

# Hierarchical Frequency-Decomposition Graph Neural Networks for Road Network Representation Learning

Jingtian Ma<sup>1, 3</sup>, Jingyuan Wang<sup>1, 2, 3, 4\*</sup>, Leong Hou U<sup>5</sup>

<sup>1</sup>School of Computer Science and Engineering, Beihang University, Beijing, China

<sup>2</sup>School of Economics and Management, Beihang University, Beijing, China

<sup>3</sup>MITT Key Laboratory of Data Intelligence and Management, Beihang University, Beijing, China

<sup>4</sup>MOE Engineering Research Center of Advanced Computer Application Technology, Beihang University, China

<sup>5</sup>University of Macau, Macau SAR, China

## Abstract

Road networks are critical infrastructures underpinning intelligent transportation systems and their related applications. Effective representation learning of road networks remains challenging due to the complex interplay between spatial structures and frequency characteristics in traffic patterns. Existing graph neural networks for modeling road networks predominantly fall into two paradigms: spatial-based methods that capture local topology but tend to over-smooth representations, and spectral-based methods that analyze global frequency components but often overlook localized variations. This spatial-spectral misalignment limits their modeling capacity for road networks exhibiting both coarse global trends and fine-grained local fluctuations. To bridge this gap, we propose HiFiNet, a novel hierarchical frequency-decomposition graph neural network that unifies spatial and spectral modeling. HiFiNet constructs a multi-level hierarchy of virtual nodes to enable localized frequency analysis, and employs a decomposition–updating–reconstruction framework with a topology-aware graph transformer to separately model and fuse low- and high-frequency signals. Theoretically justified and empirically validated on multiple real-world datasets across four downstream tasks, HiFiNet demonstrates superior performance and generalization ability in capturing effective road network representations.

## 1 Introduction

The road network serves as the fundamental component of intelligent transportation systems (ITS), which supports a wide range of traffic-related applications such as traffic forecasting (Ji et al. 2020; Jiang et al. 2023a; Ji et al. 2025), trajectory inference (Wu et al. 2019b; Guo et al. 2020), and urban planning (Wang et al. 2021b; Chen et al. 2024). To facilitate these applications, it is crucial to build effective and generalizable representations of road networks. A widely adopted approach is to model the road network as a graph, where nodes denote road segments and edges reflect topological connectivity. Learning expressive representations for such road network graphs remains a core challenge.

Early studies for modeling road networks often rely on random walk-based approaches (Perozzi, Al-Rfou, and Skiena 2014; Grover and Leskovec 2016), which generate

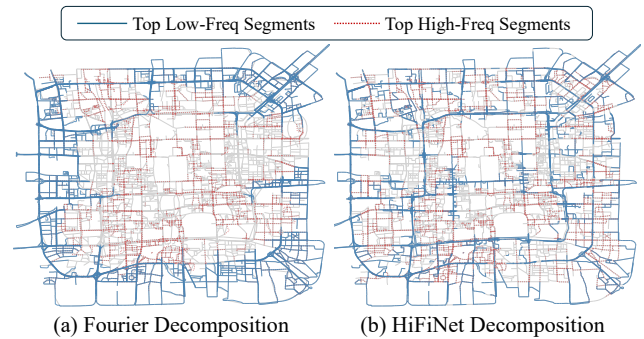


Figure 1: The frequency decomposition of Fourier Transform and HiFiNet on traffic flow signals.

node sequences through random walks and treat them as sentences to learn node embeddings. While simple and scalable, these methods neglect node attributes and fail to capture structural semantics such as frequency-aware patterns.

To address these limitations, graph neural networks (GNNs) have emerged as powerful tools for graph representation learning. Broadly, existing GNNs can be categorized into two paradigms: *spectral-based* and *spatial-based* approaches. Spectral methods (Bruna et al. 2013; Defferrard, Bresson, and Vandergheynst 2016) are grounded in graph signal processing theory, defining graph convolutions via the eigenbasis of the Laplacian to analyze node features in the frequency domain. In contrast, spatial-based GNNs (Ji et al. 2022; Zhang et al. 2024; Han et al. 2025) perform message passing by directly aggregating features from local neighborhoods, enabling efficient and inductive learning on graphs with varying topology. More recently, transformer-based architectures have been extended to graph domains (Dwivedi and Bresson 2020; Ying et al. 2021), incorporating global attention mechanisms along with structural or positional encodings to capture long-range dependencies beyond local neighborhoods.

However, despite these advances, these two paradigms remain largely *spatial-spectral misalignment*. Spatial-based models capture local structural patterns but inherently behave as low-pass filters (Wu et al. 2019a; Bastos et al. 2022), leading to over-smoothing and poor global expressiveness.

\*Corresponding author.

Spectral-based models, while offering theoretical insights through global frequency decomposition, often overlook localized variations. Consequently, there is a critical gap: a unified framework that simultaneously models spatial structures and frequency components is still lacking, which limits the expressive capacity of current models for road networks that exhibit both coarse global patterns and fine-grained local variations. As illustrated in Fig. 1(a), applying the Fourier Transform to traffic flow signals reveals distinct frequency characteristics: low-frequency edges (blue solid lines) generally appear in peripheral areas, whereas high-frequency edges (red dotted lines) concentrate in city centers. This reflects a typical urban pattern—peripheral regions are shaped by regular commuting patterns, while central areas exhibit more fluctuation due to diverse land use. However, exceptions may exist: some peripheral roads maintain stable flow, while certain inner-city segments show significant volatility. These observations support the presence of spatial-spectral misalignment in existing approaches, where frequency patterns are not well localized, leading to suboptimal representation of complex spatial structures.

To address these challenges, we propose a novel *Hierarchical Frequency-Decomposition Network (HiFiNet)* for road network representation learning. Our framework integrates spatial and spectral modeling in a unified architecture through two key innovations: First, we introduce a three-level hierarchy by clustering road segments into *localities* and *regions*, each represented by virtual nodes. This hierarchy captures multi-scale spatial semantics enables localized graph signal decomposition. As shown in Fig. 1(b), low-frequency edges arise not only at city margins but also along inner ring roads where traffic evolves in a smooth and stable manner. We theoretically prove that this hierarchical projection naturally separates low- and high-frequency components, mitigating over-smoothing and enhancing representational diversity. Second, we design a frequency-decomposition learning module following a *decomposition–updating–reconstruction* paradigm. It explicitly models low- and high-frequency signals and updates them via a topology-aware graph transformer, capturing both smooth global trends and sharp local variations. These enriched components are then fused into discriminative representations under a unified loss framework that promotes consistency across frequencies and scales.

Our contributions are summarized as follows:

- We propose *HiFiNet*, a unified spatial-spectral framework that integrates hierarchical graph modeling with localized frequency decomposition for road network representation learning.
- We design a multi-level hierarchy that not only captures spatial locality but also enables frequency separation, a property we theoretically validate in our framework.
- We develop a frequency decomposition module that jointly models low- and high-frequency graph signals, enhancing the expressiveness of learned representations.
- Extensive experiments on real-world datasets across four downstream tasks demonstrate the superior performance and generalization ability of our approach.

## 2 Related Work

Our work is related to the following research directions:

**Road Network Modeling.** Road network modeling focuses on capturing the structural and semantic characteristics of urban road systems. Early methods such as DeepWalk (Perozzi, Al-Rfou, and Skiena 2014) and Node2vec (Grover and Leskovec 2016) used random walks to learn shallow representations, but lacked the capacity to incorporate node attributes. With the advent of GNNs, models (Wang et al. 2020; Chen et al. 2021; Wang et al. 2021a; Mao et al. 2022; Jiang et al. 2023b; Ji et al. 2023; Zhang and Long 2023; Zhou et al. 2024; Ma et al. 2025) leveraged neighbor aggregation for representation learning. To model long-range dependencies, hierarchical approaches such as HRNR (Wu et al. 2020) introduced pooling operations. However, most GNN-based models suffer from over-smoothing and often overlook high-frequency components that are critical for preserving fine-grained road patterns.

**Graph Spectral Theory.** Graph spectral methods analyze structural properties through spectral decomposition of matrices like the Laplacian. Classical models such as ChebNet (Defferrard, Bresson, and Vandergheynst 2016), GCN (Kipf and Welling 2017), and CayleyNet (Levie et al. 2018) inherently perform low-pass filtering, which aids denoising but causes over-smoothing. Recent studies (Wu et al. 2019a; Zhu et al. 2021; Bo et al. 2021) propose frequency-aware GNNs that balance low- and high-frequency components for better expressiveness. However, most spectral methods focus on node-level tasks, and their potential for modeling hierarchical and functional patterns in road networks remains underexplored.

**Graph Neural Networks.** GNNs have achieved remarkable success in graph representation learning. Representative models include GCN (Kipf and Welling 2017), GAT (Velickovic et al. 2017), and GraphSAGE (Hamilton, Ying, and Leskovec 2017), which follow a message-passing paradigm. However, these methods tend to behave as low-pass filters, causing over-smoothing (Nt and Maehara 2019). To better capture global context, recent work explores Graph Transformers (Dwivedi and Bresson 2020; Ying et al. 2021; Wu et al. 2022), which use attention mechanisms to learn long-range dependencies. Yet, how to balance local structural priors and global flexibility remains an open challenge.

## 3 Preliminaries

In this section, we introduce the notations used throughout the paper and formally define our task.

**Definition 1 (Road Network)** A road network is modeled as a directed graph  $\mathcal{G} = \langle \mathcal{S}, \mathbf{A}_{\mathcal{S}} \rangle$ , where  $\mathcal{S}$  denotes the set of  $N_{\mathcal{S}}$  road segments, and  $\mathbf{A}_{\mathcal{S}} \in \mathbb{R}^{N_{\mathcal{S}} \times N_{\mathcal{S}}}$  is the binary adjacency matrix, with  $\mathbf{A}_{\mathcal{S}}[i, j] = 1$  indicating a directed connection from segment  $s_i$  to  $s_j$ , and 0 otherwise.

**Definition 2 (Segment Signal)** The segment signal matrix  $\mathbf{X}_{\mathcal{S}} \in \mathbb{R}^{N_{\mathcal{S}} \times d_0}$  encodes raw attributes (e.g., road class, lane number, traffic flow), where each row  $\mathbf{x}_{\mathcal{S}}^i$  corresponds to segment  $s_i$ . From the graph signal processing view,  $\mathbf{X}_{\mathcal{S}}$  can be

decomposed into low-frequency components (smooth global patterns) and high-frequency components (local variations).

To facilitate frequency-aware hierarchical modeling, we introduce two virtual node types: *localities* and *regions*, and organize the road network as a three-level hierarchy: *segment*  $\rightarrow$  *locality*  $\rightarrow$  *region*.

**Definition 3 (Locality)** A locality  $l \in \mathcal{L}$  refers to a group of spatially adjacent road segments that collectively serve a specific traffic-related function (e.g., overpass, intersection), where  $\mathcal{L}$  denotes the set of  $N_L$  localities.

**Definition 4 (Region)** A region  $r \in \mathcal{R}$  consists of multiple localities and represents a broader urban area with a specific functional role (e.g., residential, commercial zone), where  $\mathcal{R}$  denotes the set of  $N_R$  regions.

**Definition 5 (Hierarchical Road Network)** We define a hierarchical road network as  $\mathcal{H} = \langle \mathcal{V}, \mathcal{E} \rangle$ , where  $\mathcal{V} = \mathcal{S} \cup \mathcal{L} \cup \mathcal{R}$  denote the set of all nodes, and  $\mathcal{E} = \{\mathbf{A}_S, \mathbf{A}_L, \mathbf{A}_R, \mathbf{A}_{SL}, \mathbf{A}_{LR}\}$  includes adjacency matrices for: (1) segment–segment ( $\mathbf{A}_S$ ), (2) locality–locality ( $\mathbf{A}_L$ ), (3) region–region ( $\mathbf{A}_R$ ), (4) segment–locality ( $\mathbf{A}_{SL}$ ), and (5) locality–region ( $\mathbf{A}_{LR}$ ) relations.

Unlike segment adjacency matrix, the matrices  $\mathbf{A}_L$ ,  $\mathbf{A}_R$ ,  $\mathbf{A}_{SL}$ , and  $\mathbf{A}_{LR}$  are treated as learnable associations. In particular,  $\mathbf{A}_{SL}$  and  $\mathbf{A}_{LR}$  act as *segment-to-locality* and *locality-to-region* assignment matrices that support hierarchical aggregation. With these definitions, we are now ready to formally define our task.

**Definition 6 (Road Network Representation Learning)** Given a road network  $\mathcal{G}$  and segment signal matrix  $\mathbf{X}_S$ , the objective is to construct the hierarchical road network  $\mathcal{H}$  and learn a  $d$ -dimensional embedding  $\mathbf{h}_m \in \mathbb{R}^d$  for each node  $m \in \mathcal{V}$ , where  $d \ll |\mathcal{V}|$ . The learned representations are expected to preserve both low- and high-frequency semantics and generalize across downstream traffic tasks.

## 4 Hierarchical Frequency-Decomposition Network

In this section, we propose the *Hierarchical Frequency-Decomposition Network (HiFiNet)* for road network representation learning. As illustrated in Fig. 2, HiFiNet constructs a three-level hierarchical architecture to capture the multi-scale structure of road networks. We theoretically demonstrate that this design enables effective decomposition of *low-frequency* and *high-frequency* components in road signals. These components are then explicitly modeled and fused to produce the final representations that capture both smooth and variant patterns of the road network.

### 4.1 Hierarchical Architecture Modeling

**Contextual Embedding for Road Segments.** Road segments are often associated with rich contextual attributes. To incorporate this auxiliary information, we map these attributes to latent embedding vectors for individual segments.

Given a segment  $s_i$ , its raw attribute vector  $\mathbf{x}_S^i \in \mathbb{R}^{d_0}$  (Def. 2) includes four key attributes: segment ID, lane number (LN), segment length (SL), and geographical location

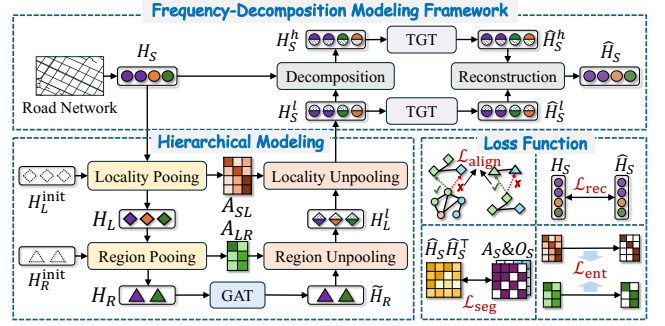


Figure 2: The overall framework of HiFiNet.

(longitude and latitude, denoted as LL). We map each attribute (or its discretized bin) to a learnable embedding, denoted as  $e_{ID}^i$ ,  $e_{LN}^i$ ,  $e_{SL}^i$ , and  $e_{LL}^i$ , respectively. The resulting contextual embedding of  $s_i$  is:

$$\mathbf{v}_S^i = e_{ID}^i \parallel e_{LN}^i \parallel e_{SL}^i \parallel e_{LL}^i, \quad (1)$$

where “ $\parallel$ ” denotes vector concatenation.

Stacking the contextual embeddings of all segments yields the matrix  $\mathbf{V}_S \in \mathbb{R}^{N_S \times d'}$ , where  $d'$  is the total dimension of the concatenated embeddings. We then apply a two-layer feed-forward network (FFN) with nonlinearity to obtain the initial segment feature matrix:

$$\mathbf{H}_S = \text{FFN}(\mathbf{V}_S), \quad (2)$$

where  $\mathbf{H}_S \in \mathbb{R}^{N_S \times d}$  denotes the initial segment features used in the subsequent hierarchical modeling.

**Locality Graph Construction.** Localities serve as clusters of segments to capture local connectivity patterns such as intersections. A key component is the *segment-to-locality* assignment matrix  $\mathbf{A}_{SL}$ , which encodes how segments are grouped into localities based on structural similarity.

We assume that each segment is softly assigned to a locality, with contributions weighted by importance. To model this varying importance, we employ a cross-attention mechanism to capture the interactions between segments and localities. Let  $\mathbf{H}_L^{\text{init}} \in \mathbb{R}^{N_L \times d}$  be a randomly initialized learnable matrix representing locality embeddings. The assignment matrix is computed as

$$\mathbf{A}_{SL} = \text{softmax} \left( \frac{(\mathbf{H}_S \mathbf{W}_S)(\mathbf{H}_L^{\text{init}} \mathbf{W}_L)^\top}{\sqrt{d}} \right), \quad (3)$$

where  $\text{softmax}(\cdot)$  denotes the row-wise normalization, and  $\mathbf{W}_S$ ,  $\mathbf{W}_L$  are learnable projection matrices. Each entry in  $\mathbf{A}_{SL}$  models the conditional probability of assigning segment  $s_i$  to locality  $l_j$ :  $\mathbf{A}_{SL}[i, j] = \Pr(l_j | s_i)$ .

Given  $\mathbf{A}_{SL}$ , we compute locality features by aggregating segment features with residual connections:

$$\mathbf{H}_L = \mathbf{A}_{SL}^\top \mathbf{H}_S + \mathbf{H}_L^{\text{init}}. \quad (4)$$

The locality adjacency matrix is constructed as:

$$\mathbf{A}_L = \mathbf{A}_{SL}^\top \mathbf{A}_S \mathbf{A}_{SL}, \quad (5)$$

which can be interpreted as

$$\mathbf{A}_L[i, j] = \sum_{s_m, s_n} \Pr(l_i | s_m) \cdot \Pr(l_j | s_n) \cdot \mathbf{A}_S[m, n]. \quad (6)$$

**Region Graph Construction.** Urban areas often consist of regions with distinct functional roles (Yuan, Zheng, and Xie 2012). We build a region-level graph over localities using a learnable *locality-to-region* assignment matrix  $\mathbf{A}_{LR}$ .

Let  $\mathbf{H}_R^{\text{init}} \in \mathbb{R}^{N_R \times d}$  denote the initial region embeddings. The assignment matrix is computed similarly via attention:

$$\mathbf{A}_{LR} = \text{softmax} \left( \frac{(\mathbf{H}_L \mathbf{W}_L)(\mathbf{H}_R^{\text{init}} \mathbf{W}_R)^\top}{\sqrt{d}} \right), \quad (7)$$

where  $\mathbf{W}_L$  and  $\mathbf{W}_R$  are learnable projection matrices and each entry  $\mathbf{A}_{LR}[j, k] = \Pr(r_k | l_j)$ .

The region features and adjacency matrix are constructed by aggregating locality features and connections:

$$\mathbf{H}_R = \mathbf{A}_{LR}^\top \mathbf{H}_L + \mathbf{H}_R^{\text{init}}, \quad \mathbf{A}_R = \mathbf{A}_{LR}^\top \mathbf{A}_L \mathbf{A}_{LR}, \quad (8)$$

We theoretically demonstrate that our hierarchical structure exhibits favorable spectral properties. Specifically:

**Theorem 1** *Let  $\mathbf{A}_{XY} \in \mathbb{R}^{N_Y \times N_X}$  denote an assignment matrix satisfying the equi-partition and row-normalization properties. Then, the projection of graph signals from the original graph  $X$  to the coarsened graph  $Y$  approximately preserves the low-frequency energy while attenuating high-frequency components.*

This theorem indicates that the proposed hierarchical projection naturally acts as a spectral low-pass filter, preserving smooth signal components while suppressing high-frequency noise.

**Low-frequency Feature Propagation.** Once the *segment-locality-region* hierarchical structure is established, we adopt a top-down message propagation strategy to propagate these preserved low-frequency features.

We begin by applying a standard Graph Attention Network (GAT) (Velickovic et al. 2017) to update the node features in the region-level graph, allowing each region node to capture coarse-grained global contextual information:

$$\tilde{\mathbf{H}}_R = \text{GAT}(\mathbf{H}_R, \mathbf{A}_R). \quad (9)$$

Next, since regions aggregate multiple localities, they capture coarse-grained low-frequency patterns. We propagate this information to localities via the assignment matrix  $\mathbf{A}_{LR}$ , followed by a GAT module for refinement:

$$\tilde{\mathbf{H}}_L^l = \mathbf{A}_{LR} \tilde{\mathbf{H}}_R, \quad \mathbf{H}_L^l = \text{GAT}(\tilde{\mathbf{H}}_L^l, \mathbf{A}_L), \quad (10)$$

where  $\mathbf{H}_L^l \in \mathbb{R}^{N_L \times d}$  is the low-frequency locality features.

Similarly, segment-level low-frequency features are obtained by aggregating from localities via  $\mathbf{A}_{SL}$ :

$$\tilde{\mathbf{H}}_S^l = \mathbf{A}_{SL} \mathbf{H}_L^l, \quad \mathbf{H}_S^l = \text{GAT}(\tilde{\mathbf{H}}_S^l, \mathbf{A}_S), \quad (11)$$

where  $\mathbf{H}_S^l \in \mathbb{R}^{N_S \times d}$  is the low-frequency segment features.

This top-down unpooling process effectively suppresses high-frequency fluctuations and retains low-frequency components, providing essential input for the subsequent frequency decomposition modeling module.

## 4.2 Frequency Decomposition Modeling

We adopt a *decomposition-updating-reconstruction* framework to explicitly separate and model the low-frequency and high-frequency components of segment features, which enables more effective representation learning.

**Decomposition Stage.** Since the original segment features  $\mathbf{H}_S$  obtained by Eq. (2) contain both low-frequency and high-frequency components, we can obtain the high-frequency part by subtracting the low-frequency signal:

$$\mathbf{H}_S^h = \mathbf{H}_S - \mathbf{H}_S^l, \quad (12)$$

where  $\mathbf{H}_S, \mathbf{H}_S^l, \mathbf{H}_S^h \in \mathbb{R}^{N_S \times d}$  denote the original, low-frequency, and high-frequency segment features separately.

**Updating Stage.** As road segment graphs often involve a large number of nodes, limiting aggregation to local neighborhoods may fail to capture long-range dependencies, leading to over-smoothing. We thus propose a topology-aware graph transformer (TGT) that integrates global attention with local structure to update both frequency components.

Taking the low-frequency features as an example, we have

$$\tilde{\mathbf{H}}_S^l = \text{TGT}(\mathbf{H}_S^l, \mathbf{A}_S), \quad (13)$$

where  $\tilde{\mathbf{H}}_S^l \in \mathbb{R}^{N_S \times d}$  is the updated low-frequency feature matrix, and  $\mathbf{A}_S$  is the segment-level adjacency matrix.

The TGT module consists of  $N$  blocks. We set  $\mathbf{H}_S^{l,0} = \mathbf{H}_S^l$ , and for each block  $i = 0, \dots, N-1$ , we compute

$$\mathbf{Q}_{l,i} = \mathbf{H}_S^{l,i} \mathbf{W}_q^{l,i}, \quad \mathbf{K}_{l,i} = \mathbf{H}_S^{l,i} \mathbf{W}_k^{l,i}, \quad \mathbf{V}_{l,i} = \mathbf{H}_S^{l,i} \mathbf{W}_v^{l,i}, \quad (14)$$

where  $\mathbf{W}_q^{l,i}, \mathbf{W}_k^{l,i}$ , and  $\mathbf{W}_v^{l,i}$  are learnable matrices.

Next, we integrate global attention and local topology via

$$\text{ATT}_{l,i} = \alpha \cdot \text{softmax} \left( \frac{\mathbf{Q}_{l,i} \mathbf{K}_{l,i}^\top}{\sqrt{d}} \right) + (1 - \alpha) \cdot \mathbf{A}_S, \quad (15)$$

where  $\alpha$  is a learnable parameter balancing global and local information. The features are updated as

$$\tilde{\mathbf{H}}_S^{l,i} = \text{LayerNorm} \left( \text{ATT}_{l,i} \mathbf{V}_{l,i} + \mathbf{H}_S^{l,i} \right), \quad (16)$$

$$\mathbf{H}_S^{l,i+1} = \text{LayerNorm} \left( \text{FFN}(\tilde{\mathbf{H}}_S^{l,i}) + \tilde{\mathbf{H}}_S^{l,i} \right), \quad (17)$$

where  $\text{LayerNorm}(\cdot)$  denotes the layer normalization operation. After processing all blocks, we obtain  $\tilde{\mathbf{H}}_S^l = \mathbf{H}_S^{l,N}$ .

Similarly, the high-frequency features are updated through the same TGT process, yielding

$$\tilde{\mathbf{H}}_S^h = \text{TGT}(\mathbf{H}_S^h, \mathbf{A}_S). \quad (18)$$

**Reconstruction Stage.** After updating both components, we reconstruct the final segment features by combining the low-frequency and high-frequency signals:

$$\hat{\mathbf{H}}_S = \beta \cdot \tilde{\mathbf{H}}_S^l + (1 - \beta) \cdot \tilde{\mathbf{H}}_S^h, \quad (19)$$

where  $\hat{\mathbf{H}}_S \in \mathbb{R}^{N_S \times d}$  is the reconstructed segment feature matrix, and  $\beta$  is a learnable parameter that balances the contributions of low-frequency and high-frequency information.

## 4.3 Model Training

Our model involves various learnable parameters, including node representations (e.g.,  $\mathbf{H}_*$ ), assignment matrices (e.g.,  $\mathbf{A}_*$ ) within the hierarchical structure, and trainable parameters for frequency decomposition modeling. To jointly optimize these components while satisfying theoretical constraints, we design a set of tailored loss functions that jointly optimize the model.

**Alignment Loss.** Since the hierarchical structure performs node aggregation, the feature of each child node should closely resemble that of its corresponding parent, while remaining distinct from non-parent nodes. We design a contrastive loss to encourage this property:

$$\begin{aligned}\mathcal{L}_{\text{align}}^{SL} &= -\frac{1}{N_S} \sum_{i=1}^{N_S} \log \left( \frac{\exp(\text{sim}(\mathbf{h}_S^i, \mathbf{h}_L^{p(i)})/\tau)}{\sum_{j=1}^{N_L} \exp(\text{sim}(\mathbf{h}_S^i, \mathbf{h}_L^j)/\tau)} \right), \\ \mathcal{L}_{\text{align}}^{LR} &= -\frac{1}{N_L} \sum_{j=1}^{N_L} \log \left( \frac{\exp(\text{sim}(\mathbf{h}_L^j, \mathbf{h}_R^{p(j)})/\tau)}{\sum_{k=1}^{N_R} \exp(\text{sim}(\mathbf{h}_L^j, \mathbf{h}_R^k)/\tau)} \right), \\ \mathcal{L}_{\text{align}} &= \frac{1}{2} (\mathcal{L}_{\text{align}}^{SL} + \mathcal{L}_{\text{align}}^{LR}),\end{aligned}\quad (20)$$

where  $\mathbf{h}_S^i$ ,  $\mathbf{h}_L^j$ , and  $\mathbf{h}_R^k$  denote the features of the  $i$ -th segment,  $j$ -th locality, and  $k$ -th region, respectively;  $p(i)$  and  $p(j)$  represent the parent node indices;  $\tau$  is a temperature parameter controlling distribution sharpness; and  $\text{sim}(\cdot, \cdot)$  denotes the cosine similarity function.

**Reconstruction Loss.** To ensure the frequency decomposition module effectively retains key information, we require that the reconstructed segment features remain consistent with the original segment features:

$$\mathcal{L}_{\text{rec}} = \frac{1}{N_S} \sum_{i=1}^{N_S} \left\| \hat{\mathbf{h}}_S^i - \mathbf{h}_S^i \right\|_2^2, \quad (21)$$

where  $\hat{\mathbf{h}}_S^i$  and  $\mathbf{h}_S^i$  denote the reconstructed and original features of the  $i$ -th segment, respectively.

**Semantic Loss.** To ensure that the reconstructed segment features capture the semantic structure of the road network, we align their pairwise similarities with the relational structure, which is defined by the combination of static topology and dynamic origin-destination (OD) flow:

$$\mathcal{L}_{\text{sem}} = \frac{1}{N_S^2} \left\| \hat{\mathbf{H}}_S \hat{\mathbf{H}}_S^T - (\lambda \mathbf{A}_S + (1 - \lambda) \mathbf{O}_S) \right\|_F^2, \quad (22)$$

where  $\mathbf{O}_S$  denotes the normalized OD matrix derived from trajectory data, and  $\lambda$  is a balancing coefficient.

**Entropy Loss.** The preservation of low-frequency components in hierarchical modeling relies on the assignment matrices satisfying theoretical constraints. To promote this, we minimize the entropy of the assignment distributions:

$$\begin{aligned}\mathcal{L}_{\text{ent}}^{SL} &= -\frac{1}{N_S} \sum_{i=1}^{N_S} \sum_{j=1}^{N_L} \mathbf{A}_{SL}[i, j] \log(\mathbf{A}_{SL}[i, j]), \\ \mathcal{L}_{\text{ent}}^{LR} &= -\frac{1}{N_L} \sum_{j=1}^{N_L} \sum_{k=1}^{N_R} \mathbf{A}_{LR}[j, k] \log(\mathbf{A}_{LR}[j, k]), \\ \mathcal{L}_{\text{ent}} &= \frac{1}{2} (\mathcal{L}_{\text{ent}}^{SL} + \mathcal{L}_{\text{ent}}^{LR}).\end{aligned}\quad (23)$$

This encourages the assignments to be sharp, ensuring clearer hierarchical structure.

Finally, the overall loss function is formulated as

$$\mathcal{L} = \gamma_1 \mathcal{L}_{\text{align}} + \gamma_2 \mathcal{L}_{\text{rec}} + \gamma_3 \mathcal{L}_{\text{sem}} + \gamma_4 \mathcal{L}_{\text{ent}}, \quad (24)$$

where  $\gamma_1$ ,  $\gamma_2$ ,  $\gamma_3$ , and  $\gamma_4$  are hyperparameters that control the relative contributions of each term.

## 5 Experiments

In this section, we conduct experiments to demonstrate the effectiveness of our proposed model.

### 5.1 Experimental Setup

**Construction of the Datasets.** To evaluate the performance of our model, we use three real-world public datasets collected from Beijing (*BJ*), Chengdu (*CD*), and Xi'an (*XA*), which are major metropolitan areas in China. For all datasets, we collect road network information from *OpenStreetMap*. The *BJ* dataset contains taxi trajectory data sampled every minute, while the *CD* and *XA* datasets are sampled every 2–4 seconds. We perform map matching (Yang and Gidofalvi 2018) by aligning GPS points to road segments, which transforms the trajectory data into segment sequences. We then split the sequences into individual trajectories using the provided boundary indicators. For all downstream tasks, we divide each dataset into training, validation, and test sets with a ratio of 7:1:2.

**Methods to Compare.** In our experiments, we consider three types of baselines for a comprehensive comparison:

- Random Walk-based Models:* These methods learn node embeddings by generating random walk sequences on the graph and applying shallow embedding techniques. Representative baselines include *DeepWalk* (Perozzi, Al-Rfou, and Skiena 2014), *IRN2Vec* (Wang et al. 2019), and *Toast* (Chen et al. 2021).

- GNN-based Models:* These baselines leverage message passing neural networks to aggregate local or hierarchical information from neighbors. We consider *GCN* (Kipf and Welling 2017), *DGI* (Velickovic et al. 2019), *GeomGCN* (Pei et al. 2020), *DiffPool* (Ying et al. 2018), and *HRNR* (Wu et al. 2020).

- Graph Transformer-based Models:* These methods apply transformer architectures to graph data to capture both local structure and global dependencies. We include *GT* (Dwivedi and Bresson 2020), *Graphormer* (Ying et al. 2021), and *NodeFormer* (Wu et al. 2022).

**Evaluation Tasks.** We evaluate the learned road segment representations on four traffic-related tasks: (1) next location prediction, which aims to predict the next road segment based on historical trajectories; (2) label classification, where each segment is assigned a semantic label; (3) destination prediction, which infers the final destination from a partial trajectory; and (4) route planning, which reconstructs the full path between a source and destination.

**Evaluation Metrics.** We use task-specific evaluation metrics as follows. For next location and destination prediction, we treat them as ranking tasks and report top-1 and top-5 accuracy, denoted by  $ACC@1$  and  $ACC@5$ . For label classification, we report  $F1$ -score and  $AUC$ . The former balances precision and recall of binary classification, and the latter computes the area under the ROC curve. For route planning, we evaluate the predicted route  $r'$  and the actual route  $r$ , which share the same source and destination. We compute  $F1$ -score based on overlapping locations:  $P = \frac{|r \cap r'|}{|r'|}$ ,

Task	Dataset	Metric	DeepWalk	IRN2vec	Toast	GCN	DGI	Geom-GCN	DiffPool	HRNR	GT	Graph-ormer	Node-Former	HiFiNet
Next Location Prediction	BJ	ACC@1↑	0.383	0.371	0.391	0.387	0.381	0.391	0.398	0.412	0.362	0.374	0.371	<b>0.426</b>
		ACC@5↑	0.527	0.498	0.542	0.517	0.535	0.526	0.532	<u>0.556</u>	0.483	0.511	0.502	<b>0.587</b>
	CD	ACC@1↑	0.403	0.324	0.369	0.388	0.390	0.398	0.409	<u>0.420</u>	0.379	0.383	0.381	<b>0.442</b>
		ACC@5↑	0.556	0.454	0.542	0.552	0.538	0.546	0.556	<u>0.571</u>	0.506	0.516	0.515	<b>0.665</b>
	XA	ACC@1↑	0.346	0.324	0.335	0.333	0.375	0.346	0.358	<u>0.376</u>	0.316	0.310	0.318	<b>0.399</b>
		ACC@5↑	0.461	0.457	0.460	0.455	0.480	0.476	0.487	<u>0.500</u>	0.456	0.457	0.475	<b>0.546</b>
Label Classification	BJ	F1↑	0.676	0.733	0.679	0.790	0.797	0.773	0.769	0.819	0.821	0.817	<u>0.823</u>	<b>0.838</b>
		AUC↑	0.825	0.836	0.825	0.849	0.861	0.846	0.831	0.885	0.883	0.882	<u>0.887</u>	<b>0.906</b>
	CD	F1↑	0.702	0.686	0.645	0.716	0.726	0.719	0.702	0.747	0.763	0.752	<u>0.772</u>	<b>0.796</b>
		AUC↑	0.721	0.706	0.712	0.734	0.737	0.735	0.735	0.782	0.805	0.798	<u>0.835</u>	<b>0.869</b>
	XA	F1↑	0.626	0.627	0.640	0.646	0.659	0.651	0.650	0.694	<u>0.705</u>	0.688	0.701	<b>0.720</b>
		AUC↑	0.639	0.629	0.653	0.661	0.672	0.665	0.686	0.716	<u>0.736</u>	0.708	0.733	<b>0.811</b>
Destination Prediction	BJ	ACC@1↑	0.229	0.215	0.271	0.232	0.262	0.242	0.242	<u>0.277</u>	0.273	0.270	0.275	<b>0.297</b>
		ACC@5↑	0.321	0.313	0.396	0.352	0.366	0.357	0.365	<u>0.401</u>	0.396	0.392	0.399	<b>0.428</b>
	CD	ACC@1↑	0.187	0.235	0.239	0.251	0.171	0.267	0.270	0.281	0.282	0.282	<u>0.284</u>	<b>0.295</b>
		ACC@5↑	0.362	0.346	0.342	0.377	0.321	0.394	0.393	0.407	0.405	0.403	<u>0.409</u>	<b>0.426</b>
	XA	ACC@1↑	0.167	0.210	0.198	0.217	0.175	0.226	0.232	0.256	0.254	0.248	<u>0.260</u>	<b>0.291</b>
		ACC@5↑	0.289	0.305	0.306	0.334	0.315	0.352	0.353	0.375	0.364	0.361	<u>0.379</u>	<b>0.430</b>
Route Planning	BJ	F1↑	0.304	0.287	<u>0.325</u>	0.299	0.297	0.305	0.305	0.324	0.301	0.312	0.303	<b>0.339</b>
		EDT↓	8.151	8.853	7.899	8.241	8.114	8.138	8.142	<u>7.833</u>	8.232	8.083	8.108	<b>7.773</b>
	CD	F1↑	0.319	0.316	<u>0.390</u>	0.335	0.339	0.330	0.346	0.374	0.332	0.345	0.343	<b>0.498</b>
		EDT↓	8.006	8.011	<u>7.294</u>	7.869	7.818	7.737	7.655	7.350	7.773	7.673	7.715	<b>7.171</b>
	XA	F1↑	0.276	0.259	<u>0.332</u>	0.282	0.286	0.281	0.291	0.321	0.280	0.291	0.289	<b>0.377</b>
		EDT↓	8.585	9.157	8.205	8.866	8.611	8.621	8.519	<u>8.130</u>	8.732	8.529	8.659	<b>7.980</b>

Table 1: Performance comparison across four tasks on three datasets. Higher is better for all metrics except EDT. **Bold** indicates the best result, and underline indicates the second-best.

$R = \frac{|r \cap r'|}{|r|}$ , and  $F1 = \frac{2PR}{P+R}$ . Additionally, we report the edit distance (*EDT*), which measures the minimum number of edit operations required to transform  $r'$  into  $r$ .

## 5.2 Results and Analysis

Table 1 summarizes the performance of all baselines and our proposed model across four tasks and three datasets.

First, random walk-based methods perform poorly on label classification and destination prediction, but show relatively better results on next location prediction and route planning. This is likely due to their reliance on local graph topology, which aligns well with tasks that emphasize short-range transitions. Toast, an enhanced variant incorporating trajectory information, achieves strong performance on route planning. However, these methods fail to capture node-level attributes and long-range dependencies, limiting their generalization on more semantic or global tasks.

Second, GNN-based models generally outperform random walk-based ones, benefiting from their ability to integrate graph features and learn node interactions through deep message passing. Nevertheless, models such as GCN and DGI still struggle with long-range reasoning. Geom-GCN improves representation learning by introducing spatial priors, while DiffPool extends the receptive field via hierarchical pooling. Among these, HRNR achieves the strongest performance by leveraging dual semantic-guided assignment matrices to enhance hierarchical representation.

Third, graph transformer-based models demonstrate the

opposite performance pattern to random walk-based methods: they perform better on label classification and destination prediction, but worse on next location prediction and route planning. This can be attributed to their global attention mechanisms, which favor tasks requiring holistic semantic context, but may over-smooth or overlook local connectivity patterns critical for sequential prediction.

Finally, our proposed model HiFiNet consistently outperforms all baselines across tasks and datasets. By incorporating a hierarchical architecture and frequency-aware decomposition, HiFiNet effectively captures both local and global structures. The explicit separation of low- and high-frequency components preserves multi-scale semantics, enabling more expressive and generalizable representations for various downstream tasks.

## 5.3 Ablation Study

HiFiNet contains two key components: a hierarchical architecture and a frequency-decomposition module. We design five model variants to assess the contribution of each component, and conduct ablation studies on the *Beijing* dataset (similar trends are observed on others and omitted for brevity): (1) *NL*: without the locality level; (2) *NR*: without the region level; (3) *NB*: without both locality and region levels; (4) *NLF*: without the low-frequency component; and (5) *NHF*: without the high-frequency component.

As shown in Fig. 3, the performance ranking is: *NB* < *NL* < *NR* < *HiFiNet* and *NLF* < *NHF* < *HiFiNet*.

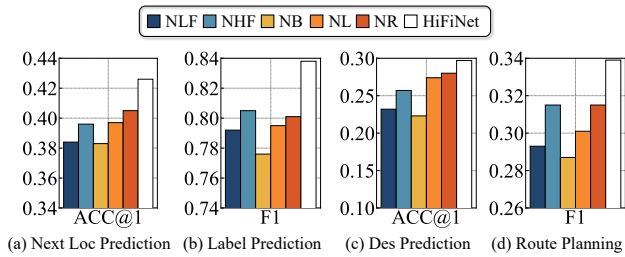


Figure 3: Ablation study of our model on Beijing dataset.

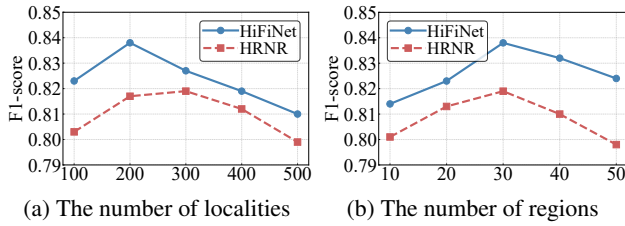


Figure 4: Parameter sensitivity of our model on Beijing dataset on label classification task.

The first group validates the role of hierarchy—removing both levels leads to the largest drop, and localities contribute more than regions, likely due to their finer granularity. The second group highlights the effectiveness of frequency decomposition. Removing either frequency component degrades performance, confirming their complementarity. Low-frequency components appear more critical, as they encode global structural semantics essential for context-aware tasks, while high-frequency features enhance fine-grained local discrimination.

#### 5.4 Parameter Sensitivity

In addition to model components, several hyperparameters require tuning in our model. We conduct sensitivity analyses on the *Beijing* dataset on the label classification task.

Specifically, we vary the number of locality nodes  $N_L$  in  $\{100, 200, 300, 400, 500\}$  and the number of region nodes  $N_R$  in  $\{10, 20, 30, 40, 50\}$ . As shown in Fig. 4, the model performs best when  $N_L = 200$  and  $N_R = 30$ . We observe that performance initially improves as the number of locality or region nodes increases, due to finer-grained representations and improved structural abstraction. However, overly large values lead to degraded performance, possibly due to increased noise or over-segmentation. Since regions are formed by aggregating finer-grained localities, it is reasonable to use more localities than regions. Overall, the model shows stable performance across a wide range of parameter settings, highlighting its robustness and applicability.

#### 5.5 Qualitative Analysis

To evaluate the representational quality of HiFiNet, we visualize the t-SNE (Maaten and Hinton 2008) projections of road segment embeddings under different frequency configurations, as shown in Fig. 5. The baseline model HRNR

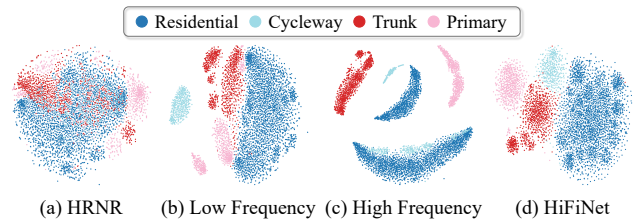


Figure 5: The t-SNE visualization of road network representations under different frequency configurations.

produces entangled embeddings with no clear separation between road types. In contrast, HiFiNet’s low-, high-, and fused-frequency representations exhibit clearer structural distinctions. For road types such as *residential* and *cycleway*, which have smoother and more regular patterns, low-frequency features form compact clusters. Their high-frequency counterparts, however, tend to fragment due to local variability and noise. Conversely, for more dynamic road types like *trunk* and *primary*, high-frequency features better capture structural complexity, while low-frequency features result in overlap or dispersion. Overall, the fused representation combines global and local information, yielding distinct and semantically coherent clusters across all road types. These results qualitatively demonstrate the advantage of frequency-aware hierarchical modeling.

## 6 Conclusion

In this paper, we introduce HiFiNet, a novel framework that unifies spatial and spectral modeling for road network representation learning. By constructing a multi-level hierarchical graph, HiFiNet enables localized frequency decomposition, capturing both coarse spatial semantics and fine-grained spectral variations. We propose a decomposition–updating–reconstruction paradigm to explicitly model low- and high-frequency components and integrate them into expressive node representations. We theoretically demonstrate that the hierarchical projection naturally acts as a spectral low-pass filter, separating frequency components and mitigating over-smoothing. Experiments on multiple real-world datasets and tasks demonstrate the robustness and generalization of HiFiNet. Beyond road networks, our unified spatial-spectral approach offers new insights for graph learning in broader spatio-temporal domains. We hope this work inspires future research into frequency-aware models that leverage structured hierarchies to better align spatial and spectral perspectives.

## Acknowledgments

Jingyuan Wang’s work was partially supported by the National Natural Science Foundation of China (No. 72222022, 72171013, 72242101) and the Fundamental Research Funds for the Central Universities (JKF-2025017226182). Leong Hou U’s work was partially supported by the Science and Technology Development Fund Macau SAR (0003/2023/RIC, 0052/2023/RIA1, 0031/2022/A, 001/2024/SKL for SKL-IOTSC).

## References

- Bastos, A.; Nadgeri, A.; Singh, K.; Kanezashi, H.; Suzumura, T.; and Mulang, I. O. 2022. How expressive are transformers in spectral domain for graphs? *arXiv preprint arXiv:2201.09332*.
- Bo, D.; Wang, X.; Shi, C.; and Shen, H. 2021. Beyond low-frequency information in graph convolutional networks. In *Proceedings of the AAAI conference on artificial intelligence*, volume 35, 3950–3957.
- Bruna, J.; Zaremba, W.; Szlam, A.; and LeCun, Y. 2013. Spectral networks and locally connected networks on graphs. *arXiv preprint arXiv:1312.6203*.
- Chen, W.; Huang, H.; Liao, S.; Gao, F.; and Biljecki, F. 2024. Global urban road network patterns: Unveiling multiscale planning paradigms of 144 cities with a novel deep learning approach. *Landscape and Urban Planning*, 241: 104901.
- Chen, Y.; Li, X.; Cong, G.; Bao, Z.; Long, C.; Liu, Y.; Chandran, A. K.; and Ellison, R. 2021. Robust Road Network Representation Learning: When Traffic Patterns Meet Traveling Semantics. In *Proceedings of the 30th ACM International Conference on Information & Knowledge Management*, 211–220.
- Defferrard, M.; Bresson, X.; and Vandergheynst, P. 2016. Convolutional neural networks on graphs with fast localized spectral filtering. *Advances in neural information processing systems*, 29.
- Dwivedi, V. P.; and Bresson, X. 2020. A generalization of transformer networks to graphs. *arXiv preprint arXiv:2012.09699*.
- Grover, A.; and Leskovec, J. 2016. node2vec: Scalable feature learning for networks. In *Proceedings of the 22nd ACM SIGKDD international conference on Knowledge discovery and data mining*, 855–864.
- Guo, S.; Chen, C.; Wang, J.; Ding, Y.; Liu, Y.; Xu, K.; Yu, Z.; and Zhang, D. 2020. A force-directed approach to seeking route recommendation in ride-on-demand service using multi-source urban data. *IEEE Transactions on Mobile Computing*, 21(6): 1909–1926.
- Hamilton, W.; Ying, Z.; and Leskovec, J. 2017. Inductive representation learning on large graphs. In *Advances in neural information processing systems*, 1024–1034.
- Han, C.; Wang, J.; Wang, Y.; Yu, X.; Lin, H.; Li, C.; and Wu, J. 2025. Bridging traffic state and trajectory for dynamic road network and trajectory representation learning. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 39, 11763–11771.
- Ji, J.; Wang, J.; Huang, C.; Wu, J.; Xu, B.; Wu, Z.; Zhang, J.; and Zheng, Y. 2023. Spatio-temporal self-supervised learning for traffic flow prediction. In *Proceedings of the AAAI conference on artificial intelligence*, volume 37, 4356–4364.
- Ji, J.; Wang, J.; Jiang, Z.; Jiang, J.; and Zhang, H. 2022. STDEN: Towards physics-guided neural networks for traffic flow prediction. In *Proceedings of the AAAI conference on artificial intelligence*, volume 36, 4048–4056.
- Ji, J.; Wang, J.; Jiang, Z.; Ma, J.; and Zhang, H. 2020. Interpretable spatiotemporal deep learning model for traffic flow prediction based on potential energy fields. In *2020 IEEE international conference on data mining (ICDM)*, 1076–1081. IEEE.
- Ji, J.; Zhang, W.; Wang, J.; and Huang, C. 2025. Seeing the unseen: Learning basis confounder representations for robust traffic prediction. *arXiv Preprint*.
- Jiang, J.; Han, C.; Zhao, W. X.; and Wang, J. 2023a. Pdfformer: Propagation delay-aware dynamic long-range transformer for traffic flow prediction. In *Proceedings of the AAAI conference on artificial intelligence*, volume 37, 4365–4373.
- Jiang, J.; Pan, D.; Ren, H.; Jiang, X.; Li, C.; and Wang, J. 2023b. Self-supervised trajectory representation learning with temporal regularities and travel semantics. In *2023 IEEE 39th international conference on data engineering (ICDE)*, 843–855. IEEE.
- Kipf, T. N.; and Welling, M. 2017. Semi-supervised classification with graph convolutional networks. *ICLR*.
- Levie, R.; Monti, F.; Bresson, X.; and Bronstein, M. M. 2018. Cayleynets: Graph convolutional neural networks with complex rational spectral filters. *IEEE Transactions on Signal Processing*, 67(1): 97–109.
- Ma, J.; Wang, J.; Zhao, W. X.; Liu, G.; and Wen, X. 2025. Spatio-Temporal Data Enhanced Vision-Language Model for Traffic Scene Understanding. *arXiv preprint arXiv:2511.08978*.
- Maaten, L. v. d.; and Hinton, G. 2008. Visualizing data using t-SNE. *Journal of machine learning research*, 9(Nov): 2579–2605.
- Mao, Z.; Li, Z.; Li, D.; Bai, L.; and Zhao, R. 2022. Jointly contrastive representation learning on road network and trajectory. In *Proceedings of the 31st ACM International Conference on Information & Knowledge Management*, 1501–1510.
- Nt, H.; and Maehara, T. 2019. Revisiting graph neural networks: All we have is low-pass filters. *arXiv preprint arXiv:1905.09550*.
- Pei, H.; Wei, B.; Chang, C.-C.; Lei, Y.; and Yang, B. 2020. Geom-GCN: Geometric Graph Convolutional Networks. In *ICLR*, 4800–4810.
- Perozzi, B.; Al-Rfou, R.; and Skiena, S. 2014. Deepwalk: Online learning of social representations. In *Proceedings of the 20th ACM SIGKDD international conference on Knowledge discovery and data mining*, 701–710.
- Velickovic, P.; Cucurull, G.; Casanova, A.; Romero, A.; Lio, P.; and Bengio, Y. 2017. Graph attention networks. *ICLR*, 1(2).
- Velickovic, P.; Fedus, W.; Hamilton, W. L.; Liò, P.; Bengio, Y.; and Hjelm, R. D. 2019. Deep graph infomax. *ICLR (poster)*, 2(3): 4.
- Wang, J.; Jiang, J.; Jiang, W.; Li, C.; and Zhao, W. X. 2021a. Libcity: An open library for traffic prediction. In *Proceedings of the 29th international conference on advances in geographic information systems*, 145–148.
- Wang, J.; Lin, X.; Zuo, Y.; and Wu, J. 2021b. DGeye: Probabilistic risk perception and prediction for urban dangerous

goods management. *ACM Transactions on Information Systems (TOIS)*, 39(3): 1–30.

Wang, M.-x.; Lee, W.-C.; Fu, T.-y.; and Yu, G. 2019. Learning Embeddings of Intersections on Road Networks. In *SIGSPATIAL*, 309–318.

Wang, M.-X.; Lee, W.-C.; Fu, T.-Y.; and Yu, G. 2020. On representation learning for road networks. *ACM Transactions on Intelligent Systems and Technology (TIST)*, 12(1): 1–27.

Wu, F.; Souza, A.; Zhang, T.; Fifty, C.; Yu, T.; and Weinberger, K. 2019a. Simplifying graph convolutional networks. In *International conference on machine learning*, 6861–6871. Pmlr.

Wu, N.; Wang, J.; Zhao, W. X.; and Jin, Y. 2019b. Learning to Effectively Estimate the Travel Time for Fastest Route Recommendation. In *CIKM*, 1923–1932.

Wu, N.; Zhao, X. W.; Wang, J.; and Pan, D. 2020. Learning effective road network representation with hierarchical graph neural networks. In *Proceedings of the 26th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*, 6–14.

Wu, Q.; Zhao, W.; Li, Z.; Wipf, D. P.; and Yan, J. 2022. Nodeformer: A scalable graph structure learning transformer for node classification. *Advances in Neural Information Processing Systems*, 35: 27387–27401.

Yang, C.; and Gidofalvi, G. 2018. Fast map matching, an algorithm integrating hidden Markov model with precomputation. *IJGIS*, 32(3): 547–570.

Ying, C.; Cai, T.; Luo, S.; Zheng, S.; Ke, G.; He, D.; Shen, Y.; and Liu, T.-Y. 2021. Do transformers really perform badly for graph representation? *Advances in neural information processing systems*, 34: 28877–28888.

Ying, Z.; You, J.; Morris, C.; Ren, X.; Hamilton, W.; and Leskovec, J. 2018. Hierarchical graph representation learning with differentiable pooling. In *Advances in neural information processing systems*, 4800–4810.

Yuan, J.; Zheng, Y.; and Xie, X. 2012. Discovering regions of different functions in a city using human mobility and POIs. In *Proceedings of the 18th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, KDD ’12, 186–194. New York, NY, USA: Association for Computing Machinery. ISBN 9781450314626.

Zhang, L.; and Long, C. 2023. Road network representation learning: A dual graph-based approach. *ACM Transactions on Knowledge Discovery from Data*, 17(9): 1–25.

Zhang, W.; Wang, J.; Yang, Y.; et al. 2024. VecCity: A taxonomy-guided library for map entity representation learning. *arXiv preprint arXiv:2411.00874*.

Zhou, H.; Huang, W.; Chen, Y.; He, T.; Cong, G.; and Ong, Y. S. 2024. Road network representation learning with the third law of geography. *Advances in Neural Information Processing Systems*, 37: 11789–11813.

Zhu, M.; Wang, X.; Shi, C.; Ji, H.; and Cui, P. 2021. Interpreting and unifying graph neural networks with an optimization framework. In *Proceedings of the web conference 2021*, 1215–1226.