FedGTP:

Exploiting Inter-Client Spatial Dependency in **Fed**erated **G**raph-based **T**raffic **P**rediction

29 Aug 2024

北京航空航天大学
BEIHANG UNIVERSITY
1952

UNIVERSIDADE DE MACAU
仁義禮知信
澳門大學

CityU
香港城市大學
City University of Hong Kong

KDD2024
BARCELONA, SPAIN

ADS TRACK

FedGTP

Exploiting Inter-Client Spatial Dependency in **Fed**erated **G**raph-based **T**raffic **P**rediction

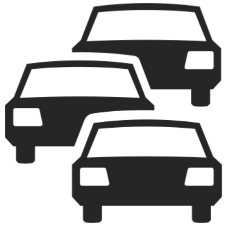
by:

Linghua Yang(Beihang University);
Wantong Chen(Beihang University);
Xiaoxi He(University of Macao);
Shuyue Wei(Beihang University);
Yi Xu(Beihang University);
Zimu Zhou(City University of Hong Kong);
Yongxin Tong(Beihang University);

KDD2024 BARCELONA, SPAIN

Traffic Prediction

Tasks



Traffic Flow Prediction

Traffic Demand Prediction

Traffic Speed Prediction

Road Occupancy Prediction

Applications

Reducing Congestion

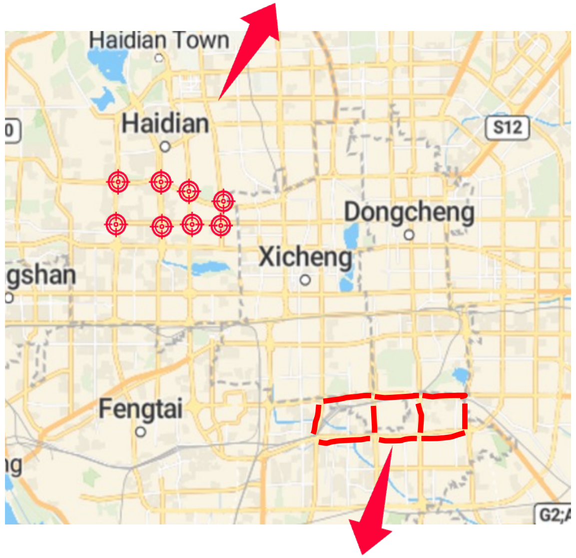
Enhancing Road Safety

Optimizing Urban Mobility

Protecting Environment

Graph-based Traffic Prediction

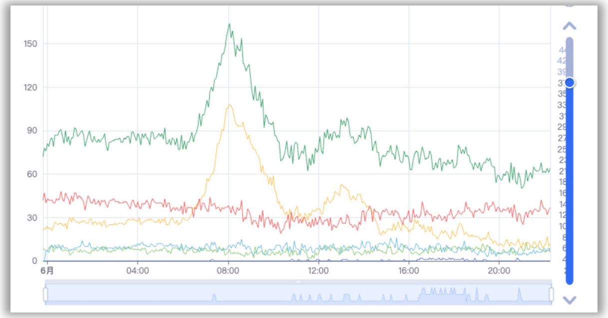
Type1: Crossing Point as Unit



Type2: Road Segment as Unit

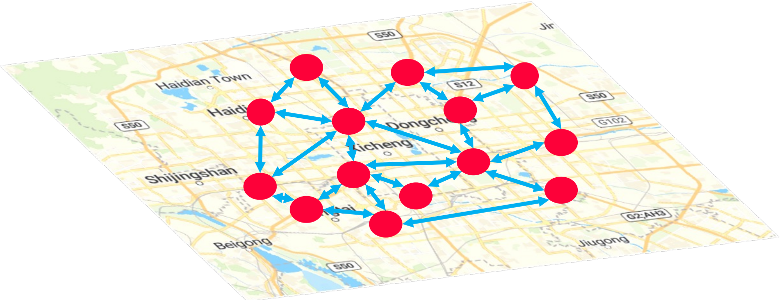
Unit of Prediction

Temporal



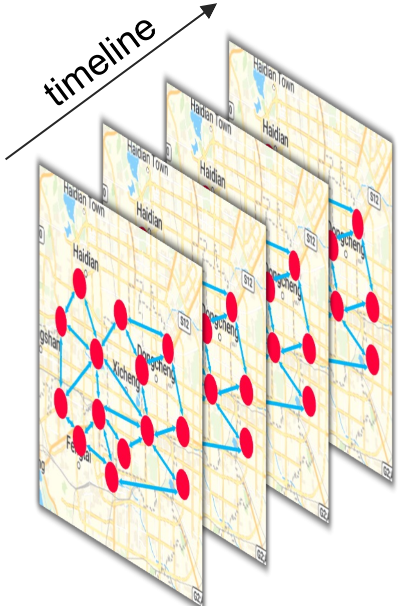
Historical traffic series within each unit
(traffic speed at crossing, traffic flow in road segments…)

Spatial



Spatial dependency between each unit
(geographical connections, semantic association…)

Spatiotemporal Graph



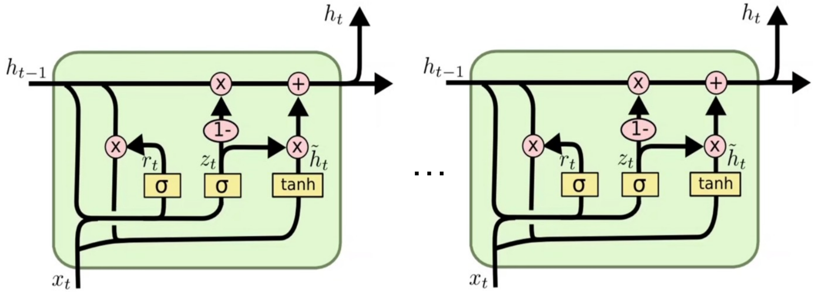
Nodes: Unit of Prediction with Temporal Data
Edges: Spatial Dependency between Units

STGNNs!

KDD2024
BARCELONA, SPAIN

General STGNNs

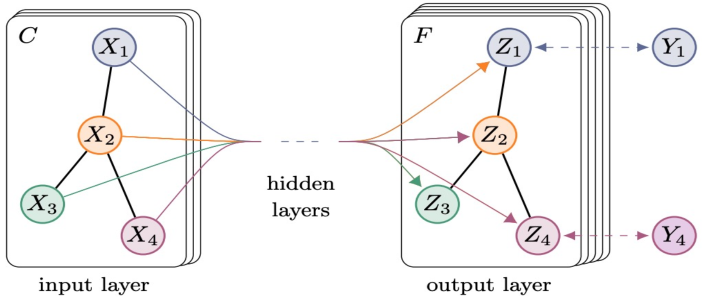
Temporal Module



Gated Recurrent Unit (GRU)

$$z_t = \sigma([X_t || h_{t-1}] \cdot W_z)$$
$$r_t = \sigma([X_t || h_{t-1}] \cdot W_r)$$
$$\hat{h}_t = tanh([X_t || (r_t \odot h_{t-1})] \cdot W_{\hat{h}})$$
$$h_t = z_t \odot h_{t-1} + (1 - z_t) \odot \hat{h}_t$$

Spatial Module



GCN

$$X^{(l)} = \left(I_N + D^{-\frac{1}{2}} A D^{-\frac{1}{2}}\right) X^{(l-1)} W$$

One typical STGNN

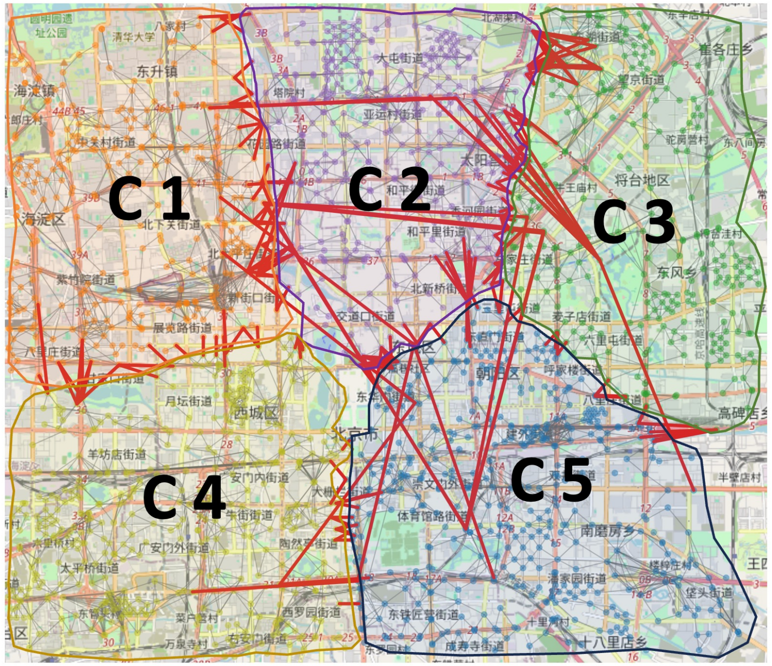
$$z_t = \sigma\left(\left(I_N + D^{-\frac{1}{2}} A D^{-\frac{1}{2}}\right) [X_t || h_{t-1}] \cdot W_z\right)$$
$$r_t = \sigma\left(\left(I_N + D^{-\frac{1}{2}} A D^{-\frac{1}{2}}\right) [X_t || h_{t-1}] \cdot W_r\right)$$
$$\hat{h}_t = tanh\left(\left(I_N + D^{-\frac{1}{2}} A D^{-\frac{1}{2}}\right) [X_t || (r_t \odot h_{t-1})] \cdot W_{\hat{h}}\right)$$
$$h_t = z_t \odot h_{t-1} + (1 - z_t) \odot \hat{h}_t$$

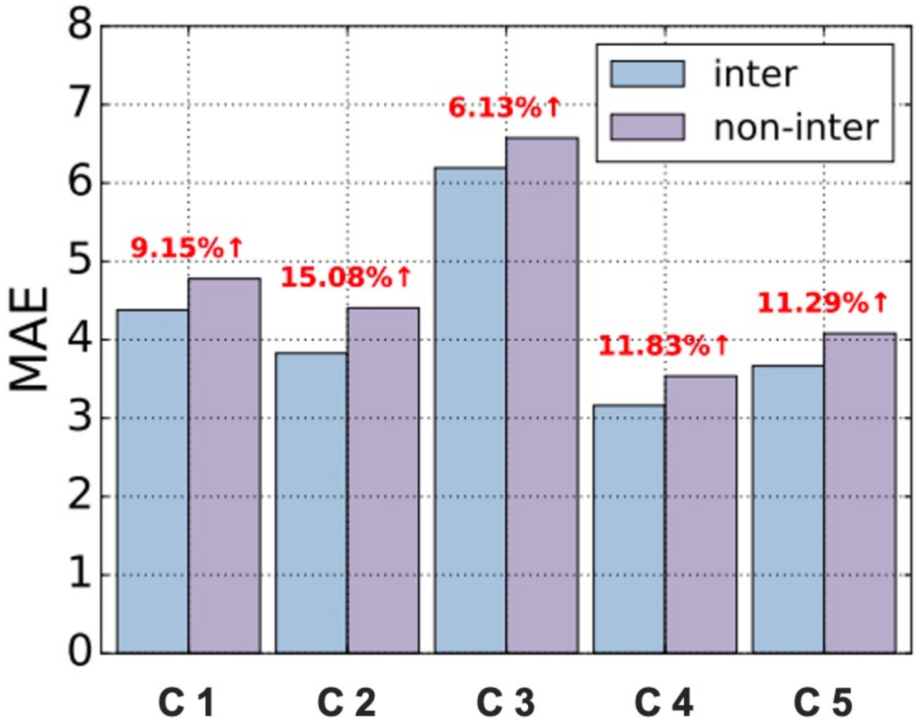
KDD2024
BARCELONA, SPAIN

Spatial Dependency

Spatial dependency
is important
(Take geographical connections
as an example…)

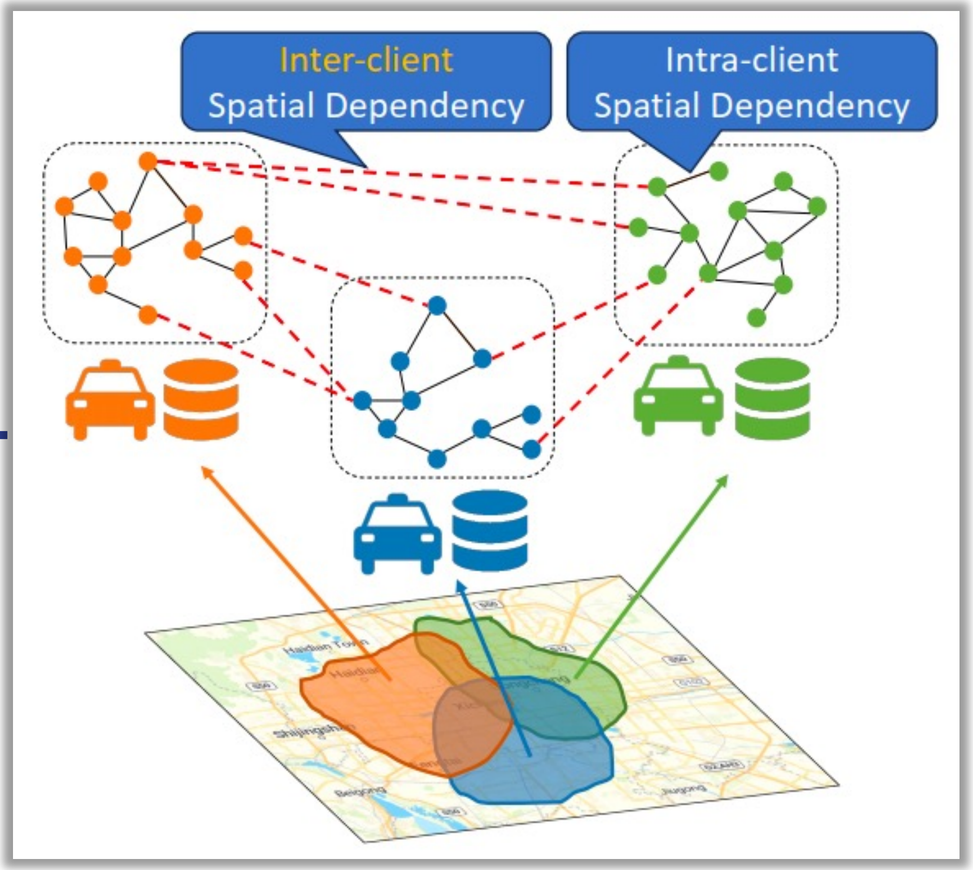
Inter-cluster
connections





KDD2024
BARCELONA, SPAIN

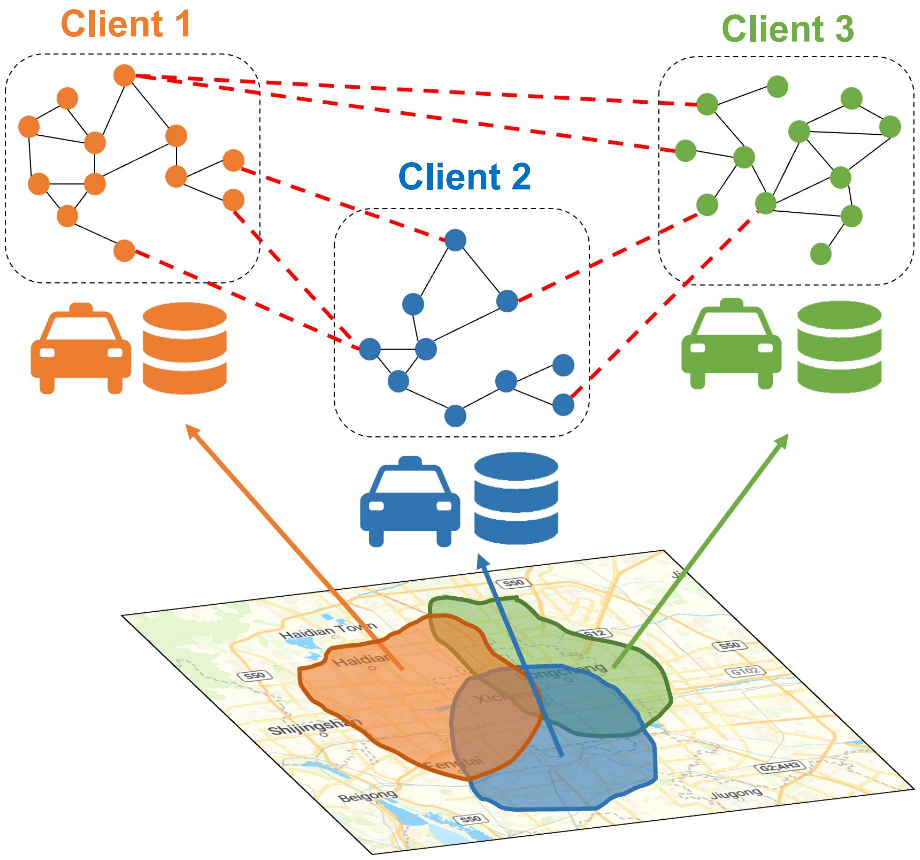
Privacy Constraints



Spatiotemporal Graph

Fragmentation

Decentralized STGNNs



Fragmented Spatiotemporal Graph

One typical Gate

$$z_t = \sigma\left(\left(I_N + D^{-\frac{1}{2}} A D^{-\frac{1}{2}}\right)[X_t || h_{t-1}] \cdot W_z\right)$$

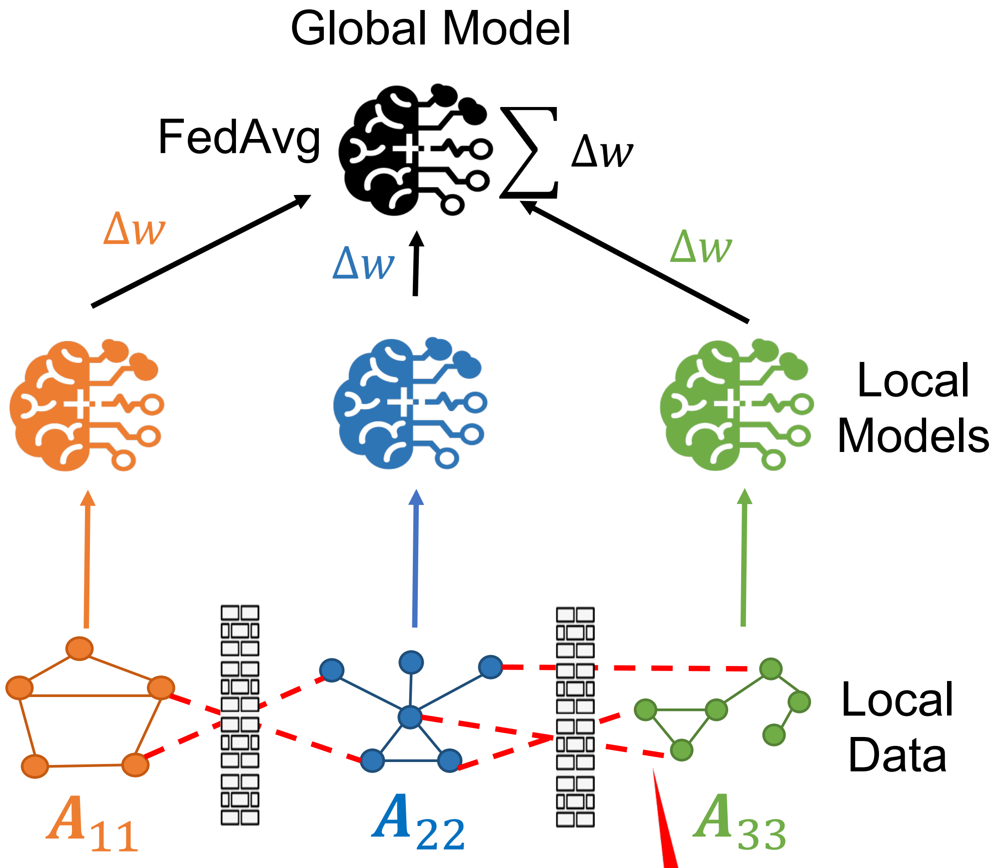
$$A = \begin{bmatrix} A_{11} & A_{12} & A_{13} \\ A_{21} & A_{22} & A_{23} \\ A_{31} & A_{32} & A_{33} \end{bmatrix}$$

Inter-client
spatial dependency

Intra-client
spatial dependency

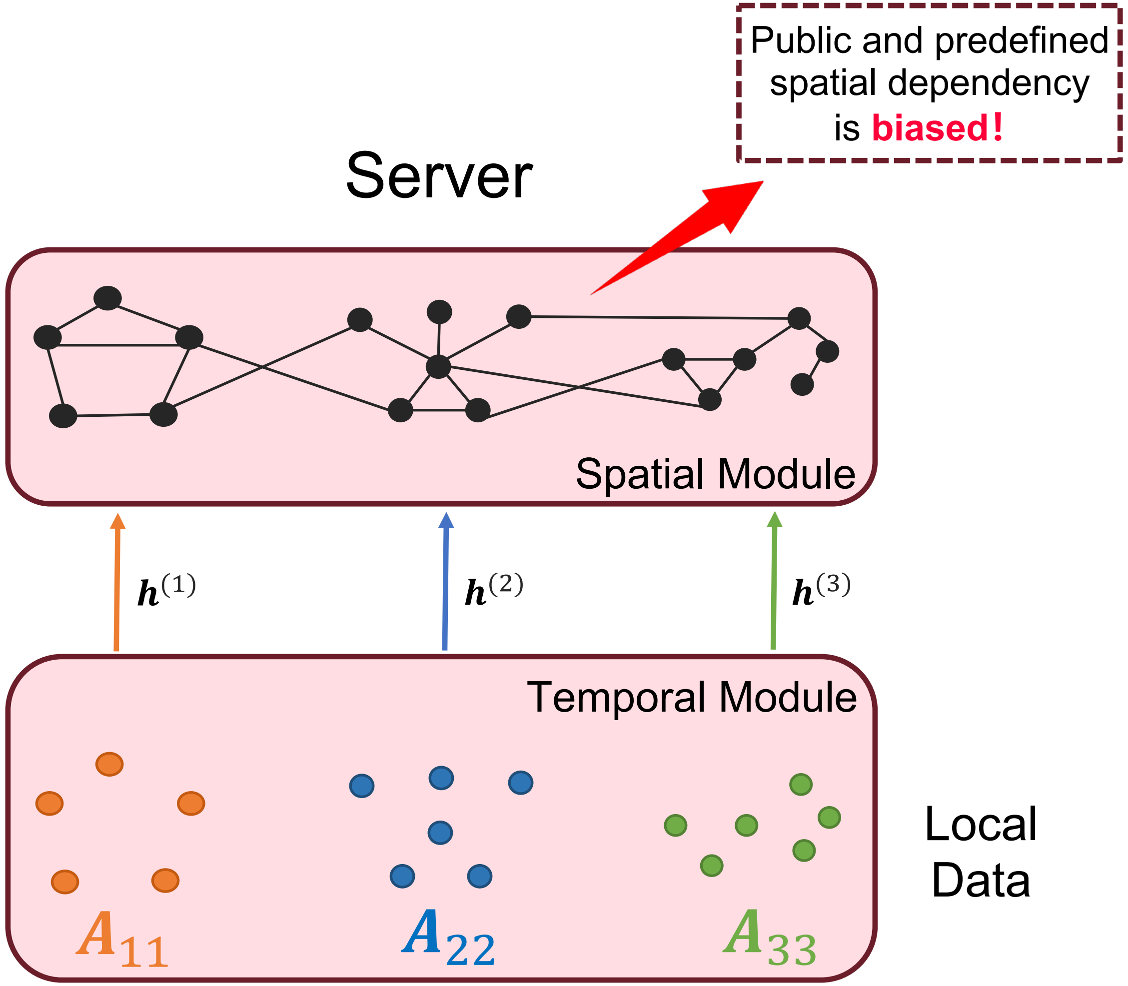
KDD2024
BARCELONA, SPAIN

Existing FedSTGNNs



FedSTGNNs
overlooking inter-client spatial dependencies

Inter-client spatial dependency is **unexploited !**



FedSTGNNs
using only the predefined and public
part of spatial dependencies

KDD2024
BARCELONA, SPAIN

Inter-Client Spatial Dependency Reconstruction

$$\boldsymbol{A} = \begin{bmatrix} A_{11} & A_{12} & A_{13} \\ A_{21} & A_{22} & A_{23} \\ A_{31} & A_{32} & A_{33} \end{bmatrix}$$

Inter-client spatial dependency

Intra-client spatial dependency

Predefined GCN

$$\boldsymbol{z}_t = \sigma\left(\left(\boldsymbol{I}_N + \boldsymbol{D}^{-\frac{1}{2}}\boldsymbol{A}\boldsymbol{D}^{-\frac{1}{2}}\right)[\boldsymbol{X}_t || \boldsymbol{h}_{t-1}] \cdot \boldsymbol{W}_z\right)$$

$$\widetilde{\boldsymbol{A}} = \begin{bmatrix} \boldsymbol{E}_1 \\ \boldsymbol{E}_2 \\ \boldsymbol{E}_3 \end{bmatrix} \cdot \begin{bmatrix} \boldsymbol{E}_1 \\ \boldsymbol{E}_2 \\ \boldsymbol{E}_3 \end{bmatrix}^\top = \begin{bmatrix} \boldsymbol{E}_1 \cdot \boldsymbol{E}_1^\top & \boldsymbol{E}_1 \cdot \boldsymbol{E}_2^\top & \boldsymbol{E}_1 \cdot \boldsymbol{E}_3^\top \\ \boldsymbol{E}_2 \cdot \boldsymbol{E}_1^\top & \boldsymbol{E}_2 \cdot \boldsymbol{E}_2^\top & \boldsymbol{E}_2 \cdot \boldsymbol{E}_3^\top \\ \boldsymbol{E}_3 \cdot \boldsymbol{E}_1^\top & \boldsymbol{E}_3 \cdot \boldsymbol{E}_2^\top & \boldsymbol{E}_3 \cdot \boldsymbol{E}_3^\top \end{bmatrix}$$

Inter-client spatial dependency

Intra-client spatial dependency

Adaptive GCN

$$\boldsymbol{z}_t = \sigma\left(\left(\boldsymbol{I}_N + \sigma(\boldsymbol{E} \cdot \boldsymbol{E}^\top)\right)[\boldsymbol{X}_t || \boldsymbol{h}_{t-1}] \cdot \boldsymbol{W}_z\right)$$

$$\sigma(\cdot) = softmax(ReLU(\cdot))$$

Inter-Client Spatial Dependency Reconstruction

$$\boldsymbol{A} = \begin{bmatrix} A_{11} & A_{12} & A_{13} \\ A_{21} & A_{22} & A_{23} \\ A_{31} & A_{32} & A_{33} \end{bmatrix}$$

Inter-client
spatial dependency

Predefined GCN

$$\boldsymbol{z}_t = \sigma\left(\left(I_N + \boldsymbol{D}^{-\frac{1}{2}}\boldsymbol{A}\boldsymbol{D}^{-\frac{1}{2}}\right)[\boldsymbol{X}_t||\boldsymbol{h}_{t-1}] \cdot \boldsymbol{W}_z\right)$$

$\boldsymbol{E}_i$ will reveal the intra-client
spatial dependency by $\boldsymbol{E}_i \cdot \boldsymbol{E}_i^{\top}$
Cannot be directly shared !

Normalization seems
hard to compute separately !

$$\widetilde{\boldsymbol{A}} = \begin{bmatrix} \boldsymbol{E}_1 \\ \boldsymbol{E}_2 \\ \boldsymbol{E}_3 \end{bmatrix} \cdot \begin{bmatrix} \boldsymbol{E}_1 \\ \boldsymbol{E}_2 \\ \boldsymbol{E}_3 \end{bmatrix}^{\top} = \begin{bmatrix} \boldsymbol{E}_1 \cdot \boldsymbol{E}_1^{\top} & \boldsymbol{E}_1 \cdot \boldsymbol{E}_2^{\top} & \boldsymbol{E}_1 \cdot \boldsymbol{E}_3^{\top} \\ \boldsymbol{E}_2 \cdot \boldsymbol{E}_1^{\top} & \boldsymbol{E}_2 \cdot \boldsymbol{E}_2^{\top} & \boldsymbol{E}_2 \cdot \boldsymbol{E}_3^{\top} \\ \boldsymbol{E}_3 \cdot \boldsymbol{E}_1^{\top} & \boldsymbol{E}_3 \cdot \boldsymbol{E}_2^{\top} & \boldsymbol{E}_3 \cdot \boldsymbol{E}_3^{\top} \end{bmatrix}$$

Adaptive GCN

$$\boldsymbol{z}_t = \sigma\left(\left(I_N + \sigma(\boldsymbol{E} \cdot \boldsymbol{E}^{\top})\right)[\boldsymbol{X}_t||\boldsymbol{h}_{t-1}] \cdot \boldsymbol{W}_z\right)$$

$$\sigma(\cdot) = softmax\big(ReLU(\cdot)\big)$$

Inter-client
spatial dependency

Intra-client
spatial dependency

KDD2024
BARCELONA, SPAIN

Decomposition Mechanism

$$\widetilde{A} = \begin{bmatrix} E_1 \\ E_2 \\ E_3 \end{bmatrix} \cdot \begin{bmatrix} E_1 \\ E_2 \\ E_3 \end{bmatrix}^\top = \begin{bmatrix} E_1 \cdot E_1^\top & E_1 \cdot E_2^\top & E_1 \cdot E_3^\top \\ E_2 \cdot E_1^\top & E_2 \cdot E_2^\top & E_2 \cdot E_3^\top \\ E_3 \cdot E_1^\top & E_3 \cdot E_2^\top & E_3 \cdot E_3^\top \end{bmatrix}$$

$$z_t = \sigma\left(\boxed{\left(I_N + \sigma(\widetilde{A})\right) \cdot [X_t || h_{t-1}]} \cdot W_z\right)$$

From one client's Perspective alone

$$Z_1 = H_1 + \sigma([E_1 \cdot E_1^\top \quad E_1 \cdot E_2^\top \quad E_1 \cdot E_3^\top]) \begin{bmatrix} H_1 \\ H_2 \\ H_3 \end{bmatrix}$$

$$\sigma(\cdot) = softmax(ReLU(\cdot))$$

Get rid of the SoftMax and separate ReLU anyway!

Plan A: SprtReLU

$$Z_1 \approx H_1 + ReLU([E_1 \cdot E_1^\top \quad E_1 \cdot E_2^\top \quad E_1 \cdot E_3^\top]) \begin{bmatrix} H_1 \\ H_2 \\ H_3 \end{bmatrix}$$

$$\approx H_1 + ReLU(E_1) \cdot ReLU\left(\sum_{j=1}^{M} E_j^\top H_j\right)$$

$\sigma(\cdot) = softmax(ReLU(\cdot))$

can be Taylor expanded into

$$\mathcal{P}_K(\cdot) = \sum_{k=0}^{K} p_k(\cdot)$$

Plan B: AdptPoLU

$$Z_1 \approx H_1 + \mathcal{P}_K([E_1 \cdot E_1^\top \quad E_1 \cdot E_2^\top \quad E_1 \cdot E_3^\top]) \begin{bmatrix} H_1 \\ H_2 \\ H_3 \end{bmatrix}$$

$$= H_1 + P_1 \cdot \mathcal{F}_K(E_1) \cdot \sum_{j=1}^{M} \left(\mathcal{F}_K^\top(E_j) \cdot H_j\right)$$

$$\mathcal{F}_K(\cdot) = \{\otimes^k(\cdot)\}_{k=0}^{K}$$

$\otimes^k(\cdot)$ transforms a matrix from $\mathbb{R}^{N\times d}$ to $\mathbb{R}^{N\times d^k}$ by applying a k-order Cartesian power on each row

KDD2024 BARCELONA, SPAIN

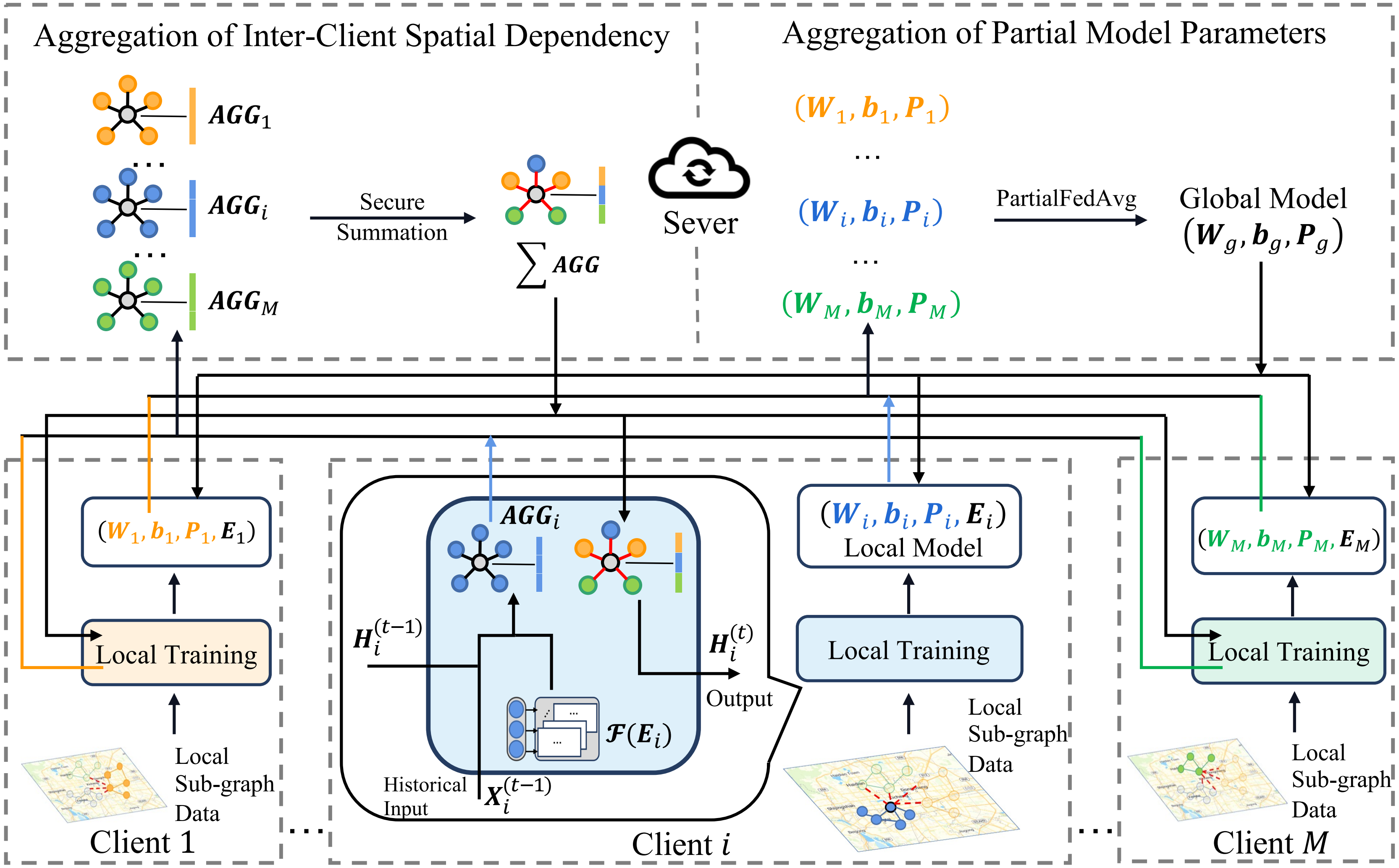
FedGTP System Overview

$$\boldsymbol{AGG}_i = \mathcal{F}_K^{\top}(\boldsymbol{E}_i)\cdot \boldsymbol{H}_i \begin{cases} = {\boldsymbol{E}_i}^{\top}\cdot \boldsymbol{H}_i & \text{, for SprtReLU} \\ = \left\{\left(\otimes^k(\boldsymbol{E}_i)\right)^{\top}\cdot \boldsymbol{H}_i\right\}_{k=0}^{K} & \text{, for AdptPoLU} \end{cases}$$

Communication overhead:

$$\begin{cases} O(M\cdot d\cdot F) & \text{, for SprtReLU} \\ O(M\cdot d^K\cdot F) & \text{, for AdptPoLU} \end{cases}$$

M : Number of Clients
d: spatial embedding size
K: order of polynomial
F: Hidden size



KDD2024
BARCELONA, SPAIN

Baseline Comparison

Table 1. Baselines overlooking inter-client spatial dependency

baseline	federated setting	task	method	RMSE	MAE	MAPE(%)
FASTGNN	PeMSD7 (4 clients, 12->9)	speed	FASTGNN	5.83	3.50	8.36
			FedGTP	**4.73**	**2.60**	**5.88**
MFVSTGNN	PEMS-BAY (8 clients, 12->12)	speed	MFVSTGNN	3.91	1.93	4.48
			FedGTP	**3.88**	**1.79**	**3.54**
	METR-LA (8 clients, 12->12)	speed	MFVSTGNN	4.45	3.35	9.42
			FedGTP	**4.41**	**3.23**	**8.75**
FLoS	PeMSD4 (5 clients, 24->12)	flow	FLoS	42.84	28.64	-
			FedGTP	**41.33**	**25.62**	-

FedGTP adaptively exploits the inter-client spatial dependency.

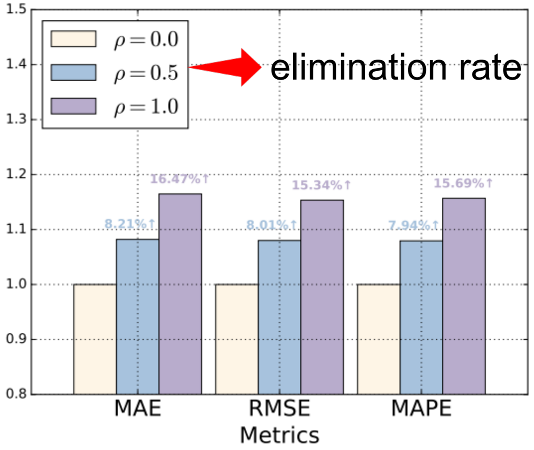
Table 2. Baselines using only the predefined and public part of spatial dependency

baseline	federated setting	task	method	RMSE	MAE	MAPE(%)
CNFGNN	PEMS-BAY (325 clients, 12->12)	speed	CNFGNN	3.7090	2.3528	4.82
			FedGTP	**3.6440**	**1.6813**	**3.35**
	METR-LA (207 clients, 12->12)	speed	CNFGNN	11.4137	7.5161	36.26
			FedGTP	**10.3978**	**4.2883**	**24.83**
FCGCN	PeMSD4 (28 clients, 6->1)	flow	FCGCN	29.6775	18.6483	22.57
			FedGTP	**26.6993**	**17.9049**	**13.99**
		speed	FCGCN	1.8777	1.0008	1.84
			FedGTP	**1.6800**	**0.9493**	**1.72**
		occ	FCGCN	0.1181	0.0066	18.92
			FedGTP	**0.0126**	**0.0064**	**16.03**
	PeMSD8 (14 clients, 6->1)	flow	FCGCN	22.4601	14.7723	11.81
			FedGTP	**20.4897**	**13.7754**	**9.09**
		speed	FCGCN	1.5570	0.8228	1.54
			FedGTP	**1.4221**	**0.7616**	**1.35**
		occ	FCGCN	0.0598	0.0058	12.91
			FedGTP	**0.0110**	**0.0054**	**10.05**
CTFL	PeMSD4 (8 clients, 12->9)	speed	CTFL-STGCN	4.87	-	4.84
			CTFL-MTGNN	4.91	-	4.79
			FedGTP	**3.42**	-	**3.78**
	PeMSD7 (8 clients, 12->9)	speed	CTFL-STGCN	5.25	-	7.08
			CTFL-MTGNN	5.23	-	7.08
			FedGTP	**4.89**	-	**6.88**

FedGTP adaptively uncovers more profound inter-client spatial dependency compared with those using only the public and predefined connections.

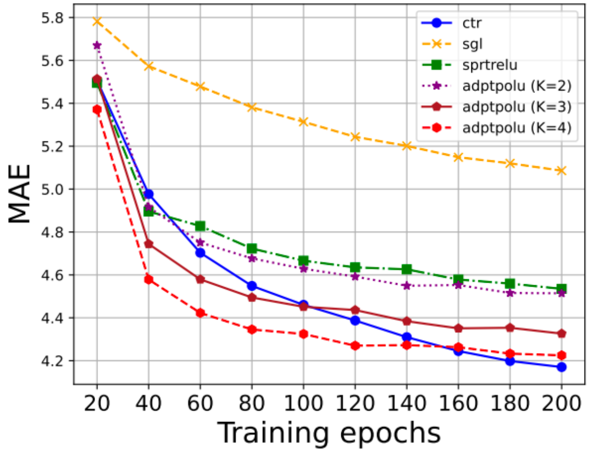
KDD2024
BARCELONA, SPAIN

Ablation Study

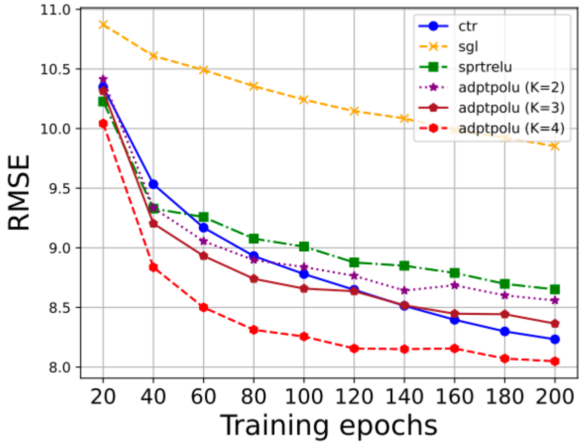


(a) Impact of inter-client spatial dependency on performance

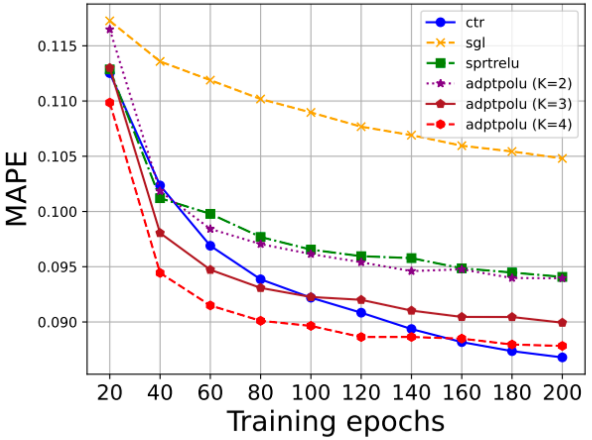
Randomly eliminate the reconstructed dependencies for each client



(b) Impact of activation decomposition on MAE



(c) Impact of activation decomposition on RMSE



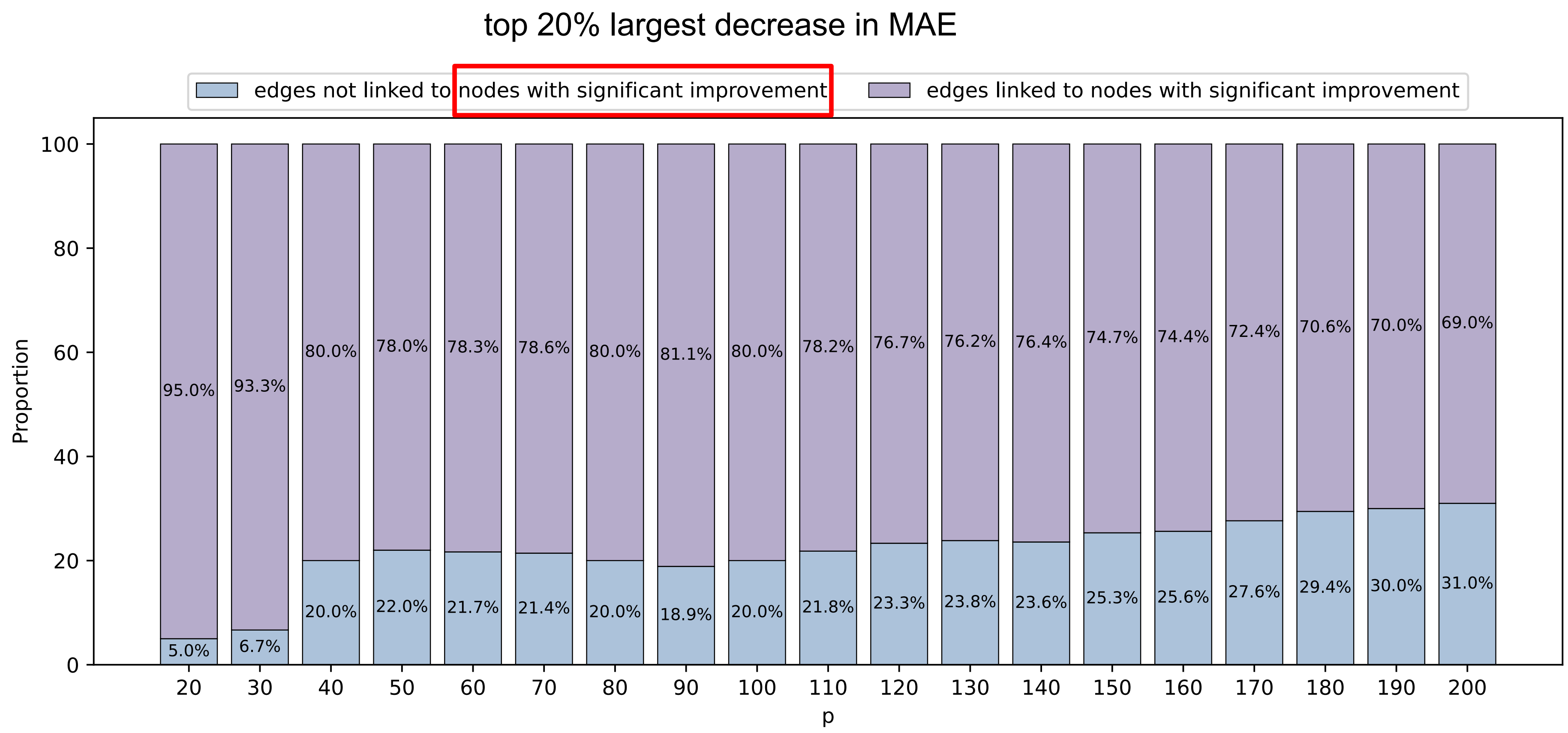
(d) Impact of activation decomposition on MAPE

- *SprtReLU* mechanism improves performance, but there still exits a gap with *ctr*
- *AdptPoLU* has similar performance with *SprtReLU* when K = 2, and performs almost close to *ctr*, when K rises to 4
- AdptPoLU sometimes beats ctr, we attribute it to the adaptive coefficients facilitating the training process

KDD2024
BARCELONA, SPAIN

Case Study

The average proportion:
$1 - (1-0.2)^2 = 36\%$



key inter-client spatial dependencies: top p edges with the largest weights

Case Study

We focus our gaze on one of the significantly benefited nodes and its associated key dependencies

KDD2024
BARCELONA, SPAIN

THANK YOU

KDD2024
BARCELONA, SPAIN