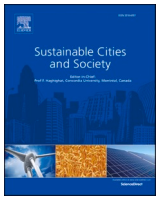
Sustainable Cities and Society 118 (2025) 106015

Contents lists available at ScienceDirect

Sustainable Cities and Society

journal homepage: www.elsevier.com/locate/scs



Predicting origin-destination flows by considering heterogeneous mobility patterns

Yibo Zhao [a,b,c], Shifen Cheng [a,b,*], Song Gao [c], Peixiao Wang [a,b], Feng Lu [a,b,d,e]

[a] State Key Laboratory of Resources and Environmental Information System, Institute of Geographic Sciences and Natural Resources Research, CAS, Beijing 100101, China
[b] University of Chinese Academy of Sciences, Beijing 100049, China
[c] Geospatial Data Science Lab, Department of Geography, University of Wisconsin-Madison, Madison, WI, 53706, USA
[d] The Academy of Digital China, Fuzhou University, Fuzhou China
[e] Jiangsu Center for Collaborative Innovation in Geographical Information Resource Development and Application, Nanjing 210023, China

ARTICLE INFO

Keywords:
Origin-destination flow
Spatial interaction
Urban mobility
Spatial heterogeneity
Imbalanced data learning
Graph attention network

ABSTRACT

The accurate prediction of origin-destination (OD) flows is essential for advancing sustainable urban mobility and supporting resilient urban planning. However, the inherent heterogeneity of mobility patterns results in complex geographic unit relations, diverse spatial organizational structures, and the long-tailed effect on OD flow distribution. This study proposes a novel OD flow prediction method based on graph-based deep learning (named as HMCG-LGBM). Specifically, 1) a modularity-based graph reconstruction strategy is presented for geographic unit relation augmentation by eliminating weak connections; 2) the heterogeneous spatial organization of OD flows is captured by combining the community detection approach and graph attention mechanism with the introduction of socio-economic and spatial features; and 3) a weighted loss function with distribution smoothing paradigm is developed to enhance the prediction for low-probability mobility events, addressing the challenges posed by long-tailed distributions. Extensive experiments conducted on real-world datasets show that the predictive performance of the proposed method is significantly improved, with the RMSE and MAE reduced from the baselines by 11.1%–33.3% and 14.1%–22.2%, respectively. The results also demonstrate the robustness of the proposed method for dealing with imbalanced OD flow distributions, providing valuable insights for spatial interaction predictive modeling in the context of sustainable urban systems.

1. Introduction

The movements of population and materials between origin and destination locations reflect the spatial interactions among geographic units, which are essential for advancing sustainable urban mobility, fostering synchronized regional development, and optimizing transportation management (Barthelemy, 2011; Liu, Yao et al., 2020). Origin-destination (OD) flow prediction involves estimating the flow volume between geographic units using historical OD flow data and the properties and relationships of these units (Barbosa et al., 2018; Rao et al., 2018). The prediction is vital for various urban applications, such as predicting future patterns of urban mobility, formulating urban development strategies, guiding traffic diversion, identifying abnormal urban incidents, and planning transportation infrastructure (Davidich et al., 2021, Liu et al., 2022; Shi, Wang, Xu, & Wang, 2022; Wang, Zhang et al., 2023). Additionally, OD flow prediction plays a supportive role in addressing societal issues such as traffic pollution and epidemic spreading, thereby supporting the development of resilient and sustainable urban systems (Jia et al., 2020; Xu et al., 2021).

Existing OD flow prediction methods include knowledge-driven theoretical models and data-driven machine learning models. Theoretical or mechanistic models are developed from physical principles and optimization theories (Lenormand et al., 2016; Schlapfer et al., 2021). These models are extensively employed to investigate the mechanisms of mobility flows, due to their simplicity and interpretability (Ren, Ercsey-Ravasz, Wang, Gonzalez, & Toroczkai, 2014; Zhao, Hu, Zeng, Chen, & Ye, 2023, Zhang & Li, 2024). Typical theoretical models include the gravity model (Ravenstein, 1885), the radiation model (Simini et al.,

* Corresponding author.
E-mail address: chengsf@lreis.ac.cn (S. Cheng).

https://doi.org/10.1016/j.scs.2024.106015
Received 20 August 2024; Received in revised form 16 November 2024; Accepted 25 November 2024
Available online 26 November 2024
2210-6707/© 2024 Elsevier Ltd. All rights are reserved, including those for text and data mining, AI training, and similar technologies.

2012), and the intervention opportunity model (Stouffer, 1940). Machine learning models include two main categories: classical machine learning models and deep learning models. The former employs traditional machine learning algorithms based on kernel or tree structures, such as support vector machines (Sana et al., 2018), gradient boosted regression trees (Friedman, 2001), and random forests (Breiman, 2001; Spadon et al., 2019). The latter typically employs neutral networks based on graph structures, such as GMEL (Liu, Miranda et al., 2020), SI-GCN (Yao et al., 2021), SpatialGAT (Cai et al., 2022), ConvGCN-RF (Yin et al., 2023), and GODDAG (Rong et al., 2023). The graph neural network model has gained popularity for predicting OD flows due to its ability to capture complex relations among geographic units and perform nonlinear fitting. Recently, efforts on trajectory modeling and OD flow prediction have been made by combining theoretical models with machine learning and generative AI approaches, such as DeepGravity (Simini et al., 2021), CATS (Rao et al., 2023), and Act2Loc (Liu et al., 2024).

However, variations in resource allocation among geographic units and disparities in movement demand typically result in heterogeneity in the mobility patterns (Yang et al., 2019; Yang et al., 2023). This heterogeneity is primarily observed in two aspects: the numerical distribution of the flow volumes and the spatial distribution of the flows. The numerical distribution heterogeneity is manifested in the statistical long-tailed distribution of flow volumes (Fig. 1(a)). Several geographic units exhibit negligible or no flow volume, whereas a minority of geographic units display considerable flow volume (Jiang et al., 2023). The spatial distribution heterogeneity refers to an imbalanced geographical distribution of OD flows, typically characterized by aggregation patterns (Fig. 1(b)). These characteristics of mobility flows pose the following challenges for current methods:

(1) The difficulties in representing the intricate geographic unit relations: The relationships between geographic units are complex and diverse, including attribute relations, spatial adjacencies, semantic relations, etc. These relationships are critical for OD flow prediction, as understanding the interaction between geographic units helps in capturing mobility patterns that cannot be directly captured from fusing attribute data alone (Yao et al., 2021; You et al., 2024). Constructing a graph structure is an effective approach for representing these relationships, where the graph nodes represent geographic units and the edge weights indicate the strength of their interactions. Current approaches for generating graph structures primarily rely on historical flow data (Cai et al., 2022), considering any flow value greater than zero as a connection between geographical units. Nevertheless, it is often the case that there is extremely little flow volume between several geographic units due to the long-tailed distribution (i.e., the head of Fig. 1(a)). This leads to the presence of weak connections in the graph (Zhuang et al., 2022). Several weak connections contribute to the complexity and size of the graph structure, thus amplifying the difficulties in representing the relationship of geographic units and the predictive capabilities of the model.

(2) The non-negligible influence of heterogeneous spatial organization: One typical type of heterogeneous spatial organization is the presence of several independent and internally cohesive subgraphs (i.e., communities) (Fig. 1(b)). Distinct subgraphs may exhibit notable variations in their mobility patterns. These variations are influenced by the socioeconomic characteristics and spatial relations of the geographic units, such as resources and policies (Yang et al., 2024). Current methods mostly apply a global model to learn the mobility patterns of OD flows, ignoring the variations across subgraphs caused by diverse socioeconomic and spatial characteristics of geographic units, resulting in limited predictive performance.

(3) The obstacle lies in capturing the long-tailed effect: The heterogeneous distribution of OD flow volumes exhibits a long-tailed effect (Fig. 1(a)). Existing research has endeavored to enhance modeling low-probability OD flows by using the weighted loss function, with the inverse of the flow distribution probabilities as weights (Cai et al., 2022). However, the density distribution of imbalanced OD flows tends to be highly discrete. The probability of certain mobility events occurring is nearly negligible, which is directly ignored in the training process when using the inverse of the distribution probabilities as weights, since their inverse is undefined. The discrete problem becomes more apparent when the mobility patterns exhibit strong heterogeneity, which makes it challenging to fully capture the long-tailed effects of OD flows.

To overcome the above challenges, this study proposes a learning framework based on community detection and graph attention mechanisms for OD flow prediction. The framework aims to make the prediction more accurate and reliable by considering the intricate relations between geographic units, the heterogeneous spatial organization, and the influence of the long-tailed effect. Notably, this study specifically focuses on inferring unobserved OD flows using regional attributes and observable OD flows, which differs from "future flow prediction," a task that aims to forecast OD flows over a specified period using historical data. The main contributions of this study are summarized as follows:

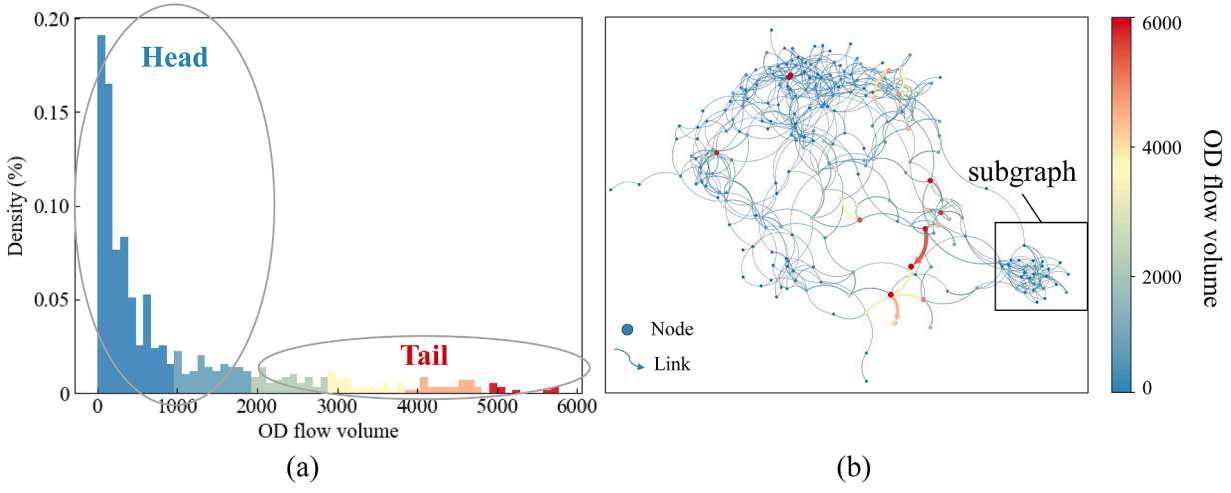


Figure 1. Diagram of heterogeneous origin - destination (OD) flows, (a) density distribution histogram of OD flow volume, (b) OD flow graph.

(1) A graph reconstruction strategy is proposed for geographic unit relation augmentation, which eliminates weak connections based on modularity. The reconstructed graph enables the efficient sharing of geographic unit characteristics in graph neural networks, thereby enhancing training efficiency and prediction quality.
(2) A heterogeneous spatial structure encoding module using graph-based deep learning is introduced, which integrates community detection and graph attention mechanisms to identify and prioritize various spatial organizations during model training. This module boosts the model's capacity to understand the heterogeneous spatial distributions of OD flows.
(3) A weighted loss function with a distribution smoothing paradigm is developed, which allows the model to learn low-probability mobility events by considering the characteristics of adjacent mobility events. This enhances the model's robustness in handling imbalanced flow distributions.
(4) Extensive experiments based on real-world datasets with different spatial scales and individuals are conducted to validate the performance of the proposed method. The results indicate that our method not only surpasses existing methods, but also exhibits the robustness for handling imbalanced OD flow distributions.

2. Preliminary

Definition 1 (Region). A region $\mathscr{R}$ is the spatial unit for carrying the mobility flows, typically represented by a city, province, or country. $\mathscr{R}$ consists of several subregions. Numerous people or objects move between subregions, forming various kinds of flows.

Definition 2 (Relation Graph). A relation graph $G = (V, E)$ represents the connection between subregions over region $\mathscr{R}$, where $V = \{v_1, \ldots, v_N\}$ is a set of subregions (nodes) with the size of $|V| = N$. $E = \{e_{ij} \mid 1 \le i, j \le N\}$ is a set of edges between two nodes. An edge represents the connection between two nodes, which can be generated by their spatial adjacency or, as referred to in this work, by the existence of the historical flows. $e_{ij} = 1$ if there are historical flows between node v_i and node v_j, else $e_{ij} = 0$.

Definition 3 (OD Flows). OD flows is a set of triplets $F = \left\{ \left(v_i, v_j, f_{ij} \right) \mid 1 \le i, j \le N \right\}$ where flow f_{ij} is the number of transitions (i.e., flow volume) from the origin v_i to the destination v_j. Note that the flow volume is asymmetrical, i.e., $f_{ij} \neq f_{ji}$.

Definition 4 (Node Indicators). Node indicators is a set of attributes $A = \{a_1, \ldots, a_M\}$ for each node (subregion). The node indicators A incorporates multiple kinds of information from socio-economic data and geographic data, which characterizes the profile of the subregion and serves as the initial node representation $H = \{h_1, \ldots, h_M\}$ of the model. The total number of attributes in node indicators is $|A| = M$.

Problem Statement. Given the relation graph $G = (V, E)$ and node indicators $A = \{a_1, \ldots, a_N\}$ over the region $\mathscr{R}$, this study aims to develop a model to predict the OD flows f_{ij} from the origin v_i to the destination v_j. A set of historical OD flows F is available for model training.

3. Methodology

The model consists of two components: a heterogeneous mobility pattern learner and a flow predictor. The overall architecture of the model is shown in Fig. 2. The heterogeneous mobility pattern learner is leveraged to generate node embeddings by modeling diverse mobility patterns. The specific steps are as follows: First, a relation graph is built using historical mobility data, whereby weak connections are eliminated via the concept of modularity. Meanwhile, the community detection algorithm Louvain is performed on the initial relation graph to assign a community label to each node. The initial node representation is composed of community indicators, socio-economic indicators, and geographic indicators. Subsequently, the graph attention network (GAT) then embeds the node representations to learn the mechanism of the mobility patterns through model training. During the training process, a bilinear function is utilized to calculate OD flows, and the node representations are iteratively updated based on a weighted Huber loss function with a distribution smoothing paradigm. The node embeddings are ultimately obtained. The flow predictor employs node embeddings, as well as spatial adjacency and community relations between nodes, into a regression model called Light Gradient Boosting Machine (LGBM) for predicting OD flows. The proposed method is named HMCG-LGBM, where HMCG represents the **H**eterogeneous **M**obility pattern learner via **C**ommunity detection and **G**AT, and LGBM represents the flow predictor.

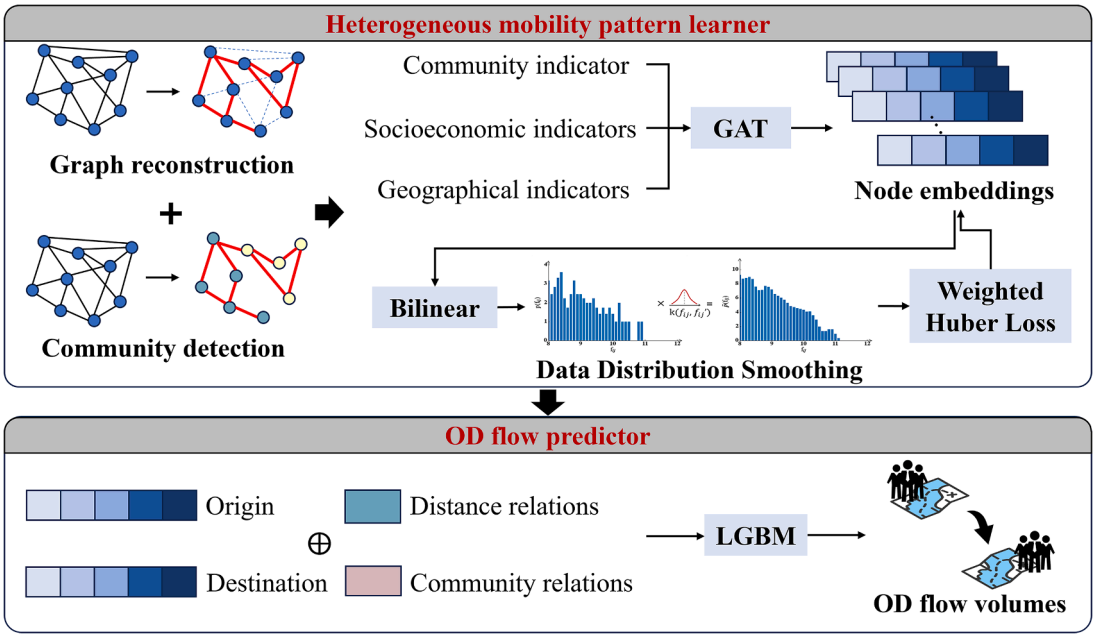


Figure 2. The overall architecture of HMCG-LGBM.

3.1. Graph reconstruction for node relation augmentation

Effectively generating graph structure is essential for the graph neural network to fully understand intricate relationships among nodes, such as spatial, attribute, and semantic relations. Historical flow-based methods are among the most popular used for constructing graph structures, due to their capacity to thoroughly record the real interactions between nodes and incorporate various types of relations (Cai et al., 2022). The approach primarily relies on the presence of historical flow data between the nodes. However, the inherent heterogeneity of OD flows indicates that most nodes possess near-zero historical flow volumes, resulting in the presence of weak connections. The weak connections contribute to the complexity and size of the relation graph. This not only exerts a negative impact on the representation of node relations but also leads to over-smoothing problems during model training.

To address this problem, the most straightforward solution to filter low flows is a threshold-based filter. However, relying solely on the absolute connection strength to eliminate low flows could damage the inherent spatial organization of the relation graph, especially in heterogeneous flow patterns where flow volumes can differ widely across subgraphs. This study develops a novel graph reconstruction strategy that aims to simplify the graph structure and augment its representation by examining the connection strengths. The core idea of this strategy is based on the principle of modularity, which is typically used to evaluate the closeness of connections between nodes in the field of complex network science (Newman & Girvan, 2004). The modularity metric measures node closeness by comparing the actual flow volume with the flow volume in a randomized network. This approach not only helps filter out weaker connections but also enhances the representation of the inherent spatial organization by identifying non-random connection patterns. Specifically, the strategy first incorporates the historical OD flow set F into the initial relation graph G, then calculates the difference q_{ij} between the actual and expected flow volume from node v_i to node v_j. The connection is removed when the actual flow volume is lower than the expected flow volume. q_{ij} is calculated as follows:

$$q_{ij} = f_{ij} + f_{ji} - \frac{f_i \times f_j}{2 \times \sum_{i,j} f_{ij}} \tag{1}$$

$$f_i = \sum_{j \le N} \left(f_{ij} + f_{ji} \right) \tag{2}$$

$$\widetilde{e}_{ij} = \begin{cases} 1, & q_{ij} < 0 \\ 0, & q_{ij} \ge 0 \end{cases} \tag{3}$$

where $\sum_{i,j} f_{ij}$ is the sum of flows of the graph, $\frac{f_i \times f_j}{2 \times \sum_{i,j} f_{ij}}$ is the expected flow volume between node v_i and v_j, f_i is the sum of flows originating and arriving at node v_i. The reconstructed graph is denoted as $\widetilde{G} = (V, \widetilde{E})$, where $\widetilde{E} = \{\widetilde{e}_{ij} \mid 1 \le i, j \le N\}$ represents a set of edges in the reconstructed graph $\widetilde{G}$. and $\widetilde{e}_{ij}$ represents the preserved connection that signifies strong relations between certain nodes. The reconstructed graph not only simplifies the structural complexity, but also augments the representation of node relations. On this basis, the graph neural network could better incorporate the node relation features and generate higher-quality node representations during model training.

3.2. Heterogeneous spatial organization encoding

Aggregation is one of the most common forms of heterogeneous mobility patterns in the real world, where graph nodes are organized into distinct groups, called communities or subgraphs (Fortunato and Hric, 2016; Shu et al., 2021). In the field of complex network science, community structures can be discovered by applying community detection algorithms (Fortunato & Hric, 2016). All nodes in the same community develop a homogeneous spatial organization and share the same community label, allowing for the model to encode the heterogeneous spatial organizations. This study leverages the Louvain algorithm, one of the most widely used approaches in the field of urban mobility (Chen et al., 2022; Jia et al., 2021), to detect the community structure. This algorithm is implemented based on modularity (Newman & Girvan, 2004), which is calculated as follows:

$$Q = \frac{1}{2\sum_{i,j} f_{ij}} \sum_{ij} \left[f_{ij} - \frac{S_i S_j}{2\sum_{i,j} f_{ij}} \right] \sigma(c_i, c_j) \tag{4}$$

where $Q \in [-0.5, 1]$ represents modularity. Q is positive when the flow volume of the community surpasses the expected flow volume obtained from random assignment. As the value of Q increases, the quality of the community improves. c_i is the community label of node v_i. $\sigma(c_i, c_j)$ represents whether node v_i and node v_j belong to the same community, $\sigma(c_i, c_j) = 1$ if they do, and $\sigma(c_i, c_j) = 0$ otherwise. S_i represents the strength of node v_i, i.e., the sum of flows of all connected edges of node v_i. In this study, the community label is regarded as a part of node indicators, enabling the model to emphasize the heterogeneous spatial organization of flows.

Moreover, considering that the formation of the spatial organization is closely related to the socio-economic characteristics and spatial configuration of geographic units, GAT (Veličković et al., 2018) is introduced to encode community labels, socio-economic indicators, and geographic indicators. GAT performs well at consolidating the neighborhood information of nodes. It considers both the node's properties and the relations between nodes when encoding node embeddings. This effectively examines how different geographic units interact and how mobility patterns work. Specifically, the graph attention layer in GAT assigns varying attention weights to a node's neighbors. The input to this layer is a set of initial node representations $H = \{h_1, \ldots, h_M\}$, and the output is a new set of node representations $H' = \{h'_1, \ldots, h'_M\}$. Unlike the original formulation by Veličković et al. (2018), our approach incorporates the edge features $D = \{d_{ij} \mid 1 \le i, j \le N\}$ where d_{ij} represent the great-circle distance between node i and j. By including these edge features directly in the attention mechanism, we aim to better capture geographic information and account for distance decay in mobility patterns (Yang et al., 2019; Schlapfer et al., 2021). Initially, a shared linear transformation, parametrized by two weight matrices W and V, is applied to each node. The attention coefficient r_{ij} for each edge is then defined as follows:

$$r_{ij} = LeakyReLU(\Theta[Wh_i \| Vd_{ij} \| Wh_j]) \tag{5}$$

where r_{ij} indicates the importance of node v_j's features to node v_i, $LeakyReLU$ is a nonlinear function (with negative input slope set to 0.2), Θ is a trainable parameter transforming the concatenated features into a scalar, $\|$ denotes concatenation. W and V are the corresponding input linear transformation's weight matrices. The attention coefficients r_{ij} are then normalized across all neighbors of node v_i using a softmax function, resulting in the attention weight α_{ij}:

$$\alpha_{ij} = softmax_j(r_{ij}) = \frac{\exp(r_{ij})}{\sum_{k \in N(i)} \exp(r_{ik})} \tag{6}$$

where α_{ij} is the attention weight of v_j to v_i. $N(i)$ represents the set of neighbors of node v_i. Note that $\alpha^l_{ij} \neq \alpha^l_{ji}$ due to the asymmetry of OD flows. Fully expanded, the attention weight α_{ij} computed by the attention mechanism is given by:

$$\alpha_{ij} = \frac{\exp\left(LeakyReLU(\Theta[Wh_i \| Vd_{ij} \| Wh_j])\right)}{\sum_{k \in N(i)} \exp\left(LeakyReLU(\Theta[Wh_i \| Vd_{ij} \| Wh_k])\right)} \tag{7}$$

Finally, the new representation h'_i of the node v_i in the next layer is calculated as a weighted sum of both its neighbor's representations and

its own representation, followed by a nonlinearity σ (e.g., ReLu):

$$h'_i = \sigma\left(\alpha_{ii} W h_i + \sum_{j \epsilon N(i)} \alpha_{ij} W h_j\right) \tag{8}$$

where W is consistent with the W in the Formula (5), ensuring a consistent linear transformation across the layer.

3.3. Long-tailed effect capturing

This section aims to achieve capturing long-tailed effect during model training. The model performs end-to-end training, where it directly learns mobility patterns from the raw data and targets the optimization objective on the loss function. By doing so, the original information can be thoroughly utilized, and the optimized process avoids manual intervention. Specifically, it allows us to target the loss function for effectively handling long-tailed flow data.

The study (Cai et al., 2022) formulates a weighted loss function, determining the loss weight by the inverse of the associated flow volume's distribution probability. According to this operation, the model can prioritize low-probability flows during the training process. However, using the inverse of the distribution probability as the loss weight may neglect the learning of certain mobility events due to the highly discrete nature of the imbalanced data distribution, especially those with probabilities close to zero, as shown in Fig. 3(a) for the case where $f_{ij} \in [10, 11]$.

Considering that the probability of a certain flow value occurring is similar to the probability of its neighboring flow values occurring (Yang et al., 2021), the Gaussian distribution smoothing paradigm is introduced to transform the discrete probability distribution into a continuous one based on a Gaussian kernel. Specifically, the Gaussian kernel is convolved with the original data density distribution to produce a smoothed data distribution (Fig. 3(b)). In the new distribution, the probability of the flow value is not only related to itself, but also affected by its neighboring flow values.

The new data distribution is defined as a smoothed density distribution with the following equation:

$$\widetilde{p}\left(f_{ij}'\right) \underset{=}{\Delta} \int_Y k\left(f_{ij}, f_{ij}'\right) p\left(f_{ij}\right) d\left(f_{ij}\right) \tag{9}$$

$$k\left(f_{ij}, f_{ij}'\right) = \exp\left(-\frac{\| f_{ij}, f_{ij}' \|^2}{2\varepsilon^2}\right) \tag{10}$$

where $p\left(f_{ij}\right)$ is the distribution probability of f_{ij} in the original dataset, $k\left(f_{ij}, f_{ij}'\right)$ is the Gaussian kernel, which portrays the similarity between the target value f_{ij}' and any other value f_{ij}. $\| f_{ij}, f_{ij}' \|$ denotes the Euclidean distance between f_{ij}' and f_{ij}. As $\| f_{ij}, f_{ij}' \|$ increases, the value of the Gaussian kernel decreases, and the similarity between the flow values decreases. The parameter ε determines the bandwidth of the kernel function, which governs the speed at which the similarity of flow values decreases with increasing distance. $\widetilde{p}\left(f_{ij}'\right)$ is the distribution probability of f_{ij}' in the smoothed density distribution. The inverse $\widetilde{p}\left(f_{ij}'\right)$ is the weight assigned to the associated flow value in the loss calculation. The loss calculation in this method is derived from the Huber function, which is defined with the following formula:

$$\mathscr{L} = \frac{1}{\widetilde{p}\left(f_{ij}\right)} \begin{cases} \frac{1}{2}\left(\widehat{f}_{ij} - f_{ij}\right)^2, if \left|\widehat{f}_{ij} - f_{ij}\right| \le \delta \\ \delta \cdot \left|\widehat{f}_{ij} - f_{ij}\right| - \frac{1}{2}\delta^2, if \left|\widehat{f}_{ij} - f_{ij}\right| > \delta \end{cases} \tag{11}$$

where $\mathscr{L}$ is the loss function, δ is the control parameter determining the threshold at which the loss function switches from mean square error (MSE) to absolute error (MAE). $\mathscr{L}$ is equivalent to MSE when $|\widehat{f}_{ij} - f_{ij}| \le \delta$. Otherwise, $\mathscr{L}$ is equivalent to MAE. The balance between MSE and MAE makes Huber more robust to outliers and mitigates the potential impact of extreme flow values on model training. Furthermore, due to the large variance of data distribution, the weight assigned to different flow volumes may vary substantially. To mitigate this issue, this study takes the square root of the original data when determining the weight of the loss function.

Furthermore, the training process updates parameters using the stochastic gradient descent method (Adam). To guide the model in learning the interactions between origins and destinations, the bilinear function is used to calculate flows, formulated as follows:

$$\widehat{f}_{ij} = (h_i)^T W_b h_j \tag{12}$$

where $\widehat{f}_{ij}$ is the calculated flow volume using bilinear function, W_b is a trainable parameter matrix with dimension N, which is used to model the interaction between the origin and the destination.

3.4. OD flow predictor

The OD flow predictor is utilized to predict the OD flow volume from

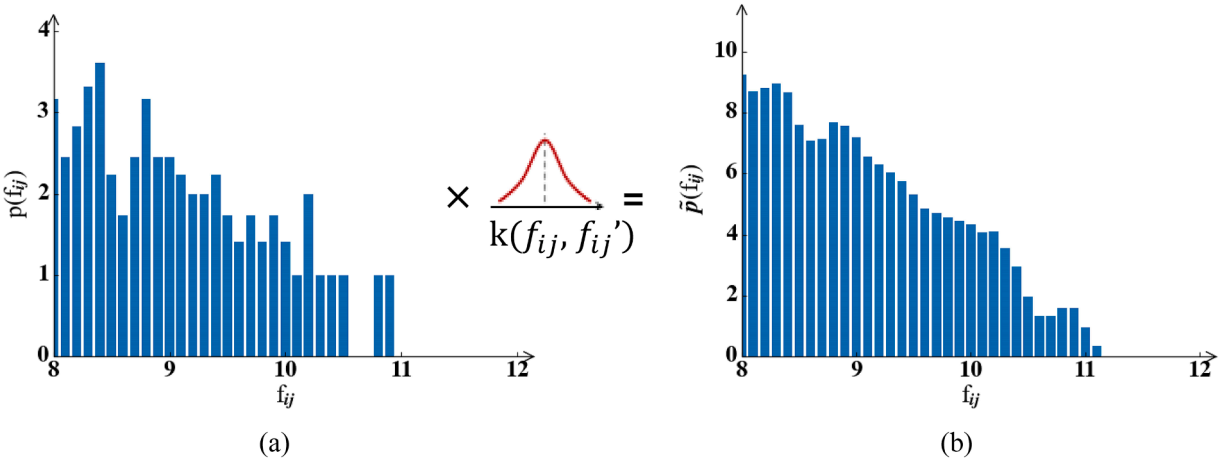


Figure 3. The distribution smoothing paradigm convolves the Gaussian kernel $k\left(f_{ij}, f_{ij}'\right)$ with the initial density distribution (a) to estimate the smoothed density distribution (b).

node v_i to node v_j. The Gradient boosting machine (GBM) (Natekin & Knoll, 2013) is served as the predictor in this study. GBM achieves the regression task by integrating multiple weak learners, which can effectively capture intricate relations and minimize the risk of overfitting. Recent works have proven that GBM (Spadon et al., 2019; Yin et al., 2023; Liu, Gong et al., 2020) is efficient for simulating OD flows as a regression function. This study uses the Light Gradient Boosting Machine (LGBM) as the predictor. Compared to other GBMs, the benefits of LGBM lie in its ability to handle large-scale datasets and high-dimensional features, which is appropriate for practical applications in the real world (Ke et al., 2017). Specifically, the prediction of OD flows is accomplished by LGBM using node embeddings, node distance relations, and node community relations as inputs. Notably, the bilinear function is not used directly as a predictor for the following reasons: (1) Its simple structure, while aiding in training and capturing origin-destination interactions, struggles to capture nonlinear relationships between flows and node features. Additionally, it overlooks other critical relational characteristics, such as spatial relationship and community relationship, which LGBM can incorporate as additional inputs for greater flexibility; (2) Separating loss computation from final prediction allows each step to focus on different objectives. The mobility pattern learner focuses on capturing heterogeneous spatial interaction patterns, while the predictor achieves stable and robust final predictions using enriched features.

4. Experimental results and discussion

In this section, extensive experiments over two real-world datasets are conducted to evaluate the performance of HMCG-LGBM, which are summarized to answer the following research questions (RQs):

RQ1. How is the overall performance of HMCG-LGBM as compared to baselines?

RQ2. How do the hyper-parameters of the model affect performance?

RQ3. Does the design of each part of the model work?

RQ4. How does the model's robustness perform when dealing with imbalanced flow volumes?

4.1. Experimental Settings

4.1.1. Data description

We collected two real-world mobility datasets with different spatial scales and individuals: the intercity truck mobility flow data in China and the intracity commuting flow data in New York City (NYC). The intercity truck mobility flow data is obtained from the trajectory data of heavy trucks in the Global Navigation Satellite System (GNSS) between April 15 and May 15, 2018. The data is collected at regular intervals of

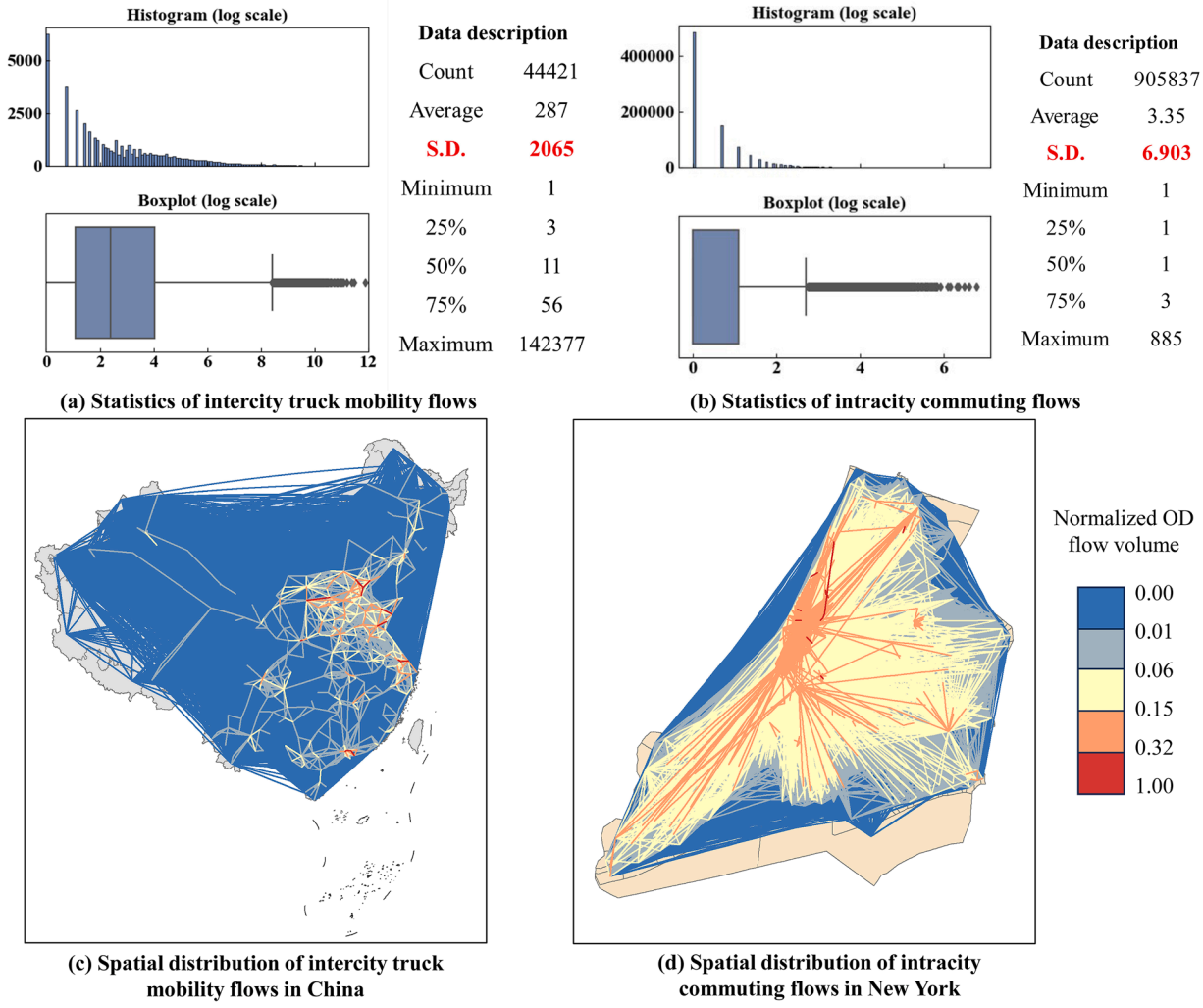


Figure 4. Overview of the two mobility datasets used in our experiments.
Notes: S.D.: Standard deviation.

2–30 seconds, capturing the location of the truck as it travels. We aggregated the trajectory data into city-based OD flow data in our previous work (Zhao, Cheng et al., 2023), which records the cities that trucks visited. The data encompasses a total of 267 cities, 44,421 OD pairs, and 12,761,311 OD trips. The intracity commuting flow data in NYC are provided by the study (Liu, Miranda et al., 2020) from the 2015 Longitudinal Employer-Household Dynamics Origin-Destination Employment Statistics (LODES) project in the US. The data documents the employment location and residence location of workers, as well as commuting flows in NYC on an annual basis. The study by Liu, Miranda et al. (2020) utilized 2010 NYC census tracts as geographical units to aggregate the commuting flow data. The data includes a total of 2,168 geographic units, 905,837 OD pairs, and 3,031,641 OD trips.

From the statistical profiles of the two datasets (Fig. 4(a), Fig. 4(b)), the density distributions of both datasets are long-tailed even after taking the logarithm of the original data. This suggests that real-world mobility patterns exhibit a heterogeneous nature. The range between the minimum and maximum values of intercity truck mobility flows is 142,376, with a standard deviation of 2,065—338.91 times higher than that of New York City commuting flow data (6.903)—indicating much greater heterogeneity. From the spatial distribution depicted in Fig. 4(c) and Fig. 4(d), both datasets exhibit a certain degree of spatial heterogeneity. The aggregation pattern of intercity truck mobility flows is more pronounced. These differences can be attributed to the broader spatial scope of truck activities, which span across cities and even nationally, and the diversity of decision-makers involved, including producers, carriers, and consumers (Zhao et al., 2024). In contrast, commuting flows are typically confined to urban areas, with individuals being the sole decision-makers. We randomly divide the dataset into a training set, a validation set, and a test set (8:1:1).

The socio-economic indicators are obtained from the statistical yearbooks of the study area: the socio-economic indicators for intercity truck mobility flows are obtained from the 2019 China Statistical Yearbook. These indicators are relevant to freight transportation, encompassing the gross national product (GDP), the GDP ratio of the primary, secondary, and tertiary industries, the imported and exported volumes of goods, the number of industrial enterprises, total retail sales of consumer goods, the number of employees for each industry, and the volume of road freight transportation, etc. The socio-economic indicators of intracity commuting flows are obtained from Liu, Miranda et al. (2020), which are based on 2015 land use and infrastructure information for NYC tax parcels, such as the number of different types of buildings, the number of buildings in each built-year interval, and the number of tax lots in different land uses. We standardize these data with a standard deviation of 1 and a mean of 0 to guarantee consistency in the indicator dimensions.

4.1.2. Parameter settings

HMCG-LGBM is implemented with PyTorch and Deep Graph Library. We used two sets of parameters for the two datasets due to the different individuals and spatial scales. For the intercity truck mobility flow dataset, the embeddings dimension is set as 64 and the number of GAT layers is set as 5. For the intracity commuting flows dataset, the embeddings dimension is set as 256 and the number of GAT layers is set as 5. The model is trained using Adam optimizer and relying on the weighted Huber loss, where key parameters include the bandwidth $\varepsilon = 2$ of the Gaussian kernel function, and the control parameter $\delta = 1$ of the Huber function.

4.2. Baselines and metrics

The following 7 models were selected as baselines for comparison with HMCG-LGBM, which can be categorized into three groups:

- Theoretical model: The gravity model (GM) (Anderson, 2011) is one of the most widely used theoretical models and is derived from Newton's law of gravity. GM suggests that the flow between two places is directly proportional to their mass (e.g., truck flow volume) and inversely proportional to their distance. The radiation model (RM) (Simini et al., 2012) treats the choice of a moving object as a process of radiation emission and absorption in physics. According to RM, a moving object tends to choose a destination where the supply exceeds the demand at the origin, while minimizing the interaction cost, such as distance, between the origin and the destination.
- Machine learning model: Random Forest (RF) (Breiman, 2001) is a decision tree-based machine learning model. Due to its robustness to missing values and resistance to overfitting, RF is widely utilized for regression problems. Numerous previous studies (Spadon et al., 2019; Yin et al., 2023) have used RF for OD flow prediction. LGBM (Ke et al., 2017) has gained popularity in the field of machine learning and is specifically employed in this study as an OD flow predictor. The ability to effectively handle large-scale datasets makes LGBM advantageous for the task in this study. The input features for the machine learning models include the attributes of the origin and destination.
- Deep learning model: The Gravity neural network (GNN) (Mozolin et al., 2000), a type of multilayer perceptron, is one of the earliest deep learning models for predicting mobility patterns. Referring to the principle of gravity model, the mass of the origin and the destination, as well as the distance between the two places are embedded in the input layer. The flow volume is predicted in the output layer. Spatial interaction graph convolutional network (SI-GCN) (Yao et al., 2021) combines graph convolution with mapping function and embeds the geographic units in a local spatial network to model OD flows from the perspective of network learning. Geo-contextual Multitask Embedding Learner (GMEL) (Liu, Miranda et al., 2020) utilizes a geo-adjacency network that contains geographic contextual information and models OD flows based on the graph attention mechanism and multitasking framework, with the GBRT model as a predictor. The HMCG model is a variant of the HMCG-LGBM, where the bilinear function (embedded in the heterogeneous mobility pattern learner) is used as the final predictor instead of LGBM. The inputs for the deep learning model include node attributes $A = \{a_1, \ldots, a_M\}$ and the initial relation graph $G = (V, E)$.

The selection of parameters for baselines is based on the grid search algorithm. To ensure the stability of the predicted results, the performance of each method is evaluated by taking the average of the results from 10 repeated runs.

The predictive performance is assessed using the following four metrics: Root mean squared error (RMSE) and mean absolute error (MAE) are common error assessment metrics, quantifying how much the predicted values deviate from the true values. RMSE is more sensitive to extreme values, thus amplifying the contribution of extreme values to the overall error. The common part of commuters (CPC) and Spearman correlation coefficient (SCC) assess the correlation between predicted values and true values. CPC is a metric of the similarity between the two flows, which has been widely used in OD flow modeling studies (Simini et al., 2021; Wang, Yao et al., 2023; Yin et al., 2023). SCC quantifies the degree of concordance between predicted values and true values based on changes in their rank.

$$RMSE = \sqrt{\frac{1}{|T|}\sum_{i,j}\left(\hat{f}_{ij} - f_{ij}\right)^2} \tag{13}$$

$$MAE = \frac{1}{|T|}\sum_{i,j}\left|\hat{f}_{ij} - f_{ij}\right| \tag{14}$$

$$CPC = \frac{2\sum_{i,j}\min\left(\hat{f}_{ij}, f_{ij}\right)}{\sum_{i,j} f_{ij} + \sum_{i,j}\hat{f}_{ij}} \tag{15}$$

$$SCC = 1 - \frac{6\sum_{i,j}\left(rg\left(\widehat{f}_{ij}\right) - rg\left(f_{ij}\right)\right)^2}{n(n^2-1)} \quad (16)$$

where $rg\left(f_{ij}\right)$ is the rank of f_{ij} in the whole dataset.

4.3. Overall performance (RQ1)

Table 1 shows the performance of the eight methods on the two datasets. Overall, the RMSE and MAE applied to the truck mobility flow data are much higher than those of the commuting flow data. This is due to the higher heterogeneity of the intercity truck mobility patterns, indicating that the high heterogeneity of mobility patterns has a noticeable impact on OD flow modeling. The RMSE of GM far exceeds that of other models in the intercity truck mobility flow data, whereas it performs well in the intracity commuting flow data. This suggests that GM has limits in datasets with strong heterogeneous mobility patterns. RM, in contrast, has relatively lower RMSE and MAE in the intercity truck mobility flow data but shows the highest errors in the intercity commuting flow data. This suggests that RM is more applicable to large spatial-scale datasets with strong heterogeneous mobility patterns. Machine learning models outperform others, indicating the effectiveness of the node indicators created in this study. It is worth noting that deep learning models underperform in the intercity truck mobility flow data, with much higher RMSE and MAE compared to the machine learning models. One possible explanation is that the initial relation graph consists of numerous weak connections, which poses a challenge for the graph neural network to function effectively. Furthermore, HMCG underperforms in both datasets, with RMSEs 36.1%–71.9% higher than those of HMCG-LGBM. The difference could result from the HMCG's reliance solely on the bilinear predictor, which struggles to capture nonlinear flow-node relationships and lacks the ability to incorporate relational features. This result suggests that separating the pattern learning and prediction stages, as in HMCG-LGBM, enhances the ability to capture complex spatial interactions and improves model accuracy.

The HMCG-LGBM demonstrates superior performance across all four metrics. In terms of the deviation of the predicted results from the true values, RMSE and MAE outperform baselines by 33.3% and 22.2% in the intercity truck mobility flow data, and by 11.0% and 14.1% in the intracity commuting flow data. In terms of the correlation between the predicted results and the true values, CPC and SCC outperform baselines by 3.1% and 0.4% in the intercity truck mobility flow data, and by 4.7% and 8.2% in the intracity commuting flow data. The results indicate that considering the heterogeneous mobility pattern can effectively improve the predictive performance. The subsequent experimental analysis is conducted using intercity truck mobility flow data due to its apparent heterogeneity.

Table 1
Overall performance compared with baselines

Dataset	Model	RMSE	MAE	CPC	SCC
Intercity truck mobility flow dataset	GM	10522.156	273.196	0.435	0.825*
	RM	718.93	121.190	0.600	0.630*
	RF	708.484	98.902	0.748	0.883*
	LGBM	665.965	81.465	0.780	0.900*
	GNN	825.408	148.396	0.633	0.424*
	SI-GCN	1004.835	229.296	0.420	0.435*
	GMEL	788.600	140.966	0.632	0.689*
	HMCG	763.552	103.961	0.610	0.612*
	HMCG-LGBM	**444.040**	**76.896**	**0.804**	**0.904***
Intracity commuting flow dataset	GM	5.752	1.894	0.663	0.597*
	RM	12.673	3.535	0.183	0.461*
	RF	5.287	2.052	0.697	0.459*
	LGBM	5.005	1.974	0.706	0.510*
	GNN	5.136	1.851	0.724	0.533*
	SI-GCN	4.954	2.069	0.708	0.566*
	GMEL	4.887	1.747	0.741	0.607*
	HMCG	5.917	1.889	0.650	0.587*
	HMCG-LGBM	**4.349**	**1.500**	**0.776**	**0.657***

* p-value < 0.001.

4.4. Hyper-parameter sensitivity analysis (RQ2)

For the hyper-parameter sensitivity analysis, we selected four main hyper-parameters, including the main parameters of GAT (embedding size, GAT layers) and Huber loss (control parameter δ of Huber function, bandwidth parameter ε of Gaussian smoothing kernel). Given that the proportion of predicted data may be a key factor influencing the results, this section also examines the effect of training set size on model performance. The impact of these hyper-parameters on the predictive performance is demonstrated in Fig. 5. Embedding size represents the depth of the node representation by vectors. A relatively low embedding size is insufficient for accurately representing node features, thereby resulting in an initial rise in model performance as the embedding size increases. An excessively large embedding size is prone to overfitting. The results show that the model performs optimally with the embedding size of 64 (Fig. 5 (a)). The number of GAT layers represents the depth of the node neighbors considered by the model. The best performance is achieved when the number of layers is 5 (Fig. 5 (b)). Insufficient layers hinder the full consideration of the node relations. The excessive number of layers leads to the consideration of an excessive number of neighbors, which in turn diminishes the node relations and increases the likelihood of over-smoothing issues. Therefore, the model's performance is reduced after the number of GAT layers exceeded 5. The bandwidth parameter ε of the Gaussian smoothing kernel represents the smoothing degree of the data distribution. An excessive smoothing degree leads to the convergence of the probability distribution of the flow volumes, which eliminates the distinctiveness of the flow volumes. Conversely, insufficient smoothing makes it challenging to address the issue of discrete distributions. In this study, the model performs optimally at $\varepsilon = 2$. The control parameter δ of Huber function determines the balance between MSE and MAE. A smaller δ indicates that Huber loss is closer to MSE, and a larger δ indicates that the loss is closer to MAE, with the former being more sensitive to outliers compared to the latter. The results show that the method performs best when $\delta = 1$ (Fig. 5 (d)). The training set size influences the completeness of the relation graph and the effectiveness of model training. To simulate different levels of data sparsity, we evaluated model performance using training set sizes ranging from 10% to 90%, while adjusting the portion of data to be imputed (Fig. 5 (e)). As expected, larger training sets yield more accurate predictions, as more observed flows enable the learner to model node characteristics and interactions more accurately. Notably, model performance stabilizes once the training set proportion exceeds 70%. For the final evaluation, we adopted an 8:1:1 split. The original intercity truck mobility data includes 44,421 OD pairs, and using the 10% (4,442 samples) for testing ensures a sufficiently large and representative sample to evaluate model performance, while still providing an adequate amount of data for training the model on heterogeneous mobility patterns.

4.5. Ablation Study (RQ3)

We designed five variants to validate the contribution of each part of the model (Fig. 6). Ours means the method proposed in this study (HMCG-LGBM). Ours-iniG builds the framework based on the initial relation graph instead of the reconstructed graph. Ours-noCom removes community labels during node encoding, which means eliminating the spatial organization learning module. Ours-noDS deletes the distribution smoothing module in the training process, and Ours-noW deletes the weight in the loss function, both of which are key steps to capture the long-tailed effect of flow distribution. To further validate the significance of long-tailed effect capturing, we design the variant Ours-noDSW, which eliminates both distribution smoothing and loss weights, i.e., the

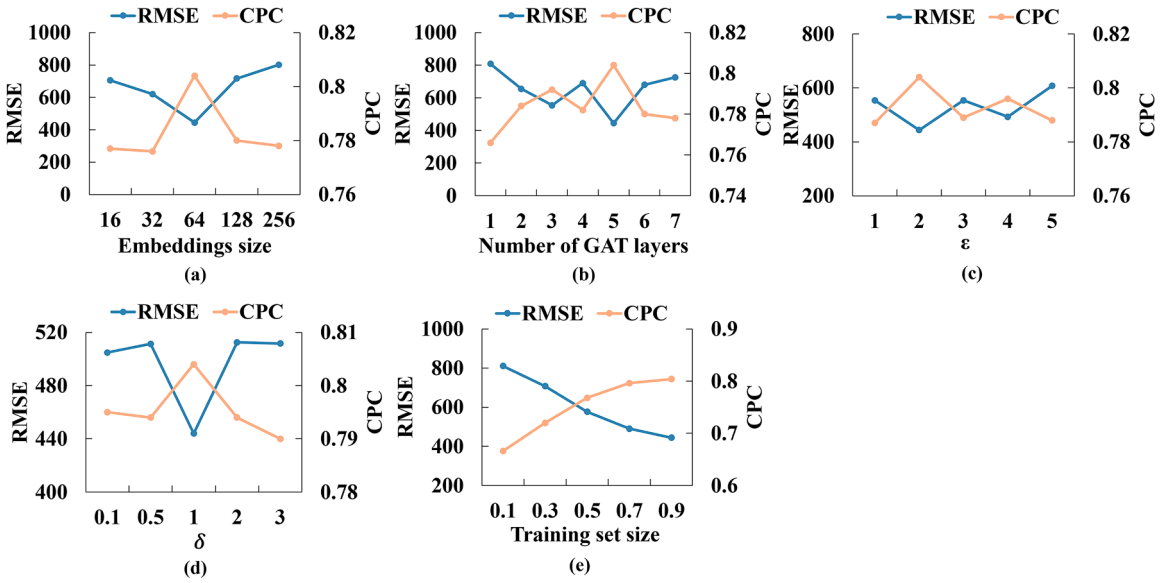


Figure 5. The results of the hyper-parameter sensitivity analysis.

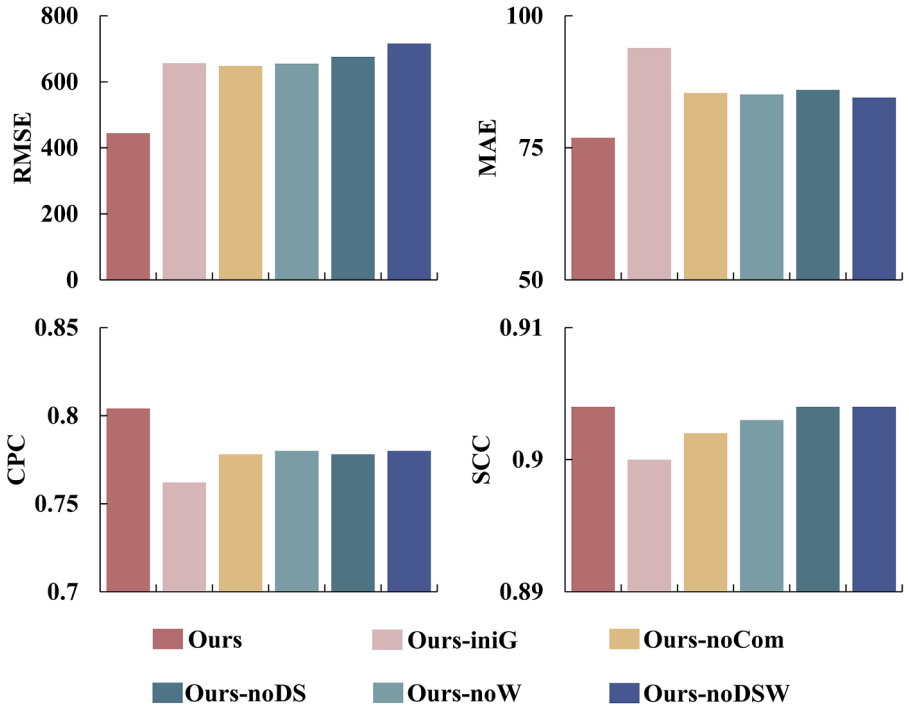


Figure 6. Ablation study of the proposed HMCG-LGBM.

Huber function is directly used as the training loss.

The results (Fig. 6) show that the absence of any part negatively affects the performance of the model. The RMSE of all variants is around 50.9% higher than that of Ours and the CPC of all variants is around 3.5% lower than Ours. These results verified the effectiveness of each part of the model. The graph reconstruction part has the most obvious impact, with Ours-iniG increasing the RMSE and MAE by 47.8% and 22.2%, respectively. The effectiveness of the distribution smoothing module shows that the distribution probability of a flow volume is not only related to itself but also to its neighboring flow volumes. This discovery offers an appropriate solution to the issue of discrete flow distribution. Furthermore, the improvement in model performance by the combination of distribution smoothing and loss weighting is noteworthy, as Ours-noDSW has the highest RMSE among all variants, with an increase of 61.2% compared to Ours.

4.6. *Robustness test for imbalanced flow distribution (RQ4)*

To verify the robustness of HMCG-LGBM in handling imbalanced flow distribution, we examined the prediction errors of HMCG-LGBM for various flow volumes and compared them with four optimal baselines (RM, LGBM, RF, GMEL) (Fig. 7). Overall, the prediction errors of all models show an upward trend as the flow volume increase. The reason is that the data follows a long-tailed distribution, as the flow volume increases, the distribution probability of the flow volume decreases. Consequently, it becomes more challenging to effectively modeling OD

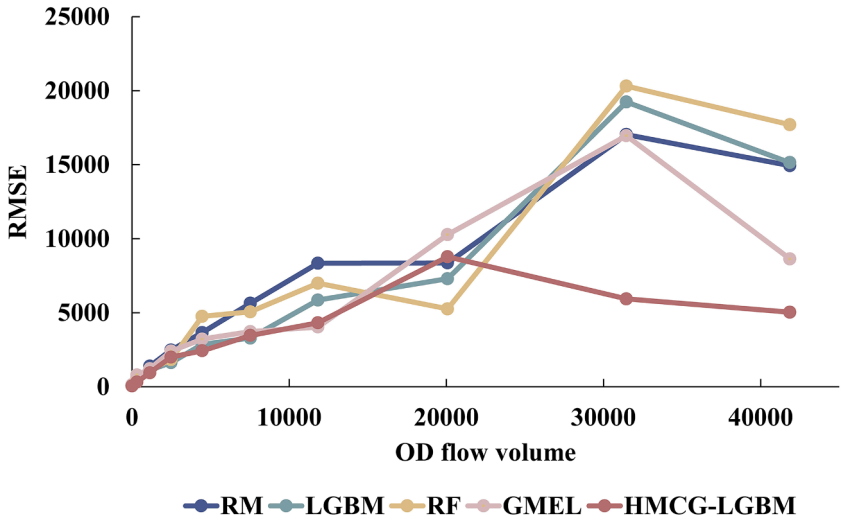


Figure 7. Prediction error for various OD flow volumes.

flows with high volume. The upward trend of RF and LGBM is the most obvious, with RMSEs of both being lower than those of the other models when the flow volume is less than 20,000 and showing a rapid upward trend after the flow volume higher than 20,000. The prediction errors (RMSEs) of RF and LGBM for the flow volume of 30,000 is 163.7% and 286.1% higher than those of flow volume of 20,000, respectively. This result indicates that the machine learning models prioritize OD flows with high distribution probability and disregard OD flows with low distribution probability, leading to diminished reliability of the prediction outcomes. The prediction error for RM and GMEL for the flow volume of 30,000 is 104.1% and 65.0% higher than for a flow volume of 20,000, respectively. In contrast, the trend of prediction error for HMCG-LGBM is significantly smoother. The RMSE of HMCG-LGBM for the flow volume of 30,000 is 32.4% lower than that for the flow volume of 20,000. The above results reveal that HMCG-LGBM can maintain a relatively consistent predictive performance, even though the experimental data distribution is highly imbalanced. This experiment validates the robustness of the proposed method in handling imbalanced flow distributions.

5. Conclusions and future work

To account for the inherent heterogeneous characteristics of real-world mobility patterns, this study proposes a novel OD flow prediction framework with graph-based learning strategies that explicitly address the impact of heterogeneous mobility patterns on the predicting process by considering the complex geographic unit relations, diverse spatial organizational structures, and long-tailed effect of flow distributions. Specifically, our proposed method HMCG-LGBM involves three key components: 1) a modularity-based graph reconstruction strategy for geographic unit relation augmentation by removing weak connections; 2) a spatial structure encoding module that incorporates community detection and graph attention mechanisms for emphasizing variations in the spatial organizations of the OD flows; and 3) a weighted loss with the distribution smoothing paradigm for training low-probability mobility events. Extensive experiments are conducted based on the intercity truck mobility flow data in China and the intracity commuting data in New York City. The results demonstrate that learning heterogeneous patterns of OD flows can effectively enhance the predictive performance of the model. The proposed method reduces the RMSE and MAE by 11.1%–33.3% and 14.1%–22.2%, respectively, compared to existing methods. The proposed method also demonstrates robust capabilities across various flow volumes when handling imbalanced flow data. In this study, the accuracy and reliability of the prediction results are enhanced by considering the heterogeneous mobility patterns of OD flows. This endows the model applicable to complex mobility scenarios. As a result, the proposed method offers insights on spatial interaction predictive modeling within sustainable urban systems, and demonstrates practical advantages in applications such as optimizing urban transportation networks and informing policy decisions based on mobility big data.

However, there are some limitations in this study: (1) While the method aims to ultimately predict OD flows by using geographic unit attributes and inter-unit relations without relying on historical flow data at detailed time intervals, a certain amount of historical data is still necessary for model training. This requirement poses a challenge for regions with limited data, particularly in capturing geographic unit relationships and modeling the spatial organization of flows. This limitation might be mitigated by using recent generative AI approaches. (2) Due to the black-box nature of the deep learning model, the knowledge obtained by the method is implicit. It is difficult to develop a full diagnosis on the mechanisms of mobility patterns in specific regions. Future research could focus on OD flow prediction in data-poor regions and the improvement of method interpretability.

CRediT authorship contribution statement

Yibo Zhao: Writing – original draft, Visualization, Formal analysis, Conceptualization. **Shifen Cheng:** Writing – original draft, Methodology, Conceptualization. **Song Gao:** Writing – review & editing. **Peixiao Wang:** Writing – original draft. **Feng Lu:** Writing – original draft, Supervision, Project administration, Funding acquisition.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

This work was supported by the the Strategic Priority Research Program of the Chinese Academy of Sciences [grant number XDB0740100-02], National Natural Science Foundation of China [grant numbers 42371469, 42101423, 42401524], and the Innovation Project of LREIS [grant numbers KPI003, YPI002]. The authors would also like to thank the anonymous referees for their helpful comments and suggestions.

Data availability

Data will be made available on request.

References

Anderson, J. E. (2011). The Gravity Model. *Annual Review of Economics, 3*(1), 133–160.

Barbosa, H., Barthelemy, M., Ghoshal, G., James, C. R., Lenormand, M., Louail, T., & … Tomasini, M. (2018). Human mobility: Models and applications. *Physics Reports, 734*, 1–74.

Barthelemy, M. (2011). Spatial networks. *Physics Reports, 499*(1-3), 1–101

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

Cai, M., Pang, Y., & Sekimoto, Y. (2022). Spatial Attention Based Grid Representation Learning For Predicting Origin–Destination Flow. In *2022 IEEE International Conference on Big Data* (pp. 485–494).

Chen, W., Chen, X., Cheng, L., Liu, X., & Chen, J. (2022). Delineating borders of urban activity zones with free-floating bike sharing spatial interaction network. *Journal of Transport Geography, 104*, Article 103442.

Davidich, N., Galkin, A., Iwan, S., Kijewska, K., Chumachenko, I., & Davidich, Y. (2021). *Monitoring of urban freight flows distribution considering the human factor, 75*. Sustainable Cities and Society, Article 103168.

Fortunato, S., & Hric, D. (2016). Community detection in networks: A user guide. *Physics Reports, 659*, 1–44.

Friedman, J. H. (2001). Greedy function approximation: a gradient boosting machine. *Annals of statistics*, 1189–1232.

Jia, J. S. S., Lu, X., Yuan, Y., Xu, G., Jia, J. M., & Christakis, N. A. (2020). Population flow drives spatio-temporal distribution of COVID-19 in China. *Nature, 582*(7812), 389-+.

Jia, T., Luo, X., & Li, X. (2021). Delineating a hierarchical organization of ranked urban clusters using a spatial interaction network. *Computers Environment and Urban Systems, 87*, Article 101617.

Jiang, X., Zhuang, D., Zhang, X., Chen, H., Luo, J., & Gao, X. (2023). Uncertainty Quantification via Spatial-Temporal Tweedie Model for Zero-inflated and Long-tail Travel Demand Prediction. In *Proceedings of the 32nd ACM International Conference on Information and Knowledge Management* (pp. 3983–3987).

Ke, G., Meng, Q., Finley, T., Wang, T., Chen, W., Ma, W., … Liu, T.-Y. (2017). Lightgbm: A highly efficient gradient boosting decision tree. *Advances in Neural Information Processing Systems, 30*.

Lenormand, M., Bassolas, A., & Ramasco, J. J. (2016). Systematic comparison of trip distribution laws and models. *Journal of Transport Geography, 51*, 158–169.

Liu, K., Jin, X., Cheng, S., Gao, S., Yin, L., & Lu, F. (2024). Act2Loc: a synthetic trajectory generation method by combining machine learning and mechanistic models. *International Journal of Geographical Information Science, 38*(3), 407–431.

Liu, L., Zhu, Y., Li, G., Wu, Z., Bai, L., & Lin, L. (2022). Online metro origin-destination prediction via heterogeneous information aggregation. *IEEE Transactions on Pattern Analysis and Machine Intelligence, 45*(3), 3574–3589.

Liu, Y., Yao, X., Gong, Y., Kang, C., Shi, X., Wang, F., & …Zhu, D. (2020). Analytical methods and applications of spatial interactions in the era of big data. *Acta Geographica Sinica, 75*(7), 1523–1538.

Liu, Z., Miranda, F., Xiong, W., Yang, J., Wang, Q., & Silva, C. (2020). Learning Geo-Contextual Embeddings for Commuting Flow Prediction. *AAAI Conference on Artificial Intelligence, 34*(01), 808–816.

Mozolin, M., Thill, J. C., & Usery, E. L. (2000). Trip distribution forecasting with multilayer perceptron neural networks: A critical evaluation. *Transportation Research Part B-Methodological, 34*(1), 53–73.

Natekin, A., & Knoll, A. (2013). Gradient boosting machines, a tutorial. *Frontiers in neurorobotics, 7*, 21.

Newman, M. E., & Girvan, M. (2004). Finding and evaluating community structure in networks. *Physical Review E, 69*(2), Article 026113.

Rao, J., Gao, S., & Zhu, S. (2023). CATS: Conditional Adversarial Trajectory Synthesis for privacy-preserving trajectory data publication using deep learning approaches. *International Journal of Geographical Information Science, 37*(12), 2538–2574.

Rao, W., Wu, Y.-J., Xia, J., Ou, J., & Kluger, R. (2018). Origin-destination pattern estimation based on trajectory reconstruction using automatic license plate recognition data. *Transportation Research Part C-Emerging Technologies, 95*, 29–46.

Ravenstein, E. G. (1885). The laws of migration. *Journal of the Royal Statistical Society, 48*(2), 167–227.

Ren, Y., Ercsey-Ravasz, M., Wang, P., Gonzalez, M. C., & Toroczkai, Z. (2014). Predicting commuter flows in spatial networks using a radiation model based on temporal ranges. *Nature Communications, 5*(9), 5347.

Rong, C., Feng, J., & Ding, J. (2023). GODDAG: Generating Origin-destination Flow for New Cities via Domain Adversarial Training. *IEEE Transactions on Knowledge and Data Engineering, 35*(10), 10048–10057.

Sana, B., Castiglione, J., Cooper, D., & Tischler, D. (2018). Using google's passive data and machine learning for origin-destination demand estimation. *Transportation Research Record, 2672*(46), 73–82.

Schlapfer, M., Dong, L., O'Keeffe, K., Santi, P., Szell, M., Salat, H., & West, G. B. (2021). The universal visitation law of human mobility. *Nature, 593*(7860), 522-+.

Shi, S., Wang, L., Xu, S., & Wang, X. (2022). Prediction of Intra-Urban Human Mobility by Integrating Regional Functions and Trip Intentions. *IEEE Transactions on Knowledge and Data Engineering, 34*(10), 4972–4981.

Shu, H., Pei, T., Song, C., Chen, X., Guo, S., Liu, Y., & …Zhou, C. (2021). L-function of geographical flows. *International Journal of Geographical Information Science, 35*(4), 689–716.

Simini, F., Barlacchi, G., Luca, M., & Pappalardo, L. (2021). A Deep Gravity model for mobility flows generation. *Nature Communications, 12*(1), 6576.

Simini, F., González, M. C., Maritan, A., & Barabási, A.-L. (2012). A universal model for mobility and migration patterns. *Nature, 484*(7392), 96–100.

Spadon, G., de Carvalho, A. C. P. L. F., Rodrigues-Jr, J. F., & Alves, L. G. A (2019). Reconstructing commuters network using machine learning and urban indicators. *Scientific Reports, 9*, 11801. Article.

Stouffer, S. A. (1940). Intervening opportunities: a theory relating mobility and distance. *American sociological review, 5*(6), 845–867.

Veličković, P., Cucurull, G., Casanova, A., Romero, A., Lio, P., & Bengio, Y. (2018). Graph attention networks. In *International Conference on Learning Representations* (pp. 1–12).

Wang, P., Zhang, Y., Hu, T., & Zhang, T. (2023). Urban traffic flow prediction: a dynamic temporal graph network considering missing values. *International Journal of Geographical Information Science, 37*(4), 885–912.

Wang, Y. X., Yao, X., Liu, Y., & Li, X. (2023). Generating population migration flow data from inter-regional relations using graph convolutional network. *International Journal of Applied Earth Observation and Geoinformation, 118*, Article 103238.

Xu, G., Lv, Y., Sun, H., Wu, J., & Yang, Z. (2021). Mobility and evaluation of intercity freight CO2 emissions in an urban agglomeration. *Transportation Research Part D: Transport and Environment, 91*, Article 102674.

Yang, X., Fang, Z., Xu, Y., Yin, L., Li, J., & Lu, S. (2019). Spatial heterogeneity in spatial interaction of human movements-Insights from large-scale mobile positioning data. *Journal of Transport Geography, 78*, 29–40.

Yang, Y., Chen, Y., Lv, Y.-Y., Jia, X.-Y., Lin, X.-J., Yan, X.-Y., & Jia, B. (2024). *Quantifying city freight mobility segregation associated with truck multi-tours behavior*. Sustainable Cities and Society, Article 105699.

Yang, Y. Z., Zha, K. W., Chen, Y. C., Wang, H., & Katabi, D. (2021). Delving into Deep Imbalanced Regression. In *Proceedings of Machine Learning Research. International Conference on Machine Learning (ICML)*. Electr Network.

Yang, Z., Chen, X., Deng, J., Li, T., & Yuan, Q. (2023). Footprints of goods movements: Spatial heterogeneity of heavy-duty truck activities and its influencing factors in the urban context. *Journal of Transport Geography, 113*, Article 103737.

Yao, X., Gao, Y., Zhu, D., Manley, E., Wang, J. E., & Liu, Y. (2021). Spatial Origin-Destination Flow Imputation Using Graph Convolutional Networks. *IEEE Transactions on Intelligent Transportation Systems, 22*(12), 7474–7484.

Yin, G. M., Huang, Z., Bao, Y., Wang, H., Li, L. N., Ma, X. L., & Zhang, Y. (2023). ConvGCN-RF: A hybrid learning model for commuting flow prediction considering geographical semantics and neighborhood effects. *Geoinformatica, 27*(2), 137–157.

You, L., Chen, Q., Qu, H., Zhu, R., Yan, J., Santi, P., & Ratti, C. (2024). FMGCN: Federated Meta Learning-Augmented Graph Convolutional Network for EV Charging Demand Forecasting. *IEEE Internet of Things Journal, 11*(14), 24452–24466.

Zhang, X., & Li, N. (2024). An activity space-based gravity model for intracity human mobility flows. *Sustainable Cities and Society, 101*, Article 105073.

Zhao, P., Hu, H., Zeng, L., Chen, J., & Ye, X. (2023). Revisiting the gravity laws of inter-city mobility in megacity regions. *Science China-Earth Sciences, 66*(2), 271–281.

Zhao, Y., Cheng, S., Liu, K., Zhang, B., & Lu, F. (2024). Intercity freight connections in China under the view of mass truck trajectories. *Cities, 150*, Article 105034.

Zhao, Y., Cheng, S., & Lu, F. (2023). Spatiotemporal interaction pattern of the Beijing agricultural product circulation. *Journal of Geographical Sciences, 33*(5), 1075–1094.

Zhuang, D., Wang, S., Koutsopoulos, H., & Zhao, J. (2022). Uncertainty quantification of sparse travel demand prediction with spatial-temporal graph neural networks. In *Proceedings of the 28th ACM SIGKDD Conference on Knowledge Discovery and Data Mining* (pp. 4639–4647).