VIRTUAL NODE IMPROVES LONG-TERM TRAFFIC PREDICTION

MIT-UF-NEU Summer Research 2024

Xiaoyang Cao, Tsinghua University

Supervisor(s):

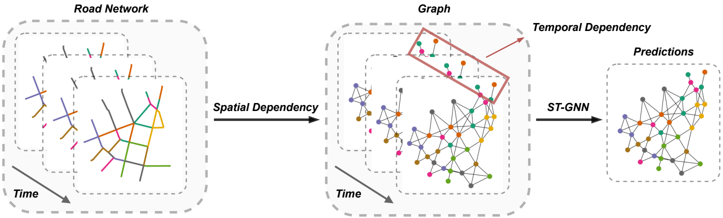
Dingyi Zhuang

Table of Contents

Traffic Prediction: Predicting future traffic conditions such as traffic flow and speed from historical traffic data.

- **Spatial Dependency:** Traffic conditions at one location are influenced by conditions at neighboring locations.
- **Temporal Dependency:** Traffic conditions at the current time are influenced by conditions from previous times.

Long-term Traffic Prediction: Time span exceeds **one hour** (Wang, Su, and Ding 2020).



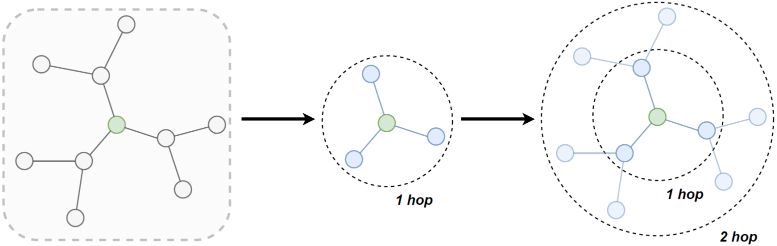
Spatial Temporal Graph Neural Networks (ST-GNNs): Simultaneously managing spatio-temporal dependencies by considering time-varying graph structures.

- **Message-passing in GNN**: GNNs operate by recursively aggregating and updating neighboring node features.

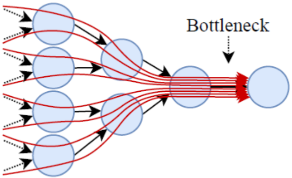
$$h_v^{(t)} = \mathrm{UPDATE}^{(t)}\left(h_v^{(t-1)}, \mathrm{AGG}^{(t)}\left(\left\{h_u^{(t-1)} \mid u \in \mathcal{N}(v)\right\}\right)\right) \quad (1)$$

$\mathcal{N}(v)$ is the set of neighboring nodes of node v.

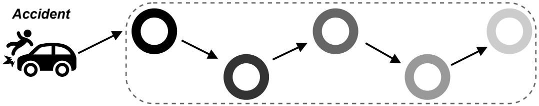
- **Bottleneck of GNNs:** information from a node's exponentially-growing receptive field is compressed into a fixed-size vector.



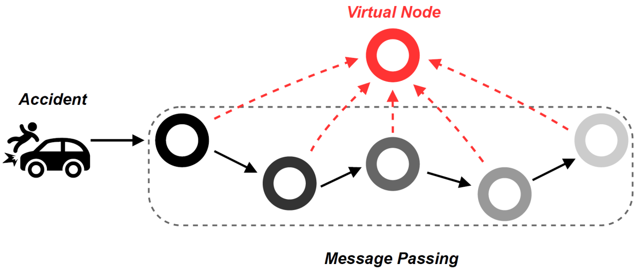
- **Over-squashing** (Alon and Yahav 2021): Information from distant nodes is overly compressed, reducing the accuracy of **long-term** forecasting.



- **A Transportation Example:** An accident occurring at the westernmost point of a road network needs to propagate through multiple intermediate points to reach the easternmost point.



Virtual Node: Virtual nodes are connected to all real nodes in the network, acting as intermediary hubs facilitating information aggregation across the entire graph.

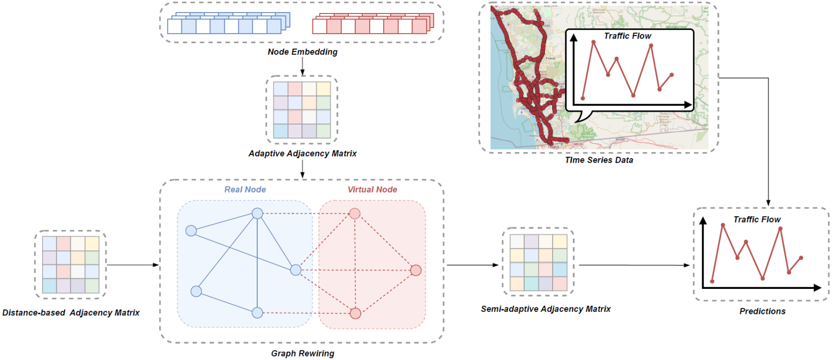


- **Mitigating over-squashing**
- **Improving long-term prediction accuracy**
- **Providing model explainability**

Table of Contents

General Framework

2 Methodology

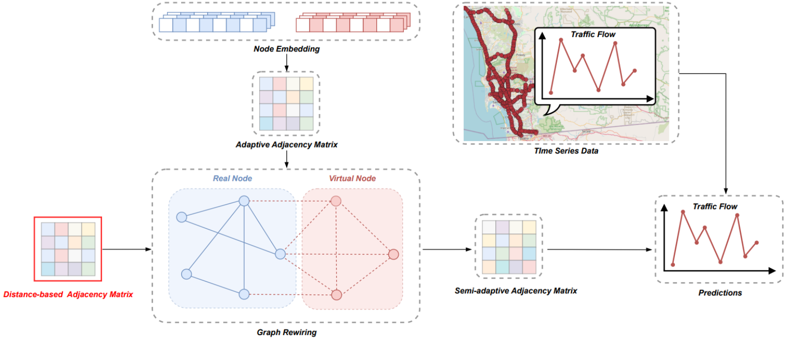


Adjacency Matrix:

- **Distance-based**: Leaverage geographical information.
- **Adaptive**: task-specific learning.
- **Semi-adaptive**: Integrate distance-based and adaptive matrix.

Distance-based Adjacency Matrix

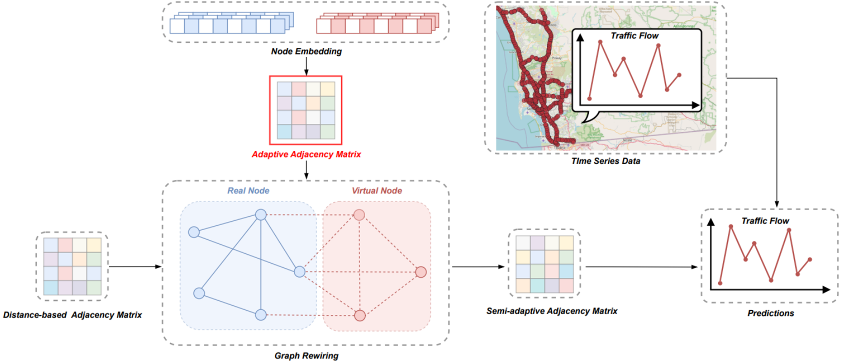
2 Methodology



- d_{ij} is the distance between nodes i and j. Smaller distance indicates stronger connection.

$$\mathbf{A}_{\text{dist, ij}} = \begin{cases} \exp\left(-\frac{d_{ij}^2}{\sigma^2}\right) & \text{if } \exp\left(-\frac{d_{ij}^2}{\sigma^2}\right) \geq r \\ 0 & \text{otherwise} \end{cases} \tag{2}$$

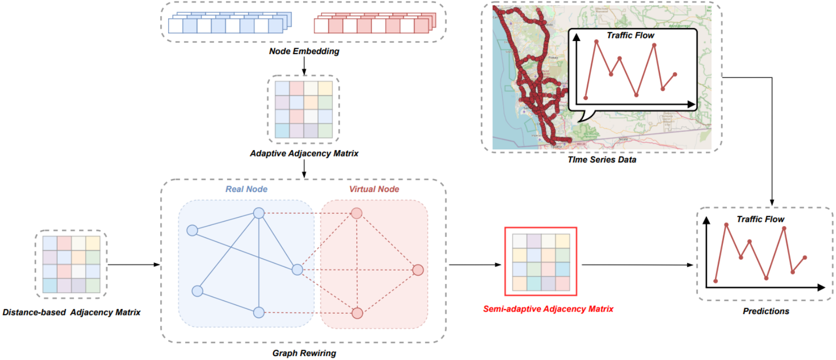
Adaptive Adjacency Matrix



- Consturcting anti-symmetric matrix for uni-directional connections (Wu et al. 2020):

$$\mathbf{A}_{\text{adapt}} = \text{ReLU}(\mathbf{E}_1 \cdot \mathbf{E}_2^T - \mathbf{E}_2 \cdot \mathbf{E}_1^T) \quad (3)$$

- $\mathbf{E}_1, \mathbf{E}_2 \in \mathbb{R}^{(|\mathcal{V}|+n_v)\times d}$ are learnable node embedding matrices. $|\mathcal{V}|$ is the number of real nodes, and n_v is the number of virtual nodes.



- Concatenating $\mathbf{A}_{dist}$ and $\mathbf{A}_{adapt}$

$$\mathbf{A}_{\text{semi}} = \begin{bmatrix} \mathbf{A}_{\text{dist}} & \mathbf{A}_{\text{adapt,real\_to\_virtual}} \\ \mathbf{A}_{\text{adapt,virtual\_to\_real}} & \mathbf{A}_{\text{adapt,virtual\_nodes}} \end{bmatrix} \tag{4}$$

Table of Contents

Dataset

The dataset used for this research is sourced from the San Diego (SD) subdataset in the LargeST benchmark dataset (Liu et al. 2024).

Attribute	**Value**
Nodes	716
Edges	17,319
Average Degree	24.2
Density	0.0338
Time Range	01/01/2017 – 12/31/2021
Sampling Rate	5 minutes
Time Frames	525,888
Data Points	0.38B

Table: Detailed Characteristics of the SD Dataset. B: billion (10^9)

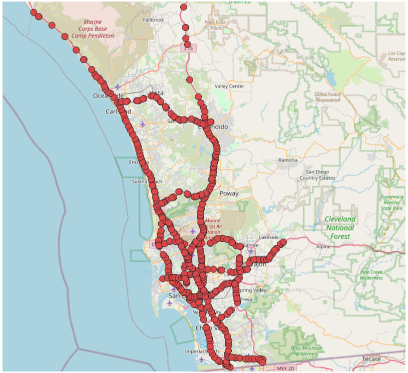


Figure: Visualization of Sensor Locations in San Diego

Key Research Questions

- **Q1:** Do virtual node alleviate over-squashing?
- **Q2:** Do virtual node improve long-term traffic prediction?
- **Q3:** How does virtual node provide explainability?

Influence score

Q1: Do virtual node alleviate over-squashing?

Definition: Layer-wise Sensitivity

Layer-wise Sensitivity[a] is defined as the L_1 norm of the gradient of the k-th layer's embedding of node x $h_x^{(k)}$ with respect to the input layer's embedding of node y $h_y^{(0)}$:

$$\text{Sensitivity}_k(x, y) = \left\| \frac{\partial h_x^{(k)}}{\partial h_y^{(0)}} \right\|_1$$

[a] Keyulu Xu et al. (2018). "Representation learning on graphs with jumping knowledge networks". In: *International conference on machine learning*. PMLR, pp. 5453–5462.

- Influence score:

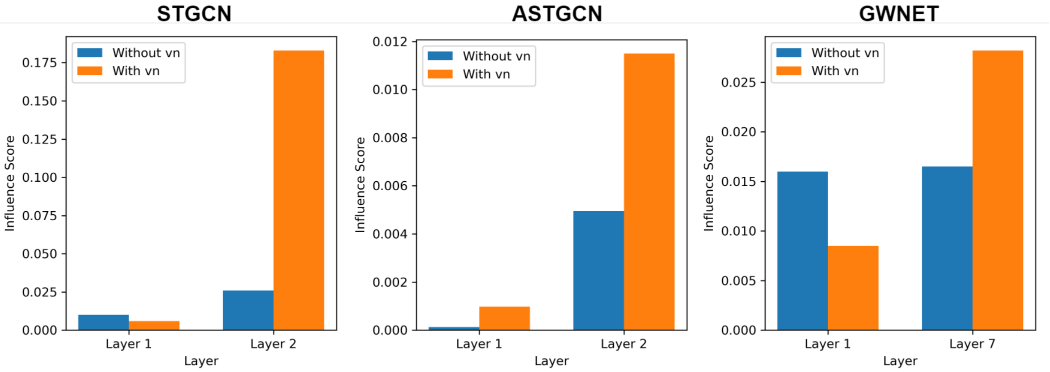
$$I_k(x_1, x_2) = \frac{\text{Sensitivity}_k(x_1, x_2) + \text{Sensitivity}_k(x_2, x_1)}{2}$$

  x_1, x_2 are the two geographically most distant nodes in the graph.

- A **lower** influence score indicates a **more severe** over-squashing problem

Impact of Virtual Nodes on Over-squashing

3 Experiment



- Adding virtual node **increases influence score as the number of layers increases.**

Long-term Performance

Q2: Do virtual node improve long-term traffic prediction?

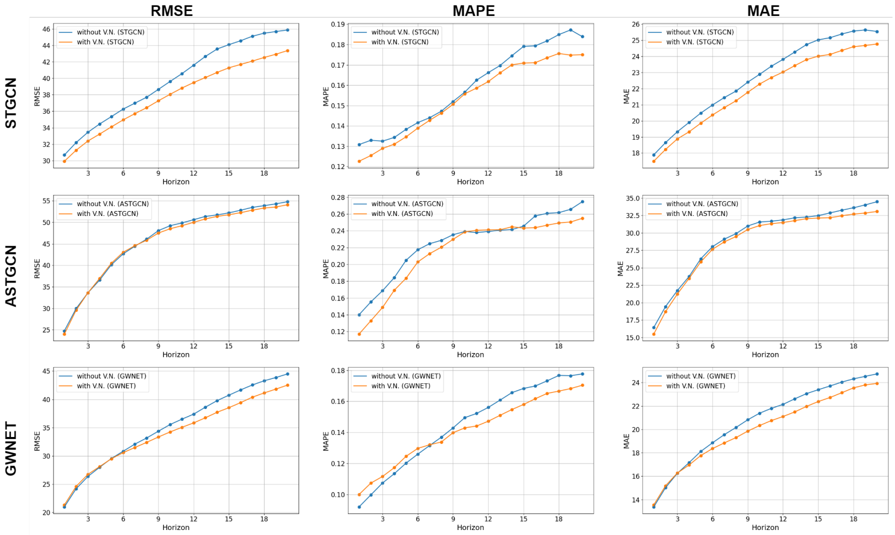
Table: Improvement in Traffic Prediction Accuracy with and without Virtual Nodes

Model	**MAE (%)**		**MAPE (%)**		**RMSE (%)**	
	Overall	**Long-term**	**Overall**	**Long-term**	**Overall**	**Long-term**
STGCN	3.22%	**3.71%**	3.31%	**4.23%**	4.81%	**6.13%**
ASTGCN	2.01%	**2.13%**	**4.67%**	3.14%	0.789%	**1.09%**
GWNET	3.32%	**3.98%**	2.43%	**5.37%**	3.19%	**4.91%**

- **Overall**: average values in 5-100 minutes; **Long-term**: average values in 75-100 minutes.
- Virtual nodes enhance traffic prediction accuracy, especially in **long-term prediction**.

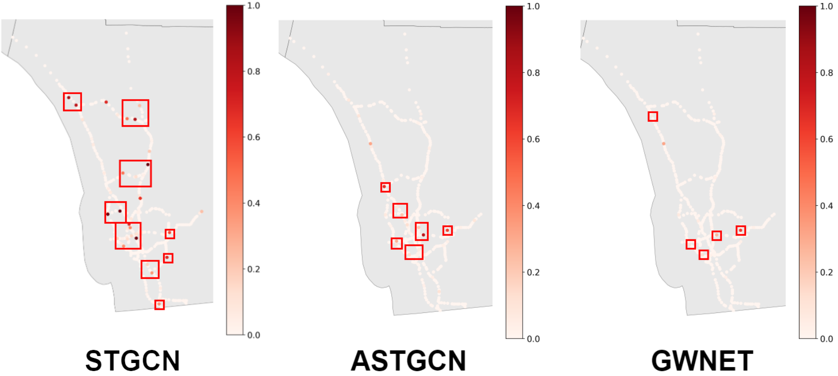
Long-term Performance

3 Experiment



Q3: How does virtual node provide explainability?

Figure: Heatmap of real to virtual adjacency matrix



- Weights at road intersections are higher.

Future Plan

- Tested models:
  - **Distance-based adjacency matrix:** STGCN, ASTGCN
  - **Distance-based + adaptive adjacency matrix:** GWNET
- Next step: Test on models with fully adaptive adjacency matrix, such as AGCRN.
- Target journal: Transportation Research Part C: Emerging Technologies

Selected References

- Alon, Uri and Eran Yahav (2021). "On the Bottleneck of Graph Neural Networks and its Practical Implications". In: *International Conference on Learning Representations*.
- Liu, Xu et al. (2024). "Largest: A benchmark dataset for large-scale traffic forecasting". In: *Advances in Neural Information Processing Systems* 36.
- Wang, Zhumei, Xing Su, and Zhiming Ding (2020). "Long-term traffic prediction based on lstm encoder-decoder architecture". In: *IEEE Transactions on Intelligent Transportation Systems* 22.10, pp. 6561–6571.
- Wu, Zonghan et al. (2020). "Connecting the dots: Multivariate time series forecasting with graph neural networks". In: *Proceedings of the 26th ACM SIGKDD international conference on knowledge discovery & data mining*, pp. 753–763.
- Xu, Keyulu et al. (2018). "Representation learning on graphs with jumping knowledge networks". In: *International conference on machine learning*. PMLR, pp. 5453–5462.

VIRTUAL NODE IMPROVES LONG-TERM TRAFFIC PREDICTION

Thank you for listening !
Any Questions ?