ABSTRACT

Urban congestion remains a persistent challenge for cities, where disruptions often propagate across road networks in complex and poorly understood ways. While recent models such as the Spatio-Temporal Graph Neural Point Process (STGNPP) provide strong performance in congestion event prediction (2), they offer limited interpretability for understanding why and how congestion spreads. Conversely, Granger causality serves as a prominent approach for uncovering directional dependencies in traffic flow (6) but struggles to scale to large, high-dimensional transportation networks. This gap highlights the need for a unified framework that combines both predictive accuracy and causal interpretability.

This research proposes a Causal-Enhanced STGNPP framework that integrates congestion event prediction with scalable causal analysis. Building on the strengths of STGNPP and inspired by recent work such as the Spatio-Temporal Granger Causality–Graph Neural Network (STGC-GNN) for traffic speed prediction (11), the proposed model introduces a Granger-based causal discovery module to identify statistically significant causal pathways of congestion propagation. Using a comprehensive traffic dataset from Xuancheng (3), this study investigates whether incorporating causal structure can improve event prediction performance and provide interpretable explanations for network-wide congestion dynamics.

Given this motivation, the project aims to address the following question: *Can causal analysis be effectively integrated into spatio-temporal neural models to enhance congestion prediction and uncover interpretable propagation mechanisms?* To answer this, the study sets two quantitative objectives: (1) improving the congestion-event Neural Point Process (NPP) learning objective (evaluated as MAE_t for the inter-event time between consecutive congestion events), and improving the duration modeling accuracy of congestion events (evaluated as MAE_d for event duration); and (2) identifying statistically significant causal relationships among road segments, whose measurability and interpretation are evaluated via spatio-temporal Granger causality analysis (see Section 5.2.2). By bridging deep learning with causal inference, this work seeks to advance both methodological development and practical decision-making for real-world transportation systems.

Causal-Enhanced Spatio-Temporal Graph Neural Point Processes for Traffic Congestion Event Prediction

Yifei Chen

Advisor
Professor Arie Shaus

A thesis submitted to the Department of Mathematics
and Statistics in partial fulfillment of the requirements for
the degree of Bachelor of Arts with Honors

Department of Mathematics and Statistics
Mount Holyoke College
South Hadley, MA 01075
May 2026

ACKNOWLEDGEMENTS

This acknowledgement is a favorite part of this thesis because it gave me the freedom to express my gratitude and my soul, something the main text was too serious to allow.

As I sit down to write this on March 28, 2026, it strikes me that this is not just the end of my thesis, but also the closing chapter of four years at college. The past years have been full of moments of challenge and growth I will never forget.

First and foremost, I am deeply grateful to my advisor, Professor Arie Shaus. Your guidance, support, and encouragement carried me through every obstacle: from picking a topic to writing the paper and polishing the final draft. As someone who had never independently handled such a large project and experiments before, your help was truly indispensable. I also thank Professors Tony Liu and Dinko Hanaan Dinko for reading this thesis and offering suggestions that improved it. I am especially grateful to Professors Timothy Chumley and Sohail Hashmi for encouraging me to write this thesis and for their steady support along the way. I thank Professor Alanna Hoyer-Leitzel for her valuable feedback on the manuscript. I also thank Jeanne Kenney, Academic Department Coordinator for Data Science, for the behind-the-scenes logistical support that made my defense and much else possible; and the staff of LITS, including those who maintain JupyterHub and Unity, for keeping the computational resources I relied on available.

To my family and friends, thank you for listening to me through every high and low, for your patience when I felt overwhelmed, and for your constant encouragement and support. Your love and presence carried me through the hardest moments and made this journey brighter, warmer, and more meaningful.

I would also like to thank my favorite singer, Eason Chan. Over these four years, music has been my companion through every swampy patch with both comfort and inspiration. I still remember your words at the concert on August 9, 2024, about remembering the joy and emotion of today for the days to come—and indeed, I do.

And finally, I want to thank myself, an ordinary soul, often fragile but resilient. The past four years were not easy, but I never gave up on any chance to grow, always getting back up after falling. If time could be divided into moments, I would express my sincere gratitude to each of my past selves for helping me grow into someone more mature and resilient.

In about ten days, five or six magnolia trees will bloom in purple and pink. I like to think of that as their farewell gift to me. I am about to leave campus, carrying with me the care of my professors, the friendship of my classmates, the company of every squirrel and bird, and the memories of George. I am heading to the colder Evanston, so I should remember to pack enough winter clothes.

Contents

List of Figures

List of Tables

GLOSSARY OF KEY TERMS AND ACRONYMS

Models	
STGNPP	Spatio-Temporal Graph Neural Point Process. A neural point process model augmented with graph-based message passing to capture spatio-temporal dependencies for congestion event prediction (2).
STGC-GNN	Spatio-Temporal Granger Causality Graph Neural Network. A traffic forecasting framework that constructs a spatio-temporal Granger causality graph and uses it within a graph neural network pipeline (11).
STGC-GNPP	Spatio-Temporal Granger Causality Graph Neural Point Process. The proposed integration of STGNPP congestion event modeling with a STGC graph(this thesis).
Graph inputs (adjacency / baselines)	
Binary connection matrix	A directed one-hop connectivity matrix derived from road network topology. It indicates physically connected upstream/downstream road segments.
Adaptive adjacency matrix	A learnable adjacency matrix in STGNPP that is computed to capture spatial dependencies.
SD graph	Spatial-distance graph whose edges/weights are constructed from geographical proximity between road segments.
Identity graph	A graph where each node is connected only to itself.
Random graph	A graph with randomized connectivity, typically matched in sparsity/degree to a target graph, used as a structural baseline.
Causal graphs (STGC variants)	
STGC	Spatio-Temporal Granger Causality. A Granger-causality-based approach for constructing a directed spatio-temporal dependency graph among road segments.
STGC-A	Speed-based STGC graph constructed using Granger causality tests on continuous speed time series.

STGC-B	Congestion-based STGC graph constructed using logistic Granger causality tests on binary congestion state time series.
Losses & metrics	
NPP loss L_{npp}	Negative log-likelihood-style loss for inter-event times, $L_{\text{npp}} = \sum_i \left(-\log \lambda(\tau_i \mid h_i) + \Lambda(\tau_i \mid h_i)\right)$.
Duration loss L_{duration}	Log-smooth duration loss, $L_{\text{duration}} = \frac{1}{N} \sum_i \log\left(1 + \left\lvert \hat{d}_i - d_i \right\rvert\right)$.
MAE$_d$	Mean Absolute Error of predicted congestion duration (in minutes).
MAE$_t$	Mean Absolute Error of predicted inter-event time (in minutes).

CHAPTER 1 INTRODUCTION

Traffic congestion has become one of the most pressing urban challenges of the 21st century. According to the Texas A&M Transportation Institute's 2021 Urban Mobility Report, traffic congestion causes more than $190 billion in economic losses annually in the United States, and the average commuter spends approximately 54 hours per year delayed in traffic (7). In China, congestion-related losses account for nearly 20% of per-capita disposable income, or approximately 5% of the national GDP (12; 13). With global urban populations projected to reach 68% of the world's population by 2050 (10), the severity and frequency of congestion are expected to further increase. Beyond economic losses, traffic congestion also leads to increased air pollution and reduced public transport reliability. These impacts make accurate congestion prediction not only a technical problem but also a critical societal and economic necessity.

However, for many years, congestion prediction, and traffic prediction in general, remains a challenging problem. This difficulty stems from the highly dynamic and multi-layered nature of urban traffic systems. Congestion is not caused by a single factor but emerges from the interaction of traffic demand, road network topology, traffic control strategies, and external influences such as weather conditions and accidents. These factors evolve over time and interact across space, resulting in nonlinear and spatially heterogeneous dynamics. In real traffic networks, congestion on one road segment can propagate to distant segments through a series of interactions, and such propagation

patterns vary over time. Therefore, congestion should be viewed as a spatio-temporal propagation process rather than a local or static phenomenon. A reliable congestion prediction framework must be capable of modeling both temporal evolution and spatial dependency simultaneously.

With the rapid deployment of intelligent transportation systems (ITS) and large-scale sensing infrastructure (16), massive volumes of traffic data have become available, which makes data-driven traffic modeling increasingly feasible. Existing traffic prediction studies can be broadly categorized into two directions (17): continuous traffic state prediction and discrete event-based prediction. Continuous prediction focuses on variables such as traffic speed, flow, or density, while discrete prediction models congestion or incidents as events.

Early research in traffic prediction mainly relied on statistical methods such as autoregressive integrated moving average (ARIMA), vector autoregression (VAR), and cellular automata (4; 18). These methods are computationally efficient and interpretable, but they generally assume linear relationships and stationary temporal patterns, which limits their ability to model complex traffic dynamics. Subsequently, machine learning techniques such as regression models, support vector regression (SVR), and ensemble methods (e.g., XGBoost) were introduced to capture nonlinear relationships (8). Although these approaches improve prediction accuracy, they still struggle to model spatial dependencies among road segments and fail to capture the structural characteristics of road networks.

With the advancement of deep learning, spatio-temporal modeling has become a major research direction in traffic prediction. Early neural-network-based studies had already demonstrated the potential of data-driven traffic prediction (5), and recurrent architectures such as gated recurrent units (GRU) and long short-term memory (LSTM) networks later became widely used for temporal traffic forecasting (14). Convolutional

neural networks (CNNs) have also been applied to traffic data by mapping road networks to grid structures, which enables the extraction of local spatial patterns. However, CNN-based methods are limited by their reliance on regular grid representations, which do not naturally align with real-world road networks. To address this limitation, graph neural networks (GNNs) have been introduced to model traffic systems as graphs, where nodes represent road segments and edges represent their connectivity. Building on this idea, spatio-temporal graph neural networks (ST-GNNs) have emerged as a family of models that combine graph convolutions with temporal modeling modules such as recurrent neural networks or attention mechanisms, and are widely regarded as a state-of-the-art paradigm for traffic forecasting (1). A representative example is the spatio-temporal graph convolutional network (STGCN) and its variants (9).

Despite these advances, most existing spatio-temporal graph-based models rely on static graph structures. Such graphs are typically constructed based on physical adjacency or distance and remain unchanged during training. However, real traffic systems exhibit dynamic and asymmetric interactions. The mutual influence between two road segments may vary over time and is often directional, depending on traffic flow patterns and singal control. Fixed graph structures are therefore insufficient to capture these dynamic spatial dependencies. This limits both prediction accuracy and interpretability.

In parallel, congestion prediction has also been studied from an event-centric perspective. Early approaches treated congestion as a classification or regression problem by discretizing time into fixed intervals (e.g., 5 or 10 minutes) and labeling congestion states accordingly (19). While such formulations are convenient, they introduced artificial temporal boundaries that do not align with the continuous nature of congestion. In reality, congestion onset and dissipation occur at irregular time points, and discretization inevitably introduces temporal distortion and information loss.

To overcome this limitation, neural point process (NPP) models have recently been introduced to traffic modeling (20). NPPs treat congestion as a continuous-time stochastic event process and learn an intensity function that characterizes the probability of congestion occurrence over time. This formulation naturally captures irregular temporal patterns and allows the modeling event duration and temporal dependency without relying on fixed time intervals. Among these approaches, the Spatial-Temporal Graph Neural Point Process (STGNPP) is an important advancement by integrating graph neural networks with neural point processes (2). In this way, this model allows simultaneous modeling of temporal event dynamics and spatial interactions within traffic networks.

Nevertheless, existing STGNPP-based and graph-based models primarily focus on learning statistical correlations rather than explicit causal relationships. In real traffic systems, congestion propagation exhibits strong causality and directionality (21). For example, congestion on upstream road segments often leads to downstream congestion after a certain time delay, while the reverse influence is usually much weaker. Such asymmetric and time-lagged dependencies reflect underlying causal mechanisms rather than simple correlations. Models based on undirected or static graphs are unable to fully capture these causal structures, which limits their effectiveness in long-term prediction and interpretability.

Recent studies in traffic speed prediction have shown that incorporating Granger causality into graph construction can significantly improve model performance by explicitly modeling directional and temporal dependencies of speed among road segments (11). Inspired by these findings, this thesis extends the idea of causality-aware modeling to congestion prediction.

Specifically, this work proposes a novel spatio-temporal congestion prediction framework that integrates neural point processes with Granger causality-based graph

learning. A spatio-temporal Granger causality graph is constructed to capture stable and directional causal relationships among road segments. This causal graph is then incorporated into a neural point process framework to model the continuous-time evolution of congestion events. By combining causal spatial modeling with event-based temporal modeling, the proposed framework is able to capture both the dynamic propagation of congestion and its underlying causal mechanisms.

The main contributions of this work are summarized as follows. First, a novel causal spatio-temporal congestion prediction framework is proposed, which integrates neural point processes with Granger causality-based graph learning. Second, the proposed method explicitly models directional and time-dependent congestion propagation, improving the interpretability of prediction. Third, this work provides a careful reimplementation and comparison of STGNPP-style baselines under different graph inputs. Fourth, experiments on real-world datasets show that, although the prediction accuracy is comparable to that of baseline models, the proposed approach provides new insights and directions for future improvement.

The remainder of this thesis is organized as follows. Section 2 reviews related work on traffic prediction and congestion modeling. Section 3 presents the motivation of this study. Section 4 describes the dataset and the corresponding exploratory data analysis. Section 5 introduces the proposed methodology and experimental setup. Section 6 presents and analyzes the experimental results. Section 7 discusses the challenges associated with incorporating Granger causality into congestion prediction, as well as potential directions for improvement. Finally, Section 8 concludes the thesis and outlines future research directions.

CHAPTER 2 LITERATURE REVIEW

This section reviews existing studies related to traffic prediction and congestion modeling. We first provide an overview of the evolution of traffic prediction methods for both continuous traffic state prediction and congestion event prediction. We then review representative spatio-temporal modeling approaches, including Spatio-Temporal Graph Convolutional Networks (STGCN) (9) and Neural Point-Process-based models. Finally, we introduce Granger causality and recent efforts to incorporate causal structures into graph-based traffic prediction frameworks, such as STGC-GNN (11), which motivates the proposed method in this work.

2.1 Traffic Prediction Tasks and Modeling Perspectives

2.1.1 Traffic Prediction Tasks

Traffic prediction aims to forecast future traffic conditions using historical transportation data. Depending on the prediction target, existing studies can generally be categorized into two groups (17):

(1) prediction of continuous traffic variables, such as traffic speed, flow, and

density, which are typically measured continuously over time. This has historically been the primary focus of traffic prediction research.

(2) prediction of discrete traffic events, particularly traffic congestion. In early studies, congestion was often treated similarly to continuous traffic variables by discretizing time into fixed intervals and labeling each interval as congested or uncongested (19). Thus it formulates a binary or multi-class classification problem. While straightforward, this approach introduces a critical limitation: artificial time discretization can obscure the true timing of congestion events, and the inherently event-driven nature of congestion is not faithfully represented. It could potentially distort temporal dependencies and causal relationships.

All in all, these two categories correspond to fundamentally different modeling objectives and have led to distinct lines of methodological development in the literature.

2.1.2 Modeling Paradigms for Traffic Predictions

Over time, traffic prediction methods have evolved from traditional statistical models to advanced neural network architectures (4; 1). Early approaches, such as autoregressive integrated moving average (ARIMA), capture temporal dependencies under linear and stationary assumptions. Classical machine learning models, including regression-based methods and support vector machines, extend this capability by modeling nonlinear relationships, but they rely heavily on handcrafted features (8). More recently, deep learning models, such as recurrent neural networks (RNNs), long short-term memory networks (LSTMs), temporal convolutional networks (TCNs), and attention-based architectures, have demonstrated strong performance in capturing complex and long-range temporal patterns (5; 14; 1).

2.1.3 Temporal and Spatial Dependencies in Traffic Prediction

No matter how prediction models evolve, whether they are designed for continuous traffic variables or discrete congestion events, two fundamental issues remain unchanged: the modeling of temporal dependency and the modeling of spatial dependency. Essentially, all traffic prediction models aim to improve the representation of these two aspects, as they jointly determine the dynamics of traffic systems.

From the temporal perspective, the key challenge lies in capturing both short-term dynamics and long-range dependencies in traffic evolution. Traffic patterns exhibit strong periodicity, such as daily and weekly cycles, as well as irregular fluctuations caused by external factors like weather and incidents. As illustrated in Fig. 2.1, the traffic states of road segments evolve over time and influence one another. From time step t_0 to $t_0 + t_1 + t_2 + t_3$, for example, road segment 5 is influenced not only by its own historical states but also by the states of other road segments. Traditional time-series models typically assume linear and stationary processes, which limits their ability to capture complex traffic dynamics. Classical machine learning models can model nonlinear relationships to some extent but still rely heavily on handcrafted features and fixed time windows. In contrast, deep learning models have shown strong capability in capturing long-range temporal dependencies and nonlinear temporal patterns (4; 14; 1).

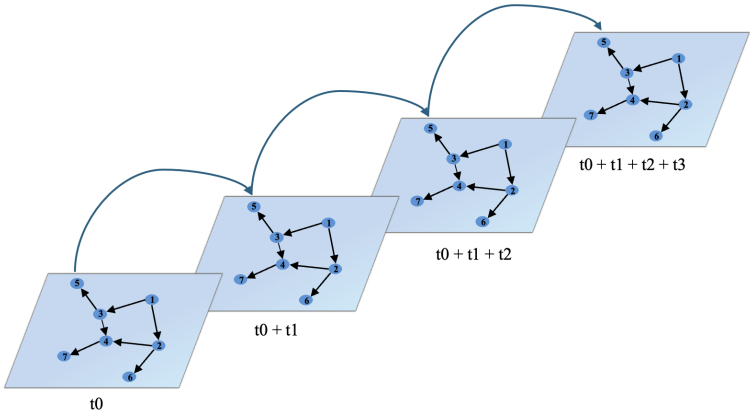


Figure 2.1: Illustration of temporal dependencies in traffic networks.

From the spatial perspective, spatial dependency modeling aims to describe how traffic conditions propagate across different locations in a road network. Unlike grid-structured data such as images, road networks exhibit irregular and heterogeneous topologies, where nodes and edges represent road segments and their connectivity. As shown in Fig. 2.2, the color of each node represents the congestion level of a road segment. At time step t_0, road segment 1 exhibits the highest level of congestion, while its directly connected neighboring segments (segments 2–5) show moderate congestion levels. At the next time step t_1, congestion propagates from segment 1 to its neighboring segments, causing segments 2–5 to become highly congested. Meanwhile, second-hop segments, specifically segments 6 and 7 experience moderate congestion. This example illustrates how congestion propagates across the network in a spatially dependent manner. Traffic states on one road segment can influence others through various factors, and such influence is often directional and time-dependent.

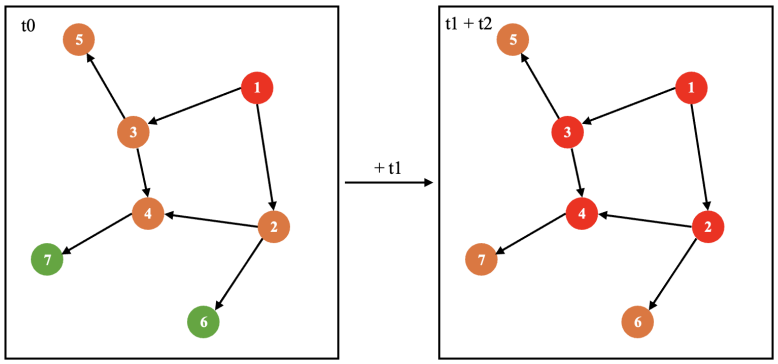


Figure 2.2: Illustration of spatial congestion propagation in a road network. Node colors represent congestion levels of road segments.

Early spatial modeling approaches typically relied on distance-based correlations or manually defined adjacency matrices, which fail to capture complex interactions in real traffic systems. In contrast, graph-based models, particularly graph neural networks (GNNs), have become the dominant paradigm for spatial modeling in traffic prediction (9; 1). These models represent road networks as graphs and learn spatial dependencies through graph convolutions or message-passing mechanisms, thereby effectively capturing both local and global spatial interactions.

2.2 Spatio-Temporal Graph-Based Traffic Prediction Models

2.2.1 Spatio-Temporal Graph Convolutional Networks (STGCN)

A landmark contribution in the domain of continuous traffic feature prediction is the Spatio-Temporal Graph Convolutional Network (STGCN) by Yu et al. (9), which was the first to employ purely convolutional architectures to capture spatio-temporal dependencies from graph-structured time series in traffic studies. STGCN represented

a significant advancement in traffic prediction by addressing the fundamental challenge of modeling spatial and temporal correlations simultaneously.

As described in (9):

$$\hat{v}_{t+1}, \ldots, \hat{v}_{t+H} = \arg \max_{v_{t+1},\ldots,v_{t+H}} \log P(v_{t+1}, \ldots, v_{t+H} \mid v_{t-M+1}, \ldots, v_t)$$

Here, v_t denotes the traffic state at time step t (e.g., velocity/speed on the road-network graph), and $\hat{v}_{t+1}, \ldots, \hat{v}_{t+H}$ are the model's predictions over H steps. Given the past M observations $v_{t-M+1}, \ldots, v_t$, the right-hand side selects future values $v_{t+1}, \ldots, v_{t+H}$ that maximize the conditional log-probability, i.e., a maximum-likelihood prediction under the model.

the model aims to predict future traffic states that maximize their conditional probability given past observations. The architecture consists of two ST-Conv blocks and a fully connected output layer. Each ST-Conv block contains three sequential components: a Temporal Gated Convolution, a Spatial Graph Convolution, and another Temporal Gated Convolution. This design allows the model to first extract short-term temporal patterns, then encode spatial dependencies across the network, and finally refine temporal dynamics. By fusing the outputs of these ST-Conv blocks, STGCN effectively approximates the conditional probability distribution of future traffic states.

However, STGCN is mainly focused on predicting continuous traffic values like speed or flow. This limitation makes it less flexible for complex traffic phenomena such as congestion propagation between traffic clusters because it struggles with sparse, irregularly distributed congestion events (28). The next subsection therefore introduces STGNPP, which can handle the sparse, irregular nature of event data and still capture the complex spatial and temporal dependencies inherent in traffic networks.

2.2.2 Spatio-Temporal Graph Neural Point Process

Unlike STGCN, which focuses on predicting continuous traffic values, STGNPP is designed to model sparse and irregularly distributed traffic congestion events (2). It models congestion as a continuous-time stochastic process via the conditional intensity function

$$\lambda(t \mid \mathcal{H}_t) = \lim_{\Delta t \to 0} \frac{P(\text{one event occurs in } [t, t + \Delta t) \mid \mathcal{H}_t)}{\Delta t},$$

which represents the instantaneous probability of a congestion event occurring at time t given the historical traffic state $\mathcal{H}_t$.

STGNPP consists of three main modules (2). First, a spatio-temporal graph learning module learns latent representations that encode both spatial and temporal structure: transformer layers with causal attention capture temporal dependencies per road, and GCN layers with mix-hop aggregation capture multi-hop spatial interactions across the network. Second, a congestion event learning module handles event sequences via a hybrid design that combines continuous state evolution between events (neural flow) with discrete updates when events occur (GRU cells). Third, an intensity function network models the conditional intensity of congestion in continuous time and supports event duration prediction. Despite its effectiveness, STGNPP relies on predefined adjacency graph structures and does not explicitly model causal relationships between road segments, which may limit its ability to capture realistic traffic propagation patterns.

Figure 2.3 summarizes the overall architecture and major components of STGNPP.

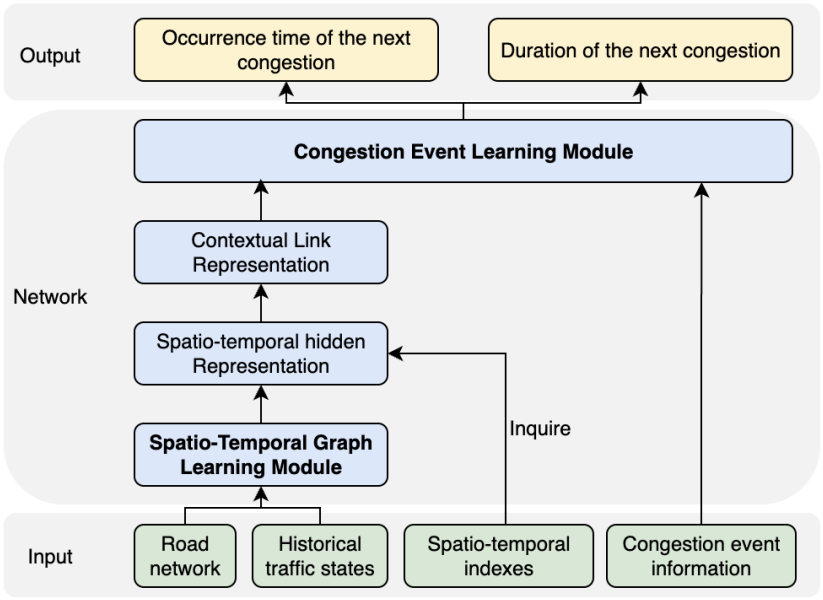


Figure 2.3: Overview of the Spatio-Temporal Graph Neural Point Process (STGNPP) architecture.

While STGNPP is highly effective for sparse, non-uniform events where STGCN struggles, the model is more complex and less interpretable. It operates like a black-box system that provides limited insight into why events occur or how they causally propagate through networks. This lack of interpretability is a significant limitation for practical applications where understanding the underlying mechanisms of congestion formation is crucial for developing effective intervention strategies.

2.3 Granger Causality and Causal Discovery

Parallel to these developments, Granger causality has become a widely used statistical framework for uncovering directional causal relationships in time series data (6). The key idea is that past values of one variable may contain predictive information about another variable beyond what is captured by its own history. In its classical

formulation, Granger causality is typically implemented through vector autoregressive (VAR) models, where nested model comparisons are used to assess whether including lagged values of a candidate predictor significantly improves forecasting performance.

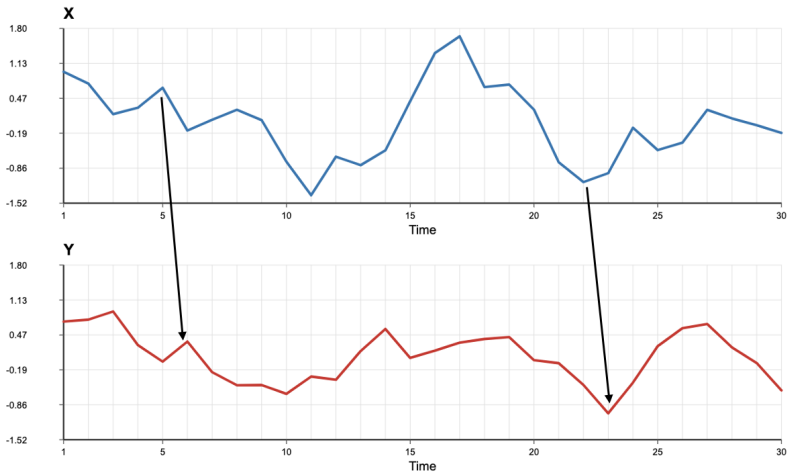


Figure 2.4: Illustration of time-lagged influence between two objects. Changes in object X propagate over time and affect the state of object Y with a temporal delay.

Over the past decades, the framework has undergone substantial methodological expansion. Recent research has extended Granger causality to high-dimensional systems, network-structured dependencies, nonlinear dynamics, and discrete-valued processes. Comprehensive reviews of these developments show both the versatility of the framework and its practical limitations, such as assumptions of linearity and stationarity (22). These extensions have significantly broadened the applicability of Granger-based methods across disciplines, including economics, neuroscience, climate science, and transportation systems.

In the context of traffic studies, this allows researchers to identify how events in one location propagate through space and time to influence conditions elsewhere. Typically, the method involves comparing models that include or exclude potential predictor variables to evaluate whether adding lagged information improves predictive

accuracy, which indicates a directional influence. To be more specific, this kind of Granger causality can be extended to spatio-temporal settings and combined with traffic speed prediction models (11).

2.3.1 GNN-based Traffic Prediction Framework with a Spatial-Temporal Granger Causality Graph

STGCN performs well in predicting traffic volume and speed (9), yet it does not capture the global and dynamic spatial dependencies in road networks necessary for long-term forecasting. A major challenge is detecting causal structures that evolve over time (11).

To address this, He et al. (2023) propose STGC-GNN, a Spatio-Temporal Granger Causality Graph Neural Network framework that integrates spatio-temporal graph learning with Granger-causality-based graph construction (11) (see Fig. 2.5). Specifically, the model first constructs a Spatio-Temporal Granger Causality (STGC) graph, where directed edges represent statistically significant causal influences between nodes, determined via time-series Granger tests.

This data-driven causal graph replaces traditional distance-based or adaptive adjacency matrices in GNNs, which allows the model to encode causal dependencies instead of fixed spatial proximity. Experiments on the METR-LA dataset demonstrate that the STGC graph captures more interpretable long-term dependencies, which is particularly effective for long-horizon forecasting and maintains short-term accuracy comparable to distance-based graphs (11).

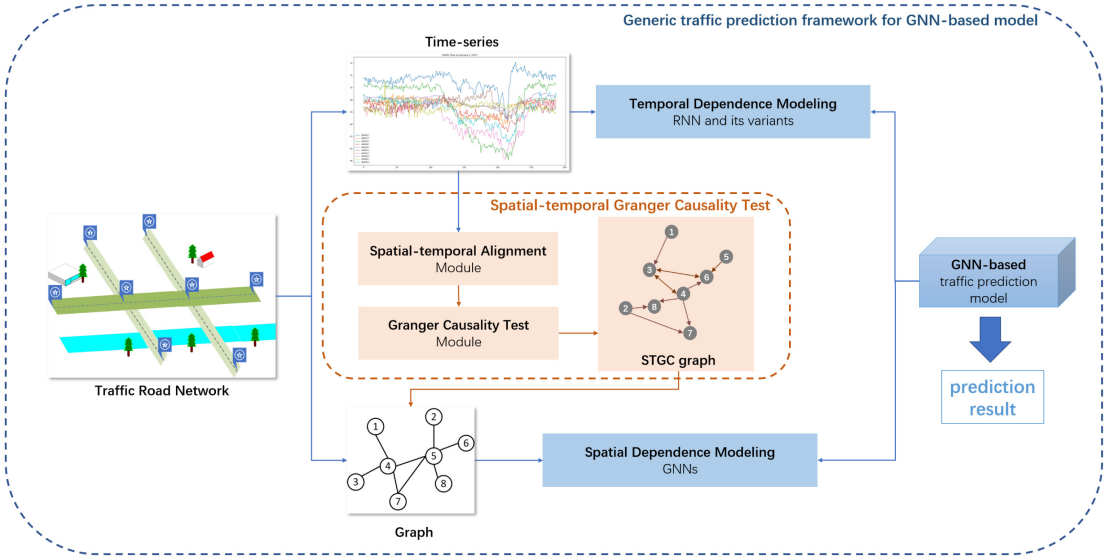


Figure 2.5: Overview of the STGC-GNN framework.

Despite these advances, existing STGC-based approaches have primarily been applied to traffic speed prediction and remain relatively underexplored in broader traffic information modeling, particularly in traffic congestion analysis. Congestion propagation exhibits complex dynamics. In particular, traffic flow theory suggests the presence of nonlinear threshold effects, where small perturbations do not trigger congestion unless a critical threshold is reached. Moreover, congestion formation involves regime shifts between different traffic states, such as transitions from free flow to synchronized flow, often driven by stochastic, nucleation-like mechanisms(15). These characteristics indicate that congestion dynamics are inherently non-smooth and event-driven, thereby requiring more targeted causal modeling strategies.

CHAPTER 3 MOTIVATION

The above review suggests that existing traffic prediction studies have made progress in modeling spatio-temporal dependencies, especially through graph neural networks and neural point processes. In particular, STGNPP is well suited for congestion event prediction because it models congestion as an event-driven process in continuous time rather than as a fixed-interval classification problem. At the same time, recent Granger-causality-based studies have shown that directional and time-lagged dependencies can provide useful information beyond static or distance-based graph structures.

Recent studies have begun exploring the integration of causal analysis with spatio-temporal graph learning. For example, the STGC-GNN model combines STGCN with Granger causality analysis to capture causal relationships in traffic dynamics. This achieved improved prediction accuracy compared with using STGCN alone (11). These studies suggest that incorporating causal discovery into spatio-temporal graph models may provide both improved predictive performance and greater interpretability.

Motivated by these insights, this study develops a unified framework for congestion prediction that jointly learns spatio-temporal representations and causal structures in an end-to-end manner. Specifically, the classical Granger causality framework is integrated with the STGNPP architecture to combine causal discovery with graph-based representation learning within a single model. Unlike existing approaches that construct

spatio-temporal graphs based on the statistical significance of traffic speed transmission, the proposed framework modifies the graph construction process to detect the statistical significance of congestion propagation. In this framework, directed edges in the causal graph are established according to whether historical congestion states at one node significantly improve the prediction of congestion states at another node.

CHAPTER 4 DATASET

4.1 Data Description

Our study is based on comprehensive traffic data from Xuancheng, a prefecture-level city in Anhui Province, China, collected through automatic vehicle identification (AVI) systems (3). Xuancheng has a permanent resident population of 2.50 million (29). These AVI systems have relatively high coverage across the road network, particularly at road intersections, where they record vehicle information and timestamps. The dataset covers the entire road network of Xuancheng, consisting of 1,156 road segments, during September 2020, with traffic flow aggregated at five-minute intervals.

The dataset contains two major components: a graph-structure layer and a temporal data layer, with a total of 5,022,132 rows in the raw form. It is constructed from resampled and simulated data derived from real traffic patterns reconstructed from single-vehicle trajectories. This design ensures privacy while maintaining the information necessary for traffic analysis. Because the dataset is generated from processed trajectory-based records rather than directly from raw sensor observations, it contains no missing values.

Only 776 out of the 1,156 road segments contain valid traffic sensor data; the remaining segments were excluded from analysis. For model training, we apply two preprocessing steps. First, road segments with multiple records at the same timestamp

(due to different turning directions at the same endpoint) are aggregated by taking the mean value across all directions, since it is difficult to attribute each direction to a specific downstream segment. This results in approximately 1.58 million rows of temporal observations. Second, for computational efficiency, we select a connected subgraph of 200 road segments from the 776 available roads using a Breadth-First Search (BFS) strategy. The full 30-day dataset is used for training and evaluation on this subnetwork.

There is no single open, universal operational definition of a binary congestion state across studies, and the Xuancheng release does not have an official congestion label in the materials we used (3). We therefore adopt the following definition. A road segment is considered congested in a given 5-minute time slot when its average speed falls below 40% of the free-flow speed for that segment, where free-flow speed is defined as the 95th percentile of observed speeds over the dataset. A congestion *event* is a period of at least 2 consecutive 5-minute slots (10 minutes) during which the segment remains congested; shorter periods are discarded. Under this definition, a small share of events last over 300 minutes (0.41% among 200 road segments, 0.69% among 720). Because approximately 94% of events last only 5–15 minutes, these long events may distort predictions. We cap all durations at 300 minutes to limit their impact.

The dataset does not distinguish recurrent congestion (predictable, capacity-binding patterns tied to stable demand) from non-recurrent congestion (shocks such as incidents or adverse weather). We therefore do not assign events to these categories in supervision. In modeling, the STGNPP backbone we adopt still modulates the event intensity with calendar context (notably time of day and day of week (2)), which allows the model to learn differences in event rates across peak versus off-peak periods. But should not be read as explicit identification of recurrent versus non-recurrent mechanisms without the missing auxiliary data.

Overall, these two layers together provide the spatio-temporal input for the prediction model: the graph-structure layer captures topological relationships between road segments (connectivity, lane configurations, physical distances); the temporal data layer contains detailed traffic measurements (vehicle counts, speed distributions, turning behaviors) that support the analysis of traffic dynamics.

A more detailed description of the dataset variables is provided in Appendix A.1.

4.2 Exploratory Data Analysis

In this section, we conduct exploratory data analysis to obtain a preliminary understanding of the dataset. Figure 4.1 displays the road network structure of Xuancheng.

Figure 4.1: Road network structure of Xuancheng.

Next, we present descriptive statistics for several traffic-related variables. Figures 4.2 and 4.3 show the distribution of average traffic flow by hour and by day of the week, which highlights clear morning and evening peaks as well as the higher

traffic volumes observed on Thursdays. To our knowledge, there is little peer-reviewed evidence showing day-of-week traffic periodicity specifically for Xuancheng or Anhui province road networks. In our sample, the Thursday-versus-other-weekday differences are small in magnitude, based on one month of aggregated indicators and a resampled time grid. So we treat any weekday ranking as descriptive and potentially sensitive to sampling noise rather than evidence of distinct geo-cultural mechanism analogous to Thursday peaks in settings where Thursday marks the end of the work week. Figures 4.4 and 4.5 report the corresponding average speeds, which drop during peak periods and remain lowest on Thursdays, mirroring the flow patterns. Finally, Figures 4.6 and 4.7 examine congestion duration / traffic flow versus *relative speed*. We adopt relative speed (speed normalized by each road segment's free-flow speed) because road segments have different speed limits; using relative speed helps reduce the confounding effect of heterogeneous speed thresholds across roads and makes comparisons across segments more meaningful.

For Figure 4.6, the relationship between congestion duration and relative speed is not visually strong: relative speed values appear fairly evenly spread across different congestion durations. One mild pattern is that very short congestion events (around 5–10 minutes) tend to exhibit relatively higher relative speed. For Figure 4.7, relative speed is roughly symmetric with respect to traffic flow and concentrates toward a mid-range band; the higher-flow observations tend to occur around the median relative-speed region rather than at extreme low or high relative speeds. Overall, these observations are somewhat different from the intuitive expectation that longer congestion events would coincide with lower relative speeds and that higher traffic flow would generally correspond to lower speed.

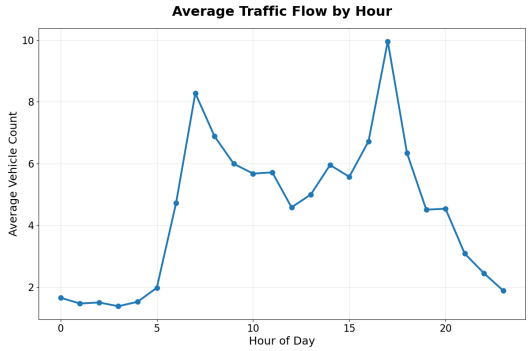


Figure 4.2: Average traffic flow by hour of day.

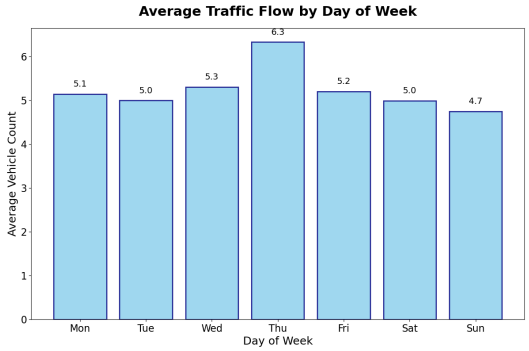


Figure 4.3: Average traffic flow by day of week.

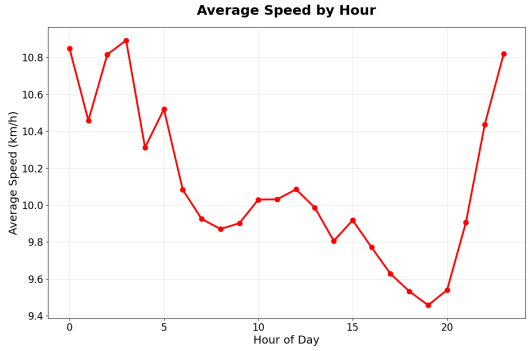


Figure 4.4: Average vehicle speed by hour of day.

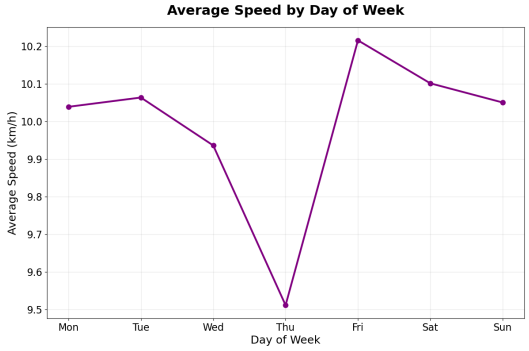


Figure 4.5: Average vehicle speed by day of week.

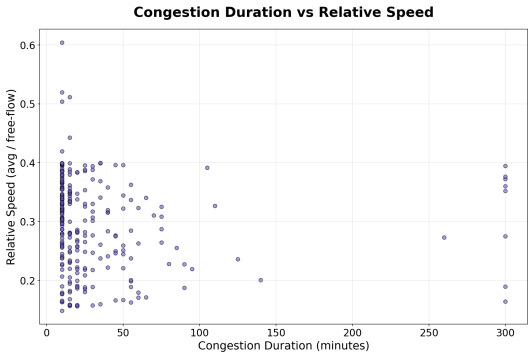


Figure 4.6: Congestion duration versus relative speed.

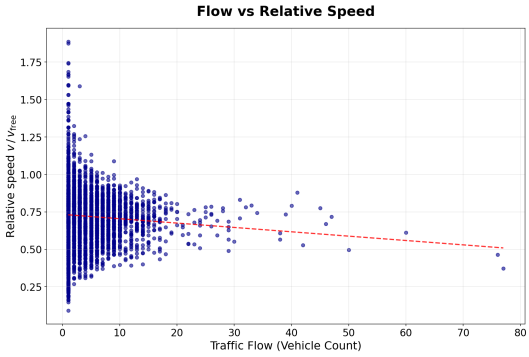


Figure 4.7: Traffic flow versus relative speed.

CHAPTER 5 METHODOLOGY

5.1 Overview

As discussed in the literature review, STGNPP is an effective model for predicting traffic congestion events using neural point processes (2), while Granger causality provides a statistical framework for identifying directional relationships between time series (6).

Building upon these two approaches, this study proposes the STGC-GNPP model (Spatio-Temporal Granger Causality Graph Neural Point Process), which integrates causal discovery with neural point process-based congestion prediction. Using the Xuancheng traffic dataset (3), the proposed framework incorporates a spatio-temporal Granger causality graph into the STGNPP architecture to capture congestion propagation patterns across the road network.

This section first introduces the overall framework of the STGC-GNPP model. We then describe the proposed causal graph construction method based on congestion propagation, followed by the experimental setup and evaluation metrics used to assess each model's performance.

5.2 Main approach

5.2.1 STGC-GNPP Framework

Our project adapts STGNPP (Section 2.2.2) and augments it with a causality graph. The introduction of stgc graph complements the traditional static spatial dependence with better causality relationship. We first re-create the original STGNPP pipeline following (2) so that subsequent comparisons isolate the impact of the causal prior rather than architectural drift. The reproduction covers the Spatio-Temporal Graph Learning Module, Congestion Event Learning Module, and Intensity Function Network.

To inject interpretability, we borrow the idea of spatio-temporal Granger causality from STGC-GNN (11) and use it as the STGC matrix, a full replacement for the adaptive adjacency.

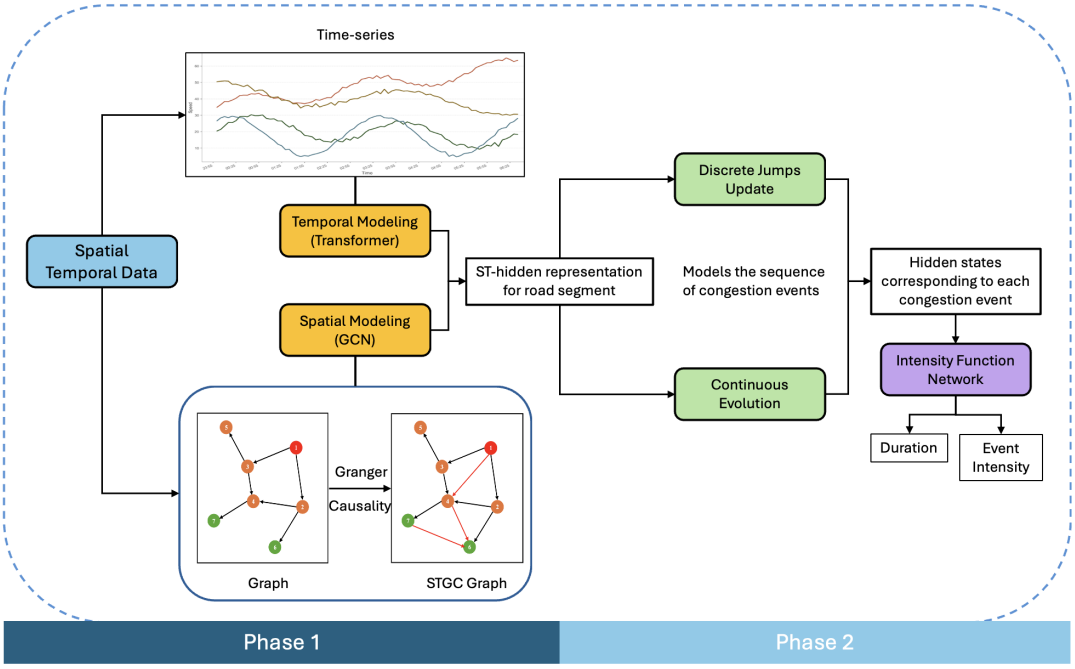


Figure 5.1: Overall framework of STGC-GNPP. Phase 1: Granger causal discovery to produce the STGC matrix. Phase 2: the STGC matrix augments the spatial graph used in STGNPP, which consists of the Spatio-Temporal Graph Learning, Congestion Event Learning, and Intensity Function Modules.

From a high-level perspective, a traffic network can be represented by two key components: the temporal evolution of traffic states and the spatial structure of the road network. Accordingly, the STGNPP framework models temporal and spatial dependencies separately (2). Temporal dependencies are captured through transformer layers with causal attention (modeling the evolution of traffic states over time) and the event modeling mechanism of the neural point process (learning the temporal patterns of congestion events). Spatial dependencies are modeled through the graph-based representation of the road network, where information is propagated between connected road segments via graph convolutional networks. The three modules of STGNPP are outlined in Section 2.2.2 (2)

Based on these spatio-temporal interactions, the framework further estimates

the Granger causality relationships between road segments and constructs a causal adjacency matrix. This causality matrix is then integrated into the model, which augments the spatial graph used for information propagation, so that the model can incorporate causal propagation patterns when predicting the occurrence and duration of congestion events. Figure 5.1 summarizes the overall pipeline of the proposed model.

5.2.2 Spatio-Temporal Granger Causality

To capture directional dependencies in traffic networks, this study adopts the spatio-temporal Granger causality (STGC) framework proposed in STGC-GNN (11). The framework is based on the transmitting causal relationship (TCR) assumption, which states that traffic information propagates from upstream road segments to downstream segments with an approximately stable velocity. Under this assumption, long-term traffic observations can be used to identify causal transmission patterns across the road network.

In practice, the STGC construction procedure consists of two main steps (11). First, a spatio-temporal alignment process is applied to estimate the transmission time between a potential cause node and a target node. This step identifies the temporal lag corresponding to the propagation of global dynamic traffic information (GDTI) from the upstream segment to the downstream segment. The time series of the two road segments are then aligned according to this estimated lag.

Second, a Granger causality test is applied to determine whether the upstream traffic information significantly improves the prediction of the downstream node (6; 11). For each pair of road segments, two autoregressive models are constructed: one that predicts the target node using only its historical observations, and another that additionally incorporates the lagged observations from the potential cause node. If

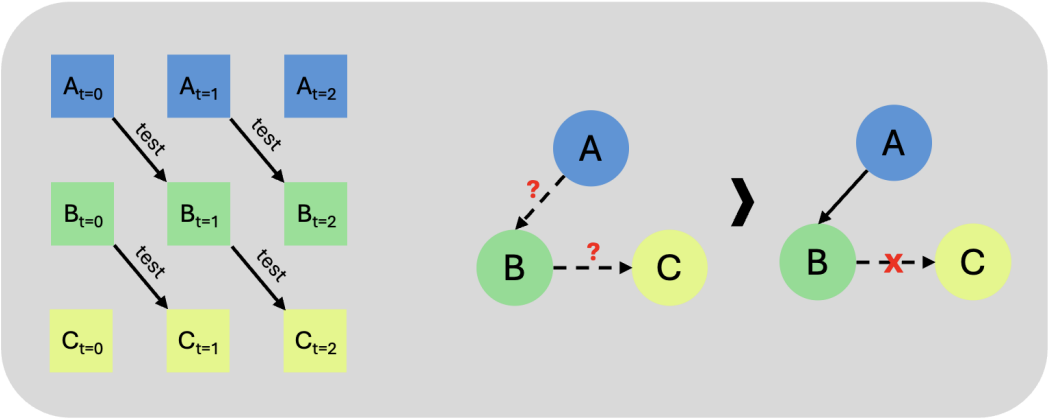


Figure 5.2: Schematic of spatio-temporal Granger causality testing with nodes A, B, and C. Causal links A→B and B→C are evaluated under the TCR assumption.

the inclusion of the upstream information significantly reduces the prediction error, the upstream segment is considered to Granger-cause the downstream segment.

Figure 5.2 and Figure 5.3 illustrate this approach with nodes A, B, and C, focusing on the causal links A→B and B→C. As shown in Figure 5.3, for the A→B case, A is the cause and B is the effect. To test whether A helps predict B, we use two input windows: A's history spans lags from $\tau + p$ down to p, where τ is the spatial–temporal lag and p is the number of past time steps; B's history uses its own past p time steps without an extra lag. An F-test is then applied to test whether the coefficients on A's lagged observations are jointly zero. Rejecting this null hypothesis indicates that A Granger-causes B.

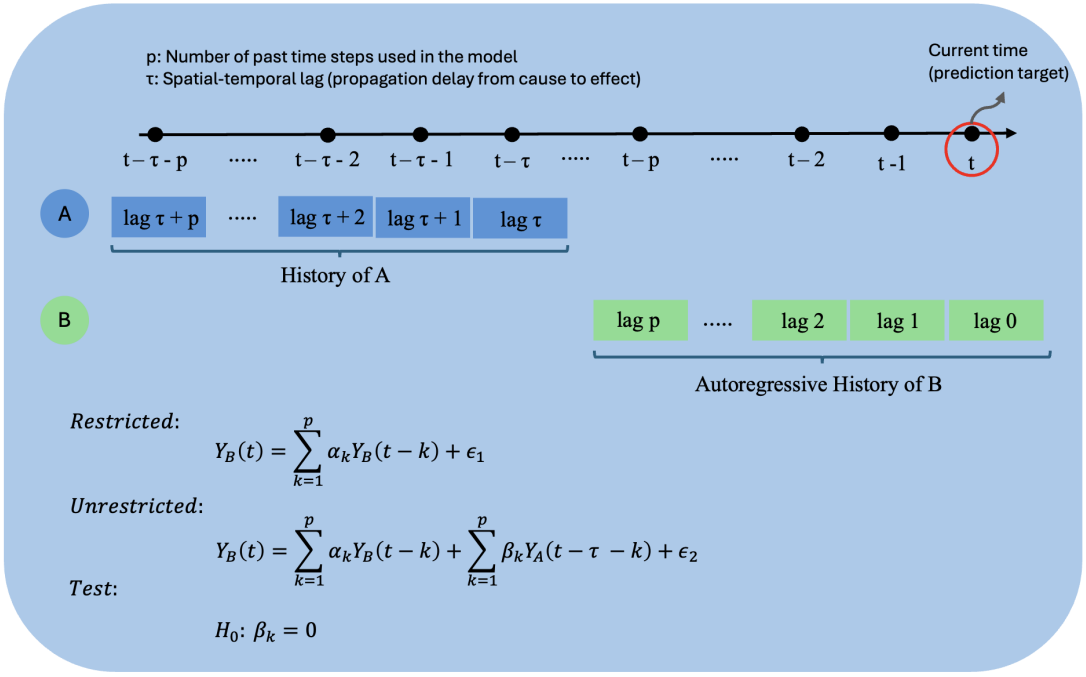


Figure 5.3: Input windows for testing A→B causality. Cause node A uses lags $\tau + p$ to p (τ = propagation delay); effect node B uses its own past p steps. An F-test compares restricted and unrestricted models.

By performing this test for all pairs of road segments, the detected causal relationships can be assembled into a directed spatio-temporal Granger causality graph (STGC graph), which represents the propagation structure of traffic information transmission across the network (11).

Figure 5.4 illustrates the overall pipeline used to construct the STGC graph. We adopt a rejection threshold of $p < 0.01$ for the Granger tests, following the STGC-GNN setting in (11). We also recognize that this is a discovery process rather than a falsification process ; so a looser threshold (e.g., $p < 0.05$) could be explored in future work.

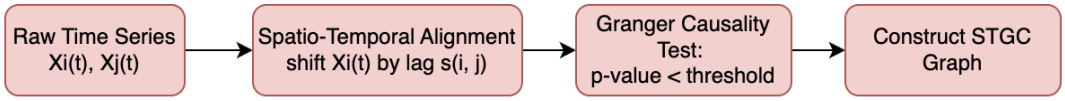


Figure 5.4: Pipeline for constructing the Spatio-Temporal Granger Causality (STGC) graph.

After the causal graph is obtained, it can be incorporated into the neural point process framework to guide information propagation during model training. As a comparison to the original STGNPP architecture in Figure 2.3, where the spatial dependency is consists of the binary matrix and an adaptive learning matrix, we replace the adaptive adjacency with the STGC graph while keeping the binary connection matrix. Figure 5.5 highlights this modification (shown in pink) (2; 11).

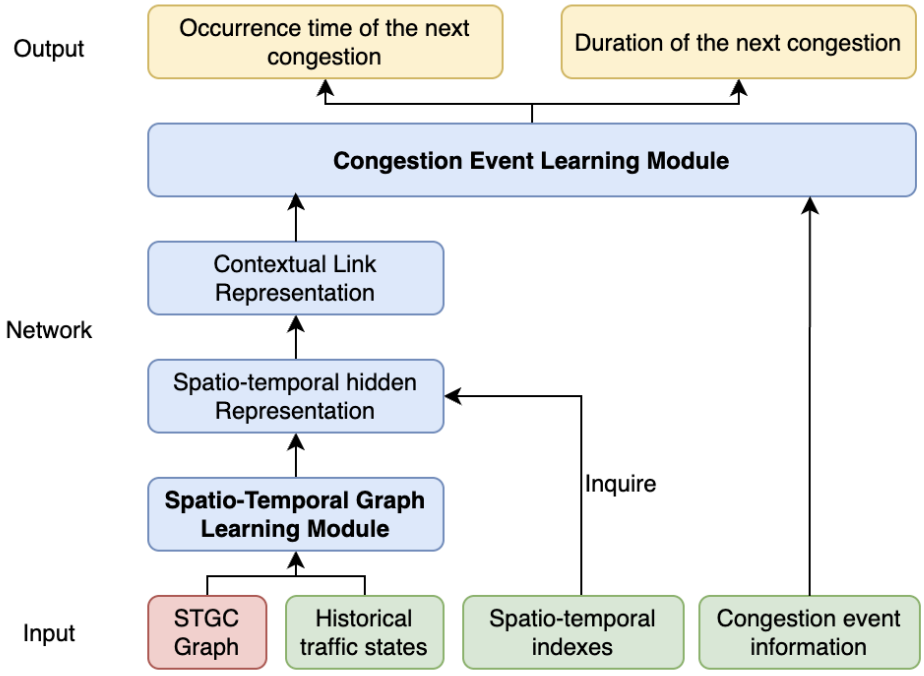


Figure 5.5: Integration of the STGC graph into the STGNPP framework.

5.2.3 Logistic spatial-temporal Granger causality

While the original STGC graph captures traffic interactions by treating speed as a proxy for traffic state and applying a vector autoregressive (VAR) framework to test Granger causal relationships, this setup may not fully match the goal of congestion prediction. Speed and congestion are closely related, but speed is only an indirect indicator. Congestion is better characterized by traffic density and flow breakdown than by speed alone (23). The same speed can mean congestion on one road and free flow on another, so speed lacks contextual meaning when taken in isolation. As a continuous variable, speed is also not well suited to represent the threshold-triggered and event-driven nature of congestion, where a transition to congested flow occurs when certain conditions are exceeded.

To better align the causal discovery process with the prediction target, we propose to directly model causal relationships based on *congestion states* rather than speed. Specifically, we adopt a **logistic Granger causality test**, which is designed to capture causal dependencies between binary congestion variables.

The key idea of logistic Granger causality is to test, for each ordered pair of road segments (A, B), whether the historical congestion status of road A provides statistically significant predictive information about the congestion status of road B, beyond what is contained in road B's own history. Let $y_{A,t} \in \{0, 1\}$ denote the congestion status of road A at time t, where 1 indicates congestion and 0 otherwise. The test compares two logistic regression models. The **restricted model** uses only the autoregression of the target road B:

$$P(y_{B,t} = 1) = \sigma\left(c + \sum_{k=1}^{p} \alpha_k y_{B,t-k}\right), \tag{5.1}$$

where $\sigma(\cdot)$ denotes the logistic sigmoid function, c is the intercept, and p represents the number of time lags. The **unrestricted model** adds the lagged congestion states of

the candidate cause road A (with spatial delay τ for propagation from A to B):

$$P(y_{B,t} = 1) = \sigma\left(c + \sum_{k=1}^{p} \alpha_k y_{B,t-k} + \sum_{k=1}^{p} \beta_k y_{A,t-\tau-k}\right). \tag{5.2}$$

Following (24), we test the null hypothesis $\beta_1 = \cdots = \beta_p = 0$ using a likelihood ratio test. If the test rejects the null, we conclude that road A Granger-causes road B.

Based on these tests, we construct a *spatio-temporal Granger causality graph*, where nodes represent road segments and directed edges indicate statistically significant congestion propagation relationships. This graph is then used as the structural foundation for modeling congestion propagation in our prediction framework.

5.3 Experimental Setup

Using the preprocessed Xuancheng dataset described in Section 4, we conducted experiments by reproducing the STGNPP model proposed by (2), which integrates the Spatio-Temporal Graph Learning Module, the Congestion Event Learning Module, and the Intensity Function Network.

For model training, we set the batch size to 16 and the number of epochs to 10. The parameter α was set to 300 (see justification below). The dataset was divided into training, validation, and test sets with a ratio of 60% / 20% / 20%. Following the setting in (2), the input time window was fixed at six hours to predict the congestion state at the next time step. The learning rate was set to 0.001, consistent with the original paper; the stride of the time window was set to one hour.

To evaluate the effectiveness of different spatial dependency structures, we constructed several types of graphs as inputs to the model (2; 11):

- **Identity Graph.** Each node is only connected to itself. This graph removes all

spatial dependencies and serves as a lower-bound baseline.

- **Spatial Distance (SD) Graph.** Edges are constructed based on geographical proximity between road segments. This graph captures spatial relationships purely from the road network topology.

- **Random Graph.** To verify that the edge connectivity in the STGC graph is meaningful rather than arbitrary, we generated 10 random graphs for each prediction horizon. These graphs match the STGC graph in structure but not in connectivity: (1) the sparsity is consistent with the STGC graph; (2) each node has the same number of neighbors as in the STGC graph; (3) neighbor node IDs are randomly assigned. Comparing the Random graph with the STGC graph isolates whether performance gains arise from the causal information encoded in the edges. The final performance is reported as the average across the 10 random graphs.

- **STGC-A (Speed-based STGC Graph).** This graph follows the original STGC construction method, where traffic speed is used to estimate Granger causal relationships between roads (11).

- **STGC-B (Congestion-based STGC Graph).** This graph represents the key methodological innovation of our work. Instead of using traffic speed, we construct the STGC graph using congestion states and the proposed logistic Granger causality test.

5.4 Training Objective and Evaluation Metrics

To evaluate the effectiveness of the proposed model, we adopt both training losses and prediction performance metrics.

Training and Validation Loss Functions During the training phase, the model is optimized using two loss components: the Neural Point Process (NPP) loss and the duration loss. Both the NPP loss and the duration loss follow the original STGNPP formulation; only the weighting coefficient α is adjusted in our experiments.

- **Neural Point Process Loss (L_{npp}).**

  The NPP loss measures how well the model captures the temporal intensity of congestion events. The model parameterizes intensity as a function of *inter-event time* τ (the time passed since the previous event). For each event i with observed inter-event time τ_i, the loss is defined as

$$L_{\text{npp}} = \sum_{i=1}^{N} \left(-\log \lambda(\tau_i \mid h_i) + \Lambda(\tau_i \mid h_i) \right), \tag{5.3}$$

  where $\lambda(\tau \mid h_i)$ denotes the instantaneous intensity at τ given the hidden state h_i, and $\Lambda(\tau_i \mid h_i) = \int_0^{\tau_i} \lambda(s \mid h_i)\, ds$ is the cumulative intensity. The first term encourages high intensity at the observed τ_i; the second term penalizes excessive intensity over $[0, \tau_i]$. On the other hand, directly minimizing MAE_t in minutes would treat inter-event prediction as a plain regression target and generally would not impose the probabilistic constraints implied by an intensity-based event model.

- **Duration Loss (L_{duration}).**

  The duration loss measures the prediction error of congestion duration. We adopt the log-smooth loss used in the STGNPP framework:

$$L_{\text{duration}} = \frac{1}{N} \sum_{i=1}^{N} \log \left(1 + \left| \hat{d}_i - d_i \right| \right), \tag{5.4}$$

  where d_i denotes the true congestion duration and $\hat{d}_i$ represents the predicted

duration. The log-smooth loss is a smoother version of absolute error and tends to be more stable to optimize under highly imbalanced duration distributions than plain MAE_d.

The overall training objective combines the two losses:

$$L_{\text{total}} = L_{\text{npp}} + \alpha \cdot L_{\text{duration}}. \tag{5.5}$$

The parameter α controls the balance between the two objectives. The original STGNPP framework uses $\alpha = 1$; in our experiments, we set $\alpha = 300$ empirically. This choice is motivated by the scale difference that emerges after adopting the log-smooth loss for duration: the duration loss magnitude becomes very small (typically two orders of magnitude smaller than the NPP loss). To balance the duration and NPP loss terms during training and prevent the duration prediction task from being underweighted, we increase α to 300.

Evaluation Metrics During the evaluation phase, we use Mean Absolute Error (MAE) to measure the prediction accuracy of congestion duration and inter-event time. Compared with the training losses, MAE provides a more direct and interpretable measure of prediction error in terms of minutes.

- **Duration Prediction Error ($\mathbf{MAE}_d$).**

$$\mathrm{MAE}_d = \frac{1}{N} \sum_{i=1}^{N} |\hat{d}_i - d_i| \tag{5.6}$$

This metric measures the average absolute difference between the predicted congestion duration $\hat{d}_i$ and the ground-truth duration d_i.

- **Inter-Event Time Prediction Error ($\mathbf{MAE}_t$).**

$$\mathrm{MAE}_t = \frac{1}{N} \sum_{i=1}^{N} |\hat{\tau}_i - \tau_i| \tag{5.7}$$

where τ_i and $\hat{\tau}_i$ denote the true and predicted inter-event times (minutes until the next congestion event), respectively.

Causal Evaluation Although our framework introduces a causal discovery component through the proposed logistic Granger causality graph, it is difficult to quantitatively evaluate causal relationships using standardized metrics such as causal alignment score or causal coverage in real-world traffic datasets. Therefore, instead of evaluating causal structures directly, we assess their effectiveness indirectly through improvements in downstream prediction performance.

Handling Duration Imbalance As noted in Section 3, more than 94% of congestion events last only 5–15 minutes; long-duration events are rare. This strong imbalance causes the model to favor predicting values close to the median duration in order to minimize the overall loss, which is unfavorable for long-duration prediction.

To mitigate this issue, we introduce two adjustments to the loss function. First, we add a **variance penalty** (*variance_factor* = 20) that encourages the model to produce a wider range of predicted durations, thus preventing collapse to the median. Second, we apply a **long-duration penalty weight**: events with true duration $\geq$ 30 minutes are assigned a weight of 3.0 in the duration loss. This weighting increases the penalty for underestimating long durations and helps the model better capture rare long events. This follows a widely used reweighting principle, as in (30).

CHAPTER 6 RESULTS

Given the experimental setting and evaluation metrics described above, we conducted experiments to evaluate the prediction performance under different historical time window sizes (2h, 4h, 6h, and 8h) and different graph structures (binary with adaptive learning, identity, SD, random, STGC-A, and STGC-B). The results are summarized in Table 6.1. The STGNPP denotes the original STGNPP backbone model, while the other variants correspond to different graph inputs.

As shown in Table 6.1, the STGNPP, SD, and Random graphs exhibit relatively similar performance across different history window settings, with no substantial differences observed. In contrast, the Identity graph consistently performs the worst, which suggests that ignoring spatial relationships significantly degrades prediction performance.

The speed-based STGC graph (STGC-A) outperforms both the Identity and Random graphs across nearly all time horizons. Outperforming the Identity graph demonstrates the importance of spatial structure, since Identity uses no spatial information. Outperforming the Random graph indicates that the causal connectivity encoded in STGC-A is meaningful; the performance gain does not arise merely from arbitrary graph connections or a certain level of sparsity. In particular, STGC-A achieves the best Val Loss and MAE_t at 6 h, the best MAE_t at 2 h, and competitive results elsewhere. Nevertheless, STGC-A does not consistently outperform STGNPP or SD; the

three achieve similar overall performance, which might suggests comparable strength in congestion prediction for this dataset.

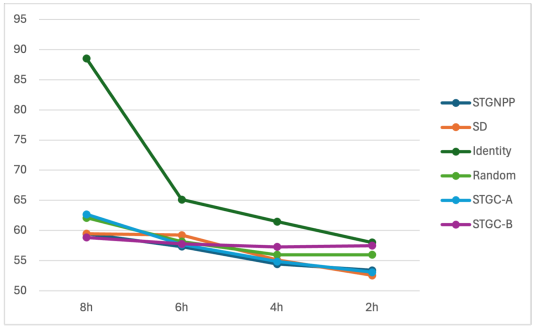
Lastly, STGC-B, the graph constructed based on congestion, does not outperform the other graph structures. Its performance remains relatively stable across history window settings. It consistently outperforms the Identity graph. This supports the importance of spatial structure, but is roughly on par with the Random graph, which may suggest that its connections are not substantially more informative than random connectivity in this setting.

Taken together, these results are consistent with a practical limitation of pairwise logistic Granger tests on sparse binary congestion sequences: when congestion is rare and many segments co-activate during peaks, the inferred edges can be noisy and weakly informative for STGNPP-style training. Therefore, for this dataset and modeling pipeline, using logistic Granger causality on raw congestion events may be a less reliable graph construction strategy than speed-based STGC or topology-based baselines.

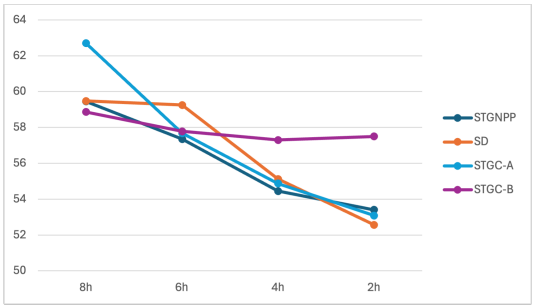
Figures 6.1, 6.2, and 6.3 below illustrate the performance of different graph structures in terms of Val Loss, MAE_t, and MAE_d under different historical time window lengths. In each figure, the left subfigure shows all six graph structures; the right subfigure focuses on STGNPP, SD, STGC-A, and STGC-B for better comparison among the main baselines and the proposed causal graph.

6.1 Trials' Details

Figures 6.4–6.6 summarize the training results obtained during the training phase. Figure 6.4 presents the validation losses for all models across 10 epochs. The validation loss corresponds to the loss computed during the model validation phase. As shown in the figure, the validation losses of all six models follow a similar trend:

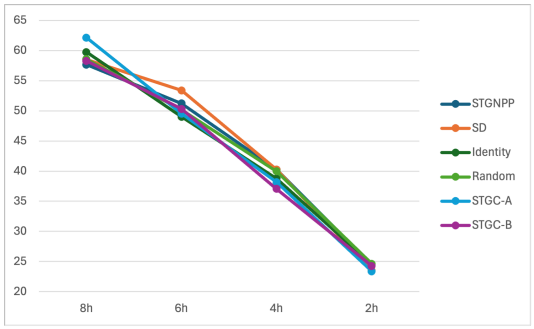


(a) Val Loss: all graph structures

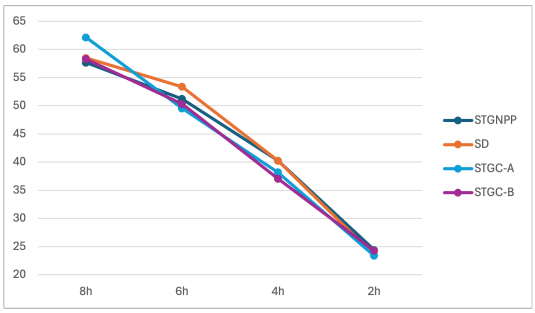


(b) Val Loss: STGNPP, SD, and STGC-A

Figure 6.1: Validation loss comparison under different graph settings.

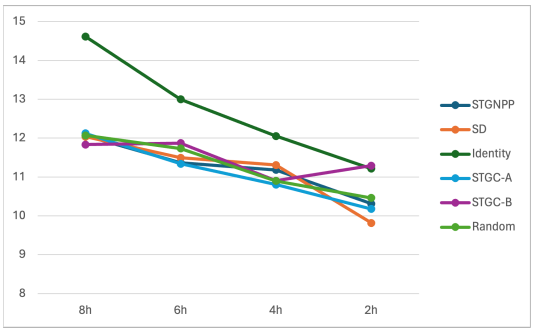


(a) MAE_t: all graph structures

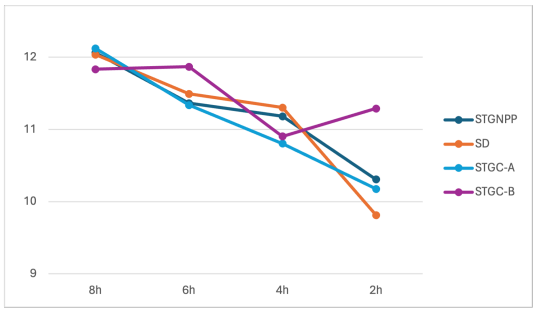


(b) MAE_t: STGNPP, SD, and STGC-A

Figure 6.2: MAE_t comparison under different graph settings.



(a) MAE_d: all graph structures



(b) MAE_d: STGNPP, SD, and STGC-A

Figure 6.3: MAE_d comparison under different graph settings.

they all start from approximately 80 and decrease at a comparable rate during training, eventually converging to very similar values. The only exception is the *Identity* graph.

For complete per-epoch numeric results (training/validation losses and MAE

values for each graph setting), see Appendix A.2.

Among the six graph structures, Identity is the only graph that does not incorporate spatial information. As a result, it exhibits a noticeably higher validation loss and a slower convergence rate compared with the other graph-based models.

As described in Section 5.4, the total loss $L_{\text{total}} = L_{\text{npp}} + \alpha L_{\text{duration}}$ combines NPP and duration components.

For the duration loss, a pattern similar to the overall validation loss can be observed. The Identity graph still decreases more slowly and remains noticeably higher than the losses of the other graph structures.

In contrast, the NPP loss shows relatively larger fluctuations across different graphs. Interestingly, the Identity graph performs comparatively better under this loss component than under the others. Across both individual loss components as well as the overall validation loss, the performances of STGC-A and STGC-B remain consistently around the average level. This suggests that incorporating causal information through STGC does not significantly degrade the model's ability to predict congestion events.

Figures 6.7, 6.8, and 6.9 visualize the predicted versus true congestion durations for STGNPP, STGC-A, and STGC-B under the 6-hour time window, comparing how tightly each model fits the ground truth. All configurations perform reasonably well for short events but struggle when the true duration becomes large. This behavior is consistent with the strong duration imbalance in our dataset (Section 3): most events last 5–15 minutes, so the models minimize loss primarily on the abundant short-duration cases. Following the adjustment described in Section 4.3 (additional penalties for duration variance), prediction improved by approximately 8%. Overall, the three models achieve similar duration and NPP loss under the 6-hour setting; visually, STGC-A tends to predict slightly higher values than STGNPP.

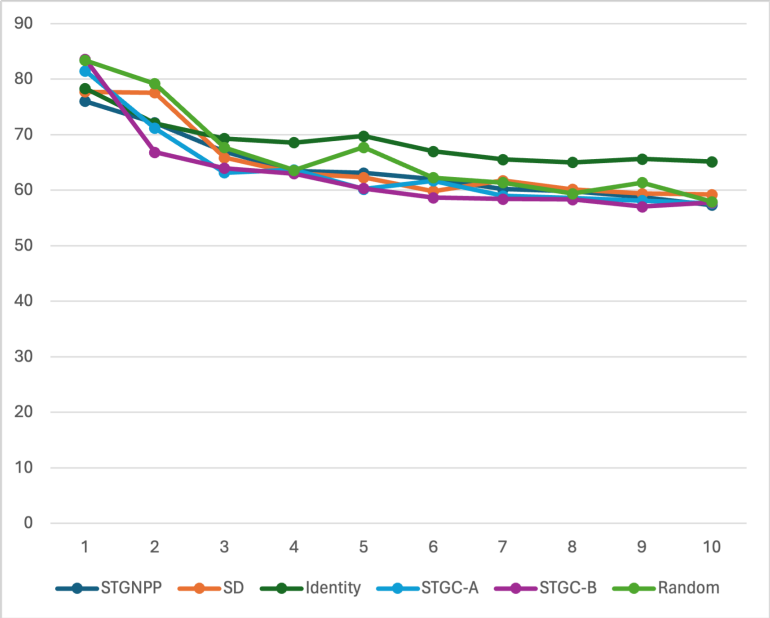


Figure 6.4: Training and validation loss over 10 epochs. The total loss is composed of $L_{\text{npp}} + \alpha L_{\text{duration}}$.

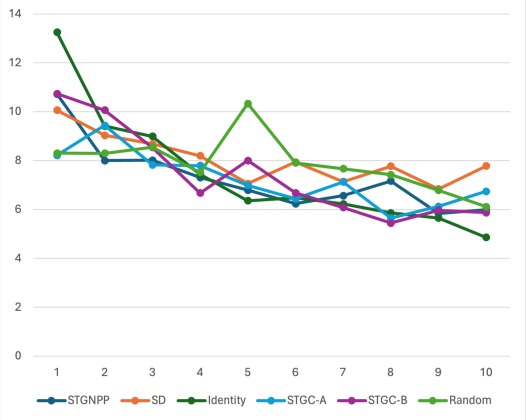


Figure 6.5: NPP loss across 10 epochs.

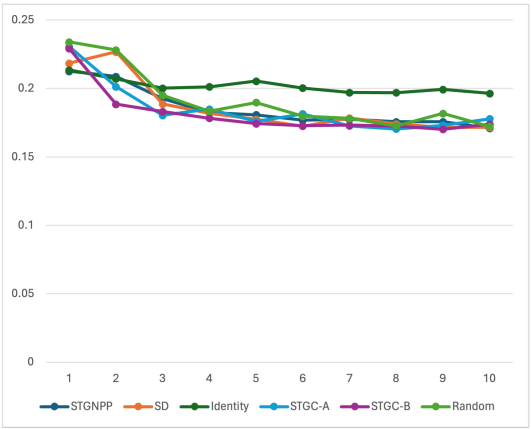


Figure 6.6: Duration loss across 10 epochs.

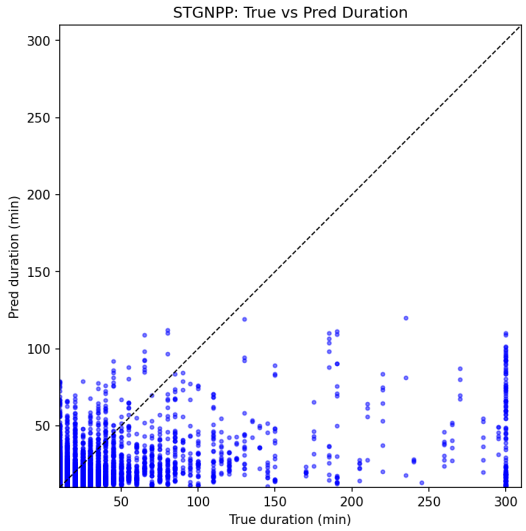


Figure 6.7: Predicted vs. true congestion duration (STGNPP).

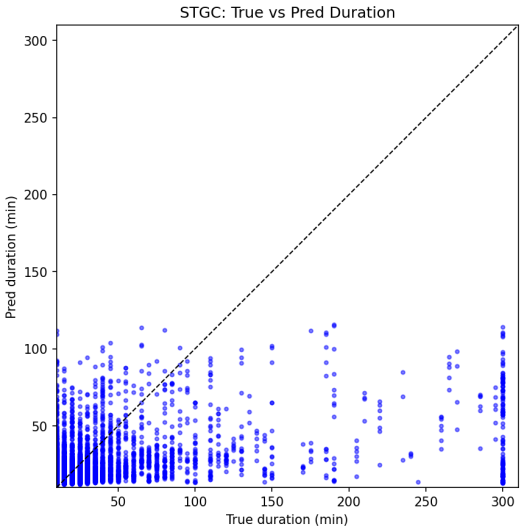


Figure 6.8: Predicted vs. true congestion duration (STGC-A).

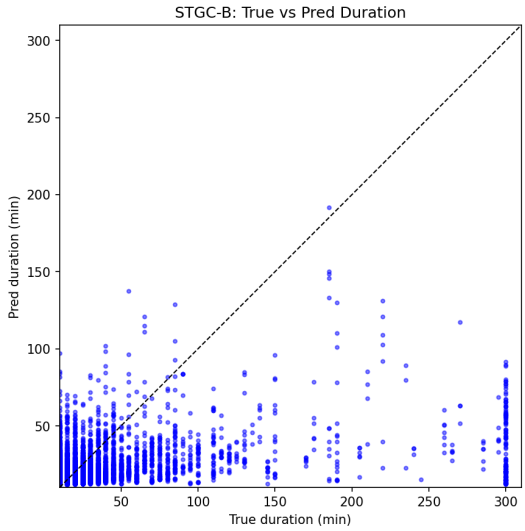


Figure 6.9: Predicted vs. true congestion duration (STGC-B).

6.2 Visualization of STGC Graph

6.2.1 Causality Interpretability

To further investigate the causal relationships between nodes, we examine whether the performance of STGC-A under the 6-hour historical window is truly at-

tributable to meaningful causal information captured by the graph, rather than randomness in the graph structure. In particular, we analyze whether the causal graph provides informative relationships that cannot be captured by other graph constructions, such as the spatial distance (SD) graph, random graph, or adaptive adjacency matrices. This analysis is conducted from two perspectives.

First, we analyze the causal directions encoded in the STGC matrix and the binary connection matrix to verify whether the directions captured by Granger causality are consistent with the physical connectivity of the road network. Since the binary matrix encodes directed one-hop neighboring relationships, we extract only the one-hop connections from the STGC matrix for a fair comparison. Figure 6.10 compares the directionality of these one-hop connections. The results show that 88.5% of the causal directions are consistent with those in the binary matrix. When directions disagree, this may reflect either (a) Granger causality inferring spurious or incorrect relationships, or (b) the traffic spillover effect: congestion can propagate through indirect paths rather than along direct physical links. For example, queue spillback causes downstream congestion to affect upstream segments, so Granger may infer a causal direction opposite to the physical road direction. The high consistency we observe helps explain why combining the STGC graph with the binary matrix can maintain prediction performance comparable to STGNPP; the STGC graph additionally provides multi-hop causal relationships beyond the one-hop structure.

Second, we visualize the prediction results of both STGNPP and STGC-A under the 6-hour historical window to identify the nodes where STGC-A achieves better prediction performance.

As shown in Figure 6.11, among all evaluated nodes, 41.3% show comparable prediction performance between the two models, 24.6% favor STGNPP, and 34% favor STGC-A. Here, we define "comparable" as $|\text{MAE}_{\text{STGC-A}} - \text{MAE}_{\text{STGNPP}}| \leq 0.5$ minutes

Figure 6.10: Consistency between causal directions and actual road connectivity.

Figure 6.11: Prediction performance comparison under the 6-hour history window. Yellow indicates equal prediction performance, red indicates better performance of STGNPP, and green indicates better performance of STGC-A.

or a relative difference $\leq 1\%$. The road segments where STGNPP performs better generally have larger average MAE values, whereas those where STGC-A performs better tend to have smaller average MAE values. This may suggests that STGC may be more suitable for relatively simple road segments while the adaptive learning matrix in STGNPP captures more complex road interactions.

Table 6.1: Validation performance under different models and time history settings

Model	Time	Metrics				
		Val Loss	Val NPP Loss	Val Dur. Loss	MAE_t	MAE_d
STGNPP	8h	59.45	5.08	0.18	57.65	12.06
	6h	57.35	5.99	0.17	51.21	11.36
	4h	54.45	5.26	0.16	40.25	11.18
	2h	53.40	4.06	0.16	24.40	10.31
SD	8h	59.48	6.05	0.18	58.49	12.03
	6h	59.26	7.78	0.17	53.38	11.49
	4h	55.13	4.60	0.16	40.24	11.30
	2h	52.57	4.23	0.16	23.49	9.81
Identity	8h	88.52	8.17	0.25	59.75	14.61
	6h	65.15	4.86	0.20	49.01	13.00
	4h	61.50	5.11	0.19	38.77	12.05
	2h	58.01	4.13	0.18	24.54	11.21
Random	8h	62.13	6.12	0.18	58.57	12.06
	6h	58.17	5.91	0.17	50.16	11.73
	4h	55.99	5.59	0.16	39.98	10.90
	2h	55.98	4.84	0.17	24.61	10.46
STGC-based Graphs						
STGC-A	8h	62.69	8.75	0.18	62.14	12.12
	6h	**57.67**	**5.87**	**0.17**	**49.52**	**11.24**
	4h	54.88	5.61	0.16	38.18	10.81
	2h	53.08	4.35	0.16	23.37	10.18
STGC-B	8h	58.86	5.13	0.18	58.29	11.83
	6h	57.78	5.48	0.17	50.31	11.86
	4h	57.29	5.60	0.17	37.05	10.90
	2h	57.50	4.55	0.17	24.24	11.29

Errors often decrease when the window is shortened from 8 h toward 2 h (notably for MAE_t).

CHAPTER 7 DISCUSSION

In this section, we analyze the performance of the proposed STGC-based graphs and discuss several possible reasons why STGC-A and STGC-B did not outperform the baseline models (STGNPP and SD) in the current experiments. Nevertheless, we argue that STGC-based graphs remain meaningful and have strong potential for improvement, particularly in terms of interpretability and structural modeling of traffic dynamics.

7.1 Shortcomings of the Proposed Models

Mismatch between STGC-A design and prediction objective. The original application scenario of STGC-A is traffic speed prediction. In the STGC-GNN framework, the objective is to model how the speeds of road segments influence the speeds of other road segments in the future. STGC-A therefore employs a Vector Autoregression (VAR) framework to test whether the past speed of one road segment Granger-causes the current speed of another.

However, the prediction task in this study focuses on congestion events, specifically the duration of congestion and the time interval between congestion occurrences. Although speed is one of the most informative variables in traffic systems, its relationship with congestion events is not perfectly aligned with this prediction objective. In practice, congestion is defined using a predefined speed threshold. As a result, varia-

tions in speed do not always directly translate into whether the congestion threshold is crossed. This mismatch between the variable used for causal discovery (speed) and the prediction target (congestion events) may limit the effectiveness of STGC-A in this task.

Difficulty of causal discovery under binary congestion modeling in STGC-B. STGC-B attempts to better align with the congestion prediction objective by converting congestion states into binary time series. Logistic regression is then used to test whether congestion on one road segment at previous time lags Granger-causes congestion on another segment.

While this formulation more directly reflects the congestion prediction task, it introduces challenges due to severe class imbalance. Congestion events occur relatively infrequently in traffic data (12.32%), which means that most time steps correspond to the non-congested state. Consequently, the time series are dominated by zeros, with relatively few positive observations. Such imbalance can make causal discovery unstable, as logistic regression models may fail to identify some true causal edges or may incorrectly infer spurious relationships.

A closely related issue for STGC-B is temporal synchronization of congestion across the network: during peak hours, congestion often emerges simultaneously on many road segments. In our dataset, approximately 23.5% of congestion events occur during morning and evening peak periods. This synchronization can induce strong co-movements among many segments even in the absence of actual physical propagation, which makes it difficult for pairwise Granger-style tests on binary congestion sequences to distinguish propagation-driven dependence from shared temporal shocks.

Together, sparsity and synchronization can result in noisy estimated graphs, which in turn limits their usefulness as spatial priors for downstream prediction models.

7.2 General Observations and Considerations

Dataset limitations. The dataset used in this study may also affect the experimental outcomes. Unlike widely used traffic datasets such as METR-LA (25) or taxi trajectory datasets from cities such as New York (26) or Beijing (27), this study uses the Xuancheng dataset, which is generated through simulation and resampling procedures intended to approximate real-world traffic patterns.

Because the STGNPP model requires explicit road connectivity information, it was not possible to directly apply the model to several commonly used datasets that lack detailed physical road network structures. The Xuancheng dataset was therefore selected as a feasible alternative. However, simulated datasets may introduce distribution shifts or distort certain traffic patterns. If congestion propagation patterns are not accurately represented, the causal structures learned by STGC may deviate from real-world traffic dynamics. Furthermore, this dataset has not yet been widely used in other traffic prediction studies, which makes it difficult to benchmark its performance against existing work.

Land use and POI categories (e.g., schools, hospitals, commercial districts) are widely recognized correlates of where and when congestion arises, but they are not included as explicit covariates in our pipeline. The Xuancheng inputs are segment-level traffic dynamics together with the road-network graph; any influence of nearby activity is therefore absorbed only indirectly, to the extent that it is already encoded in correlated speeds and flows. In this way, we cannot decompose predicted intensity or duration into a POI-attributable share versus other components, nor report a quantitative "percent explained" by those geographic variables. Answering that question would require augmenting the dataset with POI or land-use features, followed by ablation analyses, which we leave to future work.

Strength of the STGNPP baseline model. Another possible explanation for the limited performance differences is the strong modeling capability of the STGNPP baseline itself. STGNPP integrates both spatial and temporal modeling mechanisms and includes an adaptive adjacency matrix capable of learning complex interactions between road segments.

Due to this strong representation capability, the model may be less sensitive to the exact choice of graph structure, as long as the graph roughly captures the geographic connectivity among road segments. In such cases, improvements introduced by alternative graph constructions may appear relatively small. This observation suggests that further analyses, such as ablation studies, may be necessary to better isolate the contribution of different graph components.

7.3 Why STGC Remains Important and Future Directions

Despite the limitations discussed above, the STGC framework still provides meaningful benefits. One key advantage is interpretability. The adaptive adjacency matrix in STGNPP is learned purely from data; while it may capture useful patterns, it is unclear why the network learns particular edge weights, and the resulting structure lacks a statistical or physical interpretation. In contrast, STGC derives its edges from Granger causality tests, a mathematically and statistically grounded procedure, so each causal link has an explicit interpretation: past values of one segment help predict another beyond its own history. This makes the graph structure explainable rather than a black box.

A second advantage is the ability of STGC to capture multi-hop causal relation-

ships. Traditional spatial graphs have narrow scope: SD graphs rely only on geographic proximity, and binary adjacency matrices typically encode only immediate (one-hop) neighboring links. STGC can reveal causal relationships across multiple hops in the road network, allowing the model to capture propagation patterns that extend beyond direct physical neighbors.

Several directions may further improve the effectiveness of STGC-based models. One potential approach is to construct different causal graphs for different traffic regimes, such as peak and off-peak periods, which may help reduce the impact of synchronized congestion patterns. Another direction is to address the class imbalance problem in STGC-B through techniques such as class weighting or improved classification methods for causal discovery. In addition, more detailed ablation studies could be conducted to analyze how different graph structures contribute to prediction performance.

Finally, future research may explore the use of additional datasets and alternative model architectures. For example, adapting the STGNPP framework to operate with other widely used traffic datasets could provide stronger empirical validation. Such investigations may help further clarify the role of causal graphs in traffic congestion prediction and potentially unlock the full potential in prediction of STGC-based approaches.

CHAPTER 8 CONCLUSION

In this study, we investigated the role of causal graph structures in traffic congestion prediction by incorporating Spatio-Temporal Granger Causality (STGC) into the STGNPP framework. Specifically, we constructed two variants of causal graphs, STGC-A and STGC-B, and evaluated their effectiveness in predicting congestion duration and the time interval between congestion events.

Experimental results show that STGC-based graphs achieve prediction performance comparable to strong baseline models such as STGNPP and spatial distance (SD) graphs. Although STGC-A and STGC-B do not consistently outperform these baselines, they are able to maintain similar levels of predictive accuracy while introducing additional interpretability through explicitly modeled causal relationships.

Through detailed analysis, we identify several factors that may limit the performance gains of STGC in this task. These include the mismatch between speed-based causal discovery and congestion-based prediction objectives, the difficulty of causal inference under highly imbalanced binary congestion data, and the presence of synchronized congestion patterns during peak hours. In addition, the use of simulated datasets and the strong representation capability of the baseline model further reduce the observable performance differences across graph structures.

Despite these limitations, our findings highlight the value of incorporating causal structure into traffic prediction models. Unlike purely data-driven adjacency matrices,

STGC provides interpretable insights into how congestion propagates across the road network. Moreover, STGC is capable of capturing long-range, multi-hop dependencies that are difficult to model using traditional spatial graphs. This work also suggests that congestion prediction may benefit more from short-term historical information rather than long input windows, which provides useful guidance for future model design.

Future work may explore several directions to further improve the effectiveness of STGC-based approaches. These include constructing regime-specific causal graphs (e.g., peak vs. off-peak periods), addressing class imbalance in binary congestion modeling, conducting more detailed ablation studies, and evaluating the approach on additional real-world datasets. Furthermore, integrating causal graph learning more tightly with prediction objectives may help unlock the full potential of causal modeling in traffic systems.

In summary, although STGC does not significantly outperform existing graph constructions in the current setting, this study demonstrates the feasibility and value of incorporating causal graph structures into neural point-process-based traffic congestion prediction. More importantly, it suggests that future traffic prediction research should move beyond accuracy alone and place greater emphasis on interpretability and congestion propagation mechanisms. In this sense, the present work provides not only an empirical investigation, but also a meaningful step toward more explainable and theoretically grounded intelligent transportation systems.

APPENDIX A VARIABLE DEFINITIONS AND TRAINING-EPOCH LOGS

A.1 Variable Definitions

A.1.1 Graph Structure Layer Variables

Table A.1: Graph Structure Layer Variables

Variable	Description	Type
CID	Road segment identifier	Categorical
Lane Count	Number of lanes in the road segment	Numerical
Turn Direction	Direction of vehicle movement	Categorical
Downstream Road	Reference to downstream road segment	Categorical
Geometry	Geographic coordinates	Categorical
Length	Physical length of the road segment (meters)	Numerical

A.1.2 Temporal Data Layer Variables

Table A.2: Temporal Data Layer Variables

Variable	Description	Type
Detector ID	Unique identifier for traffic sensors	Categorical
Road ID	Reference to corresponding road segment	Categorical
From Time	Start time of measurement interval	Numerical
To Time	End time of measurement interval	Numerical
Interval	Duration of measurement (minutes)	Numerical
Vehicle Count	Total number of vehicles passing the sensor	Numerical
Regular Vehicle Count	Count of standard passenger vehicles	Numerical
Large Vehicle Count	Count of trucks and buses	Numerical
Arithmetic Speed	Mean speed of vehicles (km/h)	Numerical
Harmonic Speed	Harmonic mean speed of vehicles (km/h)	Numerical
Turn Direction	Direction of vehicle movement at intersections	Categorical

A.2 Training Metrics Tables Across Graph Variants

Below 6 tables are the 6 graphs (STGNPP baseline, STGNPP with SD, STGNPP with identity, STGC-GNPP with speed-based matrix, STGC-GNPP with congestion-

based matrix, STGNPP with random graph) under the 6 hour configuration in 10 epochs, which is a elaboration correspnding to the epoch loss graphs in Results, details in trials figures.

Epoch	tr. loss	tr. npp	tr. dur	val. loss	val. npp	val. dur	mae_d	mae_t
1	131.88	20.16	0.36	76.05	10.72	0.21	12.32	65.51
2	74.94	9.08	0.21	72.15	8.00	0.21	13.27	57.70
3	72.75	8.58	0.21	66.96	8.01	0.19	12.33	55.17
4	68.08	8.11	0.20	63.46	7.31	0.18	11.80	52.87
5	66.32	7.78	0.19	63.14	6.79	0.18	12.11	47.92
6	65.81	7.50	0.19	61.99	6.24	0.18	11.78	48.49
7	63.79	7.10	0.19	60.20	6.57	0.18	12.15	50.63
8	62.63	6.43	0.19	59.84	7.17	0.18	12.06	51.37
9	61.87	6.09	0.18	58.78	5.83	0.18	11.98	50.45
10	60.84	5.76	0.18	57.35	5.99	0.17	11.36	51.21

Table A.3: Training and validation metrics across epochs. (STGNPP - baseline model)

Epoch	tr. loss	tr. npp	tr. dur	val. loss	val. npp	val. dur	mae_d	mae_t
1	170.83	25.59	0.48	77.74	10.06	0.22	12.68	64.83
2	78.90	9.19	0.23	77.57	9.04	0.23	14.13	59.62
3	73.52	8.79	0.21	65.87	8.68	0.19	12.34	57.26
4	66.60	8.10	0.19	63.07	8.20	0.18	12.09	54.01
5	64.90	8.02	0.19	62.33	7.06	0.18	11.52	50.20
6	63.39	7.71	0.18	59.86	7.94	0.17	11.83	52.47
7	63.15	7.64	0.18	61.75	7.14	0.18	12.24	51.17
8	63.01	7.52	0.18	60.18	7.77	0.17	11.41	52.27
9	61.80	7.30	0.18	59.40	6.84	0.17	11.51	49.54
10	61.59	7.21	0.18	59.26	7.78	0.17	11.49	53.38

Table A.4: Training and validation metrics across epochs (STGNPP with spatial distance matrix)

Epoch	tr. loss	tr. npp	tr. dur	val. loss	val. npp	val. dur	mae_d	mae_t
1	118.99	45.13	0.23	78.35	13.25	0.21	13.07	66.19
2	73.28	10.25	0.21	72.06	9.41	0.21	13.72	63.03
3	71.27	8.91	0.21	69.33	8.99	0.20	13.14	58.35
4	69.98	8.06	0.20	68.58	7.44	0.20	13.49	52.96
5	68.87	7.19	0.20	69.79	6.36	0.21	12.84	50.45
6	69.02	6.79	0.20	67.01	6.48	0.20	12.86	50.04
7	67.05	6.34	0.20	65.55	6.23	0.20	13.07	49.98
8	66.60	5.78	0.20	65.04	5.86	0.20	13.11	48.92
9	67.09	5.61	0.20	65.63	5.65	0.20	13.31	48.88
10	65.50	5.43	0.20	65.15	4.86	0.20	13.00	49.01

Table A.5: Training and validation metrics across epochs (STGNPP with identity matrix)

Epoch	tr. loss	tr. npp	tr. dur	val. loss	val. npp	val. dur	mae_d	mae_t
1	106.39	23.88	0.26	81.51	8.21	0.23	14.09	59.24
2	74.81	9.41	0.21	71.19	9.44	0.20	11.79	61.42
3	68.94	8.67	0.19	63.16	7.82	0.18	12.10	56.06
4	65.81	7.98	0.19	63.68	7.79	0.18	12.52	53.78
5	64.25	7.50	0.18	60.21	6.98	0.17	11.54	50.98
6	63.82	6.96	0.18	61.73	6.44	0.18	12.36	50.05
7	62.79	6.87	0.18	59.01	7.12	0.17	11.84	50.86
8	60.85	6.55	0.17	58.57	5.64	0.17	11.88	48.86
9	61.22	6.21	0.18	58.13	6.12	0.17	12.02	49.39
10	59.61	6.03	0.17	57.67	6.74	0.17	11.24	49.64

Table A.6: Training and validation metrics across epochs (STGC-GNPP (speed-baesd matrix))

Epoch	tr. loss	tr. npp	tr. dur	val. loss	val. npp	val. dur	mae_d	mae_t
1	137.84	28.30	0.35	83.60	10.73	0.23	13.56	66.23
2	74.86	9.93	0.21	66.88	10.06	0.19	12.19	63.12
3	69.33	8.65	0.20	63.96	8.54	0.18	11.77	57.39
4	64.35	7.75	0.19	63.02	6.67	0.18	11.63	47.42
5	63.87	7.50	0.19	60.35	8.01	0.17	11.72	54.00
6	61.85	6.96	0.18	58.67	6.68	0.17	11.45	50.71
7	60.84	6.43	0.18	58.45	6.08	0.17	11.31	48.97
8	61.04	6.04	0.18	58.37	5.45	0.17	11.38	48.84
9	60.27	5.76	0.18	57.07	5.97	0.17	11.48	50.69
10	59.05	5.44	0.18	57.78	5.87	0.17	11.33	49.52

Table A.7: Training and validation metrics across epochs (STGC-GNPP (congestion-based matrix))

Epoch	tr. loss	tr. npp	tr. dur	val. loss	val. npp	val. dur	mae_d	mae_t
1	186.55	23.76	0.53	83.41	8.30	0.23	12.44	62.58
2	82.43	8.86	0.24	79.19	8.30	0.23	12.25	55.87
3	77.88	8.40	0.23	67.72	8.55	0.19	12.32	55.50
4	68.29	8.08	0.20	63.62	7.52	0.18	12.30	51.58
5	67.29	7.78	0.19	67.76	10.33	0.19	11.47	58.85
6	66.70	8.40	0.19	62.29	7.91	0.18	11.45	53.27
7	65.90	7.59	0.19	61.38	7.67	0.18	12.08	51.60
8	64.86	7.34	0.19	59.40	7.42	0.17	11.56	51.32
9	62.33	6.83	0.18	61.39	6.79	0.18	12.53	50.91
10	61.97	6.34	0.18	57.94	6.11	0.17	11.86	48.53

Table A.8: Training and validation metrics across epochs (STGNPP with random graph)

A.3 Code Availability

The code used for this thesis is available at `https://github.com/KatrinaC04/STGC-GNPP-clean`.

BIBLIOGRAPHY

[1] Jiang, W., & Luo, J. (2022). Graph neural network for traffic forecasting: A survey. Expert Systems with Applications, 207, 117921. https://doi.org/10.1016/j.eswa.2022.117921

[2] Jin, G., Liu, L., Li, F., & Huang, J. (2023). Spatio-Temporal Graph Neural Point Process for Traffic Congestion Event Prediction. Proceedings of the AAAI Conference on Artificial Intelligence, 37(12), 14268-14276. https://doi.org/10.1609/aaai.v37i12.26669

[3] Wang, Y., Chen, Y., Li, G., Lu, Y., He, Z., Yu, Z., & Sun, W. (2023). City-scale holographic traffic flow data based on vehicular trajectory resampling. Scientific Data, 10, 57. https://doi.org/10.1038/s41597-022-01850-0

[4] van der Bijl, B., Gijsbertsen, B., van Loon, S., Reurich, Y., de Valk, T., Koch, T., & Dugundji, E. (2022). A Comparison of Approaches for the Time Series Forecasting of Motorway Traffic Flow Rate at Hourly and Daily Aggregation Levels. Procedia Computer Science, 201, 213-222. https://doi.org/10.1016/j.procs.2022.03.030.

[5] Dougherty, M. S. (1995). A review of neural networks applied to transport.

Transportation Research Part C: Emerging Technologies, 3(4), 247–260. https://doi.org/10.1016/0968-090X(95)00009-8.

[6] Granger, C. W. J. (1969). Investigating causal relations by econometric models and cross-spectral methods. *Econometrica*, 37(3), 424–438. https://doi.org/10.2307/1912791.

[7] Schrank, D., Albert, L., Eisele, B., & Lomax, T. (2021). *2021 Urban Mobility Report*. Texas A&M Transportation Institute. https://static.tti.tamu.edu/tti.tamu.edu/documents/umr/archive/mobility-report-2021.pdf.

[8] Wu, C. H., Ho, J. M., & Lee, D. T. (2004). Travel-time prediction with support vector regression. *IEEE Transactions on Intelligent Transportation Systems*, 5(4), 276–281. https://doi.org/10.1109/TITS.2004.837813.

[9] Yu, B., Yin, H., & Zhu, Z. (2018). Spatio-temporal graph convolutional networks: A deep learning framework for traffic forecasting. *Proceedings of the 27th International Joint Conference on Artificial Intelligence*, 3634–3640. https://doi.org/10.24963/ijcai.2018/505.

[10] United Nations Department of Economic and Social Affairs. (2018). 68% of the world population projected to live in urban areas by 2050. Retrieved from https://www.un.org/uk/desa/68-world-population-projected-live-urban-areas-2050-says-un

[11] He, S., Luo, Q., Du, R., Zhao, L., He, G., Fu, H., & Li, H. (2023). STGC-GNNs: A GNN-based traffic prediction framework with a spatial–temporal Granger causality graph. *Physica A: Statistical Mechanics and its Applications*, 128913. https://doi.org/10.1016/j.physa.2023.128913

[12] J. Li & L. Zhu. *Addressing traffic management challenges and improving the operational efficiency of intelligent transportation systems*. Science and Technology Daily, June 23, 2021. Retrieved from https://www.cas.cn/cm/202106/t20210624_4794695.shtml

[13] L.-Z. Mao, H.-G. Zhu, & L.-R. Duan. "The Social Cost of Traffic Congestion and Countermeasures in Beijing." In *Sustainable Transportation Systems: Plan, Design, Build, Manage, and Maintain*, ASCE, 2012. https://doi.org/10.1061/9780784412299.0010.

[14] R. Fu, Z. Zhang, & L. Li, "Using LSTM and GRU neural network methods for traffic flow prediction," in *2016 31st Youth Academic Annual Conference of Chinese Association of Automation (YAC)*, pp. 324–328, 2016. https://doi.org/10.1109/YAC.2016.7804912.

[15] J. Sun, J. Zhang, Z. Yuan, J. Tian, & T. Wang, "A stochastic car-following model in the framework of Kerner's three-phase traffic theory," *Physica A: Statistical Mechanics and its Applications*, 2025, https://doi.org/10.1016/j.physa.2025.130798.

[16] Huang, W., Song, G., Hong, H., & Xie, K. (2017). "Next-generation innovation and development of intelligent transportation system in China." *Science China Information Sciences*, 60(11), 110201. https://doi.org/10.1007/s11432-017-9182-x.

[17] Mystakidis, A., Koukaras, P., & Tjortjis, C. (2025). "Advances in traffic congestion prediction: An overview of emerging techniques and methods." *Smart Cities*, 8(1), 25. https://doi.org/10.3390/smartcities8010025.

[18] Maerivoet, S., & De Moor, B. (2005). "Cellular automata models of road traffic." *Physics Reports*, 419(1), 1–64. https://doi.org/10.1016/j.physrep.2005.08.005.

[19] Liu, Y., & Wu, H. (2017). "Prediction of road traffic congestion based on random forest." In *2017 10th International Symposium on Computational Intelligence and Design (ISCID)*, Vol. 2, IEEE. https://doi.org/10.1109/ISCID.2017.216.

[20] Chen, R. T. Q., Amos, B., & Nickel, M. (2020). "Neural spatio-temporal point processes." arXiv preprint arXiv:2011.04583. https://doi.org/10.48550/arXiv.2011.04583

[21] Kerner, B. S., & Klenov, S. L. (2010). "A theory of traffic congestion at moving bottlenecks." *Journal of Physics A: Mathematical and Theoretical*, 43(42), 425101. https://doi.org/10.1088/1751-8113/43/42/425101.

[22] Shojaie, A., & Fox, E. B. (2022). "Granger causality: A review and recent advances." *Annual Review of Statistics and Its Application*, 9, 289–319. https://doi.org/10.1146/annurev-statistics-040120-010930.

[23] Gerlough, D. L., & Huber, M. J. (1976). *Traffic flow theory*. No. HS-006 783. Retrieved from https://trid.trb.org/View/37152

[24] Mosconi, R., & Seri, R. (2006). "Non-causality in bivariate binary time series." *Journal of Econometrics*, 132(2), 379–407. https://doi.org/10.1016/j.jeconom.2005.02.005.

[25] Li, Y., Yu, R., Shahabi, C., & Liu, Y. (2018). "Diffusion Convolutional Recurrent Neural Network: Data-Driven Traffic Forecasting." *International Conference on Learning Representations*. https://doi.org/10.48550/arXiv.1707.01926

[26] New York City Taxi and Limousine Commission (TLC) Trip Record Data. Accessed on March 23, 2026, from https://registry.opendata.aws/nyc-tlc-trip-records-pds.

[27] Lian, J., & Zhang, L. (2018). "One-month Beijing taxi GPS trajectory dataset with taxi IDs and vehicle status." *Proceedings of the First Workshop on Data Acquisition To Analysis*. https://api.semanticscholar.org/CorpusID:52976092

[28] Zhang, Y., & Zhang, J. (2026). "Congestion-Aware Traffic Forecasting with Physics-Guided Spatio-Temporal Graph Convolutional Networks." *Applied Sciences*, 16(7), 3546. https://doi.org/10.3390/app16073546.

[29] Wikipedia contributors. (n.d.). *Xuancheng*. In *Wikipedia, The Free Encyclopedia*. Retrieved April 23, 2026, from https://en.wikipedia.org/wiki/Xuancheng

[30] Zhu, M., Xia, J., Jin, X., Yan, M., Cai, G., Yan, J., & Ning, G. (2018). "Class Weights Random Forest Algorithm for Processing Class Imbalanced Medical Data." *IEEE Access*, vol. 6, pp. 4641–4652. https://doi.org/10.1109/ACCESS.2018.2789428.