Inter-Client Dependency Recovery with Hidden Global Components for Federated Traffic Prediction

Hang Zhou[1], Wentao Yu[1], Yang Wei[1], Guangyu Li[1], Sha Xu[2*], Chen Gong[3*]

[1]School of Computer Science and Engineering, Nanjing University of Science and Technology, China
[2]School of Integrated Circuits, Guangdong University of Technology, China
[3]School of Automation and Intelligent Sensing, Shanghai Jiao Tong University, China
{hangzhou, wentao.yu, csywei, guangyu.li2017}@njust.edu.cn, sally.xu@gdut.edu.cn, chen.gong@sjtu.edu.cn

Abstract

Traffic prediction plays an important role in urban management. However, existing methods rely on centralized traffic data, which may raise privacy concerns. Federated traffic prediction offers a promising solution for clients (*e.g.*, traffic management administrations) in different regions to collaboratively train models in a distributed manner without exposing private data. Nonetheless, data isolation inherently breaks the correlations between nodes (*i.e.*, traffic sensors collecting data) from different regions, which leads to the missing inter-client dependency. Consequently, current works either fail to capture the missing inter-client dependency or compromise data privacy to recover the inter-client dependency. To address this issue, we propose a novel **Fed**erated method which recovers the inter-client dependency with **HI**dden global compone**NT**s (FedHINT). We find that the traffic data from different local regions actually contain hidden global components that reflect cross-regional traffic changes. Therefore, our FedHINT aims to extract hidden global components from each client to generate proxy nodes that represent global information, which are then utilized to recover the inter-client dependency. To be specific, we employ an attention module, which is guided by the shared global queries to capture hidden global components from local traffic data, to generate proxy nodes. Subsequently, our FedHINT adaptively learns the correlations between proxy nodes and local nodes through a global encoder. During this process, the global information in proxy nodes compensate for the loss of information from cross-regional nodes, which thereby recovers the missing inter-client dependency. Intensive experiments on multiple datasets demonstrate that our FedHINT significantly outperforms the state-of-the-art methods, with an average decrease of 3.73 and 4.81 on MAE and RMSE, respectively.

Code — https://github.com/lichuan210/FedHINT

Introduction

Traffic prediction aims to forecast future traffic conditions by capturing spatial-temporal patterns from historical traffic data (Liu, Zhang, and Liu 2023). Accurate traffic prediction is crucial for urban management, as it facilitates traffic control (Guo et al. 2019), alleviates congestion (Wu et al. 2022), and supports intelligent routing (Huang et al. 2022).

Extensive deep learning-based approaches have been proposed for traffic prediction. These approaches typically employ time series models to capture temporal patterns in traffic flow and leverage Graph Neural Networks (GNNs) to model spatial patterns within traffic networks (Yu, Yin, and Zhu 2018; Geng et al. 2019; Liang et al. 2024). However, these methods require uploading traffic data from different regions to a central server for model training. In practice, traffic data are often owned by different region-specific traffic administration departments and may contain sensitive information, such as travel trajectories and vehicle locations (Li et al. 2018). Due to access restrictions and privacy issues, sharing data from different regions may be prohibited, which significantly limits the practical applications of centralized methods.

Federated Learning (FL) offers a promising solution to train models with decentralized data (McMahan et al. 2017). In this paradigm, each client independently trains a local model based on its private dataset and uploads the model parameters to a central server for federated aggregation (Yue et al. 2024; Meng et al. 2024; Huang et al. 2025; Yu et al. 2025b; Fu et al. 2025b). In federated traffic prediction, each region can be viewed as a subgraph of the global traffic network. Clients (*i.e.*, traffic administration departments of different regions in our problem) train prediction models with data collected by local traffic sensors (*i.e.*, nodes within each subgraph) and upload model parameters to a central server for collaborative learning. So far, various studies have employed the FL paradigm for traffic prediction to protect data privacy (Zhang et al. 2021a; Wang et al. 2022; Liu et al. 2023). Distinct from images or texts, where data are isolated and independent, traffic data are collected by traffic sensors within road networks (Zhang et al. 2021b). These correlations among sensors within the road network carry crucial information for changes in traffic volume. For example, correlations between regions reflect the number of vehicles traveling from one area to another. However, due to data access constraints, previous federated traffic prediction methods only capture the correlations within local traffic networks and may fail to model inter-client dependency (*i.e.*, missing correlations among cross-regional nodes), which re-

*Corresponding authors.

sults in suboptimal performance.

To recover the missing inter-client dependency, several methods propose to upload node features extracted by local models and learn complete correlations among traffic sensors on server side (Meng, Rambhatla, and Liu 2021; Yang et al. 2024). Although these approaches model the missing dependency, they require transmitting not only model parameters but also the data features of clients to server. Therefore, these additional requirements inevitably increase communication overhead and significantly raise the risk of privacy leakage. In practice, traffic data from each region contain temporal characteristics that are highly correlated with global traffic changes (Zhou et al. 2025). For example, the morning and evening commuting peaks in one region often occur simultaneously with those in other regions. Such characteristics reflect cross-regional traffic changes and can be regarded as hidden global components within each region. This insight motivates us to explore hidden global components within each client to recover the missing inter-client dependency, without exposing data features.

Therefore, in this paper, we propose a novel **Fed**erated method that recovers inter-client dependency with **HI**dden global compone**NT**s (FedHINT). Our goal is to extract hidden global components from local data to generate proxy nodes locally that represent global information, which are then utilized to recover the inter-client dependency. Specifically, we first introduce an attention module equipped with precisely designed time-shifted filters to extract temporal characteristics from local traffic data. By leveraging these temporal characteristics, the attention module employs shared global queries to identify hidden global components within local traffic data. These extracted hidden global components are then integrated to generate proxy nodes, which serve as representatives of global information. Subsequently, our FedHINT leverages the global information in proxy nodes to recover inter-client dependency by learning the correlations between proxy nodes and local nodes. Considering that the global information in proxy nodes may conflict with region-specific information, we learn different types of correlations between nodes with two independent encoders, namely global encoder and local encoder. Global encoder adaptively learns the correlations between local and proxy nodes, where the global information in proxy nodes compensate for the loss of information from cross-regional nodes so as to recover inter-client dependency. Meanwhile, local encoder focuses on mining correlations within local nodes. Finally, the model parameters for extracting hidden global components and recovering inter-client dependency are shared across clients to enhance the capability of the model to leverage global information. Thanks to effectively extracting and leveraging hidden global components, our method can successfully recover missing inter-client dependency and improve prediction performance. The contributions of this paper are summarized as follows:

- To recover inter-client dependency without compromising data privacy, we propose a novel federated traffic prediction method that leverages hidden global components within each client to recover the missing dependency.
- We propose an attention module to extract hidden global components effectively. Furthermore, we employ a global encoder to compensate for the loss of information from cross-regional nodes, which thereby recovers the missing inter-client dependency.
- Intensive experimental results demonstrate the effectiveness of the proposed FedHINT, where our method outperforms existing works with an average decrease of 3.73 and 4.81 on MAE and RMSE, respectively.

Related Work

In this section, we review the relevant works, including conventional traffic prediction and federated traffic prediction.

Conventional Traffic Prediction

Many deep learning methods perform traffic prediction by modeling both temporal patterns in traffic flow and spatial patterns within traffic networks. Models such as Recurrent Neural Network (RNN) (Shi et al. 2015), Temporal Convolutional Network (TCN) (Sutskever, Vinyals, and Le 2014), and Transformer (Jiang et al. 2023a) have been widely employed due to their ability to capture dynamic temporal patterns in time series. In addition, some studies have explored temporal information from a frequency-domain perspective. For instance, FEDFormer (Zhou et al. 2022) employs the attention mechanism in the frequency domain to capture important patterns in time series. Furthermore, TimesNet (Wu et al. 2023) leverages frequency information to explore the multi-periodicity of time series. Meanwhile, prevailing methods employ GNNs, such as Graph Convolutional Network (GCN) (Huang et al. 2021) and Graph Attention Network (GAT) (Zheng et al. 2020), to capture spatial patterns. To model dynamic spatial patterns, Graph WaveNet (Wu et al. 2019) and AGCRN (Bai et al. 2020) adaptively construct adjacency matrices instead of relying on predefined graphs. Additionally, MegaCRN (Jiang et al. 2023b) introduces a meta-graph constructed via a hyper-network. However, most existing works still rely on the centralized training paradigm. In practice, traffic data from different regions are not allowed to distribute due to privacy issues, which hinders the practical application of these methods.

Federated Traffic Prediction

Due to the advantages of privacy protection, various FL methods have been developed for traffic prediction in a distributed manner. For example, FedGRU (Liu et al. 2020) employs a clustering-based approach to aggregate local Gated Recurrent Units (GRU) for traffic forecasting. In addition, CTFL (Zhang et al. 2022) employs a divide-and-conquer strategy to reduce the communication overhead of federated traffic prediction. To address the heterogeneity of traffic data, FedTPS (Zhou et al. 2024) proposes sharing common traffic patterns while maintaining region-specific characteristics in a personalized FL manner. However, these methods fail to model inter-client dependency, as the data from other clients cannot be accessed. To tackle this issue, CNFGNN (Meng, Rambhatla, and Liu 2021) leverages a predefined graph on server to learn dependency across clients.

Analogously, FedGTP (Yang et al. 2024) adaptively learns inter-client dependency utilizing intermediate features aggregated on server. Nevertheless, except for model parameters, these methods require uploading intermediate data features, which introduces additional burdens on server and potential privacy risks. In addition, FedGCN (Hu et al. 2024) attempts to estimate the missing nodes for each client. However, it is difficult to accurately infer node information from other clients due to the non-Independent and Identically Distributed (non-IID) nature of data (Fu et al. 2025a; Yu et al. 2025a). Different from prior methods, our FedHINT leverages hidden global components within each client to recover the missing inter-client dependency without exposing data features.

Problem Definition and Preliminary

In this section, we formally define the setting for the investigated problem of federated traffic prediction and introduce the Discrete Fourier Transform (DFT) employed in our method.

Problem Definition

In federated traffic prediction, the global traffic network can be defined as a graph $\mathcal{G} = \langle \mathcal{V}, \mathcal{E} \rangle$, where $\mathcal{V}$ denotes the set of nodes and $\mathcal{E}$ denotes the set of edges. Each node corresponds to a sensor that records traffic data, and each edge depicts the relation between nodes, which can be represented by an adjacency matrix $\mathbf{A} \in \mathbb{R}^{|\mathcal{V}| \times |\mathcal{V}|}$. The notation $|\cdot|$ denotes the cardinality of a set. For given M clients possessing traffic data from M different regions, the m-th (m takes a value from $\{1, 2, \cdots, M\}$) client maintains a local subgraph $\mathcal{G}_m = \langle \mathcal{V}_m, \mathcal{E}_m \rangle$ and its corresponding dataset $\mathcal{D}_m = \{\mathbf{x}_t\}_{t=1}^{T}$, where $\mathbf{x}_t \in \mathbb{R}^{|\mathcal{V}_m|}$ represents the observed traffic data of local traffic network $\mathcal{G}_m$ at time stamp t, and T refers to the total length of recorded time series. Each client trains a local model $f_{\mathbf{W}_m}$ with learnable parameters $\mathbf{W}_m$ to predict traffic conditions for future T_2 stamps based on historical T_1 stamps, which can be represented as

$$\mathbf{x}_{t-T_1+1}, \cdots, \mathbf{x}_t \xrightarrow{f_{\mathbf{W}_m}} \mathbf{x}_{t+1}, \cdots, \mathbf{x}_{t+T_2}. \tag{1}$$

In federated traffic prediction, the server aggregates the locally trained model parameters from each client and redistributes the aggregated parameters to the clients for subsequent training. This iterative process allows for collaborative learning across clients, where optimal model parameters are obtained via the objective below:

$$\underset{\mathbf{W}_1, \cdots, \mathbf{W}_M}{\operatorname{argmin}} \sum_{m=1}^{M} \frac{|\mathcal{V}_m|}{|\mathcal{V}|} \mathcal{L}\left(\mathbf{W}_m, \mathcal{D}_m\right), \tag{2}$$

where $\mathcal{L}(\cdot, \cdot)$ represents the loss function. Note that all the mathematical notations related to the m-th client should be accompanied by the subscript m. However, for simplicity, we omit the subscript m in the following if no confusion is incurred.

Discrete Fourier Transform

The DFT plays a crucial role in digital signal processing and has been widely adopted to reveal temporal characteristics in time series (Yi et al. 2024; Piao et al. 2024). Given an input sequence $\mathbf{z} = [z_1, z_2, \cdots, z_T]^\top \in \mathbb{R}^T$, the DFT converts it into frequency-domain signal $\mathbf{z}' = [z'_1, z'_2, \cdots, z'_T]^\top \in \mathbb{C}^T$ as

$$z'_k = \sum_{t=1}^{T} z_t e^{-\frac{2\pi \mathrm{i}}{T} tk}, \quad k = 1, \cdots, T. \tag{3}$$

Here i is the imaginary unit and z'_k represents the spectral component of the sequence $\mathbf{z}$ at the frequency $\omega_k = \frac{2\pi k}{T}$. The Inverse DFT (IDFT) can be represented as

$$z_t = \frac{1}{T} \sum_{k=1}^{T} z'_k e^{\frac{2\pi \mathrm{i}}{T} tk}, \tag{4}$$

which reconstructs the time series data from the spectral components z'_k.

Methodology

This section details our proposed FedHINT, the framework of which is illustrated in Fig. 1. In the following, the critical steps will be presented by introducing the extraction of hidden global components, explaining the recovery of inter-client dependency, and describing the process of federated aggregation.

Extraction of Hidden Global Components

In federated traffic prediction, since clients cannot access cross-regional nodes that are correlated with local nodes, we propose to extract hidden global components from local data to generate proxy nodes that represent global information as substitutes for the missing cross-regional nodes. To achieve this goal, we employ an attention module, which is guided by shared global queries $\mathbf{Q} \in \mathbb{R}^{N \times d_{\text{att}}}$ to capture hidden global components from local traffic data, where N denotes the number of queries and d_{att} is the dimension of attention space. Given historical traffic data $\mathbf{X}_t = [\mathbf{x}_{t-T_1+1}, \mathbf{x}_{t-T_1+2}, \cdots, \mathbf{x}_t] \in \mathbb{R}^{|\mathcal{V}_m| \times T_1}$ over the past T_1 stamps up to time t, the keys $\mathbf{K}_t \in \mathbb{R}^{|\mathcal{V}_m| \times d_{\text{att}}}$ and values $\mathbf{V}_t \in \mathbb{R}^{|\mathcal{V}_m| \times d_{\text{att}}}$ can be obtained as

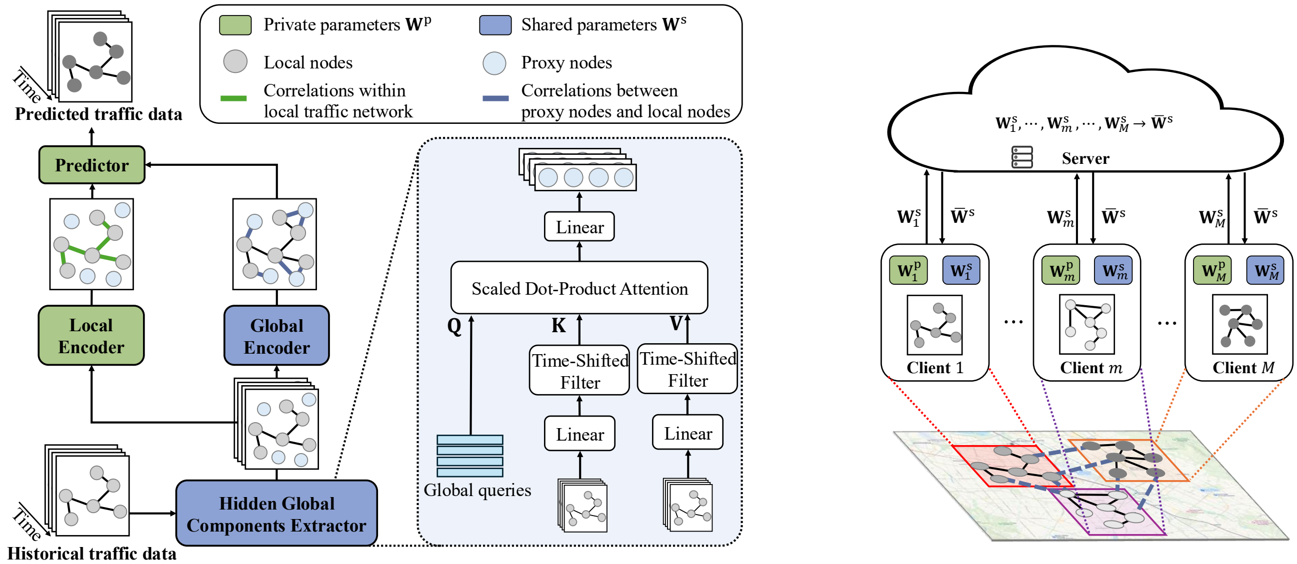
$$\mathbf{K}_t = \mathbf{X}_t \mathbf{W}_{\mathrm{K}}, \quad \mathbf{V}_t = \mathbf{X}_t \mathbf{W}_{\mathrm{V}}, \tag{5}$$

where $\mathbf{W}_{\mathrm{K}} \in \mathbb{R}^{T_1 \times d_{\text{att}}}$ and $\mathbf{W}_{\mathrm{V}} \in \mathbb{R}^{T_1 \times d_{\text{att}}}$ are learnable parameters.

To effectively identify hidden global components from local data, we propose to extract the temporal characteristics for keys and values from a frequency-domain perspective before applying the attention operation. Considering that temporal characteristics vary across different time segments, we design the time-shifted filters $\mathbf{W}_{\text{fil}} \in \mathbb{C}^{L \times d_{\text{att}}}$, where L is the number of learnable frequency filters. Taking the keys $\mathbf{K}_t$ as an example, the filter process can be denoted as

$$\tilde{\mathbf{K}}_t = \mathcal{F}^{-1}(\mathcal{F}(\mathbf{K}_t) \odot \mathrm{Expand}(\mathbf{W}_{\text{fil}}^{j}, |\mathcal{V}_m|)), \tag{6}$$

where $\mathbf{W}_{\text{fil}}^{j} \in \mathbb{C}^{d_{\text{att}}}$ represents the j-th row of $\mathbf{W}_{\text{fil}}$, the index j is computed as $t \bmod L$. In Eq. (6),



(a) The local training phase

(b) The federated aggregation phase

Figure 1: The framework of our proposed FedHINT. (a) In the training phase within each client, an attention module guided by global queries is employed to extract hidden global components from local traffic data and generate proxy nodes with global information. Subsequently, the model employs two independent encoders to learn different types of correlations between nodes, where global encoder recovers inter-client dependency by learning the correlations between proxy and local nodes. Finally, the predictor produces the predictions by leveraging the outputs from both encoders. (b) In the aggregation phase, clients from different regions upload shared parameters for aggregation to enhance the capability to leverage global information.

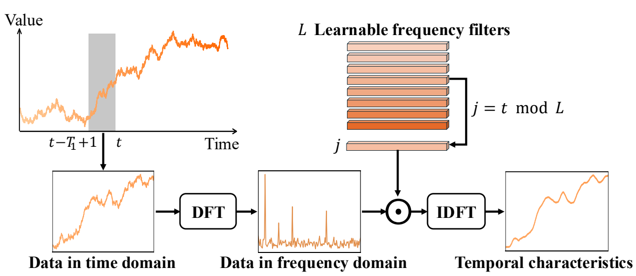


Figure 2: Illustration of time-shifted filters. The input data are first transformed into the frequency domain via DFT. The corresponding filter is then selected based on the time segment of the data to extract temporal characteristics.

$\mathrm{Expand}(\mathbf{W}_{\mathrm{fil}}^{j}, |\mathcal{V}_m|)$ extends the dimension of $\mathbf{W}_{\mathrm{fil}}^{j}$ to $\mathbb{C}^{|\mathcal{V}_m| \times d_{\mathrm{att}}}$ and $\odot$ represents the Hadamard product. The operations $\mathcal{F}$ and $\mathcal{F}^{-1}$ denote DFT and IDFT along the dimension of each row, respectively. The process of the time-shifted filters is shown in Fig. 2. In this way, the time-shifted filters can effectively capture temporal characteristics by selecting distinct filters for different time segments. Similarly, we can extract the temporal characteristics for values $\mathbf{V}_t$ and obtain $\tilde{\mathbf{V}}_t$ through the time-shifted filters. By leveraging these temporal characteristics, global queries guide the attention module to focus on hidden global components within the temporal characteristics and integrate them to generate proxy nodes with global information, which can be represented as

$$\mathbf{P}_t = \mathrm{softmax}\left(\frac{\mathbf{Q}\tilde{\mathbf{K}}_t^{\top}}{\sqrt{d_{\mathrm{att}}}}\right)\tilde{\mathbf{V}}_t, \tag{7}$$

where $\mathbf{P}_t \in \mathbb{R}^{N \times d_{\mathrm{att}}}$ is the output of the attention module. Finally, the proxy nodes are projected to the time domain through a linear transformation, which can be represented as

$$\mathbf{X}_t^{\mathrm{proxy}} = \mathbf{P}_t \mathbf{W}_{\mathrm{p}}, \tag{8}$$

where $\mathbf{W}_{\mathrm{p}} \in \mathbb{R}^{d_{\mathrm{att}} \times T_1}$ is the learnable weight matrix. By vertically concatenating local nodes $\mathbf{X}_t$ and proxy nodes $\mathbf{X}_t^{\mathrm{proxy}} \in \mathbb{R}^{N \times T_1}$, we obtain the traffic data for all nodes, denoted as $\mathbf{X}_t^{\mathrm{all}} = [\mathbf{x}_{t-T_1+1}^{\mathrm{all}}, \mathbf{x}_{t-T_1+2}^{\mathrm{all}}, \cdots, \mathbf{x}_t^{\mathrm{all}}] \in \mathbb{R}^{(|\mathcal{V}_m|+N) \times T_1}$. Furthermore, to enhance the diversity of extracted hidden global components, we introduce a regularization loss to enforce orthogonality among the global queries, which can be defined as

$$\mathcal{L}_{\mathrm{div}} = \frac{1}{N(N-1)} \sum_{i=1}^{N} \sum_{j=i+1}^{N} |\mathbf{q}_i^{\top} \mathbf{q}_j|, \tag{9}$$

where $\mathbf{q}_i \in \mathbb{R}^{d_{\mathrm{att}}}$ is the i-th query of global queries $\mathbf{Q}$.

Recovery of Inter-Client Dependency

Based on the generated proxy nodes, our FedHINT is designed to learn the correlations between proxy nodes and local nodes. During this process, the global information in proxy nodes compensate for the loss of information from cross-regional nodes and recover inter-client dependency.

Here, we adopt AGCRN (Bai et al. 2020) as the backbone of our spatial-temporal encoders, where the adaptive adjacency matrix is employed to learn the correlations between nodes. The adaptive adjacency matrix is defined as

$$\tilde{\mathbf{A}} = \mathbf{I} + \sigma(\mathbf{E}\mathbf{E}^\top), \tag{10}$$

where $\mathbf{I} \in \mathbb{R}^{(|\mathcal{V}_m|+N)\times(|\mathcal{V}_m|+N)}$ is the identity matrix and $\sigma(\cdot)$ represents the activation function. In Eq. (10), $\mathbf{E} \in \mathbb{R}^{(|\mathcal{V}_m|+N)\times e}$ is the learnable embedding matrix, where e denotes the embedding dimension of each node. The process of AGCRN can be formulated as

$$\begin{aligned}\mathbf{U}_t &= \sigma(\tilde{\mathbf{A}}[\mathbf{x}_t^{\mathrm{all}}||\mathbf{H}_{t-1}]\mathbf{E}\mathbf{W}_{\mathrm{U}} + \mathbf{E}\mathbf{B}_{\mathrm{U}}),\\ \mathbf{R}_t &= \sigma(\tilde{\mathbf{A}}[\mathbf{x}_t^{\mathrm{all}}||\mathbf{H}_{t-1}]\mathbf{E}\mathbf{W}_{\mathrm{R}} + \mathbf{E}\mathbf{B}_{\mathrm{R}}),\\ \mathbf{C}_t &= \tanh(\tilde{\mathbf{A}}[\mathbf{x}_t^{\mathrm{all}}||\mathbf{R}_t \odot \mathbf{H}_{t-1}]\mathbf{E}\mathbf{W}_{\mathrm{C}} + \mathbf{E}\mathbf{B}_{\mathrm{C}}),\\ \mathbf{H}_t &= \mathbf{U}_t \odot \mathbf{H}_{t-1} + (1-\mathbf{U}_t)\odot \mathbf{C}_t,\end{aligned} \tag{11}$$

where $\mathbf{H}_t \in \mathbb{R}^{(|\mathcal{V}_m|+N)\times h}$ is the output h-dimensional hidden state at time stamp t and $\mathbf{U}_t$, $\mathbf{R}_t$, and $\mathbf{C}_t$ represent the update gate, reset gate, and candidate state, respectively. In Eq. (11), $[\mathbf{x}_t^{\mathrm{all}}||\cdot]$ denotes the column-wise concatenation, where $\mathbf{x}_t^{\mathrm{all}}$ is first reshaped to $\mathbb{R}^{(|\mathcal{V}_m|+N)\times 1}$ before concatenation. In addition, $\mathbf{W}_{\mathrm{U}} \in \mathbb{R}^{e\times(1+h)\times h}$, $\mathbf{W}_{\mathrm{R}} \in \mathbb{R}^{e\times(1+h)\times h}$, $\mathbf{W}_{\mathrm{C}} \in \mathbb{R}^{e\times(1+h)\times h}$, $\mathbf{B}_{\mathrm{U}} \in \mathbb{R}^{e\times h}$, $\mathbf{B}_{\mathrm{R}} \in \mathbb{R}^{e\times h}$, and $\mathbf{B}_{\mathrm{C}} \in \mathbb{R}^{e\times h}$ are learnable parameters. The encoder processes the input data $\mathbf{X}_t^{\mathrm{all}}$ sequentially and produces the hidden state $\mathbf{H}_t = \mathrm{Encoder}(\mathbf{X}_t^{\mathrm{all}}, \tilde{\mathbf{A}})$ at the last stamp.

For the correlations between nodes captured in Eq. (11), each node interacts with all other nodes, including both local and proxy nodes. However, global information conveyed by correlations between proxy and local nodes may conflict with region-specific information within the local road network. How to further convey global information apart from region-specific information remains a challenge for the single encoder. Therefore, we propose to learn different types of correlations with two independent encoders, namely global encoder and local encoder. For global encoder, we define a global mask matrix $\mathbf{M}^{\mathrm{global}} \in \{0,1\}^{(|\mathcal{V}_m|+N)\times(|\mathcal{V}_m|+N)}$, where the (i,j)-th element in $\mathbf{M}^{\mathrm{global}}$ represents the mask for correlation between the i-th and the j-th nodes. The mask values for the correlations between local nodes and proxy nodes are set to 1 and 0 otherwise. Similarly, for local encoder, a local mask matrix $\mathbf{M}^{\mathrm{local}}$ is defined by setting the mask value for the correlation to 1 if both nodes are local nodes and 0 otherwise. In this way, global encoder and local encoder focus on two different types of correlations, which mitigate conflicts between global and region-specific information. The processes of two encoders can be formulated as

$$\begin{aligned}\mathbf{H}_t^{\mathrm{global}} &= \mathrm{Encoder}^{\mathrm{global}}(\mathbf{X}_t^{\mathrm{all}}, \tilde{\mathbf{A}}\odot\mathbf{M}^{\mathrm{global}}),\\ \mathbf{H}_t^{\mathrm{local}} &= \mathrm{Encoder}^{\mathrm{local}}(\mathbf{X}_t^{\mathrm{all}}, \tilde{\mathbf{A}}\odot\mathbf{M}^{\mathrm{local}}).\end{aligned} \tag{12}$$

Finally, we concatenate $\mathbf{H}_t^{\mathrm{global}}$ with $\mathbf{H}_t^{\mathrm{local}}$ and feed them into a linear predictor to obtain the prediction for all nodes $\hat{\mathbf{X}}_t^{\mathrm{all}} \in \mathbb{R}^{(|\mathcal{V}_m|+N)\times T_2}$, which can be denoted as

$$\hat{\mathbf{X}}_t^{\mathrm{all}} = [\mathbf{H}_t^{\mathrm{global}}||\mathbf{H}_t^{\mathrm{local}}]\mathbf{W}_{\mathrm{o}}, \tag{13}$$

where $\mathbf{W}_{\mathrm{o}} \in \mathbb{R}^{2h\times T_2}$ is a learnable weight matrix. By selecting the first $|\mathcal{V}_m|$ rows of $\hat{\mathbf{X}}_t^{\mathrm{all}}$, we obtain the prediction for the local traffic nodes $\hat{\mathbf{X}}_t = [\hat{\mathbf{x}}_{t+1}, \hat{\mathbf{x}}_{t+2}, \cdots, \hat{\mathbf{x}}_{t+T_2}] \in \mathbb{R}^{|\mathcal{V}_m|\times T_2}$. By following previous work (Bai et al. 2020), we adopt ℓ_1 loss function to optimize the training process. As a result, the final optimization objective of our method is represented as

$$\mathcal{L} = \sum_{t=1}^{T} |\hat{\mathbf{x}}_t - \mathbf{x}_t| + \lambda\mathcal{L}_{\mathrm{div}}, \tag{14}$$

where λ is a non-negative hyperparameter assigned to $\mathcal{L}_{\mathrm{div}}$.

Federated Aggregation

Our FedHINT aims to share the parameters for extracting hidden global components and recovering inter-client dependency, which enhances the capability of the model to leverage global information. However, due to the non-IID nature of traffic data across different regions, directly sharing all parameters may lead to suboptimal performance. To address this issue, our approach performs model aggregation in a personalized manner, where parameters corresponding to global information are shared across clients, while parameters associated with region-specific information are retained locally. The aggregation process is illustrated in Fig. 1(b), where each client trains a local model based on the traffic data collected from its respective region. For the m-th client, the model parameters can be divided into two parts, namely shared parameters $\mathbf{W}_m^{\mathrm{s}}$ (*i.e.*, the parameters of hidden global components extractor and global encoder) and private parameters $\mathbf{W}_m^{\mathrm{p}}$ (*i.e.*, the parameters of local encoder and predictor). The former aims to extract hidden global components and recover inter-client dependency, while the latter focuses on capturing correlations within the local road network and forecasting local traffic conditions. After local training on each client, the shared parameters are uploaded to the server for aggregation, which can be formulated as

$$\overline{\mathbf{W}}^{\mathrm{s}} \leftarrow \sum_{m=1}^{M} \frac{|\mathcal{V}_m|}{|\mathcal{V}|}\mathbf{W}_m^{\mathrm{s}}. \tag{15}$$

The aggregated parameters $\overline{\mathbf{W}}^{\mathrm{s}}$ are then distributed to clients for the next round of local training. Through iterative training and aggregation, this personalized aggregation strategy enables clients to generate proxy nodes based on the extracted hidden global components and leverage global information in proxy nodes to recover the missing inter-client dependency. Meanwhile, the private parameters preserve region-specific characteristics and mitigate the impact of data heterogeneity.

Experiments

To demonstrate the effectiveness of our proposed FedHINT, in this section, we perform a series of experiments on four real-world highway traffic datasets in FL scenarios.

Experimental Setup

Datasets and training details We evaluate the proposed FedHINT on four popular datasets, including *PEMS03*,

Method	*PEMS03*		*PEMS04*		*PEMS07*		*PEMS08*	
	MAE	RMSE	MAE	RMSE	MAE	RMSE	MAE	RMSE
FedGRU (Liu et al. 2020)	20.85	32.39	26.38	40.75	29.28	44.55	21.58	33.59
CNFGNN (Meng, Rambhatla, and Liu 2021)	16.89	26.33	22.17	34.14	24.78	38.24	17.47	27.29
CTFL (Zhang et al. 2022)	17.43	26.77	22.90	35.48	24.62	37.48	18.73	29.32
FedGCN (Hu et al. 2024)	15.87	23.78	20.46	30.75	22.50	35.22	16.56	24.91
FedGTP (Yang et al. 2024)	16.30	25.43	19.88	31.30	21.54	34.27	15.90	25.47
FedTPS (Zhou et al. 2024)	15.48	23.92	19.65	31.21	21.63	34.43	16.05	25.73
FedHINT (ours)	**11.95**	**19.12**	**16.11**	**26.39**	**17.19**	**28.42**	**12.38**	**20.52**

Table 1: Overall performance on four datasets. The best results are highlighted in bold.

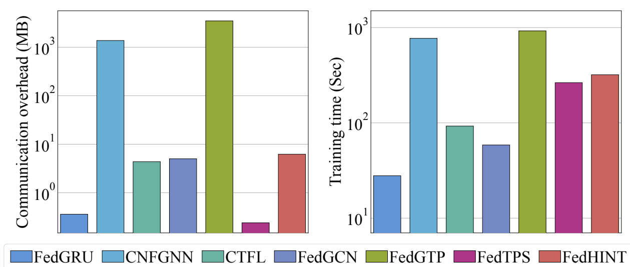


Figure 3: Comparison of the resource overhead of different methods on *PEMS03* dataset.

PEMS04, *PEMS07*, and *PEMS08*, which contain traffic flow data collected by CalTrans Performance Measurement System (PeMS) (Chen 2002). Following previous works (Yang et al. 2024; Zhou et al. 2024), we split the datasets for training, validation, and test with a ratio of $6:2:2$ and employ the METIS (Karypis 1997) algorithm to partition global traffic network into six subgraphs to simulate the FL scenarios. We use the Adam (Kingma and Ba 2015) optimizer with a learning rate of 0.003 and set the batch size to 64. The local rounds and global rounds are configured as 2 and 200, respectively. All the experiments in this work are implemented in PyTorch 1.12.1 with one GeForce RTX-3090 GPU.

Evaluation metrics Following previous work (Meng, Rambhatla, and Liu 2021), we utilize two common metrics to reveal the effectiveness of the proposed FedHINT, including Mean Absolute Error (MAE) and Root Mean Square Error (RMSE). In all experiments, we predict the traffic flow for the future 12 time stamps based on the past 12 time stamps. We evaluate the performance of the local model on each client and then average the metrics across all clients.

Main Results

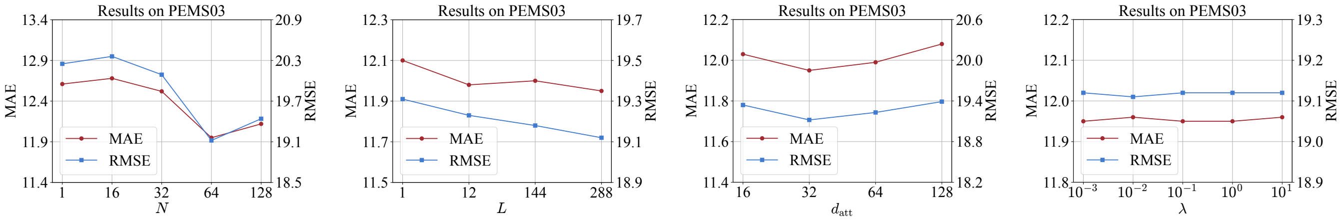
Here we conduct comprehensive experiments on four traffic prediction datasets to evaluate the performance of our FedHINT against current federated traffic prediction methods, including FedGRU (Liu et al. 2020), CNFGNN (Meng, Rambhatla, and Liu 2021), CTFL (Zhang et al. 2022), FedGCN (Hu et al. 2024), FedGTP (Yang et al. 2024), and FedTPS (Zhou et al. 2024).

As shown in Tab. 1, the proposed FedHINT significantly enhances the prediction performance and achieves superior performance compared with baseline methods. We argue that this strong performance benefits from the effective use of hidden global components, based on which proxy nodes are generated to compensate for the loss of information from cross-regional nodes so as to recover inter-client dependency. In particular, compared with the second-best methods, our approach achieves an average decrease of 3.73 on MAE and 4.81 on RMSE across four datasets.

We further investigate the resource overhead of our FedHINT, which is a crucial consideration for practical deployment in FL scenarios. Fig. 3 reports the communication overhead and training time per global round of the baseline methods and our proposed FedHINT on *PEMS03* dataset. Compared with methods that do not consider the inter-client dependency (*i.e.*, FedGRU, CTFL, and FedTPS), the proposed FedHINT introduces additional parameters to extract hidden global components and recover the missing inter-client dependency. Although this results in increased communication overhead and training time, it yields substantial improvements in prediction performance. Compared with methods that recover the inter-client dependency by uploading data features (*i.e.*, CNFGNN and FedGTP), our FedHINT extracts hidden global components from local traffic data and uploads only model parameters, which significantly reduces communication overhead. This reduction also decreases the proportion of time spent on I/O operations during training, which further decreases the overall training time.

Ablation Study

To investigate the effectiveness of different key components in our FedHINT, we conduct the following ablative experiments, including: 1) we remove the generated proxy nodes, denoted as "w/o GPN"; 2) we remove the time-shifted filters, denoted as "w/o TSF"; 3) we remove the global encoder and learn correlations between nodes with a single encoder, denoted as "w/o GED"; 4) we remove the personalized aggregation strategy and share all model parameters, denoted as "w/o PAS". Tab. 2 shows the results of ablation studies on four traffic prediction datasets. We can clearly observe that the prediction performance decreases when any component is removed, which indicates that each component significantly contributes to the final performance. For example, the model suffers considerable performance degradation when the generated proxy nodes are removed. This result indicates



(a) Analysis of N (b) Analysis of L (c) Analysis of d_{att} (d) Analysis of λ

Figure 4: Sensitivity analysis of (a) the number of global queries N, (b) the number of frequency filters L, (c) the attention dimension d_{att}, and (d) the weight λ in Eq. (14) on *PEMS03* dataset.

Method	*PEMS03*		*PEMS04*		*PEMS07*		*PEMS08*	
	MAE	RMSE	MAE	RMSE	MAE	RMSE	MAE	RMSE
w/o GPN	12.70	20.39	17.11	27.79	18.13	29.89	13.07	21.79
w/o TSF	12.11	19.28	16.32	26.58	17.42	28.68	12.49	20.54
w/o GED	12.52	20.22	17.00	27.66	17.89	29.75	13.15	21.91
w/o PAS	12.52	20.19	16.84	27.47	17.88	29.58	13.05	21.71
FedHINT	**11.95**	**19.12**	**16.11**	**26.39**	**17.19**	**28.42**	**12.38**	**20.52**

Table 2: The results of ablation studies.

that the proxy nodes generated from hidden global components can effectively represent global information, which facilitates the recovery of missing inter-client dependency and improves prediction performance.

Parametric Sensitivity

In our proposed FedHINT, there are four critical hyperparameters N, L, d_{att}, and λ that should be pre-tuned manually. In this section, we analyze the parametric sensitivity of our method to these parameters on *PEMS03* dataset. As illustrated in Fig. 4, we find that these four parameters are critical for our FedHINT to achieve good performance. To be specific, the model achieves optimal performance on *PEMS03* dataset when $N = 64$. Note that the optimal value of N is positively correlated with the number of sensors $|\mathcal{V}|$. Besides, our model achieves the best performance when the number of frequency filters L is set to 288 (*i.e.*, the total number of time stamps in a day), as this setting allows the filters to capture critical temporal characteristics across different times of the day. The best results are obtained when d_{att} is set to 32. Therefore, we adopt such a parameter configuration in our method. Furthermore, our method demonstrates stable prediction performance when λ varies from 10^{-3} to 10^{1}, and we set $\lambda = 10^{-1}$ in all experiments.

Case Study

To further explore the interpretability and effectiveness of the proxy nodes generated based on hidden global components, we analyze the cosine similarity between proxy nodes generated from local traffic data and proxy nodes derived from global traffic data on *PEMS03* dataset. As shown in Fig. 5, the similarity between locally and globally generated proxy nodes exhibits clear diagonal dominance, which indicates that the proxy nodes generated locally within each client are closely aligned with those generated from global data. This demonstrates that our FedHINT can effectively extract hidden global components that reflect cross-regional traffic changes from local data, which facilitates the generation of proxy nodes that represent global information. This analysis provides further insight into the interpretability of our FedHINT and demonstrates its ability to extract and utilize hidden global components effectively within each client.

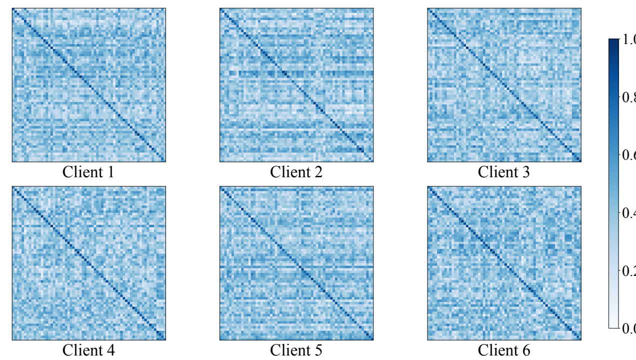


Figure 5: Similarity between locally and globally generated proxy nodes on *PEMS03* dataset.

Conclusion

In this paper, we propose FedHINT, a novel federated traffic prediction method that leverages hidden global components within local traffic data to recover the missing inter-client dependency without compromising data privacy. Different from previous works that upload data features to model the dependency across clients, the proposed FedHINT extracts hidden global components that reflect cross-regional traffic changes from local traffic data. Based on the extracted hidden global components, our FedHINT generates proxy nodes that represent global information on each client to compensate for the loss of information from cross-regional nodes. By learning the correlations between local nodes and proxy nodes, our FedHINT effectively recovers the missing inter-client dependency. Experimental evaluations conducted on four widely used datasets demonstrate that our FedHINT significantly outperforms state-of-the-art methods, with an average decrease of 3.73 and 4.81 on MAE and RMSE, respectively.

Acknowledgments

This research is supported by NSF of China (Nos: 62336003, 12371510).

References

Bai, L.; Yao, L.; Li, C.; Wang, X.; and Wang, C. 2020. Adaptive graph convolutional recurrent network for traffic forecasting. In *Advances in Neural Information Processing Systems*, 17804–17815.

Chen, C. 2002. *Freeway performance measurement system (PeMS)*. University of California, Berkeley.

Fu, L.; Deng, B.; Huang, S.; Liao, T.; Pan, S.; and Chen, C. 2025a. Less is more: Federated graph learning with alleviating topology heterogeneity from a causal perspective. In *International Conference on Machine Learning*.

Fu, L.; Huang, S.; Lai, Y.; Liao, T.; Zhang, C.; and Chen, C. 2025b. Beyond federated prototype learning: Learnable semantic anchors with hyperspherical contrast for domain-skewed data. In *Proceedings of the AAAI Conference on Artificial Intelligence*, 16648–16656.

Geng, X.; Li, Y.; Wang, L.; Zhang, L.; Yang, Q.; Ye, J.; and Liu, Y. 2019. Spatiotemporal multi-graph convolution network for ride-hailing demand forecasting. In *Proceedings of the AAAI Conference on Artificial Intelligence*, 3656–3663.

Guo, S.; Lin, Y.; Feng, N.; Song, C.; and Wan, H. 2019. Attention based spatial-temporal graph convolutional networks for traffic flow forecasting. In *Proceedings of the AAAI Conference on Artificial Intelligence*, 922–929.

Hu, N.; Liang, W.; Zhang, D.; Xie, K.; Li, K.; and Zomaya, A. Y. 2024. FedGCN: A federated graph convolutional network for privacy-preserving traffic prediction. *IEEE Transactions on Sustainable Computing*, 9(6): 925–935.

Huang, J.; Huang, Z.; Fang, X.; Feng, S.; Chen, X.; Liu, J.; Yuan, H.; and Wang, H. 2022. DuETA: Traffic congestion propagation pattern modeling via efficient graph learning for eta prediction at baidu maps. In *Proceedings of the 31st ACM International Conference on Information and Knowledge Management*, 3172–3181.

Huang, R.; Huang, C.; Liu, Y.; Dai, G.; and Kong, W. 2021. LSGCN: Long short-term traffic prediction with graph convolutional networks. In *Proceedings of the Twenty-Ninth International Joint Conference on Artificial Intelligence*, 2355–2361.

Huang, S.; Fu, L.; Liao, T.; Deng, B.; Zhang, C.; and Chen, C. 2025. FedBG: Proactively mitigating bias in cross-domain graph federated learning using background data. In *Proceedings of the Thirty-Fourth International Joint Conference on Artificial Intelligence*, 5408–5416.

Jiang, J.; Han, C.; Zhao, W. X.; and Wang, J. 2023a. PDFormer: Propagation delay-aware dynamic long-range transformer for traffic flow prediction. In *Proceedings of the AAAI Conference on Artificial Intelligence*, 4365–4373.

Jiang, R.; Wang, Z.; Yong, J.; Jeph, P.; Chen, Q.; Kobayashi, Y.; Song, X.; Fukushima, S.; and Suzumura, T. 2023b. Spatio-temporal meta-graph learning for traffic forecasting. In *Proceedings of the AAAI Conference on Artificial Intelligence*, 8078–8086.

Karypis, G. 1997. METIS: Unstructured graph partitioning and sparse matrix ordering system. *Technical Report*.

Kingma, D. P.; and Ba, J. 2015. Adam: A method for stochastic optimization. In *International Conference on Learning Representations*.

Li, L.; Liu, J.; Cheng, L.; Qiu, S.; Wang, W.; Zhang, X.; and Zhang, Z. 2018. CreditCoin: A privacy-preserving blockchain-based incentive announcement network for communications of smart vehicles. *IEEE Transactions on Intelligent Transportation Systems*, 19(7): 2204–2220.

Liang, K.; Zhou, S.; Liu, M.; Liu, Y.; Tu, W.; Zhang, Y.; Fang, L.; Liu, Z.; and Liu, X. 2024. Hawkes-enhanced spatial-temporal hypergraph contrastive learning based on criminal correlations. In *Proceedings of the AAAI Conference on Artificial Intelligence*, 8733–8741.

Liu, F.; Zhang, W.; and Liu, H. 2023. Robust spatiotemporal traffic forecasting with reinforced dynamic adversarial training. In *Proceedings of the 29th ACM SIGKDD Conference on Knowledge Discovery and Data Mining*, 1417–1428.

Liu, L.; Tian, Y.; Chakraborty, C.; Feng, J.; Pei, Q.; Zhen, L.; and Yu, K. 2023. Multilevel federated learning-based intelligent traffic flow forecasting for transportation network management. *IEEE Transactions on Network and Service Management*, 20(2): 1446–1458.

Liu, Y.; James, J.; Kang, J.; Niyato, D.; and Zhang, S. 2020. Privacy-preserving traffic flow prediction: A federated learning approach. *IEEE Internet of Things Journal*, 7(8): 7751–7763.

McMahan, B.; Moore, E.; Ramage, D.; Hampson, S.; and y Arcas, B. A. 2017. Communication-efficient learning of deep networks from decentralized data. In *Artificial Intelligence and Statistics*, 1273–1282.

Meng, C.; Rambhatla, S.; and Liu, Y. 2021. Cross-node federated graph neural network for spatio-temporal data modeling. In *Proceedings of the 27th ACM SIGKDD Conference on Knowledge Discovery and Data Mining*, 1202–1211.

Meng, L.; Liang, K.; Yu, H.; Liu, Y.; Zhou, S.; Liu, M.; and Liu, X. 2024. FedEAN: Entity-aware adversarial negative sampling for federated knowledge graph reasoning. *IEEE Transactions on Knowledge and Data Engineering*, 36(12): 8206–8219.

Piao, X.; Chen, Z.; Murayama, T.; Matsubara, Y.; and Sakurai, Y. 2024. Fredformer: Frequency debiased transformer for time series forecasting. In *Proceedings of the 30th ACM SIGKDD Conference on Knowledge Discovery and Data Mining*, 2400–2410.

Shi, X.; Chen, Z.; Wang, H.; Yeung, D.-Y.; Wong, W.-K.; and Woo, W.-c. 2015. Convolutional LSTM network: A machine learning approach for precipitation nowcasting. In *Advances in Neural Information Processing Systems*, 802–810.

Sutskever, I.; Vinyals, O.; and Le, Q. V. 2014. Sequence to sequence learning with neural networks. In *Advances in Neural Information Processing Systems*, 3104–3112.

Wang, H.; Zhang, R.; Cheng, X.; and Yang, L. 2022. Federated spatio-temporal traffic flow prediction based on graph convolutional network. In *2022 14th International Conference on Wireless Communications and Signal Processing*, 221–225.

Wu, H.; Hu, T.; Liu, Y.; Zhou, H.; Wang, J.; and Long, M. 2023. TimesNet: Temporal 2D-variation modeling for general time series analysis. In *International Conference on Learning Representations*.

Wu, Y.; Shen, C.; Chen, S.; Wu, C.; Li, S.; and Wei, R. 2022. Intelligent orchestrating of IoT microservices based on reinforcement learning. *Chinese Journal of Electronics*, 31(5): 930–937.

Wu, Z.; Pan, S.; Long, G.; Jiang, J.; and Zhang, C. 2019. Graph wavenet for deep spatial-temporal graph modeling. In *Proceedings of the Twenty-Eighth International Joint Conference on Artificial Intelligence*, 1907–1913.

Yang, L.; Chen, W.; He, X.; Wei, S.; Xu, Y.; Zhou, Z.; and Tong, Y. 2024. FedGTP: Exploiting inter-client spatial dependency in federated graph-based traffic prediction. In *Proceedings of the 30th ACM SIGKDD Conference on Knowledge Discovery and Data Mining*, 6105–6116.

Yi, K.; Fei, J.; Zhang, Q.; He, H.; Hao, S.; Lian, D.; and Fan, W. 2024. FilterNet: Harnessing frequency filters for time series forecasting. In *Advances in Neural Information Processing Systems*, 55115–55140.

Yu, B.; Yin, H.; and Zhu, Z. 2018. Spatio-temporal graph convolutional networks: A deep learning framework for traffic forecasting. In *Proceedings of the Twenty-Seventh International Joint Conference on Artificial Intelligence*, 3634–3640.

Yu, W.; Chen, S.; Tong, Y.; Gu, T.; and Gong, C. 2025a. Modeling inter-intra heterogeneity for graph federated learning. In *Proceedings of the AAAI Conference on Artificial Intelligence*, 22236–22244.

Yu, W.; Gong, C.; Han, B.; Fan, L.; and Yang, Q. 2025b. Integrating commonality and individuality for graph federated learning: A graph spectrum perspective. *Authorea Preprints*.

Yue, S.; Deng, Y.; Wang, G.; Ren, J.; and Zhang, Y. 2024. Federated offline reinforcement learning with proximal policy evaluation. *Chinese Journal of Electronics*, 33(6): 1360–1372.

Zhang, C.; Zhang, S.; James, J.; and Yu, S. 2021a. FASTGNN: A topological information protected federated learning approach for traffic speed forecasting. *IEEE Transactions on Industrial Informatics*, 17(12): 8464–8474.

Zhang, C.; Zhang, S.; Yu, S.; and Yu, J. J. 2022. Graph-based traffic forecasting via communication-efficient federated learning. In *2022 IEEE Wireless Communications and Networking Conference*, 2041–2046.

Zhang, K.; Yang, C.; Li, X.; Sun, L.; and Yiu, S. M. 2021b. Subgraph federated learning with missing neighbor generation. In *Advances in Neural Information Processing Systems*, 6671–6682.

Zheng, C.; Fan, X.; Wang, C.; and Qi, J. 2020. GMAN: A graph multi-attention network for traffic prediction. In *Proceedings of the AAAI Conference on Artificial Intelligence*, 1234–1241.

Zhou, H.; Yu, W.; Wan, S.; Tong, Y.; Gu, T.; and Gong, C. 2024. Traffic pattern sharing for federated traffic flow prediction with personalization. In *2024 IEEE International Conference on Data Mining*, 639–648.

Zhou, H.; Yu, W.; Wan, S.; Tong, Y.; Gu, T.; and Gong, C. 2025. FedTPS: Traffic pattern sharing for personalized federated traffic flow prediction. *Knowledge and Information Systems*, 67(7): 5873–5899.

Zhou, T.; Ma, Z.; Wen, Q.; Wang, X.; Sun, L.; and Jin, R. 2022. FEDformer: Frequency enhanced decomposed transformer for long-term series forecasting. In *International Conference on Machine Learning*, 27268–27286.