facilitating a hierarchical arrangement of clusters. Additionally, the

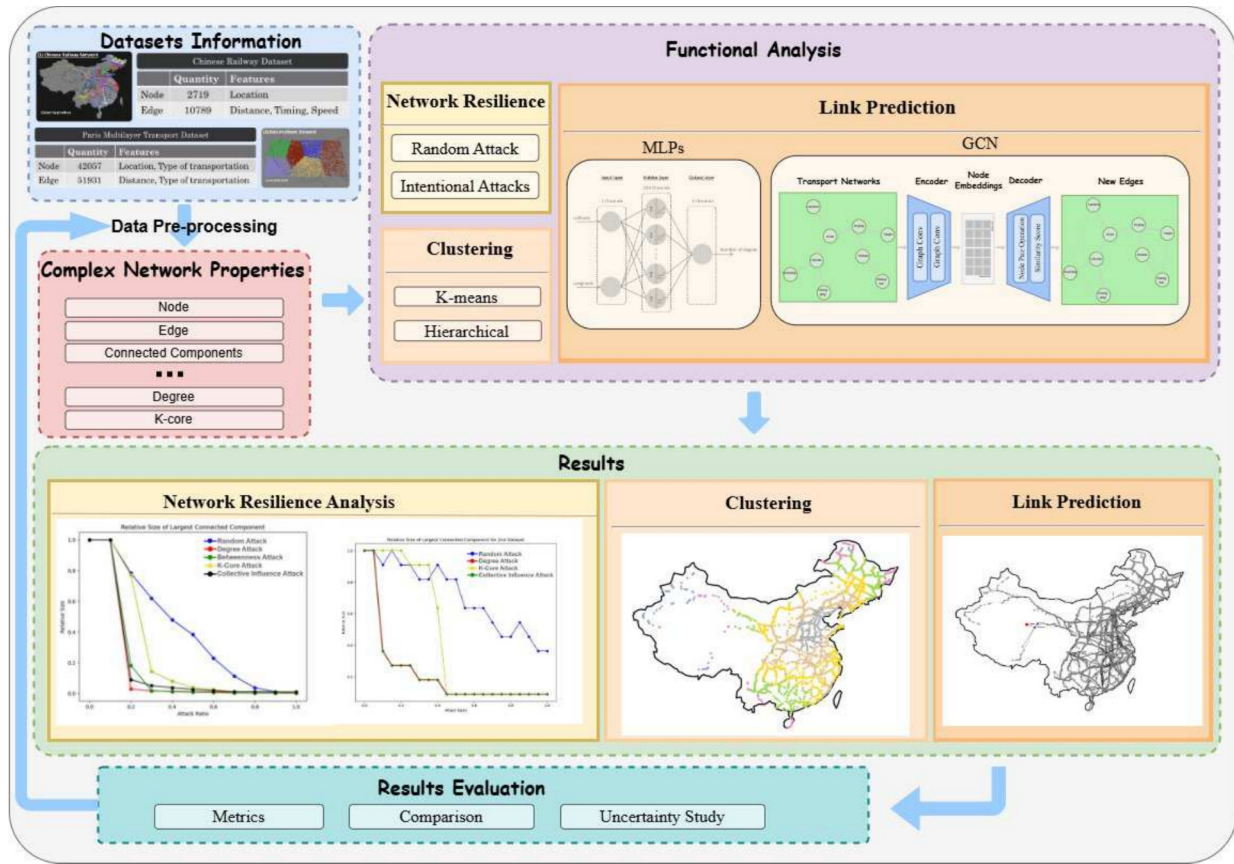


Fig. 1. The framework of Machine Learning-driven complex network analysis of transport systems.

Table 1

Key features of China railway network (left) and Paris transport network (right) datasets.

China railway network		Paris multi-modal transport network	
Feature	Type of feature	Feature	Type of feature
Train station number (node)	Node	Node ID (node_id)	Node
Train station latitude (lat)	Node	Node name (node_name)	Node
Train station longitude (lon)	Node	Node latitude (lat)	Node
State (province) of train station (state)	Node	Node longitude (lon)	Node
City of train station (city)	Node	Node type (type)	Node
Current station (st_id)	Edge	Node state (state)	Node
Target station (st_tg)	Edge	Current node (source)	Edge
Train number (train)	Edge	Target node (target)	Edge
Max speed of train (train_max_speed)	Edge	Edge type (type)	Edge
Arrival time (arr_time)	Edge	Distance of route (distance)	Edge
Departure time (dep_time)	Edge		
Target arrival time (next_arr_time)	Edge		
Stay time at station (stay_time)	Edge		
Distance of route (distance)	Edge		
Travel time (travel_time)	Edge		

Table 2

Key properties of a complex network.

Network structure	Nodes
Node number	Degree (Wang and Chen, 2003)
Edge number	K-core/shell (Kitsak et al., 2010)
The number of connected components (Strogatz, 2001)	Degree Correlation (Boccaletti et al., 2006)
The number of nodes of the largest connected component	Degree centrality (Freeman, 1977)
The number of edges of the largest connected component	Closeness centrality (Freeman, 1977)
K-core/shell (Kitsak et al., 2010)	Eigenvector centrality (Freeman, 1977)
Density (Li and Cai, 2004)	
Diameter (Lin and Ban, 2013; Li and Cai, 2004)	
Average distance (Lin and Ban, 2013)	
Efficiency (Lin and Ban, 2013; Li and Cai, 2004)	

$$clustering_bias = 1 - \frac{number_same_clusters}{total_samples} \quad (1)$$

where *number_same_clusters* refers to the number of samples assigned to the same cluster by both clustering methods, and *total_samples* refers to the total number of samples.

3.2. Network resilience

Network resilience in this paper refers to a network's ability to maintain its core functions in the face of failures, attacks, or other disruptions. It encompasses both the static characteristics and the dynamic

elbow method (Shahapure and Nicholas, 2020) is used to determine the optimal number of clusters for the algorithms. A unique integer is assigned to each cluster based on their node metrics means. Then, the clustering bias is calculated using Eq. (1),

evolution of the network's connections. This study examines their dynamic characteristics and structural evolution under escalating complex network feature attacks. To simulate different types of attacks, two strategies—random and intentional attacks—are employed, based on the features shown in Table 3.

The collective influence is calculated by the Eq. (2).

$$CI_l(i) = (k_i - 1) \sum_{j \in Ball(i,l)} (k_j - 1) \quad (2)$$

where l denotes a maximum distance or neighbourhood radius of a node i , k_i denotes the degree of node i and $Ball(i, l)$ is the set of nodes within a distance l from node i . l acts as a constraint on the nodes to consider. For a given l , the collective influence on i can be calculated.

For random attacks, nodes are gradually removed at random, while for intentional attacks, critical nodes are sequentially removed based on their importance as determined by node features. The impact on the networks is assessed alongside the attacks.

3.3. Link prediction

Link prediction predicts potential connections between nodes in a complex network by analysing node features and their relationships. A Graph Convolutional Network (GCN) (Zhang et al., 2019) is a method for semi-supervised learning on graph or complex network data, leveraging both the topology and attributes of the graph. This study uses an encoder-decoder framework (Masui, 2020) for link prediction based on a GCN model and vector similarity, as shown in the Link Prediction component of Fig. 1. The encoder is a two-layer GCN model that learns node embeddings as vector representations. The decoder then calculates a similarity score, representing the link probability, between pairs of node vectors. The GCN model captures a network's information through message-passing, transforming the link prediction task into a binary classification problem. In the context of a transport network, nodes represent stations, while edges correspond to the connections between them. A typical GCN layer performs three key functions: message passing, aggregation and update, as expressed in Eq. (3).

$$H^{(l+1)} = \sigma \left(D^{-\frac{1}{2}} A D^{-\frac{1}{2}} H^{(l)} W^{(l)} \right) \quad (3)$$

where σ represents the activation function, A is the adjacency matrix further normalized by the degree matrix D , H denotes the node representation matrix updated at each layer $l + 1$, and W is the learnable weight matrix. The final node representation matrix $H^{(l+1)}$ is then passed to the output layer. Moreover, this study employs BCEWithLogitsLoss (Paszke et al., 2019) to compute the loss based on Binary Cross Entropy (BCE),

$$\mathcal{L}_{\text{BCEWithLogits}}(x, y) = \max(x, 0) - x \cdot y + \log(1 + e^{-|x|}) \quad (4)$$

where x is the raw logit (i.e., model output before applying sigmoid) and $y \in \{0, 1\}$ is the ground truth label. This method combines an implicit

Table 3
Network attack types.

Attack type	Description
Degree (Wang and Chen, 2003)	The number of nodes connected to a node.
Betweenness centrality (Freeman, 1977)	The number of shortest paths through a node, with higher values indicating greater network efficiency. Attacking such nodes would rapidly diminish efficiency.
K-core/shell (Kitsak et al., 2010)	Measures a node's proximity to the network center. Attacking it can quickly fragment the network.
Collective influence (Zhang et al., 2016)	The influence of surrounding nodes.

Sigmoid operation and BCE loss in one single class. The primary advantage of BCEWithLogitsLoss lies in its enhanced stability compared to BCE loss, while avoiding issues such as numerical overflow and vanishing gradients. Link prediction is fundamentally a binary classification problem, where the model predicts an edge probability between two nodes. This study then applies confusion metrics such as ROC (Receiver Operating Characteristic) and Accuracy to evaluate the model's performance.

Before training, the two datasets are labeled as either positive (link exist) or negative (no link) samples, with equal representation for both classes. Both datasets represented by features derived from network properties and resilience analysis are then split into training, validation and testing datasets in a ratio of 8:1:1, with the validation set used to assess performance during training. Stochastic Gradient Descent (SGD) is employed as the optimization method.

This study also addresses the scenario of new edge and feature prediction with a new node being randomly added to a transport network. The novel framework is shown in the figure below.

Firstly, a new node being randomly added to a network, with its location specified by latitude and longitude. Then, the new node is classified into an appropriate group using the Random Forest classification algorithm (Breiman, 2001), based on province in the first dataset or node type in the second dataset. Next, the node's degree is predicted using a Multi-layer Perceptron (MLP) network with 256 hidden neurons, leveraging its location to capture the spatial distribution crucial for network dynamics (Ramchoun et al., 2016). Based on the predicted degree, the KD-Tree (i.e., k-dimensional tree) method is employed to identify the closest neighbour nodes for establishing links (Skrodzki, 2019). Finally, the features of each link are predicted using a separate MLP with two hidden layers of sizes 128 and 64. The input features and output labels for the two datasets are detailed in Table 4.

4. Results

4.1. Network properties and clustering

Table 5 summarizes the key network properties of the China railway network and the Paris multi-modal network based on Complex Network Analysis (CNA). In the China railway network, the presence of a single connectivity graph indicates that all points in the network are directly reachable. Metrics such as average distance and efficiency highlight its relative efficiency. In contrast, the Paris multi-modal network contains a greater number of nodes and edges and is divided into seven distinct connected graphs or districts. K-core/shell analysis identifies a sub-network with 20 nodes and 94 edges, highlighting a highly connected core area. The network's low density of 0.0003 indicates sparse connections among nodes, while its larger diameter reflects a significantly longer path span compared to the China railway network. The high average distance and low efficiency further underscore the Paris network's structural complexity.

Correlation analysis is conducted on all numerical values of node and edge features from the two datasets, as illustrated in Figs. 3 and 4. Figs. 3a and 4a indicate a low or negligible correlation among the nodes features, except for the moderate positive correlation between longitude and latitude observed in 3a. In 3b there are strong correlations among arrival time, departure time and next arrival time, as well as between travel time and distance.

Table 4
Features and labels for edge feature prediction using MLPs.

Datasets	Input features	Labels
China Railway Network	Latitude, longitude, province	Train speed, distance, travel time
Paris Multi-modal Transport Network	Latitude, longitude, node type	Distance, edge type

Table 5
China railway and Paris transport network properties.

Property	China railway	Paris transport
Node number	2719	12,237
Edge number	6168	21,441
The number of connected components	1	7
The number of nodes of the largest connected component	2719	11,789
The number of edges of the largest connected component	6168	20,944
The k-core of the largest connected component	Graph with 14 nodes & 64 edges	Graph with 20 nodes & 94 edges
The network density of the largest connected component	0.001669	0.000301
Diameter of the largest connected component	47	110
Average distance of the largest connected component	8.78929	39.75795
The efficiency of the largest connected component	0.13798	0.03375

This study models both datasets as directed graphs, with each connection defined by its direction of travel. Figs. 5 and 6 visualise their degree distributions on linear and log-log scales. Degree distribution is the probability distribution of the degrees over all nodes. It provides insight into the network's structure, such as whether it has many highly connected hubs or mostly evenly connected nodes. Both Figs. 5a and 6a show a long tail, exhibiting a power-law degree distribution in which only a few nodes have very large degree while the majority of nodes have a relatively small degree, unlike random networks where connections are more evenly spread. This finding aligns with findings from traditional literature (Lin and Ban, 2013; Háznagy et al., 2015; Holme et al., 2002). The power-law degree distributions indicate that both networks are efficient and scalable, but also potentially vulnerable if key hubs are disrupted (Holme et al., 2002; Wang et al., 2020). The log-log plots in the figures, which show a linear relationship between the log of node degree and frequency, further confirm this pattern.

Communities or clusters in complex networks refer to groups of nodes that are more densely connected than with the rest of the network. These communities often represent functional units or entities that share common properties or roles within the system. This study clusters nodes of the two datasets based on node closeness centrality or edge features. Fig. 7 shows the clusters obtained based on closeness centrality in the China railway network. The spatial distribution exhibits a pattern of concentric circles, radiating outward from China's urban centers toward peripheral cities. This pattern underscores the centrality and outward influence of major urban hubs within the network. Fig. 8a and b present the clusters derived based on train speed, with the optimal number of clusters determined using the elbow method. Clustering based on single feature is more effective than using threshold settings, which may adapt to the data and discover the right number of clusters, given unknown or irregular distributions. From Fig. 8a, contrary to the common perception that low-speed trains primarily serve rural locales, the findings reveal a more nuanced picture that these trains commonly operate across diverse landscapes yet there remains a need for enhanced accessibility in less developed areas. Notably, several low-speed routes are strategically designed to support tourism, connecting scenic and cultural destinations. On the other hand, Fig. 8b underscores the widespread adoption of high-speed trains in metropolitan regions, facilitating rapid and efficient connectivity across urban sectors.

Clustering of the Paris multi-modal transport network based on distance reveals two distinct clusters, i.e., short- and long-distance connections between nodes, as shown in Fig. 9a and b. The pictures show a high density of short-distance connections within central Paris while a prevalence of long-distance connections in the outskirts. Fig. 9a highlights a characteristic feature of urban cores, where frequent, short links are essential for facilitating the efficient movement of people and

goods within bustling city centers. In contrast, Fig. 9b shows the long-distance connections, predominantly represented by grey edges. These routes are concentrated on the urban periphery, indicating their role in linking outer areas of the city to its center or connecting Paris with more distant locations.

To assess the accuracy of the clustering models, this study compares the clustering outcomes from K-means and Hierarchical clustering, based on node properties from the first dataset. The results are presented in Table 6, where the largest percentage difference between the two clustering methods is observed in the clustering coefficient (Wang and Chen, 2003), whereas the methods based on eigenvector centrality show good agreement. The clustering coefficient in complex networks describes the extent to which nodes tend to cluster together, with a high value indicating strong local cohesion and a low value suggesting a more dispersed structure. Fig. 10 presents a comparison of the average clustering coefficient values in descending order with the optimal number of clusters (29), highlighting the similar trends exhibited by both methods. This study does not incorporate multiple features for community detection. Future work could explore using multiple features, which may uncover distinct patterns valuable for transport network development.

4.2. Network resilience

Fig. 11 illustrates the relative size changes of two transport networks as a function of attack ratio under different attack strategies. The relative size of a transport network is the proportion of nodes in the current largest component relative to that in the initial largest component before an attack. The China railway network disintegrates rapidly when subjected to degree-based attacks but shows greater stability under random attacks. In contrast, the Paris multi-modal transport network exhibits more moderately decreasing response across various attack types, demonstrating its resilience.

As an example of key node attack, Table 7 demonstrates the changes in network properties after the removal of a selected key node, specifically, Haikou, the capital of Hainan Island, from the China railway network and the node with the highest degree from the Paris multi-modal network. The removal of Haikou splits the China railway network into two connected subgraphs, underscoring its critical role in linking different parts of the network. The number of nodes and edges in the largest connected subgraph decreases by 7 and 12, respectively. However, the network's density, diameter, average distance, and efficiency remain largely unchanged, demonstrating its resilience. For the Paris multi-modal network, the removal of the high-degree node reduces the total number of connections by 54, while the number of connected components remains unchanged. This suggests that, despite the node's importance, its removal does not severely compromise the network's overall structure. This resilience is further supported by the minimal changes observed in network density, diameter, average distance, and efficiency.

4.3. Link prediction

This study applies an encoder-decoder model for hidden link prediction. Fig. 12 presents the predicted hidden links in red in the China

Table 6
Clustering performance bias on the China railway network.

Feature	Performance bias
Degree	2.83 %
Degree Centrality	1.43 %
Clustering coefficients	53.77 %
Closeness centrality	17.14 %
Betweenness centrality	2.17 %
Eigenvector centrality	0.04 %

Table 7

Comparison of network properties before and after the removal of a key node.

Property	Before (China)	After (China)	Before (Paris)	After (Paris)
Node number	2719	2718	12,237	12,236
Edge number	6168	6162	21,441	21,387
The number of connected graphs	1	2	7	7
The number of nodes of the largest connected component	2719	2712	11,789	11,788
The number of edges of the largest connected component	6168	6156	20,944	20,890
The k-core of the largest connected component	Graph with 14 nodes & 64 edges	Graph with 14 nodes & 64 edges	Graph with 20 nodes & 94 edges	Graph with 20 nodes & 94 edges
The network density of the largest connected component	0.001669	0.001675	0.000301	0.000301
Diameter of the largest connected component	47	47	110	110
Average distance of the largest connected component	8.78929	8.791531	39.75795	39.764247
The efficiency of the largest connected component	0.13798	0.138006	0.03375	0.033725

railway network and the Paris multi-modal transport network based on an optimal classification threshold of 0.71, above which links are classified as positive. In Fig. 12a, the predicted connections reveal critical links between Hainan Island and mainland China, further confirming the finding in network resilience analysis, as well as between southeast coastal areas and between Beijing and the northwestern regions, emphasizing the need to strengthen regional connectivity with major economic and political hubs. The predictions closely align with China's National Medium- and Long-Term Railway Network Plan (China National Development and Reform commission, 2016), focusing on regions with strong economic activity and major transport corridors. Conversely, the Paris multi-modal transport network, being densely interconnected, shows predicted links that are shorter and more evenly distributed across the network. The encoder-decoder model achieves a testing accuracy of 92.4 % for the China railway network and 88.7 % for the Paris multi-modal transport network, demonstrating its effectiveness in link prediction. The dense hidden edges predicted indicate a high transport demand between nodes, which should be prioritised for construction based on economic, geographical, and environmental considerations.

Additionally, for policy planning it is beneficial to raise the sigmoid threshold in the model for high link probabilities and to prioritize construction based on network efficiency by analysing the number against strength of hidden links between regions or clusters. Fig. 13 shows the high probability or strength links predicted, with a higher classification threshold of 0.99. The hidden links in the China railway network are all located in eastern China, a highly developed region, likely driven by strong economic and social interaction demands. It is also observed that these links form a transportation artery running from south to north across the country, highlighting the need to strengthen the connections

between southern and northern China, consistent with Chinese government's 12th Five-Year Plan for the Revitalization of Northeast China (China National Development and Reform commission, 2012). The hidden links predicted in the Paris transport network are largely located in the western region, similarly the prosperous area, which may pick up on opportunities in more developed areas for future high-value transport links. These align with the official plans, such as the Grand Paris Express (Société des grands projets, 2025), the EOLE project (CERTIFER, 2021), and various tramway projects (Wikipedia, 2025), aiming to strengthen connections within the western suburbs. In addition, an interesting link is predicted from southern Paris extending eastward toward northern Seine-et-Marne, indicating a demand for a direct connection and more balanced development between the western and eastern regions, aligning well with the Ile-de-France Mobilités 2030 plans (Ile-de-France Mobilités, 2024), which aim to improve south-north interconnections, potentially through a direct Évry-Meaux or Évry-Mitry railway. However, these predictions are based solely on network characteristics and should be interpreted with caution. Political, environmental, and social factors must also be considered when planning transport infrastructure. Furthermore, the datasets used are likely biased toward more developed areas, due to the greater number and strength of links within these regions.

Subsequently, this study randomly generates node locations, generally situated near or within a local administrative unit, and predicts new edges and features as illustrated in Fig. 2. The examples of edge features predicted by the models within China and Paris are given in Tables 8 and 9 respectively, and visualised for China railway network in Fig. 14 as a case study. From the figure, the newly generated node and its subsequently connected neighbours are located in western China. Table 8 presents the two edges to neighbours and their features predicted, i.e., Train Max Speed, Distance and Travel Time, linking a randomly generated new node and its two neighbour nodes. The predicted maximum speed is consistent with train speeds in the local region, and the average speed, calculated based on Distance and Travel Time, is around 95, well below the maximum speed. The top two nearest neighbour nodes in the China railway network are identified using the KD-Tree search. The same approach is applied to Paris transport network, as presented in Table 9, showing three neighbour nodes with two predicted connection features. The long-distance feature predicted for the new edges in the China railway network and the short-distance feature in the Paris transport network reflect the geographical and economic characteristics of the respective regions. This approach is useful for local governments in transport planning, while connecting to the national transport network, maintaining compatibility, and preserving the network's key characteristics.

The features of each link are predicted using a two-hidden-layer MLP for each network. The model performance is summarized in Table 10. The link feature prediction results indicate that MLP models trained on the Paris network outperform those trained on the China railway network, potentially attributed to the larger dataset and the more dense and complex structure of the Paris transport network.

To assess the edge feature prediction uncertainty, the study further applies a Monte Carlo approach by randomly generating node locations within mainland China and Paris region. The distribution of prediction results for each label along with their respective means in red is shown in Figs. 15 and 16.

Figs. 15 and 16 show the range of possible values for each edge feature. The distribution of predicted train speeds in the China railway network exhibits a narrow peak around the mean speed of 269.45 km/h, with significantly fewer observations at higher speeds. The histograms for predicted distance and travel time reveal a heavy right tail, indicating the presence of some trains with considerably longer travel times compared to the majority. The distributions of predicted distance and transport type in the Paris multi-modal network exhibit distinct patterns. The distance distribution shows a right-skewed tail with notable concentrations at minor distances ($0 < d < 25$) and long distances ($d >$

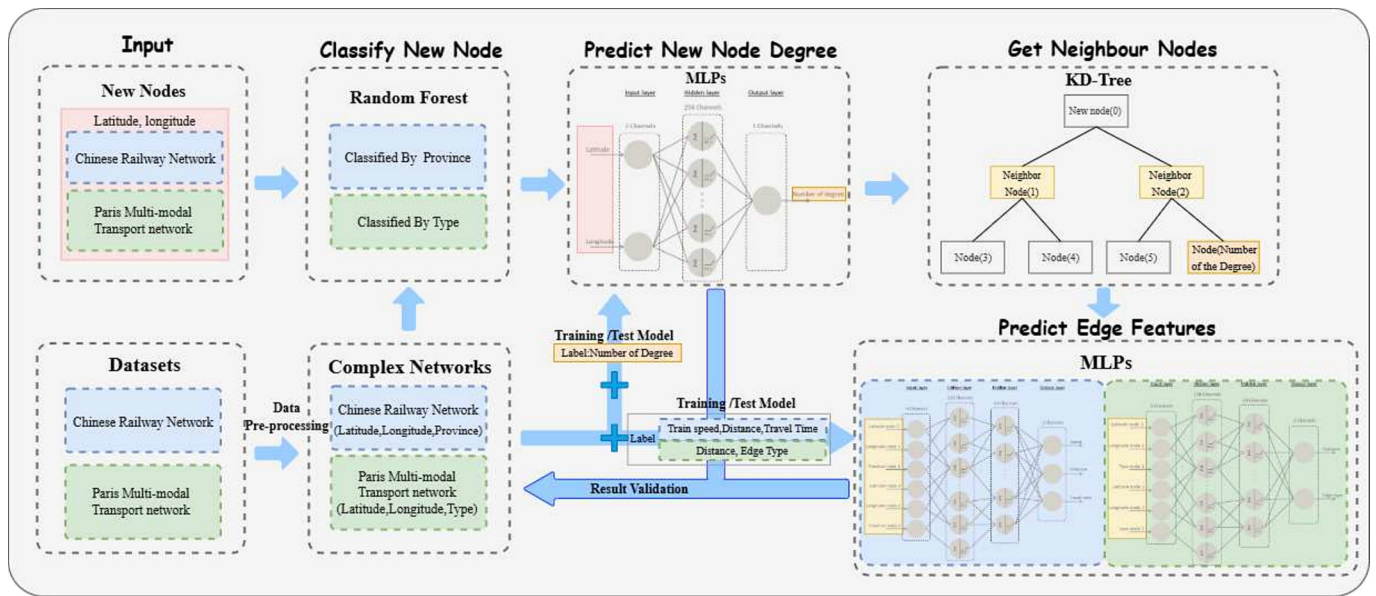


Fig. 2. The framework of new edge and feature prediction for a transport network based on random nodes.

Table 8
Edge features predicted in China railway network.

Neighbour node (No.)	Train max speed (km/h)	Distance (km)	Travel time (day)
1315	120	288.145107	0.126274
2231	120	234.668826	0.104041

Table 9
Edge features predicted in Paris multi-modal transport network.

Neighbour node (No.)	Distance (km)	Transport type
5832	0.525381	1.00 (Road)
11,904	3.105358	3.00 (Metro)
5893	0.452285	4.00 (Tram)

125), while the transport type distribution has a left-skewed tail, with transport type 4 being the most prevalent. These predicted features provide valuable insights for future transport planning.

5. Discussion

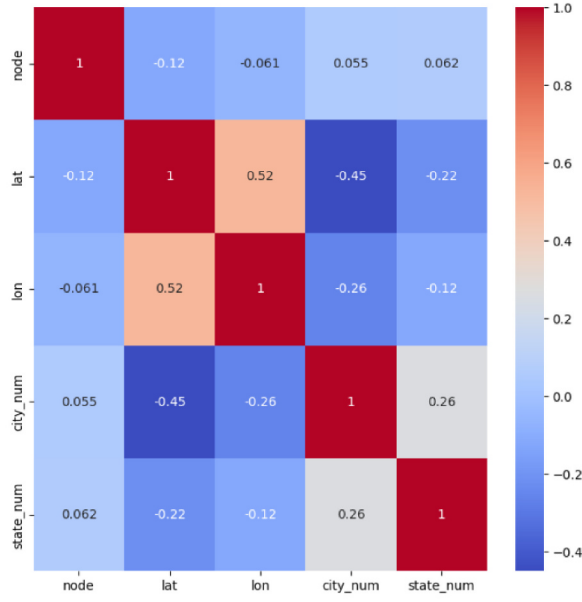
This study develops a novel and generic Machine Learning-driven framework that integrates Complex Network Theory and Machine Learning methods including Neural Network models for a comprehensive analysis of transport systems in various scenarios, involving four key functional development and analysis, namely, network characterization, resilience assessment, community detection, and node and edge prediction. The framework is subsequently applied to analyse two distinct transport systems, i.e., the China railway network and the Parisian multi-modal transport system. The research findings provide nuanced insights into network structure, resilience, and connectivity, highlighting the vital importance of structured analysis using an integrated Machine Learning and complex networks approach.

Complex network properties, such as the number of connected graphs, feature correlation, and degree distribution, are calculated and analysed to study the features and structure of transport systems. Comparing the number of connected graphs reveals that the China railway network is more integrated, while the Paris multi-modal transport network is localized, reflecting its independent economic and social divisions. The degree distribution plots indicate that both networks

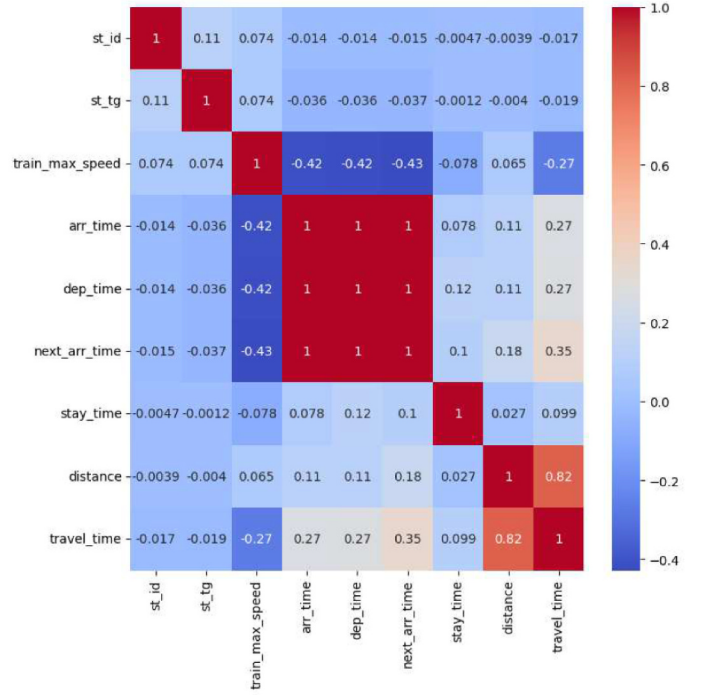
follow a power-law distribution, with a few highly connected hub nodes and many sparsely connected peripheral nodes, a common feature of scale-free networks. Such hubs likely represent key transport centers within the network. This type of network structure is commonly observed in transport systems. An analysis of K-cores shows that the number of transport centers in the China railway network is comparable to that of the Paris network, though its density of the largest connected component is significantly higher. Furthermore, differences in efficiency and network diameter highlight the contrasting characteristics of a national railway system and a multi-modal urban transport network, where the China railway network covers fewer stations en route to destinations than the Paris multi-modal network. Correlation analysis indicates a low or negligible correlation among the nodes features, except for the moderate positive correlation between longitude and latitude in China railway network, and strong correlations among arrival time, departure time and next arrival time, as well as between travel time and distance in Paris multi-modal network, instrumental for the subsequent Feature Engineering process.

The clustering analysis of transport networks based on Machine Learning methods provides insights into their structure and organization, revealing distinct geographic and operational differences. Different clustering models are developed using K-means and Hierarchical clustering methods, by leveraging various node and edge properties. In the China railway network, the node clustering based on closeness centrality indicates a hierarchical structure, where central cities act as hubs connecting to surrounding regions. The edge clustering based on train speed shows that high-speed trains are mainly concentrated in urban areas, despite their potential benefits for rural regions with longer distances between stations. These findings highlight the diverse roles of train services in China and the strategic planning behind its railway infrastructure, supporting transport, economic growth, and regional connectivity. In the Paris multi-modal transport network, the edge clustering by distance shows high connectivity and density in the central area, with long-distance connections more common on the outskirts. A comparison between K-means and Hierarchical clustering techniques reveals that both produce similar clustering results across various metrics, with the clustering coefficient showing significant differences only due to a shift in the number of clusters evidenced in Fig. 10.

The study of complex network resilience is vital for ensuring the stability, efficiency, and safety of modern transport systems. This paper presents two methods, i.e., attack simulation and key node removal, to

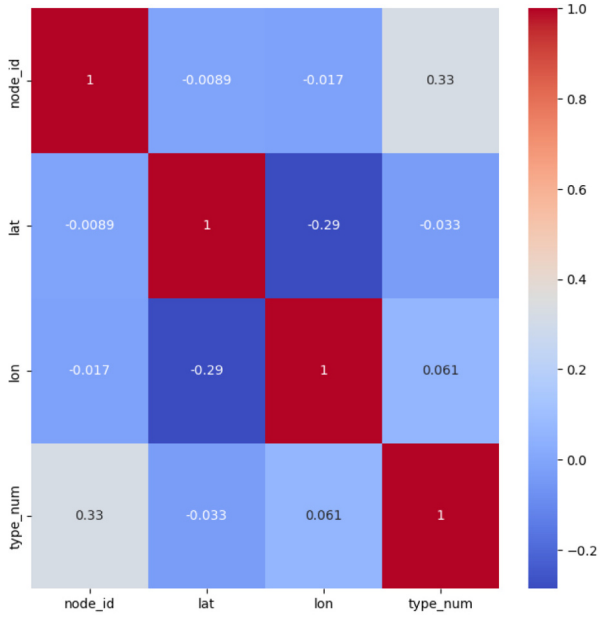


(a) Node features

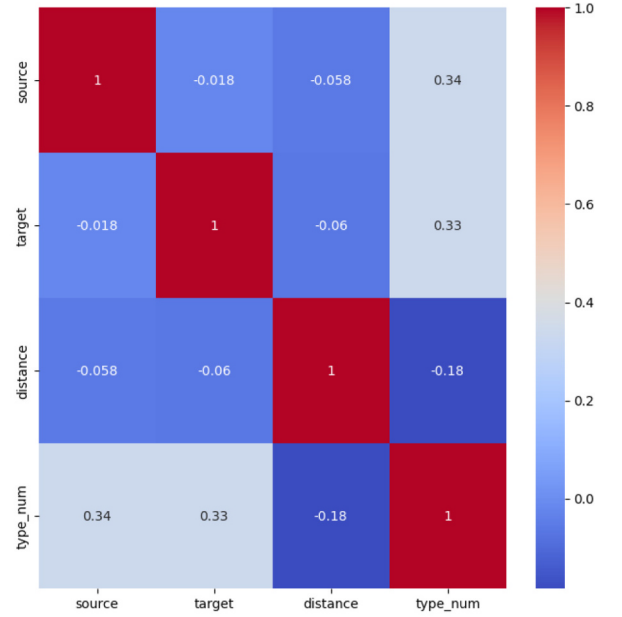


(b) Edge features

Fig. 3. Pearson's correlation coefficients between China railway network features.



(a) Node features



(b) Edge features

Fig. 4. Pearson's correlation coefficients between Paris multi-modal transport network features.

evaluate a network's resilience, which are then applied to two distinct scenarios, i.e., China and Paris transport networks. The removal of a key node (i.e., Haikou) with a moderate degree in the China railway network causes the network to split into two connected subgraphs, although other network properties show minimal change. This underscores the vulnerability of the China railway network's resilience in the area, attributed to the power-law effect, in line with the findings in the literature (Holme et al., 2002; Wang et al., 2020). In contrast, the

removal of the highest-degree node in the Paris multi-modal transport network has little impact on network properties, with the number of connected subgraphs remaining unchanged, indicating the network's high resilience. The attack simulation in this study utilizes both intentional attacks such as degree, betweenness, K-core, and collective influence attacks as well as random attacks. Simulation results show that intentional attacks inflict more damage on the China railway network than random attacks at initial stage, potentially due to the crucial role of

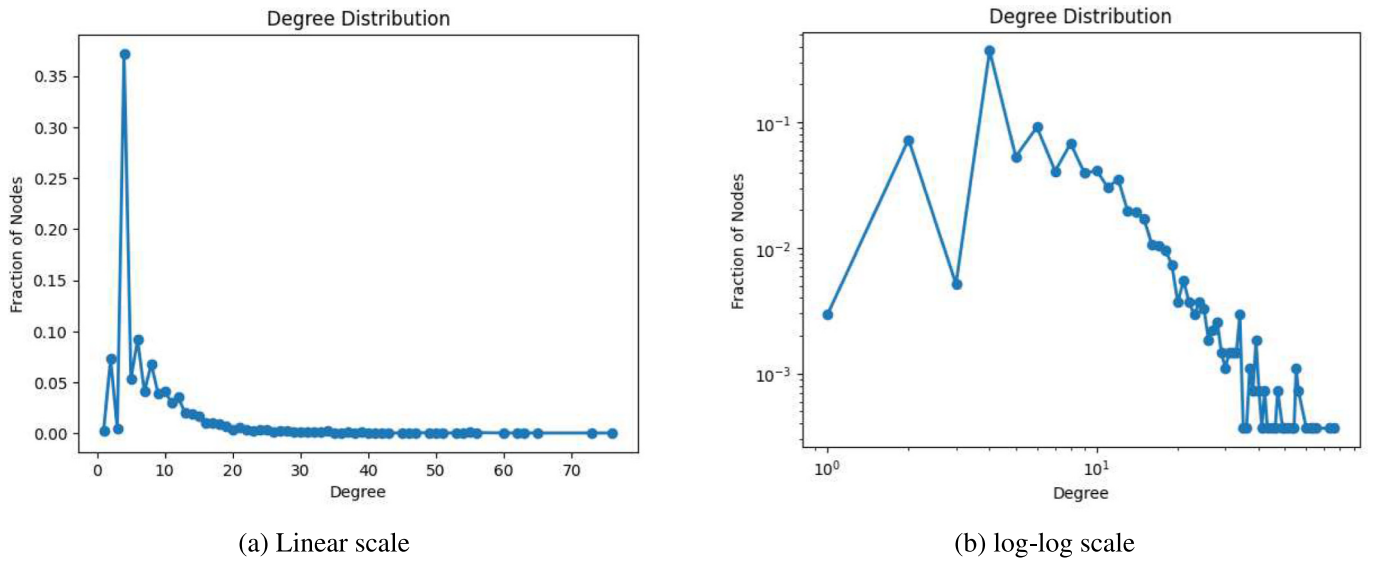


Fig. 5. Degree distribution plots of China railway network.

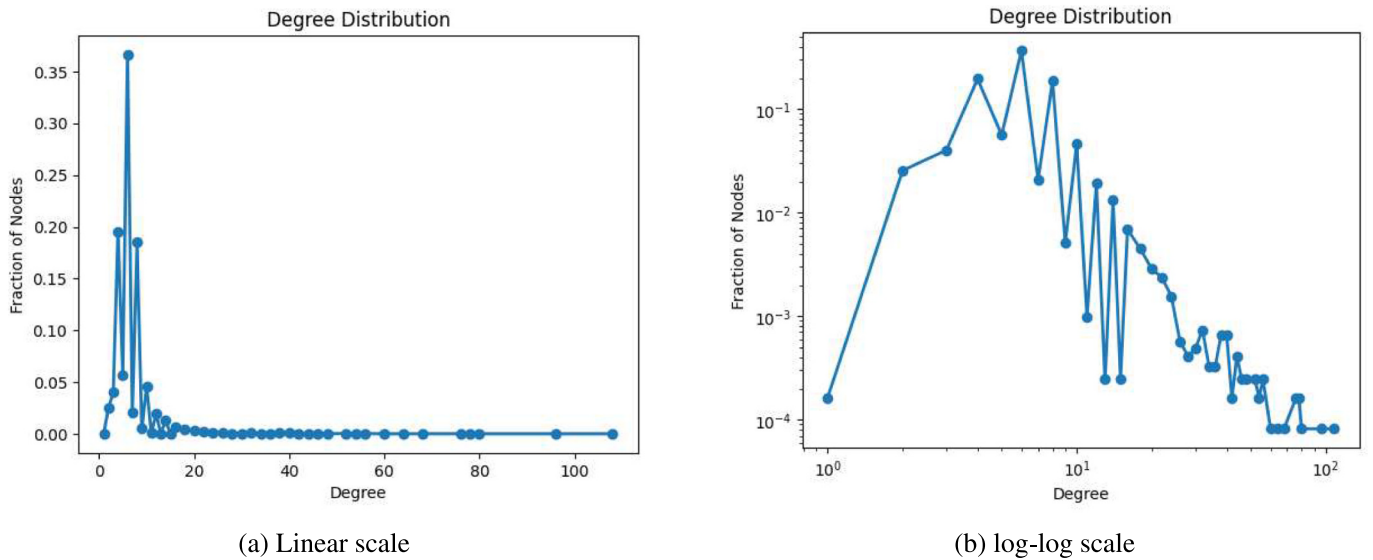


Fig. 6. Degree distribution plots of Paris multi-modal transport network.

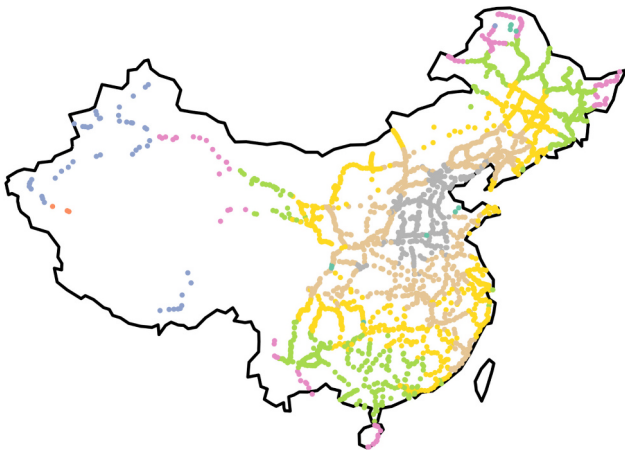


Fig. 7. Clustering based on closeness centrality in China railway network.

hubs, while both types of attacks have low effects on the Paris transport system, further demonstrating the resilience of the network, primarily due to its multi-modal structure. Moreover, compared to the study by Berche et al. (Berche et al., 2009) on the resilience of the Paris transport network, this study shows that the network is more resilient under degree attacks, although it remains the least resilient under the other attack strategies applied in this study. A potential reason for this difference is that Berche et al. considered only the bus and subway networks of Paris, whereas a higher-modality network demonstrates greater robustness (Gao et al., 2011). As societies become increasingly interconnected and reliant on transport networks, understanding how these systems respond to disruptions caused by natural disasters, accidents, or human factors becomes crucial. Resilient transport networks can absorb shocks, maintain functionality during adverse events, and recover quickly, thereby minimizing economic losses and social disruption. This study provides methods and insights for assessing and designing more robust, adaptive transport networks. Nonetheless, additional comparisons with existing literature and thorough model sensitivity analyses to validate and reinforce the robustness of the

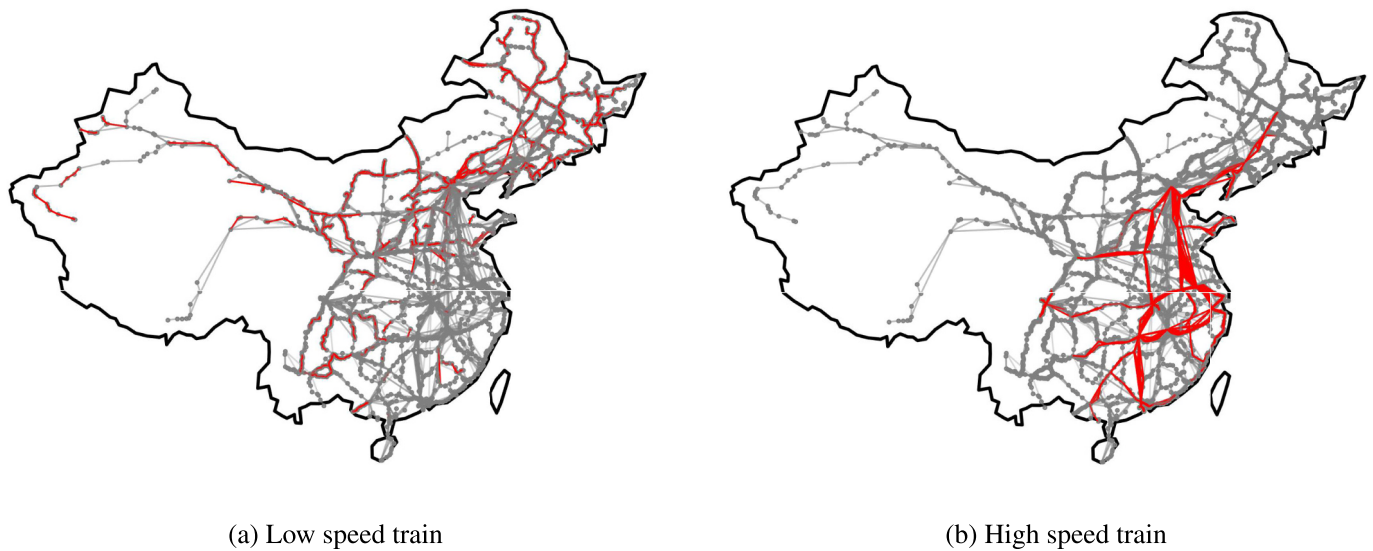


Fig. 8. Edge clustering by route in China railway network.

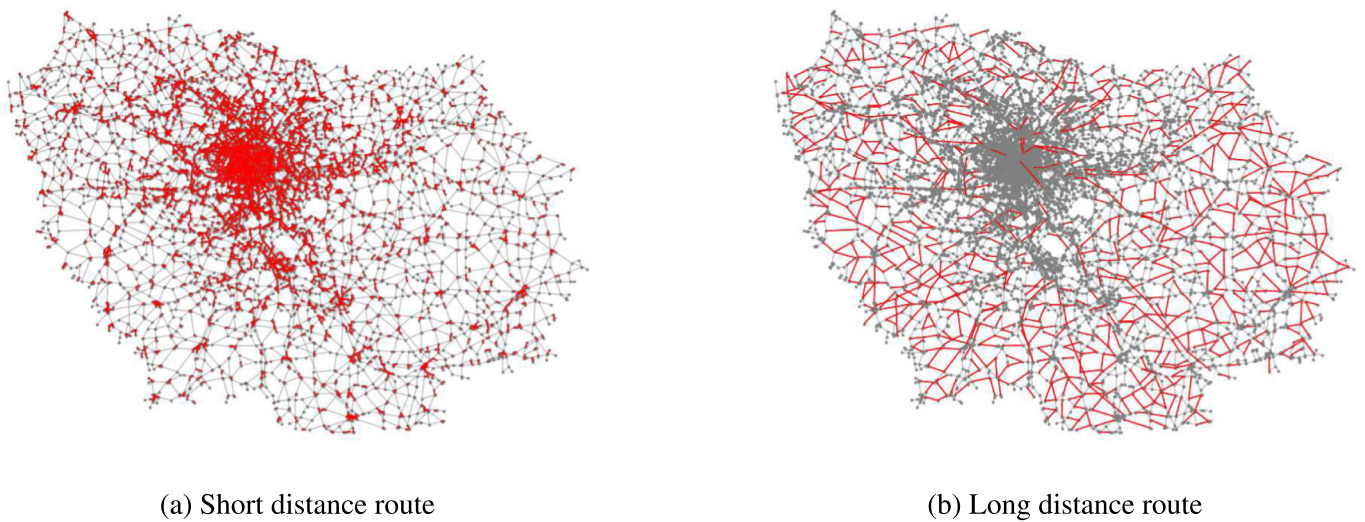


Fig. 9. Edge clustering by distance in Paris multi-modal transport network.

results, are necessary.

This research then develops link prediction models for detecting hidden connections and predicting future links in networks, by leveraging state-of-the-art Graph Convolutional Neural Networks (GCNs) and Multi-Layer Perceptron Networks (MLPs). The data used is augmented with negative samples during the training of the link prediction models. This approach enhances the model's performance by providing additional training data and enabling the observation of network patterns with predicted links. For the China railway network, the prediction results suggest a potential expansion of the high-speed railway connecting Hainan Island to the mainland China, aligning with the finding from network resilience analysis. Despite geographical constraints, the feasibility of such an expansion is supported by Hainan's recent development plans (Xinhua News Agency, 2020). The predicted connections also reveal critical links between southeast coastal areas and between Beijing and the northwestern regions, emphasizing the need to strengthen regional connectivity, which closely aligns with China's National Medium- and Long-Term Railway Network Plan (China National Development and Reform commission, 2016). Notably, these findings are derived solely from the available railway data. For the Paris multi-modal transport network, the predicted links are short and

numerous, reflecting the city's dense node distribution and the regionally segmented nature of its public transport system in the dataset. Then, the study raises the sigmoid threshold from 0.71 to 0.99 in the link prediction model for high probability links, showing the hidden links predicted largely located in more developed regions for both networks, likely driven by strong economic and social interaction demands. However, the datasets are likely biased toward more developed areas, due to the greater number and strength of links within these regions. It is also observed that the hidden links in China network form a transportation artery running from south to north across the country, highlighting the need to strengthen the connections between southern and northern China, consistent with Chinese government's 12th Five-Year Plan for the Revitalization of Northeast China (China National Development and Reform commission, 2012), and an interesting link from southern Paris extending eastward toward northern Seine-et-Marne, indicating a demand for a direct connection and more balanced development between the western and eastern regions, aligning well with the Ile-de-France Mobilités 2030 plans (Ile-de-France Mobilités, 2024). The link feature prediction results indicate that MLP models trained on the Paris network outperform those trained on the China railway network. This may be attributed to the larger dataset and the more dense and

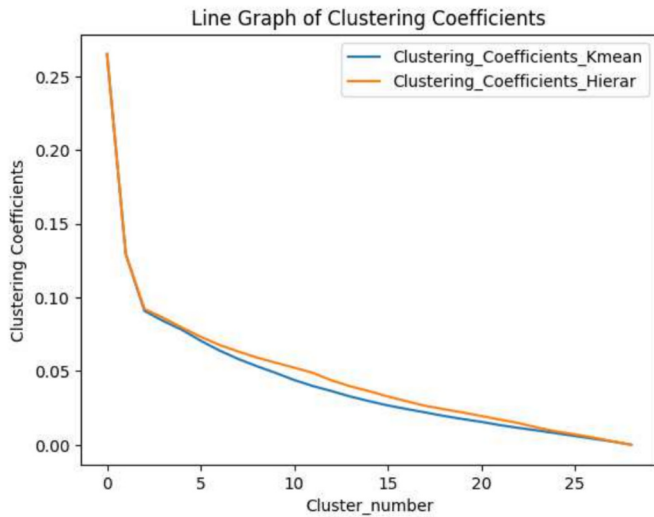


Fig. 10. Comparison of average values between two clustering techniques by clustering coefficients.

complex structure of the Paris transport network. However, further research is needed to explore other potentially influential factors, such as differences in data features and model hyperparameters.

Subsequently, this study randomly generates node locations and predicts new edges and their features using a novel approach based on KD-Tree and MLPs. The results indicate that the predicted maximum speeds within the China railway network align closely with actual train speeds in the corresponding local regions. The exhibition of long-distance and short-distance characteristics of the predicted new edges within the China and Paris transport networks respectively, accurately reflect the geographical and economic features of their corresponding regions. This approach provides an effective tool for local governments in transport planning, particularly for integrating new constructions into the national network.

Overall, the establishment of these new links could significantly enhance interregional accessibility, economic integration, and social cohesion. The findings, which align well with existing development patterns, offer a practical guidance for potential extensions of the current transport network. Furthermore, the link and feature prediction methods developed using GCN, KD-Tree, and MLPs demonstrate strong

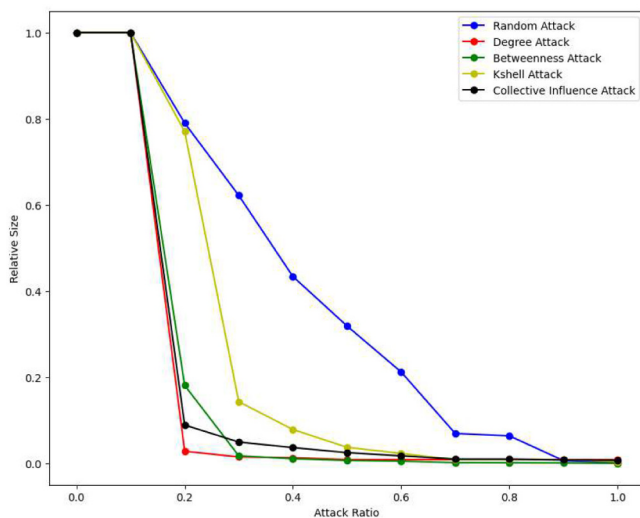
potential for understanding and anticipating the evolution of transport systems. These methods identify structurally important but currently missing connections and help prioritize the planning of new transport lines. Nevertheless, it is important to note that, while the predicted links appear reasonable, practical considerations from economic, geographical, environmental, and social perspectives must also be taken into account in transport planning and construction. Additionally, the link prediction models may require further validation and refinement through historical data analysis and expert input.

The novel framework provides a foundation for further theoretical developments in transport systems and practical applications across diverse use cases and other network domains. The study primarily focuses on methodological research into complex transport network analysis using complex network theory and Neural Networks, with two specific scenarios. It is not intended as a comprehensive comparison of the China railway network and the Paris transport network, but rather as a means of testing the generality of the proposed framework. Any comparative analysis in the study should therefore be interpreted cautiously, as a national railway system cannot be directly equated with a multi-modal urban transport system. In addition, incorporating richer node and edge information such as geographical and economic factors as well as road conditions and traffic flows into the two scenarios may further enhances the analytical and predictive capabilities of the framework. The results may also be improved through refinement and optimization of its components.

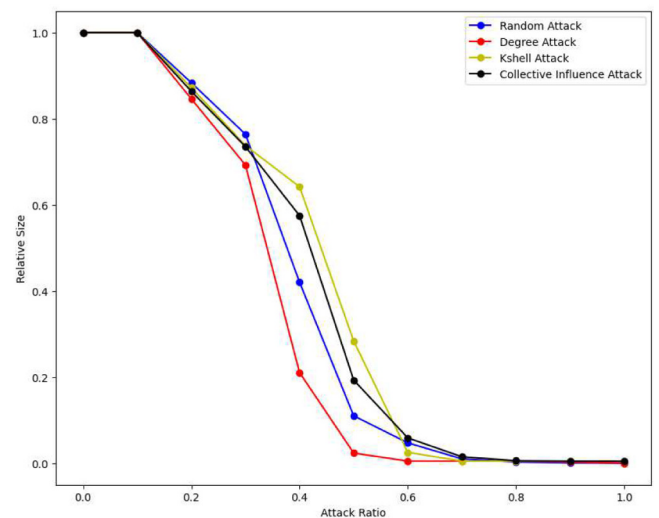
6. Conclusion and future work

This research develops a novel and generic Machine Learning-driven framework for the comprehensive complex network analysis of transport systems, integrating theoretical Machine Learning and complex network analysis methods in four key perspectives, namely, Network property analysis, community detection, resilience assessment and link prediction. It is applied to two real-world transport network scenarios, i. e., the China railway network and the Parisian multi-modal transport system. Specifically, the study analyses the properties, clustering, resilience, and potential links of these two networks to support transport management and planning.

The analysis of network properties for the China railway network and the Paris multi-modal network reveals distinct complex network features in service purpose, network scale, density, and efficiency, yet both demonstrate a power-law distribution in their node degrees. The

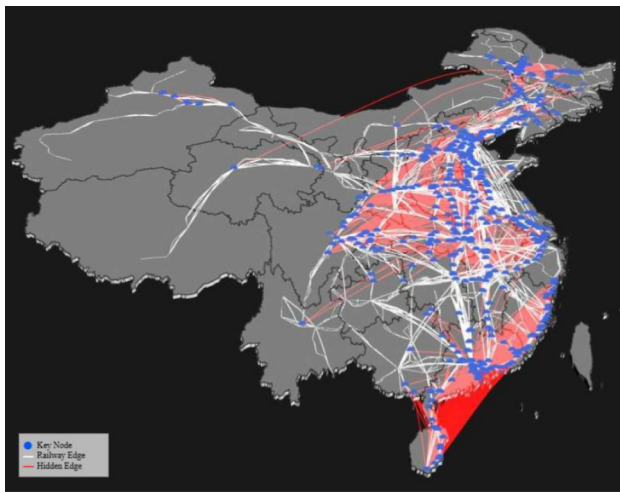


(a) State change of China network

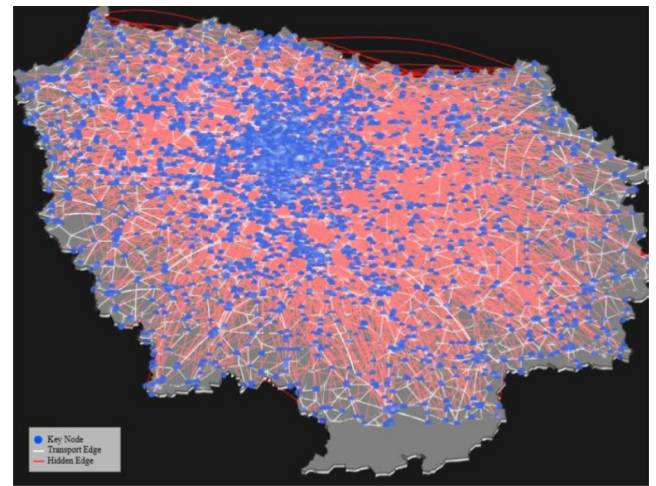


(b) State change of Paris network

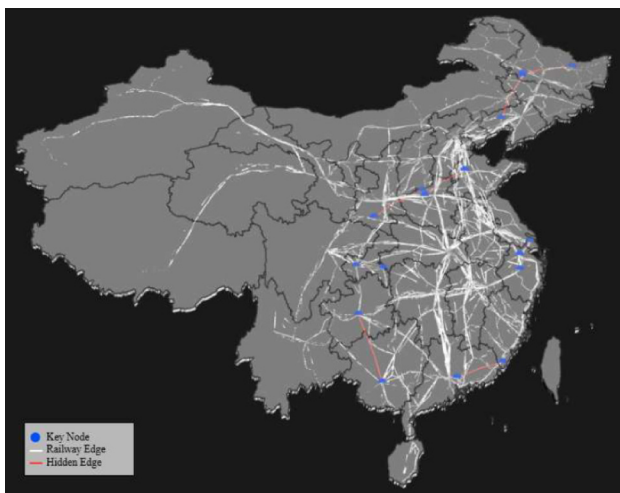
Fig. 11. State change of two transport networks under simulated attacks.



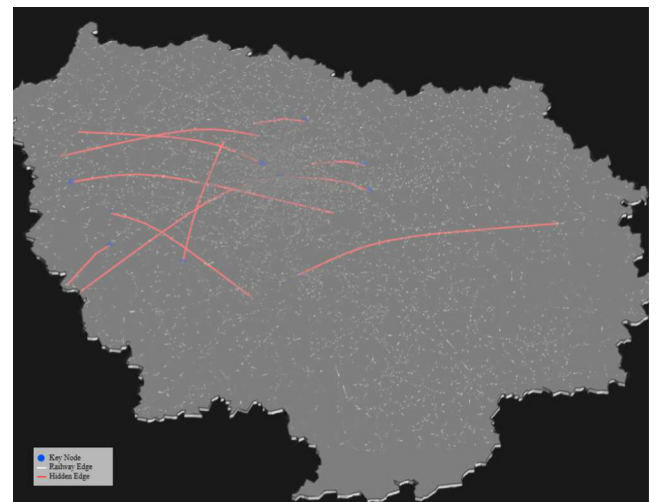
(a) Predicted hidden edges in the China railway network



(b) Predicted hidden edges in the Paris transport network

Fig. 12. Link prediction using the encoder-decoder model ($T = 0.71$).

(a) Predicted hidden edges in the China railway network



(b) Predicted hidden edges in the Paris transport network

Fig. 13. Link prediction using the encoder-decoder model ($T = 0.99$).

clustering results based on K-means, Hierarchical clustering and various node and edge properties exhibits a pattern of concentric circles, radiating outward from urban centers toward peripheral cities in China railway network, while a high density of short-distance connections within central Paris and a prevalence of long-distance connections in the outskirts. Meanwhile, a comparison of K-means and Hierarchical clustering techniques reveals comparable performance across various complex network properties. Network resilience analysis examines the two networks' dynamic characteristics and structural evolution under escalating node attacks, with two strategies, i.e., random and intentional attacks. The attack simulation results show that intentional attacks cause more damage to the China railway network at the initial stage due to the importance of hubs, while both attack types have minimal effects on the Paris transport system, highlighting its resilience, mainly due to its multi-modal structure. This research also develops link prediction models for detecting hidden connections and predicting future links in networks based on Graph Convolutional Neural Networks and Multi-Layer Perceptron Networks. The predictions reveal critical links between Hainan Island and mainland China as well as between Beijing and the northwestern regions in China railway network, while shorter and

more evenly distributed predicted links across Paris multi-modal network. The results of this study align closely with existing literature and official transport development plans.

Overall, the study advances understanding of transport networks' static and dynamic behaviors, and aiding strategic planning and optimization in transport management. Moreover, the case studies of the China railway and Paris transport networks may offer new insights into the advancement of Machine Learning-driven complex network analysis for transport systems across different economies. Future work will incorporate additional network features to improve model accuracy, investigate the impact of node deletions on shortest paths, and conduct uncertainty and sensitivity analyses across more transport network scenarios to strengthen the framework's robustness.

CRediT authorship contribution statement

Yuqin Xia: Writing – original draft, Software, Methodology, Investigation, Formal analysis. **Kewei Wang:** Writing – original draft, Software, Methodology, Formal analysis, Data curation. **Purin Tanirat:** Writing – original draft, Software, Methodology, Formal analysis. **Bryan**

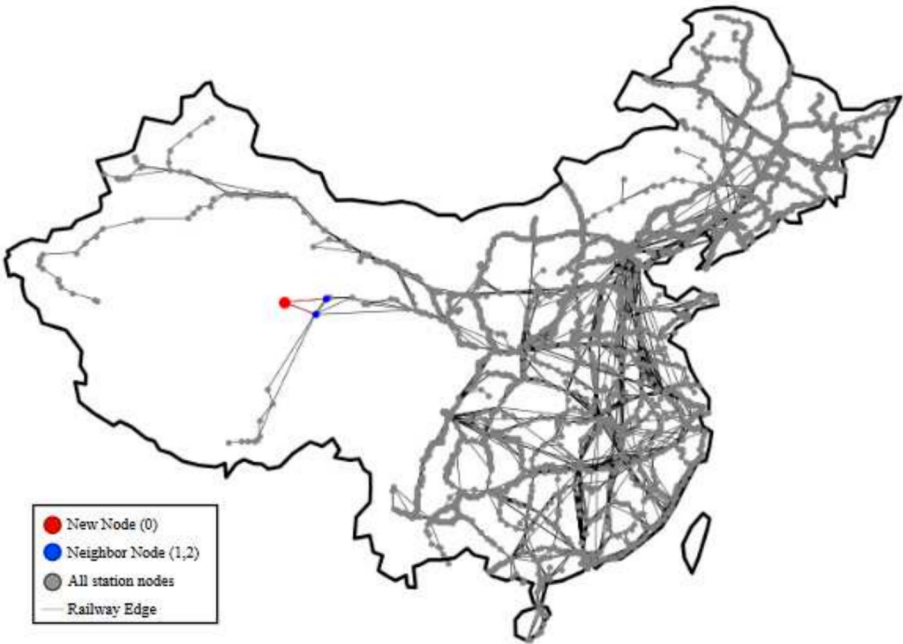


Fig. 14. Visualised edge feature prediction example for the China railway network.

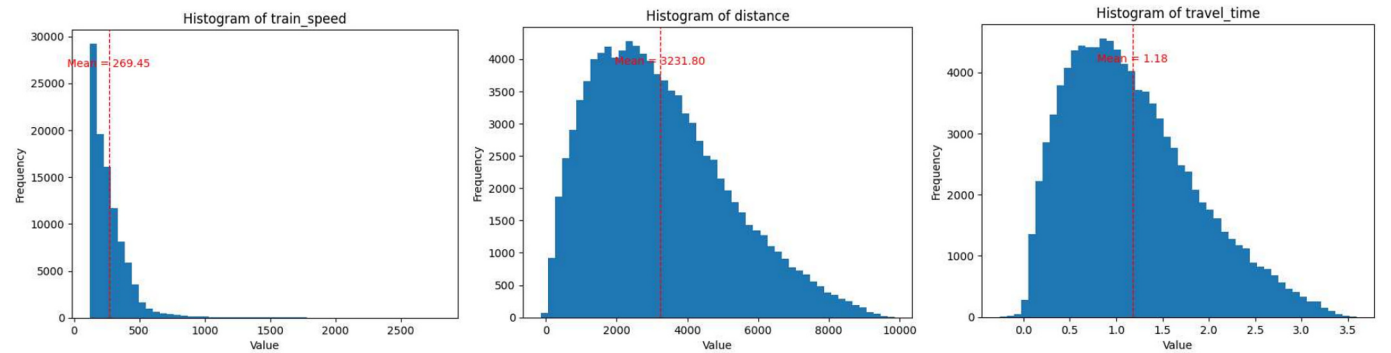


Fig. 15. Predicted edge feature distributions of China railway network.

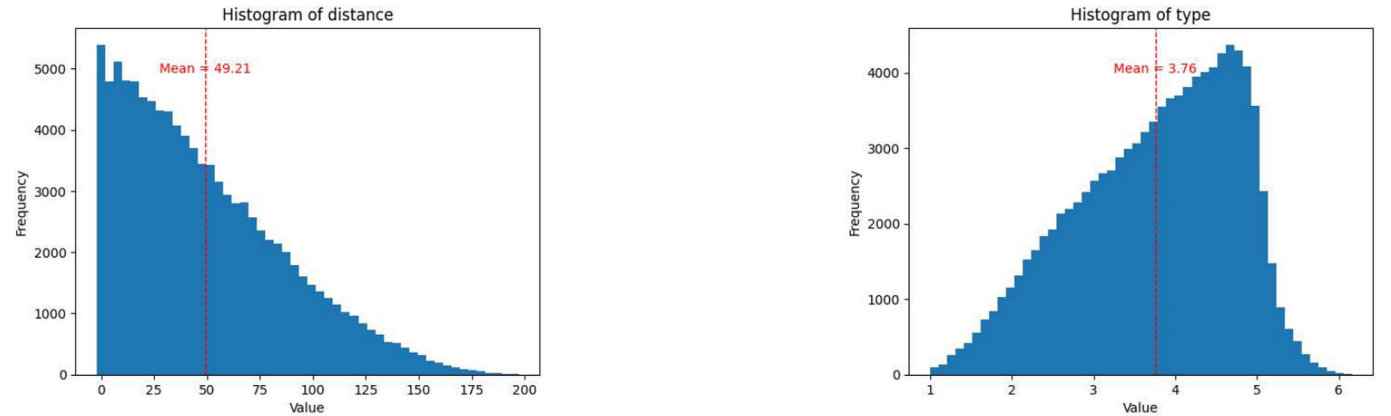


Fig. 16. Predicted edge feature distributions of Paris multi-modal transport network.

Lee: Writing – original draft, Software, Methodology, Formal analysis. Conceptualization.
Irene Moulitsas: Supervision, Resources, Project administration, Conceptualization.
Jun Li: Writing – review & editing, Validation, Supervision, Resources, Project administration, Methodology,

Table 10

Feature prediction accuracies of the MLP models.

Dataset	Prediction accuracy				
	Degree	Distance	Travel time	Train speed	Edge type
China Railway	54.63 %	38.88 %	89.94 %	65.57 %	–
Paris Transport	71.66 %	90.88 %	–	–	99.92 %

Data availability

Data will be made available on request.

References

- Agrawal, Smita, Patel, Atul, 2020. Clustering algorithm for community detection in complex network: a comprehensive review. *Recent Adv. Comput. Sci. Commun.* 13 (4), 542–549 (Formerly: *Recent Patents on Computer Science*).
- Ait Rai, K., Machkour, M., Antari, J., 2023. Influential nodes identification in complex networks: a comprehensive literature review. *Beni-Suef Univ. J. Basic Appl. Sci.* 12, 18.
- Asgari, Fereshteh, Sultan, Alexis, Xiong, Haoyi, Gauthier, Vincent, El-Yacoubi, Mounim A., 2016. CT-mapper: mapping sparse multimodal cellular trajectories using a multilayer transportation network, 95, 69–81.
- Behrooz, Hojat, Hayeri, Yeganeh M., 2022. Machine learning applications in surface transportation systems: a literature review. *Appl. Sci.* 12 (18).
- Berche, Bertrand, Von Ferber, Christian, Holovatch, Taras, Holovatch, Yu, 2009. Resilience of public transport networks against attacks. *Eur. Phys. J. B* 71, 125–137.
- Boccaletti, S., Latora, V., Moreno, Y., Chavez, M., Hwang, D.U., 2006. Complex networks: structure and dynamics, 424 (4), 175–308.
- Breiman, Leo, 2001. Random forests. *Mach. Learn.* 45 (1), 5–32.
- CERTIFER, 2021. Eole Project. <https://www.certifer.eu/achievement/france-eole-proje-ct/>. Published: 2021-05.
- China Academy Of Railway Sciences Corporation Limited, 2022. 12306 China Railway. <https://www.12306.cn/en/index.html>. Published: 2022-10-06.
- China National Development and Reform commission, 2012. The 12th Five-Year Plan for the Revitalization of Northeast China. <https://www.ndrc.gov.cn/>. Published: 2012-03-22.
- China National Development and Reform commission, 2016. China's National Medium-and Long-Term Railway Network Plan. https://www.ndrc.gov.cn/fggz/fzzlgh/gjjz_xgh/200806/P020191104623823270325.pdf.
- China State Railway Group Corporation Limited, 2022. China Railway. <http://www.china-railway.com.cn/english/>. Published: 2022-10-06.
- Crucitti, Paolo, Latora, Vito, Marchiori, Massimo, 2004. Model for cascading failures in complex networks, 69 (4), 045104. Publisher: American Physical Society.
- Cubukcu, K. Mert, 2021. Using circuitry as a network efficiency measure: the example of Paris. *Spat. Inf. Res.* 29 (2), 163–172.
- Gao, Jianxi, Buldyrev, Sergey V., Havlin, Shlomo, Stanley, H. Eugene, 2011. Robustness of a network of networks, 107 (19), 195701. Publisher: American Physical Society.
- Gao, J., Liu, X., Li, D., Havlin, S., 2015. Recent progress on the resilience of complex networks. *Energies* 8 (10), 12187–12210.
- Freeman, Linton C., 1977. A set of measures of centrality based on betweenness, 40 (1), 35–41. Publisher: [American Sociological Association, Sage Publications, Inc.].
- Gleeson, James P., Onaga, Tomokatsu, Fennell, Peter, Cotter, James, Burke, Raymond, O'Sullivan, David J.P., 2021. Branching process descriptions of information cascades on twitter, 8 (6). Publisher: Oxford Academic.
- Greene, Derek, Cunningham, Pádraig, Mayer, Rudolf, 2008. Unsupervised learning and clustering. In: Cord, Matthieu, Cunningham, Pádraig (Eds.), *Machine Learning Techniques for Multimedia: Case Studies on Organization and Retrieval, Cognitive Technologies*. Springer, pp. 51–90.
- Haghighan, Seyed Arman, Geroliminis, Nikolas, Akbarzadeh, Meisam, 2021. Community detection in large scale congested urban road networks. *PLoS One* 16 (11), e0260201.
- Háznagy, Andor, Fi, István, London, András, Nemeth, Tamás, 2015. Complex network analysis of public transportation networks: a comprehensive study. In: 2015 International Conference on Models and Technologies for Intelligent Transportation Systems (MT-ITS). IEEE, pp. 371–378.
- Holme, Petter, Kim, Beom Jun, Yoon, Chang No, Han, Seung Kee, 2002. Attack vulnerability of complex networks, 65 (5), 056109. Publisher: American Physical Society.
- Huang, Shao Lun, Chen, Kwang Cheng, 2015, September. Information Cascades in Social Networks Via Dynamic System Analyses, pp. 1262–1267. ISBN: 9781467364324. Publisher: Institute of Electrical and Electronics Engineers Inc.
- Huang, Liping, Yang, Yongjian, Gao, Hepeng, Zhao, Xuehua, Du, Zhanwei, 2018. Comparing community detection algorithms in transport networks via points of interest. *IEEE Access* 6, 29729–29738.
- Île-de France Mobilités, 2024. Ile-de-France Mobilités 2030 Plans - Regional Initiatives. https://www.iledefrance.fr/sites/default/files/2024-04/Plan_Mobilités_6_P_DMIF_Stratégie.pdf. Published: 2024-02-06.
- Jiang, Weiwei, Luo, Jiayun, 2022. Graph neural network for traffic forecasting: a survey. *Expert Syst. Appl.* 207, 117921.
- Jiao, Jingjuan, Wang, Jiaoe, Jin, Fengjun, 2017. Impacts of high-speed rail lines on the city network in China. *J. Transp. Geogr.* 60, 257–266.
- Kitsak, Maksim, Gallos, Lazaros K., Havlin, Shlomo, Liljeros, Fredrik, Muchnik, Lev, Stanley, H. Eugene, Makse, Hernán A., 2010. Identification of influential spreaders in complex networks, 6 (11), 888–893. Number: 11 Publisher: Nature Publishing Group.
- Kölker, Katrin, Stoebe, Emily, Lütjens, Klaus, 2025. In the flow of connectivity: a dynamic network analysis of europe's air transportation system. *J. Air Transp. Res. Soc.* 4, 100055.
- Kumar, Ajay, Singh, Shashank Sheshar, Singh, Kuldeep, Biswas, Bhaskar, 2020. Link prediction techniques, applications, and performance: a survey. *Phys. A: Stat. Mech. Appl.* 553, 124289.
- Li, Wei, Cai, Xu, 2004. Statistical analysis of airport network of China. *Phys. Rev. E* 69 (4), 046106.
- Li, Zhongcan, Huang, Ping, Wen, Chao, Dong, Wei, Ji, Yindong, Rodrigues, Filipe, 2024. Railway network delay evolution: a heterogeneous graph neural network approach. *Appl. Soft Comput.* 159, 111640.
- Liang, Xiaozhen, Wang, Yingying, Yang, Mingge, 2024. Systemic modeling and prediction of port container throughput using hybrid link analysis in complex networks. *Systems* 12 (1).
- Lin, Jingyi, Ban, Yifang, 2013. Complex network topology of transportation systems. *Transp. Rev.* 33 (6), 658–685.
- Masui, Tomonori, 2020. Graph Neural Networks with PYG on Node Classification, Link Prediction, and Anomaly Detection: Pytorch Geometric Implementations on Major Graph Problems.
- Newman, M.E.J., 2003. The structure and function of complex networks, 45 (2), 167–256. Publisher: Society for Industrial and Applied Mathematics Publications.
- Oubaalla, Wafaa, Benhlila, Laila, 2018. An overview of community detection methods in transportation networks. In: *Proceedings of the 12th International Conference on Intelligent Systems: Theories and Applications*, pp. 1–6.
- Paszke, Adam, Gross, Sam, Massa, Francisco, Lerer, Adam, Bradbury, James, Chanan, Gregory, Killeen, Trevor, Lin, Zeming, Gimelshein, Natalia, Antiga, Luca, et al., 2019. Pytorch: an imperative style, high-performance deep learning library. In: *Advances in Neural Information Processing Systems*, 32. Curran Associates, Inc.
- Rahmani, Saeed, Baghbani, Asiye, Bouguila, Nizar, Patterson, Zachary, 2023. Graph neural networks for intelligent transportation systems: a survey. *IEEE Trans. Intell. Transp. Syst.* 24 (8), 8846–8885.
- Ramchoun, Hassan, Ghanou, Youssef, Ettaouil, Mohamed, Idrissi, Mohammed Amine Janati, 2016. Multilayer perceptron: architecture optimization and training. *Int. J. Interact. Multimedia Artif. Intell.* 4 (1). Accepted: 2021-07-07T10:37:59Z.
- Schaller, Thibaud, Li, Jun, Jenkins, Karl W., 2025. A data-driven approach for automatic aircraft engine borescope inspection defect detection using computer vision and deep learning. *J. Exp. Theor. Anal.* 3 (1).
- Shahapure, Ketan Rajshekhar, Nicholas, Charles, 2020. Cluster quality analysis using silhouette score. In: 2020 IEEE 7th International Conference on Data Science and Advanced Analytics (DSAA), pp. 747–748.
- Skródzki, Martin, 2019. The K-D Tree Data Structure and a Proof for Neighborhood Computation in Expected Logarithmic Time. volume abs/1903.04936.
- Société des grands projets, 2025. Grand Paris Express. <https://www.grandparisexpress.fr/benefit-metro-ile-de-france>. Accessed: 2025-04-20.
- Srinivasan, Srinikethan Madapuzi, Truong-Huu, Tram, Gurusamy, Mohan, 2019. Machine learning-based link fault identification and localization in complex networks. *IEEE Internet Things J.* 6 (4), 6556–6566.
- Strogatz, Steven H., 2001. Exploring complex networks. *Nature* 410 (6825), 268–276.
- Wang, Wei, Cai, Kaiquan, Du, Wenbo, Wu, Xin, Tong, Lu Carol, Zhu, Xi, Cao, Xianbin, 2020a. Analysis of the chinese railway system as a complex network. *Chaos, Solitons Fractals* 130, 109408.
- Wang, Jiaoe, Mo, Huihui, Wang, Fahui, Jin, Fengjun, 2011. Exploring the network structure and nodal centrality of china's air transport network: a complex network approach, 19 (4), 712–721.
- Wang, Liang, Xue, Xiaolong, Xue, Weirui, Yuan, Hongping, 2020. Measuring the resilience of China railway network. In: *Construction Research Congress 2020*. American Society of Civil Engineers, Reston, VA, pp. 278–284.
- Wang, Xiao Fan, Chen, Guanrong, 2003. Complex networks: small-world, scale-free and beyond, 3 (1), 6–20. Conference Name: IEEE Circuits and Systems Magazine.
- Wei, Liu, Ling, Chen, 2020. Link prediction in complex networks. *Inf. Control.* 49 (1), 1–23.
- Wikipedia, 2025. Tramways in Île-de-France. https://en.wikipedia.org/wiki/Tramways_in_%C3%8Ede-France. Accessed: 2025-04-20.
- Xinhua News Agency, June 2020. Hainan Free Trade Port Construction General Plan. Retrieved from. The Central People's Government of the People's Republic of China. English version available at Xinhua Net <http://www.xinhuanet.com/english/2020-06/02/c.139105952.htm>.
- Yao, Xin, Gao, Yong, Zhu, Di, Manley, Ed, Wang, Jiaoe, Liu, Yu, 2021. Spatial origin-destination flow imputation using graph convolutional networks. *IEEE Trans. Intell. Transp. Syst.* 22 (12), 7474–7484.
- Zhang, Si, Tong, Hanghang, Xu, Jiejun, Maciejewski, Ross, 2019. Graph convolutional networks: a comprehensive review. *Comput. Soc. Netw.* 6 (1), 1–23.
- Zhang, Dong-ming, Du, Fei, Huang, Hongwei, Zhang, Fan, Ayyub, Bilal M., Beer, Michael, 2018. Resiliency assessment of urban rail transit networks: Shanghai metro as an example, 106, 230–243.

- Zhang, Zi-Ke, Liu, Chuang, Zhan, Xiu-Xiu, Lu, Xin, Zhang, Chu-Xu, Zhang, Yi-Cheng, 2016. Dynamics of information diffusion and its applications on complex networks, 651, 1–34.
- Zheng, Henghe, Zhao, Haitao, Ahmadi, Gholamreza, 2024. Towards improving community detection in complex networks using influential nodes. *J. Complex Netw.* 12 (1), cnae001.
- Zhou, Fan, Yang, Qing, Zhong, Ting, Chen, Dajiang, Zhang, Ning, 2020. Variational graph neural networks for road traffic prediction in intelligent transportation systems. *IEEE Trans. Industr. Inform.* 17 (4), 2802–2812.



DSPACE

<https://dspace.org/>

Machine Learning driven complex network analysis of transport systems

Xia, Yuqin; Wang, Kewei; Tanirat, Purin; Lee, Bryan; Moulitsas, Irene ...

2025-07

Elsevier

<https://doi.org/10.1016/j.jtrangeo.2025.104270>

Xia Y, Wang K, Tanirat P, et al., (2025) Machine Learning driven complex network analysis of transport systems. Journal of Transport Geography, Volume 127, July 2025, Article number 104270