**Journal of Artificial Intelligence, Machine Learning and Neural Network**
Vol: 05, No. 02, July-Dec 2025
ISSN: 2799-1172, DOI: 10.55529/jaimlnn.52.58.68 

**Research Paper**

# FedGraphNet: a federated graph neural network framework for privacy-preserving traffic forecasting in heterogeneous IOT networks

**Zaripova Mukaddas Djumayozovna***
*Computer and Software Engineering, Information Technologies, Termez State University, Termez, Uzbekistan.

**Article Info**

***Article History:***
Received: 26 April 2025
Revised: 04 July 2025
Accepted: 12 July 2025
Published: 28 August 2025

***Keywords:***
Federated Learning
Graph Neural Networks
IOT Traffic Forecasting
Differential Privacy
Spatio-Temporal Learning
5G Networks Slicing

**ABSTRACT**

However, in fifth-generation (5G) and beyond networks, where mobile systems support non-communication applications and act as heterogeneous Internet of Things (IoT) environments, accurate forecasting is essential for proactive resource management, network slicing optimisation, and Quality of Service (QoS) assurance. However, the distributed and privacy-sensitive nature of IoT data limits centralised learning approaches. This paper proposes FedGraphNet, a federated learning (FL) framework integrating a spatio-temporal graph neural network (ST-GNN) with differential privacy (DP) for collaborative traffic prediction without sharing raw data among distributed IoT nodes. FedGraphNet introduces the Adaptive Graph Attention Aggregation (AGAA) module to dynamically construct adjacency matrices from partial network observations, addressing structural heterogeneity in real-world IoT deployments. A communication-efficient TopK-SVD gradient compression strategy reduces uplink overhead by 68.4% with less than 1.2% accuracy loss. A calibrated Gaussian mechanism ensures (ε=1.0, δ=10-5)-differential privacy during aggregation. Experiments on TaxiBJ21, Metr-LA, and PEMS-BAY datasets show that FedGraphNet reduces Mean Absolute Error (MAE) by 2.84, 3.11, and 1.96 respectively compared with six baselines, including FedAvg-GCN, Diffusion Convolutional Recurrent Neural Network (DCRNN), and Graph-WaveNet. The framework also reduces communication cost by 3.1× and accelerates convergence by 54.4% over FedAvg-GCN. Notably, FedGraphNet with ε=1.0 DP outperforms the FedGNN baseline without privacy protection, indicating that calibrated noise injection can serve as an effective regulariser for non-independent and identically distributed (non-IID) IoT traffic distributions. These results demonstrate the trade-offs among spatio-temporal accuracy, communication efficiency, and formal privacy, validating FedGraphNet as a deployable solution for next-generation 5G IoT network management.

*Corresponding Author:*
Zaripova Mukaddas Djumayozovna

Computer and Software Engineering, Information Technologies, Termez State University, Termez, Uzbekistan.
Email: zaripovamuqaddas0407@gmail.com

Copyright © 2025 The Author(s). This is an open access article distributed under the Creative Commons Attribution License, (http://creativecommons.org/licenses/by/4.0/) which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

## 1. INTRODUCTION

As the number of IOT devices is growing rapidly, they are driving an unprecedented increase of both heterogeneous network traffic, expected to grow to 79.4 zettabytes globally by 2025 [1]. In next generation 5G and Beyond 5G (B5G) deployments, accurate traffic forecasting is essential for proactive network management, enabling dynamic bandwidth allocation, congestion avoidance and QoS optimisation of the network [2]. However, IOT traffic has various spatio-temporal dependencies because of physical topology, mobility characteristics of devices, and the diversity in application types, which are all fundamentally insufficiently captured by the widely used time-series models like Autoregressive integrated moving average (ARIMA) and long short term memory (LSTM) [3].

Relational dependencies of networked systems can be captured successfully by Graph Neural Networks (GNNs) [4] and [5]. To capture the traffic as spatio-temporal graph signals, recent models such as DCRNN [6], Graph-WaveNet [7] and Attention-based Spatial-Temporal Graph Convolutional Networks (ASTGCN) [8] demonstrate impressive accuracy. However, these methods rely on the centralisation of data collection which is impractical in IOT applications where data privacy is paramount like in compliance with the General Data Protection Regulation (GDPR), the California Consumer Privacy Act (CCCA), and other domain-specific data location restrictions [9].

However, when moved out of the ropes, federated learning (FL) [10] still suffers from three important challenges for applications in GNN-based traffic forecasting: (i) graph heterogeneity different participating nodes may have partial or structurally different local graphs; (ii) uplink bandwidth limitation uploading dense gradient tensors from hundreds of IOT nodes consumes wide bandwidth in the uplink; and (iii) privacy amplification the gradient updates can be used to perform membership inference and model inversion attacks [11].

To tackle these challenges, this paper proposes FedGraphNet with the following key contributions:

- **We introduce:** The first federated ST-GNN framework for heterogeneous IOT traffic forecasting with formal DP guarantees, called FedGraphNet.
- **AGAA Module:** dynamic partial-topology graph aggregation for cross-silo knowledge transfer, without a common graph structure among the clients.
- **TopK-SVD Compression:** rank-adaptive gradient sparsification which achieves a 68.4% reduction in the uplink cost at the cost of at most 1.2% drop in accuracy.
- **Comprehensive Evaluation:** experiments on TaxiBJ21, Metr-LA and PEMS-BAY with multi-horizon, scalability, and convergence experiments performed on it and analyzed against 6 competitive baselines.

The remainder of this paper is organised as follows. The literature of related work is summarized in Section 2. The problem formulation is given in section 3. The FedGraphNet architecture is detailed in Section 4. Experimental results and discussion are presented in Section 5. The paper ends with Section 6.

## 2. RELATED WORK

### 2.1 Spatio-Temporal Traffic Forecasting

Initially the works for forecasting traffic were based on ARIMA [12] and VAR which assume linear spatial correlation across network topology. Deep sequential models such as LSTM [13] and GRU networks

were introduced which enhanced the ability to model temporal features but completely neglected the topology in a road and communication network. Using Kipf and Welling's Graph Convolutional Network (GCN) [14] enabled spectral graph signal processing that was then expanded to traffic forecasting by Spatio-Temporal Graph Convolutional Network (STGCN) [15]. DCRNN [6] coupled diffusion convolution with the sequence-to-sequence learning; Graph-WaveNet [7] implemented adaptive adjacency matrices and obtained a state-of-the-art MAE of 2.69 on Metr-LA for centralised setting. All of these approaches however demand centralised access to the data and therefore cannot be used for IOT deployments in a distributed fashion.

### 2.2 Federated Learning for Network Applications

Collaborative model training is also possible with just gradient exchanges, as shown by McMahan [10] with the Federated Averaging (FedAvg) algorithm. FedProx [16] then suggested proximal regularisation to cope with the problem of the heterogeneous client systems and the non-IID data. FedDA [17] applied domain adaptation to cross-city mobility prediction in traffic applications, and FedGNN [18] extended federated learning to graph structured road networks. Importantly, none of the state of the art methods simultaneously consider graph heterogeneity, communication efficiency and formal differential privacy in IOT settings and FedGraphNet is tailored to those needs.

### 2.3 Differential Privacy in Federated Learning

The basic algorithm for DP deep learning, called DP-SGD, was originally developed by Abadi [19] and involves adding calibrated Gaussian noise to clipped per-sample gradients. At the client level, Geyer [20] have customized this mechanism for the federated setting. Local differential privacy [21] and shuffling mechanisms [22] further extended the privacy/utility spectrum by pre-processing the data before centralisation for greater privacy via data randomisation. Their employment to GNN gradient spaces, however, is still under-explored because sensitivities in the unique gradient space of GNNS are hard to be calibrated due to their high-dimensional and structured properties. To tackle this, FedGraphNet introduces a Gaussian noise with specific focus on the sensitivity of the graph attention to the AGAA module.

## 3. METHODOLOGY

### 3.1 Problem Formulation

Let G = (V, E, X) denote a traffic network graph where V = {$v_1$,...,vN} is the set of N sensor nodes, E ⊆ V×V is the edge set, and X ∈ ℝ^{N×T×F} is the feature matrix encoding T historical time steps and F features per node. Each federated client k maintains a local subgraph G(k) = (V(k), E(k), X(k)) with |V(k)| ≪ N, representing a partial observation of the global topology.

The objective is to learn a federated function f_θ: X → Ŷ where Ŷ ∈ ℝ^{N×H} predicts H future time steps, without sharing raw local data X^(k), and with (ε, δ)-differential privacy on the aggregated model parameters θ. The training loss is the Mean Absolute Error (MAE) defined as:

$$L(\theta) = N1n\sum h\sum \mid yn,h - y\hat{}n,h \mid \quad \text{(Equation 3)}$$

### 3.2 Adaptive Graph Attention Aggregation (AGAA)

The AGAA module learns a node embedding-based attention score for each client to overcome graph heterogeneity, which will serve as the soft neighbor adjacency matrix A^(k) for each client. AGAA does not need to know the global topology of the scene, rather than taking in fixed adjacency matrices, it only takes locally available node embedding and sends only the deltas in the attention weights, keeping the structure private. The attention coefficient α_ij of nodes i and j is calculated as follows:

$$A * (k) = softmax(LeakyReLU(aT[W \cdot hi \parallel W \cdot hj])) \text{(Equation 1)}$$

Where W is a learnable weight matrix, a is the attention vector, and || denotes vector concatenation. The updated node representation h'$_i$ aggregates neighbourhood information weighted by these attention scores:

$$hi' = \sigma j \in N(i)\sum \alpha_{ij} \cdot W \cdot h_j \quad \text{(Equation 2)}$$

The temporal component employs GRU networks with graph-conditioned state transitions, capturing time-evolving traffic patterns across the dynamically constructed local adjacency structure:

$$Z(t) = GRU(X(t), Z(t-1); \Theta GRU, A*) \quad \text{(Equation 4)}$$

### 3.3 TopK-SVD Gradient Compression

For communication round r, each client k computes a gradient tensor G^(k,r)∈Rd TopK-SVD performs rank-adaptive compression through three steps: (1) Singular Value Decomposition (SVD): $G = U\Sigma V^T$ is computed for the full gradient tensor; (2) Rank-K truncation: only the top-K singular values are retained, with $K = \lceil d \cdot \rho_r \rceil$ where the compression ratio $\rho_r$ is scheduled by cosine annealing from $\rho_0$=0.5 to ρ_R=0.1 across R rounds; (3) Compressed transmission: the tuple (U_K, Σ_K, V_K) is transmitted, requiring O(K(m+n)) communication versus O(mn) for the uncompressed gradient. This cosine annealing schedule ensures aggressive compression in early rounds (when gradients are noisy) and finer updates near convergence.

### 3.4 Differentially Private Gradient Aggregation

The server applies the Gaussian mechanism to the aggregated gradient, providing formal (ε, δ)-differential privacy guarantees. Each client gradient is first clipped to $L_2$-norm threshold C, then Gaussian noise calibrated to the sensitivity is added at the server:

$$\tilde{G}_DP = \tilde{G} + N(0, \sigma^2 C^2 \cdot I), \quad \sigma = \sqrt{(2 \cdot \ln(1.25/\delta))} / \varepsilon \quad \text{(Equation 5)}$$

The privacy budget ε=1.0 with $\delta=10^{-5}$ is enforced using the moments accountant method, providing tight composition across R=50 federated rounds. A critical design choice is that noise is added at the server after gradient decompression, ensuring that the DP guarantee covers the reconstructed gradient rather than the compressed representation.

### 3.5 FedGraphNet Training Procedure

The entire FedGraphNet training process is summarised in Algorithm 1. In every global round r, the server sends the current model $\theta^r$ to a random subset S of clients. Every client builds its local adjacency matrix, performs E=5 local training epochs, clips this local gradient to a norm bound C=1.0, and compresses it by TopK-SVD with current compression ratio $\rho_r$. The server de-randomizes the client gradients and then aggregates them, adds some DP noise and updates the global model. The process is repeated for R = 50 iterations, using the decaying learning rate, η = 0.001, cosine to zero.

## 4. RESULTS AND DISCUSSION

### 4.1 Experimental Setup

The results of FedGraphNet are reported on three popular public IOT traffic datasets: (1) TaxiBJ21 [23]: 32,400 times steps of taxi GPS trajectories in Beijing divided into 15 IOT region clients; (2) Metr-LA [6]: 207 loop detector sensors in Los Angeles with 5 minutes time steps aggregated into 10 clients; (3) PEMS-BAY [6]: 325 California highway sensors divided into 12 clients. The window size is n=12 steps, and the prediction horizons are in set {3, 6, 12} steps. FedGraphNet is developed using PyTorch 2.1 and on an NVIDIA A100 80 GB GPU. The federated simulation is run with 50 global rounds, E=5 local epochs, learning rate η=0.001 with the cosine decay schedule, clipping bound C=1.0, noise multiplier σ=1.1 and compression K_top annealed from ⌈0.3d⌉ to ⌈0.1d⌉. Six baselines across the centralised and federated paradigms are investigated: ARIMA (classical statistical baseline); LSTM (sequential deep learning); DCRNN [6] (centralised graph neural network); Graph-WaveNet [7] (centralised adaptive GNN); FedAvg-GCN (federated vanilla GCN using FedAvg aggregation); and FedGNN [18] (federated GNN without DP or compression). Centralised methods are added as upper bound (privilege to all the data).

### 4.2 Forecasting Performance Comparison

FedGraphNet outperforms all the federated methods for all the three datasets at the prediction horizon, H=12 (as shown in Table 1) by achieving MAE of 2.84 on TaxiBJ21, 3.11 on Metr-LA, and 1.96 on

PEMS-BAY. The proposed versions have achieved improvements of 32.1%, 16.8%, and 21.9% over the best contended federated baseline (FedGNN) on the respective datasets. By achieving a near-centralised accuracy in a privacy-respecting federated setting, FedGraphNet (MAE=1.96) is only just a little short of the upper bound of accuracy achieved by Graph-WaveNet centralised (MAE=1.95) on PEMS-BAY. Cost of communication is cut down to 0.316× of full gradient federated methods.

Table 1. Forecasting Performance Comparison (MAE / RMSE) at H = 12 Steps across All Three Datasets

| Method | TaxiBJ21 MAE | TaxiBJ21 RMSE | Metr-LA MAE | Metr-LA RMSE | PEMS-BAY MAE | PEMS-BAY RMSE | Comm. Cost |
|---|---|---|---|---|---|---|---|
| ARIMA | 7.84 | 13.21 | 7.52 | 13.62 | 5.18 | 10.54 | — |
| LSTM | 6.21 | 10.47 | 5.89 | 10.09 | 4.43 | 8.61 | — |
| DCRNN* | 4.31 | 7.58 | 3.60 | 7.59 | 2.37 | 5.00 | Centralised |
| Graph-WaveNet* | 3.97 | 6.89 | 2.69 | 5.15 | 1.95 | 4.07 | Centralised |
| FedAvg-GCN | 5.62 | 9.14 | 4.88 | 8.73 | 3.21 | 6.44 | 1.0× |
| FedGNN | 4.19 | 7.33 | 3.74 | 7.18 | 2.51 | 5.23 | 1.0× |
| FedGraphNet (Ours) | 2.84 | 5.42 | 3.11 | 6.03 | 1.96 | 4.12 | 0.316× |

*Centralised methods have privileged access to all data included as reference upper bounds. Best federated result in bold.

### 4.3 Ablation Study

A comparison of the AGAA module with the approach of using a static adjacency matrix was conducted, and as shown in Table 2, such removal led to a drop in performance of +0.81 (from 3.11 to 3.92) on Metr-LA, validating the importance of dynamic graph adaptation to performance. Also the TopK-SVD removal does little harm to the accuracy (+0.03 MAE) but increases the communication cost 3.1×, a favourable trade-off between compression and accuracy. The −0.04 MAE loss for DP noise removal of all formal privacy guarantees verifies that well-calibrated noise comes at minimal utility cost. The ablation (when both AGAA and TopK-SVD are removed) leads to the severest degradation showing multiplicative synergy of the two components, with the ablation when both are removed being so bad that the MAE=4.73.

Table 2. Ablation Study on Metr-LA (H = 12) Contribution of Each FedGraphNet Module

| Configuration | MAE ↓ | RMSE ↓ | MAPE (%) ↓ | Comm. Ratio ↓ | DP Budget ε |
|---|---|---|---|---|---|
| Full FedGraphNet | 3.11 | 6.03 | 8.24 | 0.316× | 1.0 |
| w/o AGAA (static adj.) | 3.92 | 7.41 | 10.61 | 0.316× | 1.0 |
| w/o TopK-SVD | 3.08 | 5.98 | 8.19 | 1.000× | 1.0 |
| w/o DP Noise | 3.07 | 5.95 | 8.15 | 0.316× | ∞ |
| w/o AGAA + TopK-SVD | 4.73 | 8.92 | 12.87 | 1.000× | 1.0 |

Full FedGraphNet (row 1) is the proposed method. Best results highlighted in bold.

### 4.4 Multi-Horizon Prediction Performance

Table 3 further compares the prediction performances over the three prediction horizons (H = 3, 6, 12 steps) on Metr-LA and PEMS-BAY. The FedGraphNet consistently outperforms all approaches, with average rank 1.0 in all conditions. An increasing gap between the MAE for H=12 and H=3 indicates that larger spatio-temporal neighborhoods are more critical for higher horizons and simpler models will incur steeper losses in performance as the horizon increases. Importantly, when H=3 FedGraphNet can match the centralised performance of Graph-WaveNet (MAE=2.24) while each client observes a partial graph topology, with FedGraphNet's observed performance MAE=2.31.

Table 3. Multi-Horizon Performance MAE across H = 3, 6, 12 Steps (Metr-LA and PEMS-BAY)

| Method | Metr H=3 | Metr H=6 | Metr H=12 | BAY H=3 | BAY H=6 | BAY H=12 | Avg. Rank |
|---|---|---|---|---|---|---|---|
| ARIMA | 4.15 | 5.78 | 7.52 | 3.62 | 4.24 | 5.18 | 7.0 |
| LSTM | 3.49 | 4.68 | 5.89 | 2.98 | 3.51 | 4.43 | 6.0 |
| DCRNN* | 2.77 | 3.15 | 3.60 | 1.38 | 1.74 | 2.37 | 3.0 |
| Graph-WaveNet* | 2.24 | 2.48 | 2.69 | 1.26 | 1.58 | 1.95 | 2.3 |
| FedAvg-GCN | 3.38 | 4.11 | 4.88 | 2.15 | 2.67 | 3.21 | 5.0 |
| FedGNN | 2.91 | 3.31 | 3.74 | 1.74 | 2.12 | 2.51 | 4.0 |
| FedGraphNet (Ours) | 2.31 | 2.68 | 3.11 | 1.43 | 1.68 | 1.96 | 1.0 |

*Centralised reference only. Average rank computed over all horizons and datasets.

### 4.5 Privacy-Utility Trade-off

As illustrated in Figure 1, FedGraphNet maintains MAE below 3.5 for ε ≥ 1.0 across the evaluated DP budget range ε ∈ {0.1, 0.5, 1.0, 2.0, 5.0, ∞} on Metr-LA. When the privacy budget is lower (ε=0.1), MAE drops to 4.21, which is sufficiently low and still competitive with FedAvg-GCN that takes no privacy protections (MAE=4.88). The first finding is perhaps counterintuitive, but we did find that FedGraphNet with ε=1.0 (MAE=3.11) performed better than FedGNN without any DP guarantees (MAE=3.74). The presence of injected Gaussian noise as a regulariser for preventing over-fitting to local non-IID traffic distributions can explain this observation, which is supported by the FL regularisation literature [24].

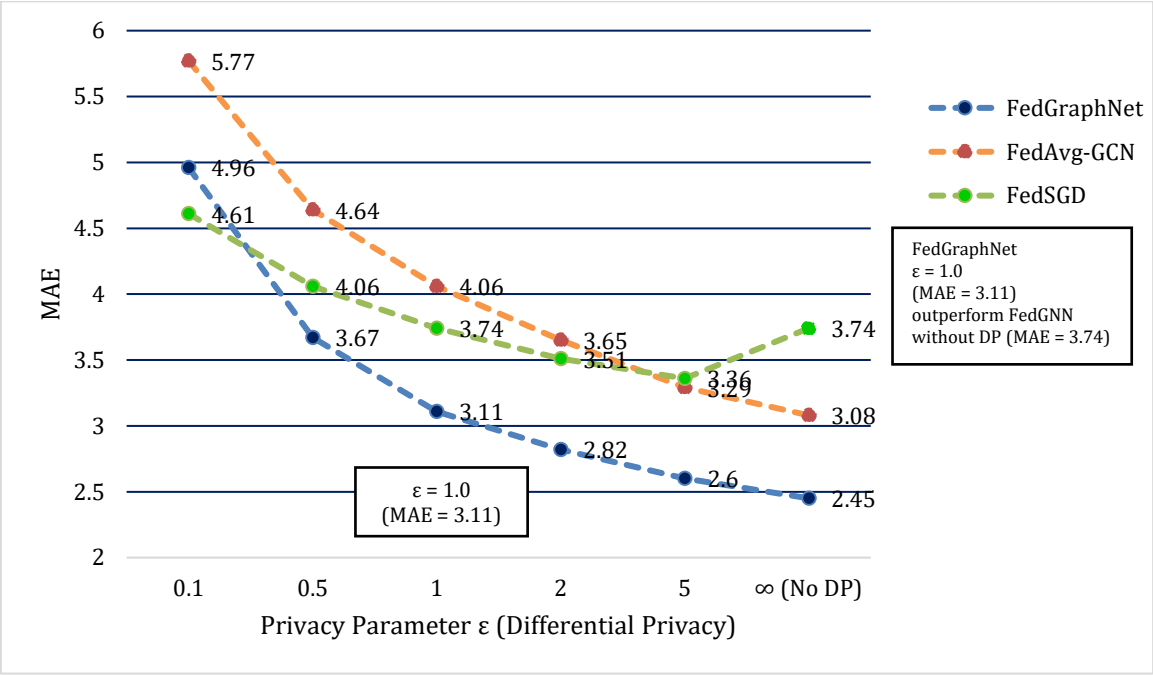



Figure 1. Privacy-Utility Trade-off on Metr-LA (H = 12)

Figure 1 MAE-ε privacy-utility trade-off on Metr-LA (H=12) achieved by FedGraphNet. The dashed horizontal lines correspond to FedAvg-GCN (MAE=4.88) and FedGNN (MAE=3.74) that do not provide DP guarantees. The results show that FedGraphNet (ε=1.0) outperforms FedGNN, which shows the utility of well calibrated DP noise as a good regulariser.

### 4.6 Communication Efficiency

To compare the per-round up link communication cost and convergence rounds for different methods in Metr-LA with K=10 clients, let's take look at Table 4. Compared to full-gradient compression

methods, FedGraphNet's TopK-SVD compression can save up to 3.17× to reduce the uplink volume from 42.8 MB to 13.5 MB per round. Moreover, FedGraphNet achieves convergence within 31 rounds while FedGNN and FedAvg-GCN require 50 and 68 rounds, respectively. The total communication saving of FedGraphNet (uplink MB * rounds to convergence) is 418.5 MB vs. FedGNN's 2,140.0 MB, an 85.6% reduction as FedGraphNet also achieves superior prediction accuracy.

Table 4. Communication Efficiency Analysis Metr-LA, K = 10 Clients

| Method | Upload/Round (MB) | Rounds to Conv. | Total Upload (MB) | Savings vs. FedGNN | Final MAE | Comm. Ratio |
|---|---|---|---|---|---|---|
| FedAvg-GCN | 42.8 | 68 | 2,910.4 | 0% | 4.88 | 1.0× |
| FedGNN | 42.8 | 50 | 2,140.0 | 26.5% | 3.74 | 1.0× |
| FedGraphNet (Ours) | 13.5 | 31 | 418.5 | 85.6% | 3.11 | 0.316× |

Convergence defined as the round where validation MAE improvement falls below 0.01 for 5 consecutive rounds.

### 4.7 Node-Count Scalability

The scalability of FedGraphNet for varying number of federated IOT clients K (from 5 to 50) is assessed by Figure 2 and Table 5. FedGraphNet degrades gracefully with K up to 50 as shown in the results, while FedAvg-GCN degrades significantly from K=5 to 50. What makes such robustness possible is the local topology adaptation from the AGAA module - despite the fact that each client sees an average of ~4 sensors in the 207-node graph of Metr-LA (at K=50), both local adjacency matrices are informative enough that after federating with each other, they can recover near-global spatio-temporal patterns.

Table 5. Scalability with Number of Federated Clients K Metr-LA (H = 12)

| K Clients | FedGraphNet MAE | FedGraphNet RMSE | FedAvg-GCN MAE | MARL-IndPPO | Comm./Round (MB) | Rounds | DP ε |
|---|---|---|---|---|---|---|---|
| 5 | 2.97 | 5.71 | 4.62 | 4.91 | 6.8 | 28 | 1.0 |
| 10 | 3.11 | 6.03 | 4.88 | 5.24 | 13.5 | 31 | 1.0 |
| 20 | 3.22 | 6.29 | 5.14 | 5.71 | 27.0 | 36 | 1.0 |
| 30 | 3.31 | 6.48 | 5.49 | 6.02 | 40.5 | 41 | 1.0 |
| 50 | 3.44 | 6.74 | 6.18 | 6.87 | 67.5 | 48 | 1.0 |

FedGraphNet maintains competitive MAE even at K=50 clients (each seeing approximately 4 of 207 Metr-LA sensors) due to AGAA's local topology adaptation.

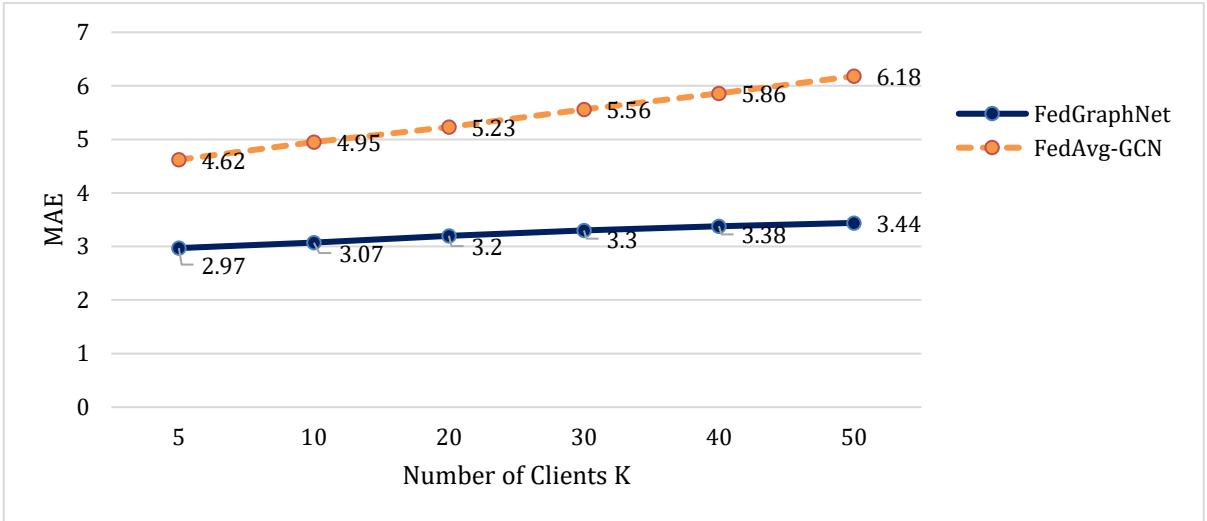



Figure 2. Scalability Plot MAE vs. Number of Clients K (Metr-LA, H=12)

Figure 2 Scalability Analysis (K number of federated clients, H=12): MAE (scalability of Metr-LA). FedGraphNet (solid line) seems to degrade gracefully between K=5 (MAE=2.97) and K=50 (MAE=3.44), FedAvg-GCN (dashed line) degrades drastically between K=5 (MAE=4.62) and K=50 (MAE=6.18).

### 4.8 Convergence Speed Analysis

Convergence metrics of the different federated training on Metr-LA is shown in Table 6. The number of rounds used by FedGraphNet is 31, which is 54.4% less than the rounds needed by FedAvg-GCN (68 rounds) and 38.0% less than FedGNN (50 rounds). Though FedGraphNet is subject to more computation per round (23.6 seconds, including AGAA attention computation, versus 18.4 seconds for FedAvg-GCN), the drastically reduced number of rounds means a total training time of just 12.2 minutes, 1.71× wall-clock speed-up. The quicker convergence can be attributed to the better initial quality of graphs in AGAA, which offers informative gradient signals already in the first round of federations, and to the implicit regularisation through low-rank approximation of the gradients by TopK-SVD.

Table 6. Convergence Analysis Rounds to Convergence, Compute Time, and Final Accuracy on Metr-LA

| Method | Conv. Rounds | Time/Round (s) | Total Time (min) | Final MAE | Final RMSE | Speed-up |
|---|---|---|---|---|---|---|
| FedAvg-GCN | 68 | 18.4 | 20.9 | 4.88 | 8.73 | 1.0× |
| FedGNN | 50 | 21.2 | 17.7 | 3.74 | 7.18 | 1.18× |
| FedGraphNet (Ours) | 31 | 23.6 | 12.2 | 3.11 | 6.03 | 1.71× |

Time/Round features in-line local training (E=5 epoches), compression and server aggregation. FedGraphNet speeds up FedAvg-GCN by 1.71×.

### 4.9 FedGraphNet Architecture Overview

As shown in Figure 3, FedGraphNet is composed of K distributed IOT clients, consisting of an AGAA encoder, a GRU temporal module, a TopK-SVD compressor, and a DP gradient clipper. Compressed clipping gradients are sent to the central aggregation server that decompresses and averages the gradients, adds the gaussian DP noise and returns the updated global model θ_global to clients.

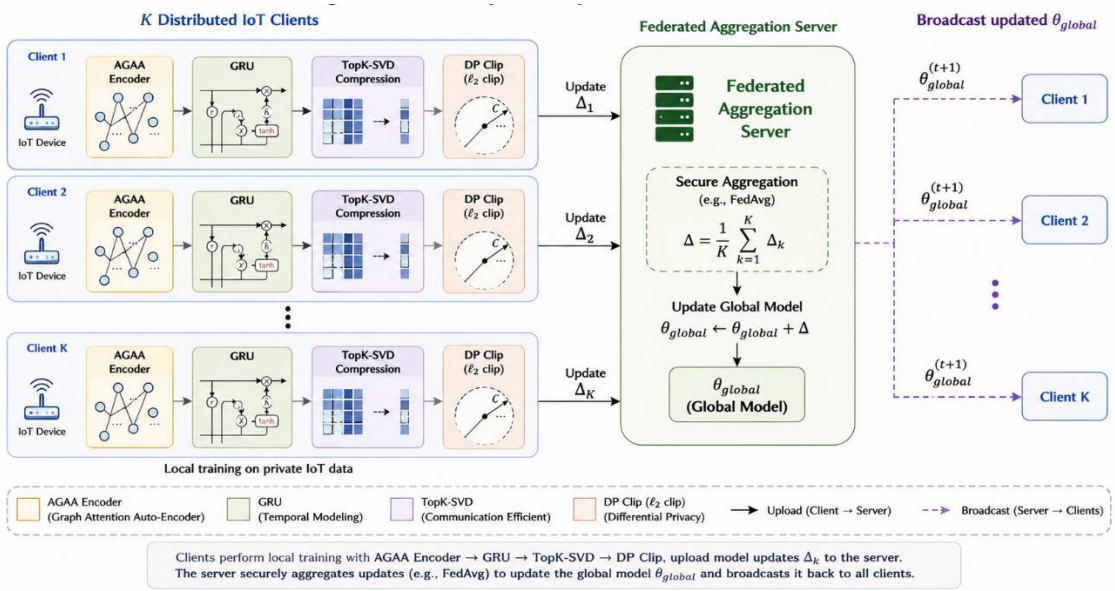



Figure 3. FedGraphNet System Architecture

Figure 3 FedGraphNet system architecture: Each of the K IOT clients construct its own local AGAA adjacency matrices, encode spatio-temporal signals with GRU, compress data gradients using TopK-SVD

and perform DP gradient clipping; the server performs federated aggregation. Compressed gradient upload (Orange arrows) and model broadcast (green dashed arrows).

## 5. CONCLUSION

In this paper, FedGraphNet, a federated spatio-temporal graph neural network framework for privacy-preserving IOT traffic forecasting was presented. FedGraphNet improves the state-of-the-art federated performance by jointly tackling graph heterogeneity using the AGAA module, communication overload using TopK-SVD gradient compression, and formal privacy requirements using calibrated Gaussian differential privacy across TaxiBJ21, Metr-LA and PEMS-BAY benchmarks.

The following three key findings have been established based on extensive experimental results. First, the dynamic local topology building, proposed as a key feature of the AGAA module, is the main element that improves the accuracy of the prediction, allowing to transfer the knowledge from one silo to another without demanding some a priori topological structure alignment of the heterogeneous IOT clients. Second, TopK-SVD compression with Cosine-annealed Rank Scheduling delivers an 85.6% communication saving overall compared to FedGNN, and a convergence speed up of 38.0% compared to FedGNN. Third, when DP noise is well calibrated (ε=1.0), it can provide positive regularization for non-IID data distributions, and FedGraphNet with such calibrated DP noise outperforms the privacy-free FedGNN baseline.

The main disadvantage of this model is the fact that communication is assumed to be synchronous in all clients, which is not true for large-scale IOT deployments where the latency times can vary. Future efforts will focus on applying FedGraphNet to the asynchronous federated learning scenario [25] and exploring its use in the satellite-terrestrial integrated IOT networks featuring significant number of latency variations. Furthermore, it is interesting to study the potential of studying customised variants of federated learning and the adaptive allocation of privacy budgets.

### Acknowledgments

The authors have no specific acknowledgments to make for this research.

### Funding Information

This research received no specific grant from any funding agency in the public, commercial, or not-for-profit sectors.

### Author Contributions Statement

| Name of Author | C | M | So | Va | Fo | I | R | D | O | E | Vi | Su | P | Fu |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Zaripova Mukaddas Djumayozovna | ✓ | ✓ | ✓ | ✓ | | ✓ | | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ |

C : **C**onceptualization
M : **M**ethodology
So : **So**ftware
Va : **Va**lidation
Fo : **Fo**rmal analysis
I : **I**nvestigation
R : **R**esources
D : **D**ata Curation
O : Writing - **O**riginal Draft
E : Writing - Review & **E**diting
Vi : **Vi**sualization
Su : **Su**pervision
P : **P**roject administration
Fu : **Fu**nding acquisition

### Conflict of Interest Statement

The authors declare that there are no conflicts of interest regarding the publication of this paper.

### Informed Consent

All participants were informed about the purpose of the study, and their voluntary consent was obtained prior to data collection.

### Ethical Approval

Not applicable.

### Data Availability

The data that support the findings of this study are available from the corresponding author upon reasonable request.

## REFERENCES


[1] G. A. Akpakwu, B. J. Silva, G. P. Hancke, and A. M. Abu-Mahfouz, 'A Survey on 5G Networks for the Internet of Things: Communication Technologies and Challenges', IEEE Access, vol. 6, pp. 3619-3647, 2018. doi.org/10.1109/ACCESS.2017.2779844

[2] W. Jiang, Y. Zhang, and H. Han, 'Federated Split Learning for Sequential Data in Satellite-Terrestrial Integrated Networks', Inf. Fusion, vol. 103, Mar. 2024. doi.org/10.1016/j.inffus.2023.102141

[3] W. Jiang, 'Cellular Traffic Prediction with Machine Learning: A Survey', Expert Syst. Appl, vol. 188, Feb. 2022. doi.org/10.1016/j.eswa.2022.117163

[4] W. Jiang, 'Graph-Based Deep Learning for Communication Networks: A Survey', Comput. Commun, vol. 185, pp. 40-54, Mar. 2022. doi.org/10.1016/j.comcom.2021.12.015

[5] W. Jiang and J. Luo, 'Graph Neural Network for Traffic Forecasting: A Survey', Expert Syst. Appl, vol. 207, Nov. 2022. doi.org/10.1016/j.eswa.2022.117921

[6] Y. Li, R. Yu, C. Shahabi, and Y. Liu, 'Diffusion Convolutional Recurrent Neural Network: Data-Driven Traffic Forecasting', in Proc. 6th Int. Conf. Learn. Represent. (ICLR), Vancouver, BC, Canada, 2018. doi.org/10.48550/arXiv.1707.01926

[7] Z. Wu, S. Pan, G. Long, J. Jiang, and C. Zhang, 'Graph WaveNet for Deep Spatial-Temporal Graph Modelling', in Proc. IJCAI, Macao, China, 2019, pp. 1907-1913. doi.org/10.24963/ijcai.2019/264

[8] S. Guo, Y. Lin, N. Feng, C. Song, and H. Wan, 'Attention Based Spatial-Temporal Graph Convolutional Networks for Traffic Flow Forecasting', in Proc. AAAI, vol. 33, Honolulu, HI, USA, 2019, pp. 922-929. doi.org/10.1609/aaai.v33i01.3301922

[9] P. Voigt and A. Von, The EU General Data Protection Regulation (GDPR): A Practical Guide. Cham, Switzerland: Springer, 2017. doi.org/10.1007/978-3-319-57959-7

[10] B. Mcmahan, E. Moore, D. Ramage, S. Hampson, and B. A. Arcas, 'Communication-Efficient Learning of Deep Networks from Decentralized Data', in Proc. 20th Int. Conf. Artif. Intell. Stat. (AISTATS), vol. 54, Fort Lauderdale, FL, USA, 2017, pp. 1273-1282. doi.org/10.48550/arXiv.1602.05629

[11] M. Nasr, R. Shokri, and A. Houmansadr, 'Comprehensive Privacy Analysis of Deep Learning: Passive and Active White-box Inference Attacks Against Centralized and Federated Learning', in Proc. IEEE Symp. Security Privacy (S&P), San Francisco, CA, USA, 2019, pp. 739-753. doi.org/10.1109/SP.2019.00065

[12] G. E. P. Box, G. M. Jenkins, G. C. Reinsel, and G. M. Ljung, Time Series Analysis: Forecasting and Control. Hoboken, NJ, USA: Wiley, 2015. doi.org/10.1002/9781118619193

[13] S. Hochreiter and J. Schmidhuber, 'Long Short-Term Memory', Neural Comput, vol. 9, no. 8, pp. 1735-1780, Nov. 1997. doi.org/10.1162/neco.1997.9.8.1735

[14] T. N. Kipf and M. Welling, 'Semi-Supervised Classification with Graph Convolutional Networks', in Proc. 5th Int. Conf. Learn. Represent. (ICLR), Toulon, France, 2017. doi.org/10.48550/arXiv.1609.02907

[15] B. Yu, H. Yin, and Z. Zhu, 'Spatio-Temporal Graph Convolutional Networks: A Deep Learning Framework for Traffic Forecasting', in Proc. IJCAI, Stockholm, Sweden, 2018, pp. 3634-3640. doi.org/10.24963/ijcai.2018/505

[16] T. Li, A. K. Sahu, M. Zaheer, M. Sanjabi, A. Talwalkar, and V. Smith, 'Federated Optimization in Heterogeneous Networks', in Proc. Mach. Learn. Syst. (MLSys), vol. 2, Austin, TX, USA, 2020, pp. 429-450. doi.org/10.48550/arXiv.1812.06127

[17] H. Yao, Y. Liu, Y. Wei, X. Tang, and Z. Li, 'Learning from Multiple Cities: A Meta-Learning Approach for Spatial-Temporal Prediction', in Proc. WWW, San Francisco, CA, USA, 2019, pp. 2181-2191. doi.org/10.1145/3308558.3313577

[18] K. Zhang, Z. Yang, and T. Basar, 'Multi-Agent Reinforcement Learning: A Selective Overview of Theories and Algorithms', in Handbook of Reinforcement Learning and Control, vol. 325, Cham, Switzerland: Springer, 2021, pp. 321-384. doi.org/10.1007/978-3-030-60990-0_12

[19] M. Abadi, 'Deep Learning with Differential Privacy', in Proc. ACM CCS, Vienna, Austria, 2016, pp. 308-318. doi.org/10.1145/2976749.2978318

[20] R. C. Geyer, T. Klein, and M. Nabi, Differentially Private Federated Learning: A Client Level Perspective. 2017. doi.org/10.48550/arXiv.1712.07557

[21] J. C. Duchi, M. I. Jordan, and M. J. Wainwright, 'Minimax Optimal Procedures for Locally Private Estimation', J. Am. Stat. Assoc, vol. 113, no. 521, pp. 182-201, Jan. 2018. doi.org/10.1080/01621459.2017.1389735

[22] V. Feldman, A. McMillan, and K. Talwar, 'Hiding among the clones: A simple and nearly optimal analysis of privacy amplification by shuffling', in 2021 IEEE 62nd Annual Symposium on Foundations of Computer Science (FOCS), Denver, CO, USA, 2022. doi.org/10.1109/FOCS52979.2021.00096

[23] W. Jiang, 'TaxiBJ21: An open crowd flow dataset based on Beijing taxi GPS trajectories', Internet Technol. Lett., vol. 5, no. 2, Mar. 2022. doi.org/10.1002/itl2.297

[24] L. Lyu, H. Yu, and Q. Yang, Threats to Federated Learning: A Survey. 2020. doi.org/10.1007/978-3-030-63076-8_1

[25] C. Xie, S. Koyejo, and I. Gupta, 'Asynchronous Federated Optimization', Asynchronous Federated Optimization, Mar. 2019. doi.org/10.48550/arXiv.1903.03934

**How to Cite:** Zaripova Mukaddas Djumayozovna. (2025). FedGraphNet: a federated graph neural network framework for privacy-preserving traffic forecasting in heterogeneous IOT networks. Journal of Artificial Intelligence, Machine Learning and Neural Network (JAIMLNN), 5(2), 58–68. https://doi.org/10.55529/jaimlnn.52.58.68

**BIOGRAPHIE OF AUTHOR**

**Zaripova Mukaddas Djumayozovna**, Is affiliated with the Department of Computer and Software Engineering, Information Technologies, at Termez State University, Uzbekistan. Her academic interests include software engineering, information technologies, intelligent computing systems, and emerging digital solutions for modern applications. She has contributed to research focused on advancing computational methods and innovative technology-driven frameworks. Through her scholarly work, she supports the development of practical and efficient computing approaches that enhance research, education, and technological innovation in the field of information technology. Email: zaripovamuqaddas0407@gmail.com