ESTIMATION OF TRAFFIC VOLUMES IN UNMONITORED ROADS

Jaime Navajas Diez

Universidad Pontificia Comillas ICAI, Spain

Abstract

Urban traffic monitoring systems often rely on costly sensor networks, which cannot feasibly cover every street. This work proposes a graph-based approach to estimate vehicle volumes on unmonitored roads using Graph Neural Networks (GNNs). The street network of Barcelona is modeled as a graph where traffic volume estimation is framed as a node-level regression task. Several architectures are explored, including GraphSAGE, GIN, Correct & Smooth (C&S), and a custom GIN-based model with reconstruction tasks (DAGI). Baseline comparisons and extensive experiments demonstrate that GNNs can achieve reasonable accuracy even with limited monitored data. The results suggest a feasible reduction in sensor deployment without severely compromising traffic information quality, supporting more sustainable and cost-efficient mobility systems.

Keywords: Graph Neural Networks (GNN), Unmonitored Roads, Traffic, Urban Graphs, Prediction, Neural Networks.

Highlights

- A graph-based method is proposed to estimate vehicle volumes on unmonitored urban streets.
- The Barcelona Road network is modeled using both node- and edge-level graph representations.
- Several GNN architectures are evaluated, including GraphSAGE, GIN, and Correct & Smooth (C&S).
- A custom model (DAGI), combining GIN with Jumping Knowledge and reconstruction tasks, is introduced.
- Mean Absolute Error (MAE) is used to assess prediction performance across various test ratios.
- Baseline comparisons demonstrate that some GNN-based models outperform naive estimation methods.
- Results suggest a significant potential to reduce the number of physical sensors required in cities.
- This approach contributes to more sustainable, cost-effective, and scalable urban mobility monitoring systems.

1. Introduction

Monitoring urban traffic is essential for optimizing mobility policies, reducing emissions, and improving citizens' quality of life. However, installing sensors on every street is economically and technically unfeasible. This limitation has sparked interest in alternative methods to infer traffic conditions using partial data. Graph Neural Networks (GNNs) have emerged as a powerful tool to model complex urban networks and infer missing information based on spatial relationships and known observations.

This work explores the use of GNN-based models to estimate vehicle volumes on unmonitored streets in the city of Barcelona. The goal is to evaluate how accurately traffic can be predicted using graph-based learning approaches and how much physical sensor deployment could potentially be avoided.

1.1. Motivation

Urban traffic monitoring is a critical component of smart mobility systems, helping authorities manage congestion, reduce emissions, and make data-driven decisions. However, deploying physical sensors across every street is costly and impractical, especially in large cities. This constraint creates information gaps that limit the potential of traffic prediction systems.

Email address: 201908637@alu.comillas.edu (Jaime Navajas Diez)

Graph Neural Networks (GNNs) offer a promising alternative by enabling the estimation of traffic conditions in unmonitored streets using the topology of the road network and partial sensor data. The motivation behind this project is to explore whether GNN-based approaches can provide accurate predictions while reducing dependency on expensive infrastructure. From a broader perspective, this work aligns with the goals of promoting sustainable urban planning and cost-effective public service management.

1.2. Objectives

The primary objective of this project is to evaluate how well Graph Neural Networks can estimate traffic volumes on unmonitored streets by learning from partially observed traffic data and the topology of the road network. The study is conducted using the urban layout of Barcelona, a complex and realistic testbed for validating GNN performance in real-world conditions.

The work explores different graph formulations and GNN architectures to identify which combinations yield better estimations and under what conditions. In addition, the project aims to determine how the proportion of monitored versus unmonitored streets affects prediction accuracy. This contributes to the long-term goal of minimizing infrastructure requirements while still achieving meaningful insights into urban mobility patterns.

Main objectives:

- Develop a graph-based representation of the Barcelona Road network.
- Convert traffic data into graph formats suitable for node-level and edge-level regression.
- Design and test several GNN models, including GraphSAGE, GIN, DAGI, and Correct & Smooth.
- Implement baseline methods for comparison, including global and local mean strategies.
- Evaluate model performance using Mean Absolute Error (MAE) across various test splits.
- Investigate the impact of graph topology and monitor coverage on model accuracy.
- Assess the feasibility of reducing sensor deployment without compromising prediction quality.
- Contribute to the development of more sustainable and efficient traffic monitoring systems.

2. Related Work

Urban traffic prediction has traditionally been addressed through a combination of statistical models and sensor-based infrastructure. However, the sparsity and cost of real-world traffic data have led researchers to explore alternative methods capable of generalizing across networks with incomplete or missing information.

Early approaches in this space applied neural models to traffic forecasting using traditional feature-based learning methods [20][21], or sequence-based architectures such as Recurrent Neural Networks (RNNs) for trajectory prediction [22]. As data availability increased, so did the complexity of the models, moving toward graph-based representations to better capture the topological structure of urban networks.

Graph Neural Networks (GNNs) have emerged as a robust framework for learning over structured data. Their capacity to leverage neighborhood information makes them particularly well suited for tasks where spatial dependencies are critical. Foundational models in the field, such as GraphSAGE [19] and GIN [18], introduced scalable ways to generate node embeddings and proved the expressive power of message passing mechanisms. These models laid the groundwork for a growing body of research applying GNNs to transportation problems.

Specifically, GNNs have been adopted for imputing missing information in graphs. For instance, MRAP [2] proposes a multi-relational propagation scheme to complete node attributes, which is applicable to heterogeneous or partially labeled graphs. This is especially relevant in traffic networks, where many road segments lack sensors. Other strategies like Correct and Smooth (C&S) [11] decouple prediction and propagation by combining a simple base model with iterative graph-based error correction and label smoothing. C&S has shown strong performance in sparse-label and semi-supervised scenarios.

Another relevant technique is NoGE, introduced in [17], which reframes imputation as a message-passing task on missing node features. Its structure is particularly suited for sparse sensing environments and aligns with the goal of estimating unknown traffic volumes based on topological relationships.

In this project, we also introduce a custom model named DAGI (Deep Auxiliary GIN with Jumping Knowledge), which builds on the GIN architecture and incorporates Jumping Knowledge mechanisms to aggregate multi-hop information. In addition, DAGI integrates reconstruction tasks as auxiliary objectives, which aim to reinforce structural learning in low-label regimes. The model is specifically designed for edge-level prediction reformulated as node-level regression through extended graph formulations.

GNN-based imputation methods have also been explored in biomedical and scientific applications. Wang et al. [13] proposed a graph-based strategy for imputing brain measurements across datasets, illustrating the flexibility of these models across domains. Similarly, Bayram et al. [2] and Huang et al. [11] have developed graph frameworks focused on recovering missing or incomplete node features with high accuracy.

In the context of traffic networks, Liu et al. [12] implemented GraphSAGE for segment-level traffic speed forecasting with sparse data, highlighting its inductive capabilities. Their work addresses conditions similar to those faced in real urban settings, where only partial sensor coverage is available.

Earlier works have also explored interpolation techniques using spatial statistical models and spatio-temporal kriging [5], as well as deep learning methods applied to network-wide trajectory prediction [1] and hybrid optimization-learning schemes [8].

This aligns with a previous work [10] that also investigates the use of GNNs for static traffic volume estimation rather than time-series forecasting. In this setup, the goal is not to predict future flows, but to infer current traffic conditions on unmonitored streets based on available sensor data and the network topology.

Overall, the literature shows a shift from sensor-dependent, temporally driven models to graph-based learning and imputation approaches that can operate under limited data availability. This study contributes to this line of research by benchmarking existing models such as C&S, MRAP, NoGE and GraphSAGE, and by proposing DAGI, a new architecture tailored for semi-supervised traffic volume estimation on real-world urban networks.

3. Problem Description

This work addresses the challenge of estimating vehicle volume on streets that are not equipped with monitoring devices. In urban settings, only a limited portion of the road network is instrumented with traffic sensors due to the high cost of deployment and maintenance. As a result, traffic information is often incomplete, limiting the potential for comprehensive mobility analysis and planning.

To overcome this limitation, the road network is represented as a graph, where intersections are treated as nodes and streets as edges. The goal is to estimate traffic volume on streets without sensors by leveraging the information available on monitored streets, as well as the topological structure of the network.

Two graph representations are considered in this study:

- **Segment Graph:** In this formulation, each street segment is modeled as a node, and connections between adjacent streets are represented as edges. This structure enables the use of standard node-based GNN models but requires reinterpreting the graph so that each node represents a street rather than an intersection.

- **Extended Graph:** This approach creates artificial nodes to represent each directed street segment, embedding the traffic volume and other features into them. These artificial nodes are connected to the corresponding intersection nodes, enabling edge-level information to be processed as node attributes. This transformation makes the task compatible with standard node regression architectures in GNNs.

The traffic estimation problem is framed as a semi-supervised regression task, where only a portion of the graph has known traffic volumes, and the remaining values must be predicted. This setup allows the use of Graph Neural Networks to infer missing data by exploiting both the observed values and the structural relationships within the urban network.

By comparing different formulations and neural architectures, this study aims to determine how accurately traffic can be estimated in unmonitored streets, and under what conditions sensor deployment could be optimized or reduced.

4. Methodology

The methodology of this study is based on formulating the traffic volume estimation task as a semi-supervised node regression problem over urban graphs. The main goal is to infer vehicle volumes on unmonitored streets by leveraging the structure of the city's road network and partial volume measurements from sensor-equipped streets.

To this end, several Graph Neural Network (GNN) models are implemented and evaluated using two types of graph representations: segment graphs and extended graphs, each offering a different abstraction of the urban environment. These formulations allow the transformation of edge-level predictions into node-level tasks, which are compatible with most GNN architectures.

The study includes both classic and novel GNN-based approaches, as well as baseline methods for comparison:

- **GraphSAGE:** An inductive framework that learns node embeddings by aggregating features from neighboring nodes. This model is suited for generalizing to unseen graph components and has been widely used in traffic-related tasks.

- **GIN (Graph Isomorphism Network):** A strong baseline in graph learning, known for its expressive power. It processes node features through multiple layers and is used as a core component in other models within this study.

- **Correct and Smooth (C&S):** A hybrid method that applies a base prediction using a simple regressor (e.g., MLP or GAT), followed by correction and smoothing steps based on the graph structure. This decoupled strategy enhances performance in sparse-label settings.

- **NoGE:** A model specifically designed for missing feature imputations in graphs. It treats unobserved data as latent variables and propagates information through message-passing layers to recover missing values.

- **DAGI (Deep Auxiliary GIN with Jumping Knowledge):** A custom architecture developed, based on GIN. It incorporates Jumping Knowledge to merge information from multiple layers and includes graph reconstruction tasks as auxiliary learning objectives. These tasks aim to improve the model's capacity to learn from limited labeled data by enforcing structural consistency.

Each model is trained using only partial information from the graph, simulating realistic scenarios where traffic data is available for a limited set of streets. The models are evaluated based on their ability to generalize and accurately estimate traffic volume on the unmonitored parts of the network.

In addition to GNNs, two baseline strategies are implemented for comparison:

- **Global Mean:** Predicts the unknown traffic volume using the average of all known values in the network.

- **Local Mean:** Uses the average traffic volume of neighboring nodes (or edges) as the prediction for unmonitored nodes (or edges).

To quantitatively assess model performance, the **Mean Absolute Error (MAE)** is used as the primary evaluation metric, enabling a consistent comparison across models and test conditions.

5. Tools & Libraries

The implementation of this project relies on a modular architecture that combines graph preprocessing, model definition, training, and evaluation in a reproducible pipeline. The workflow was designed to facilitate experimentation with various Graph Neural Network (GNN) models, while allowing flexibility in data handling and graph construction.

The development was carried out entirely in Python, using a combination of specialized libraries for graph learning and machine learning:

- **PyTorch:** Used as the core deep learning framework to define and train all models.

- **PyTorch Geometric (PyG):** A high-level extension for PyTorch that provides modules and utilities for building GNNs. PyG was used to implement standard architectures like GIN and GraphSAGE, as well as the custom DAGI model.

- **Scikit-learn:** Utilized baseline regressors (e.g., linear models, tree-based regressors) and evaluation metrics.

- **NetworkX:** Used for graph construction, manipulation, and basic visualization.

- **Pandas and NumPy:** Employed for data manipulation, loading, and preprocessing.

Hyperparameter optimization was performed using manual search.

All experiments were executed in local environments, using GPUs or CPU. The project was version-controlled using Git and documented via Jupyter notebooks.

6. Experimental Setup

To evaluate the performance of different GNN models in the task of traffic volume estimation, a series of controlled experiments were designed. The evaluation focused on measuring the predictive accuracy of each model under varying levels of data availability, simulating real-world scenarios where only a portion of the street network is monitored.

Graph Data

All experiments were performed on the Barcelona Road network, formed in two ways:

- **Segment Graph:** Each node represents a street segment, and edges represent connectivity between adjacent segments.

- **Extended Graph:** Artificial nodes were created to represent directed street segments, allowing traffic volume (originally an edge attribute) to be predicted via node regression.

Each graph contains real traffic volume data assigned to a subset of streets, with remaining volumes treated as unknown.

Label Masking Strategy

To simulate partial monitoring, only a percentage of the streets were assumed to be observed in each experiment. Test sets were generated by randomly selecting a proportion of the labeled nodes (or edges), while ensuring that no pair of bidirectional edges was simultaneously hidden, to avoid underdetermined configurations.

The percentage of hidden labels varied between 2% and 30%, with ten repetitions per split to account for variance due to random selection. This strategy enabled the analysis of how model performance degrades as fewer monitored streets are available.

Baselines

Two simple baseline estimators were used:

- **Global Mean:** Assigns the average of all known volumes as the prediction for unknown segments.

- **Local Mean:** Assigns the average of the neighboring known volumes.

These baselines serve as reference points for comparing the added value of graph-based learning.

Model Configuration

All GNN models were trained using supervised node regression. Each model used the known traffic volumes as supervision and was validated on the masked test set. Hyperparameters such as learning rate, hidden dimensions, dropout, number of layers, and aggregation type were manually tuned per model.

Models were trained for a fixed number of epochs (typically between 150 and 1000, depending on convergence speed), and evaluated on the test set using the Mean Absolute Error (MAE) as the main metric.

7. Results *(See Annex for detailed metrics)*

7.1. Individual Results

GIN

Two versions of GIN were evaluated: one trained for 200 epochs and another extended to 600 epochs. The longer training consistently improved results across all test percentages, but neither version was able to surpass the performance of Correct and Smooth (C&S) or DAGI. The average MAE remained in the range of 1004–1117, depending on test split and training duration. Although GIN provided stable performance and competitive results in mid-range splits (e.g., 10%–20%), its expressiveness was not sufficient to compensate for the lack of supervision in more sparse scenarios.

GraphSAGE

GraphSAGE delivered stable yet moderate results, with MAE values ranging between 1162 and 1267. Its performance remained relatively constant regardless of the proportion of hidden labels, suggesting robustness but limited adaptability. Despite its inductive capabilities, GraphSAGE was consistently outperformed by Correct and Smooth variants.

NoGE

NoGE achieved improved accuracy compared to standard GNNs in most test cases, with MAEs ranging from 1088 to 1194. While it never achieved the best result in any single split, it showed consistent improvements over GraphSAGE and performed on par with the GIN models. Its main limitation was the lack of significant gains as label sparsity increased, possibly due to its reliance on initial feature completeness.

DAGI

DAGI obtained consistently better performance than GIN, GraphSAGE, and NoGE, with MAEs ranging from 1100 to 1212, depending on the test percentage. In any case, it was not the best model overall, as it was surpassed in all test splits by the Correct and Smooth variants.

Correct and Smooth (C&S)

C&S was the best-performing family of models across all test scenarios. When applied to MLP, MLP Linear, and GAT base predictors, it consistently reduced the MAE by 50% or more in comparison to the raw predictions.

C&S + MLP reached the lowest MAE overall: 462 at 4% test, and remained under 500 in most splits.

C&S + MLP Linear also performed quite well, with MAEs as low as 833–914 depending on the test.

C&S + GAT, both with 25 and 5 epochs, showed robust results post-correction, although the raw (pre-C&S) GAT models were among the worst performers prior to smoothing.

The effectiveness of C&S is especially notable in scenarios, where traditional GNNs tend to struggle. Its correction and smoothing phases demonstrated strong regularization and error propagation across the graph structure, validating previous findings in the literature [11].

7.2. Global Comparison

The overall results of the study show that model simplicity combined with graph-aware correction mechanisms outperforms more sophisticated architectures in the task of estimating traffic volumes on unmonitored streets.

The Correct and Smooth (C&S) strategy was by far the most effective across all test splits. Regardless of the base regressor used (MLP, Linear MLP, or GAT), the post-processing stages of correction and smoothing significantly reduced the error, often halving the MAE compared to the raw predictions. In particular, C&S with MLP achieved the best performance globally, reaching MAE values as low as 462 at 18% test, and remaining consistently below 500 up to 30% of missing data.

Surprisingly, simpler models combined with correction (like C&S + MLP) were able to outperform deeper architectures such as GIN, NoGE, and DAGI. These more sophisticated GNNs despite incorporating mechanisms like message passing, jumping knowledge, or auxiliary reconstruction failed to deliver competitive performance in this specific setting. In fact, DAGI did not surpass even the simplest baselines, suggesting that model complexity may not be beneficial when input features are limited.

A key insight from the experiments is that the lack of informative node and edge attributes in the graph (beyond topology and traffic volume) limits the ability of complex models to generalize effectively. In such low-feature scenarios, approaches like C&S, which focus on refining basic predictions through the graph's structure, prove more effective than deep architectures that rely on rich feature sets.

Finally, baseline methods based on global and local means, although clearly outperformed by C&S, performed better than expected, especially when compared to models like DAGI and NoGE. This reinforces the idea that simple aggregation methods can still be useful when feature quality is poor.

In summary, the experiments demonstrate that the most successful strategies are not necessarily the most complex, but those that best exploit the available structure, even with minimal supervision.

8. Discussion & Analysis

The results obtained in this study offer several important insights regarding the practical application of Graph Neural Networks (GNNs) for estimating traffic volumes on unmonitored streets in urban settings.

First, the performance gap between different model families reveals that structural refinement strategies like Correct and Smooth (C&S) are better suited to scenarios with limited sensor coverage and sparse graph attributes. Despite its simplicity, C&S consistently outperformed more complex models by focusing on propagation of label information rather than learning deep feature representations.

This outcome challenges the common assumption that more sophisticated architectures (e.g., GIN, DAGI, NoGE) will necessarily outperform simpler alternatives. While such models are theoretically powerful and designed for generalization, their advantage diminishes in practical conditions where node and edge features are minimal, as is often the case in urban traffic networks. Without access to rich metadata (e.g., speed limits, number of lanes, neighborhood type), the expressive capacity of deep models cannot be fully exploited.

Moreover, the DAGI model, which incorporates

Jumping Knowledge and auxiliary reconstruction tasks, failed to improve upon even the most basic baselines. This highlights a key limitation: when structural and feature information is poor, model complexity may introduce more noise than benefit, particularly if the task is formulated as node regression based on limited supervision.

Another significant observation is the robustness of the C&S framework, especially when applied to weak base predictors such as shallow MLPs or undertrained GATs. The ability of C&S to improve performance even from very poor initial predictions (as seen in the GAT variants) underlines its potential for deployment in resource-constrained or data-scarce environments.

The sensitivity to test size also provides useful guidance. While all models degraded as the proportion of unknown volumes increased, C&S variants showed graceful performance deterioration, whereas standard GNNs and DAGI exhibited steeper drops or plateaued early.

In terms of hypothesis validation, the study confirms that leveraging the graph topology is essential for inferring traffic in unmonitored areas. However, it also shows that not all GNN-based strategies benefit equally from the topology alone. Methods that explicitly enforce local consistency (as in C&S) are more effective in sparse-label regimes than those relying solely on learned embeddings.

Finally, the fact that simple baselines like Local Mean outperformed complex GNNs like DAGI and NoGE in several test splits indicate that a clear understanding of data limitations is critical when selecting models for real-world applications. Investing in model sophistication without addressing data sparsity or enriching the feature space may lead to counterproductive results.

9. Future Work

The results obtained throughout this study point to several promising research directions that could enhance both the accuracy and applicability of traffic volume estimation using graph-based models.

A key extension involves incorporating temporal dynamics into the estimation process. While this project focused on static daily volumes, traffic patterns are time-dependent. Future work should aim to predict hourly volumes by incorporating temporal sequences into the model, potentially using spatio-temporal GNNs or recurrent architectures adapted to graph inputs.

Additionally, the models used here relied almost exclusively on graph topology and a single numerical feature (weight) and the label (traffic volume). This limitation affected the performance of more sophisticated architecture, as shown in the results. A clear avenue for improvement lies in enriching the input data with additional node and edge features, such as the number of lanes, average speeds, land use, or known congestion levels. These attributes would allow expressive models like DAGI or NoGE to reach their full potential.

Another future direction is to apply these techniques to new real-world urban networks, beyond the Barcelona graph used here. Testing on networks of different size, structures, and sensor distribution such as Madrid or global cities would provide a clearer picture of the generalizability of the results obtained. This could also include cities with less regular topologies or networks designed for public transport.

A particularly relevant research line is the optimal sensor placement problem. Rather than predicting volumes from a fixed set of sensors, future work could develop algorithms that identify the most valuable subset of streets to monitor, minimizing the number of sensors required while keeping the estimation error within acceptable bounds. This optimization could be formulated as a budget-constrained sensor allocation problem.

Further work could also explore the estimates of emissions maps from predicted volumes. Integrating traffic forecasts with environmental models would provide urban planners with valuable tools for sustainability planning and air quality assessment.

Finally, a more theoretical direction involves understanding which topological structures or graph formulations (e.g., segment-based vs. node-based) better support the task of traffic volume estimation. This includes comparing different formulations across the same city or exploring how to design synthetic graphs that maximize estimation accuracy under constraints of sparsity and noise.

In summary, while this project demonstrates that GNN-based imputation is a feasible and efficient method for estimating traffic volumes in real urban settings, its practical deployment still depends on further advances in temporal modeling, feature engineering, urban generalization, and strategic data acquisition.

10. Conclusion

This project sets out to explore the feasibility of estimating traffic volumes on unmonitored streets using graph-based learning approaches, with a focus on Graph Neural Networks (GNNs) and their capacity to generalize from partial observations in real urban environments.

The work addressed a relevant and practical problem: the limited spatial coverage of physical sensors in traffic monitoring systems. By representing the road network as a graph and formulating the estimation task as a semi-supervised regression problem, it was possible to test various architectures on the real case of the city of Barcelona.

The experimental results confirm that leveraging the topology of the urban network significantly improves the accuracy of traffic volume estimation compared to baseline methods. However, they also show that model complexity does not necessarily translate into better performance. While advanced GNN models like GIN, NoGE, and DAGI were designed to exploit structural patterns, they failed to outperform simpler strategies under sparse labeling and low-feature conditions.

In contrast, Correct and Smooth (C&S) proved to be the most effective approach across all scenarios. Its decoupled architecture combining a simple base regressor with graph-aware error propagation consistently achieved the lowest Mean (MAE), especially in highly sparse setups. These findings validate the hypothesis that graph structure alone can compensate for the absence of dense sensor data, provided that the model architecture is adapted to propagate information efficiently.

Nevertheless, the study also revealed critical limitations. The lack of rich node and edge attributes hindered the performance of expressive architectures and highlighted the need for better feature engineering. Furthermore, the results suggest that performance plateaus beyond a certain threshold of missing data, indicating that not all streets are equally predictable through structural inference alone.

In conclusion, this work demonstrates that semi-supervised graph learning is a viable and scalable solution for estimating traffic volumes in partially monitored cities. It provides a strong foundation for future research in the direction of temporal modeling, sensor optimization, and sustainable urban planning. More broadly, it underscores the importance of aligning model complexity with data availability, a principle applicable across many domains of deep learning.

11. Sustainability and SDG Alignment

The work carried out in this project contributes directly to several of the United Nations' Sustainable Development Goals (SDGs), by addressing the need for more intelligent, efficient, and environmentally conscious mobility systems.

SDG 3 – Good Health and Well-Being

By enabling the estimation of traffic volumes without requiring full sensor coverage, the proposed methodology can support the identification of high-congestion areas and traffic-related health risks. This facilitates better-informed public policies aimed at reducing exposure to air pollution and noise — two critical factors for urban health.

SDG 11 – Sustainable Cities and Communities

The development of scalable, data-efficient methods for urban traffic estimation contributes to the broader goal of building inclusive, safe, and sustainable cities. This project provides a framework that can help cities monitor mobility more cost-effectively, even in underserved or low-income areas with limited infrastructure.

SDG 13 – Climate Action

By reducing the need for widespread physical sensor deployment, and enabling smarter mobility planning, this approach supports the design of strategies aimed at lowering emissions and optimizing traffic flow. Accurate traffic estimation is a key input for carbon footprint modeling and environmental impact assessments.

In summary, the results of this work are not only relevant from a technical perspective but also offer a pathway toward more sustainable and equitable urban mobility, aligning with global efforts to mitigate climate change and improve quality of life in cities.

Acknowledgements

Supervisor: Luis Francisco Sánchez Merchante.
Entity: ICAI – Universidad Pontificia Comillas.

References

[1] *«Network-Wide Vehicle Trajectory Prediction in Urban Traffic Networks using Deep Learning», ResearchGate, doi: 10.1177/0361198118794735*

[2] *E. Bayram, A. Garcia-Duran, y R. West, «Node Attribute Completion in Knowledge Graphs with Multi-Relational Propagation», 10 de noviembre de 2020, arXiv: arXiv:2011.05301. doi: 10.48550/arXiv.2011.05301*

[3] *W. Feng et al., «Graph Random Neural Network for Semi-Supervised Learning on Graphs», 21 de septiembre de 2021, arXiv: arXiv:2005.11079. doi: 10.48550/arXiv.2005.11079*

[4] *S. Shirakawa, S. Ogino, y T. Nagao, «Graph structured program evolution», en* Proceedings of the 9th annual conference on Genetic and evolutionary computation, *GECCO '07. New York, NY, USA: Association for Computing Machinery, jul. 2007, pp. 1686–1693. doi: 10.1145/1276958.1277290*

[5] *«Forecasting Network Data Spatial Interpolation of Traffic Counts from Texas Data», ResearchGate, doi: 10.3141/2105-13*

[6] *R. K. C. Chan, J. M.-Y. Lim, y R. Parthiban, «A neural network approach for traffic prediction and routing with missing data imputation for intelligent transportation system»,* Expert Syst. Appl.*, vol. 171, p. 114573, jun. 2021, doi: 10.1016/j.eswa.2021.114573*

[7] *«Traffic Estimation for Proactive Freeway Traffic Control». Disponible en: https://journals.sagepub.com/doi/epdf/10.3141/1679-11*

[8] *Y. Djenouri, A. Belhadi, G. Srivastava, y J. C.-W. Lin, «Hybrid graph convolution neural network and branch-and-bound optimization for traffic flow forecasting»,* Future Gener. Comput. Syst.*, vol. 139, pp. 100–108, feb. 2023, doi: 10.1016/j.future.2022.09.018*

[9] *D. Zhou, O. Bousquet, T. Lal, J. Weston, y B. Schölkopf, «Learning with Local and Global Consistency», en* Advances in Neural Information Processing Systems*, MIT Press, 2003. Disponible: https://proceedings.neurips.cc/paper_files/paper/2003/hash/87682805257e619d49b8e0dfdc14affa-Abstract.html*

[10] *A. Moreno Grande, «Estimación de tráfico en vías no monitorizadas mediante el uso de grafos y redes neuronales». 2 de junio de 2023*

[11] *Q. Huang, H. He, A. Singh, S.-N. Lim, y A. R. Benson, «Combining Label Propagation and Simple Models Out-performs Graph Neural Networks», 2 de noviembre de 2020, arXiv: arXiv:2010.13993. doi: 10.48550/arXiv.2010.13993*

[12] *J. Liu, G. P. Ong, y X. Chen, «GraphSAGE-Based Traffic Speed Forecasting for Segment Network With Sparse Data»,* IEEE Trans. Intell. Transp. Syst.*, vol. 23, n.º 3, pp. 1755–1766, mar. 2022, doi: 10.1109/TITS.2020.3026025*

[13] *Y. Wang, W. Peng, S. F. Tapert, Q. Zhao, y K. M. Pohl, «Imputing Brain Measurements Across Data Sets via Graph Neural Networks», 19 de agosto de 2023, arXiv: arXiv:2308.09907. doi: 10.48550/arXiv.2308.09907*

[14] *bstabler, Repositorio Github bstabler. [En línea]. Disponible en: https://github.com/bstabler/TransportationNetworks*

[15] *L. F. Sánchez Merchante, Repositorio Github lmerchante. [En línea]. Disponible en: https://github.com/lmerchante*

[16] *TensorFlow GNN. [En línea]. Disponible en: https://www.tensorflow.org/gnn*

[17] *«Papers with Code - Node Co-occurrence based Graph Neural Networks for Knowledge Graph Link Prediction». Disponible en: https://paperswithcode.com/paper/nodeco-occurrence-based-graph-neural7*

[18] *K. Xu, W. Hu, J. Leskovec, and S. Jegelka, "How Powerful are Graph Neural Networks?" en* Proc. Int. Conf. Learn. Representations (ICLR)*, 2019*

[19] *W. Hamilton, Z. Ying, and J. Leskovec, "Inductive Representation Learning on Large Graphs," en* Proc. Adv. Neural Inf. Process. Syst. (NeurIPS)*, 2017, pp. 1024–1034*

[20] *G.-L. Chang, et al., "Traffic forecasting using neural networks,"* Transportation Research Record*, no. 1678, pp. 93–104, 1999*

[21] *S. Ishak, S. Kothuri, and C. Alecsandru, "Performance evaluation of short-term traffic prediction models,"* Journal of Transportation Engineering*, vol. 129, no. 6, pp. 664–672, 2003*

[22] *J. Choi and D. Kim, "Deep Learning-Based Vehicle Trajectory Prediction with Recurrent Neural Networks,"* Sensors*, vol. 18, no. 12, pp. 1–15, 201*

Annex (Results Metrics)

Table 1: Summary table of Results

Test	GINN (200 epochs) Mean MAE	GINN (600 epochs) Mean MAE	GraphSage Mean MAE	NoGE Mean MAE	DAGI Mean MAE	C&S MLP Mean MAE		C&S MLP Linear Mean MAE		C&S GAT 25 epochs Mean MAE		C&S GAT 5 epcohs Mean MAE	
						After C&S	Before C&S	After C&S	Before C&S	After C&S	Before C&S	After C&S	Before C&S
2%	1117	1089	1267	1194	1212	483	1217	912	1259	1243	3166	797	1137
4%	1072	1023	1175	1122	1194	467	1194	833	1131	1184	3416	798	1134
6%	1080	1033	1205	1136	1113	494	1234	885	1186	1211	3265	846	1242
8%	1077	1024	1203	1133	1134	485	1243	914	1228	1337	4170	818	1206
10%	1049	1007	1175	1110	1117	483	1255	882	1181	1326	4091	830	1213
12%	1040	999	1173	1105	1101	483	1255	887	1194	1262	3614	829	1192
14%	1052	1006	1165	1099	1124	484	1216	890	1188	1145	3137	817	1166
16%	1052	1017	1164	1095	1100	487	1232	896	1198	1278	3845	813	1181
18%	1041	1019	1163	1092	1094	462	1192	878	1193	1171	2846	832	1180
20%	1037	1017	1162	1088	1120	478	1203	890	1181	1329	4032	818	1176
22%	1065	1004	1171	1093	1137	474	1234	896	1192	1229	3397	814	1184
24%	1051	1016	1169	1094	1106	485	1226	884	1176	1330	4188	823	1184
26%	1051	1014	1170	1089	1112	485	1248	911	1230	1312	3932	834	1199
28%	1054	1004	1170	1093	1123	467	1211	889	1186	1167	2888	820	1188
30%	1049	1029	1178	1096	1144	474	1205	889	1189	1241	3576	819	1174

Chart 1: Summary Chart of Results