

# Unlocking Dynamic Inter-Client Spatial Dependencies: A Federated Spatio-Temporal Graph Learning Method for Traffic Flow Forecasting

Feng Wang<sup>1,2</sup>, Tianxiang Chen<sup>1,2</sup>, Shuyue Wei<sup>1,3</sup>, Qian Chu<sup>1,2</sup>,  
Yi Zhang<sup>4\*</sup>, Yifan Sun<sup>5\*</sup>, Zhiming Zheng<sup>1,2</sup>

<sup>1</sup>Beijing Advanced Innovation Center for Future Blockchain and Privacy Computing

<sup>2</sup>School of Artificial Intelligence, Beihang University, China

<sup>3</sup>School of Computer Science and Engineering, Beihang University, China

<sup>4</sup>School of Mathematics, Renmin University of China, China

<sup>5</sup>Center for Applied Statistics and School of Statistics, Renmin University of China, China

{fengwang, txchen, weishuyue, qianchu}@buaa.edu.cn, {ethanzhang, sunyifan}@ruc.edu.cn, zzheng@pku.edu.cn

## Abstract

Spatio-temporal graphs are powerful tools for modeling complex dependencies in traffic time series. However, the distributed nature of real-world traffic data across multiple stakeholders poses significant challenges in modeling and reconstructing inter-client spatial dependencies while adhering to data locality constraints. Existing methods primarily address static dependencies, overlooking their dynamic nature and resulting in suboptimal performance. In response, we propose **Federated Spatio-Temporal Graph with Dynamic Inter-Client Dependencies (FedSTGD)**, a framework designed to model and reconstruct dynamic inter-client spatial dependencies in federated learning. FedSTGD incorporates a federated nonlinear computation decomposition module to approximate complex graph operations. This is complemented by a graph node embedding augmentation module, which alleviates performance degradation arising from the decomposition. These modules are coordinated through a client-server collective learning protocol, which decomposes dynamic inter-client spatial dependency learning tasks into lightweight, parallelizable subtasks. Extensive experiments on four real-world datasets demonstrate that FedSTGD achieves superior performance over state-of-the-art baselines in terms of RMSE, MAE, and MAPE, approaching that of centralized baselines. Ablation studies confirm the contribution of each module in addressing dynamic inter-client spatial dependencies, while sensitivity analysis highlights the robustness of FedSTGD to variations in hyperparameters.

## Introduction

Traffic Flow Forecasting (TFF) is a cornerstone of Intelligent Transportation Systems. The performance of TFF critically depends on the modeling of complex spatial dependencies within transportation systems. These dependencies are naturally encoded as graphs, where nodes denote sensors and edges signify their spatial relationships. Due to their capacity for capturing such spatial dependencies, graph neural networks have become a dominant approach for TFF (Fang et al. 2021; Li and Zhu 2021; Kong, Guo, and Liu 2024).

Spatial dependencies in transportation systems are typically dynamic. To validate this, we examine passenger flow data from the Hangzhou Metro, and observe variations in spatial dependencies between weekdays and weekends, and even within weekdays (Figure 1(a)). Due to the abundant information they contain, dynamic spatial dependencies offer potential to enhance the accuracy of TFF models.

The modeling of dynamic spatial dependencies in real-world applications is often hindered by data fragmentation across diverse stakeholders, such as governmental agencies and private enterprises, collectively referred to as clients (Zhang et al. 2021). This fragmentation is exacerbated by stringent data privacy regulations, such as the General Data Protection Regulation (EP and Council 2016), which restrict data aggregation. Therefore, the distributed nature of traffic data introduces a critical distinction between intra-client and inter-client dynamic spatial dependencies (Ge et al. 2024; Yang et al. 2024; Zhou et al. 2024), as illustrated in Figure 1(b). While dynamic intra-client dependencies can be modeled locally, the data fragmentation renders the modeling of dynamic inter-client spatial dependencies intractable.

Federated Learning (McMahan et al. 2017) has become a foundational paradigm for collaborative training under data locality, attracting broad research interest across domains (Wei et al. 2025). However, existing federated graph learning models either overlook inter-client spatial modeling (e.g., FLoS (Wang et al. 2022)) or treat these dependencies as static (e.g., FedGTP (Yang et al. 2024)). Such limitations curtail model expressivity and, in turn, predictive performance. To demonstrate this, we evaluate representative models on the HZMetro dataset (Figure 1(c)). Our proposed FedSTGD, which explicitly models dynamic inter-client spatial dependencies, achieves significant gains, approaching the centralized baselines. However, modeling such dependencies in federated learning confronts **three key challenges**.

First, efficiently decomposing inter-client nonlinear operations is challenging. Modeling such dependencies requires complex nonlinear operations over distributed client data, which cannot be performed in federated settings due to data locality constraints. Existing methods often rely on polynomial approximations (Chrysos et al. 2021; Fan et al. 2023; Yang et al. 2024), but this entails projections into high-

\*Corresponding authors.

Copyright © 2026, Association for the Advancement of Artificial Intelligence (www.aaai.org). All rights reserved.

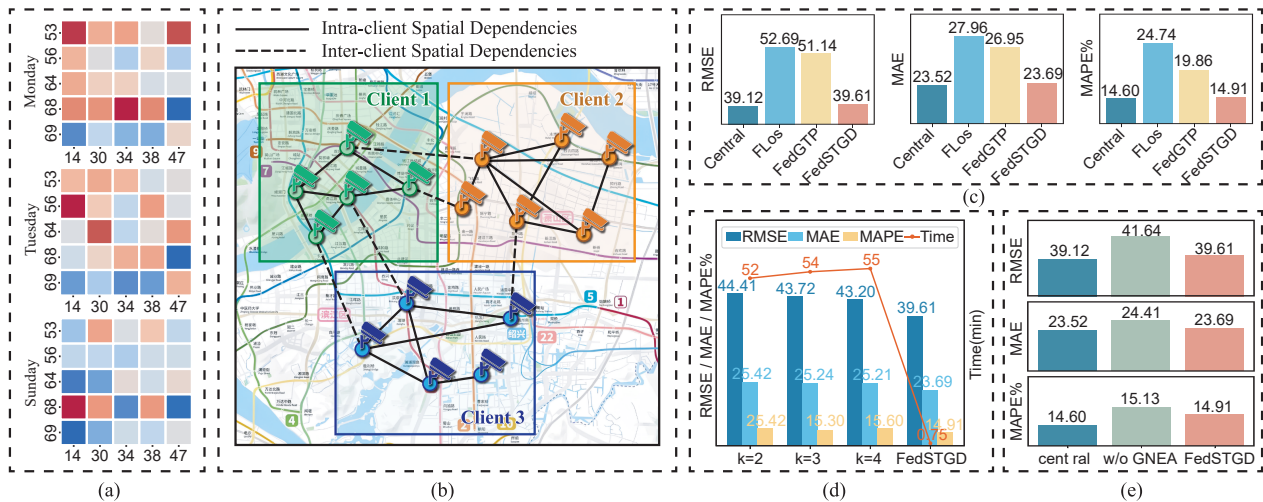


Figure 1: Experimental results on the HZMetro dataset. (a) Visualization of dynamic inter-client spatial dependencies; (b) Schematic illustration of intra-client and inter-client spatial dependencies; (c) Performance comparison between FedSTGD and other mainstream methods; (d) Comparison of Federated Nonlinear Computation Decomposition in FedSTGD against AdptPoLU in other mainstream methods; (e) Ablation study on Graph Node Embedding Augmentation module.

dimensional spaces where dimensionality scales exponentially with polynomial degree, leading to high overhead. Our empirical evaluation on HZMetro dataset (Figure 1(d)) reveals that this method incurs high overhead with minimal performance gains, while ours entails negligible overhead.

Secondly, the representational capacity of graph node embeddings may be insufficient in federated settings. This limitation stems from the decomposition of inter-client nonlinear operations, which compromises the expressive power of the embeddings and results in degraded federated performance. We confirm this limitation through an ablation study on the HZmetro dataset (Figure 1(e)), where a variant without our embedding augmentation falls short of centralized performance, while its inclusion yields performance close to that of centralized baselines.

Third, executing inter-client graph operations remains a formidable challenge. Modeling these dependencies requires computations that span multiple clients, necessitating the decomposition of such operations into parallelizable sub-tasks suitable for distributed execution.

To this end, we introduce FedSTGD (**F**ederated **S**patio-**T**emporal **G**raph with **D**ynamic **I**nter-**C**lient **D**ependencies), a novel framework designed to enhance TFF by explicitly modeling dynamic inter-client spatial dependencies. FedSTGD consists of three key components: (1) *Module of federated non-linear computation decomposition* decomposes non-linear computations across clients; (2) *Module of graph node embedding augmentation* enriches the representational capacity of node embeddings; (3) *Module of client-server collective learning* coordinates dynamic inter-client spatial dependencies modeling between clients and central server. The contributions of our work are summarized as follows:

- We make the first attempt to model and reconstruct dynamic inter-client spatial dependencies in a federated spatio-temporal graph model by introducing a novel re-

covery mechanism under the data locality constraint.

- We propose a novel framework, dubbed FedSTGD. This framework employs a federated nonlinear computation decomposition module to approximate complex inter-client nonlinear computations as lightweight, client-side operations. To address resulting approximation errors, we integrate a graph node embedding augmentation module for enhancing local representations. On this basis, a client-server collective learning module enables the recovery of dynamic inter-client spatial dependencies by decomposing the tasks into lightweight, parallelizable subtasks amenable to distributed execution across clients.
- Extensive evaluations on four real-world datasets show that FedSTGD delivers state-of-the-art performance, outperforming diverse baseline methods and closely approximating the performance of centralized baselines. Ablation studies confirm the effectiveness of individual components, while hyperparameter sensitivity analyses affirm the framework’s robustness.

## Related Work

We review related work on federated spatio-temporal graph-based TFF from three key perspectives: methods that rely on predefined graphs, publicly accessible graphs, and that focus on modeling inter-client spatial dependencies.

**Predefined Graph** This line of research leverages predefined graph structures to model inter-client and intra-client dependencies. For instance, FLoS (Wang et al. 2022) and MFVSTGNN (Liu et al. 2023) use connectivity matrices to characterize spatial dependencies. However, such approaches yield only rough representations of spatio-temporal dependencies, thereby constraining their predictive performance.

**Publicly Accessible Graphs** This line of research leverages publicly accessible graphs to model inter-client and intra-client dependencies. For instance, CNFGNN (Meng, Rambhatla, and Liu 2021) extracts cross-node dependencies at the server level through predefined graph topologies, while FedSTN (Yuan et al. 2023) separately captures long- and short-term spatio-temporal patterns within a shared road network. Similarly, FedAGCN (Qi et al. 2023) and FCGCN (Xia, Jin, and Chen 2023) assume a globally shared spatial topology and employ community detection for subgraph partitioning; CTFL (Zhang et al. 2022) extends this paradigm by assuming identical graph structures across clients. Although these methods partially capture inter-client spatial dependencies, their dependence on global topology assumptions often proves impractical in real-world settings.

**Modeling of Inter-Client Spatial Dependency** This line of research focuses on modeling inter-client spatial dependencies in federated learning. Existing methods approximate these dependencies in various ways: FASTGNN (Zhang et al. 2021) introduces noise from differential privacy mechanism; DSTGCRN (Pham et al. 2025) and FedSTG (Zhang et al. 2025) employ graph-aware neighborhood aggregation; FUELS (Liu et al. 2025) applies contrastive dual semantic alignment; and FedGTP (Yang et al. 2024) reconstructs static spatial relations. However, these methods either overlook the dynamic nature of inter-client spatial dependencies or achieve only partial approximations.

## Preliminaries

In this section, we define the key components and objectives of our study.

**Definition 1.** A spatio-temporal graph for a traffic network at time  $t$  is denoted by  $G^t = (V, E, A^t)$ , where  $V$  is the set of  $N = |V|$  nodes,  $E$  is the set of edges reflecting connectivity, and  $A^t \in \mathbb{R}^{N \times N}$  is a dynamic adjacency matrix.

**Definition 2.** The graph signal for  $G^t = (V, E, A^t)$  at time  $t$  is denoted by  $X^t \in \mathbb{R}^{N \times d}$ , where  $d$  is the feature dimensionality per node.

**Definition 3.** Spatio-temporal graph-based TFF is to learn a function  $f(\cdot; \theta)$  that maps  $T_{in}$  input graph signals  $\mathcal{X} = (X^{t-T_{in}+1}, \dots, X^t)$  and graph structures  $\mathcal{G} = (G^{t-T_{in}+1}, \dots, G^t)$  to  $T_{out}$  future graph signals  $\mathcal{Y} = (X^{t+1}, \dots, X^{t+T_{out}})$ . The parameters  $\theta$  are optimized by minimizing a loss function  $\mathcal{L}$  over the training data distribution  $\mathcal{D}$ :

$$\theta^* = \operatorname{argmin}_{\theta} \mathbb{E}_{(\mathcal{X}, \mathcal{Y}) \sim \mathcal{D}} [\mathcal{L}(f(\mathcal{G}, \mathcal{X}; \theta), \mathcal{Y})]. \quad (1)$$

An example of a spatial-temporal graph model for TFF is the Time-aware Graph Convolutional Recurrent Network (TGCRN) proposed by (Ma et al. 2024).

**Example 1.** The TGCRN model operates as follows:

$$\mathcal{A}_\nu = \langle E_\nu, E_\nu^\top \rangle, \quad (2)$$

$$\eta_\tau^t = \langle E_\tau^t, E_\tau^{t-1\top} \rangle, \quad (3)$$

$$\mathcal{A}_\rho = \tanh(\langle X^{(t)}, X^{(t)\top} \rangle), \quad (4)$$

$$A^t = (\mathbf{1}_{N \times N} + \alpha \sigma(\mathcal{A}_\rho)) \odot (\mathcal{A}_\nu + \eta_\tau^t \cdot \mathbf{1}_{N \times N}), \quad (5)$$

$$\hat{E}^t = [E_\nu \| (\mathbf{1}_{N \times 1} \otimes E_\tau^t)], \quad (6)$$

$$z^t = \sigma(\mathcal{A}^t I_1^t \hat{E}^t W_z + \hat{E}^t b_z), \quad (7)$$

$$r^t = \sigma(\mathcal{A}^t I_1^t \hat{E}^t W_r + \hat{E}^t b_r), \quad (8)$$

$$\hat{h}^t = \tanh(\mathcal{A}^t I_2^t \hat{E}^t W_{\hat{h}} + \hat{E}^t b_{\hat{h}}), \quad (9)$$

$$h^t = (1 - z^t) \odot h^{t-1} + z^t \odot \hat{h}^t. \quad (10)$$

Here:  $\odot$  is the Hadamard product;  $\otimes$  is the Kronecker product;  $\mathbf{1}_{p \times q}$  denotes the all-ones matrix;  $I_1^t = [X^t \| h^{t-1}]$  and  $I_2^t = [X^t \| r^t \odot h^{t-1}]$  are the gated recurrent unit (GRU) inputs;  $E_\nu \in \mathbb{R}^{N \times d_N}$  is the learnable graph node embeddings;  $E_\tau^t \in \mathbb{R}^{1 \times d_T}$  is the learnable time vectors;  $\mathcal{A}_\nu, \mathcal{A}_\rho \in \mathbb{R}^{N \times N}$  are the self-learned adjacency matrix and the periodic discriminant matrix;  $\eta_\tau^t$  is a scalar trend factor;  $\sigma(\cdot)$  is the sigmoid activation;  $\alpha$  is a hyperparameter;  $(W_z, W_r, W_{\hat{h}})$  and  $(b_z, b_r, b_{\hat{h}})$  are learnable weight matrices and bias vectors, respectively.

The spatial-temporal graph-based TFF can be extended to a federated setting with a set of clients  $\mathcal{C} = \{C_1, \dots, C_M\}$  that partition the global graph, where each client  $C_i$  possesses a local graph signal  $X_i^t \in \mathbb{R}^{N_i \times d}$  and a subgraph  $G_i^t = (V_i, E_i, A_{ii}^t)$ , comprising  $N_i = |V_i|$  nodes, intra-client edges  $E_i$ , and an adjacency matrix  $A_{ii}^t \in \mathbb{R}^{N_i \times N_i}$ .

**Definition 4.** Federated spatio-temporal graph-based TFF is to learn a function  $f(\cdot; \theta)$  that takes, for each client  $C_i$ ,  $T_{in}$  graph signals  $\mathcal{X}_i = (X_i^{t-T_{in}+1}, \dots, X_i^t)$  and graph structures  $\mathcal{G}_{ii} = (G_{ii}^{t-T_{in}+1}, \dots, G_{ii}^t)$  as input and predicts  $T_{out}$  future graph signals  $\mathcal{Y}_i = (X_i^{t+1}, \dots, X_i^{t+T_{out}})$ . The parameters  $\{\theta_1, \dots, \theta_M\}$ , encompassing both global and private parameters, are optimized by minimizing a loss function  $\mathcal{L}$  across the clients, weighted by their node sizes:

$$\{\theta_1^*, \dots, \theta_M^*\} = \operatorname{argmin}_{\theta_1, \dots, \theta_M} \sum_{i=1}^M \frac{N_i}{N} \mathbb{E}_{(\mathcal{X}_i, \mathcal{Y}_i) \sim \mathcal{D}_i} [\mathcal{L}(f(\mathcal{G}_{ii}, \mathcal{X}_i; \theta), \mathcal{Y}_i)] \quad (11)$$

where  $\mathcal{D}_i$  is the training data distribution for client  $C_i$ .

## Methodology

In this section, we introduce the FedSTGD framework. We start with the federated formation of FedSTGD, then describe three modules for client-side spatio-temporal modeling, followed by the federated model training procedure, and conclude with time and communication complexity analysis. The architecture is shown in Figure 2, with pseudocode provided in Appendix A of the supplementary material.

### Federated Formulation of FedSTGD.

FedSTGD leverages the TGCRN model as its backbone, noted for its state-of-the-art prediction accuracy via learning of dynamic spatial dependencies. In the following, we first describe the client-side partitioning of model components, followed by a concise overview of client-side spatio-temporal modeling and server-side model aggregation.

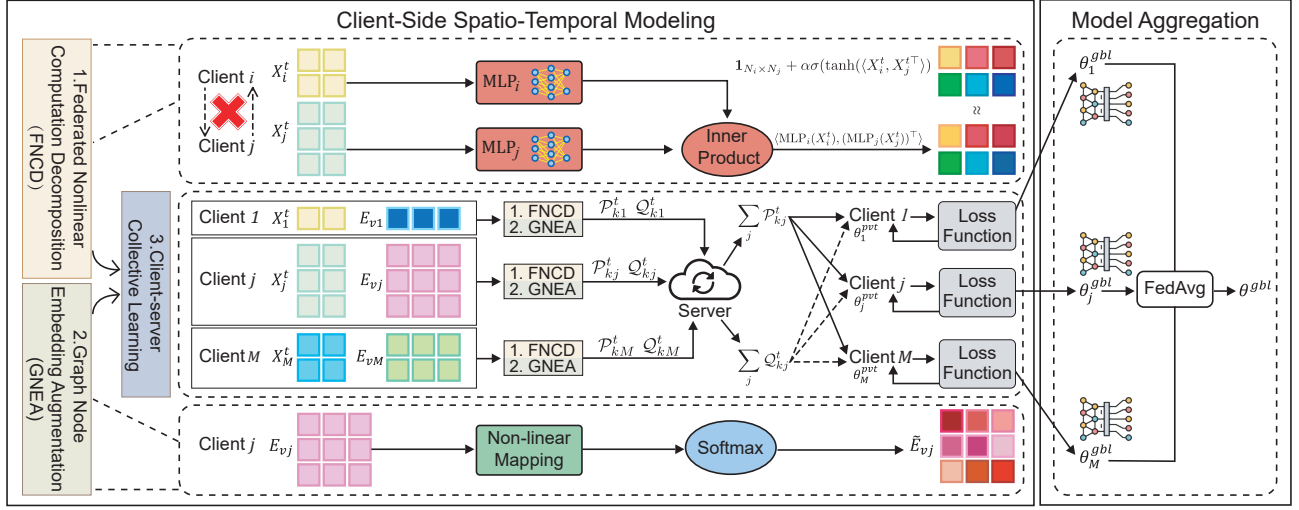


Figure 2: Overview of the Proposed FedSTGD Framework

**1. Client-Side Data, Embedding, and Spatial Dependency Partitioning.** To reformulate the model defined in Eqs. (2)-(10) to accommodate federated learning, we partition the model’s node features, embeddings, GRU inputs, and adjacency matrix across the clients set  $\mathcal{C}$ . Specifically, the node features  $X^t$ , node embeddings  $E_\nu$ , time embedding  $\hat{E}^t$ , and the GRU inputs  $I_k^t$  ( $k = 1, 2$ ) are partitioned into client-specific blocks:  $\{X_j^t\}_{j=1}^M$ ,  $\{E_{\nu j}\}_{j=1}^M$ ,  $\{\hat{E}_j^t\}_{j=1}^M$ , and  $\{I_{k j}^t\}_{j=1}^M$  respectively. Additionally, the time-dependent adjacency matrix is partitioned into  $\mathcal{A}^t = [\mathcal{A}_{ij}^t]_{i,j=1}^M$ , where the diagonal blocks  $\mathcal{A}_{ii}^t$  represent **dynamic intra-client spatial dependencies**, and the off-diagonal blocks  $\mathcal{A}_{ij}^t$  ( $i \neq j$ ) encode **dynamic inter-client spatial dependencies**.

**2. Client-Side Spatio-Temporal Modeling.** Building upon the aforementioned partitioning and in accordance with the model specified in Eqs. (2)–(10), we formulate the client-side spatio-temporal modeling as follows:

$$L_{ki}^t = \sum_{j=1}^M \left\{ \left[ (\mathbf{1}_{N_i \times N_j} + \alpha \sigma(\tanh(\langle X_i^t, X_j^{t\top} \rangle)) \odot (\langle E_{\nu i}, E_{\nu j}^\top \rangle + \eta_{\tau i}^t \cdot \mathbf{1}_{N_i \times N_j}) \right] \cdot I_{k j}^t \right\}, \quad (12)$$

$$z_i^t = \sigma(L_{i1}^t \hat{E}_i^t W_{zi} + \hat{E}_i^t b_{zi}), \quad (13)$$

$$r_i^t = \sigma(L_{i1}^t \hat{E}_i^t W_{ri} + \hat{E}_i^t b_{ri}), \quad (14)$$

$$\hat{h}_i^t = \tanh(L_{i2}^t \hat{E}_i^t W_{hi} + \hat{E}_i^t b_{hi}), \quad (15)$$

$$h_i^t = (1 - z_i^t) \odot h_i^{t-1} + z_i^t \odot \hat{h}_i^t. \quad (16)$$

where  $W_{zi}$ ,  $W_{ri}$ ,  $W_{hi}$ ,  $b_{zi}$ ,  $b_{ri}$ ,  $b_{hi}$ , and  $\eta_{\tau i}^t$  are client-side global learnable parameters;  $E_{\nu i}$  is client-side private learnable parameters.

**Remark 1.** Learning dynamic inter-client spatial dependencies in FedSTGD is **non-trivial**. This is because learning such dependencies requires inter-client data operations compounded by nonlinear operations (Eq. (12)), which can

not be done directly under data locality constraints. Therefore, we will introduce **three dedicated modules** in the subsequent analysis to address this issue.

**3. Server-side Model Aggregation.** The server optimizes the global model parameters through the FedAvg algorithm.

### Federated Nonlinear Computation Decomposition

In this module, we address the client-level decomposition of nonlinear operations, i.e.,

$$\mathbf{1}_{N_i \times N_j} + \alpha \sigma(\tanh(\langle X_i^t, X_j^{t\top} \rangle)). \quad (17)$$

Computing Eq. (17) requires cross-client inner products before applying nonlinearity. This process raises privacy concerns, as it shares data across clients to compute  $\langle X_i^t, X_j^t \rangle$ . To mitigate this, we try to leverage the one-way property of nonlinear functions, where the forward mapping from data  $X_i$  to its transformed representation is efficient, but the inverse mapping is computationally infeasible or noisy, thus protecting against data inference (Tian, Weng, and Feng 2024). However, the non-commutative nature of these operations renders the direct reversal of the nonlinear transformation and inner product infeasible.

Inspired by the universal approximation theorem (Lu et al. 2021), which establishes that the multi-layer perceptron (MLP) can approximate any continuous function with arbitrary precision given sufficient layers and neurons, we propose a federated nonlinear computation decomposition module. In this module, each client locally trains an MLP, and the inner product  $\langle \text{MLP}_i(X_i^t), (\text{MLP}_j(X_j^t))^\top \rangle$  is formulated to approximate  $\mathbf{1}_{N_i \times N_j} + \alpha \sigma(\tanh(\langle X_i^t, X_j^{t\top} \rangle))$ . This approximation supports the subsequent decomposition of federated computations. Formally, this relationship is expressed as:

$$\mathbf{1}_{N_i \times N_j} + \alpha \sigma(\tanh(\langle X_i^t, X_j^{t\top} \rangle)) \approx \langle \text{MLP}_i(X_i^t), (\text{MLP}_j(X_j^t))^\top \rangle. \quad (18)$$

Here,  $\text{MLP}_i$  represents an  $L$ -layer MLP with public learnable parameters  $W_{\text{MLP}_i}$  and  $b_{\text{MLP}_i}$ . During training, these

parameters are iteratively optimized to enhance the approximation accuracy of  $\text{MLP}_i(X_i^t)$  as specified in Eq. (18).

### Graph Node Embedding Augmentation

In this module, we propose a graph node embedding augmentation strategy to mitigate the degradation in graph representation capacity induced by the previous federated nonlinear computation decomposition. Formally, let  $E_{vi} \in \mathbb{R}^{N \times d}$  denote the node embedding for client  $i$ . Our augmentation module unfolds in two key steps:

**1. Non-Linear Mapping.** We apply a nonlinear transformation to the embedding matrix  $E_{vi}$  to enhance its feature expressiveness:

$$E'_{vi} = \mathcal{F}(E_{vi}; W_{NL}), \quad (19)$$

where  $\mathcal{F}(\cdot)$  is a nonlinear mapping function (e.g., an MLP, kernel function, or attention mechanism),  $W_{NL} \in \mathbb{R}^{d \times d'}$  are global learnable parameters in  $\mathcal{F}(\cdot)$ , and  $d'$  denotes the output dimensionality. This step injects additional representational capacity into the graph node embedding.

**2. Softmax Normalization.** To eliminate scale discrepancies across the augmented graph node embeddings of different clients, we normalize the transformed embedding matrix row-wise using the softmax function:

$$\tilde{E}_{vi} = \text{Softmax}(E'_{vi}) \quad (20)$$

where the softmax operation is performed along the feature dimension, ensuring that each row sums to one.

### A Client-server Collective Learning Protocol

In this module, we propose a client-server collective learning protocol to facilitate the distributed execution of client-side spatio-temporal modeling, as formulated in Eqs. (12)–(16). To this end, we first substitute Eqs. (18) and (20) into Eq. (12), resulting in the following reformulation:

$$\begin{aligned} \tilde{L}_{ki}^t = & \quad (21) \\ & \underbrace{\sum_{j=1}^M [\langle \text{MLP}_i(X_i^t), (\text{MLP}_j(X_j^t))^\top \rangle \odot (\eta_\tau^t \cdot \mathbf{1}_{N_i \times N_j})]}_{P_{ki}^t} \cdot I_{kj}^t \\ & + \underbrace{\sum_{j=1}^M [\langle \text{MLP}_i(X_i^t), (\text{MLP}_j(X_j^t))^\top \rangle \odot \langle \tilde{E}_{vi}, \tilde{E}_{vj}^\top \rangle]}_{Q_{ki}^t} \cdot I_{kj}^t. \end{aligned}$$

Consequently, the evaluation of Eq. (21) reduces to the computation of  $P_{ki}^t$  and  $Q_{ki}^t$ .

The first component,  $P_{ki}^t$ , can be reformulated as:

$$P_{ki}^t = \eta_{\tau i}^t \cdot \left[ \text{MLP}_i(X_i^t) \cdot \sum_{j=1}^M ((\text{MLP}_j(X_j^t))^\top \cdot I_{kj}^t) \right]. \quad (22)$$

Therefore, the computation of  $P_{ki}^t$  proceeds as follows:

- **Step 1.** Each client  $C_j$  locally computes  $\mathcal{P}_{kj}^t = (\text{MLP}_j(X_j^t))^\top \cdot I_{kj}^t$  and sends it to the server.

- **Step 2.** The server aggregates these terms by computing  $\sum_{j=1}^M \mathcal{P}_{kj}^t$  and broadcasts the sum back to each client.
- **Step 3.** Each client  $C_i$  calculates  $P_{ki}^t = \eta_{\tau i}^t \cdot \text{MLP}_i(X_i^t) \cdot \sum_{j=1}^M \mathcal{P}_{kj}^t$ .

The component  $Q_{ki}^t$  is more challenging to compute in a distributed fashion due to the coupling of inner and Hadamard product operations across different clients' feature matrices and node embeddings. To address this, we introduce an algebraic transformation that decouples the client-side computations. This transformation is formalized in Theorem 1.

**Theorem 1.** Let  $w_{i1}, w_{i2}, \dots, w_{id}$  be the columns of  $\text{MLP}_i(X_i^t) \in \mathbb{R}^{N_i \times d}$  and  $v_{i1}, v_{i2}, \dots, v_{id_N}$  be the columns of  $\tilde{E}_{vi} \in \mathbb{R}^{N_i \times d_N}$ . Define  $\Gamma : \mathbb{R}^{N_i \times d} \times \mathbb{R}^{N_i \times d_N} \rightarrow \mathbb{R}^{N_i \times (d_N \cdot d)}$  as

$$\Gamma(\text{MLP}_i(X_i^t), \tilde{E}_{vi}) = [w_{i1} \odot v_{i1}, \dots, w_{i1} \odot v_{id_N}, w_{i2} \odot v_{i1}, \dots, w_{i2} \odot v_{id_N}, \dots, w_{id} \odot v_{i1}, \dots, w_{id} \odot v_{id_N}].$$

Then, we have:

$$\begin{aligned} \langle \text{MLP}_i(X_i^t), (\text{MLP}_j(X_j^t))^\top \rangle \odot (\tilde{E}_{vi} \cdot \tilde{E}_{vj}^\top) \\ = \Gamma(\text{MLP}_i(X_i^t), \tilde{E}_{vi}) \cdot \Gamma(\text{MLP}_j(X_j^t), \tilde{E}_{vj})^\top. \end{aligned}$$

The proof of Theorem 1 is shown in Appendix B in supplementary materials.

Based on Theorem 1, we reformulate component  $Q_{ki}^t$  as:

$$\begin{aligned} Q_{ki}^t = & \Gamma(\text{MLP}_i(X_i^t), \tilde{E}_{vi}) \\ & \times \sum_{j=1}^M \left( \Gamma(\text{MLP}_j(X_j^t), \tilde{E}_{vj})^\top \cdot I_{kj}^t \right). \quad (23) \end{aligned}$$

This permits the computation of  $Q_{ki}^t$  as follows:

- **Step 1.** Each client  $C_j$  computes  $\mathcal{Q}_{kj}^t = \Gamma(\text{MLP}_j(X_j^t), \tilde{E}_{vj})^\top \cdot I_{kj}^t$  and sends it to the server.
- **Step 2.** The server aggregates these terms to form  $\sum_{j=1}^M \mathcal{Q}_{kj}^t$  and broadcasts the sum.
- **Step 3.** Each client  $C_i$  calculates  $Q_{ki}^t = \Gamma(\text{MLP}_i(X_i^t), \tilde{E}_{vi}) \cdot \sum_{j=1}^M \mathcal{Q}_{kj}^t$ .

Following the steps above, client  $i$  computes  $\tilde{L}_{ki}^t$  while adhering to data locality constraints. Subsequently, client  $i$  evaluates Eqs. (13)–(16), thereby finishing the client-side spatio-temporal modeling.

### Training Process

After the client-side spatio-temporal modeling, each client evaluates the loss function and optimizes both global and private model parameters using stochastic gradient descent. The private parameters remain local, whereas the global parameters are uploaded to the server for aggregation via the FedAvg algorithm.

### Time and Communication Complexity Analysis

**Time Complexity.** The time complexity of the FedSTGD is  $O(N \cdot d_N \cdot d^2)$ , which scales linearly with the graph size  $N$ , indicating efficient performance for large-scale graphs.

**Communication Complexity.** The communication complexity of FedSTGD is given by  $O(|\theta| \cdot M \cdot R_g + (M \cdot d_N \cdot d^2) \cdot R_l)$ , where  $R_g$  is the number of global rounds and  $R_l$  is the number of local rounds per client. The first term reflects model aggregation across clients, while the second captures client-side spatio-temporal modeling.

## Experiments

In this section, we evaluate the performance of FedSTGD on a series of experiments over four real-world datasets, which are summarized to answer the following research questions:

- **RQ1:** How does the TFF performance of FedSTGD compare to various baseline models?
- **RQ2:** What is the contribution of each designed module to the overall model performance?
- **RQ3:** How do variations in hyperparameters influence the performance of the model?

### Experimental Setup

**Datasets** To evaluate the performance of FedSTGD, we utilize four real-world datasets: HZMetro (Liu et al. 2022), SHMetro (Liu et al. 2022), NYC-Bike<sup>1</sup>, and NYC-Taxi<sup>2</sup>. Each dataset offers unique traffic flow dynamics. Detailed descriptions of these datasets are provided in Table 1.

Datasets	Nodes	Size	TimeStep	TimeRange
HZMetro	80	2.35M	15 min	Jan 1-25, 2019
SHMetro	288	811.8M	15 min	Jul 1-Sep 30, 2016
NYC-Bike	250	30.7M	30 min	Apr 1-Jun 30, 2016
NYC-Taxi	266	35M	30 min	Apr 1-Jun 30, 2016

Table 1: Descriptions of the Datasets.

**Implementation Details** We conducted all experiments on a computational platform equipped with an Intel Xeon Gold 5215 CPU (2.50 GHz) and an NVIDIA A100 40 GB GPU. The Adam optimizer was employed, configured with an initial learning rate of  $10^{-3}$ , an  $L_2$  regularization penalty of  $10^{-4}$ , and a learning rate decay factor of 0.3 applied at epoch milestones [5, 20, 40, 70, 90]. A batch size of 16 was used throughout. For the graph-based model, the hyperparameter  $\alpha$  was set to 0.3, with node embedding dimension  $d_N$  and temporal embedding dimension  $d_T$  configured to 64. The federated nonlinear computation decomposition module employed an MLP architecture, structured with an initial layer mapping inputs to four neurons, followed by a ReLU activation function, and a final Softmax layer to produce output probabilities. The same MLP configuration was adopted for the nonlinear function in the graph node embedding augmentation.

**Evaluation Metrics** Performance was evaluated using the following metrics: Root Mean Square Error (RMSE), Mean Absolute Error (MAE), Mean Absolute Percentage Error (MAPE).

<sup>1</sup><https://www.citibikenyc.com/system-data>

<sup>2</sup><https://www.nyc.gov/site/tlc/about/tlc-trip-record-data.page>

**Baselines.** We compare FedSTGD with the following 9 baselines belonging to three classes. (1) **predefined graph-based methods** (FLoS (Wang et al. 2022), MFVST-GNN (Liu et al. 2023)); (2) **publicly accessible graph-based methods** (CTFL (Zhang et al. 2022), FCGCN (Xia, Jin, and Chen 2023)); (3) **inter-client spatial dependency recovery methods** (FASTGNN (Zhang et al. 2021), FedGTP (Yang et al. 2024), FedSTG (Zhang et al. 2025), FUELS (Liu et al. 2025), DSTGCRN (Pham et al. 2025)). Details are provided in Appendix C in the supplementary materials.

### Performance Comparison (RQ1)

Table 2 presents the results from federated graph-based baselines. The best results are highlighted in bold. Several key conclusions can be drawn:

- FedSTGD consistently outperforms all competing methods across every evaluation metric and dataset examined.
- Among the predefined graph-based methods, several approaches rely on predefined connectivity matrices to characterize spatial dependencies, resulting in poor forecasting performance.
- Among the publicly accessible graph-based methods, many overlook privacy concerns by treating graph structure information as publicly available. However, the inherent limitations of the graph models employed lead to consistently poor performance.
- Among the inter-client spatial dependency modeling methods, some of them attempt to approximate the inter-client spatial dependency, yet empirical results indicate insufficient approximation capabilities. The FedGTP framework fully reconstructs inter-client spatial dependencies, but its focus on static dependencies constrains its predictive accuracy, rendering it inadequate for capturing dynamic spatial interactions.

### Ablation Study (RQ2)

To assess the impact of individual components within FedSTGD, we conduct an ablation study by constructing several variants of the framework:

- **w/o GNEA:** This variant ablates the modeling of graph node embedding augmentation.
- **w/o -,All:** This variant ablates all the modeling of inter-client spatial dependencies, both static and dynamic.
- **w/o Dyn,Dyn:** This variant ablates the modeling of inter-client and intra-client dynamic spatial dependencies.
- **w/o Dyn,All:** This variant ablates the modeling of intra-client dynamic spatial dependencies and inter-client spatial dependencies.
- **w/o All,All:** This variant ablates all the modeling of inter-client spatial dependencies.

Table 3 presents the comparison of FedSTGD and its variants on HZMetro and NYC-Taxi datasets. From this comparison, we can draw several conclusions: (1) FedSTGD consistently achieves the best performance relative to its variants, underscoring the effectiveness of its full configuration. (2)

Method	HZMetro			SHMetro			NYC Bike			NYC Taxi		
	RMSE	MAE	MAPE%	RMSE	MAE	MAPE%	RMSE	MAE	MAPE%	RMSE	MAE	MAPE%
MFSTGNN	58.90	34.03	22.50	78.96	35.85	19.35	2.95	1.86	19.90	19.08	9.49	18.92
FLoS	52.69	27.96	20.93	75.38	35.31	19.97	2.89	1.81	18.92	17.79	10.23	18.83
FCGCN	54.67	29.06	23.73	79.33	36.61	22.04	2.87	1.62	18.70	18.77	10.63	18.56
CTFL	54.41	28.99	21.60	84.30	41.79	20.30	2.89	1.65	18.71	17.03	8.74	16.94
FASTGNN	68.37	36.45	23.42	101.22	49.74	24.96	2.90	1.68	18.68	19.17	10.26	19.87
FedSTG	56.31	29.76	22.01	68.26	31.91	22.41	2.86	1.81	18.81	15.49	7.50	18.63
DSTGCRN	56.12	29.13	22.76	71.01	33.60	22.98	3.04	1.89	19.87	26.15	9.32	20.65
FUELS	54.98	30.81	22.49	66.57	29.23	21.67	2.86	1.72	18.92	12.19	6.56	17.90
FedGTP	51.14	26.95	19.86	64.92	26.88	18.04	2.85	1.72	18.63	10.74	5.23	17.52
<b>FedSTGD</b>	<b>39.61</b>	<b>23.69</b>	<b>14.91</b>	<b>50.08</b>	<b>24.24</b>	<b>17.36</b>	<b>2.73</b>	<b>1.61</b>	<b>18.52</b>	<b>9.03</b>	<b>4.93</b>	<b>17.17</b>

Table 2: Comparison of performance on the traffic flow forecasting task between FedSTGD and baselines.

Model	HZMetro			NYC Taxi		
	RMSE	MAE	MAPE%	RMSE	MAE	MAPE%
w/o GNEA	41.64	24.41	15.13	9.98	5.36	19.03
w/o All,All	62.57	27.93	21.94	17.91	10.43	20.82
w/o -,All	43.23	25.21	15.60	9.58	5.24	18.21
w/o Dyn,Dyn	51.14	26.95	19.86	10.74	5.23	14.12
w/o Dyn,All	52.66	26.50	21.67	9.87	5.49	18.91
<b>FedSTGD</b>	<b>39.61</b>	<b>23.69</b>	<b>14.91</b>	<b>9.03</b>	<b>4.93</b>	<b>17.17</b>

Table 3: Ablation study on HZMetro and NYC Taxi datasets.

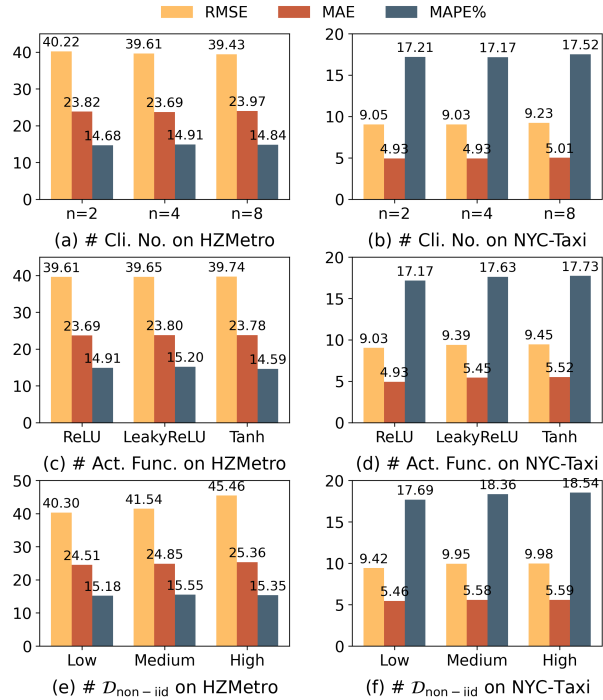


Figure 3: Parameters Sensitivity Analysis on HZMetro and NYC-Taxi Datasets.

The variant “w/o GNEA” underperforms FedSTGD, demonstrating the critical role of the graph node embedding augmentation module. (3) The variant “w/o All,All” underperforms FedSTGD, demonstrating the critical role of spatial modeling. (4) The variant “w/o -,All” underperforms FedSTGD, emphasizing the importance of modeling inter-client spatial dependencies. (5) The variants “w/o Dyn,Dyn” and “w/o Dyn,All” underperform FedSTGD, underscoring the necessity of modeling dynamic spatial dependencies at both inter-client and intra-client levels.

### Sensitivity Analysis (RQ3)

Figure 3 presents the results of a hyperparameter sensitivity analysis for our FedSTGD framework on the HZMetro and NYC-Taxi datasets. This analysis involves varying the number of clients across the set  $\{2, 4, 8\}$ , evaluating different activation functions within the MLP—including ReLU, LeakyReLU, Tanh—and considering three distinct levels of data heterogeneity (i.e.,  $\mathcal{D}_{non-iid}$ ). Several key insights can be drawn from these experiments: (1) The model’s performance demonstrates overall robustness as the number of clients and the degree of data heterogeneity increase, though with a modest decline. (2) Variations in the MLP activation functions have a negligible impact on the model’s outcomes. These findings emphasize the stability of FedSTGD under diverse hyperparameter settings, underscoring its suitability for scalable applications in federated learning environments.

### Conclusion

In this paper, we propose FedSTGD, a novel federated spatio-temporal graph learning framework that effectively captures dynamic inter-client spatial dependencies while adhering to data locality constraints. Comprehensive experiments demonstrate the superior performance of FedSTGD over state-of-the-art baselines, underscoring the critical role of reconstructing dynamic spatial dependencies in federated learning. Ablation studies validate the efficacy of each module in modeling dynamic inter-client spatial dynamics, whereas sensitivity analyses affirm the robustness of FedSTGD to hyperparameter variations.

## Acknowledgments

This work is supported by the fundamental research funds for the central universities, and the research funds of Renmin University of China (23XNL014).

## References

- Chrysos, G. G.; Moschoglou, S.; Bouritsas, G.; Deng, J.; Panagakos, Y.; and Zafeiriou, S. 2021. Deep polynomial neural networks. *IEEE transactions on pattern analysis and machine intelligence*, 44(8): 4021–4034.
- EP and Council. 2016. Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data, and repealing Directive 95/46/EC (General Data Protection Regulation). <https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX%3A32016R0679>. Accessed: 2025-07-13.
- Fan, F.-L.; Li, M.; Wang, F.; Lai, R.; and Wang, G. 2023. On expressivity and trainability of quadratic networks. *IEEE Transactions on Neural Networks and Learning Systems*.
- Fang, Z.; Long, Q.; Song, G.; and Xie, K. 2021. Spatial-temporal graph ode networks for traffic flow forecasting. In *Proceedings of the 27th ACM SIGKDD conference on knowledge discovery & data mining*, 364–373.
- Ge, J.; Xu, G.; Lu, J.; Xu, C.; Sheng, Q. Z.; and Zheng, X. 2024. FedAGA: A federated learning framework for enhanced inter-client relationship learning. *Knowledge-Based Systems*, 286: 111399.
- Kong, W.; Guo, Z.; and Liu, Y. 2024. Spatio-temporal pivotal graph neural networks for traffic flow forecasting. In *Proceedings of the AAAI conference on artificial intelligence*, volume 38, 8627–8635.
- Li, M.; and Zhu, Z. 2021. Spatial-temporal fusion graph neural networks for traffic flow forecasting. In *Proceedings of the AAAI conference on artificial intelligence*, volume 35, 4189–4196.
- Liu, L.; Chen, J.; Wu, H.; Zhen, J.; and Li, G. 2022. Physical-Virtual Collaboration Modeling for Intra- and Inter-Station Metro Ridership Prediction. *IEEE Transactions on Intelligent Transportation Systems*, 23(4): 3377–3391.
- Liu, L.; Tian, Y.; Chakraborty, C.; Feng, J.; and Pei, Q. 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, Q.; Sun, S.; Liang, Y.; Liu, M.; and Xue, J. 2025. Personalized federated learning for spatio-temporal forecasting: A dual semantic alignment-based contrastive approach. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 39, 12192–12200.
- Lu, L.; Jin, P.; Pang, G.; Zhang, Z.; and Karniadakis, G. E. 2021. Learning nonlinear operators via DeepONet based on the universal approximation theorem of operators. *Nature machine intelligence*, 3(3): 218–229.
- Ma, M.; Hu, J.; Jensen, C. S.; Teng, F.; Han, P.; Xu, Z.; and Li, T. 2024. Learning time-aware graph structures for spatially correlated time series forecasting. In *2024 IEEE 40th International Conference on Data Engineering (ICDE)*, 4435–4448. IEEE.
- 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. PMLR.
- Meng, C.; Rambhatla, S.; and Liu, Y. 2021. Cross-Node Federated Graph Neural Network for Spatio-Temporal Data Modeling. In *KDD '21: The 27th ACM SIGKDD Conference on Knowledge Discovery and Data Mining*, 1202–1211. ACM. ISBN 978-1-4503-8332-5.
- Pham, T.; Furno, A.; Chamroukhi, F.; and Oukhellou, L. 2025. Federated Dynamic Modeling and Learning for Spatiotemporal Data Forecasting. *arXiv preprint arXiv:2503.04528*.
- Qi, T.; Chen, L.; Li, G.; Li, Y.; and Wang, C. 2023. FedAGCN: A Traffic Flow Prediction Framework Based on Federated Learning and Asynchronous Graph Convolutional Network. *Applied Soft Computing*, 138: 110175.
- Tian, Y.; Weng, H.; and Feng, Y. 2024. Towards the Theory of Unsupervised Federated Learning: Non-asymptotic Analysis of Federated EM Algorithms. In *International Conference on Machine Learning*, 48226–48279. PMLR.
- Wang, H.; Zhang, R.; Cheng, X.; and Yang, L. 2022. Federated Spatio-Temporal Traffic Flow Prediction Based on Graph Convolutional Network. *2022 14th International Conference on Wireless Communications and Signal Processing (WCSP)*, 221–225.
- Wei, S.; Tong, Y.; Zhou, Z.; Xu, Y.; Gao, J.; Wei, T.; He, T.; and Lv, W. 2025. Federated reasoning LLMs: a survey. *Frontiers Comput. Sci.*, 19(12): 1912613.
- Xia, M.; Jin, D.; and Chen, J. 2023. Short-Term Traffic Flow Prediction Based on Graph Convolutional Networks and Federated Learning. *IEEE Transactions on Intelligent Transportation Systems*, 24(1): 1191–1203.
- Yang, L.; Chen, W.; He, X.; Wei, S.; and Xu, Y. 2024. FedGTP: Exploiting Inter-Client Spatial Dependency in Federated Graph-Based Traffic Prediction. In *KDD '24, KDD '24*, 6105–6116. Association for Computing Machinery. ISBN 979-8-4007-0490-1.
- Yuan, X.; Chen, J.; Yang, J.; Zhang, N.; and Yang, T. 2023. FedSTN: Graph Representation Driven Federated Learning for Edge Computing Enabled Urban Traffic Flow Prediction. *IEEE Transactions on Intelligent Transportation Systems*, 24(8): 8738–8748.
- Zhang, C.; Zhang, S.; Yu, J. J. Q.; and Yu, S. 2021. 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 (WCNC)*, 2041–2046.

Zhang, Y.; Wang, X.; Yu, X.; Yang, K.; Zhou, Z.; and Wang, Y. 2025. FedSTG: Breaking through Spatio-Temporal Data Silos with Federated Graph Learning. In *Companion Proceedings of the ACM on Web Conference 2025*, 1534–1538.

Zhou, F.; Liu, S.; Fujita, H.; Hu, X.; Zhang, Y.; Wang, B.; and Wang, K. 2024. Fault diagnosis based on federated learning driven by dynamic expansion for model layers of imbalanced client. *Expert Systems with Applications*, 238: 121982.