Regional Transferability of Graph Neural Networks for Traffic Forecasting

How does the GNN traffic forecasting model, trained with long-horizon historical data from one region, perform in regions lacking historical traffic data, and how are these performance variations correlated with spatial differences among the regions?

Author: Ivans Kravcevs
I.Kravcevs@student.tudelft.nl

Supervisor: Elena Condeguti

1 Background

- Modern traffic management systems require effective models for traffic forecasting
- Graph Neural Networks (GNN) show one of the best performance in traffic forecasting due to the ability to capture spatiotemporal dependencies [1]
- Transferability in traffic forecasting involves model training on one region's data and its application in another region
- It is assumed that GNN performance is strongly dependent on the spatial structure of the training and transfer traffic sensor graph [2]

2 Research questions

- What is the performance of the GNN model in the traffic forecasting of the training region?
- What is the performance of the same model on the unexplored structurally different regions?
- How does the structural difference between training and transfer regions correlate with the model's performance in the transfer region?

3 Methodology

- **DCRNN** [3] model was used as the main model and it was trained over a **10-sensor** set and a **50-sensor** set in the region of Los Angeles (**METR-LA**)
- Masked Mean Absolute Error (MAE) is used as the main **performance measure**, and Root Mean Squared Error (RMSE) is an additional
- Models were **transferred** to regions with a similar amount of sensors in the San Jose area (**PEMS-BAY**)
- **Distance** between areas was measured using Frobenius distance, Absolute Sum distance, and Cosine distance operating on **graph adjacency matrixes**
- Multiple **masks** (0, 20000 and 40000) were introduced to cover missing distance values in a graph adjacency matrix
- The **correlation** between distance and model performance was measured using Pearson coefficient
- **Bucketed simulated annealing (BSA)** approach was introduced to find graphs with diverse distances

$$\mathrm{FroD}(A,B)=\sqrt{\sum_{i=1}^{m}\sum_{j=1}^{n}|a_{ij}-b_{ij}|^2}$$

$$\mathrm{AbsSum}(A,B)=\sum_{i=1}^{m}\sum_{j=1}^{n}|a_{ij}-b_{ij}|$$

where:

A, B : *compared matrices*
m, n : *amount of rows and columns in the matrices*
a_{ij}, b_{ij} : *values of a and b in the row i and column j*

$$\mathrm{CosD}(A,B)=\frac{1}{2}\left(\frac{1}{n}\sum_{i=1}^{n}1-\frac{\mathbf{A}_{i\cdot}\cdot\mathbf{B}_{i\cdot}}{\|\mathbf{A}_{i\cdot}\|\|\mathbf{B}_{i\cdot}\|}+\frac{1}{m}\sum_{j=1}^{m}1-\frac{\mathbf{A}_{\cdot j}\cdot\mathbf{B}_{\cdot j}}{\|\mathbf{A}_{\cdot j}\|\|\mathbf{B}_{\cdot j}\|}\right)$$

where:

$\mathbf{A}_{i\cdot}, \mathbf{B}_{i\cdot}$: *the row vectors from matrices A and B*
$\mathbf{A}_{\cdot j}, \mathbf{B}_{\cdot j}$: *the column vectors from matrices A and B*

Figure 1: Graph distance metrics

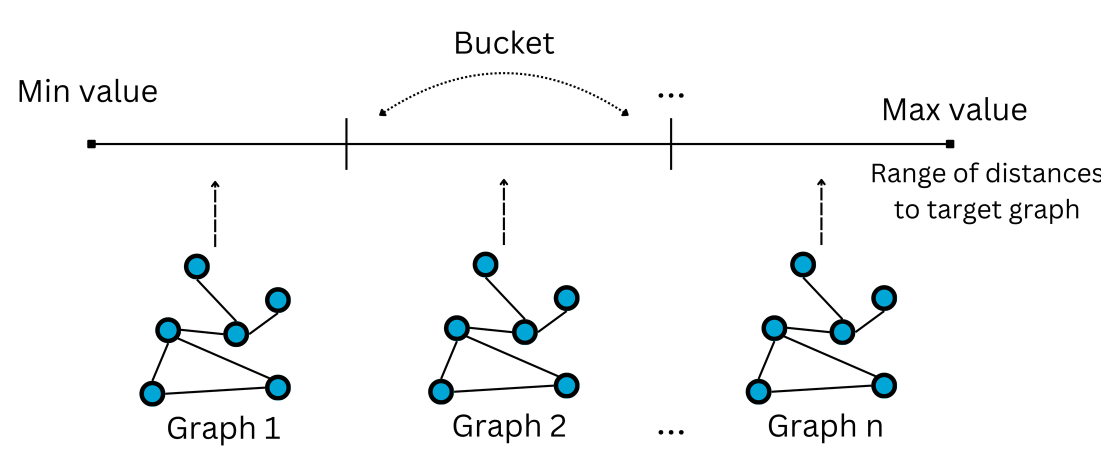

Figure 2: Bucketed simulated annealing visualization

5 Conclusions

- GNN performance is highly dependent on the selected training region
- Models trained on low spatial correlation regions transfer better by avoiding overfitting
- The structural differences between the training and transfer regions are not strongly correlated with the model's performance.
- The current graph distance metrics mostly capture incorrect regional spatial patterns

6 Future work

- Training the model for different datasets and regional scenarios
- Exploration of all the proposed metric configurations using BSA
- Exploration of other possibilities for the mask values and graph distance metrics
- Deeper exploration of the cosine metrics for the transferability

4 Experiments and Results

Model performance in the training region

Dataset	MAE	RMSE
Full dataset	3.60	7.59
50-sensor subset	**2.92**	**6.43**
10-sensor subset	**4.75**	**14.56**

Table 1: Model performance for 1-hour predictions on sensors with historical data in METR-LA area

Transferred model performance in comparison with other models

Model	MAE	RMSE
STGCN	6.53	10.07
FC-LSTM	4.69	8.48
GMAN	4.05	7.57
DCRNN	3.3	6.91
50-sensor subset DCRNN	**4.74±0.02**	**9.96 ±0.03**
10-sensor subset DCRNN	**3.75±0.04**	**7.78±0.07**

Table 2: Transfered model performance on sensor sets in PEMS-BAY. The first 4 rows represent models trained on the full dataset [3]. The last 2 rows represent the average performance of the model tested on subsets of the dataset with the standard error

Example of transfered model predictions

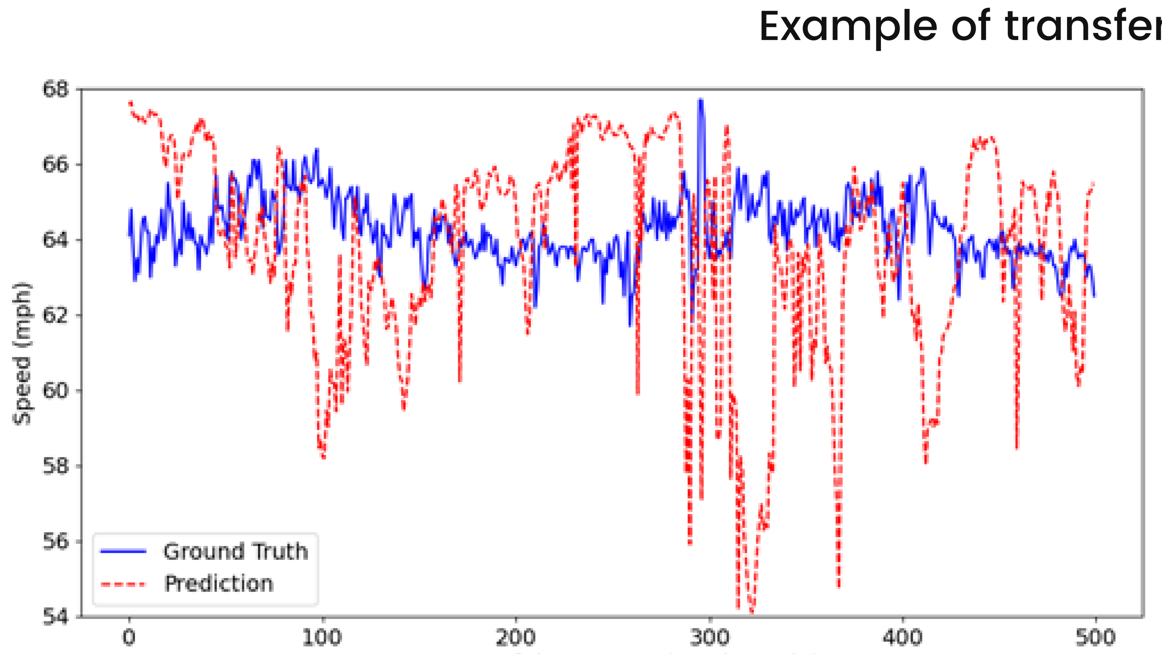

Figure 3: DCRNN predictions using 50-sensor model for one specific sensor, showing ground truth (blue) versus predictions (red)

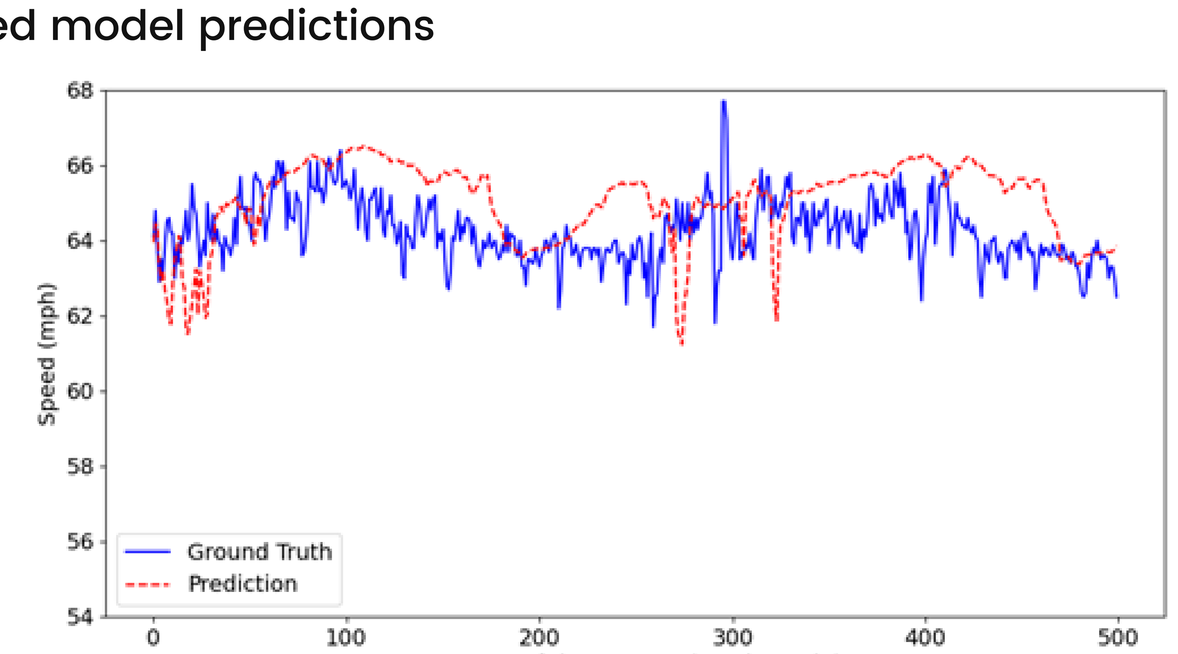

Figure 4: DCRNN predictions using 10-sensor model for one specific sensor, showing ground truth (blue) versus predictions (red)

Correlation between graph distance and performance:

- Correlation values are weak to moderate
- Model error is bigger for graphs with smaller FroD and AbsSum distance (negative correlation)
- Weak positive correlation can be observed for CosD

Metric	50-sensor model	10-sensor model
CosD (20000 mask)	0.23	0.11
CosD (40000 mask)	0.4	0.35
FroD (0 mask)	-0.55	-0.32
AbsSum (0 mask)	-0.37	-0.24

Table 5: Correlation between graph distances and performance of the 50-sensor and 10-sensor models for the transfer graphs selected using BSA

Example of best-observed correlation

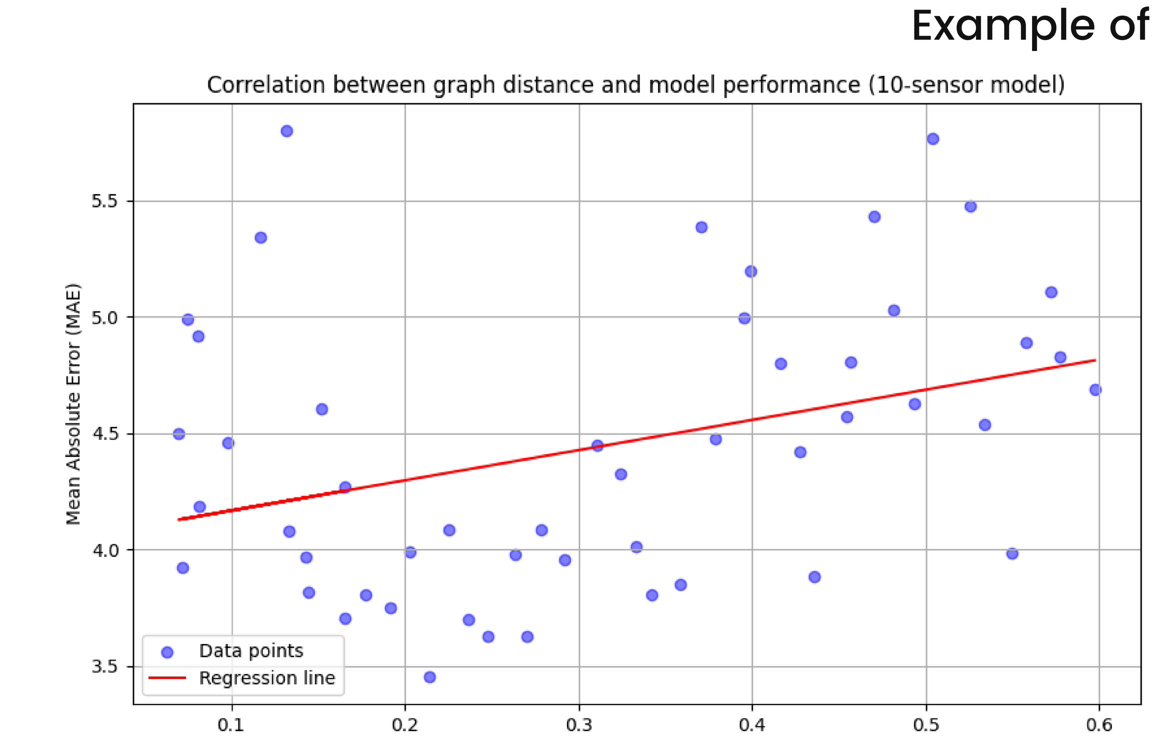

Figure 4: Correlation between graph distance(measured in CosD with mask 40000) and performance of the 10-sensor model

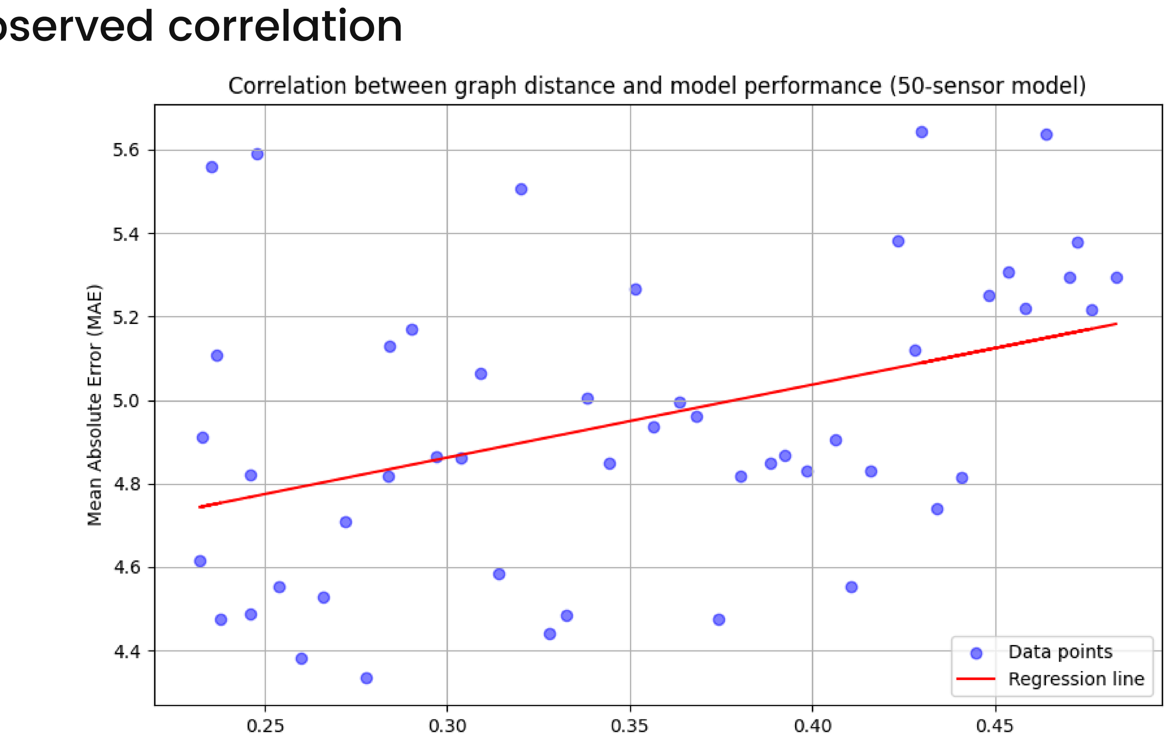

Figure 5: Correlation between graph distance(measured in CosD with mask 40000) and performance of the 50-sensor mode

References

[1] W. Jiang and J. Luo, 'Graph Neural Network for Traffic Forecasting: A Survey', Expert Systems with Applications, vol. 207, p. 117921, Nov. 2022, doi: 10.1016/j.eswa.2022.117921.
[2] T. Mallick, P. Balaprakash, E. Rask, and J. Macfarlane, 'Transfer Learning with Graph Neural Networks for Short-Term Highway Traffic Forecasting', in 2020 25th International Conference on Pattern Recognition (ICPR), Milan, Italy: IEEE, Jan. 2021, pp. 10367–10374. doi: 10.1109/ICPR48806.2021.9413270.
[3] Y. Li, R. Yu, C. Shahabi, and Y. Liu, 'Diffusion Convolutional Recurrent Neural Network: Data-Driven Traffic Forecasting'. arXiv, Feb. 22, 2018. doi: 10.48550/arXiv.1707.01926.