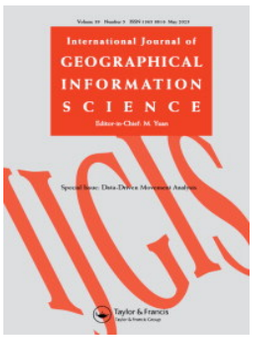

International Journal of Geographical Information Science

ISSN: (Print) (Online) Journal homepage: www.tandfonline.com/journals/tgis20

Capturing spatial heterogeneity of population-level human mobility via a prior-guided graph neural network

Zhipeng Wang, Tong Zhang, Jing Li, Peixiao Wang & Yu'ang Zhu

To cite this article: Zhipeng Wang, Tong Zhang, Jing Li, Peixiao Wang & Yu'ang Zhu (21 Apr 2025): Capturing spatial heterogeneity of population-level human mobility via a prior-guided graph neural network, International Journal of Geographical Information Science, DOI: 10.1080/13658816.2025.2493215

To link to this article: https://doi.org/10.1080/13658816.2025.2493215

INTERNATIONAL JOURNAL OF GEOGRAPHICAL INFORMATION SCIENCE
https://doi.org/10.1080/13658816.2025.2493215

Taylor & Francis
Taylor & Francis Group

RESEARCH ARTICLE

Capturing spatial heterogeneity of population-level human mobility via a prior-guided graph neural network

Zhipeng Wang[a], Tong Zhang[a], Jing Li[b], Peixiao Wang[c] and Yu'ang Zhu[a]

[a]State Key Laboratory of Information Engineering in Surveying, Mapping and Remote Sensing, Wuhan University, Wuhan, China; [b]Department of Geography and the Environment, University of Denver, Denver, CO, USA; [c]State Key Laboratory of Resources and Environmental Information System, Institute of Geographic Sciences and Natural Resources Research, Chinese Academy of Sciences, Beijing, China

ABSTRACT

Modeling population-level human mobility has been attracting multidisciplinary research attention due to its profound implications for sustainable urban development. However, previous studies have often neglected the explicit consideration of spatial heterogeneity of travel demand, which limits their abilities to accurately estimate mobility flows. In this study, we introduce a prior-guided, data-driven human mobility model that integrates the position of origins and destinations, spatial travel patterns, and physical models as priors to capture spatial heterogeneity of human mobility. Specifically, we introduce the concept of 'relative attractiveness' to emulate the underlying driving force for the formation of spatial heterogeneity in human mobility. To learn the embeddings of 'relative attractiveness', we propose a suite of methods that integrate prior knowledge and graph neural networks, mainly including a relative position encoding module to encode the position of different origin-destination (OD) pairs relative to the entire geographical space and a message-passing method inspired by the classical physical models to simulate the mechanisms of mobility flow generation. Finally, a gradient boosting regression tree is trained to generate the mobility flow based on the learned embeddings. Extensive experiments on two real-world datasets have showed our model outperforms state-of-the-art data-driven mobility models in terms of accuracy and generalization.

ARTICLE HISTORY
Received 6 November 2024
Accepted 9 April 2025

KEYWORDS
Human mobility modeling; graph neural network; position encoding; spatial heterogeneity; prior-guided

1. Introduction

For more than a century, population-level human mobility modeling has long been attracting multi-disciplinary efforts in physics, geography, transportation, and computer science, aiming at building universal models for travel flow generation and prediction (Barbosa *et al.* 2018). The travel flow generation problem focuses on generating the mobility flows using only the information of node attributes, such as population and land use information, as illustrated in Figure 1.

CONTACT Tong Zhong zhangt@whu.edu.cn

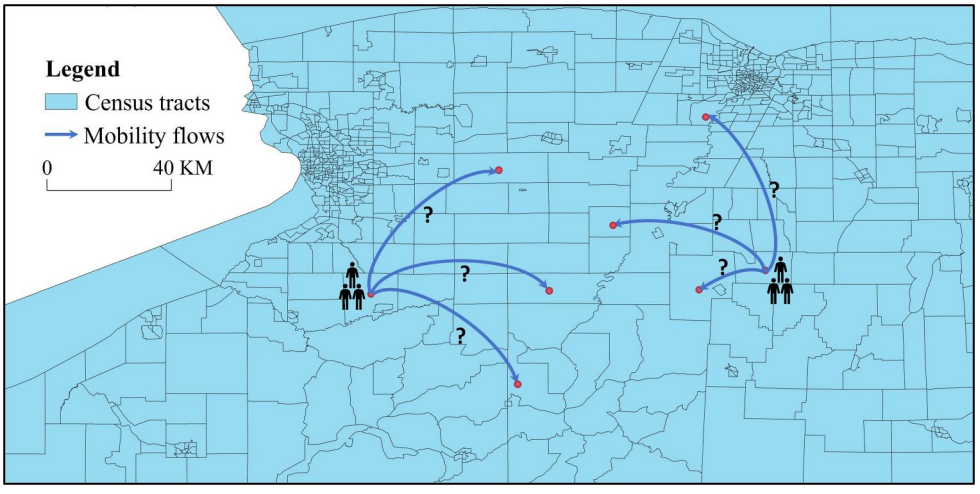

Figure 1. Using the Western Region of New York State as a geographic background to explain the travel flow generation problem.

In order to describe human mobility dynamics, several well-known physical mobility models have been developed, including the traditional gravity model (Zipf 1946, Wilson 1969), the intervening opportunities model (Stouffer 1940), and the radiation model (Simini *et al*. 2012). With parsimonious and intuitive forms, physical mobility models have been widely used to underpin many real-world applications, such as transportation planning (de Dios Ortúzar 2011), economic geography (Duenas *et al*. 2013), business geography (Reilly 1931). However, these physical models oversimplify the real mechanisms of how travel flows are generated, and therefore often fail to recover empirically observed travel demand patterns (Masucci *et al*. 2013, Barbosa *et al*. 2018).

Recently, deep learning-driven human mobility models have gained momentum due to their ability to generalize collective spatio-temporal movement patterns, such as population migration characteristics and traffic flow patterns, from massive observational mobility data (Luca *et al*. 2021, Gu *et al*. 2024). Since deep learning-driven models are typically trained on sufficient data samples collected from a specific city, they often perform better than traditional physical models in flow generation and prediction for that city (Yin *et al*. 2023, Shi *et al*. 2024). However, when trained deep learning models are applied to other cities, significant performance degradation would be observed. Transfer learning can help improve generalization by extracting abstract mobility knowledge from a source city and transferring it to target cities (Wang *et al*. 2019). Transfer learning methods rely heavily on the data from the source cities, and the transferred knowledge is typically represented as neural embeddings, which are difficult to understand, validate, and reuse. Given that traditional physical models and deep learning-driven methods have their own strengths and weaknesses, the integration of them may complement each other and become a possible solution towards a universal and rigorous mobility model. This idea has been explored by integrating traditional physical models with nonlinear neural networks to better describe the nonlinear relationship between mobility flows and different geographic features (Simini *et al*. 2021).

This idea coincides with the recently emerging research paradigm of 'physics-informed machine learning (PIML)', which encourages the learned model to produce more physically consistent results, with higher training efficiency and better generalizability than regular machine learning models (Karniadakis *et al*. 2021). However, the existing relevant studies often build their models based on off-the-shelf deep neural networks and have not devoted much effort to developing generalizable network architectures to embed necessary and comprehensive prior knowledge. Therefore, there is still much room for improvement in terms of accuracy and generalizability in developing physics-informed data-driven human mobility models.

Despite considerable efforts in developing physical and data-driven models, the goal of building a universal population-level mobility model remains elusive. The challenge in building such an accurate, universal, and robust mobility model lies in the complexity of human travel behavior, which is influenced by many geographic, demographic, social and economic factors as well as individual preferences (Mwale *et al*. 2022). This complexity results in heterogeneous spatial patterns of aggregated travel demand. For instance, With the roughly the same population distribution and travel distance, two origin-destination pairs usually have different travel flows because they are located in different geographic areas of the city. Intuitively, flow generation between origins and destinations is driven by the features of origins and destinations, their geographical locations, external factors (e.g., weather) as well as the complex interactions between these factors. As these factors and interactions are location dependent, spatial heterogeneity of human mobility is inevitable. This hard-to-quantify spatial heterogeneity may explain the poor performance of traditional physical models (Masucci *et al*. 2013) and the poor generalization performance of deep learning-driven models. We argue that in order to build a universal mobility model that yields satisfactory performance at different spatial scales, across different cities and zoning schemes, the spatial heterogeneity of travel demand should be well accounted for.

Although a few studies have attempted to address the heterogeneity issue in human mobility (Zhou *et al*. 2023, Tang *et al*. 2024), most of them still fall short in modeling the spatial heterogeneity of human mobility from two perspectives: 1) They are still not spatially explicit models. The spatial heterogeneity of travel demand is essentially related to geographic positions of origins and destinations (Fotheringham *et al*. 1996). Nevertheless, there are almost no studies that have explicitly modeled the effects of positions on the performance of human flow generation. 2) prior knowledge, such as classical physical models, geographic positions and travel expenses, is not sufficiently leveraged and incorporated into the current data-driven human mobility models. Therefore, the optimization of these model cannot be correctly constrained by prior knowledge. Trained models usually fail to generalize the spatial heterogeneity of travel demand in space, thereby leading to poor performance when being used for flow generation.

Aiming to address the above two research gaps and better capture the spatial heterogeneity in human mobility, this study introduces a prior-guided data-driven human mobility model based on the popular graph neural network (GNN) architecture (Battaglia *et al*. 2018, Wu *et al*. 2021), which is named Prior-Guided Mobility Flow Generation Network (PG-MFG). Three types of priors are incorporated into the model:

the positions of origins and destinations, average travel expenses, and physical models (the gravity model and radiation model) (Zipf 1946, Simini *et al*. 2012). The main contributions of this study can be summarized as follows:

- We propose a prior-guided, data-driven human mobility model, PG-MFG, designed to capture the spatial heterogeneity of human mobility. To the best of our knowledge, we are the first to explicitly account for spatial heterogeneity in human mobility modeling by combining comprehensive prior knowledge with data-driven GNNs;
- We develop a suite of deep learning methods that effectively integrate comprehensive prior knowledge to correctly constrain the optimization of the proposed models;
- Within the framework of PG-MFG, based on the concept of 'relative attractiveness', we extend the vanilla GNN with a message-passing method to learn the embeddings of 'relative attractiveness', simulating the mechanisms underlying mobility flow generation as reflected in physical models;
- Extensive experiments show that, compared to state-of-the-art baselines, the proposed models require far fewer training samples to achieve state-of-the-art performance. It also excels in performing well when tested on datasets with different mobility patterns and geographical contexts from the datasets they are trained on.

2. Motivation and problem formulation

2.1. Spatial heterogeneity of human mobility

We illustrate spatial heterogeneity of human mobility through an example. As depicted in Figure 2, a city exhibits a relatively severe segregation between employment and housing, with work and entertainment opportunities predominantly clustered in the central business district. As shown in Figure 2(a), three OD pairs, which have similar populations and travel distances, exhibit different travel flows due to their distinct geographical locations within the city. Figure 2(b) illustrates that the average travel expenses for each origin are often heterogeneous as the distributions of travel distances and locations of destinations are considerably diverse. As shown in Figure 2(c), $O1 \rightarrow D1$ and $O2 \rightarrow D2$ are two overlapping OD pairs with reverse directions. Since $D1$ is located in the most objectively attractive location in the city, the attractiveness of $D1$ to $O1$ is greater than that of $D2$ to $O2$, resulting in a significant disparity in the flow of the two OD pairs.

Based on the above analysis, we argue that the key to capture the spatial heterogeneity of human mobility lies in the accurate measurement of the attractiveness of destinations for the origins, which is a comprehensive assessment of the factors influencing human travel demand. Unlike objective attractiveness, which can be expressed in terms of the number of opportunities (e.g., job opportunities) available in a location to satisfy an individual's travel purposes (Simini *et al*. 2012), the attractiveness we refer to describes the relative relationship between origins and destinations. Thus, we define it as 'relative attractiveness'. Drawing from a review of existing research on the factors influencing travel demand (Boyce and Williams 2015, Sharma 2019, Schläpfer

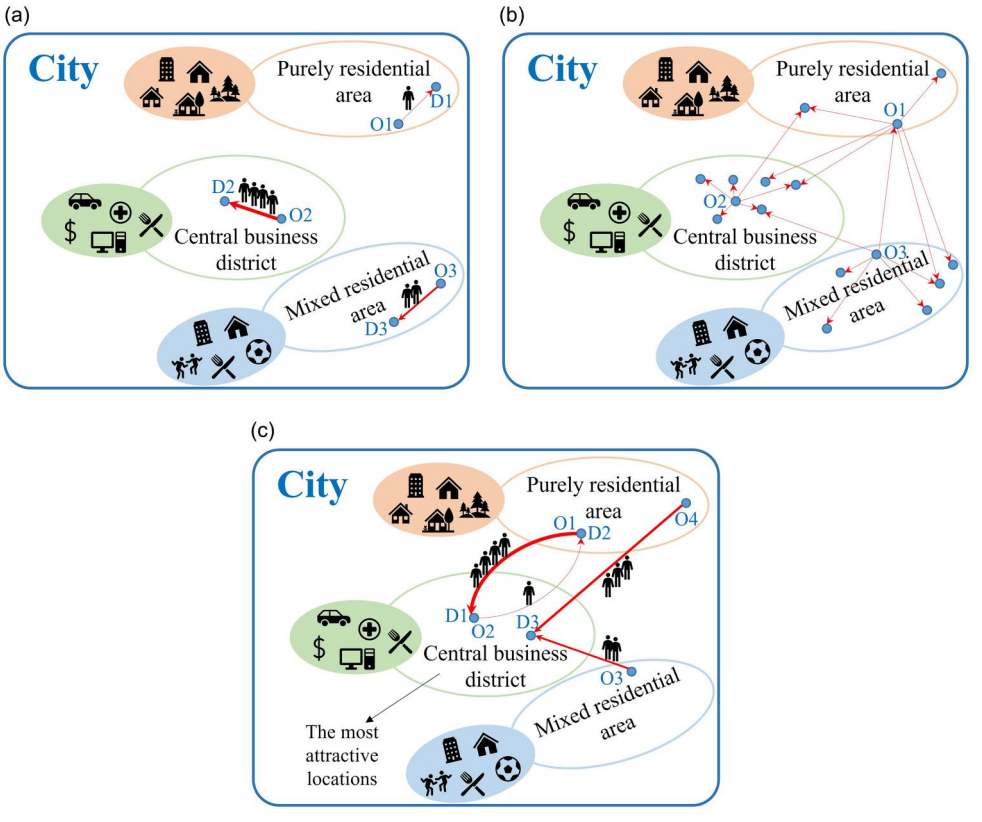

Figure 2. Examples highlighting the spatial heterogeneity of human mobility. (a) The impact of location on spatial heterogeneity of human mobility. (b) Heterogeneous travel expenses due to the travel distance from origins. (c) The impact of attractiveness on spatial heterogeneity of human mobility.

et al. 2021, Mwale *et al*. 2022), and focusing solely on demographic and positional attributes, we identify four types of key factors that influence travel demand and are essential for a comprehensive measurement of 'relative attractiveness': position, average travel expense, opportunities and distance. To reflect these factors, we select four types of prior knowledge and integrate them into a GNN architecture to learn the embeddings of 'relative attractiveness', as illustrated in Figure 3. The specific integration method is detailed in the Methodology Section. Ultimately, the integrated prior GNN model generates embeddings that represent 'relative attractiveness', which are used to generate mobility flows.

2.2. Definitions and problem formulation

Definition 1. Geographical region. A large geographical space is partitioned into non-overlapping geographical regions denote as $\mathcal{R} = \{r_1, r_2, ..., r_n\}$. In this study, we use census tracts as geographic regions, and consider them to be the basic geographic units of generating human mobility.

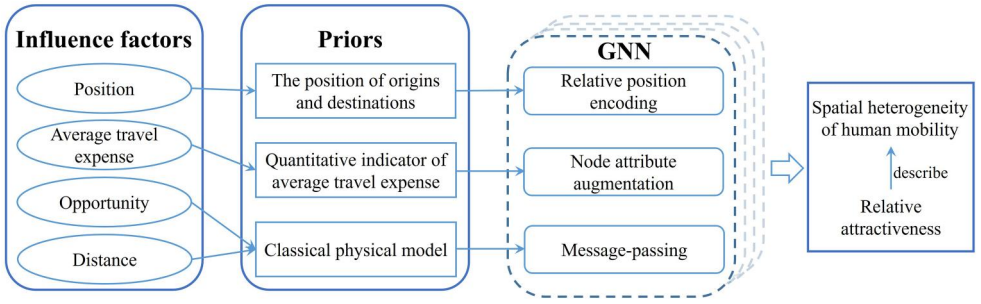

Figure 3. An overall process that describes the spatial heterogeneity of human mobility.

Definition 2. Mobility flow. The mobility flow F_{ij} is the total number of population movements between origin region r_i and destination region r_j. In addition, the out-flow from an origin, denote as O_i, represents the total number of population movements originating from region r_i.

Definition 3. Mobility flow network. The mobility flow network is a weighted directed graph $\boldsymbol{G} = (\boldsymbol{V}, \boldsymbol{E}, \boldsymbol{A})$ where nodes $\boldsymbol{V}$ of the graph represent regions $\mathcal{R}$, The directed links $\boldsymbol{E} = \{e_{ij} | 1 \leq i, j \leq n\}$ describe potential travel demand from origin nodes to destination nodes, $\boldsymbol{A} = \{a_1, a_2, \ldots, a_n\}$ is the set of geographical features (e.g. socio-economic, demographic, point of interest) of geographical regions that serve as the node attributes. In this study, we only use demographic data as the node attributes, posing minimum data requirement of mobility modeling.

Problem

Given $\boldsymbol{G} = (\boldsymbol{V}, \boldsymbol{E}, \boldsymbol{A})$ of a specific geographical space $\mathcal{R}_1$,our goal is to develop a data-driven model to generate mobility flow F_{ij} between all origin nodes and all destination nodes in $\mathcal{R}_1$. The trained model developed in $\mathcal{R}_1$ can be applied to generate mobility flow in another geographical space $\mathcal{R}_2$.

3. Methodology

3.1. A Prior-guided population-level flow generation framework

Figure 4 illustrates the proposed model framework named PG-MFG, which is composed of six main components and converts the task of mobility flow generation between origins and destinations into two sub-tasks: one is to generate the total out-flow from origins, and the other is to allocate the total out-flow of the origin to each of its destinations.

- Initially, in component 1, a weighted directed mobility flow network graph is constructed, with each geographic region as a node, the population of the region as node attribute, and the potential travel demand between two regions as edge.
- For the former sub-task: a multilayer perceptron uses population and distance data to generate the total out-flow of all origins in component 2 (Section 3.3).

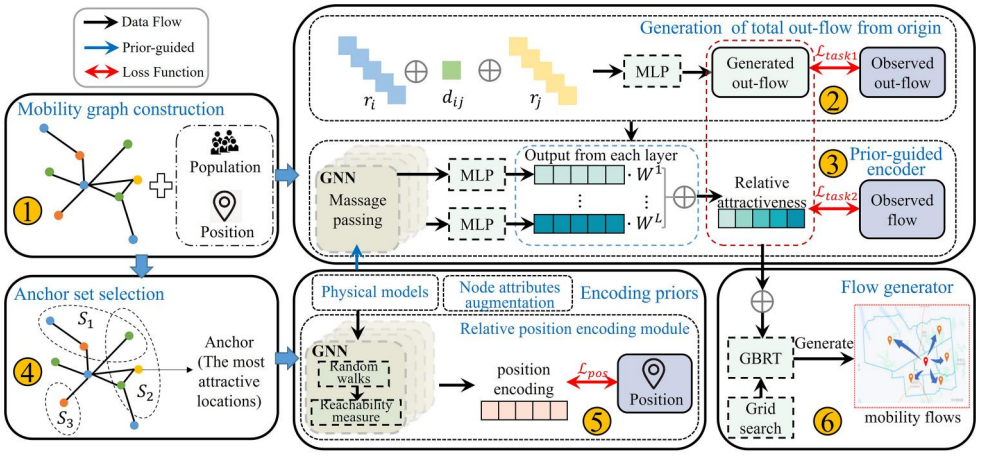

Figure 4. The framework of the proposed prior-guided data-driven human mobility model.

- For the latter sub-task: a prior-guided graph encoder is designed to integrate the physical model priors with a GNN, as well as fuse the encoded priors to learn the embedding of 'relative attractiveness' in component 3 (Section 3.4). The results obtained in component 2 provide an initialization embedding of the origin for the message-passing process in component 3.
- In component 4, in order to encode positional priors, multiple sets of anchors $S = \{S_1, S_2, \ldots, S_k\}$ are selected among the most objectively attractive locations in the study region.
- In component 5, preferential random walks are performed on the mobility flow network to measure the reachability between nodes and anchor sets, thereby encoding the relative positions of origin-destination pairs to the anchor sets (Section 3.2.2). Moreover, node attribute augmentation is used to encode average travel expense priors (Section 3.2.3). To encode physical model priors, a message-passing method that simulates the generation mechanism of flow allocation is designed (Section 3.4).
- Finally, in component 6, a gradient boosting regression tree (GBRT) (Fredman 2001, Prettenhofer and Louppe 2014) is trained to generate mobility flows between all origins and destinations (section 3.5). The final loss function is composed of two sub-task losses and a positional encoding loss (Section 3.6).

3.2. Encoding priors

3.2.1. Physical model priors

The physical model priors used in this study are the singly-constrained gravity model (Zipf 1946, Wilson 1969) and the radiation model (Simini *et al*. 2012), as shown in Equations (1) and (2) respectively.

$$Y_{ij} = O_i P_{ij} = O_i \frac{m_j^{\beta_1} f(D_{ij})}{\sum_k m_k^{\beta_1} f(D_{ik})} \tag{1}$$

$$Y_{ij} = O_i P_{ij} = O_i \frac{n_i n_j}{(n_i + s_{ij})(n_i + n_j + s_{ij})} \tag{2}$$

where O_i means the total out-flow of region r_i, m_j and n_j is the number of population and opportunities in region r_j, respectively. β_1 is a learnable parameter, $f(D_{ij})$ is a distance-dependent decreasing function, s_{ij} is the number of opportunities in a range centered on r_i with radius D_{ij} (Excluding n_i and n_j).

In order to better integrate the radiation model in this study, Eq. (2) is transformed into the following form:

$$Y_{ij} = O_i n_i \left[\frac{1}{n_i + s_{ij}} - \frac{1}{n_i + n_j + s_{ij}} \right] \tag{3}$$

The above two physical models both reflect the allocation idea of the total out-flow flow into destination. We use this flow allocation idea to design the proposed models, ensuring that the optimization of the model can be correctly constrained by physical model priors.

3.2.2. Encoding relative position priors

In this study, the 'relative positions' refers to the positions of OD pairs in relation to the most objectively attractive locations within the study region. The variation in relative positions is a crucial factor contributing to spatial heterogeneity of human mobility.

(1) the architecture of RPE. Figure 5 illustrates an overview of the RPE module's general architecture. The module consists of two layers, each involving a distinct selection of k anchor sets $S = \{S_1, S_2, ..., S_k\}$, which represent the sets of the most objectively attractive locations within the study region (Section 3.2.2-(2)). In each layer, we first compute the comprehensive reachability Me_i^l between node v and each anchor in the anchor set S_i by the comprehensive reachability computation function $F\left(v, u, \mathrm{h}_v^l, \mathrm{h}_u^l\right)$ (Sections 3.2.2-(3) and (4)). To further measure the comprehensive reachability between node v and the entire anchor-set S_i, we employ a comprehensive reachability

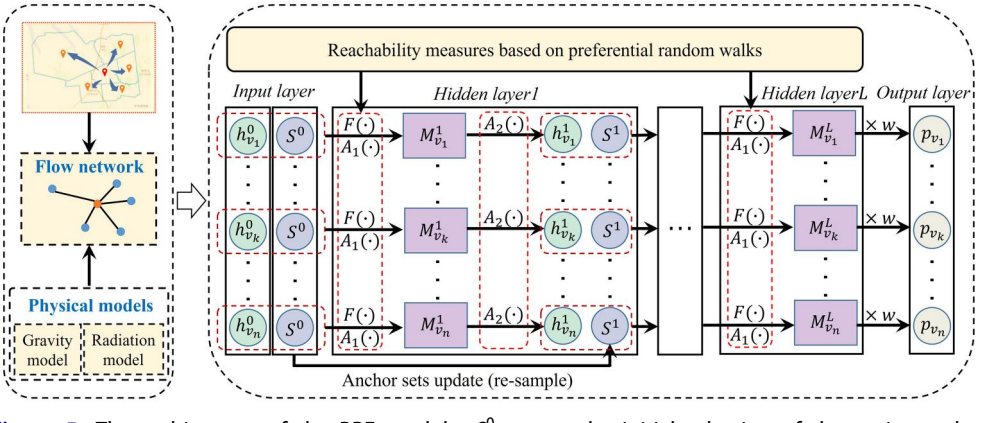

Figure 5. The architecture of the RPE module. S^0 means the initial selection of the entire anchor set $S = \{S_i\}$ at the first layer of the module.

aggregation function A_1 to generate message $M_v[i]^l$. The message $M_v[i]^l$ in $M_v{}^l$ means the comprehensive reachability computed by node v with respect to the i-th anchor set S_i in the entire anchor set S. Subsequently, to measure the overall comprehensive reachability between node v and the entire anchor set S, we utilize another aggregation function A_2 to aggregate the messages $M_v[i]^l$ (Section 3.2.2-(4)), and these aggregated messages serve as inputs to the subsequent hidden layer. In the final layer, a non-linear transformation is applied to derive the relative position encoding $p_v \in \mathbb{R}^k$, using trainable weights $w \in \mathbb{R}^r$ and a non-linearity function σ.

(2) Anchor set selection. We use the most objectively attractive locations in the study region as anchors, which are reference points for relative position encoding. Specifically, we first calculate the sum of total in-flow and the population of each node, which serve as proxies to reflect the number of opportunities and objective attractiveness of the node. We observe that a small number of nodes attract a large proportion of travels. We select these attractive nodes as the most objectively attractive locations in the study region. Finally, we randomly sample k anchor sets $S = \{S_1, S_2, ..., S_k\}$ (S denotes the entire anchor set) from these most objectively attractive locations for each layer of the PRE module.

(3) a reachability measure based on preferential random walks. To capture the positions of various nodes in relation to the anchors, it is necessary to establish the connection between nodes and anchors in the network to manifest the positional disparities of different nodes. In this study, we represent this connection by measuring the reachability between nodes and anchors. We estimate the reachability of all nodes by means of preferential random walks, which are deemed capable of approximating population-level human mobility (Yan *et al.* 2017).

'Preference' refers to the transition probability during the random walks process. Concretely, a random walk with walk length $Walk_L$, which start from node v_i and transfer to adjacent node u through an edge $e = (v_i, u)$ based on a transition probability $p(e)$. The calculation equation for the transition probability is as follows:

$$p(e) = \frac{w_e}{\sum_{\forall e' \in N(v_i)} w_{e'}} \tag{4}$$

where w_e means the weight of edge e, $N(v_i)$ means the all-outgoing edges from v_i. In this study, a physical model is used to generate an initial flow network with the generated flow intensity as the weight of edge.

We introduce a reachability measure from v to u, denoted as $m(v, u)$, defined as follows:

$$m(v, u) = \frac{\sum_{i=1}^{Walk_N} count_i(v, u)}{Walk_L \times Walk_N} \tag{5}$$

where $count_i(v, u)$ represents the number of times a random walker accesses node u in the i-th random walk starting from node v, and $Walk_N$ represents the number of random walks start from node v.

(4) Computing comprehensive reachability. The comprehensive reachability computation function F combines the reachability in both directions. The equation is written as follows:

$$F\left(v, u, h_v^l, h_u^l\right) = \left(\left(m(v,u)h_v^l\right)||\left(m(u,v)h_u^l\right)\right) \tag{6}$$

where $||$ is the concatenation operation, h_v^l is attributes of node v.

We use the MEAN aggregation function, which is permutation invariant, to instantiate two comprehensive reachability aggregation functions, A_1 and A_2, as shown in Figure 5. The equations are as follows:

$$A_1\left(Me_i{}^l\right) = \frac{1}{a_i}\sum_{i=1}^{a_i} Me_i{}^l \tag{7}$$

$$A_2\left(M_v{}^l\right) = \frac{1}{k}\sum_{i=1}^{k} M_v[i]^l \tag{8}$$

where a_i is the number of anchors in the i-th anchor set S_i, and k is the total number of anchor sets.

3.2.3. Encoding average travel expense priors by node attribute augmentation

The model we proposed is built upon the GNN architecture (Battaglia *et al*. 2018). However, the traditional message-passing based GNNs may struggle to effectively use node attribute (i.e., population), edge attribute (i.e., distance), and graph structure to capture the underlying average travel expenses. To address this, we need to incorporate average travel expense priors into GNN. We apply a simple and efficient *node attribute augmentation* to model the node level travel pattern by defining a quantitative indicator that can measure average travel expenses as an attribute of origin nodes.

The definition of average travel expenses λ_i for origin node r_i is as follows:

$$\lambda_i = \frac{\sum_{j\in\mathbb{N}(i)} pop_i \times pop_j \times d_{ij}}{D(i)} \tag{9}$$

where pop_i is the population of r_i, d_{ij} is the Euclidean distance between r_i and r_j, $D(i)$ means the degree of the node r_i, $\mathbb{N}(i)$ represents the neighborhoods of node r_i. λ_i is utilized in the subsequent computation of 'relative attractiveness' embeddings.

3.3. Generating total out-flow for each origin

In this section, we generate the total out-flow for each origin, which lays the groundwork for the subsequent flow allocation among various destinations.

We select six factors that empirically contribute to the total out-flow of origin, i.e., the population of origin O, pop_O; average travel expenses for origin O, λ_O; the total population of all destination from O, $pop_{\{D\}}$; the sum of the distances between O and each destination from O, $d_{O,\{D\}}$; the sum of the ratio of the population of each destination to the distance between it and the corresponding origin O, $R_{O,\{D\}}$; and comprehensive out-flow direction of the origin O, $\overrightarrow{F_O} = \sum_{D\in\mathbb{N}(O)} \overrightarrow{p_D} - \overrightarrow{p_O}$. $\mathbb{N}(O)$ represents the neighborhood of origin O, $\overrightarrow{p_O}$ means the relative position encoding of origin O.

Then, the above six factors are used to train a multilayer perceptron to generate the total out-flow $\hat{O}$ for each origin, the equation is as follows:

$$\hat{O}_o = MLP\left(pop_O||\lambda_O||pop_{\{D\}}||d_{O,\{D\}}||R_{O,\{D\}}\Big|\Big|\overrightarrow{Fo}\right) \tag{10}$$

where *MLP* represents a multilayer perceptron, || is the concatenation operation.

3.4. Prior-guided graph encoder for learning 'relative attractiveness' embeddings

As shown in Figure 6, the graph encoder is based on the mobility flow network and consists of multiple layers. In each layer, we design a message-passing method to learn the 'relative attractiveness' embeddings for the edges in the mobility flow network. Then, the final 'relative attractiveness' embeddings are obtained by weighting the sum of the learned embeddings of each layer with learnable weights W.

3.4.1. Initialization of embeddings within the mobility flow network

We initialize the embeddings of the origin and destination nodes separately. The specific initialization of embeddings within the mobility flow network is as follows:

$$a_O^0 = \hat{O}_o; a_D^0 = pop_D; a_{\overline{D}}^0 = pop_{\overline{D}} \tag{11}$$

$$E_{O\to D}^0 = d_{OD}; E_{O\to\overline{D}}^0 = d_{O\overline{D}} \tag{12}$$

where $\overline{D}$ is all destinations node within the neighborhood of O, a_O^0 means the attributes of O, $\hat{O}_o$ is the generated total out-flow, pop_D means the population of D, $E_{O\to D}^0$ denotes the attribute of edge between O and D, d_{OD} denotes the Euclidean distance between O and D.

3.4.2. Simulating mobility flow generation through message-passing

The message-passing process primarily involves node aggregation, node update, edge update and position update. The first three computations are guided by physical models, with the computation proceeding from the node level to the edge level. As shown in Figure 7, the core approach involves decomposing the equations of physical models (gravity model or radiation model) to reflect the flow generation mechanisms embodied in these models. This decomposition is designed to ensure that the resulting equations progressively emulate the flow generation process. The decomposed equations

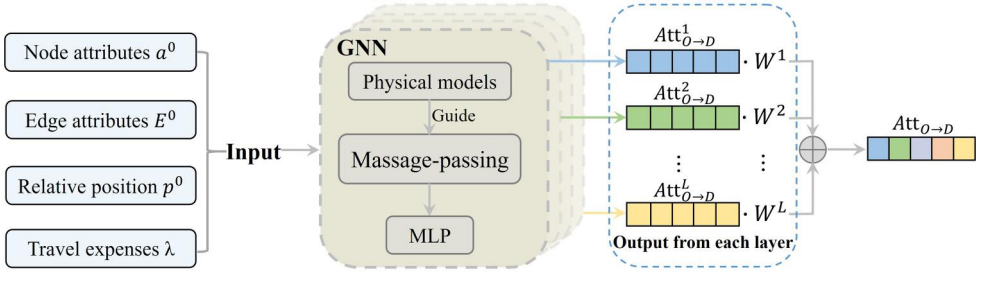

Figure 6. The architecture of the prior-guided graph encoder.

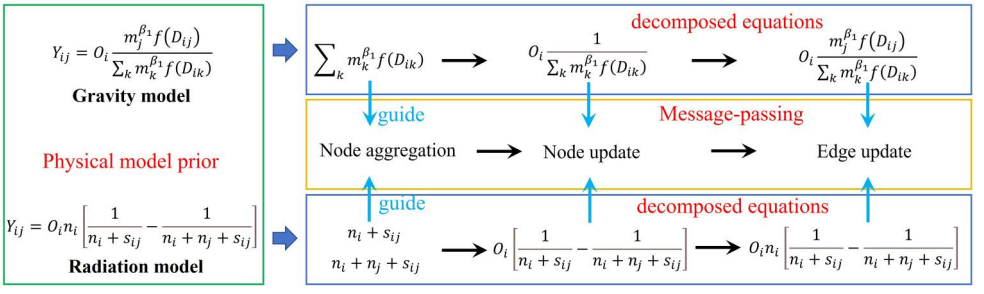

Figure 7. The overview of message-passing process guided by physical model priors.

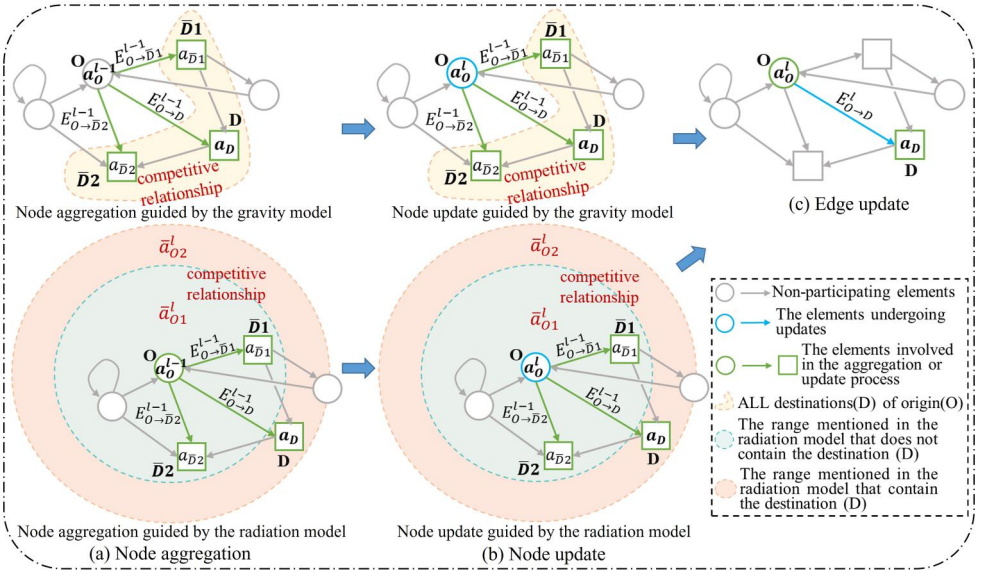

Figure 8. The message-passing process that emulates the generation of mobility flow. There is a competitive relationship among the destinations of the same origin. The variables used in Figure are from Equations (13) to (20).

are then utilized to guide the computation of message aggregation and update functions, which take advantage of the powerful nonlinear fitting capabilities of deep networks. Consequently, the message-passing process we have designed is capable of simulating mobility flow generation.

Specifically, Figure 8 provides a visual representation of the elements within an exemplar mobility flow network graph that participates in each of these computational steps in the message-passing process. For instance, the node aggregation process involves the embedding of destination nodes and the edge embedding between origin node and different destination nodes. Note the steps of node aggregation and update have two versions since they are guided by two different physical models.

(1) Node aggregation. The purpose of node aggregation is to aggregate the node embeddings and edge embeddings in the node's neighborhood. In this study, for origin nodes, we use node aggregation to emulate the aggregation of 'relative

opportunity' factors that contribute to the generation of travel demand. For destination nodes, we do not perform node aggregation operations.

We believe that the impact of the number of opportunities and distance at the destination on the origin should be comprehensively considered. For example, although the numerous opportunities at destination a can be attractive to origin b, the long distance between a and b can reduce this attractiveness. We propose a simple concept of 'relative opportunity' RO to describe this phenomenon, and the calculation equation is as follows:

$$RO_{O\to D} = \frac{a_D^0}{E_{O\to D}^0} \tag{13}$$

Since the RO is a simplified concept of 'relative opportunity', we further integrate it into the message-passing process to construct a nonlinear representation of nodes and edges within the mobility flow network.

Then, two specific node aggregation methods are developed based on two physical models. The physical meaning of node aggregation is to aggregate the relative opportunities of all destination node within the neighborhood of origin node and measure the total 'relative opportunities' available to the origin from all competing destinations, as shown in Figure 8(a). These destination nodes are in competition with each other, i.e., the greater the 'relative opportunities' of the destination, the stronger its competitiveness.

• *Node aggregation guided by the gravity model.* According to the '$\sum_k m_k^{\beta_1} f(D_{ik})$' part of Equation (1), the equation for aggregation is as follows:

$$\overline{a}_O^l = \rho\left(a_{\overline{D}}^{l-1}, E_{O\to\overline{D}}^{l-1}\right) = MLP_1\left(\sum_{k\in\overline{D}} \frac{a_k^{l-1}}{E_{O\to k}^{l-1}}\right) = MLP_1\left(\sum_{k\in\overline{D}} RO_{O\to k}^{l-1}\right) \tag{14}$$

where $\overline{a}_O^l$ denotes the aggregated message of the origin node at l-th layer, MLP_1 is a multilayer perceptron.

• *Node aggregation guided by the radiation model.* According to the '$n_i + s_{ij}$' and '$n_i + n_j + s_{ij}$' part of Equation (3), each node r_i needs to aggregate 'relative opportunity' messages within its two neighborhoods, where one neighborhood (Nei1) is a range centered on O with a radius of the distance between O and D and does not contain destination D, the other neighborhood (Nei2) is based on neighbourhood1 containing destination D. The equation for aggregation is as follows:

$$\overline{a}_{O1}^l = \rho\left(a_{\text{Nei1}}^{l-1}, E_{O\to\text{Nei1}}^{l-1}\right) = MLP_1\left(\sum_{k\in\text{Nei1}} \frac{a_k^{l-1}}{E_{O\to k}^{l-1}}\right) = MLP_1\left(\sum_{k\in\text{Nei1}} RO_{O\to k}^{l-1}\right) \tag{15}$$

$$\overline{a}_{O2}^l = \rho\left(a_{\text{Nei2}}^{l-1}, E_{O\to\text{Nei2}}^{l-1}\right) = MLP_1\left(\sum_{k\in\text{Nei2}} \frac{a_k^{l-1}}{E_{O\to k}^{l-1}}\right) = MLP_1\left(\sum_{k\in\text{Nei2}} RO_{O\to k}^{l-1}\right) \tag{16}$$

where ρ is a node aggregation operation, $\overline{a}_{O1}^l$ and $\overline{a}_{O2}^l$ denotes the aggregated message of the origin node in the two neighborhoods, respectively.

(2) Node update. We employ initial node embeddings and message derived from node aggregation to perform node updates, as shown in Figure 8(b). The updated

node embeddings are utilized to represent the travel demand of the O for the unit 'relative opportunity' (1/*ALL_RO*) (guided by the gravity model) or a traveler originating from O (guided by the radiation model), after the O perceives the all 'relative opportunities' *ALL_RO* that may be available.

• *Node update guided by the gravity model.* When guided by the gravity model, the physical meaning of node update is to estimate travel demand resulting from a unit relative opportunity (1/*ALL_RO*), which reflects the influence that each unit of relative opportunity has on shaping travel choices.

According to the "$O_i \frac{1}{\sum_k m_k^{\beta_1} f(D_{ik})}$" part of equation (1), the equation for node update is as follows:

$$a_O^l = \phi^a\left(a_O^{l-1}, \overline{a}_O^l\right) = MLP_2\left(a_O^{l-1} \otimes \frac{1}{\overline{a}_O^l}\right) \tag{17}$$

where a_O^l is the updated node embedding, ϕ^a is a node update operation that acts on each origin node in graph, $\otimes$ denotes Hadamard product operation.

• *Node update guided by the radiation model.* When guided by the radiation model, the physical meaning of node update is to estimate travel demand of a traveler originating from O who selecting the opportunity at D instead of opportunities available in other neighborhood.

According to the "$O_i\left[\frac{1}{n_i+s_{ij}} - \frac{1}{n_i+n_j+s_{ij}}\right]$" part of equation (3), the equation for node update is as follows:

$$a_O^l = \phi^a\left(a_O^{l-1}, \overline{a}_{O1}^l, \overline{a}_{O2}^l\right) = MLP_2\left(a_O^{l-1} \otimes \left(\frac{1}{\overline{a}_{O1}^l} - \frac{1}{\overline{a}_{O2}^l}\right)\right) \tag{18}$$

(3) Edge update. As shown in Figure 8(c), the edge update process involves the updated embedding of the O node, the embedding of the D node and the pre-update edge embedding. The physical meaning of edge update is to update the overall travel demand estimate from O to D when only 'relative opportunities' are considered.

• *Edge update guided by the gravity model.* According to the complete equation (1), the equation for edge update is as follows:

$$E_{O\to D}^l = \phi^E\left(a_O^l, a_D^{l-1}, E_{O\to D}^{l-1}\right) = MLP_3\left(MLP_1\left(\frac{a_D^{l-1}}{E_{O\to D}^{l-1}}\right) \otimes a_O^l\right) \tag{19}$$

where $\frac{a_D^{l-1}}{E_{O\to D}^{l-1}} = RO_{O\to D}^{l-1}$.

The above equation expresses the selection of relative opportunities from O to D.

• *Edge update guided by the radiation model.* According to the complete Equation (3) and in order to maintain a consistent understanding and application of n_i with Equations (15) and (16), the equation for edge update is a special form of Equation (19).

$$E_{O\to D}^l = \phi^E\left(a_O^l, a_{D(O)}^{l-1}, E_{O\to D(O)}^{l-1}\right) = MLP_3\left(MLP_1\left(\frac{a_{D(O)}^{l-1}}{E_{O\to D(O)}^{l-1}}\right) \otimes a_O^l\right) \tag{20}$$

where $D(O)$ denotes that O is acting as destination at this time.

The equation mentioned above expresses the travel demand of all travelers from O who select the opportunities at D rather than those in other neighborhood.

(4) Updating positional embeddings. To integrate the positional embedding information from neighboring nodes into the positional embedding of a given node, we update the positional embedding of each region node r_i by the weighted sum of all relative differences $(p_i - p_j)_{\forall j}$, which indicates the relative position relationship between OD pair $(i \rightarrow j)$ and the entire anchor set. The weights are obtained through the transformation of edge embedding m_{ij}, which carry the information of the whole graph. The equations for positional embedding update are as follows and are based on the work of Satorras *et al*. (2021).

$$p_{ij}^{l} = \phi_p\left(p_i^{l-1}, p_j^{l-1}\right) = p_i^{l-1} - p_j^{l-1} \tag{21}$$

$$m_{ij} = \phi_e\left(a_i^{l-1}, a_j^{l-1}, ||p_{ij}^{l}||^2, e_{ij}^{l-1}\right) \tag{22}$$

$$p_i^{l} = \phi_x\left(p_i^{l-1}, p_j^{l-1}, m_{ij}\right) = p_i^{l-1} + C\sum_{j\neq i} p_{ij}^{l} \cdot \phi_m(m_{ij}) \tag{23}$$

where p_i is the relative position encoding of r_i, e_{ij} is the Euclidean distance between O and D, C equals $1/(M-1)$ and M is the number of nodes in the graph, ϕ_e is edge operation, and $\phi_m : \mathbb{R}^N \rightarrow \mathbb{R}^1$ can transform the m_{ij} into a scalar value.

(5) Computation of 'relative attractiveness' embeddings. We argue that the core driven factor of human travel demand within a region is the attractiveness of other regions to that region. The spatial heterogeneity in travel demand is due to differences in the attractiveness that depends on both origin and destination. Therefore, to further describe the spatial heterogeneity in human travel demand, we propose the concept of 'relative attractiveness', which describes the attractiveness of a destination for its origins based on a comprehensive measurement of factors that affect human travel choices, including position, average travel expense, opportunities (pop_O and pop_D) and distance (d_{OD}). After the integration of priors with GNN, the position is represented as p_{OD}^{l}, the average travel expenses is represented as λ_O, and the combined influence of opportunities and distance is represented as $E_{O\rightarrow D}^{l}$. The equation for calculating 'relative attractiveness' is as follows:

$$p_{OD}^{l} = \phi_p\left(p_O^{l-1}, p_D^{l-1}\right) = p_O^{l-1} - p_D^{l-1} \tag{24}$$

$$Att_{O\rightarrow D}^{l} = \varphi\left(E_{O\rightarrow D}^{l}, p_{OD}^{l}, \lambda_O\right) = MLP_4\left(pop_O||pop_D||d_{OD}||E_{O\rightarrow D}^{l}||p_{OD}^{l}||\lambda_O\right) \tag{25}$$

$$Att_{O\rightarrow D} = \sum_{l\in L} W^{l} \cdot Att_{O\rightarrow D}^{l} \tag{26}$$

where p_{OD}^{l} means the position of the OD pair relative to the entire study region, $Att_{O\rightarrow D}^{l}$ denotes the 'relative attractiveness' embeddings computed in l-th layer, $Att_{O\rightarrow D}$ denotes the final 'relative attractiveness' embeddings, W is learnable weights, $||$ is the concatenation operation.

3.5. Generator using GBRT for estimating mobility flow

We use the learned 'relative attractiveness' embedding in the encoder between any origin node O and destination node D, concatenate the generated total outflow from O, as input to a GBRT, and finally generate the mobility flow between O and D. The generator is formularized as follow:

$$\hat{F}_{OD} = GBRT\left(Att_{O\to D}||\hat{O}_o\right) \tag{27}$$

where $||$ is the concatenation operation. GBRT is an ensemble learning algorithm that combines the optimization strategy of gradient descent with the robust modeling power of decision trees (Safavian and Landgrebe 1991).

3.6. Loss function

We acquire structural embeddings through message passing method and position embeddings via relative location encoding. As the learning of these two types of embeddings is separate, the final loss function needs to integrate the losses of two embedding tasks: a task loss ($\mathcal{L}_{Task1}$ or $\mathcal{L}_{Task2}$) and a positional encoding loss $\mathcal{L}_{Pos}$. The MSELoss (Mean squared Error loss) is employed for the task loss, and the positional encoding loss is computed using the Laplacian eigenvector loss (Belkin and Niyogi 2003), which forces positional encoding to form a coordinate system constrained by the graph topology (Dwivedi *et al*. 2021). The equation is as follows,

$$\mathcal{L}_{Task1} = \sum_i \left(O_i - \hat{O}_i\right)^2; \mathcal{L}_{Task2} = \sum_{i,j} \left(F_{ij} - \hat{F}_{ij}\right)^2 \tag{28}$$

$$\mathcal{L}_{Pos} = tr\left(p^T L p\right) = \sum_{i,j} ||p_i - p_j||^2 \cdot e^{-||x_i - x_j||^2} \tag{29}$$

$$\mathcal{L}_{total1} = \mathcal{L}_{Task1} + \alpha_1 \mathcal{L}_{Pos}; \mathcal{L}_{total2} = \mathcal{L}_{Task2} + \alpha_2 \mathcal{L}_{Pos} \tag{30}$$

where L is the Laplacian matrix, x_i is the real coordinate position of the region r_i, α_1 and α_2 are hyper-parameters greater than 0, $||\cdot||^2$ is the Frobenius norm, $\mathcal{L}_{total1}$ and $\mathcal{L}_{total2}$ are the final loss of both tasks, respectively. O_i and $\hat{O}_i$ denote the real and generated total out-flow from origins, F_{ij} and $\hat{F}_{ij}$ denote the real and generated flow intensity data.

4. Experiments

4.1. Datasets

We chose New York state and Pennsylvania state as the study area. Census tracts were taken as the basic areal unit. The data for each study area contains OD flow data, population and coordinates of the census tracts.

The data for New York State comes from the work by Simini *et al*. (2021) which proposed an advanced deep gravity model for mobility flow generation. To compare with the deep gravity model, we conducted experiments on the New York State dataset using experimental settings similar to that reported in Simini *et al*. (2021).

The data for Pennsylvania state were obtained from an OD flow open data (Kang *et al*. 2020). The mobility flows between census tracts provided by the open data are extracted from millions of anonymous cell phone users visiting various places, and has a high correlation with the American Commuting Survey (ACS) commuting flows patterns, which indicates the high reliability of the produced data.

4.2. Baselines

We compared three categories of baseline methods to prove the effectiveness of the proposed models. The first category is classical physical models, including the Gravity Model (GM) (Zipf 1946, Wilson 1969) and the Radiation Model (RM) (Simini *et al*. 2012). The second category is traditional machine learning methods, including the Random Forest (RF) (Breiman *et al*. 2001) and the Gradient Boosting Regression Tree (GBRT) (Prettenhofer and Louppe 2014). The third category is deep learning models, including the GNN_based Model (GNNM) (Luo and Chen 2024) and the Deep Gravity (DG) (Simini *et al*. 2021).

4.3. Evaluation metrics

We used Normalized Root Mean Squared Error (NRMSE), Pearson correlation (Corr), Common Part of Commuters (CPC) and Jesen-Shannon Divergence (JSD) as evaluation metrics to compare the performance of the baselines and the proposed models. CPC is the most commonly used metric for mobility flow generation models (Simini *et al*. 2021). JSD is used to evaluate the similarity between the overall distribution of real flow data and generated flow data. The higher the CPC, Corr and the lower the NRMSE, JSD, the better the performance.

4.4. Model settings

In this section, we introduce the parameter settings of the proposed models in the experiments. The number of graph encoder layers was set to 2 for all models using GNN and the estimators of RF and GBRT was set to 100. We used the RMSprop optimizer with momentum 0.9 and learning rate $5{\cdot}10^{-6}$ for all deep learning related models. Besides, we also divided the total OD pairs into nearly 1:1 as the training and test set. In the relative position encoder module, the number of random walks $Walk_N$ and the length of each random walk $Walk_L$ were set to 100/50 and 100/40 for the New York and Pennsylvania dataset, respectively.

4.5. Results and analysis

4.5.1. Evaluation of the performance of mobility flow generation

We develop two models for mobility flow generation: PG-MFG(GM), informed by the Gravity Model, and PG-MFG, informed by the Radiation Model. As depicted in Table 1, the evaluation results yielded the following key findings:

Table 1. Overall performance on real-world datasets considering training data size. The best results are presented in bold italics, while the top results from the baseline model are underlined.

		Training data size					Mean of metrics			
		75%	60%	50%	30%	15%	CPC	NRMSE	Corr.	JSD
New York	CPC									
GM		0.416	0.376	0.340	0.402	0.275	0.362	12.226	0.045	0.635
RM		0.491	0.565	0.406	0.393	0.395	0.450	11.602	0.102	0.594
RF		0.718	0.703	0.702	0.698	0.648	0.694	1.973	0.937	0.093
GBRT		0.725	0.714	0.712	0.715	0.677	0.709	1.625	0.956	0.088
GNNM		0.713	0.691	0.703	0.629	0.638	0.675	2.789	0.873	0.081
DG		0.724	0.685	0.682	0.671	0.655	0.683	2.014	0.921	0.342
PG-MFG(GM)		0.806	0.792	0.781	0.776	0.770	0.785	1.204	0.976	0.047
PG-MFG(RM)		0.798	0.794	0.788	0.787	0.785	***0.790***	***1.196***	***0.977***	***0.045***
Pennsylvania	CPC									
GM		0.197	0.192	0.192	0.213	0.210	0.201	15.307	0.054	0.652
RM		0.234	0.237	0.237	0.244	0.214	0.233	14.856	0.108	0.613
RF		0.686	0.673	0.667	0.659	0.639	0.665	2.207	0.969	0.115
GBRT		0.695	0.682	0.675	0.667	0.651	0.674	2.133	0.972	0.108
GNNM		0.655	0.644	0.637	0.645	0.623	0.641	3.462	0.916	0.102
DG		0.687	0.671	0.658	0.664	0.628	0.662	2.230	0.956	0.361
PG-MFG(GM)		0.753	0.747	0.744	0.745	0.727	0.743	1.809	0.978	0.082
PG-MFG(RM)		0.759	0.751	0.748	0.742	0.732	***0.746***	***1.795***	***0.980***	***0.079***

The proposed models consistently outperform all baseline models across all evaluation metrics. Based on the mean of these metrics, the proposed models demonstrated significant improvements over the best-performing baseline: the NRMSE was reduced by more than 25% in New York and 15% in Philadelphia, and the CPC saw an increase of over 10% in both states. A investigation into the role of each type of prior in the proposed models is provided in Appendix A. Additionally, Appendix B offers a comparative visualization of the generated mobility flows by both the proposed models and the baseline models.

Traditional physical models exhibited the poorest performance, with the highest NRMSE of 12.226 in New York and 15.307 in Philadelphia, and the lowest CPC of 0.362 and 0.201, respectively. These models' simplicity hinders their ability to accurately capture the complex spatial heterogeneity inherent in human mobility.

Deep learning models, when not leveraging sufficient prior knowledge of human mobility, generally fail to outperform traditional machine learning models. The mean of evaluation metrics from the two study areas indicate that the RF and GBRT models achieve higher CPC and lower NRMSE than the GNNM and DG. This is likely due to the strong spatial heterogeneity of human mobility, which can lead to overfitting in deep learning models when addressing the problem of mobility flow generation.

A deep learning model that does not incorporate information from surrounding nodes when modeling mobility flows between origin-destination pairs is likely to perform poorly in terms of network statistics similarity. The comparison between the GNNM and DG models reveals that while the DG model has a slightly higher CPC than GNNM, its JSD is significantly higher. This indicates that the DG model is unable to sufficiently capture the spatial heterogeneity inherent in human mobility, thereby limiting its capacity to generate a flow distribution that closely resembles the real mobility flow network.

Although the performance of all non-physical models declines with a decrease in the proportion of training data, the proposed models exhibit relatively good stability, indicating superior generalization capabilities. GNNM and DG, being deep learning models, possess a larger number of parameters that require optimization, thus necessitating more extensive training data. Consequently, their performance is more sensitive to the amount of training data available. The proposed models, however, takes full advantage of necessary prior knowledge, making it more adept at capturing the intrinsic pattern of human mobility and allowing it to achieve better performance even with a reduced training dataset.

4.5.2. Evaluation of the model's capability to capture spatial heterogeneity of human mobility

To test our model's capacity to capture this spatial heterogeneity, we evaluated the performance of the proposed models and all baseline models on generating mobility flow for three distinct sets of OD pairs selected in the entire test dataset, which have similar population distributions and travel distances (the population difference between various nodes is less than 100, and the travel distance difference between distinct OD pairs is less than 1 km):

1. OD pairs with the same origin but different destinations (Same O & Different D);
2. OD pairs with the same destination but different origins (Different O & Same D);
3. OD pairs with the different origins and different destinations (Different O & Different D).

From the evaluation results in Table 2, we have the following findings:

The proposed models steadily achieve the best performance in mobility flow generation evaluation for three types of OD pairs, indicating that our models have a stronger ability to capture spatial heterogeneity of human mobility than all baseline

Table 2. The performance comparison of mobility flow generation for specific OD pairs.

	Same O & Different D				Different O & Same D				Different O & Different D			
	NRMSE	Corr	CPC	JSD	NRMSE	Corr	CPC	JSD	NRMSE	Corr	CPC	JSD
New York												
GM	11.682	0.032	0.395	0.593	11.627	0.037	0.408	0.584	13.547	0.030	0.228	0.634
RM	11.467	0.096	0.452	0.559	11.398	0.102	0.471	0.548	13.498	0.046	0.238	0.605
RF	2.239	0.889	0.658	0.108	2.205	0.887	0.670	0.104	1.296	0.925	0.798	0.045
GBRT	1.981	0.907	0.673	0.095	1.921	0.903	0.686	0.091	0.967	0.953	0.819	0.039
GNNM	2.164	0.882	0.692	0.073	2.113	0.880	0.702	0.069	1.649	0.881	0.762	0.051
DG	2.316	0.875	0.652	0.365	2.264	0.879	0.668	0.358	1.208	0.921	0.796	0.132
PG-MFG(GM)	1.214	0.967	0.741	0.059	1.147	0.968	0.751	0.056	0.632	0.984	0.875	0.024
PG-MFG(RM)	***1.185***	***0.968***	***0.750***	***0.058***	***1.119***	***0.969***	***0.761***	***0.054***	***0.614***	***0.985***	***0.877***	***0.023***
Pennsylvania												
GM	12.679	0.038	0.304	0.663	12.415	0.057	0.335	0.654	14.968	0.150	0.052	0.872
RM	13.942	0.026	0.266	0.645	13.883	0.051	0.274	0.641	12.594	0.275	0.307	0.623
RF	3.067	0.876	0.519	0.203	2.846	0.877	0.539	0.191	0.598	0.965	0.891	0.029
GBRT	3.046	0.877	0.530	0.196	2.838	0.877	0.546	0.186	0.572	0.968	0.898	0.028
GNNM	3.159	0.859	0.504	0.204	3.012	0.848	0.522	0.193	1.025	0.929	0.801	0.058
DG	3.280	0.834	0.495	0.206	3.156	0.830	0.496	0.201	0.563	0.968	0.901	0.029
PG-MFG(GM)	2.307	0.932	0.597	0.146	2.135	0.931	0.608	0.139	***0.229***	***0.994***	***0.949***	***0.007***
PG-MFG(RM)	***1.479***	***0.982***	***0.767***	***0.074***	***1.431***	***0.980***	***0.773***	***0.070***	0.241	0.989	0.943	0.008

models. From an overall trend in the evaluation metrics, it is observed that our models perform better on the Type 3 OD pairs compared to the first two. This is primarily because the first two types of OD pairs share a common origin (O) or destination (D), leading to a high competitive or collaborative relationship between OD pairs. Consequently, the spatial heterogeneity in human mobility within the first two types is not only determined by the spatial position differences of O or D but is also influenced by the complex dependencies between different OD pairs' mobility flows. In contrast, the Type 3 OD pairs exhibits weaker interdependencies, and their spatial heterogeneity is predominantly determined by the comprehensive spatial position differences of O and D. Moreover, our models exhibit slightly improved performance on Type 2 OD pairs compared to Type 1. This suggests that when the D is the same, the variations in travel demands from different O are relatively minor. However, when the O is the same, the differences in the attractiveness of various D become more pronounced. Consequently, the influence of the destination on the spatial heterogeneity of travel demand is more substantial.

The GNNM outperforms other baselines on the first two types of OD pairs. This is primarily because GNNM aggregates messages from neighboring OD pairs, thereby accounting for dependencies between them to some extent. However, for Type 3, GNNM does not show a performance advantage. This is attributed to the lack of specific position awareness in GNNM, which hinders its ability to capture spatial heterogeneity in human mobility for OD pairs with substantial spatial position differences.

We further validate the impact of our message-passing design and relative location encoding on capturing spatial heterogeneity of human mobility (see Appendix C for details).

4.5.3. Evaluation of generalization capability

To further verify the model's generalization ability, we assessed its transferability by evaluating its performance when trained in one study area and applied to another.

As Table 3 indicates, models trained in New York exhibited a notable decline in performance when applied in Pennsylvania. This degradation is primarily attributed to the differences in human mobility patterns between the two states. The proposed prior-guided deep learning models surpasses all others in transfer performance. It achieves over a 12% improvement in CPC in New York (test area) and more than a 7% improvement in Pennsylvania (test area) when compared to the best-performing baseline. This

Table 3. Mobility flow generation results on test areas.

	New York → Pennsylvania				Pennsylvania → New York			
Model	NRMSE	Corr	CPC	JSD	NRMSE	Corr	CPC	JSD
GM	15.847	0.043	0.150	0.682	13.025	0.069	0.382	0.601
RM	14.596	0.051	0.229	0.629	11.879	0.075	0.419	0.584
RF	3.133	0.942	0.559	0.186	1.636	0.954	0.655	0.089
GBRT	2.273	0.950	0.571	0.172	1.634	0.957	0.664	0.080
GNNM	4.285	0.888	0.507	0.168	3.331	0.913	0.630	0.076
DG	4.915	0.944	0.465	0.476	4.084	0.924	0.597	0.389
PG-MFG(GM)	2.179	***0.979***	0.612	0.155	1.521	0.967	0.743	0.062
PG-MFG(RM)	***2.113***	0.977	***0.625***	***0.154***	***1.414***	***0.969***	***0.776***	***0.054***

reaffirms that our model's ability to capture the spatial heterogeneity of human mobility is superior, thereby enhancing the model's generalization capability.

In line with the results from Table 1, the RF and GBRT models demonstrate stronger generalization ability compared to the GNNM and DG models. Moreover, the GNNM model outperforms the DG model in generalization, with its JSD being the best among all baseline models. This indicates that deep learning models leveraging GNNs have an edge in capturing spatial heterogeneity in human mobility patterns when compared to non-GNN deep learning models.

5. Discussion

This study draws on the research paradigm of Physics-Informed Machine-Learning (Karniadakis *et al*. 2021) and introduces a mobility flow generation model that integrates comprehensive prior guidance. The primary reason for our models' superiority lies in their explicit consideration of inherent spatial heterogeneity in human mobility based on the integration of prior knowledge and data-driven model. In contrast, most existing studies neglect this critical spatial heterogeneity, that is, they overlook the fundamental source of errors in human mobility generation, which contributes to the performance bottleneck in current models.

Specifically, first, spatial heterogeneity of human mobility is inherently location-dependent (Lomi 1995). Traditional physical models, which ignore positional information, often underperform in regions with strong spatial heterogeneity of human mobility. Data-driven models such as RF, GBRT, and DG (Simini *et al*. 2021) also overlook position information. Although these models significantly enhance their nonlinear fitting capabilities leveraging ensemble learning or deep learning (Rong *et al*. 2023b), they are at risk of overfitting due to the strong spatial heterogeneity of human mobility. Compared to the DG model, GNN-based models excel in capturing dependencies between regional nodes and exhibit stronger inductive learning capabilities (Hamilton *et al*. 2017). However, most GNN-based models (Liu *et al*. 2020, Luo and Chen 2024) learn node embeddings by encoding information about their local neighborhoods (Nishad *et al*. 2020). These approaches, which rely solely on neighborhood information, fail to capture the position of origins and destinations within the broader context of the mobility flow network graph structure (You *et al*. 2019). Compared with above studies, the relative position encoding developed in this study endows the proposed models with the ability to be spatially explicit and to capture the global positions of origins and destinations, thereby overcoming the limitations of existing GNN-based models. The positional awareness capability assists the model in learning the nonlinear relationships between mobility flows and their driving factors (e.g., opportunity and distance), while considering positions.

Secondly, existing models lack a fundamental understanding of the underlying reasons for the formation of spatial heterogeneity in human mobility. Traditional physical models cannot adequately account for the underlying mechanism of mobility flow generation for specific OD pairs. Most data-driven models simply leverage the features of OD pair (e.g., regional attributes) and real mobility flow for end-to-end learning (Pourebrahim *et al*. 2019, Simini *et al*. 2021, Rong *et al*. 2023a). Although these models

consider an increasing number of driving factors contributing to mobility flow, they do not collectedly describe these factors based on the underlying reasons for the spatial heterogeneity of human mobility, making them struggle to sufficiently capture the spatial heterogeneity shown in Figure 2(c). This study argues that the collective influence of these driving factors ultimately forms the relative attraction of destinations to origins, and this differential in relative attractiveness is the underlying reason for the spatial heterogeneity of human mobility. Consequently, this study proposes the concept of 'relative attractiveness' and fully integrates prior knowledge and GNN architecture to obtain its embedding. This approach enables the proposed models to focus on learning differences in relative attractiveness to capture the spatial heterogeneity of complex and diverse human mobility. Furthermore, the models proposed in this study is a general framework and can incorporate more influencing factors to achieve a better performance in human flow generation.

6. Conclusion

Constructing a universal population-level mobility flow generation model presents significant challenges, primarily due to the inherent spatial heterogeneity in human mobility, which stems from the complexity of human travel behavior. Overly simple physical models and recently proposed data-driven machine learning models that suffer from poor generalization, both have limited expressive power to capture this spatial heterogeneity. In this study, we introduce a prior-guided, data-driven model named PG-MFG, which addresses these limitations through a combination of prior knowledge and GNN architecture. We conducted extensive experiments using real-world human mobility datasets from New York and Pennsylvania. The results demonstrate that the proposed PG-MFG model not only achieves superior accuracy and generalization performance but also has a greater ability to capture spatial heterogeneity compared to all the baseline models.

There are still some limitations. This study considers a limited scope of prior knowledge. Future work could incorporate a wider spectrum of prior knowledge and explore its integration with deep learning approaches to develop more universally applicable human mobility generation models.

Acknowledgments

The numerical calculations in this paper have been done on the supercomputing system in the Supercomputing Center of Wuhan University.

Disclosure statement

No potential conflict of interest was reported by the author(s).

Funding

This work was supported by the National Natural Science Foundation of China under [Grant NOs. 42371470 and 42401524] and China National Postdoctoral Program for Innovative Talents under [Grant NO. BX20230360].

Notes on contributors

Zhipeng Wang is a PhD candidate at the State Key Laboratory of Information Engineering in Surveying, Mapping and Remote Sensing (LIESMARS), Wuhan University, Wuhan, China. His research interests include spatially explicit machine learning and geospatial artificial intelligence (GeoAI) approaches to human mobility modeling. His contribution to this paper includes the conception of the research idea, software, visualization, and writing of the manuscript.

Tong Zhang is a Professor with the State Key Laboratory of Information Engineering in Surveying, Mapping and Remote Sensing (LIESMARS), Wuhan University, Wuhan, China. His research topics include urban computing and machine learning. His contribution to this paper includes conceptualization, manuscript revision, experimental analysis, and funding acquisition.

Jing Li is an associate professor in the Department of Geography and the Environment at University of Denver. Her research areas are geovisualization, high performance geocomputation, and spatiotemporal data modeling and analysis. Her contribution to this paper includes the supervision of the experiments, research design, and the writing of the manuscript.

Peixiao Wang is a Postdoctoral Fellow from State Key Laboratory of Resources and Environmental Information System, Institute of Geographic Sciences and Nature Resources Research, Chinese Academy of Sciences. His research topics include spatiotemporal data mining, and spatiotemporal prediction, especially focus on spatiotemporal prediction of transportation systems. His contribution to this paper includes experimental analysis and validation.

Yu'ang Zhu is a Master candidate at the State Key Laboratory of Information Engineering in Surveying, Mapping and Remote Sensing (LIESMARS), Wuhan University, Wuhan, China. His research interests include spatiotemporal data mining and human mobility modeling. His contribution to this paper includes experimental analysis and manuscript revision.

ORCID

Zhipeng Wang http://orcid.org/0009-0007-1547-1779
Tong Zhang http://orcid.org/0000-0002-0683-4669
Peixiao Wang http://orcid.org/0000-0002-1209-6340

Data and codes availability statement

The data and codes that support the findings of this study are available in 'figshare.com' with the identifier https://doi.org/10.6084/m9.figshare.27642891.

References

Barbosa, H., *et al.*, 2018. Human mobility: models and applications. *Physics Reports*, 734, 1–74.

Battaglia, P.W., *et al.*, 2018. Relational inductive biases, deep learning, and graph networks. arxiv preprint arxiv:1806.01261.

Belkin, M., and Niyogi, P., 2003. Laplacian eigenmaps for dimensionality reduction and data representation. *Neural Computation*, 15 (6), 1373–1396.

Boyce, D.E., and Williams, H.C., 2015. *Forecasting urban travel: Past, present and future*. Cheltenham: Edward Elgar Publishing.

Breiman, L., 2001. Random forests. *Machine Learning*, 45 (1), 5–32.

de Dios Ortúzar, J., and Willumsen, L., 2011. *Modeling Transport*, New York: John Wiley and Sons Ltd.

Dueñas, M., and Fagiolo, G., 2013. Modeling the international-trade network: A gravity approach. *Journal of Economic Interaction and Coordination*, 8 (1), 155–178.

Dwivedi, V.P., *et al.*, 2021. Graph neural networks with learnable structural and positional representations. arxiv preprint arxiv:2110.07875.
Fotheringham, A.S., *et al.*, 1996. The geography of parameter space: an investigation of spatial non-stationarity. *International Journal of Geographical Information Systems*, 10 (5), 605–627.
Friedman, J.H., 2001. Greedy function approximation: a gradient boosting machine. *Annals of Statistics*, 29 (5), 1189–1232.
Gu, X., *et al.*, 2024. Predicting the network shift of large urban agglomerations in China using the deep-learning gravity model: a perspective of population migration. *Cities*, 145, 104680.
Hamilton, W., *et al.*, 2017. Inductive representation learning on large graphs. *Advances in Neural Information Processing Systems*, 30, 1024–1034.
Kang, Y., *et al.*, 2020. Multiscale dynamic human mobility flow dataset in the US during the COVID-19 epidemic. *Scientific Data*, 7 (1), 390.
Karniadakis, G., *et al.*, 2021. Physics-informed machine learning. *Nature Reviews Physics*, 3 (6), 422–440.
Liu, Z., *et al.*, 2020. April. Learning geo-contextual embeddings for commuting flow prediction. *In Proceedings of the AAAI conference on artificial intelligence* (Vol. 34, No. 01, pp. 808–816).
Luca, M., *et al.*, 2021. A survey on deep learning for human mobility. *ACM Computing Surveys*, 55 (1), 1–44.
Lomi, A., 1995. The population ecology of organizational founding: Location dependence and unobserved heterogeneity. *Administrative Science Quarterly*, 40 (1), 111–144.
Luo, M., and Chen, Y., 2024. Simulating inter-city population flows based on graph neural networks. *Geocarto International*, 39 (1), 2331223.
Masucci, A., *et al.*, 2013. Gravity versus radiation models: On the importance of scale and heterogeneity in commuting flows. *Physical Review. E, Statistical, Nonlinear, and Soft Matter Physics*, 88 (2), 022812.
Mwale, M., *et al.*, 2022. Factors that affect travel behaviour in developing cities: A methodological review. *Transportation Research Interdisciplinary Perspectives*, 16, 100683.
Nishad, S., *et al.*, 2020. Graphreach: Position-aware graph neural network using reachability estimations. arxiv preprint arxiv:2008.09657.
Prettenhofer, P., and Louppe, G., 2014. February. Gradient boosted regression trees in scikit-learn. In PyData 2014.
Pourebrahim, N., *et al.*, 2019. Trip distribution modeling with Twitter data. *Computers, Environment and Urban Systems*, 77, 101354.
Reilly, W., 1931. *The law of retail gravitation*. New York: W. J. Reilly.
Rong, C., *et al.*, 2023a. GODDAG: Generating origin-destination flow for new cities via domain adversarial training. *IEEE Transactions on Knowledge and Data Engineering*, 35 (10), 10048–10057.
Rong, C., *et al.*, 2023b. OD network generation via gravity-guided GAN. arXiv:2306.03390v1.
Safavian, S.R., and Landgrebe, D., 1991. A survey of decision tree classifier methodology. *IEEE Transactions on Systems, Man, and Cybernetics*, 21 (3), 660–674.
Satorras, V.G., *et al.*, 2021. E (n) equivariant graph neural networks. In *International conference on machine learning* (pp. 9323–9332). PMLR.
Schläpfer, M., *et al.*, 2021. The universal visitation law of human mobility. *Nature*, 593 (7860), 522–527.
Sharma, S., 2019. Factors influencing travel behavior and mode choice. *Geospatial World*
Shi, Q., *et al.*, 2024. A fusion model of temporal graph attention network and machine learning for inferring commuting flow from human activity intensity dynamics. *International Journal of Applied Earth Observation and Geoinformation*, 126, 103610.
Simini, F., *et al.*, 2012. A universal model for mobility and migration patterns. *Nature*, 484 (7392), 96–100.
Simini, F., *et al.*, 2021. A deep gravity model for mobility flows generation. *Nature Communications*, 12 (1), 6576.
Stouffer, S.A., 1940. Intervening opportunities: a theory relating mobility and distance. *American Sociological Review*, 5 (6), 845–867.

Tang, J., *et al.*, 2024. Predicting human mobility flows in response to extreme urban floods: A hybrid deep learning model considering spatial heterogeneity. *Computers, Environment and Urban Systems*, 113, 102160.
Yan, X.Y., *et al.*, 2017. Universal model of individual and population mobility on diverse spatial scales. *Nature Communications*, 8 (1), 1639.
Yin, G., *et al.*, 2023. ConvGCN-RF: A hybrid learning model for commuting flow prediction considering geographical semantics and neighborhood effects. *GeoInformatica*, 27 (2), 137–157.
You, J., *et al.*, 2019. Position-aware graph neural networks. In *International conference on machine learning* (pp. 7134–7143). PMLR.
Wang, L., *et al.*, 2019. Cross-city transfer learning for deep spatio-temporal prediction. In *Proceedings of the 28th International Joint Conference on Artificial Intelligence* (IJCAI), 1893–1899.
Wilson, A.G., 1969. The use of entropy maximising models, in the theory of trip distribution, mode split and route split. *Journal of Transport Economics and Policy*, 3 (1), 108–126.
Wu, Z., *et al.*, 2021. A comprehensive survey on graph neural networks. *IEEE Transactions on Neural Networks and Learning Systems*, 32 (1), 4–24.
Zipf, G., 1946. The P1 P2/D hypothesis: On the intercity movement of persons. *American Sociological Review*, 11 (6), 677–686.
Zhou, Z., *et al.*, 2023. Predicting collective human mobility via countering spatiotemporal heterogeneity. *IEEE Transactions on Mobile Computing*.

Appendices

A. Ablation studies

We delve into the role of each type of prior knowledge played within our models. To do this, we systematically removed key components: (1) the node attribute augmentation module (no NA), (2) the relative position encoding module (no RPE), (3) the message-passing module (no MP), (4) all three modules simultaneously (no ALL). For each variation, we assessed the models' performance. The outcomes of these evaluations are depicted in Figure A1 and A2.

Node attribute augmentation

we first assessed the impact of incorporating the node attribute augmentation module. Based on the CPC and NRMSE metrics, this augmentation boosts the quality of the generated mobility

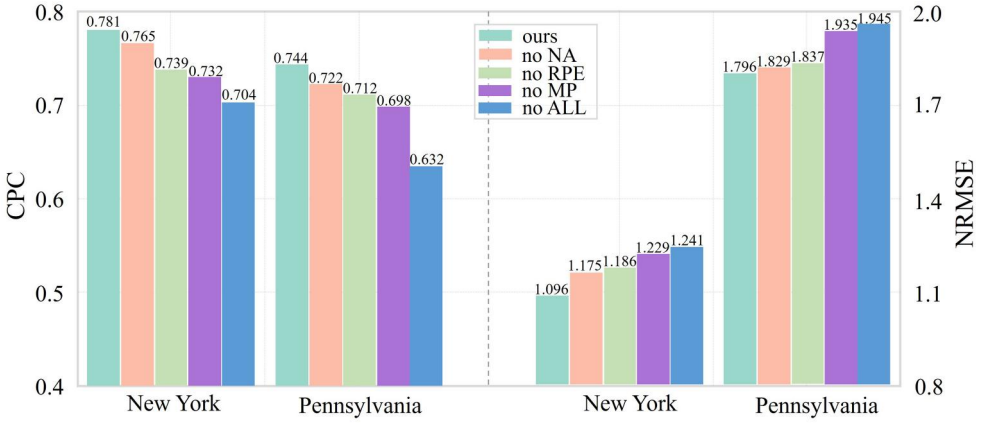

Figure A1. Ablation study on PG-MFG(GM).

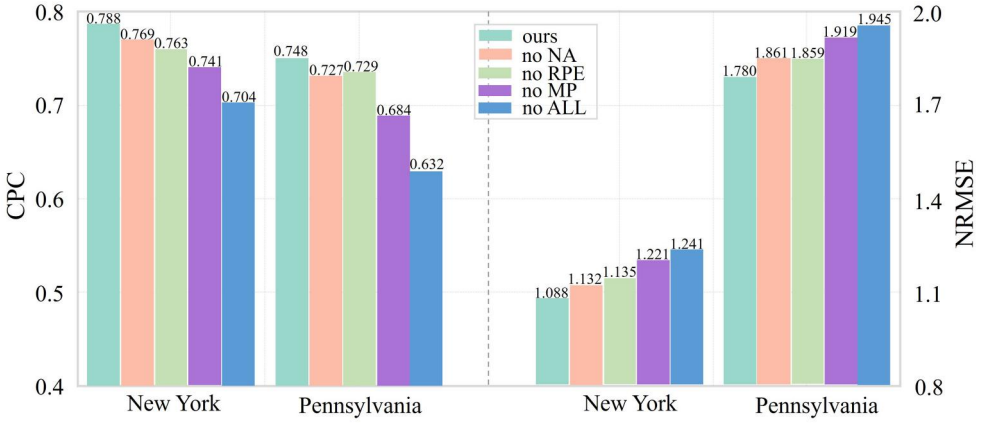

Figure A2. Ablation study on PG-MFG(RM).

flow data across both study areas. The introduction of the node attribute augmentation module alone yields a performance improvement of approximately 2% to 3%. This augmentation enriches the node attributes, providing a more informative embedding that enables the models to learn the differences in travel patterns among various origin nodes.

Relative position encoding

Previous studies have often overlooked the explicit modeling of positional effects on human flow generation, despite the fact that spatial positioning is crucial for capturing the heterogeneity in human mobility. To address this, we integrated a relative position encoding module to equip our models with positional awareness of both origin and destination nodes. The results depicted in Figure A1 and A2 indicate that this addition alone can enhance performance by approximately 3% to 6%. Moreover, the findings reveal that the relative position encoding module has a more pronounced impact on the performance of PG-MFG(GM) compared to PG-MFG(RM). This disparity could stem from the fact that the gravity model (GM) largely disregards the spatial position disparities of origin and destination nodes. In contrast, the radiation model (RM) implicitly accounts for these spatial position differences by factoring in the number of opportunities present in varying neighborhoods of the origin node.

Message-passing

Experiments conducted with a model that solely incorporates the designed message-passing method reveal significant gains in performance. This method stands out as the most impactful enhancement to our models, offering performance improvements ranging from approximately 6% to 10%. These results underscore the effectiveness of combining the strengths of physical models and deep learning models in generating human mobility flow.

All

The objective of this experiment was to assess the overall performance enhancement of the model design when compared to a variant without any prior knowledge enhancements. As illustrated in Figures A1 and A2, the incorporation of the model designs yields substantial benefits, achieving an

overall performance improvement ranging from approximately 9% to 18%. This result further confirms the effectiveness of the prior-guided deep learning paradigm, especially in tackling the complex spatial heterogeneity that is the key challenge of mobility flow generation.

B. Visual comparisons

The visualization of generated mobility flows by both the proposed models and baseline models, using the New York dataset, is presented in Figure B1. Generally speaking, the proposed models exhibit smaller absolute residuals between the generated and observed mobility flows compared to the baseline models, and the mobility flow network formed by the generated flows more closely resembles the actual network. As depicted in Figures B1(b) and B1(c), the proposed models are particularly skilled at accurately generating flows for longer distances, while the less accurate flows are predominantly found in medium to short-distance OD pairs. Both the proposed models and baseline models tend to perform less well in regions with high mobility flow intensity, such as New York City, due to the larger population and more complex travel behaviors present in these areas, leading to increased spatial heterogeneity in human mobility. However, when compared to the best-performing baseline model (GBRT), the proposed models show superior performance in regions with dense mobility flows, particularly near New York City (marked with green circles) (Figure B1(b)-(d)), demonstrating their enhanced capability to capture the complex spatial heterogeneity of human mobility.

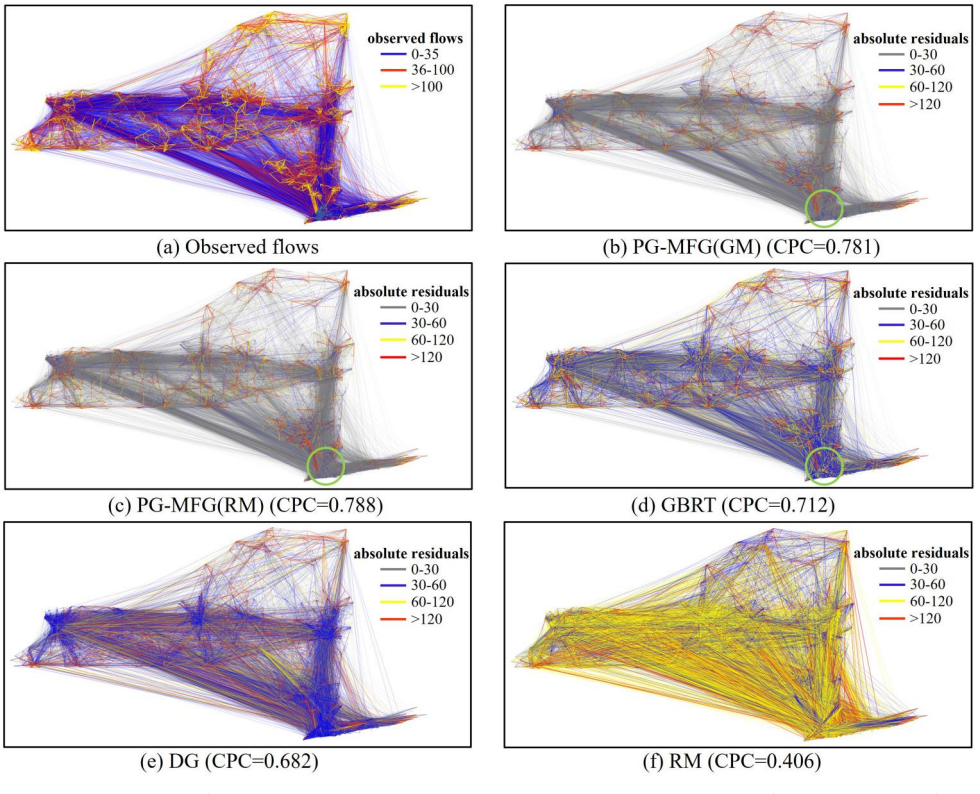

Figure B1. Mobility flow generation in New York State. (a): visualization of the observed flows. (b)-(f): the difference between the mobility flows generated by the proposed models/baseline models and the observed flows. CPC indicates the Common Part of Commuters.

Table C1. The performance of the proposed model variants in generating mobility flows for specific OD pairs.

Evaluation metrics	NRMSE	Corr	CPC	JSD
Same O & different D				
PG-MFG(GM)—no ALL	1.894	0.836	0.626	0.132
PG-MFG(GM)—only MP	1.715	0.842	0.651	0.094
PG-MFG(GM)—only PRE	1.500	0.854	0.663	0.097
Different O & same D				
PG-MFG(GM)—no ALL	1.822	0.837	0.640	0.113
PG-MFG(GM)—only MP	1.589	0.846	0.656	0.088
PG-MFG(GM)—only PRE	1.453	0.851	0.675	0.092
Different O & different D				
PG-MFG(GM)—no ALL	1.177	0.941	0.795	0.084
PG-MFG(GM)—only MP	1.113	0.946	0.811	0.042
PG-MFG(GM)—only PRE	0.938	0.964	0.842	0.038

C. The performance of the model variants in generating mobility flows for specific OD pairs

We developed three variants of the PG-MFG(GM) model: (1) a variant with all prior knowledge-based modules removed (no ALL); (2) a variant that adds only our designed message-passing module to the Variant1 (only MP); and (3) a variant that adds only the relative position encoding module to the Variant1 (only PRE). We then evaluated the flow generation performance of these variants on the three types of OD pairs using the New York dataset.

As Table C1 demonstrates, when compared to Variant (1), the performance of the other two variants showed improvement across all four evaluation metrics. This enhancement indicates that both of our model designs effectively augment the model's capability to capture the spatial heterogeneity of human mobility. Notably, the relative position encoding module provides a greater enhancement to this capability. This is primarily because geographical position is a crucial factor associated with spatial heterogeneity, as noted by Fotheringham *et al*. (1996). Our findings also underscore the importance of incorporating spatial position in generative models of human activity.