Spatiotemporal graph diffusion for network wide travel speed gap filling

Guillaume Neven[a,b,*], Nicolas Saunier[a,b], and Lijun Sun[b,c]

[a]*Civil, Geological and Mining Engineering Department Polytechnique Montréal, Montréal, Canada*
[b]*Interuniversitary Research Centre on Entreprise Networks, Logistics and Transportation, Montréal, Canada*
[c]*Civil Engineering Department, McGill University, Montréal, Canada*
[*]*Corresponding author. E-mail address: guillaume.neven@polymtl.ca.*

1 Introduction & Literature Review

Data are a key element in modern transportation systems, thanks to the development of data-driven methods for diagnosis, operation and forecasting. Yet, the data needed to understand the traffic behaviour on a city scale is often missing, due to privacy, lack of funding or technology. It is therefore crucial to have good estimation methods, to provide insightful information even where no or few data were recorded.

Researchers have tested and adapted traditional approaches in recent years for the gap filling task. Cui et al. (2020) proposed a Markov model, where the missing values, and after some time the future values, are inferred from the past values. With a similar approach, Ma et al. (2022) introduced a multi-attention based model, both temporal and spatial, to predict city-wide bus travel time estimation. Others tried to cope with missing data via clever filling procedures. For example, Zhao et al. (2023) developed a procedure called Decay Unrolling (DU), where missing road links are filled by looking at neighbouring data, both in terms of space and time. If neighbouring data are missing, the procedure is repeated.

Many other gap filling procedures have been proposed, among which some rely already on a generative model, even though most such generative models are applied to images rather than graphs. First, Tashiro et al. (2021) adapted a Denoising Diffusion Probabilistic Models (DDPM) to impute multivariate time series. The idea is no longer to generate images from scratch, but to fill the blanks. To capture the inference capacities, a small section of known data are put aside on every batch. The reconstruction loss is then computed on this hidden part only. Recently Liu et al. (2023) proposed a similar approach, taking in additional geographical features, namely the distance between detectors, to further improve model performance.

For the travel time prediction task, Rasul et al. (2021) and Tashiro et al. (2021) developed respectively TimeGrad and CSDI, both diffusion frameworks for time series prediction. Later, Lin et al. (2024) proposed a similar approach, with the difference being that they transformed the travel time series to the Fourier domain, did the prediction in the Fourier domain, and then transformed the data back to temporal data. Their deterministic performance was better than their probabilistic ones, yet no comparison to any deterministic baseline was provided.

Albeit developed for a different task, where all the missing time steps are the ones in the future, the methodologies are similar to the task at hand. Yet, generative models for prediction revealed to be a bit disappointing compared to traditional Deep Learning methods Yu et al. (2018); Zhang et al. (2020, 2019). However, the adaptation of those models for gap filling task is challenging, and the solution may lie in generative models.

In this paper, a new model, called DIFF-STGCN, is introduced and tested against the current state-of-the-art for gap filling of travel speed ratio for the City of London. This new model will aim not only to beat the current baseline, but also to leverage the scalability of graphs to allow a city wide prediction.

2 Methodology

The proposed DIFF-STGCN model is based on Denoising Diffusion Probabilistic Models (DDPMs) (Ho et al., 2020), which learn data distributions through iterative noising and denoising processes. This section briefly outlines the core principles of DDPMs and introduces the spatiotemporal graph adaptation used in our approach.

The forward diffusion process transforms an initial distribution of link travel times $q(\mathbf{Y}_0) \in \mathbb{R}^{T\times E}$ (where T is the number of time steps and E is the number of edges in graph $\mathscr{G}$) into a noisy distribution $q(\mathbf{Y}_s)$ after s steps. This is defined as:

$$q(\mathbf{Y}_s \mid \mathbf{Y}_{s-1}) = \mathscr{N}\left(\sqrt{1-\beta_s}\mathbf{Y}_{s-1}, \beta_s \mathbf{I}\right), \tag{1}$$

where β_s is the noise schedule. By iterating this process S time, $q(\mathbf{Y}_S \mid \mathbf{Y}_0)$ can be expressed as:

$$q(\mathbf{Y}_S \mid \mathbf{Y}_0) = \mathscr{N}\left(\sqrt{\bar{\alpha}_s}\mathbf{Y}_0, (1-\bar{\alpha}_s)\mathbf{I}\right), \tag{2}$$

with $\bar{\alpha}_s = \prod_{s=1}^{S}(1-\beta_s)$. The reverse (denoising) process $q(\mathbf{Y}_{s-1} \mid \mathbf{Y}_s)$ is approximated using a neural network θ:

$$p_\theta(\mathbf{Y}_{s-1} \mid \mathbf{Y}_s) = \mathscr{N}\left(\mu_\theta(\mathbf{Y}_s, s), \sigma_s^2 \mathbf{I}\right), \tag{3}$$

where μ_θ is the predicted mean by the neural network θ and σ_s^2 is the posterior variance. The model is trained to minimize the difference between the predicted and actual noise, focusing on unvisited edges using a mask Δ.

Unlike traditional DDPMs for images, our approach represents the transport network as a graph, where nodes correspond to road edges and features represent travel times over t consecutive steps. This graph-based structure allows the model to handle non-grid-like road networks and varying numbers of neighbors per node.

- **Training**: A random spatiotemporal graph $\mathbf{Y}_0$ is sampled, and noise equivalent to s steps is added to create $\mathbf{Y}_s$. The model ε_θ predicts the noise ε added to unvisited edges, and the loss is computed as:

$$\nabla_\theta \left\| \varepsilon^{\sim\Delta} - \varepsilon_\theta\left(\mathbf{Y}_s, s, \mathscr{G}, \mathbf{Y}_0^{\Delta}\right)^{\sim\Delta} \right\|^2. \tag{4}$$

- **Inference**: Starting from Gaussian noise $\mathbf{Y}_S$, the model iteratively denoises the graph by predicting and removing noise at each step s until $s = 0$, generating a plausible travel time distribution for missing edges.

By operating at the neighborhood level, the model avoids computationally expensive network-wide operations, enabling scalable predictions for large networks.

3 Preliminary Results

To test the newly introduced framework, data from the TRAFFIC4CAST (Neun et al., 2023) competition were used. The data, provided by HERE, come from floating car data for the City of London. The full network is composed of 197k links. At this time, the model has been tested on a reduced dataset (968 links in downtown London) which correspond to the limit of scalability of the CSDI baseline. Each spatiotemporal graph is composed of 32 successive 15-min time bins (8 h). Only the 4 middle 15-min bins are filled, the last 3.5 h and next 3.5 h are used as inputs.

The model and the baseline have been tested on a week of data. This week had on average 44.23% of missing values, of top of which 16.73% have been hidden for testing, for a total of 60.96% of missing values. The error is computed for the hidden values as the mean absolute error (MAE).

The results for the reduced dataset for a test day are shown in figure 1. The DIFF-STGCN demonstrates its abilities to fill the missing travel speed ratio. By iterating the sampling process, i.e. starting with different noise, one can create a distribution of predictions. Indeed, the derived confidence level is shown on figure 1. When a single time step is missing, the model is often more confident. Meaning, that among the 100 samples created, most predicted similar travel ratios. On the other hand, consecutive missing values often lead to more uncertain predictions. The predictions sometime are unstable, as shown be the highest quantiles. This should be fixed by tuning the hyperparameters, controlling the level of noise added. The CSDI baseline seems to be less responsive to slight variations in the speed ratio, better capturing the big picture rather than individual events.

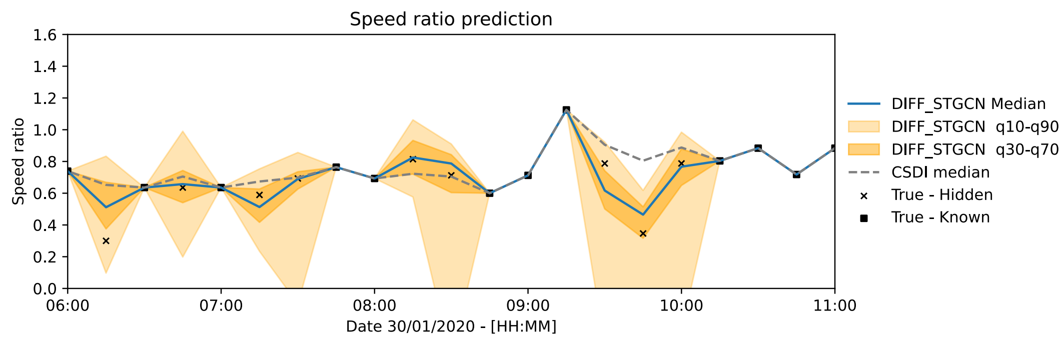

Figure 1: Missing values imputation for a test morning, with the proposed model median value in blue, and the 10th to 90th and 30th to 70th quantiles in orange. The CSDI baseline is in grey, and the ground truth in black. Crosses represent known data hidden to the models, while squares are known to the models.

To test the scalability of the model, the number of edges was increased to all the major links in the greater London area, using the open street map categories starting from *secondary_link* and up, amounting to 47'256 links. Finally, it was increased to the full network, i.e. 132'414 links.

Table 1 presents the CSDI baseline and the proposed model for the downtown area, as well as the performance for larger networks. Unfortunately, the model complexity of the CSDI baseline limits the number of edges: therefore, no baseline results can be provided for larger networks. Here lies the true improvement of the proposed model, leveraging graph structure to allow full network prediction. Not only is the proposed model better than the baseline on the small network, but its performance slightly degrades on the larger one, 47 times larger, but it is still better than the baseline. Finally, it comes as a surprise that the model inferring the full 132k links actually performs the best. It has to be noted that tuning is still in process and might improve results for all network sizes.

	CSDI	DIFF-STGCN			
Metric	Downtown	Downtown	Major	All	Downtown*
MAE	0.1975	0.1618	0.1801	**0.1057**	0.1498
RMSE	0.3266	0.2862	0.2949	**0.1883**	0.2155

Table 1: Imputation performance for different network sizes (* Downtown links predicted by the full network model)

Yet, it is interesting to compare the inference from the baseline and the one produce by the full network on the same edges. The last column of table 1 exhibits the performance for the same downtown links. It seems the full-size model benefits from the additional information and outperforms the baseline. Indeed, the major performance gap could have been explained by a multitude of low traffic link, easier to predict. Yet, figure 2 presents the score per road category and their representation. It seems the score is consistent across all category, and residential road, not present in the Major and Downtown subset, actually have the worst performance.

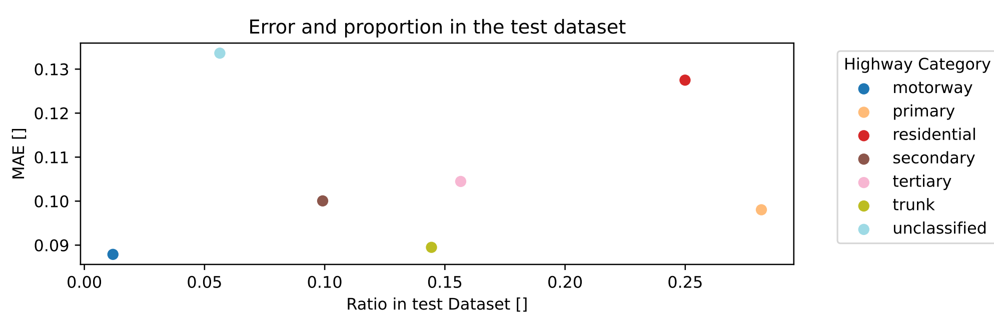

Figure 2: Error per category and their share of the network in the test dataset

The authors realize that the increase of loss present from downtown to major followed by a consequent drop in error is somewhat surprising. The same script was use for all different network size. A possibility lies in the lack of tuning, both for the baseline and DIFF-STGCN, which could increase performances for both. In the coming months, an extensive tuning will be performed to confirm our results. Either way, it will be further analyses.

4 Conclusion and Future work

This paper shows that leveraging the network graph can outperform the current state of the art, while allowing for a scalable model. As the model operate on a graph structure, it allows a much higher scalability. DIFF-STGCN seems to be able to perform a full network inference, allowing a better understanding and management of road networks. Also, both the baseline and the proposed model will be applied to another dataset in the City of Toronto for further tests.

Acknowledgement

The data used in this study were obtains during the TRAFFIC4CAST NeurIPS2022 competition (Neun et al., 2023). The authors also wish to acknowledge the support of NVIDIA who provided a RTX-8000 GPU used in this project. The project has been funded by MITACS in collaboration with Transcality AG and the City of Toronto.

References

Cui, Z., Lin, L., Pu, Z., and Wang, Y. (2020). Graph markov network for traffic forecasting with missing data. *Transportation Research Part C: Emerging Technologies*, 117:102671.

Ho, J., Jain, A., and Abbeel, P. (2020). Denoising diffusion probabilistic models. (arXiv:2006.11239). arXiv:2006.11239 [cs, stat].

Lin, L., Shi, D., Han, A., and Gao, J. (2024). Specstg: A fast spectral diffusion framework for probabilistic spatio-temporal traffic forecasting. (arXiv:2401.08119). arXiv:2401.08119 [cs].

Liu, M., Huang, H., Feng, H., Sun, L., Du, B., and Fu, Y. (2023). PriSTI: A Conditional Diffusion Framework for Spatiotemporal Imputation. arXiv:2302.09746 [cs].

Ma, J., Chan, J., Rajasegarar, S., and Leckie, C. (2022). Multi-attention graph neural networks for city-wide bus travel time estimation using limited data. *Expert Systems with Applications*, 202:117057.

Neun, M., Eichenberger, C., Martin, H., Spanring, M., Siripurapu, R., Springer, D., Deng, L., Wu, C., Lian, D., Zhou, M., Lumiste, M., Ilie, A., Wu, X., Lyu, C., Lu, Q.-L., Mahajan, V., Lu, Y., Li, J., Li, J., Gong, Y.-J., Grötschla, F., Mathys, J., Wei, Y., Haitao, H., Fang, H., Malm, K., Tang, F., Kopp, M., Kreil, D., and Hochreiter, S. (2023). Traffic4cast at neurips 2022 – predict dynamics along graph edges from sparse node data: Whole city traffic and eta from stationary vehicle detectors.

Rasul, K., Seward, C., Schuster, I., and Vollgraf, R. (2021). Autoregressive denoising diffusion models for multivariate probabilistic time series forecasting. (arXiv:2101.12072). arXiv:2101.12072 [cs].

Tashiro, Y., Song, J., Song, Y., and Ermon, S. (2021). CSDI: Conditional Score-based Diffusion Models for Probabilistic Time Series Imputation. arXiv:2107.03502 [cs, stat].

Yu, B., Yin, H., and Zhu, Z. (2018). Spatio-temporal graph convolutional networks: a deep learning framework for traffic forecasting. *Proceedings of the Twenty-Seventh International Joint Conference on Artificial Intelligence*.

Zhang, C., Yu, J. J., and Liu, Y. (2019). Spatial-temporal graph attention networks: a deep learning approach for traffic forecasting. *IEEE Access*, 7:166246–166256.

Zhang, J., Zheng, Y., Sun, J., and Qi, D. (2020). Flow Prediction in Spatio-Temporal Networks Based on Multitask Deep Learning. *IEEE Transactions on Knowledge and Data Engineering*, 32(3):468–478. Conference Name: IEEE Transactions on Knowledge and Data Engineering.

Zhao, S., Zhao, D., Liu, R., Xia, Z., Chen, B., and Chen, J. (2023). Gmat-du: Traffic anomaly prediction with fine spatiotemporal granularity in sparse data. *IEEE Transactions on Intelligent Transportation Systems*, 24(11):13503–13517.