



北京工业大学  
BEIJING UNIVERSITY OF TECHNOLOGY

# Memory-augmented dynamic graph convolution networks for traffic data imputation with diverse missing patterns

(用于具有多样缺失模式的交通数据插补的内存增强动态图卷积网络)

汇报人：闫林枝

# CONTENT\_

01 背景\_

02 方法\_

03 实验\_

04 结论\_

P

ART ONE

---

背景

01

# 背景

- 来自真实交通系统的交通数据在交通研究和应用中发挥着至关重要的作用，如预测交通状况、规划驾驶路线和优化交通流量；
- 数据丢失是一个不可避免的问题；
- 数据缺失问题严重影响了交通状况的实时监控，并进一步限制了其他下游应用；
- 因此，如何对缺失数据进行估计，即对交通数据进行补全就成为一个关键问题。

# 现有方法以及局限性

- 现有的方法无法有效地整合和利用来自不同节点在不同时间点的全局时空信息。这意味着这些方法可能在捕捉和利用整个交通网络中的时空依赖关系方面存在不足，从而影响数据插补的效果；
- 未考虑交通系统或交通数据的独特特征：动态空间依赖关系和相关缺失模式。

✓ 针对上面存在的这些问题，提出了本文的深度学习架构：  
记忆增强动态图卷积网络（**MDGCN**）

# 缺失模式

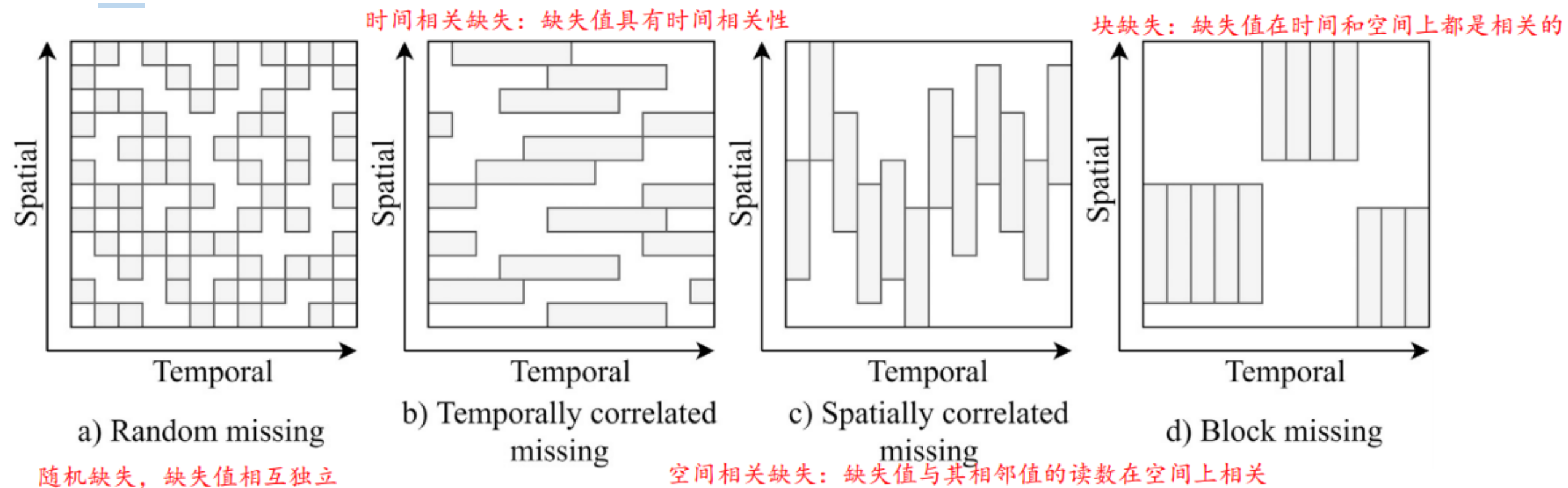


Fig. 1. Patterns of missing data.

# 预备知识

- 交通网络：加权有向图 $G=(V, E, A)$
- $A$ ：加权邻接矩阵，表示节点之间的接近性。

➤ 地理接近度

交通模式相似度

$$a_{g,ij} = \exp(-(\frac{dist_{ij}}{\delta_g})^2),$$

$$a_{s,ij} = \exp(-(\frac{\|p_i - p_j\|}{\delta_s})^2),$$

$$A_{ij} = \begin{cases} 1 & a_{g,ij} > \kappa_g \text{ and } a_{s,ij} > \kappa_s, \\ 0 & a_{g,ij} \leq \kappa_g \text{ or } a_{s,ij} \leq \kappa_s, \end{cases}$$

- 节点  $i$  和节点  $j$  在地理上接近并且在交通模式上相似， $A_{ij} = 1$ ；
- 节点  $i$  和节点  $j$  在地理上不接近或在交通模式上不相似，则为0.

# 问题定义

$x_t^v$ : 时间 $t$ 时, 在节点 $v$ 处的观测值

$$X_t = \{x_t^1, x_t^2, \dots, x_t^N\} \quad \tilde{X}_{1:T} = F(X_{1:T}, G).$$

➤ 要解决的**问题**, 给定观测数据  $X_1: X_T$  和图结构  $G$ , 交通数据插值的任务是学习一个函数  $F$ , 该函数能够在时间间隔  $[1, T]$  上对整个交通数据序列进行预测。

■ 至于这个观测值, 有可能是交通流量, 也有可能是车速, 更有可能是交通拥堵程度, 本文之后实验中的数据集是车速。

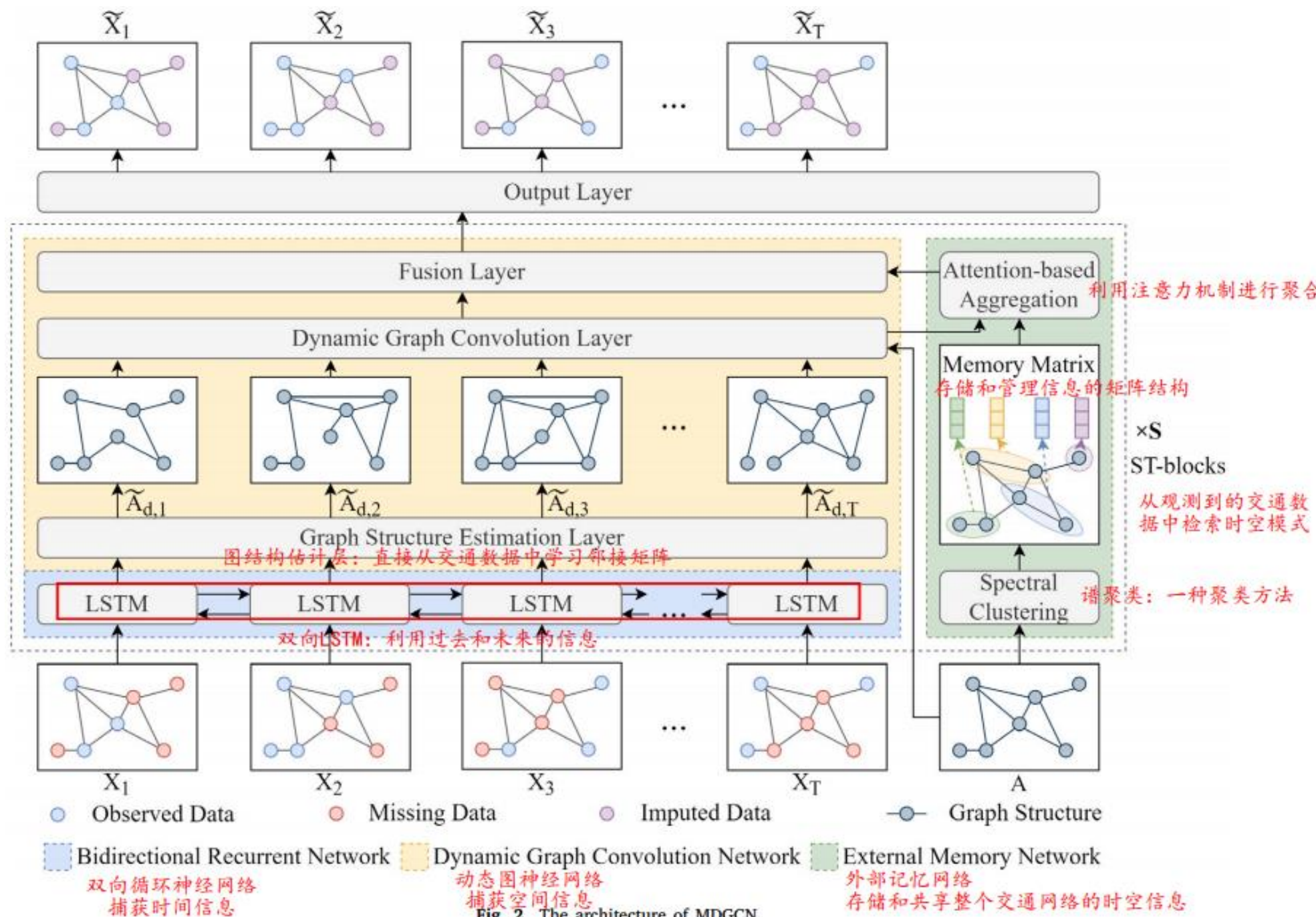


P

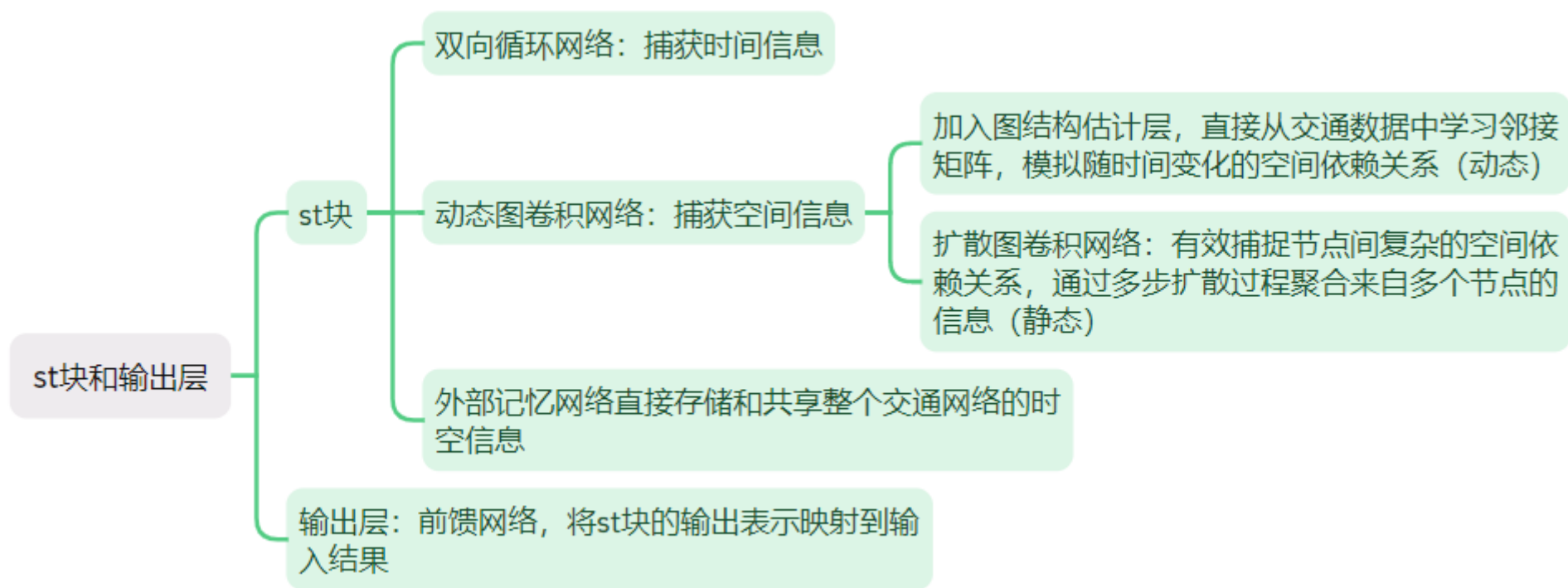
ART THREE

方法

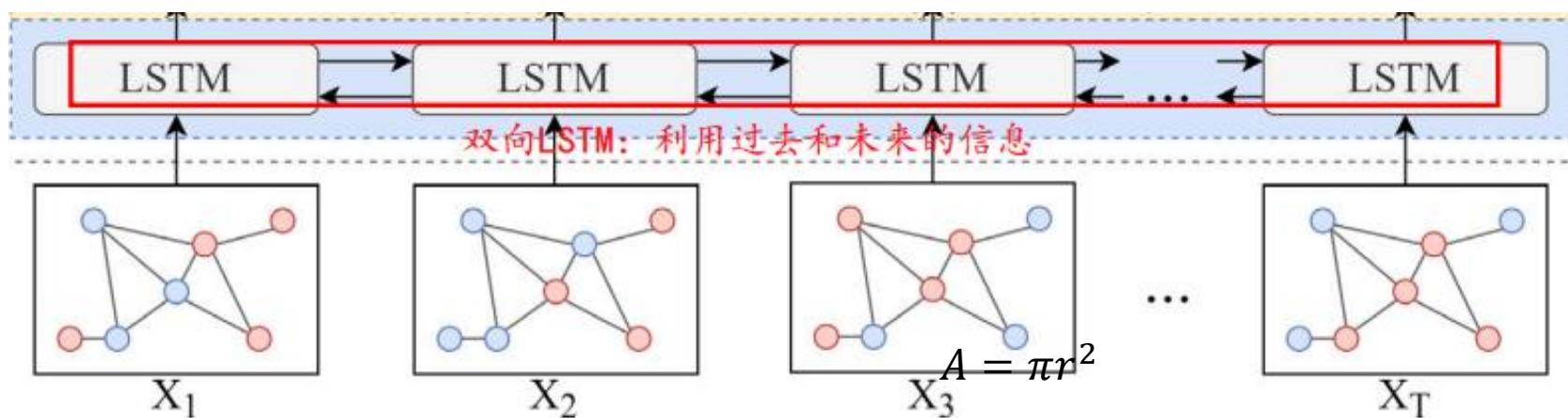
# MDGCN



# MDGCN组成



# BLSTM捕获时间序列信息



- 由两个LSTM网络组成，一个处理缺失数据之前的数据，另一个处理缺失数据之后的数据；

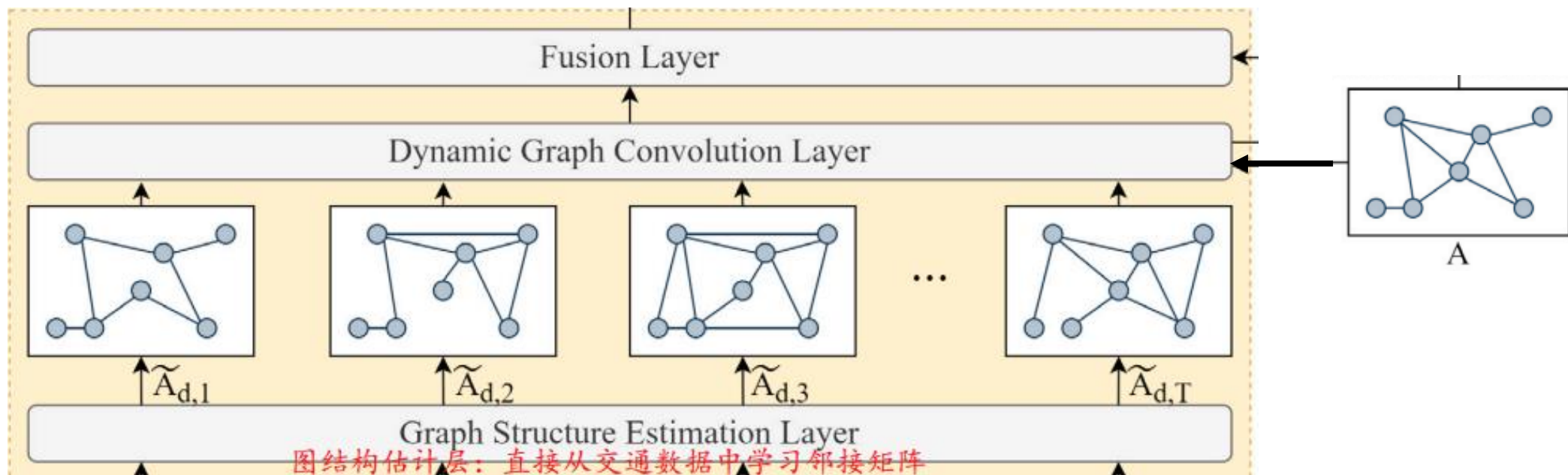
正向  $h_{f,t}, c_{f,t} = LSTM(z_l^v, \text{上一个时间步的隐藏状态向量 } h_{f,t-1}, \text{上一个时间步的单元记忆状态向量 } c_{f,t-1})$

反向  $h_{b,t}, c_{b,t} = LSTM(z_l^v, h_{b,t+1}, c_{b,t+1})$

$$z_{l,t}^v = W_l[h_{f,t}; h_{b,t}] + \beta_l$$

- 最后的输出只使用了隐藏状态向量：因为这些向量已经包含了通过LSTM单元传递的所有重要信息，而单元记忆状态向量主要用于在LSTM内部保存和更新长期依赖信息。

# 动态图卷积网络捕获空间信息



➤ **图结构评估层:** 通过 BLSTM 输出的节点特征, 动态地生成邻接矩阵

$$\tilde{A}_{d,t} = \text{SoftMax}(\text{ReLU}(Z_{l,t}W_d^1 + \beta_d^1)W_d^2 + \beta_d^2)$$

# 动态图卷积网络捕获空间信息

- **扩散图卷积网络**: 有效捕捉节点间复杂的空间依赖关系, 通过多步扩散过程聚合来自多个节点的信息, 从而提高预测或插补任务的效果。

$$DG\!C\!N(H) = \sum_{k=1}^K \left( F_k(A_f) H \theta_f^k + F_k(A_b) H \theta_b^k \right)$$

- ✓ **动态图卷积网络**: 结合上面两者, 将静态和动态的扩散结果结合起来, 全面地捕捉时空依赖关系。

$$DDG\!C\!N(Z_{l,t}) = \sum_{k=1}^K \left( F_k(A_f) Z_{l,t} \theta_f^k + F_k(A_b) Z_{l,t} \theta_b^k + F_k(\tilde{A}_{d,t}) Z_{l,t} \theta_t^k \right)$$

## 动态图卷积网络捕获空间信息

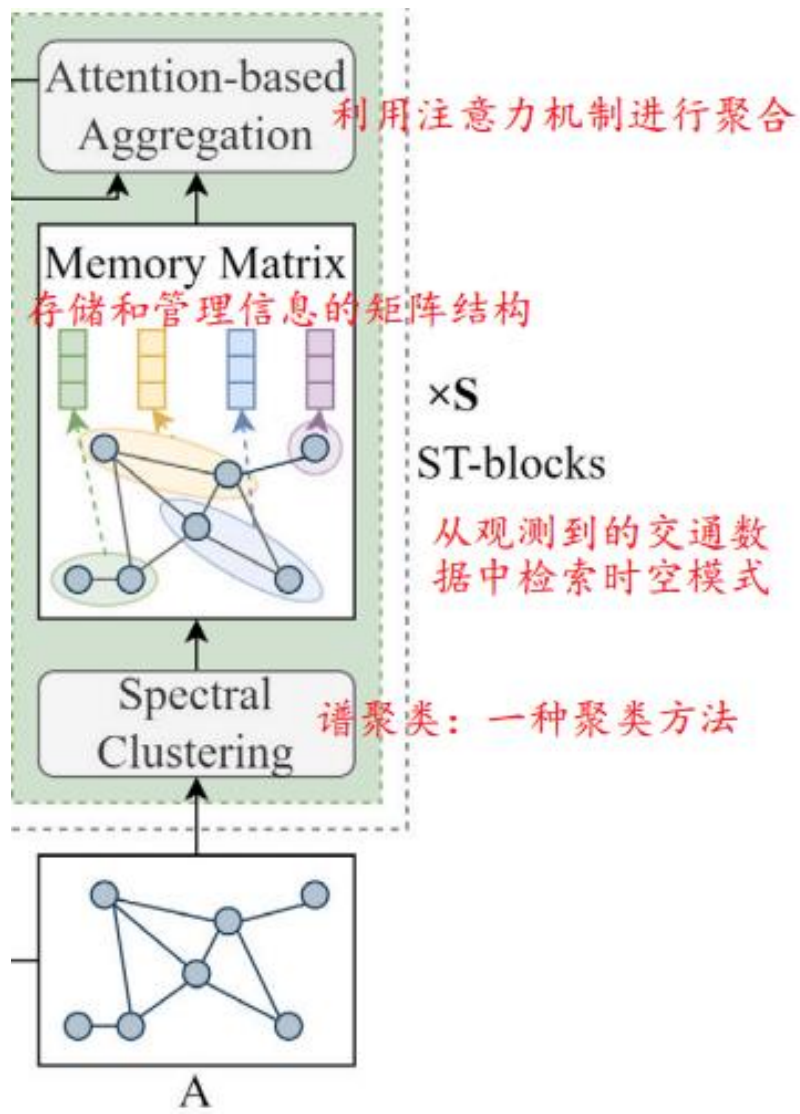
➤ 为了并行处理多个时间步，将DDGCN扩展到多维向量

$$\tilde{\mathbf{A}}_{d,1:T} = \text{SoftMax}(\text{ReLU}(\mathbf{Z}_{l,1:T} \mathbf{W}_d^1 + \boldsymbol{\beta}_d^1) \mathbf{W}_d^2 + \boldsymbol{\beta}_d^2),$$

$$\mathbf{Z}_{g,1:T} = \sum_{k=1}^K \left( F_k(\mathbf{A}_{f,1:T}) \otimes \mathbf{Z}_{l,1:T} \boldsymbol{\theta}_f^k + F_k(\mathbf{A}_{b,1:T}) \otimes \mathbf{Z}_{l,1:T} \boldsymbol{\theta}_b^k + F_k \otimes (\tilde{\mathbf{A}}_{d,1:T}) \mathbf{Z}_{l,1:T} \boldsymbol{\theta}_t^k \right)$$



# 外部记忆网络直接模拟全球时空信息



- 首先用谱聚类的方法将所有节点分为Q个组；
- 构建记忆矩阵M，每行表示一个组的共享信息；
- 为了捕捉交通数据的动态变化，使用注意力机制自适应地将不同组的模式知识传递给目标节点；

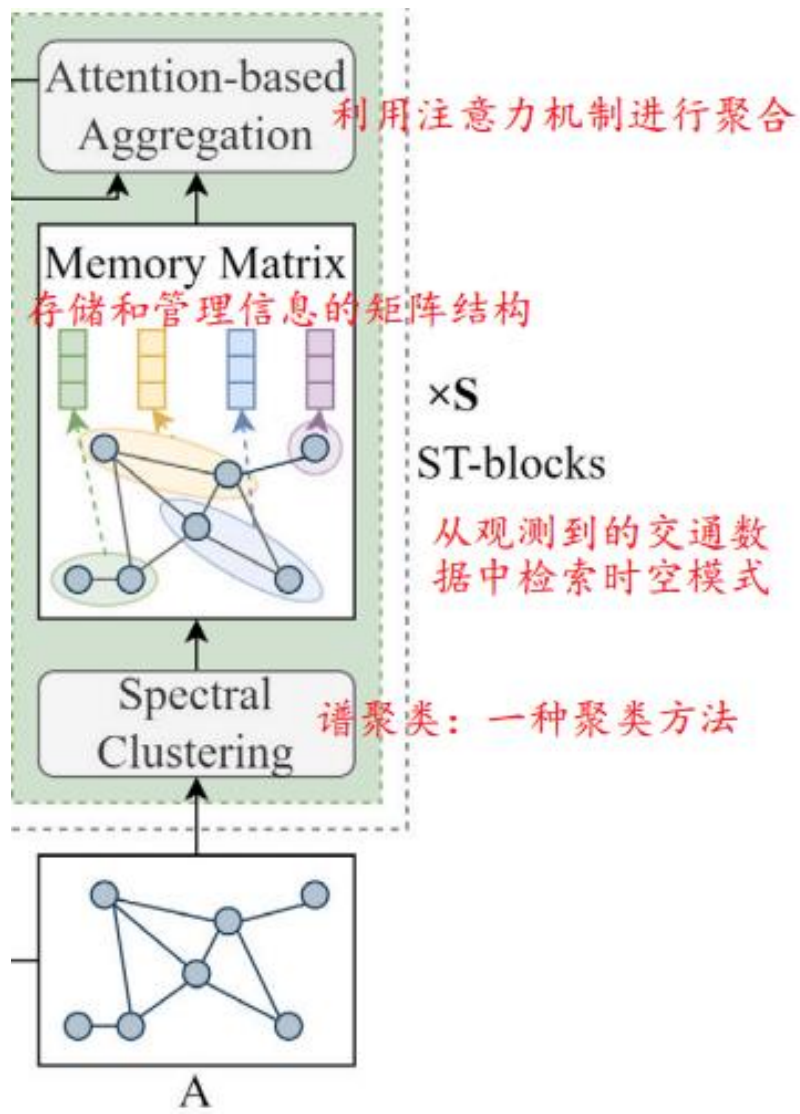
$$\mathbf{q}_t^v = \mathbf{z}_{r,t}^v \mathbf{W}_q + \mathbf{b}_q.$$

$$\mathbf{s}_t^v = \text{SoftMax}(\mathbf{q}_t^v \mathbf{M}^T),$$

$$\mathbf{g}_t^v = \sum_{g=1}^Q \mathbf{s}_{t,g}^v \mathbf{M}_g,$$



# 外部记忆网络直接模拟全球时空信息



- 将不同模块学习到的时空信息进行融合，生成下一层或最终的输出；

$$\mathbf{z}_{o,t}^v = \text{LayerNorm}(\text{ReLU}(\mathbf{W}_g \text{concat}(\mathbf{g}_t^v, \mathbf{z}_{r,t}^v) + \beta_g)),$$

- 定义聚类损失，确保注意力权重与预定义的类别一致，从而保证模型的稳定性和准确性；

$$L_{cluster} = - \sum_{i=1}^S \sum_{t=1}^T \sum_{v=1}^N o_v \log(s_{t,i}^v),$$

# 模型设置

➤ 损失函数：聚类损失+重构误差

$$L_{cluster} = - \sum_{i=1}^S \sum_{t=1}^T \sum_{v=1}^N o_v \log(s_{t,i}^v),$$

$$L_{mse} = \sum_{v=1}^N \sum_{t=1}^T (\tilde{x}_t^v - \hat{x}_t^v)^2$$

真实值    估计值

$$L(\theta) = L_{mse} + \lambda L_{cluster},$$

# 模型设置

## ➤ 训练数据的生成

---

**Algorithm 1:** Generating training samples

---

**Input** : historical data  $X \in \mathbb{R}^{N \times P}$  over period  $[1, P]$  for training, missing type  $Y$ , the size of imputation window  $T$ , batch size  $B$ , training iteration  $I$

**for**  $i$  **in**  $\{1, 2, \dots, I\}$  **do**

    Randomly choose a time point  $t$  within range  $[1, P - T]$ ;

    Obtain sampled data  $X_{sample} = X_{t:t+T}$  from  $X$ ;

**for**  $j$  **in**  $\{1, 2, \dots, B\}$  **do**

        Randomly generate a missing ratio  $r$  within range  $(0, 1)$ ;

        Generate a mask matrix  $E_{sample}$  according to  $Y$  and  $r$ ;

$X_{train} = X_{sample} \odot E_{sample}$ ;

**end**

    Use sets  $\{X_{train}^{1:B}\}$  to train DSTCGN;

**end**

---

随机生成一个训练子集  
用于模型训练的二值掩码矩阵

P

ART FOUR

实验

# 数据

- ✓ **METR-LA**: 洛杉矶高速公路网络上的207个环路传感器收集的交通速度数据, 定义 $\kappa g=0.1$ ,  $\kappa s=0.1$ ;
- ✓ **INRIX-SEA**: 城市道路网络上收集的交通速度数据, 选取了223条相连道路组成的路网样本进行实验, 定义 $\kappa g=0.$ ,  $\kappa s=0.8$ ;

# 不同缺失模式和缺失比率

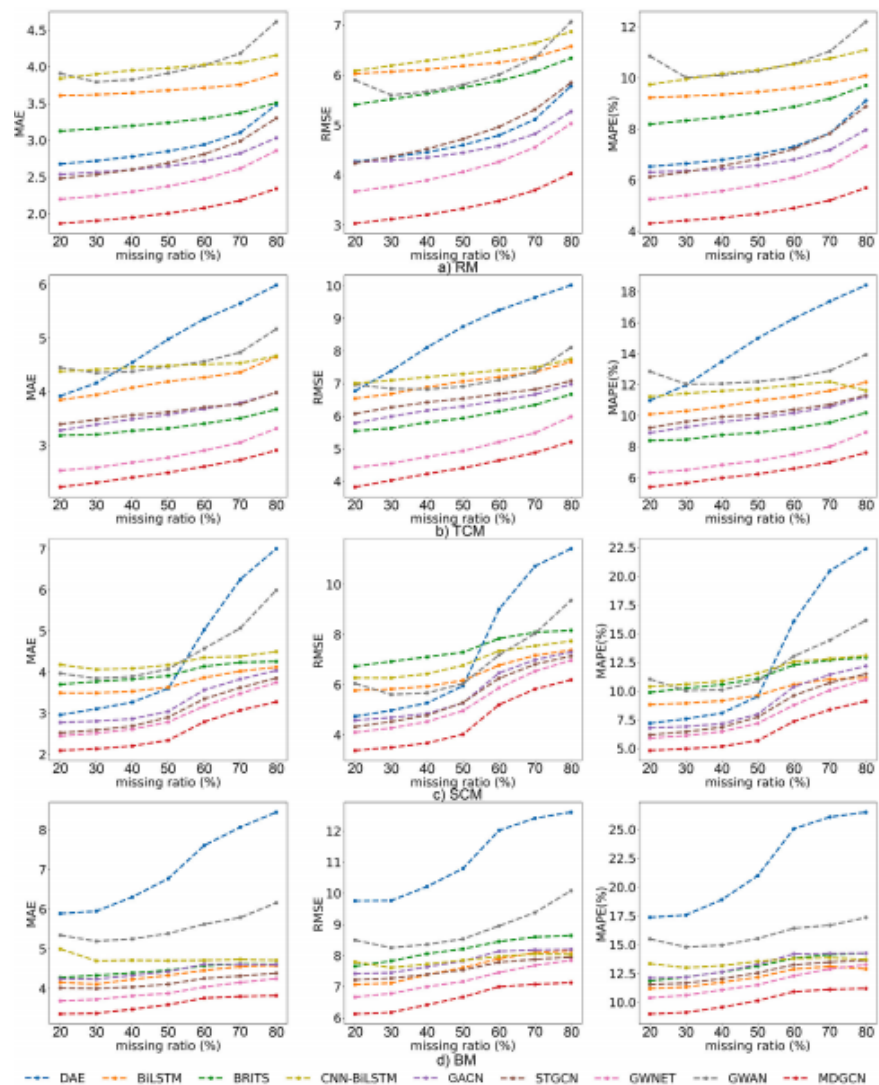


Fig. 3. Performance comparison for different missing patterns on METR-LA data. Results averaged over 3 independent runs.

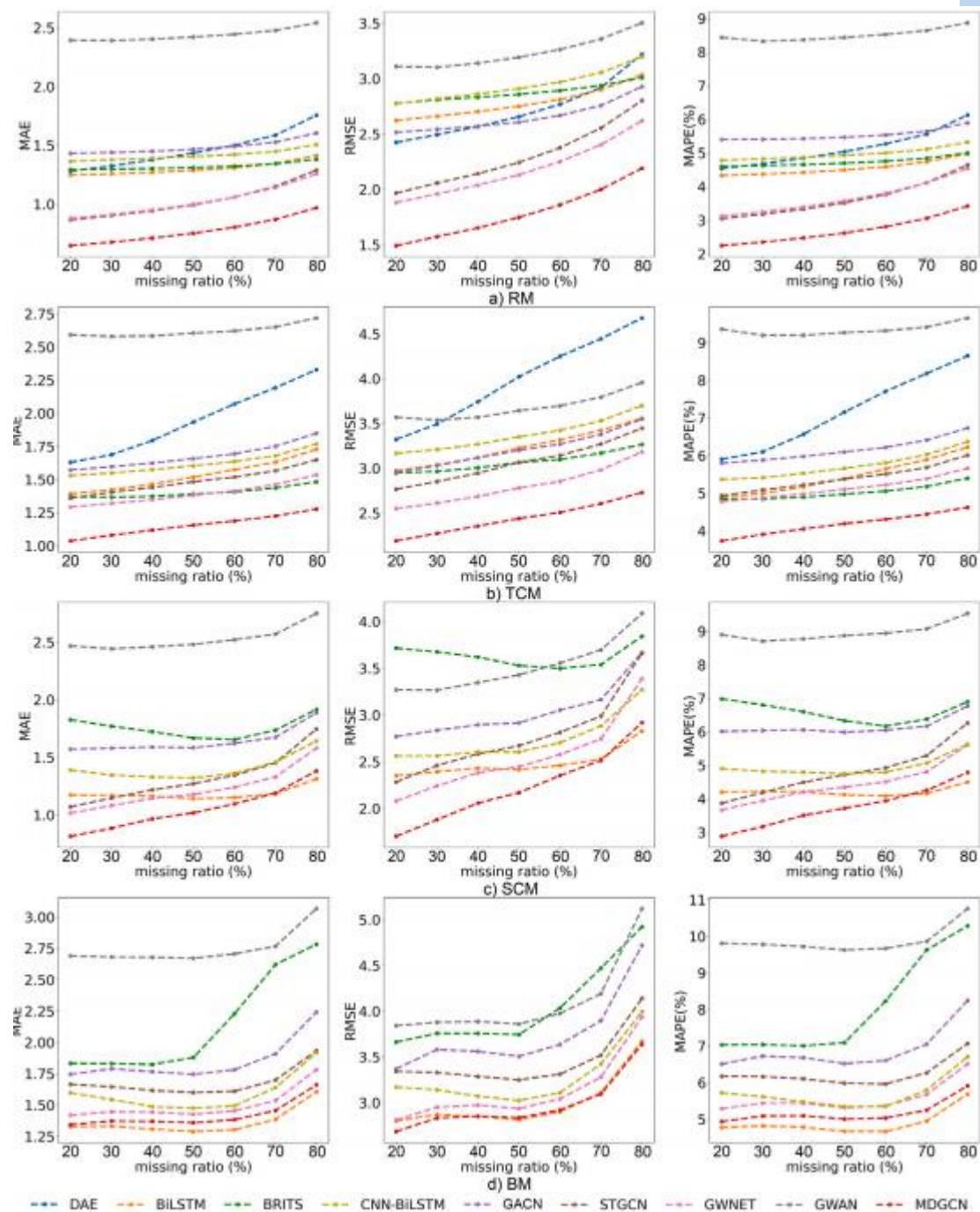


Fig. 4. Performance comparison for different missing patterns on INRIX-SEA data. Results averaged over 3 independent runs. The results of DAE for SCM and BM on INRIX-SEA data is presented separately in Appendix B.



# 结果示例

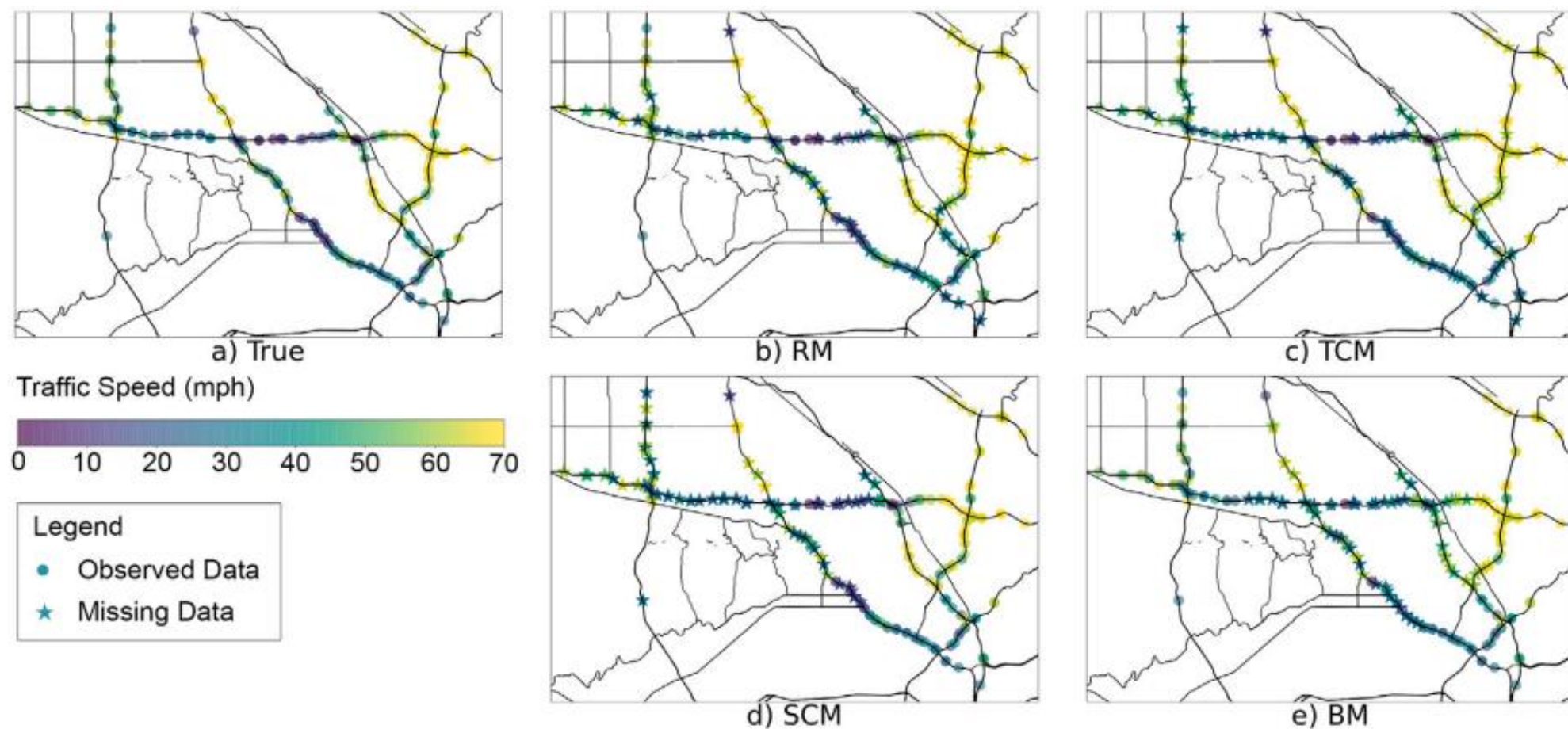


Fig. 5. The imputation result of the METR-LA data at 18:00 on 2012-05-24 using MDGCN.

# 外部记忆网络的影响

**Table 2**  
Performance (RMSE/MAPE) comparison of different variants of the external memory network on METR-LA data. Results averaged over 3 independent runs.

Model		Missing ratios						
		20%	30%	40%	50%	60%	70%	80%
RM	-Memory	3.072/4.37	3.157/4.49	3.246/4.60	3.373/4.76	3.527/4.98	3.753/5.30	4.096/5.80
	-Cluster	3.062/4.35	3.147/4.47	3.235/4.57	3.360/4.73	3.513/4.95	3.735/5.26	4.075/5.75
	-Attn	3.056/4.33	3.142/4.45	3.229/4.55	3.355/4.72	3.507/4.94	3.729/5.25	4.066/5.75
	MDGCN	3.033/4.30	3.120/4.42	3.207/4.52	3.331/4.68	3.483/4.90	3.700/5.20	4.035/5.69
TCM	-Memory	3.963/5.83	4.161/5.96	4.384/6.20	4.576/6.47	4.819/6.83	5.057/7.23	5.388/7.84
	-Cluster	3.878/5.39	4.064/5.63	4.271/5.96	4.448/6.20	4.680/6.55	4.919/6.95	5.240/7.55
	-Attn	3.858/5.38	4.041/5.64	4.245/5.98	4.420/6.23	4.651/6.59	4.890/6.98	5.218/7.58
	MDGCN	3.828/5.36	4.024/5.61	4.231/5.94	4.410/6.19	4.643/6.55	4.879/6.94	5.212/7.55
SCM	-Memory	3.401/4.91	3.521/5.06	3.695/5.27	4.073/5.86	5.483/8.12	6.087/9.13	6.419/9.83
	-Cluster	3.394/4.85	3.513/5.00	3.683/5.20	4.050/5.76	5.330/7.64	5.873/8.45	6.194/9.10
	-Attn	3.380/4.87	3.500/5.02	3.672/5.23	4.027/5.77	5.209/7.47	5.806/8.46	6.189/9.24
	MDGCN	3.379/4.83	3.500/4.97	3.671/5.18	4.023/5.69	5.205/7.36	5.826/8.38	6.195/9.12
BM	-Memory	6.253/9.49	6.347/9.64	6.592/10.13	6.833/10.75	7.231/11.73	7.289/11.84	7.333/11.90
	-Cluster	6.222/9.38	6.296/9.52	6.521/9.96	6.776/10.63	7.150/11.62	7.175/11.64	7.243/11.73
	-Attn	6.109/9.03	6.172/9.17	6.409/9.59	6.645/10.18	6.987/11.02	7.067/11.26	7.148/11.40
	MDGCN	6.119/9.01	6.173/9.13	6.414/9.56	6.660/10.13	7.001/10.92	7.072/11.10	7.132/11.20



# 动态邻接矩阵的影响

Table 3

Performance (RMSE/MAPE) Comparison of Different Adjacency Matrix Configurations on METR-LA data. Results averaged over 3 independent runs.

Model		Missing ratios						
		20%	30%	40%	50%	60%	70%	80%
RM	Fix-only	3.102/4.41	3.191/4.53	3.283/4.64	3.418/4.81	3.582/5.05	3.822/5.38	4.194/5.92
	Fix-SAAM	3.090/4.39	3.176/4.50	3.269/4.62	3.397/4.78	3.554/5.01	3.785/5.34	4.131/5.87
	GAT	3.473/5.00	3.546/5.10	3.631/5.20	3.759/5.35	3.917/5.57	4.160/5.89	4.571/6.48
	Dynamic-only	3.108/4.41	3.195/4.53	3.285/4.64	3.410/4.80	3.565/5.03	3.785/5.33	4.119/5.83
	Fix-dynamic	<u>3.033/4.30</u>	<u>3.120/4.42</u>	<u>3.207/4.52</u>	<u>3.331/4.68</u>	<u>3.483/4.90</u>	<u>3.700/5.20</u>	<u>4.035/5.69</u>
TCM	Fix-only	4.154/5.86	4.447/6.29	4.745/6.78	4.992/7.19	5.254/7.61	5.510/8.10	5.850/8.79
	Fix-SAAM	3.970/5.51	4.188/5.79	4.431/6.18	4.635/6.49	4.887/6.87	5.139/7.31	5.486/7.95
	GAT	4.832/6.93	5.194/7.52	5.501/8.11	5.779/8.61	6.017/9.02	6.262/9.55	6.553/10.20
	Dynamic-only	3.968/5.54	4.135/5.77	4.350/6.12	4.534/6.38	4.746/6.71	4.980/7.12	5.304/7.73
	Fix-dynamic	<u>3.828/5.41</u>	<u>4.027/5.65</u>	<u>4.231/5.98</u>	<u>4.410/6.24</u>	<u>4.643/6.59</u>	<u>4.879/6.99</u>	<u>5.212/7.61</u>
SCM	Fix-only	3.420/4.96	3.545/5.12	3.719/5.34	4.120/6.00	5.773/8.78	6.570/10.42	6.903/11.14
	Fix-SAAM	3.383/4.86	3.502/5.01	3.675/5.22	4.053/5.80	5.421/7.93	6.101/9.06	6.482/9.90
	GAT	3.665/5.37	3.782/5.51	3.958/5.73	4.368/6.38	5.989/9.21	6.793/10.87	7.177/11.86
	Dynamic-only	3.483/5.02	3.609/5.17	3.783/5.39	4.202/6.00	5.350/7.66	5.878/8.54	<u>6.180/9.19</u>
	Fix-dynamic	<u>3.379/4.83</u>	<u>3.500/4.97</u>	<u>3.671/5.18</u>	<u>4.023/5.69</u>	<u>5.205/7.36</u>	<u>5.826/8.38</u>	6.195/9.12
BM	Fix-only	6.662/10.41	6.730/10.55	6.947/11.01	7.262/11.85	7.707/13.18	7.836/13.49	7.948/13.72
	Fix-SAAM	6.264/9.54	6.359/9.67	6.559/10.08	6.781/10.68	7.244/11.93	7.374/12.22	7.465/12.37
	GAT	6.962/11.14	7.044/11.28	7.327/12.02	7.527/12.71	8.060/14.23	8.186/14.43	8.214/14.45
	Dynamic-only	6.350/9.63	6.394/9.69	6.625/10.08	6.835/10.67	7.108/11.40	7.170/11.58	7.222/11.64
	Fix-dynamic	<u>6.119/9.01</u>	<u>6.173/9.13</u>	<u>6.414/9.56</u>	<u>6.660/10.13</u>	<u>7.001/10.92</u>	<u>7.072/11.10</u>	<u>7.132/11.20</u>

# 动态邻接矩阵的影响

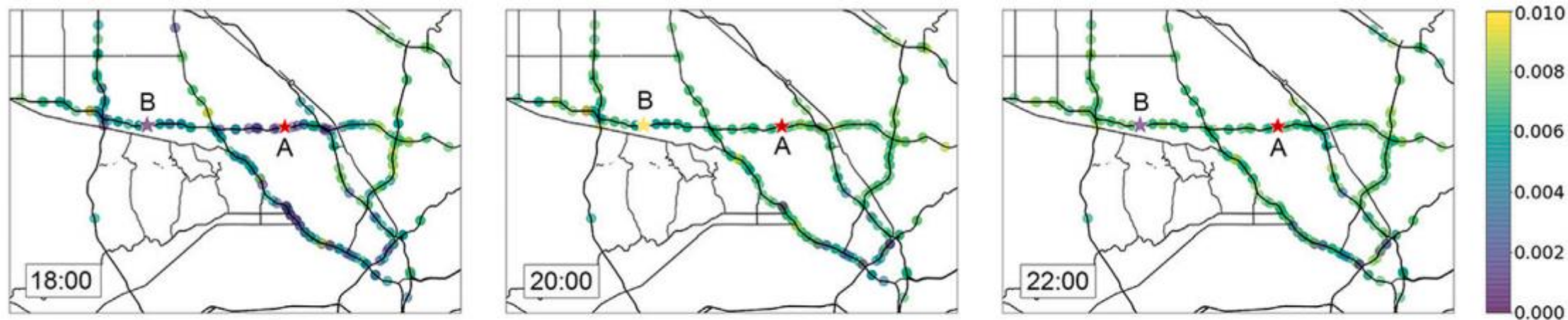


Fig. 6. Transition weights between sensor A and the other sensors in METR-LA data from 18:00 to 22:00 on 2012-05-24. Sensor B is marked as an example to highlight the dynamic transition weights between sensors A and B.

✓ 证实了捕捉动态空间依赖性的必要性。

P

ART FIVE

---

结论

05

# 结论

## ✓ 填补了两个研究空白

- ❑ 不能直接利用来自不同时间的不同节点的全球时空信息；
- ❑ 没有很好地处理交通数据的动态空间依赖性和相关缺失模式。

## ✓ 提出新的深度学习框架MDGCN

- ❑ 由多个时空块组成，以实现时空信息的捕获；
- ❑ 每个块包含捕获时间相关性的双向循环层和捕获空间相关性的扩散图卷积层；
- ❑ 引入了一个外部存储网络来存储和共享不同节点的全局时空信息；
- ❑ 提出了一种图结构估计技术，直接从交通数据中学习动态空间依赖关系。

# 感谢观看！

汇报人：闫林枝

> GOODBYE <