



FedSTG: Breaking through Spatio-Temporal Data Silos with Federated Graph Learning

Yudong Zhang

School of Artificial Intelligence and
Data Science, University of Science
and Technology of China
Hefei, China
zyd2020@mail.ustc.edu.cn

Xu Wang*

School of Software Engineering,
University of Science and Technology
of China
Hefei, China
Suzhou Institute for Advanced
Research, University of Science and
Technology of China
Suzhou, China
wx309@ustc.edu.cn

Xuan Yu

School of Software Engineering,
University of Science and Technology
of China
Hefei, China
yx2024@mail.ustc.edu.cn

Kuo Yang

School of Artificial Intelligence and
Data Science, University of Science
and Technology of China
Hefei, China
yangkuo@mail.ustc.edu.cn

Zhengyang Zhou

School of Software Engineering,
University of Science and Technology
of China
Hefei, China
Suzhou Institute for Advanced
Research, University of Science and
Technology of China
Suzhou, China
zzy0929@ustc.edu.cn

Yang Wang*

School of Software Engineering,
University of Science and Technology
of China
Hefei, China
Suzhou Institute for Advanced
Research, University of Science and
Technology of China
Suzhou, China
angyan@ustc.edu.cn

Abstract

Spatio-temporal graph (STG) forecasting plays a crucial role in various urban computing applications driven by Web systems. However, conventional STG forecasting methods, which rely on centralized data processing, have posed obstacles to real-world applications in numerous data silo scenarios. Given this, some researchers have preliminarily explored incorporating federated learning into STG forecasting. Though promising, the existing methods encounter two critical issues, i.e., *dynamics* and *heterogeneity* in STGs. To this end, we propose a novel personalized Federated learning framework for Spatio-Temporal Graph forecasting (called **FedSTG**), along with an evolutionary graph learning module and a personalized federated aggregation algorithm based on evolving temporal patterns in the framework. Our approach enhances the STG learning capabilities in the federated paradigm, yield addressing the problem of STG forecasting in data silo scenarios. Extensive experiments on four real-world datasets in STG forecasting have been conducted to demonstrate the superiority of our approach.

*Corresponding authors.

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for components of this work owned by others than the author(s) must be honored. Abstracting with credit is permitted. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from permissions@acm.org.

WWW Companion '25, Sydney, NSW, Australia

© 2025 Copyright held by the owner/author(s). Publication rights licensed to ACM.
ACM ISBN 979-8-4007-1331-6/25/04
<https://doi.org/10.1145/3701716.3715562>

CCS Concepts

• **Information systems** → **Spatial-temporal systems**; **Data mining**; • **Computing methodologies** → **Distributed computing methodologies**.

Keywords

Spatio-temporal graph; data silos; federated learning

ACM Reference Format:

Yudong Zhang, Xu Wang, Xuan Yu, Kuo Yang, Zhengyang Zhou, and Yang Wang. 2025. FedSTG: Breaking through Spatio-Temporal Data Silos with Federated Graph Learning. In *Companion Proceedings of the ACM Web Conference 2025 (WWW Companion '25)*, April 28-May 2, 2025, Sydney, NSW, Australia. ACM, New York, NY, USA, 5 pages. <https://doi.org/10.1145/3701716.3715562>

1 Introduction

Spatio-Temporal Graph (STG) is an effective tool for modeling the physical world with structural information, especially in web networks and smart cities. Therefore, research on STG forecasting has attracted much attention from enormous researchers in the past decades [10, 13]. Graph neural networks (GNNs) are at the forefront of deep learning techniques capable of learning structural information from complex topologies of data, which have been widely applied in various STG forecasting tasks [9]. Specifically, STGCN [12] applies ChebNet graph convolution to extract spatio-temporal correlations in traffic data. Graph WaveNet [11] captures the spatial correlation with a diffusion convolution layer and learns the temporal correlation using TCN. Evidently, the aforementioned methods highly rely on large-scale data for centralized training,

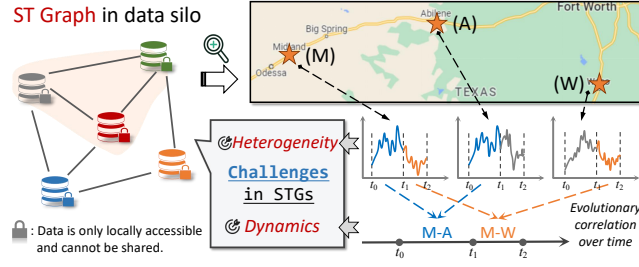


Figure 1: A case of STG (COVID-19 cases in Texas hospitals). The data distribution presents heterogeneity in spatio-temporal dimension, and the structural relationship between nodes also evolves dynamically over time.

which inevitably puts forward strict requirements for the centralized management of data. Nowadays, with the increasing voice of data isolation, data is often held by various distributed nodes and cannot be shared with each other, which poses a new challenge for the realistic application of STG forecasting.

In response to this, some work has preliminarily incorporated Federated Learning (FL) into STG forecasting to enable distributed nodes to collaboratively train models without raw data sharing [14]. CNFGNN [7] aggregates local temporal embeddings uploaded from clients and employs GNNs to obtain spatial embeddings, which are sent back to the corresponding clients for forecasting. SFL [1] learns both the global and personalized models simultaneously using client-wise relation graphs and clients' private data. Though intriguing, there are still two critical **challenges** that remain to be further addressed: $\hookrightarrow$ **Distribution heterogeneity**: As depicted in Fig. 1, the distribution of data in STGs is inconsistent in temporal and spatial dimensions, and as a result, a unified global model cannot be applicable to the local data of each node. Existing FL work, like CNFGNN, mostly uses Federated Averaging (FedAvg) [6] to train a shared global model for all nodes through equally aggregating the parameters of the client models, so that the models in all clients are indistinguishable, which fail to achieve local model personalization and lead to undesirable performance. $\hookrightarrow$ **Dynamic graph structure**: In addition to the pre-defined structure based on geographical proximity, there is also a potential dynamic structural relationship among nodes that evolves over time. A sample illustration is given in Fig. 1, the strong correlation (i.e., connectivity in STG) among these three nodes changes from "M-A" ($t_0 \sim t_1$) to "M-W" ($t_1 \sim t_2$). Previous FL methods only focus on capturing static graph structure, which discards the evolutionary patterns of STG structure, thereby limiting the propagation of global temporal patterns in space and the capacities of federated STG learning.

Consequently, developing a novel personalized FL framework, which simultaneously solves the problem of distribution heterogeneity and dynamic graph structure, would drastically promote the effectiveness and practicability of STG forecasting in data silo scenarios. Meanwhile, in real-world federated spatio-temporal scenarios, each node is an isolated organization holding only its own data. Unlike high-performance central servers, the nodes have limited computing and storage capacity and do not support massive

data processing. Generally, each node can only access recent few-shot spatio-temporal data (e.g., several months of data), and such incremental data hinders an obstacle to long-term spatio-temporal pattern modeling. According to Fig. 1, we discover that the evolutionary rule of long-term spatio-temporal patterns will bring insights into understanding spatio-temporal heterogeneity and dynamics. Therefore, we will start with long-term spatio-temporal pattern mining to address the above challenges.

To this end, we propose a novel personalized Federated learning framework for Spatio-Temporal Graph forecasting (called **FedSTG**). Targeting the first challenge, we conduct Federated Aggregation with Temporal Pattern of clients (FedATP), which can adaptively aggregate parameters based on temporal pattern characteristics to obtain the personalized client model. To overcome the second issue, we introduce an evolutionary graph learning module based on dynamic graph structure in the server-side model, which can facilitate the capacity of graph structure modeling in STGs. The main **contributions** of this paper lie in four aspects:

- ❶ We propose a novel federated STG learning paradigm. Our FedSTG provides an ingenious marriage of FL and STG forecasting, and extends STGs to data silo scenarios.
- ❷ We develop a personalized aggregation algorithm called FedATP dedicated to STGs. It alleviates the problem of data heterogeneity and facilitates local model personalization.
- ❸ We extract complex graph structure among clients from temporal evolution perspectives, which emphasizes dynamic inter-client topology and enhances STG modeling in FL setting.
- ❹ The experiments show impressive results brought by our novel setting of FL for STG forecasting, where our FedSTG can consistently outperform other baselines.

2 Preliminary

Definition 1 (Spatio-Temporal Graph). Given the data held by N correlated organizations (nodes). We denote the topological structure of nodes as a weighted graph $G = (V, E, A)$, where V is the set of $|V| = N$ nodes¹, E is the set of edges, and $A \in \mathbb{R}^{N \times N}$ is a weighted adjacency matrix representing the proximities among nodes, e.g., measured by the distance of node pairs. The incremental data in organization i is denoted as $\mathbf{X}_{\tau,i} \in \{\mathbf{X}_{1,i}, \mathbf{X}_{2,i}, \dots, \mathbf{X}_{\mathcal{T},i}\}$, where $\tau \in \{1, \dots, \mathcal{T}\}$ is the time interval, meaning the data keeps updating over time. Notably, $\mathbf{X}_{\tau,i} \in \mathbb{R}^{T \times D}$ is only locally accessible by organization i , where D is the feature dimension and T denotes the past time steps in a time interval.

Problem 1 (Optimization Objective in FL). Our goal is to forecast the future STG data held by the N organizations, i.e., learning a function f_{θ_i} that can predict future H time steps based on the past T time steps. In FL setting, the objective function is as follows:

$$\min_{\theta_{\tau,1}, \theta_{\tau,2}, \dots, \theta_{\tau,N}} \frac{1}{N} \sum_{i=1}^N \mathcal{L}_i(f_{\theta_{\tau,i}}), \quad (1)$$

where $\theta_{\tau,i}$ is the parameters of the local model in client i in the τ -th interval, $\mathcal{L}_i$ denotes the loss function of local model in client i .

3 Methodology

The core idea of our proposal is to establish an effective personalized model for each client without data sharing as illustrated in Fig. 2.

¹In this paper, each node denote a data silo, corresponding to a client in FL system.

Firstly, We preserve and update evolutionary temporal patterns in clients using temporal pattern bank. Next, we extract the static and dynamic graph structure among clients on the server side. Then, we conduct parameter aggregation based on the temporal patterns of clients to obtain personalized models applicable to local tasks.

3.1 Temporal Pattern Modeling

Long- and short-term temporal patterns exist simultaneously in spatio-temporal data. Generally, each client can only access recent few-shot data, which hinders an obstacle to long-term pattern modeling. Inspired by Fig. 1, we discover that the evolutionary rule of long-term patterns will bring insights into the solution of heterogeneity and dynamics. Therefore, we start with long-term temporal pattern mining to facilitate downstream tasks.

3.1.1 Temporal Pattern Bank. We adopt a Temporal Pattern Bank (TP-Bank) module to store and update the significant patterns. TP-Bank is a parameterized embedding matrix $\mathbb{B} \in \mathbb{R}^{K \times d}$, where K represents the number of temporal pattern categories and d is the dimension of the pattern representation. In particular, we employ GRU encoder depicted in Fig. 2 to obtain the short-term temporal embedding $\mathbf{h}_{\tau,i}$ of client i in time interval τ , and then we conduct query projection with $\mathbf{h}_{\tau,i}$ to get the query vector for the attention mechanism:

$$\mathbf{h}_{\tau,i} = \text{GRU}_i(\mathbf{X}_{\tau,i}, \mathbf{h}_{\tau,i}^{(0)}), \quad \mathbf{q}_{\tau,i} = \mathbf{W}_{\tau}^T \mathbf{h}_{\tau,i} + \mathbf{b}_{\tau}^T, \quad (2)$$

where $\mathbf{h}_{\tau,i}^{(0)}$ is a zero-valued initial hidden state vector.

Next, we calculate the similarity score between the query vector $\mathbf{q}_{\tau,i}$ and the patterns stored in the TP-Bank $\mathbb{B}_{\tau,i}$ by the dot product operation and extract the corresponding patterns from $\mathbb{B}_{\tau,i}$:

$$\begin{aligned} p_{\tau,i}(k) &= \frac{\exp(\langle \mathbf{q}_{\tau,i}, \mathbb{B}_{\tau,i}(k) \rangle)}{\sum_{k'=1}^K \exp(\langle \mathbf{q}_{\tau,i}, \mathbb{B}_{\tau,i}(k') \rangle)}, \\ \mathbf{z}_{\tau,i} &= \sum_{k=1}^K p_{\tau,i}(k) * \mathbb{B}_{\tau,i}(k), \end{aligned} \quad (3)$$

where $\mathbb{B}_{\tau,i}(k)$ denotes the k -th pattern category.

3.1.2 Optimization. Then, we concatenate the long-term representation $\mathbf{z}_{\tau,i}$ with the short-term representation (output of GRU decoder) of client i , and input the enhanced representation to a fully connected layer to predict future values in the next H steps. We use mean absolute error (MAE) to measure the predicted results:

$$\mathcal{L}_p(\theta_{\tau,i}) = \|\hat{\mathbf{X}}_{\tau,i} - \mathbf{X}_{\tau,i}\|, \quad (4)$$

where $\hat{\mathbf{X}}_{\tau,i}$, $\mathbf{X}_{\tau,i}$ are predicted value and ground truth of client i .

Notably, TP-Bank can be automatically updated over time end-to-end, adopting parameter regularization inspired by [2]. The core idea is to preserve parameters with high weights and allow unimportant parameters to be flexible for updating. Concretely, we first compute the attention coefficient between the data in the current interval $\mathbf{X}_{\tau,i}$ and pattern knowledge in the last interval $\mathbb{B}_{\tau-1,i}$ to obtain optimization matrix $\Omega_{\tau,i}$, and then utilize $\Omega_{\tau,i}$ to constrain the optimization direction of pattern updating in TP-Bank:

$$\begin{aligned} \mathcal{L}_k &= \|\mathbb{B}_{\tau,i} - \mathbb{B}_{\tau-1,i} \Omega_{\tau,i}\|^2, \\ \Omega_{\tau,i} &= \sigma \left((\mathbf{X}_{\tau,i} \mathbf{U}_1)^\top \mathbf{U}_2 (\mathbf{U}_3 \mathbb{B}_{\tau-1,i}) \right), \end{aligned} \quad (5)$$

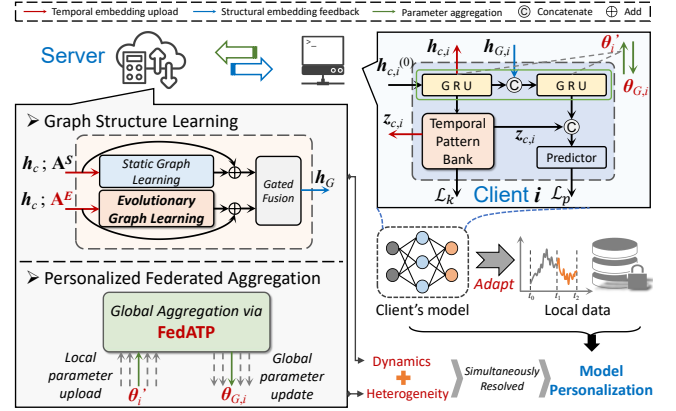


Figure 2: Overview of FedSTG: Personalized federated learning framework for STG forecasting across data silos.

where $\mathbf{U}_1$, $\mathbf{U}_2$ and $\mathbf{U}_3$ are the learnable parameters.

In this way, the long-term temporal pattern is retained and updated adaptively according to the current spatio-temporal context. Thus, the final optimization objective of client i is:

$$\min_{\theta_{\tau,i}} \mathcal{L}_p(\theta_{\tau,i}) + \alpha \mathcal{L}_k(\theta_{\tau,i}), \quad (6)$$

where α is the regularization weights for balancing loss.

3.2 Graph Structure Learning

As graph structure constantly evolves depicted in Fig. 1, modeling both static and dynamic graph structures is crucial for STG learning. We propose an evolutionary graph learning module based on the variation of temporal distribution, to effectively implement the modeling of federated dynamic graphs without raw and extra data.

3.2.1 Static and Evolutionary Graph Construction. Primarily, we construct the pre-defined (static) and dynamic graph to represent the structural relationships among nodes. We compute the connectivity (edge weight) $\mathbf{A}_{i,j}^S$ between node v_i and v_j ($v_i, v_j \in V, i \neq j$) via a threshold Gaussian kernel weighting function for the static graph. For the evolutionary graph, we obtain the connectivity $\mathbf{A}_{i,j}^E$ by calculating the similarity of temporal patterns in node v_i and v_j :

$$\begin{aligned} \mathbf{A}_{i,j}^S &= \begin{cases} \exp\left(-\frac{d(i,j)^2}{\sigma^2}\right), & d(i,j) \leq \kappa \\ 0, & \text{otherwise} \end{cases}, \\ \mathbf{A}_{i,j}^E &= \begin{cases} \text{Sim}(\mathbf{z}_i, \mathbf{z}_j), & \text{Sim}(\mathbf{z}_i, \mathbf{z}_j) \geq \beta \\ 0, & \text{otherwise} \end{cases}, \end{aligned} \quad (7)$$

where $d(\cdot, \cdot)$ is Euclidean distance, $\text{Sim}(\cdot, \cdot)$ is similarity measure, κ and β are threshold parameters to control neighborhood size, and $\mathbf{z}_i, \mathbf{z}_j$ are the long-term patterns of v_i and v_j stored in TP-Bank.

3.2.2 Graph Fused Learning. Next, we capture the structural information in STG with temporal embeddings $\mathbf{h}_c = \{\mathbf{h}_{c,1}, \dots, \mathbf{h}_{c,N}\}$ uploaded by clients. By performing two parallel Graph Convolutional Networks (GCNs) on both static and evolutionary graphs, we obtain

²We omit τ for succinctness in the following.

Table 1: Average performance comparison (MAE and RMSE) of FedSTG and all alternative baselines on the four datasets.

(a) Air Quality Forecasting Task.							(b) Epidemic Spread Forecasting Task.						
Datasets	Methods	Settings	6-step ahead		12-step ahead		Datasets	Methods	Settings	6-step ahead		12-step ahead	
			MAE	RMSE	MAE	RMSE				MAE	RMSE	MAE	RMSE
KnowAir	GRU	Centralized	21.33	28.29	29.73	38.22	Covid-TX	GRU	Centralized	33.82	53.86	46.81	77.64
	STGCN	Centralized	20.82	27.59	27.81	37.47		STGCN	Centralized	32.83	51.79	45.38	73.60
	TrafficStream	Centralized	19.55	25.45	26.11	34.48		TrafficStream	Centralized	28.19	44.34	39.70	60.86
	FedGRU	Federated	23.48	31.69	31.38	41.32		FedGRU	Federated	43.82	68.91	61.71	94.59
	SFL	Federated	22.39	30.49	31.04	40.96		SFL	Federated	39.99	65.18	57.02	92.00
	CNFGNN	Federated	22.08	30.26	30.64	39.59		CNFGNN	Federated	38.13	61.63	52.61	88.73
	FedSTG (ours)	Federated	20.79	28.36	29.14	38.12		FedSTG (ours)	Federated	35.90	57.22	48.70	84.93
AirBJ	GRU	Centralized	23.44	41.17	33.94	53.90	Covid-CA	GRU	Centralized	212.82	332.22	278.30	425.85
	STGCN	Centralized	21.17	36.27	28.01	48.15		STGCN	Centralized	201.24	316.64	266.25	417.47
	TrafficStream	Centralized	17.59	31.69	31.71	42.94		TrafficStream	Centralized	181.43	286.72	240.61	393.54
	FedGRU	Federated	26.91	48.55	38.72	60.13		FedGRU	Federated	240.16	379.54	318.50	520.94
	SFL	Federated	25.66	46.31	37.23	57.81		SFL	Federated	232.28	371.21	311.09	505.65
	CNFGNN	Federated	25.21	45.49	36.63	56.89		CNFGNN	Federated	228.19	358.58	295.51	468.32
	FedSTG (ours)	Federated	23.27	40.98	32.86	52.98		FedSTG (ours)	Federated	220.31	346.39	286.79	443.76

two structural embeddings of complementary information:

$$\begin{aligned} \mathbf{h}_G^S &= \sigma((\mathbf{A}^S + \mathbf{I}^S)\mathbf{h}_c^{(I)}\mathbf{W}_1^{(I)} + \mathbf{b}_1^{(I)}), \\ \mathbf{h}_G^E &= \sigma((\mathbf{A}^E + \mathbf{I}^E)\mathbf{h}_c^{(I)}\mathbf{W}_2^{(I)} + \mathbf{b}_2^{(I)}) \end{aligned} \quad (8)$$

where $\mathbf{W}^{(I)}, \mathbf{b}^{(I)}$ are learnable, and $\mathbf{I}^S, \mathbf{I}^E$ are identity matrices.

Then, we adopt the gated fusion [15] to adaptively consider static and dynamic effects. The two structural embeddings are fused as:

$$\begin{aligned} \mathbf{h}_G &= \mathbf{g} \odot \mathbf{h}_G^S + (1 - \mathbf{g}) \odot \mathbf{h}_G^E, \\ \mathbf{g} &= \sigma(\mathbf{h}_G^S \mathbf{W}_{g,1} + \mathbf{h}_G^E \mathbf{W}_{g,2} + \mathbf{b}_g), \end{aligned} \quad (9)$$

where $\mathbf{h}_G$ is global structural embedding calculated by the server and fed back to clients, $\mathbf{g}$ is the gate, $\mathbf{W}_{g,1}, \mathbf{W}_{g,2}, \mathbf{b}_g$ are learnable.

3.3 Personalized Federated Aggregation

To enable model to adapt to local tasks, we propose a Federated Aggregation algorithm based on Temporal Pattern characteristics of clients, namely **FedATP**, which can adaptively aggregate partial parameters from local models and perform a personalized update for each client. Primarily, we calculate the similarity of temporal patterns in each node-pairs, and then compute the relevance score to obtain the relevance matrix of the clients' temporal patterns. If the long-term patterns of clients possess a higher similarity, the spatio-temporal characteristics are also of greater relevance. Therefore, we perform parameter aggregation based on the relevance matrix for each client. The process of FedATP is formulated as:

$$\theta_{G,i} \leftarrow \sum_{j=1}^N \mathbf{W}_{ij} \theta'_j, \quad \text{where } \mathbf{W}_{ij} = \frac{\exp\{\text{Sim}(\mathbf{z}_i, \mathbf{z}_j)\}}{\sum_{j=1}^N \exp\{\text{Sim}(\mathbf{z}_i, \mathbf{z}_j)\}} \quad (10)$$

where θ'_j is the parameter of GRU in client j , and $\theta_{G,i}$ is the updated parameter after aggregation. Notably, we only aggregate and update the parameter of GRU in each client, and the other parameters in local models (TP-Bank and predictor) are not involved, which remedies the issues of model personalization and overfitting.

Table 2: Detailed information of the evaluated datasets.

Properties	Datasets			
	KnowAir	AirBJ	Covid-TX	Covid-CA
Data Type	Air quality	Air quality	Epidemic spread	Epidemic spread
# of Nodes	184	35	251	55
# of Time span	01/2016-12/2016	01/2016-12/2016	01/2020-12/2020	01/2020-12/2020
Time step	3 hour	1 hour	1 day	1 day

4 Experiments

4.1 Experimental Settings

Tasks & Datasets. To investigate the effectiveness of FedSTG, we evaluate FedSTG and baselines on four datasets in STG forecasting: **Air quality forecasting:** 1) KnowAir is a PM_{2.5} concentration dataset that covers 184 cities in China [8]. 2) AirBJ is an air quality dataset [4] that contains the concentration of PM_{2.5} held by 35 stations in Beijing. **Epidemic spread forecasting:** 3) Covid-TX is a COVID-19 spread dataset that covers 251 counties (hospitalizations) in Texas [3]. 4) Covid-CA is a COVID-19 spread dataset that covers 55 counties in California [3]. Details are provided in Table 2.

Baselines. We compare FedSTG with two kinds of methods: **Centralized methods:** GRU, STGCN [12], TrafficStream [2]. **Federated methods:** FedGRU [5], SFL [1], CNFGNN [7]. We use $T=12$ historical time steps to predict the next $H=12$ time steps.

4.2 Experimental Results

Result 1 (Improvement). The average performances are summarized in Table 1. For the federated approach, the performance of FedSTG surpasses the others by a large margin across all datasets. **Result 2 (Ablation).** We conduct ablation studies with 6 variants. **① Effect of Evolutionary Graph Learning:** (i) without Graph Structure Learning (**w/o GSL**), (ii) without Evolutionary Graph Learning (**w/o EGL**). **② Effect of FedATP:** (iii) a variant adopts FedAvg (**FedSTG-Avg**), (iv) a variant aggregates all local parameters (**FedSTG-EA**). **③ Effect of TP-Bank:** i) a variant removes TP-Bank (**w/o TP+Avg**), ii) a variant omits TP-Bank, and aggregation

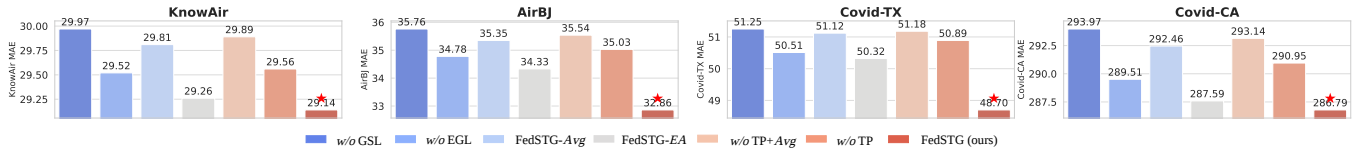


Figure 3: Ablation analysis on 12-step ahead forecasting across all datasets. MAE is reported.

is conducted with temporal embedding $\mathbf{h}_t$ (**w/o TP**). The results in Fig. 3 prove that all components in FedSTG play an important role.

5 Conclusion

In this paper, we propose a novel federated spatio-temporal graph learning framework (FedSTG) for cross-silo STG forecasting. We address two critical issues that have not been well-solved in federated STG learning, i.e., *heterogeneity* and *dynamics*. Extensive experiments have been conducted to demonstrate the superiority of FedSTG. Therefore, our work not only extends STG learning to data silo scenarios, but also improves the level of Web services.

Acknowledgments

This paper is partially supported by the National Natural Science Foundation of China (No.12227901), the Project of Stable Support for Youth Team in Basic Research Field from the Chinese Academy of Sciences (No.YSBR-005), the Natural Science Foundation of Jiangsu Province (No.BK20240460), and the open fund of the State Key Laboratory of Resources and Environmental Information System.

References

- [1] Fengwen Chen, Guodong Long, Zonghan Wu, Tianyi Zhou, and Jing Jiang. 2022. Personalized Federated Learning With a Graph. In *Proceedings of the Thirty-First International Joint Conference on Artificial Intelligence, IJCAI-22*, Lud De Raedt (Ed.). International Joint Conferences on Artificial Intelligence Organization, 2575–2582. doi:10.24963/ijcai.2022/357 Main Track.
- [2] Xu Chen, Junshan Wang, and Kunqing Xie. 2021. TrafficStream: A Streaming Traffic Flow Forecasting Framework Based on Graph Neural Networks and Continual Learning. In *Proceedings of the Thirtieth International Joint Conference on Artificial Intelligence, IJCAI-21*. 3620–3626.
- [3] Yuzhou Chen, Ignacio Segovia-Dominguez, Baris Coskunuzer, and Yulia Gel. 2022. TAMP-S2GCNets: coupling time-aware multipersistance knowledge representation with spatio-supra graph convolutional networks for time-series forecasting. In *International Conference on Learning Representations*.
- [4] Yuxuan Liang, Songyu Ke, Junbo Zhang, Xiuwen Yi, and Yu Zheng. 2018. Geoman: Multi-level attention networks for geo-sensory time series prediction. In *IJCAI*, Vol. 2018. 3428–3434.
- [5] Yi Liu, JQ James, Jiawen Kang, Dusit Niyato, and Shuyu Zhang. 2020. Privacy-preserving traffic flow prediction: A federated learning approach. *IEEE Internet of Things Journal* 7, 8 (2020), 7751–7763.
- [6] Brendan McMahan, Eider Moore, Daniel Ramage, Seth Hampson, and Blaise Agüera y Arcas. 2017. Communication-efficient learning of deep networks from decentralized data. In *Artificial intelligence and statistics*. PMLR, 1273–1282.
- [7] Chuizheng Meng, Sirisha Rambhatla, and Yan Liu. 2021. Cross-node federated graph neural network for spatio-temporal data modeling. In *Proceedings of the 27th ACM SIGKDD Conference on Knowledge Discovery & Data Mining*. 1202–1211.
- [8] Shuo Wang, Yanran Li, Jiang Zhang, Qingye Meng, Lingwei Meng, and Fei Gao. 2020. Pm2.5-gnn: A domain knowledge enhanced graph neural network for pm2.5 forecasting. In *Proceedings of the 28th international conference on advances in geographic information systems*. 163–166.
- [9] Xu Wang, Pengkun Wang, Binwu Wang, Yudong Zhang, Zhengyang Zhou, Lei Bai, and Yang Wang. 2024. Latent Gaussian Processes Based Graph Learning for Urban Traffic Prediction. *IEEE Transactions on Vehicular Technology* 73, 1 (2024), 282–294. doi:10.1109/TVT.2023.3307755
- [10] Xu Wang, Hongbo Zhang, Pengkun Wang, Yudong Zhang, Binwu Wang, Zhengyang Zhou, and Yang Wang. 2023. An observed value consistent diffusion model for imputing missing values in multivariate time series. In *Proceedings of the 29th ACM SIGKDD Conference on Knowledge Discovery and Data Mining*. 2409–2418.
- [11] Zonghan Wu, Shirui Pan, Guodong Long, Jing Jiang, and Chengqi Zhang. 2019. Graph wavenet for deep spatial-temporal graph modeling. In *Proceedings of the 28th International Joint Conference on Artificial Intelligence*. 1907–1913.
- [12] Bing Yu, Haoteng Yin, and Zhanxing Zhu. 2018. Spatio-temporal graph convolutional networks: a deep learning framework for traffic forecasting. In *Proceedings of the 27th International Joint Conference on Artificial Intelligence*. 3634–3640.
- [13] Yudong Zhang, Pengkun Wang, Binwu Wang, Xu Wang, Zhe Zhao, Zhengyang Zhou, Lei Bai, and Yang Wang. 2024. Adaptive and Interactive Multi-Level Spatio-Temporal Network for Traffic Forecasting. *IEEE Transactions on Intelligent Transportation Systems* 25, 10 (2024), 14070–14086.
- [14] Yudong Zhang, Xu Wang, Pengkun Wang, Binwu Wang, Zhengyang Zhou, and Yang Wang. 2024. Modeling Spatio-Temporal Mobility Across Data Silos via Personalized Federated Learning. *IEEE Transactions on Mobile Computing* 23, 12 (2024), 15289–15306.
- [15] Chuanpan Zheng, Xiaoliang Fan, Cheng Wang, and Jianzhong Qi. 2020. Gman: A graph multi-attention network for traffic prediction. In *Proceedings of the AAAI conference on artificial intelligence*, Vol. 34. 1234–1241.