ICML | 2022

Thirty-ninth International Conference on Machine Learning

# DSTAGNN: Dynamic Spatial-Temporal Aware Graph Neural Network for Traffic Flow Forecasting

Shiyong Lan[1], Yitong Ma[1], Weikang Huang[1], Wenwu Wang[2], Hongyu Yang[1], Piaoyang Li[1]

[1] Sichuan University, China [2] University of Surrey, UK

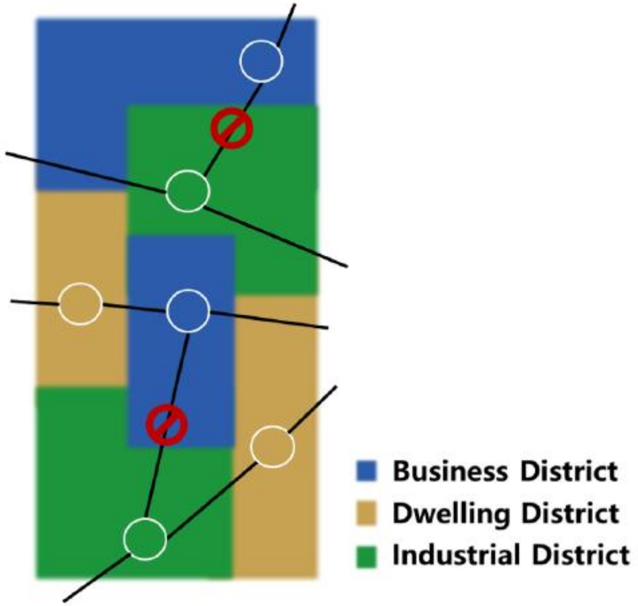
# Content

1. Introduction
2. Methodology
3. Experimental Result
4. Conclusion

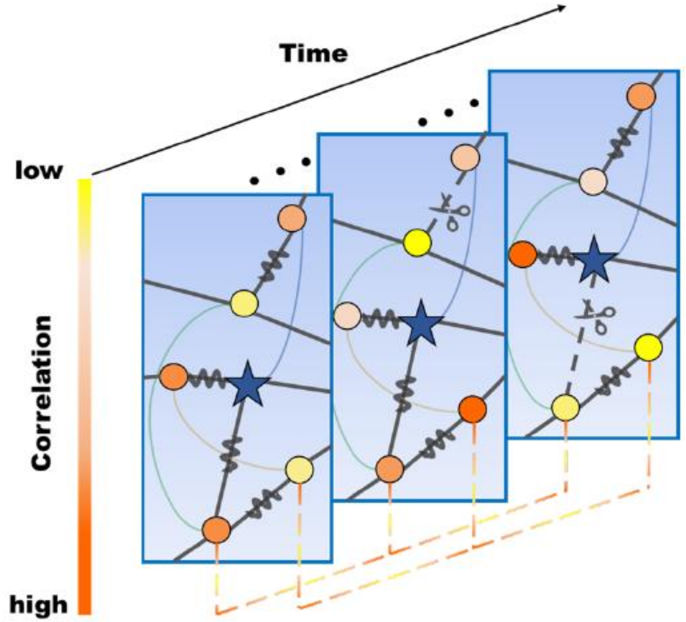
# 1 Introduction

# 1 Introduction

Due to the presence of complex dynamic spatial-temporal dependencies within a road network, achieving highly accurate traffic flow prediction is a challenging task.




(a) Road network of a certain zone.




(b) Dynamic spatial-temporal relevance in traffic flow data.

Fig 1. Dynamic spatial-temporal correlations in real world traffic data.

# 1 Introduction – Research Questions

- Spatially similar urban functional areas in the road network have remarkably similar traffic flow patterns even if they are far away, which demands **simultaneously capturing wide-scale local and global spatial relevance**.

- The interweaving effect of long-term dynamic similar patterns and short-term random irregular patterns in the time dimension is bound to require **adaptively focusing on temporal dependence in a wide range**.

# 1 Introduction – Proposed Ideas

- **A novel graph** to capture dynamic attributes of spatial association among nodes by mining from their historic traffic flow data directly, without using a predefined static adjacency matrix.

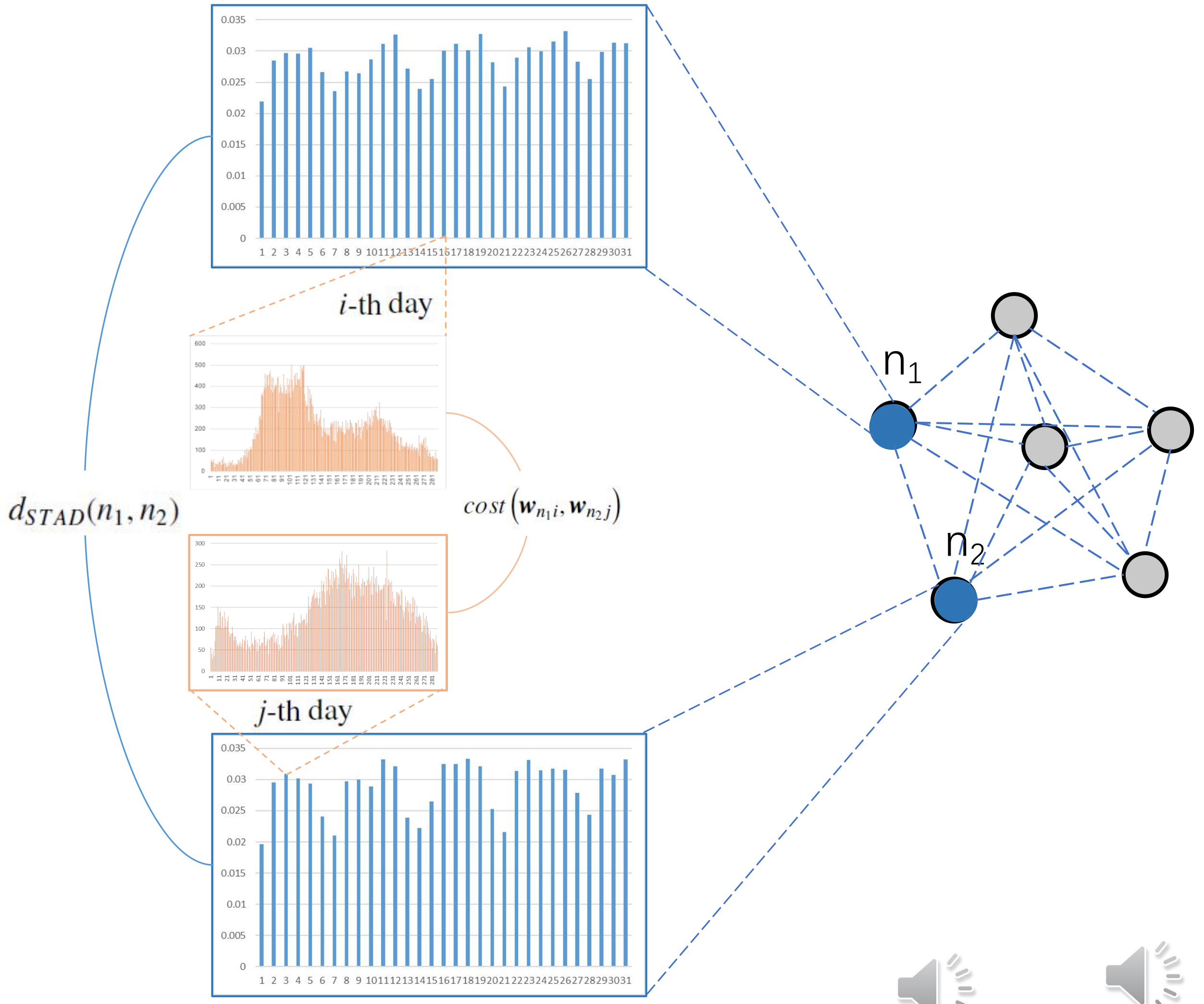
- **A new spatial-temporal attention module** to exploit the dynamic spatial correlation within multi-scale neighborhoods based on multi-order Chebyshev polynomials in GCN, meanwhile, the wide range of temporal dependency is exploited by the multi-head self-attention.

- **An improved gated convolution module**, which can further enhance the awareness of the model to dynamic temporal dependency within the road network.

# 2 Methodology

# 2 Methodology

## 2.1 Spatial-Temporal Aware Graph Construction



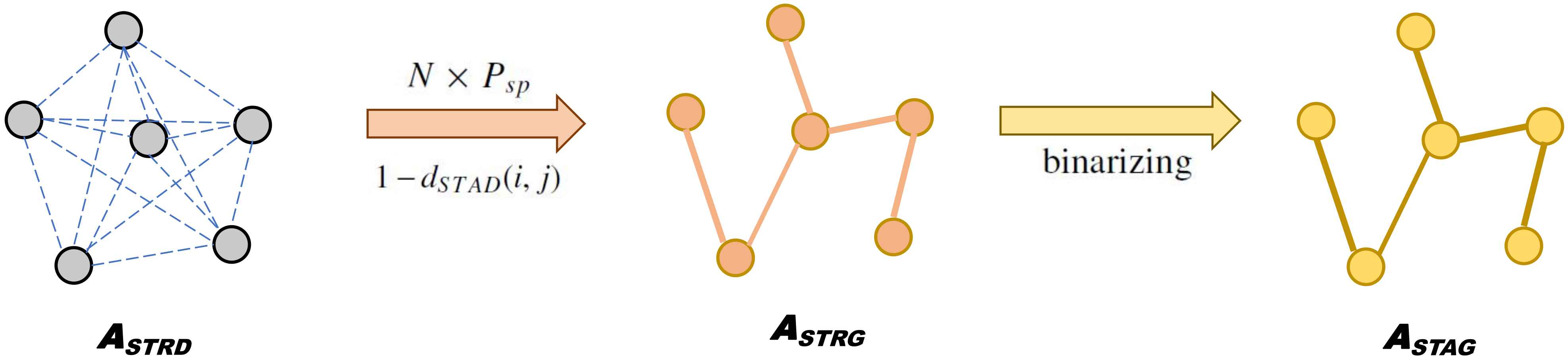

$$\boldsymbol{X}_n^f = (\boldsymbol{w}_{n1}, \boldsymbol{w}_{n2}, \ldots, \boldsymbol{w}_{nD}),\ \boldsymbol{w}_{nd} \in \mathbb{R}^{d_t},\ \text{where } d \in [1,\ D] \quad (1)$$

$$m_{nd} = \frac{\|\boldsymbol{w}_{nd}\|_2}{Z_n}, \quad Z_n = \sum_{d=1}^{D} \|\boldsymbol{w}_{nd}\|_2 \quad (2)$$

$$P_n\{X_d = m_{nd}\} \quad (3)$$

$$cost\left(\boldsymbol{w}_{n_1 i}, \boldsymbol{w}_{n_2 j}\right) = 1 - \frac{\boldsymbol{w}_{n_1 i}^{\top} \cdot \boldsymbol{w}_{n_2 j}}{\left\|\boldsymbol{w}_{n_1 i}\right\|_2 \times \left\|\boldsymbol{w}_{n_2 j}\right\|_2} \quad (4)$$

$$\begin{aligned} d_{STAD}(n_1, n_2) &\triangleq STAD\left(\boldsymbol{X}_{n_1}, \boldsymbol{X}_{n_2}\right) = \\ &\inf_{\gamma \in \Pi[P_{n_1}, P_{n_2}]} \int_x \int_y \gamma(x, y) \left(1 - \frac{\boldsymbol{w}_{n_1 x}^{\top} \cdot \boldsymbol{w}_{n_2 y}}{\sqrt{\boldsymbol{w}_{n_1 x}^{\top} \boldsymbol{w}_{n_1 x}} \times \sqrt{\boldsymbol{w}_{n_2 y}^{\top} \boldsymbol{w}_{n_2 y}}}\right) dx dy \\ \text{s.t.} &\int \gamma(x, y) dy = \frac{\left\|\boldsymbol{w}_{n_1 x}\right\|_2}{\sum_{x=1}^{D} \left\|\boldsymbol{w}_{n_1 x}\right\|_2}, \int \gamma(x, y) dx = \frac{\left\|\boldsymbol{w}_{n_2 y}\right\|_2}{\sum_{y=1}^{D} \left\|\boldsymbol{w}_{n_2 y}\right\|_2} \end{aligned} \quad (5)$$


$N \times P_{sp}$
$1 - d_{STAD}(i, j)$
binarizing
$A_{STRD}$
$A_{STRG}$
$A_{STAG}$

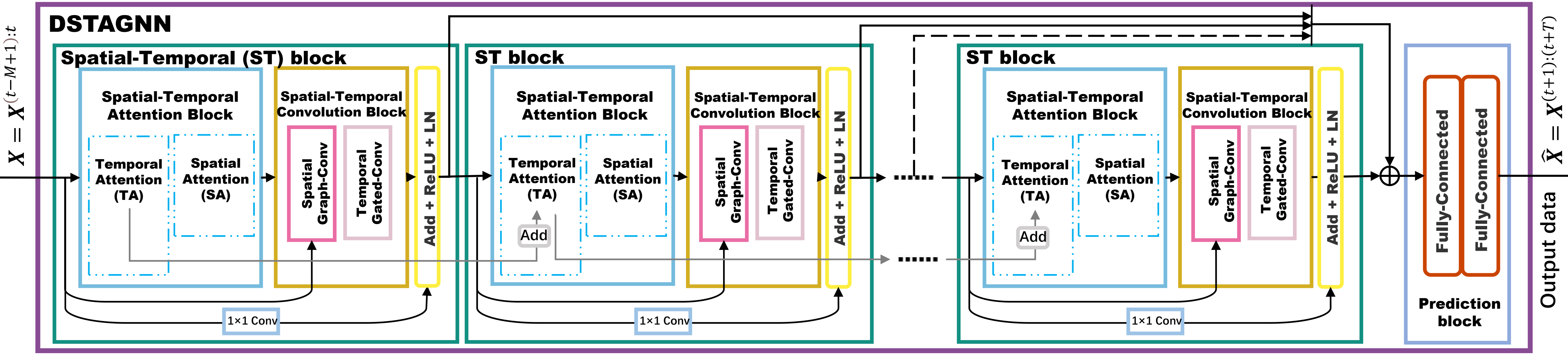



Fig 2(a) . Overall structure of DSTAGNN, consisting of several Spatial-Temporal (ST) blocks and a prediction block.

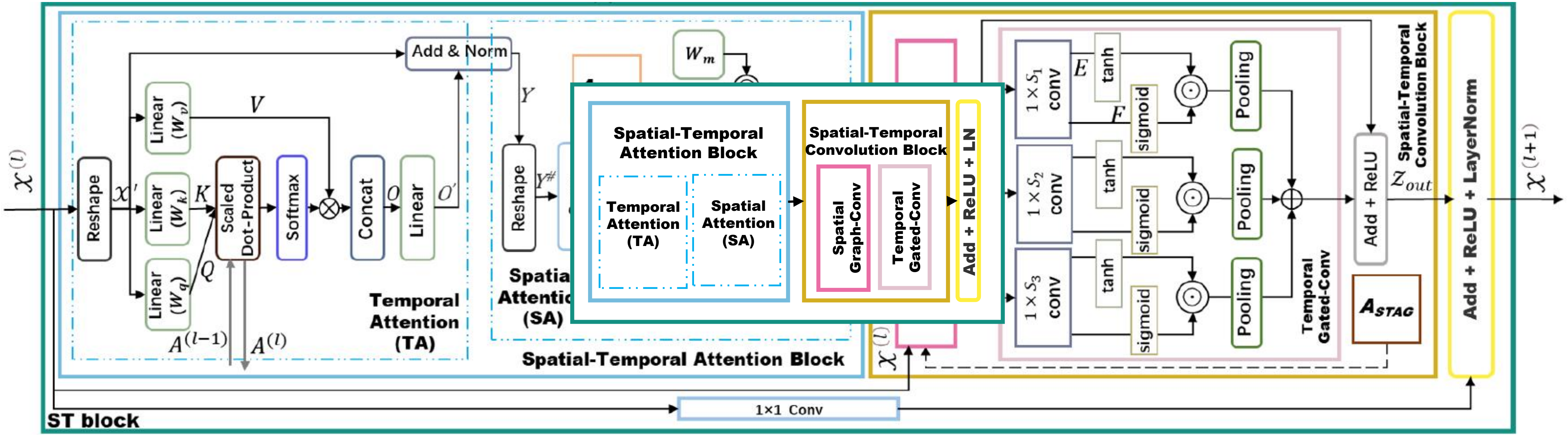



Fig 2(b). Detail of the Spatial-Temporal block

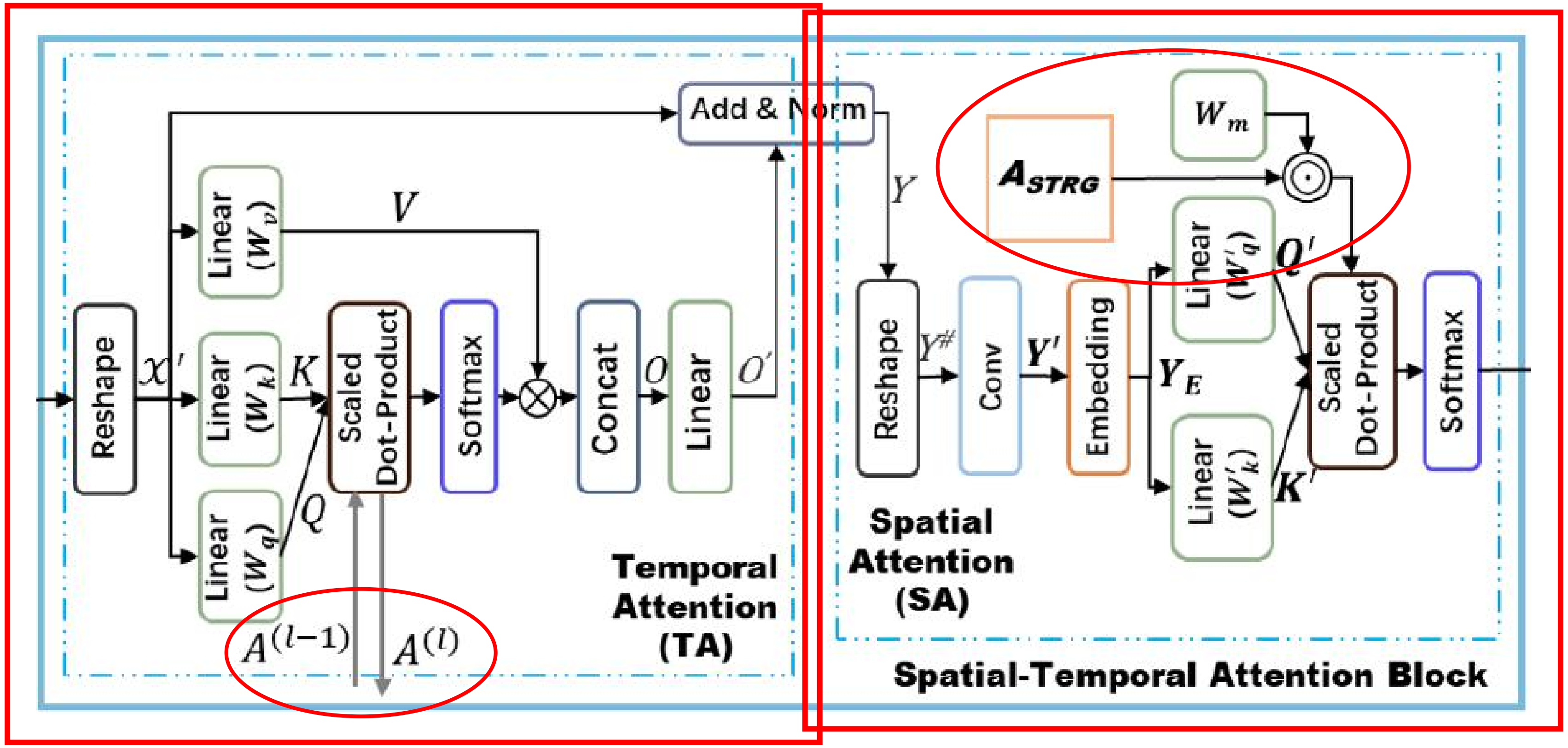



Fig 3. Detail of the Spatial-Temporal Attention Block

$$g_\theta * Gx = g_\theta(\mathbf{L})x = \sum_{k=0}^{K-1} \boldsymbol{\theta}_k \left(T_k(\tilde{\mathbf{L}}) \odot \boldsymbol{P}^{(k)}\right) x$$

$$Z_{out}^{(l)} = \text{M-GTU}(\mathcal{Z}^{(l)}) = \text{ReLU}(\text{Concat}(\text{Pooling}(\Gamma_1 *_\tau \mathcal{Z}^{(l)}), \text{Pooling}(\Gamma_2 *_\tau \mathcal{Z}^{(l)}), \text{Pooling}(\Gamma_3 *_\tau \mathcal{Z}^{(l)})) + \mathcal{Z}^{(l)})$$

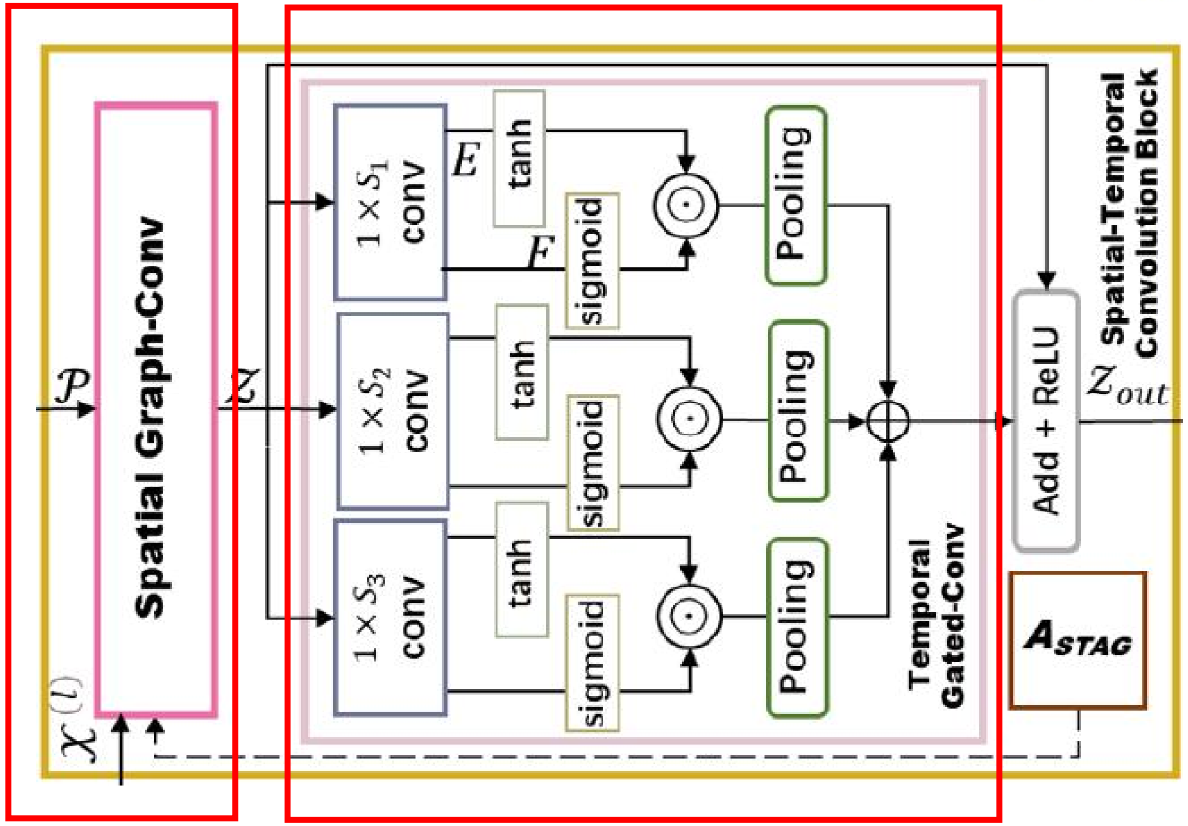



Fig 4. Detail of the Spatial-Temporal Convolution Block

# 3 Experimental Result

**Evaluation Metric:** The mean absolute error (MAE), the mean absolute percentage error (MAPE), and the root mean squared error (RMSE) are used to measure the performance of models.

**Baseline Methods**: We compare our DSTAGNN with ten baseline methods, STFGNN, AGCRN, STGODE, Z-GCNETs etc.

| DATASETS | NODES | EDGES | TIMESTEPS | MISSINGRATIO |
|---|---|---|---|---|
| PEMS03 | 358 | 547 | 26208 | 0.672% |
| PEMS04 | 307 | 340 | 16992 | 3.182% |
| PEMS07 | 883 | 866 | 28224 | 0.452% |
| PEMS08 | 170 | 295 | 17856 | 0.696% |

Table 1. Description and statistics of datasets

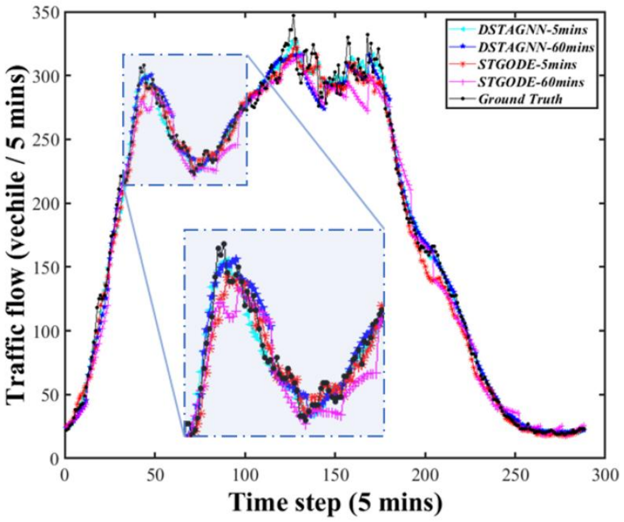
# 3 Experimental Result

| Datasets | Metric | *FC-LSTM | *TCN | DCRNN | STGCN | ASTGCN(r) | STSGCN | AGCRN | STFGNN | STGODE | Z-GCNETs | DSTAGNN-G | **DSTAGNN** |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| PEMS03 | MAE | 21.33 | 19.31 | 18.18 | 17.49 | 17.69 | 17.48 | *15.98 | 16.77 | 16.50 | *16.64 | 15.61 | **15.57** |
| | MAPE(%) | 22.33 | 19.86 | 18.91 | 17.15 | 19.40 | 16.78 | *15.23 | 16.30 | 16.69 | *16.39 | 14.79 | **14.68** |
| | RMSE | 35.11 | 33.24 | 30.31 | 30.12 | 29.66 | 29.21 | *28.25 | 28.34 | 27.84 | *28.15 | 27.23 | **27.21** |
| PEMS04 | MAE | 26.24 | 23.11 | 24.70 | 22.70 | 22.93 | 21.19 | 19.83 | 19.83 | 20.84 | 19.50 | 19.41 | **19.30** |
| | MAPE(%) | 19.30 | 15.48 | 17.12 | 14.59 | 16.56 | 13.90 | 12.97 | 13.02 | 13.77 | 12.78 | 12.84 | **12.70** |
| | RMSE | 40.49 | 37.25 | 38.12 | 35.55 | 35.22 | 33.65 | 32.26 | 31.88 | 32.82 | 31.61 | 31.63 | **31.46** |
| PEMS07 | MAE | 29.96 | 32.68 | 25.30 | 25.38 | 28.05 | 24.26 | *22.37 | 22.07 | 22.59 | *21.77 | 21.67 | **21.42** |
| | MAPE(%) | 14.34 | 14.22 | 11.66 | 11.08 | 13.92 | 10.21 | *9.12 | 9.21 | 10.14 | *9.25 | 9.06 | **9.01** |
| | RMSE | 43.94 | 42.23 | 38.58 | 38.78 | 42.57 | 39.03 | *36.55 | 35.80 | 37.54 | *35.17 | 35.04 | **34.51** |
| PEMS08 | MAE | 22.20 | 22.69 | 17.86 | 18.02 | 18.61 | 17.13 | 15.95 | 16.64 | 16.81 | 15.76 | 15.90 | **15.67** |
| | MAPE(%) | 15.02 | 14.04 | 11.45 | 11.40 | 13.08 | 10.96 | 10.09 | 10.60 | 10.62 | 10.01 | 9.97 | **9.94** |
| | RMSE | 33.06 | 35.79 | 27.83 | 27.83 | 28.16 | 26.80 | 25.22 | 26.22 | 25.97 | 25.11 | 25.24 | **24.77** |

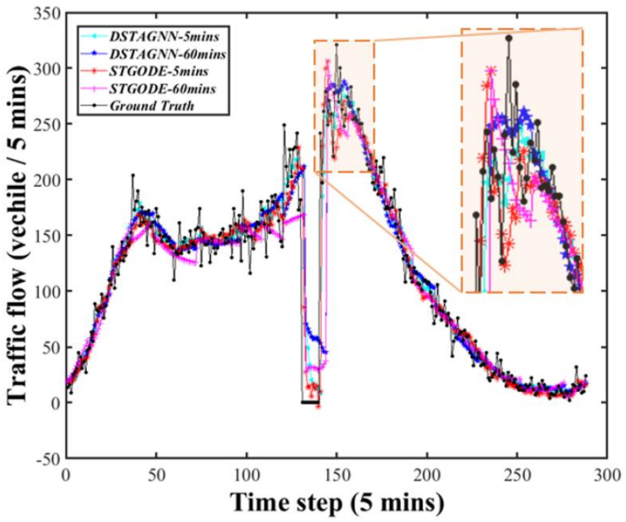
* denotes re-implementation or re-training. ____ denotes the best indicator in baselines.

Table 2. Performance comparison of our DSTAGNN and baseline models on PEMS datasets.

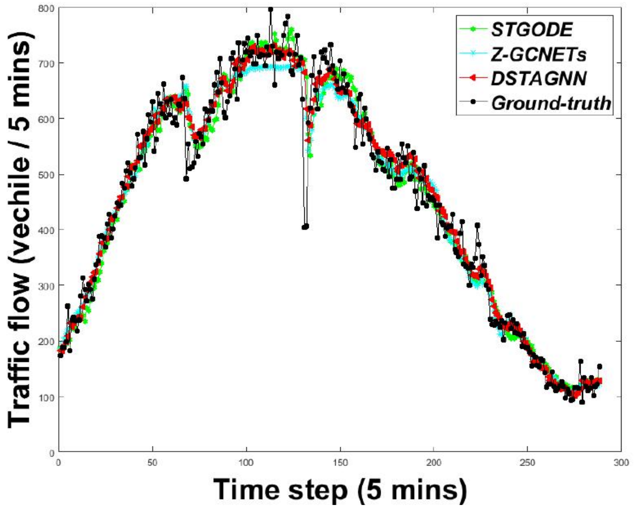
# 3 Experimental Result



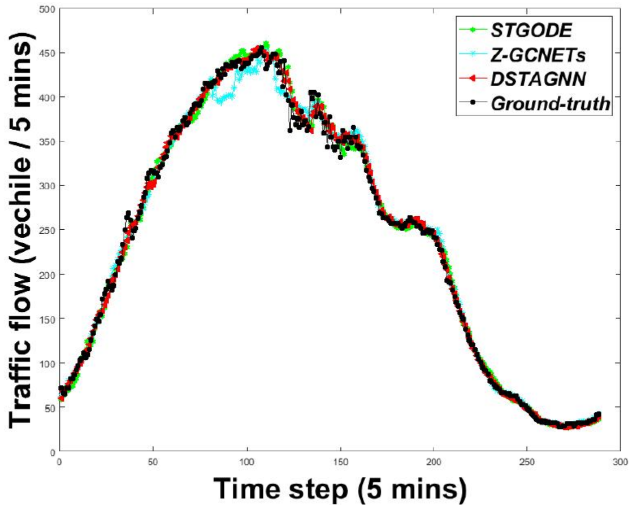

(a) Node #148 in PEMS04



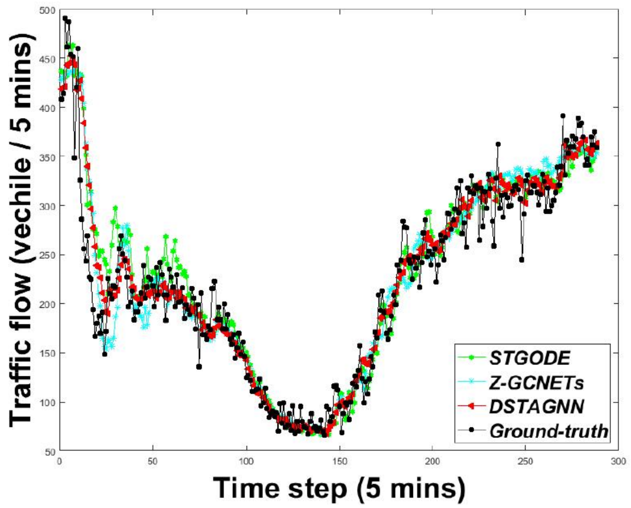

(b) Node #22 in PEMS04




(c) Node #111 in PEMS04

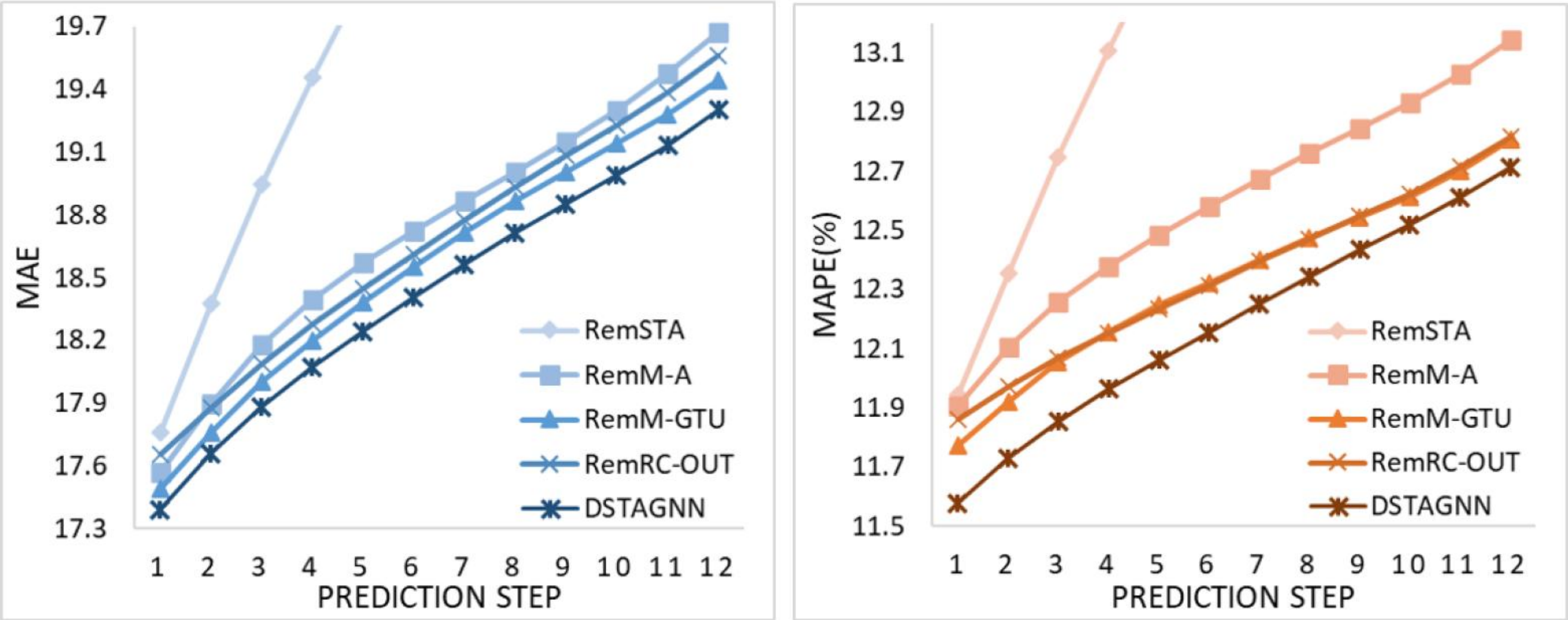



(d) Node #261 in PEMS04




(e) Node #15 in PEMS08

Fig 4. Comparison of prediction curves

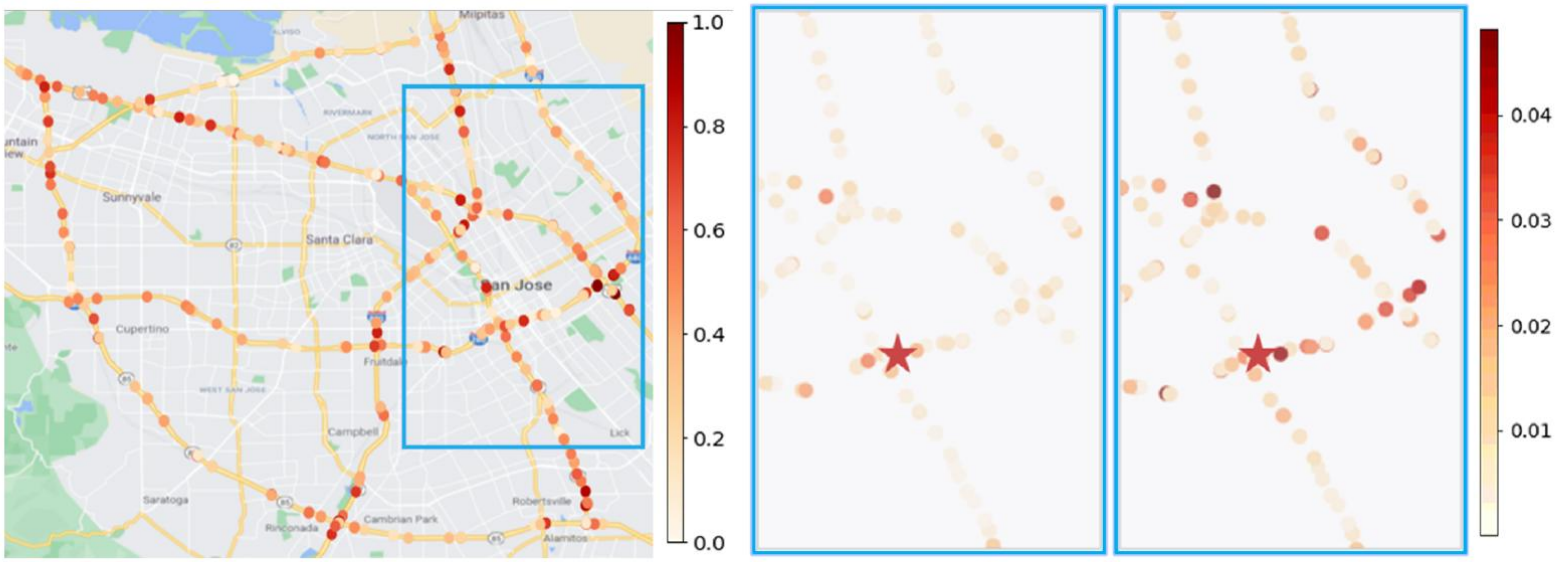



(a) MAE per prediction step (b) MAPE per prediction step

Fig 5. Ablation experiment of module effectiveness

# 3 Experimental Result

## Visualization of spatial-temporal dependency



(a) (b)

Fig 6. Spatial-temporal dependency obtained by DSTAGNN on the PEMS-BAY dataset. (a) is the global self-attention from the 1st attention head. (b) are the dependencies between the target node (red star) and its surrounding nodes obtained by the 2nd and 3rd attention heads. The zone of (b) corresponds to the blue box area in (a).

# 4 Conclusion

# 4 Conclusion

We have presented a novel deep learning framework DSTAGNN for traffic flow prediction.

- Our DSTAGNN has utilized spatial-temporal aware distance (STAD) derived from historic traffic data without relying on a predefined static adjacency matrix.
- With this method, the representation of the internal dynamic association attributes between nodes of the road network can be enhanced effectively. In addition, graph convolution operated on the Spatial-Temporal Aware Graph (STAG) generated from STAD can reduce the dependency on prior information of the road network.
- Combined with our spatial-temporal attention module and multi-receptive field gated convolution, our DSTAGNN further boosts the awareness of dynamic spatial-temporal dependency in time series data.
- Therefore, our DSTAGNN achieves new state-of-the-art performance on the four public data sets for traffic flow prediction, compared to several recent baseline methods.

# Thanks