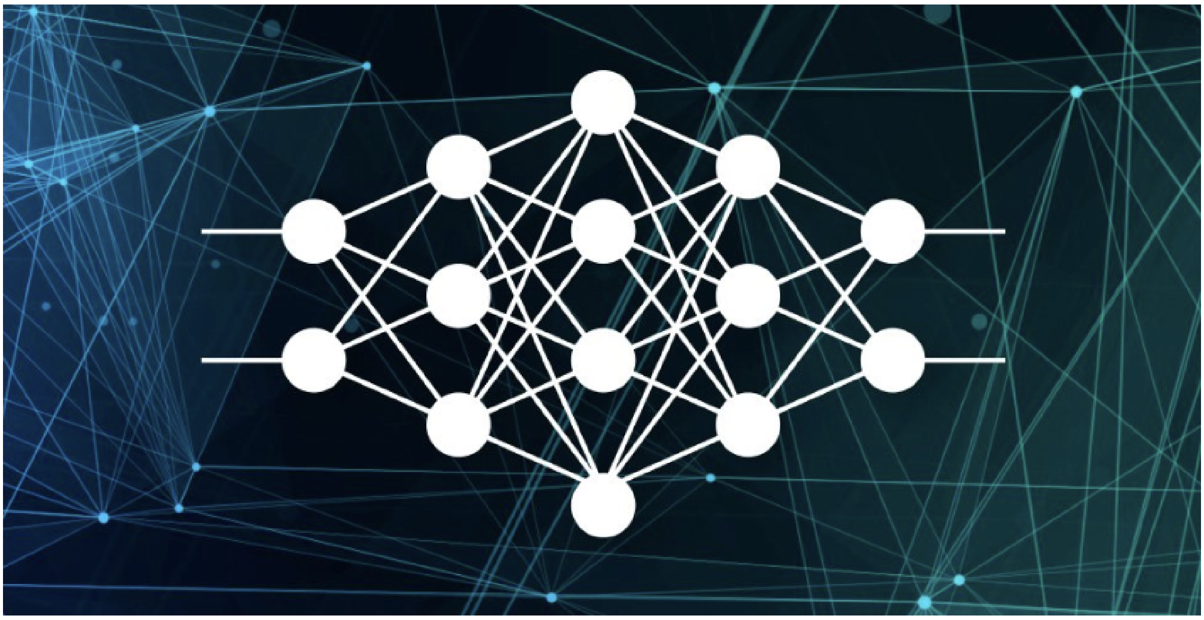
Enhancing Freight Transportation Predictions in the Netherlands Using Spatial-Temporal Graph Neural Networks

Rafal Tarczynski
student #4439899

1st supervisor: Martijn Meijers
2nd supervisor: Azarakhsh Rafiee
3rd supervisor: Amin Jalilzadeh
TNO supervisor: Finn Winkelmann

15 January, 2024



Contents

1. Introduction

Freight transportation plays a pivotal role in sustaining economic activities, with road freight remaining the dominant mode within the Netherlands despite evolving logistics trends. According to the TNO report *Decamod: Toolbox voor rekenen aan CO_2-reductie in transport en logistiek* (2020) TNO (2020), road transport is projected to carry a slightly reduced share of total freight—declining from 75% in 2014 to 73% by 2030. However, this modest decrease does not diminish its environmental impact. Road freight is expected to contribute approximately 81% of the total CO_2 emissions from freight transport by 2030, a disproportionately high figure considering its share of ton-kilometres. This discrepancy arises from the lower energy efficiency of road vehicles compared to rail and inland shipping, coupled with the prevalence of short-haul, high-frequency trips, particularly within urban areas. More details of vehicle, freight type classification and logistic segments can be found in appendices 6, 7 and 8.

Furthermore, 73% of road freight is dedicated to domestic transportation, where modal shifts to rail or waterways are often impractical. This sector's rigidity amplifies the importance of exploring optimisation strategies. While advancements in alternative transport modes are essential, the greatest potential for immediate CO_2 reduction lies within the road freight sector. Measures such as route efficiency could yield significant environmental benefits. Even incremental improvements can result in substantial emission reductions due to the scale of road freight operations.TNO (2020)

Many complex relationships can be effectively represented through knowledge graphs, which serve as powerful structures for modifying, enhancing, and generating new graphs (Martin and Reichmann, 2024). Their versatility has been increasingly leveraged in solving real-world problems, particularly within scientific and industrial domains. For instance, NVIDIA has harnessed the potential of Graph Neural Networks (GNNs) to optimise physical structures for additive manufacturing, leading to significant advancements in lattice structure simulation and predictive modelling Jain et al. (2024). This approach has demonstrated how graph-based models can streamline design processes, improve material efficiency, and reduce computational costs in complex engineering tasks.

Similarly, researchers from institutions including Google DeepMind developed the Graph Networks for Materials Exploration (GNoME) framework, which utilises GNNs to evaluate material stability based on structural and compositional properties. By scaling the training of these networks, the GNoME framework has achieved remarkable generalisation capabilities, enabling the discovery of over 2.2 million stable crystal structures and significantly enhancing the efficiency of materials discovery (Merchant et al., 2023). These examples underscore the transformative potential of graph-based models in diverse applications, from optimising manufacturing processes to accelerating scientific discoveries.

Although not directly connected to my graduation topic, those novelties present the potential of Graph Neural Networks (GNNs). My graduation research explores the use GNNs with an additional temporal dimension, in combination with the TNO's Digital Twin platform to visualise the results and enhance our understanding of the relationship between city morphology and transportation networks. This study aims to model the impact of certain elements of urban tissue morphology on road freight transportation flow patterns.

The study will also involve training machine learning models on existing data from the Rotterdam-The Hague Metropolitan Area and Amsterdam, focusing on how these cities' unique morphological structures influence traffic and freight flow. These models will then be applied to new urban areas to predict transportation flows based on their own distinct morphological characteristics. This approach aims to reveal the influence of urban structure on transportation dynamics and provide predictive insights for city planners when evaluating transportation flow under varying morphological scenarios.

Within the context of the MSc Geomatics program at TU Delft, the study embraces skills in spatial data analysis, geospatial datasets, and the application of machine learning techniques in geographical contexts directly corresponds with several core courses in the program. Notably, the course *Machine Learning for the Built Environment* (GEO5017) explores the foundations of machine learning methodologies, *Python Programming for Geomatics* (GEO1005) develops proficiency in programming skills essential for geospatial data processing and analysis. Additionally, *Sensing Technologies* (GEO1001) provides foundational knowledge in data acquisition methods, which is crucial for understanding and implementing various sensing techniques in geospatial research. Geo Database Management Systems (GEO1006) provided insights in data storing and handling, which will be crucial for managing big datasets for this project.

These courses collectively equip students with the skills necessary to manage, analyse, and interpret complex spatial data, which are essential for the successful execution of this research.

1.1. Hypothesis

Freight transportation models predominantly rely on road network structures and traffic data to predict flows. However, urban morphology and socio-economic factors play a fundamental role in shaping transportation patterns. This research hypothesises that integrating building attributes and socio-economic data into Graph Neural Network (GNN) training will enhance the predictive accuracy of freight transportation flow models.

Rationale Buildings influence freight movement through their function, density, and spatial distribution. For example, commercial zones experience higher freight activity than residential areas, and large industrial facilities often generate specific freight demand patterns. In addition, residential and low-density areas, which generally experience lower economic activity, are often less attractive for road infrastructure planning and logistics optimisation. As a result, these areas may have lower freight transport intensity, leading to spatial disparities in freight accessibility. Incorporating such data into GNN training may help reveal hidden patterns in freight distribution, particularly in areas where infrastructure development and transport accessibility are constrained.

Traditional transportation forecasting models often struggle in data-scarce environments, requiring extensive calibration and domain-specific knowledge. By incorporating additional urban features, this study aims to bridge these gaps and provide a more adaptable, data-driven approach. Graph-based models, particularly Temporal Graph Neural Networks (TGNNs), are well-suited to capturing dynamic relationships between spatial entities, making them a promising choice for this task.

2. Related work

Transportation is essential to urban economic health, yet forecasting transportation flows remains a challenge, especially in regions with limited data. TNO's Digital Twin and MASS-GT framework integrates datasets such as traffic, and environmental data, offering a comprehensive digital twin to simulate the effects of planning decisions on urban infrastructure. However, it lacks predictive capabilities that relate to city elements such as buildings, and requires much data and calibration to be applied in different contexts.

By utilising GNNs' ability to model networked relationships, the proposed approach aims to improve long-term transportation forecasting and support decision-making processes even in data-scarce environments.

The application of Graph Neural Networks (GNNs) to urban systems has gained momentum due to their ability to model spatial and relational dependencies in graph-structured data. This section summarises key advancements relevant to freight transportation and urban morphology modelling, highlighting how they inform the research.

2.1. Traffic Flow Prediction with GNNs

GNNs have demonstrated significant potential in traffic flow prediction. Xiong et al. (2024) introduced the Gated Fusion Adaptive Graph Neural Network (GFAGNN), which dynamically learns spatial and temporal correlations in traffic data by employing adaptive graph convolution and attention mechanisms. While this approach addresses dynamic traffic conditions, its primary focus on sensor-based data limits its applicability to morphological influences.

2.2. Graph Convolutional Networks in Road Networks

Jepsen et al. (2020) proposed Relational Fusion Networks, tailored for road network tasks such as travel time estimation and speed prediction. This method highlights the advantages of edge-centric learning in road networks and addresses challenges like low-density connectivity and sharp boundaries in network homophily.

2.3. Quantifying Urban Spatial Homogeneity

Xue et al. (2021) presented a graph-based machine learning approach to measure spatial homogeneity in urban road networks. By analysing topological similarities across subnetworks, their work reveals the interplay between urban morphology and socioeconomic indicators. However, it focuses on structural homogeneity without leveraging predictive modelling for transportation flows.

2.4. General Advances in GNNs

Wu et al. (2019) provided a comprehensive survey of GNNs, categorising their evolution into recurrent, convolutional, and spatial-temporal models. Their work highlights the adaptability of GNNs to non-Euclidean data, which is pivotal for modelling urban systems. Similarly, Battaglia et al. (2018) introduced relational inductive biases as a framework to enhance generalisation in structured data, emphasising the significance of graph networks in relational reasoning.

2.5. Knowledge Gap

Despite significant advancements in urban freight transportation modelling, a critical gap persists in the integration of urban morphology - specifically, the spatial configuration and land use patterns of urban areas into these models. Traditional freight transport models often emphasize economic factors, infrastructure capacity, and policy impacts, yet they frequently overlook how the physical layout of a city influences freight movement and demand. This oversight is particularly pronounced in data-scarce environments, where limited access to detailed transportation and logistics data hampers the development of accurate and responsive models.

Urban morphology plays a pivotal role in shaping transportation dynamics. The arrangement of roads, the distribution of commercial and residential zones, and the density of urban development directly affect freight routes, delivery efficiency, and overall logistics planning. For instance, the phenomenon of logistics sprawl, characterised by the relocation of logistics facilities from inner urban areas to suburban zones, has been observed to increase truck travel distances and associated emissions (Aljohani and Thompson, 2016). Understanding these spatial nuances is essential for creating models that accurately reflect real-world freight movements.

In parallel, the emergence of machine learning techniques, particularly Spatio-Temporal Graph Neural Networks (STGNNs), offers promising avenues for modelling complex spatiotemporal relationships in transportation systems. STGNNs are adept at capturing both the spatial dependencies inherent in transportation networks and the temporal dynamics of traffic flows (Jiang et al., 2023). Their application has shown potential in traffic prediction tasks, leveraging the topological structure of road networks to forecast traffic conditions accurately.

Despite these technological advancements, there remains a paucity of research focused on the application of STGNNs to model the influence of urban morphology on freight transportation. The proposed research aims to bridge this gap by developing an STGNN-based framework that incorporates urban form characteristics into freight transport models. By doing so, it seeks to enhance the predictive accuracy of these models and provide valuable insights for urban planners and policymakers striving to design efficient and resilient urban freight systems.

Addressing this knowledge gap is crucial for several reasons. Firstly, integrating urban morphology into freight models can lead to more sustainable and efficient logistics operations by aligning transportation planning with the physical realities of urban environments. Secondly, in data-scarce settings, leveraging the structural information of urban form can serve as a proxy for missing data, thereby improving model robustness and applicability. Ultimately, this research endeavours to contribute to the development of holistic urban freight transport models that are both data-efficient and sensitive to the spatial intricacies of urban landscapes.

3. Research Questions

Freight transportation in urban areas is a multifaceted and dynamic process influenced by a range of factors including infrastructure, socio-economic conditions, and environmental policies TNO (2020). A significant challenge in this domain is the scarcity of comprehensive and high-quality data, which limits the accuracy of predictive models Wu et al. (2019). Temporal Graph Neural Networks (TGNNs) provide a promising framework to model these complex interactions over time, leveraging both spatial and temporal dependencies to enhance predictive capabilities Xiong et al. (2024); Jepsen et al. (2020).

To address the overarching research question:

> **"To what extent can insights into urban morphology, modeled with Temporal Graph Neural Networks, enhance the accuracy and adaptability of freight transportation predictions in the Netherlands?"**

The following sub-questions are proposed to guide the investigation:

- **What are the suitable GNN architectures for this task?**
  Identifying and evaluating the most effective Graph Neural Network (GNN) architectures is crucial for capturing the interactions between urban morphology and freight transportation dynamics. Existing studies indicate that models such as Gated Fusion

Adaptive Graph Neural Networks (GFAGNN) and Spatio-Temporal Graph Convolutional Networks (STGCN) offer advantages in handling spatial and temporal dependencies Xiong et al. (2024); Yu et al. (2018). Furthermore, advancements in spatial-temporal neural networks have led to the development of hybrid models such as Diffusion Convolutional Recurrent Neural Networks (DCRNN) and Spatio-Temporal Graph Convolutional Networks (STGCN), which integrate convolutional structures to enhance temporal learning capabilities Li et al. (2018). Among these, STGCN-based frameworks have shown significant promise in addressing the complexities of dynamic freight transportation networks due to their ability to model sequential dependencies effectively Yu et al. (2018); Wu et al. (2019).

- **What are the relations between road and building networks, and how are they defined?**
  Road networks exhibit complex relational structures, particularly in terms of connectivity, accessibility, and spatial organisation. While freight transportation patterns are influenced by urban spatial structure, the interaction between road and building networks remains an area requiring further exploration (Xue et al., 2021; Jepsen et al., 2020). This study seeks to formalise these relationships using graph-based methodologies.

  To analyse these interactions, the six-position postal code areas (PC6) from CBS are used as spatial aggregation units for extracting and structuring urban features. Each PC6 area represents a high-resolution spatial unit containing approximately 20 households, providing a detailed framework for analysing socio-economic and spatial patterns. Within each PC6 boundary, building attributes - such as land use, density, height, and function, are aggregated and linked to the corresponding road network features.

  By incorporating socio-economic data from CBS, including income levels, employment rates, and business activity, contextual relationship is established between the built environment and transportation infrastructure. This approach allows for a more refined analysis of how urban morphology influences freight flows, contributing to the development of graph-based models.

- **What are significant time frames for the temporal aspect of such neural networks?**
  The temporal granularity of transportation models is essential for capturing relevant freight movement patterns. Prior studies emphasize the importance of selecting appropriate time windows for prediction, particularly in highly dynamic urban environments Li et al. (2018); Rozemberczki et al. (2020). Based on the length of prediction, traffic forecast is generally classified into two scales: **short-term** (5 - 30 min) and **medium-to-long-term** (over 30 min) Yu et al. (2018). Most prevalent statistical approaches, such as linear regression, perform well on short-interval forecasts. However, due to the uncertainty and complexity of traffic flow, these methods are less effective for relatively long-term predictions Yu et al. (2018). STGCN is suitable for both short-term (5-30 min) and medium-to-long-term (30+ min) predictions. The convolution-based architecture avoids error accumulation issues seen in recurrent models, making it a robust choice for freight transportation forecasting Yu et al. (2018).

3.0.1. Relevance of the Research Question

Freight transportation plays a crucial role in urban sustainability, economic growth, and environmental impact. The optimisation of freight movement within cities can lead to reduced congestion, lower emissions, and improved efficiency in logistics. However, traditional transportation models often fail to incorporate the spatial-temporal complexity of urban networks, leading to suboptimal predictions and planning strategies Yu et al. (2018).

Academically, this research contributes to the field of Graph Neural Networks (GNNs) by extending their application to dynamic urban transportation modelling. While GNNs have been extensively studied in static graph scenarios, their integration into temporal urban systems remains an open challenge. This study aims to bridge that gap by leveraging Spatio-Temporal Graph Convolutional Networks (STGCN) to model freight movement patterns more accurately.

From a societal perspective, the outcomes of this research can aid policymakers, urban planners, and logistics companies in making data-driven decisions. By improving predictive models, cities can design better freight policies, optimise delivery routes, and mitigate the negative externalities associated with urban freight transport. Ultimately, this research seeks to enhance the adaptability and accuracy of freight transportation forecasts, contributing to smarter and more sustainable urban logistics systems.

4. Methodology

4.1. Internship at TNO

My internship at TNO, within the Sustainable Transport and Logistics (STL) unit, provided an essential foundation for my graduation research. While working with MASS-GT, a Multi-Agent Simulation System for Goods Transport, I focused on validating and refining the Firm Synthesiser module, which generates synthetic firm distributions based on urban data. This involved comparing synthetic outputs with real-world datasets from the Municipality of Amsterdam using spatial analysis techniques such as clustering, density estimation, and point-to-point distance calculations.

However, during this process, a key realisation emerged: while existing predictive freight transport models effectively simulate logistics flows, they rely primarily on road networks, traffic and partly on socio-economic data, largely overlooking the structural influence of the built environment. Freight movements are not dictated by road networks alone; they are fundamentally shaped by urban morphology - building functions, density, land use, and socio-economic activity. Despite working with advanced simulation models, the absence of cityscape features in predictive frameworks created a blind spot in forecasting accuracy and making the model less adaptive for new contexts.

This gap became the foundation of my graduation research. Building on my experience with MASS-GT and VMA, my project aims to integrate urban morphology and socio-economic data into GNNs to enhance freight transportation forecasting. By moving beyond purely infrastructure-based predictions and embedding spatial context into machine learning models, I aim to capture deeper patterns in freight movement that remain unaccounted for in conventional simulations. My internship not only introduced me to the technical methodologies needed for this research but also highlighted the critical need for a paradigm shift in how urban logistics models are designed.

4.1.1. MASS-GT + VMA

The integration of the Multi-Agent Simulation System for Goods Transport (MASS-GT) with the Verkeersmodel Amsterdam (VMA) enhances freight transport modelling by combining agent-based freight demand simulation with large-scale traffic network models. This integration allows for improved accuracy across macro, meso, and micro levels, leveraging the strengths of each model.

History and Development MASS-GT was developed at Delft University of Technology (TU Delft) and has been applied in various European research projects, including its role as the Tactical Freight Simulator (TFS) in the HARMONY project under Horizon 2020 (De Bok et al., 2021). The model was designed to provide detailed freight demand forecasting and logistics behaviour simulation, enabling better transport policy analysis, originally for Metropolregio Rotterdam Den Haag (MRDH) (see fig. 1). VMA, on the other hand, is a well-established transport model for the Amsterdam region, primarily focusing on passenger traffic to support infrastructure planning Gemeente Amsterdam (2019).

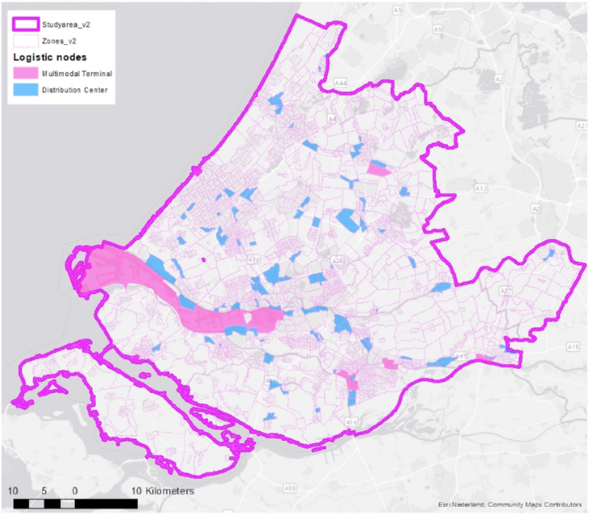


Figure 1: Study area: the province of South Holland in The Netherlands.

Modelling Approach and Scale Integration MASS-GT and VMA operate at different spatial and temporal scales, making their integration essential for a comprehensive understanding of freight transport dynamics:

- - **Macro-Level (Strategic):** MASS-GT simulates freight demand by modelling firm activities, shipment generation, and logistics decision-making processes.
- - **Meso-Level (Tactical):** The integration allows for route choice modelling, vehicle type selection, and tour formation using OD matrices derived from MASS-GT outputs.
- - **Micro-Level (Operational):** VMA refines the spatial distribution of freight trips, providing insights into traffic flow, network congestion, and vehicle interactions.

Transport Flow Types As illustrated in Figure 2 and detailed in Table 9 (appendix D), transport flows describe the movement of goods between different logistical entities, including producers, distribution centres (DCs), transshipment terminals (TTs), and consumers. Internal flows occur within the modelled region, such as direct deliveries from producers to consumers (Flow Type 1), shipments from producers to distribution centres (Flow Type 2), and movements between DCs (Flow Type 5). Additionally, intermediary flows involve logistical hubs such as transshipment terminals, where goods are rerouted between producers, distribution centres, and consumers, as seen in Flow Types 4, 6, 7, 8, and 9. External flows extend beyond the modelled area, representing exchanges between external producers, consumers, and logistical hubs, as classified under Flow Types 10, 11, and 12. These classifications provide insight into the structure of freight movement and form the basis for understanding freight demand and transport network efficiency.

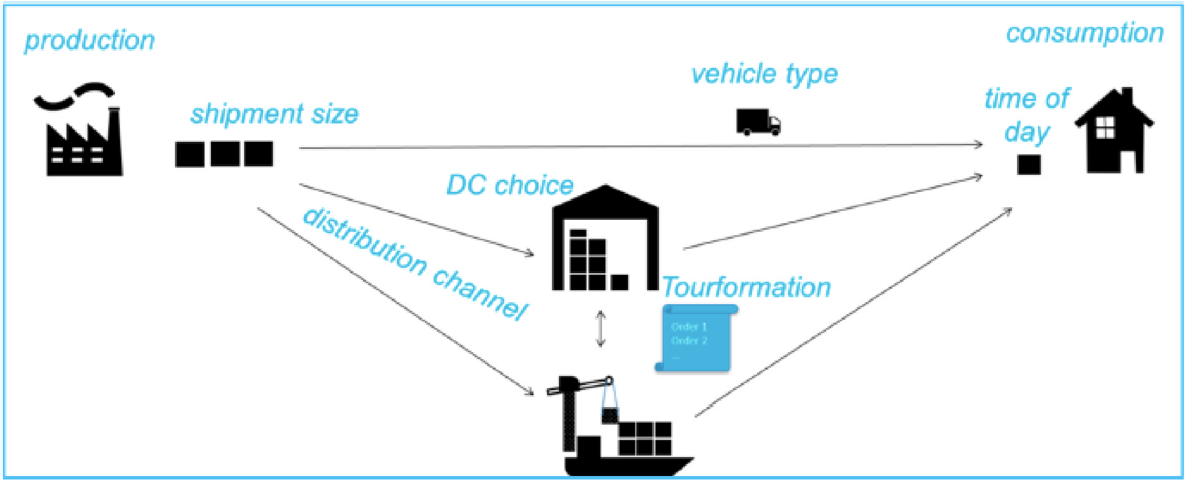


Figure 2: Conceptual model for logistic choices in MASS-GT

Data Sources and Calibration The integration of MASS-GT and VMA is achieved through a variety of real-world datasets. Among others, the following datasets were used:

- - **CBS Freight Trip Data (VESDI):** The dataset includes extensive truck trip diaries collected from transport management systems, providing empirical data on vehicle movements (De Bok et al., 2022).
- - **BasGoed National Freight Model:** A large-scale Dutch freight demand model, used to calibrate freight volumes and economic flows at the national level (De Bok et al., 2017).
- - **Firm Population Data from CBS:** Business registry data that defines the distribution and characteristics of firms, crucial for synthetic firm population modeling (, CBS).
- - **VMA Traffic Flow Data:** Observed vehicle counts and congestion patterns in the Amsterdam region, used to validate micro-level freight movements from ANPR dataset.

4.1.2. TNO's Digital Twin

TNO's Digital Twin is a high-fidelity simulation framework that models the spatial and temporal distribution of freight transport in urban environments. The system integrates Origin-Destination (OD) matrices from MASS-GT and VMA to allocate freight and passenger traffic onto specific roads in Amsterdam. This allocation process generates a highly detailed dataset, mapping vehicle flows across different road segments on an hourly basis for an average working day.

One of the key advantages of utilising the Digital Twin's output for this study is its reliability as a ground-truth reference. While raw datasets such as OD matrices and network attributes provide valuable insights, they remain incomplete representations of real-world freight movement. The Digital Twin refines these inputs by incorporating the most advanced transport models currently available, ensuring a realistic and data-informed allocation of freight flows. By leveraging this enriched dataset, the proposed Graph Neural Network (GNN) model can be trained and validated with a high degree of accuracy.

Another crucial aspect of the Digital Twin's output is its capacity to fill gaps in data availability. Real-world freight transport data often suffers from inconsistencies due to missing or fragmented datasets. The DT mitigates these limitations by combining multiple validated models to produce a cohesive and structured freight allocation dataset, making it a more robust and scalable alternative to relying solely on raw sensor or survey data.

At present, a screenshot of the Digital Twin's interface and outputs cannot be provided, as access to the GitLab repository containing the necessary files is still pending. However, the process for obtaining access is already in progress, and once available, visual representations and precise data structure of the DT's outputs will be included in the final documentation.

4.1.3. Urban Morphology and Road Networks

Urban morphology is the study of the physical form of cities, focusing on the patterns and processes of their formation and transformation over time. It examines the spatial structure and character of urban areas by analysing their component parts and the relationships between them. The main elements of urban form include:

- **Streets**: The network of pathways facilitating movement and access within the city.
- **Plots**: The subdivisions of land delineating property boundaries.
- **Buildings**: The structures erected within plots, serving various functions.

These elements interact to shape the urban landscape, influencing both its functionality and aesthetics (Silva, 2022; Kropf, 2017).

Similarly, road networks form the backbone of urban transportation, dictating how goods and people move within a city. They consist of:

- **Roads**: Pathways designated for vehicular traffic, varying in scale from local streets to highways.
- **Intersections**: Junctions where two or more roads meet, regulating traffic flow.
- **Interchanges**: Complex intersections, often involving bridges or tunnels, allowing for the free flow of traffic between major roads.

The design and organisation of road networks significantly impact urban accessibility, mobility, and the overall efficiency of transportation systems (Bell and Iida, 1997; World Bank).

Application to this research In existing freight transportation models, road networks are typically the primary focus, while the influence of urban morphology remains underexplored. However, urban morphology plays a crucial role in shaping freight flow patterns. The density, function, and spatial arrangement of buildings influence where goods are delivered, how freight routes are structured, and the overall efficiency of urban logistics.

This research aims to integrate urban morphology features into Graph Neural Network (GNN) models to enhance freight transportation forecasting. By incorporating variables such as building footprints, land use types, and plot density, the model will account for the spatial constraints and economic activity that drive freight movement.

4.2. Literature Review

4.2.1. Knowledge Graphs and GNNs Architectures

Knowledge Graphs in Freight Transportation Knowledge Graphs (KGs) are structured representations of relationships between entities, capturing complex interactions in a machine-readable format. They consist of nodes representing entities (e.g., road segments, intersections, freight hubs) and edges encoding relationships (e.g., traffic flow, logistics connections, regulatory constraints) (Hogan et al., 2021).

Graph Neural Networks for Freight Flow Prediction Graph Neural Networks (GNNs) are deep learning architectures designed to process graph-structured data, making them well-suited for transportation modeling. Unlike traditional machine learning models, which treat input data as independent observations, GNNs leverage relational dependencies between nodes and edges to capture spatial and temporal patterns (Wu et al., 2021).

In this study, the objective is to use GNNs to predict freight flow distributions based on the constructed Knowledge Graph (KG).

GNN Architectures Several GNN architectures are suitable for modelling freight transport networks:

- **Graph Convolutional Networks (GCNs)**: Aggregate information from neighbouring nodes to learn spatial relationships between road segments and hubs (Kipf and Welling, 2017).
- **Graph Attention Networks (GATs)**: Apply attention mechanisms to weigh the importance of different node connections, allowing for more refined predictions (Veličković et al., 2018).
- **Graph Recurrent Neural Networks (GRNNs)**: Incorporate temporal dependencies into freight flow predictions by modelling sequential interactions within the transport network (Seo et al., 2018).
- **Graph Transformers**: Adapt transformer-based architectures for graph data, enhancing the model's ability to capture complex spatial dependencies (Dwivedi and Bresson, 2021).
- **Spatio-Temporal Graph Convolutional Networks (STGCNs)**: Extend traditional GCNs by incorporating time series data, making them ideal for modelling dynamic freight flows and predicting future traffic states based on historical data (Yu et al., 2018).

Chosen Model: Spatio-Temporal Graph Convolutional Networks (STGCNs) For this study, STGCNs are chosen due to their ability to model both the spatial dependencies of road networks and the temporal dynamics of freight flows. Unlike static GCNs, which capture spatial relationships but lack temporal awareness, STGCNs leverage both graph convolution operations and temporal convolution layers to capture patterns over time.

The core mechanism behind STGCNs consists of two primary components:

1. Spatial Graph Convolutions: These layers capture the connectivity between road segments by propagating information across the graph structure. Each road segment (node) aggregates information from its neighbours, allowing the model to understand how congestion or freight flows propagate in the network.
2. Temporal Convolutions: Instead of relying on recurrent architectures (such as GRNNs), STGCNs use 1D convolution operations over time series data, enabling efficient learning of temporal dependencies in freight movement.

This dual approach makes STGCNs particularly well-suited for freight transport modelling, as freight flows are highly dynamic, varying by time of day, congestion patterns, and regulatory constraints. Additionally, STGCNs have demonstrated strong performance in traffic forecasting tasks, outperforming traditional time-series models like LSTMs in terms of efficiency and scalability (Yu et al., 2018).

Input Vector Data

This section provides an overview of the datasets used in this study, their sources and formats. The datasets cover road networks, socio-economic data, building information and logistics infrastructure for freight modelling.

4.1. Datasets Overview

The datasets used in this research originate from various sources, including open data portals and proprietary datasets. Table 1 provides a summary of the datasets.

Name	**Format**	**Scope**	**Source**
Road Networks	Shapefile (.shp)	MRDH, Amsterdam, NL,BE,DE highways	TNO
Socio-economic Data (PC6, PC4)	GeoPackage (.gpkg), CSV	Netherlands	CBS
Building Data	format to be specified	Netherlands	Amin Jalilzadeh
Distribution Centers	Shapefile (.shp)	MRDH, Amsterdam	TNO
Parcel Depots	Shapefile (.shp)	Netherlands	Delivery Companies
Transshipment Terminals	Shapefile (.shp)	MRDH	TNO

Table 1: Overview of vector datasets used in this study

4.2. Road Network Data

The road network data consists of multiple datasets, covering both local roads and major highways. The datasets are provided in vector format as shapefiles and contain attributes such as road type, number of lanes, and speed limits. In some sets, a specified number of different type of vehicles is also assigned.

4.2.1. Dataset Details

- **Name:** Road Network - roads and junctions
- **Format:** Shapefile (.shp)
- **Scope:** MRDH, Amsterdam, NL, BE, DE highways
- **Features:** Road type, number of lanes, speed limit
- **Purpose:** Used to model transport infrastructure and define road connectivity for freight flow simulation.

4.3. Socio-economic Data

The socio-economic dataset is provided by CBS and contains statistical information aggregated at different spatial levels, including PC6 and PC4 areas (see fig. 5.

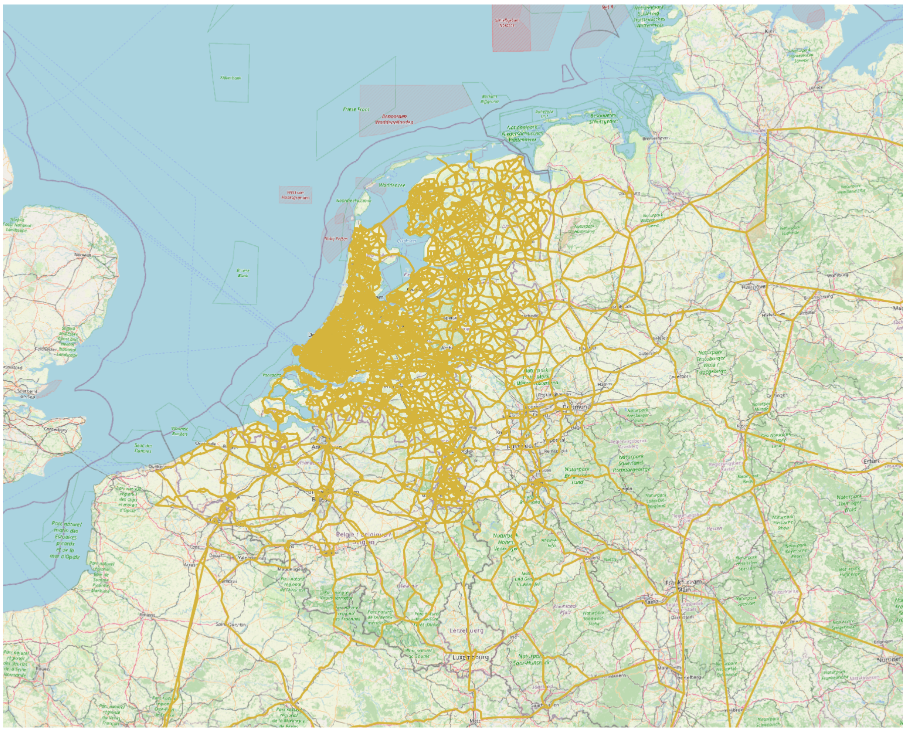


Figure 3: Road network dataset previewed in Qgis.

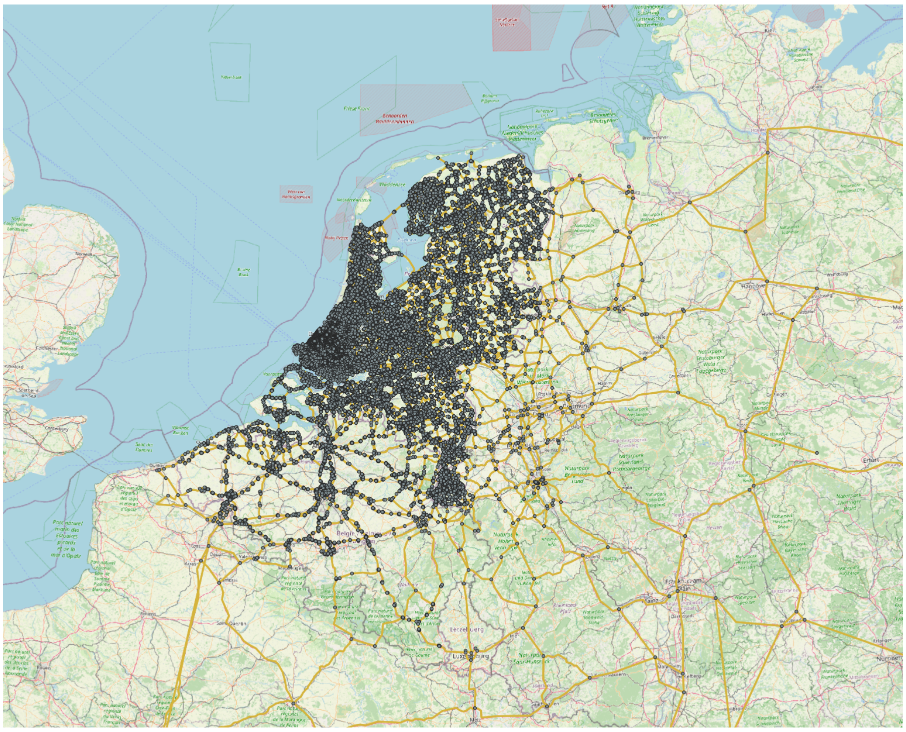


Figure 4: Junctions dataset previewed in Qgis.

4.3.1. Dataset Details

- **Name:** CBS Socio-economic Data
- **Format:** GeoPackage (.gpkg), CSV

- **Scope:** Netherlands
- **Features:** pc6 data
- **Purpose:** Used to link socio-economic conditions to freight activity and demand modelling

Figure 5: PC6 areas, excerpt from Amsterdam, previewed in Qgis.

4.4. Building Data

The building dataset provides spatial, functional and energy attributes of buildings across the Netherlands.

4.4.1. Dataset Details

- **Name:** Building Data
- **Format:** to be specified, csv/neo4j?
- **Scope:** Netherlands
- **Features:** Age, occupancy, footprint, height, energy label
- **Purpose:** Used to estimate land use and freight demand/routing based on building function

4.5. Logistics Infrastructure Data

This section includes datasets related to freight logistics hubs such as distribution centres, parcel depots, and transshipment terminals (see fig. 6).

4.5.1. Distribution Centers

- **Name:** Distribution Centers
- **Format:** Shapefile (.shp)
- **Scope:** MRDH, Amsterdam
- **Features:** location
- **Purpose:** Used to analyse freight movement between logistics hubs

4.5.2. Parcel Depots

- **Name:** Parcel Depots
- **Format:** Shapefile (.shp)
- **Scope:** Netherlands
- **Features:** Company, surface area, region
- **Purpose:** Used to model last-mile delivery operations

4.5.3. Transshipment Terminals

- **Name:** Transshipment Terminals
- **Format:** Shapefile (.shp)
- **Scope:** MRDH
- **Features:** location, area, perimeter
- **Purpose:** Used to model multimodal freight transfers

4.6. Machine Learning Model Development

The development of the machine learning model follows a structured, multi-phase approach to assess the impact of different data components on freight flow prediction. The training process focuses on an average working day, with hourly freight flow data for a selected district in Amsterdam, which will be determined based on the output from TNO's Digital Twin once access to the repository is granted.

Training Phases The model will undergo three sequential training phases, each incorporating an increasing amount of information to analyse its effect on predictive accuracy:

1. **Phase 1 – Road Network Only:** The first phase will train the Spatio-Temporal Graph Convolutional Network (STGCN) using only road network data (RN). This serves as a baseline to evaluate how well the network structure alone can inform freight flow predictions.

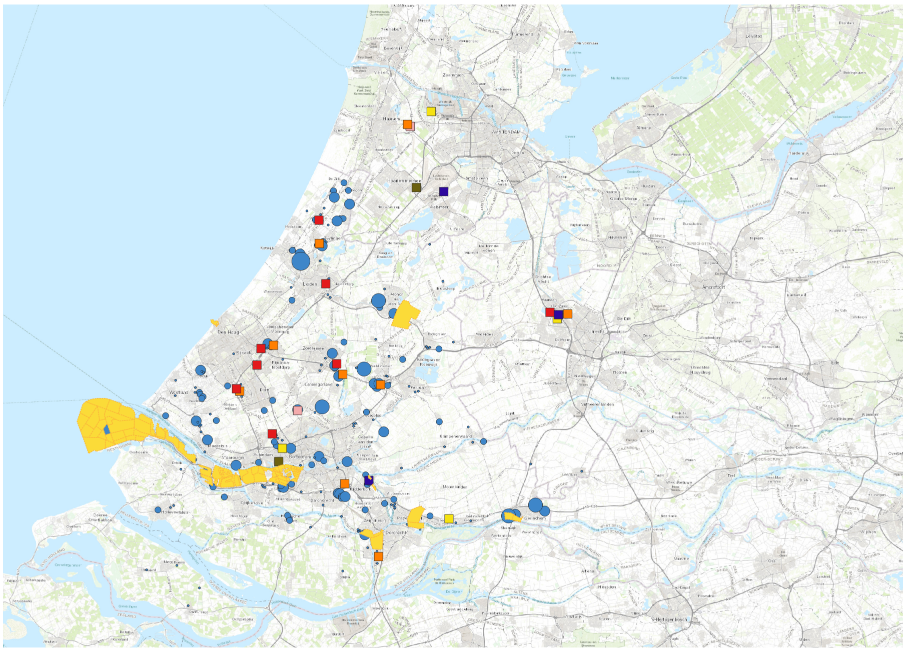


Figure 6: Logistic Infrastructure

2. **Phase 2 – Road Network + 70% DT Output:** In the second phase, the training dataset is expanded to include 70% of TNO's Digital Twin output, which provides allocated vehicle flows per road segment per hour. This integration enhances the model's ability to capture real-world freight distributions and adds the temporal dimension to the network.

3. **Phase 3 – Full Dataset (RN + DT + Urban Morphology):** The final training phase incorporates urban morphology data, such as building densities, land use types, and socio-economic attributes, along with the previous datasets. This phase evaluates the extent to which urban morphology improves the predictive capabilities of the model.

Justification for Multi-Phase Training This incremental training methodology allows for a structured evaluation of how different input features contribute to prediction performance. By first training on road network data alone, the model establishes a spatial baseline. Introducing freight allocation data from the Digital Twin improves predictions by incorporating traffic flow information. Finally, integrating urban morphology characteristics accounts for land use effects, which may further refine accuracy by linking freight activity to built environment patterns.

Feature vs. Dimension

An important consideration in the development of the machine learning model is how different data components are incorporated i.e. either as features or as additional dimensions in the input tensor. The choice between these approaches can significantly influence the model's ability to capture complex spatial and temporal dependencies, impacting predictive accuracy and computational efficiency.

Defining Features vs. Dimensions Features represent descriptive attributes of entities (e.g., road segments, vehicles, or socio-economic indicators). These attributes are included as addi-

tional node or edge features in the model, providing context without altering the data structure.
Dimensions define structural axes along which the data is indexed, typically representing spatial, temporal, or categorical variations. Organising data along these dimensions allows the model to explicitly learn patterns and dependencies across them.

Evaluating Location as a Feature or a Dimension A key design choice is how to incorporate location - such as province or city, in the input representation. Two alternative approaches exist:

- Location as a Feature: Treating location as a feature means including it as an attribute for each entity (e.g., node or edge). This approach allows the model to learn location-specific variations while maintaining flexibility in data structure.
- Location as a Dimension: Incorporating location as a dedicated dimension restructures the tensor to separate data along different locations. This enables operations such as convolutions or attention mechanisms to capture spatial dependencies explicitly.

The suitability of each approach depends on the structure of the data and the modelling objectives. For structured spatial data (e.g., grid-based datasets), location as a dimension can enhance the model's ability to learn local interactions. Conversely, in irregular networks (e.g., road graphs), treating location as a feature allows the model to infer spatial relationships more flexibly.

Application to Freight Transport Modelling In the context of freight transport, different components lend themselves more naturally to being features or dimensions. Freight transport is inherently temporal, making time a natural choice as a tensor dimension. Vehicle Type Categories can be encoded as a categorical dimension, allowing the model to distinguish between different transport modes. On the other hand, attributes such as road type, speed limits, or economic factors are best treated as features since they provide descriptive context rather than defining an inherent structural axis.

Proposed Empirical Evaluation Given the potential influence of this decision, an empirical evaluation will be conducted to assess the impact of treating city location as a feature versus a dimension. The methodology will involve:

1. Feature-Based Approach: Training a model where location is included as a node or edge attribute.
2. Dimension-Based Approach: Structuring the input tensor to separate data by location as an additional axis.
3. Benchmarking: Comparing model performance across key metrics such as predictive accuracy, computational efficiency, and the ability to capture spatial dependencies.

Conclusion The decision to represent location as a feature or a dimension has implications for model design and performance. By systematically evaluating both approaches, this study aims to determine the most effective strategy for modelling freight transport dynamics, ensuring that spatial and temporal dependencies are captured optimally.

4.7. Validation Testing

To ensure the robustness of the proposed model, a systematic validation approach is adopted, leveraging existing simulation outputs and real-world traffic sensor data. The validation consists of two key components: comparing model predictions with the TNO Digital Twin and using high-resolution highway sensor data for evaluation.

Validation Using TNO's Digital Twin TNO's Digital Twin serves as the most precise available benchmark for freight vehicle allocation across the road network. The DT model integrates results from MASS-GT and VMA, processing them into a refined vehicle distribution model. To validate the proposed Graph Neural Network model, 30% of unseen data from the DT will be reserved as a validation set. By comparing the model's predictions against this high-fidelity dataset, it is possible to assess the accuracy of freight flow estimations at various road segments.

Validation Using Highway Sensor Data Highway sensor data provides another high-precision validation method, as it offers minute-level vehicle counts for different vehicle types at multiple sensor locations. This dataset consists of matrices where each row corresponds to a specific sensor, and each column represents a vehicle count per time interval.

Performance Metrics To quantitatively evaluate model accuracy, the following metrics will be used:

- **Mean Absolute Error (MAE):** Measures the absolute differences between predicted and actual vehicle counts.
- **Root Mean Squared Error (RMSE):** Provides insight into prediction variance by penalising larger errors.
- **R-squared (R^2) Score:** Assesses the proportion of variance explained by the model compared to ground-truth data.

These metrics will be computed separately for TNO's DT validation dataset and highway sensor data to assess model generalisation across different traffic scenarios.

Expected Challenges Validation accuracy may be influenced by differences in sensor placement, traffic anomalies, and missing data. Additionally, while TNO's Digital Twin provides a highly detailed freight movement prediction, it remains a simulated dataset rather than direct ground truth. As a result, highway sensor data will serve as an additional real-world validation step.

4.8. Scalability Plan

Definition of Scalability in This Context Scalability refers to the model's ability to generalise across different urban environments, transport networks, and data availability levels while maintaining predictive accuracy. This includes:

- **Geographical scalability**: Applying the model to different cities and regions.
- **Computational scalability**: Handling larger datasets efficiently.
- **Temporal scalability**: Adapting to changing freight dynamics over time.

Steps to Ensure Scalability in the Development Process

To enable the model's transferability from a single city to multiple regions, the following strategies will be incorporated into the research and implementation phases:

Modular Data Inputs To ensure the model remains adaptable across different urban environments, data inputs are sourced from standardised datasets such as OpenStreetMap (OSM), CBS, and European Commission datasets, alongside proprietary data from TNO. Instead of relying on a single predefined dataset, a modular approach allows the integration of multiple data sources, ensuring flexibility in regions with varying levels of data availability.

The integration of urban morphology, transport networks, and firm distributions is managed through a graph database framework such as Neo4j, which enables efficient querying and relationship modelling across diverse datasets. Urban morphology data, including building footprints, land use classifications, and street layouts, is extracted from sources such as OSM and 3D BAG. Transport network data, sourced from OpenStreetMap and TNO's Digital Twin, ensures accurate representation of road infrastructure and freight corridors. Additionally, firm-level logistics behaviour is modelled using freight activity data from CBS and proprietary TNO datasets.

By leveraging Neo4j's graph-based storage and query capabilities, the model dynamically integrates new data sources without requiring extensive manual adjustments. This modular pipeline ensures scalability, enabling the model to be applied across multiple urban regions while maintaining accuracy and efficiency.

Flexible Knowledge Graph Construction: To ensure the model remains scalable and adaptable across different urban environments, the Knowledge Graph (KG) is designed as a dynamic structure rather than a static predefined dataset. This is achieved by implementing an automated pipeline that extracts urban morphology and freight flow data from multiple sources, including OpenStreetMap (OSM), 3D BAG, CBS freight statistics, and TNO's Digital Twin.

Unlike traditional graph structures, which rely on predefined relationships, the KG dynamically integrates new data points by adjusting its schema based on available transport and infrastructure data. This is facilitated through a graph database framework such as Neo4j, which enables efficient querying, relationship modelling, and real-time updates as new freight transport patterns emerge.

The pipeline consists of three key stages:

- **Graph Construction:** Nodes (e.g., road segments, logistic hubs, firms) and edges (e.g., freight flows, road connectivity) are extracted dynamically from external data sources.
- **Attribute Enrichment:** Nodes and edges are enriched with urban morphology attributes, including land use, building density, street width, and transport regulations.
- **Schema Adaptation:** The KG structure evolves over time by integrating new spatial and freight-related constraints, making it transferable across different urban settings.

By leveraging Neo4j's graph database capabilities, the KG can efficiently store, query, and update transport relationships while maintaining scalability across different cities and countries. This ensures that the model can be applied beyond its initial test environment, making it a valuable asset for large-scale freight transport analysis.

Model Training on Multiple Urban Areas

Instead of training GNN only on one city, the training dataset will cover different parts of Amsterdam, including areas with diverse urban structures (e.g., dense city centre, industrial zones, suburban areas).
The model will first be trained on sections of Amsterdam and fine-tuned using limited data from other cities such as Rotterdam and The Hague.

Application to Different Cities and Countries

Scale	Implementation Strategy
City-Level	Train the model on different **districts of Amsterdam**, validate on **Rotterdam and The Hague**.
Regional-Level	Extend the model to **South Holland**, covering multiple urban typologies (e.g., port areas, suburban logistics hubs).
National-Level	Integrate with **BasGoed**, the Dutch national freight model, and validate results on Utrecht and Eindhoven.
International-Level	Apply the model to cities with different urban structures (e.g., **Warsaw, Berlin, London**) using European open datasets.

Table 2: Scalability Plan for Expanding the GNN Model to Multiple Regions

The main scope of this graduation project is to cover the **City-Level**. The remaining levels can be seen as extensions planned for future work (see fig. 2).

Value for TNO and Future Research

Scalability ensures that the model is not only useful for a single city but can be adapted for:

- Supporting decision-making across multiple urban areas, not just Amsterdam.
- Higher prediction accuracy
- Using the model where existing data is not sufficient to predict the freight logistics

By ensuring scalability in the development process, this research enables broader applicability, making the model a valuable asset for freight transport analysis at multiple scales.

4.9. Expected Results

The final outcome of this research is the development of a Spatio-Temporal Graph Convolutional Network (STGCN) capable of predicting freight traffic dynamics at an hourly resolution for an urban environment at least on a district scale. When provided with road network and building data, the model will generate a graph representation of the city that captures freight flow distributions over time. This framework will allow for dynamic prediction of freight movement based on spatial structure and temporal variations, rather than relying solely on historical transport data.

A key expectation is that the model will be scalable and generalisable across different cities, particularly in locations where transportation data is scarce or incomplete. By leveraging urban morphology and road network information, the model will be able to generate predictions even in data-limited settings, making it applicable for urban planning and freight management strategies. The predicted freight flows will be represented as dynamic graphs, where each node corresponds to a specific location in the urban freight network, while edges represent estimated movements between these points. Traffic volume and other contextual information will be integrated as evolving attributes within the graph, ensuring that predictions align with real-world conditions.

One of the primary advantages of this approach is that the model will be inherently city-sensitive, adapting to the spatial and infrastructural characteristics of different urban environments. Unlike traditional forecasting methods that primarily depend on statistical relationships within existing transport datasets, this approach will directly incorporate city elements like PC6 and building data into the prediction process. The ability to model freight traffic through structural urban data rather than relying exclusively on trip-based surveys or OD matrices will enable more robust and transferable applications.

The integration of recent advances in STGCN research is expected to enhance predictive accuracy. The final deliverables will include a fully trained STGCN model capable of predicting freight traffic flow, a scalable framework that can be applied across different cities, and graph-based visualisations illustrating how freight movement evolves over time.

Since freight transport is inherently dynamic and influenced by a range of spatial and temporal factors, the ability of the model to capture variations in freight distribution will be critical. The model is expected to identify key transport corridors and bottlenecks. As a result, the proposed system will not only contribute to academic research on freight transport modelling but also provide a valuable decision-support tool for urban mobility planning.

5. Time planning

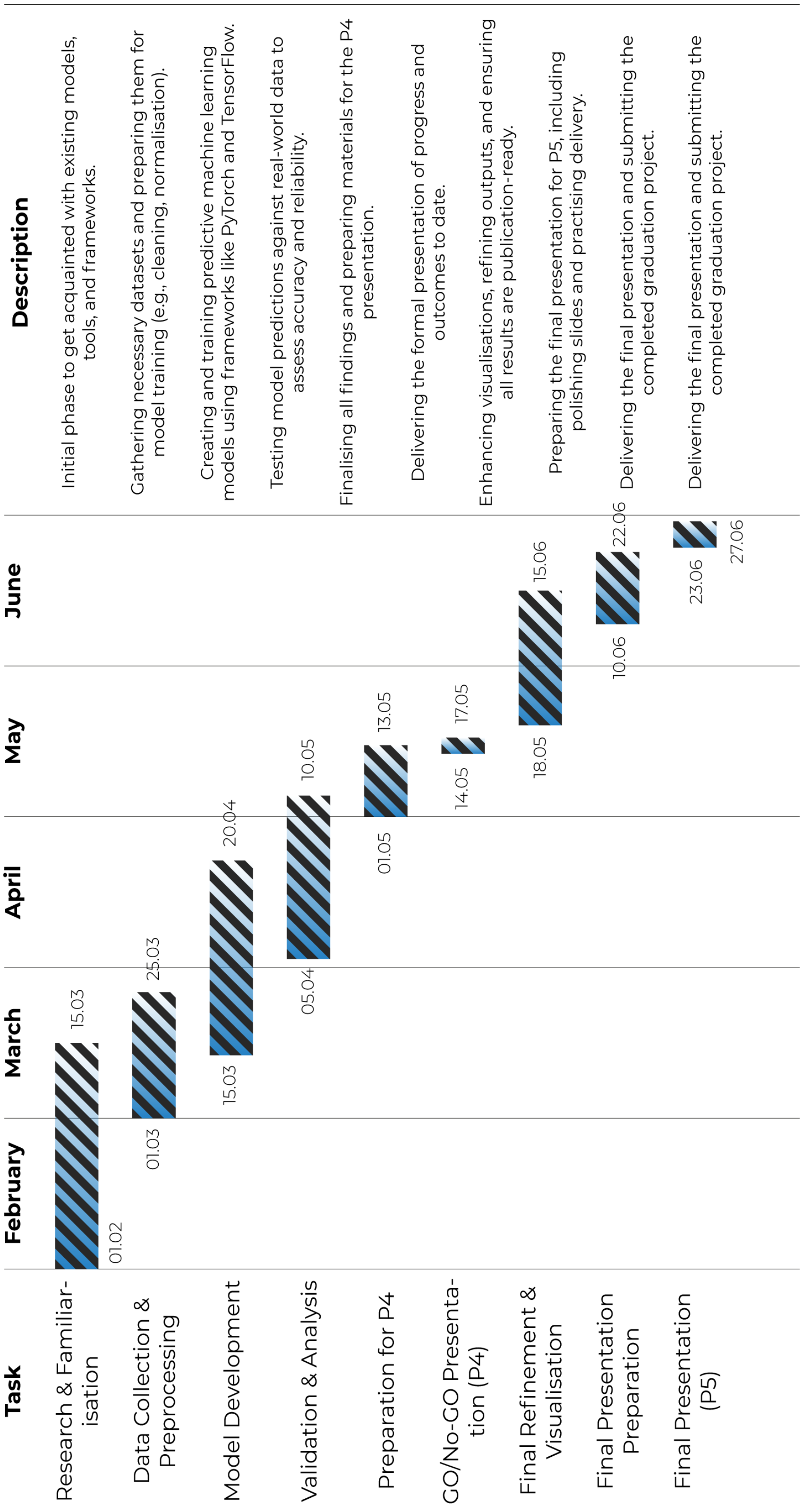


Figure 7: Graduation Planning Gantt chart

5.1. February – Research & Familiarisation (Week 1-4)

16-29 February: The initial prototype of the model is developed, and the first datasets are introduced for testing. The focus begins with a selected district of Amsterdam, allowing for an early evaluation of freight flow predictions. Initial testing is conducted to determine which building-related features contribute most effectively to the model, while spatial aggregation techniques using the PC6 dataset are refined. This phase lays the foundation for structured data integration and further improvements in model accuracy.

5.2. March – Data Collection & Preprocessing (Week 5-8)

1-15 March: The first half of March is dedicated to acquiring and structuring critical datasets, including highway sensor data for validation. Data preprocessing tasks such as cleaning, normalisation, and format conversion are carried out to ensure compatibility with the machine learning framework. During this stage, initial tests are performed to verify the integrity and consistency of the collected data.

16-31 March: The integration of MASS-GT and VMA outputs becomes the primary focus. Custom scripts are developed to efficiently merge various datasets, ensuring that spatial and temporal attributes remain aligned. The consistency of these datasets is rigorously tested, preparing them for direct input into the model training pipeline.

5.3. March-April – Model Development (Week 9-12)

1-15 April: The initial architecture for the STGCN model is implemented. The first phase of training begins, using only road network data to establish a baseline model. Core components such as graph convolutions, temporal dependencies, and loss functions are developed and tested.

16-30 April: The second training phase incorporates 70% of TNO's Digital Twin output, refining predictions by adding dynamic freight allocation data. Model hyperparameters are tuned, and various graph architectures are tested to improve predictive accuracy.

5.4. April-May – Validation & Analysis (Week 13-16)

1-10 May: Model evaluation begins with the use of Mean Absolute Error (MAE), Root Mean Squared Error (RMSE), and R^2 scores. A comparative study is conducted to measure improvements gained by incorporating additional datasets.

11-17 May: The final phase of training includes urban morphology data, allowing for a comprehensive evaluation of its impact on freight transport predictions. The model is tested against real-world datasets to validate its generalisation capabilities.

5.5. May – Preparation for P4 & GO/NO-GO Presentation (Week 17-18)

18-30 May: The research progress is compiled into a formal presentation for the P4 (GO/NO-GO) review. Key findings are summarised, model performance results are visualised, and a structured report is prepared. Presentation slides are refined for clarity and effectiveness.

5.6. June – Final Refinement, Visualisation & P5 Preparation (Week 19-22)

1-15 June: Model predictions are enhanced with improved visualisation techniques. Graph-based results are refined, and final validation steps are conducted. A detailed comparative analysis of different training phases is documented.

16-22 June: The final presentation (P5) is prepared, including a well-structured slideshow and supporting documentation. Practice sessions are conducted to ensure clarity in delivery.

5.7. June – Final Presentation (P5) & Thesis Submission (Week 23-24)

23-27 June: The final research findings are presented during the P5 session. The completed thesis, along with the supporting datasets, models, and documentation, is submitted for evaluation. Final reflections on the project are documented, and potential future research directions are outlined.

6. Tools and Datasets for Development

Tools

Category	Tools and Description
Python Libraries	• **PyTorch:** Deep learning library for graph-based neural networks. • **TensorFlow:** Machine learning framework for advanced model development. • **Scikit-learn:** Essential for preprocessing, evaluation, and simpler models. • **Spektral:** Specialised library for graph neural network models. • **NetworkX:** Graph creation, analysis, and manipulation. • **PyTorch Geometric (PyG):** Optimised library for graph neural networks. • **Deep Graph Library (DGL):** Scalable alternative for graph-based learning. • **SHAP (SHapley Additive exPlanations):** Game theoretic approach to explain the output of any machine learning model
Spatial Data Tools	• **Geopandas:** For handling geospatial data and spatial joins. • **Shapely:** Geometry manipulation and spatial analysis. • **OSMnx:** Extraction of road and spatial networks. • **Fiona:** Reading and writing geospatial file formats.

Table 3: Python Libraries and Spatial Data Tools.

Category	Tools and Description
Machine Learning Utilities	• **XGBoost:** Gradient boosting for tabular data. • **LightGBM:** Fast and efficient boosting alternative. • **Optuna:** Automated hyperparameter tuning.
3D and Geospatial Tools	• **GDAL/OGR:** Handling raster and vector geospatial formats. • **CesiumJS:** Interactive 3D visualisation for spatial data. • **kepler.gl:** Browser-based geospatial visualisation.

Table 4: Machine Learning Utilities and 3D Geospatial Tools.

Category	Tools and Description
Data Management	• **PostGIS:** Spatial database extension for PostgreSQL. • **Neo4j:** Graph database.
Visualisation Tools	• **Matplotlib:** General-purpose plotting library. • **Seaborn:** Statistical visualisations built on Matplotlib. • **Plotly:** Interactive and web-based visualisations. • **Gephi:** Visualising and analysing graph structures.
Temporal Data Libraries	• **PyTorch Geometric Temporal:** Spatio-temporal graph neural network library. • **Statsmodels:** Time-series analysis and statistical modelling. • **Prophet:** Forecasting library for temporal data.

Table 5: Data Management, Visualisation, and Temporal Tools.

Datasets Overview

- **Primary Datasets (Described more in detail in section 4 Methodology):**
  - Road networks, MASS-GT + VMA + TNO's DT provided by TNO.
  - Buildings' data including features as: age, occupancy, function, footprint size, height, energy label provided by the courtesy of Amin Jalilzadeh
  - VESDI dataset

- **Additional Data Sources (if necessary and possible to obtain):**
  - **Road Networks:**
    * NWB
    * OSM (OpenStreetMap)
    * PDOK
    * Kadaster
  - **Buildings:**
    * 3D BAG
    * BAG
    * Kadaster 3D
    * CityGML

6.1. Data for Learning and Testing Purposes

In the beginning, I plan to conduct tests with synthetic data created by me, to see if my functions work. Additionally, to test my network with known datasets I plan to use datasets described in appendix E.

References

K. Aljohani and R. G. Thompson. Impacts of logistics sprawl on the urban environment and logistics: Taxonomy and review of literature. *Journal of Transport Geography*, 57:255–263, 2016. doi: 10.1016/j.jtrangeo.2016.08.009. URL https://doi.org/10.1016/j.jtrangeo.2016.08.009.

P. W. Battaglia, J. B. Hamrick, V. Bapst, et al. Relational inductive biases, deep learning, and graph networks. *arXiv preprint arXiv:1806.01261*, 2018.

M. G. Bell and Y. Iida. *Transportation Network Analysis*. John Wiley & Sons, 1997.

G. Boeing. OSMnx: A Python package to work with graph-theoretic OpenStreetMap street networks. *Journal of Open Source Software*, 2(12):215, 2017. doi: 10.21105/joss.00215. URL https://doi.org/10.21105/joss.00215.

G. Boeing. Osmnx: Python package to work with openstreetmap data. https://github.com/gboeing/osmnx, 2023. Accessed: 2025-01-18.

S. N. (CBS). Dutch firm population database. Technical report, CBS, 2020. URL https://www.cbs.nl/en-gb.

M. De Bok, B. Wesseling, J. Kiel, and J. Francke. An exploration of freight transport forecasts for the netherlands with basgoed. In *Proceedings of the European Transport Conference*, 2017. URL https://aetransport.org/public/downloads/6MAFo/5422-59b6eb44e03aa.pdf.

M. De Bok, L. Tavasszy, A. Nadi, and S. Thoen. Application of the harmony tactical freight simulator to a case study for zero emission zones in rotterdam. In *Proceedings of the European Transport Conference*, 2021. URL https://research.tudelft.nl/en/publications/application-of-the-harmony-tactical-freight-simulator-to-a-case-s.

M. De Bok, L. Tavasszy, and S. Thoen. Simulation of the impacts of a zero-emission zone on freight deliveries in urban areas using an agent-based model. *Transportation Research Record*, 2675(7):1–12, 2022. doi: 10.1177/03611981211012694. URL https://journals.sagepub.com/doi/abs/10.1177/03611981211012694.

V. P. Dwivedi and X. Bresson. A generalization of transformer networks to graphs. In *Proceedings of the AAAI Conference on Artificial Intelligence*, 2021. URL https://arxiv.org/abs/2012.09699.

Gemeente Amsterdam. Uitgangspunten Verkeersmodel Amsterdam 3.0. Technical report, 2019. URL https://maps.amsterdam.nl/verkeersprognoses/UitgangspuntenVMA3.0_v1_1.pdf.

C. L. Giles, K. D. Bollacker, and S. Lawrence. Citeseer: An automatic citation indexing system. In *Proceedings of the third ACM conference on Digital libraries*, pages 89–98. ACM, 1998.

D. Grattarola et al. Spektral datasets. https://graphneural.network/datasets/, 2023. Accessed: 2025-01-18.

A. Hogan, E. Blomqvist, M. Cochez, C. D'Amato, G. Melo, C. Gutiérrez, J. F. Sequeda, M. Maleshkova, J. E. Labra Gayo, R. Navigli, S. Neumaier, A.-C. Ngonga Ngomo, A. Polleres, S. M. Rashid, A. Rula, L. Schmelzeisen, E. Jiménez-Ruiz, T. Soru, A. Splendiani, and F. M. Suchanek. Knowledge graphs. *ACM Computing Surveys*, 54(4):1–37, 2021. doi: 10.1145/3447772. URL https://dl.acm.org/doi/10.1145/3447772.

A. Jain, E. Haghighat, M. Nabian, and S. Nelaturi. Using graph neural networks for additive manufacturing, 2024. URL https://developer.nvidia.com/blog/using-graph-neural-networks-for-additive-manufacturing/.

T. S. Jepsen, C. S. Jensen, and T. D. Nielsen. Graph convolutional networks for road networks. In *Proceedings of the 33rd International Conference on Artificial Neural Networks*, 2020.

F. Jiang, J. Ma, C. J. Webster, A. J. F. Chiaradia, Y. Zhou, Z. Zhao, and X. Zhang. Spatio-temporal graph neural networks for predictive learning in urban computing: A survey. *IEEE Transactions on Knowledge and Data Engineering*, 2023. doi: 10.1109/TKDE.2023.3333824. URL https://doi.org/10.1109/TKDE.2023.3333824.

T. N. Kipf and M. Welling. Semi-supervised classification with graph convolutional networks. In *Proceedings of the International Conference on Learning Representations (ICLR)*, 2017. URL https://arxiv.org/abs/1609.02907.

K. Kropf. *The Handbook of Urban Morphology*. Wiley, 2017. URL https://search.ebscohost.com/login.aspx?direct=true&scope=site&db=nlebk&db=nlabk&AN=1611815.

Y. Li, R. Yu, C. Shahabi, and Y. Liu. Diffusion convolutional recurrent neural network: Data-driven traffic forecasting. In *International Conference on Learning Representations (ICLR)*, 2018.

A. Martin and N. Reichmann. What is a knowledge graph?, 2024. URL https://www.esri.com/arcgis-blog/products/arcgis-enterprise/data-management/what-is-a-knowledge-graph/.

A. McCallum, K. Nigam, J. Rennie, and K. Seymore. Automating the construction of internet portals with machine learning. *Information Retrieval*, 3(2):127–163, 2000.

A. Merchant, S. Batzner, S. S. Schoenholz, M. Aykol, G. Cheon, and E. D. Cubuk. Scaling deep learning for materials discovery. *Nature*, 624:80–85, 2023. doi: 10.1038/s41586-023-06735-9. URL https://www.nature.com/articles/s41586-023-06735-9.

B. Rozemberczki et al. Pytorch geometric temporal: Spatiotemporal signal processing with neural machine learning models. https://github.com/benedekrozemberczki/pytorch_geometric_temporal, 2020. Accessed: 2025-01-18.

P. Sen, G. Namata, M. Bilgic, L. Getoor, B. Galligher, and T. Eliassi-Rad. Collective classification in network data. *AI magazine*, 29(3):93–93, 2008.

Y. Seo, M. Defferrard, P. Vandergheynst, and X. Bresson. Structured sequence modeling with graph convolutional recurrent networks. *Proceedings of the International Conference on Neural Information Processing Systems (NeurIPS)*, 2018. URL https://arxiv.org/abs/1612.07659.

V. M. A. d. O. Silva. *Urban Morphology: An Introduction to the Study of the Physical Form of Cities*. Springer, second edition edition, 2022. doi: 10.1007/978-3-030-92454-6.

TNO. Decamod: Toolbox voor rekenen aan co$_2$-reductie in transport en logistiek. Technical Report TNO-2020-R11938, TNO, 2020. URL https://publications.tno.nl/publication/34637806/N2wlcA/TNO-2020-R11938.pdf.

P. Veličković, G. Cucurull, A. Casanova, A. Romero, P. Liò, and Y. Bengio. Graph attention networks. In *Proceedings of the International Conference on Learning Representations (ICLR)*, 2018. URL https://arxiv.org/abs/1710.10903.

World Bank. Roads, total network (km). *World Development Indicators: Metadata Glossary*. Retrieved from https://databank.worldbank.org/metadataglossary/world-development-indicators/series/IS.ROD.TOTL.KM.

Z. Wu, S. Pan, F. Chen, et al. A comprehensive survey on graph neural networks. *Journal of LaTeX Class Files*, XX(XX):1–41, 2019.

Z. Wu, S. Pan, G. Long, J. Jiang, C. Chang, and S. Y. P. Zhang. A comprehensive survey on graph neural networks. *IEEE Transactions on Neural Networks and Learning Systems*, 32 (1):4–24, 2021. doi: 10.1109/TNNLS.2020.2978386. URL https://ieeexplore.ieee.org/document/9099040.

L. Xiong, X. Yuan, Z. Hu, X. Huang, and P. Huang. Gated fusion adaptive graph neural network for urban road traffic flow prediction. *Neural Processing Letters*, 56:9, 2024. doi: 10.1007/s11063-024-11479-2.

J. Xue, N. Jiang, S. Liang, et al. Quantifying spatial homogeneity of urban road networks via graph neural networks. *Nature Machine Intelligence*, 2021. URL https://arxiv.org/abs/2101.00307. Accepted for publication.

B. Yu, H. Yin, and Z. Zhu. Spatio-temporal graph convolutional networks: A deep learning framework for traffic forecasting. In *Proceedings of the Twenty-Seventh International Joint Conference on Artificial Intelligence, IJCAI-18*, pages 3634–3640. International Joint Conferences on Artificial Intelligence Organization, 7 2018. doi: 10.24963/ijcai.2018/505. URL https://doi.org/10.24963/ijcai.2018/505.

7. Appendices

A. Vehicle Types

ID	Is Freight Type	Available in Parcel Module	Description
0	1	0	Truck (small)
1	1	0	Truck (medium)
2	1	0	Truck (large)
3	1	0	Truck+trailer (small)
4	1	0	Truck+trailer (large)
5	1	0	Tractor+trailer
6	1	0	Special vehicle
7	0	1	Van
8	0	1	LEVV
9	0	0	Moped

Table 6: Vehicle Type Classification

B. NSTR Goods Classification

ID	Description
0	Agricultural products and live animals
1	Foodstuffs and animal fodder
2	Solid mineral fuels
3	Petroleum products
4	Ores and metal waste
5	Metal products
6	Crude and manufactured minerals, building materials
7	Fertilizers
8	Chemicals
9	Machinery, transport equipment, manufactured articles and miscellaneous articles
-1	Empty

Table 7: NSTR Goods Classification

C. Logistic Segment Classification

ID	Description
0	Food (general cargo)
1	Miscellaneous (general cargo)
2	Temperature controlled
3	Facility logistics
4	Construction logistics
5	Waste
6	Parcel (consolidated flows)
7	Dangerous
8	Parcel (deliveries)

Table 8: Logistic Segment Classification

D. Transport Flow Classification

ID	Is External	Description
1	0	Producer to Consumer
2	0	Producer to DC
3	0	DC to Consumer
4	0	Producer to TT
5	0	DC to DC
6	0	TT to Consumer
7	0	DC to TT
8	0	TT to DC
9	0	TT to TT
10	1	External Producer/Consumer to/from Producer/Consumer
11	1	External Producer/Consumer to/from DC
12	1	External Producer/Consumer to/from TT

Table 9: Transport Flow Classification

E. Datasets for Learning and Testing

1. **Cora, CiteSeer, and PubMed**
    - **Purpose:** These datasets are commonly used for evaluating node classification, semi-supervised learning, and understanding graph structures in citation networks, thereby strengthening foundational knowledge about Graph Neural Networks (GNNs) McCallum et al. (2000); Giles et al. (1998); Sen et al. (2008).
    - **Relevance:** They serve as benchmarks for model architectures and semi-supervised tasks involving small to medium-sized graphs.
    - **Available in:** Spektral, PyTorch Geometric, Deep Graph Library (DGL) Grattarola et al. (2023).
2. **Open Street Map (OSM) Graphs**

- **Purpose:** These datasets represent transportation networks, focusing on spatial and connectivity structures, making them ideal for road networks, urban planning, and freight transportation analysis Boeing (2017).
- **Key Features:**
  - Nodes represent intersections or waypoints.
  - Edges represent streets or paths.
  - Attributes include geographic location, type of road, and distance.
- **Relevance:** They align with studies on urban morphology and transportation systems.
- **Datasets:**
  - **OSMnx Python package:** Allows extraction of road networks for any city or region Boeing (2023).
- **Available in:** OSMnx.

3. **METR-LA and PEMS-BAY**
   - **Purpose:** These datasets provide spatio-temporal graph data, where nodes are sensors placed on roads, and edges represent the connectivity of road segments. They are ideal for testing temporal forecasting models and integrating graph data with time-series models Li et al. (2018).
   - **Key Features:**
     - **METR-LA:**
       * 207 sensors.
       * Traffic data collected over several months.
       * Time-varying attributes like speed, flow, and occupancy.
     - **PEMS-BAY:**
       * 325 sensors.
       * Similar structure to METR-LA.
   - **Relevance:** These datasets are useful for integrating dynamic data and testing how well Graph Neural Networks (GNNs) adapt to time-series forecasting tasks.
   - **Available in:** DCRNN Resources, PyTorch Geometric Temporal Rozemberczki et al. (2020).