Discovering the Precursors of Traffic Breakdowns Using Spatiotemporal Graph Attribution Networks

Zhaobin Mo[a], Xiangyi Liao[b], Dominik A. Karbowski[c], Yanbing Wang[d,*]

[a] *Department of Civil Engineering and Engineering Mechanics, Columbia University*
[b] *School of Economics, Minzu University of China*
[c] *Vehicle and Mobility Systems Department, Argonne National Laboratory*
[d] *School of Sustainable Engineering and the Built Environment, Arizona State University*

Abstract

Understanding and predicting the precursors of traffic breakdowns is critical for improving road safety and traffic flow management. This paper presents a novel approach combining spatiotemporal graph neural networks (ST-GNNs) with Shapley values to identify and interpret traffic breakdown precursors. By extending Shapley explanation methods to a spatiotemporal setting, our proposed method bridges the gap between black-box neural network predictions and interpretable causes. We demonstrate the method on the Interstate-24 data, and identify that road topology and abrupt braking are major factors that lead to traffic breakdowns.

Keywords: Traffic Breakdown, Spatiotemporal Graph Neural Network, Attribution Network

1. Introduction

Traffic breakdowns, characterized by sudden congestion and reduced vehicle speeds, can lead to severe accidents and increased travel times. Identifying the contributing factors enables the development of predictive models to mitigate these events.

Several methods have been developed to identify and predict traffic breakdowns. Statistical estimators and probabilistic models analyze transitional events, with one approach using statistical estimators to assess breakdown probability by classifying these occurrences [1]. Machine learning techniques, such as artificial neural networks, have also shown promise for modeling abrupt traffic transitions [2]. However, a key limitation of current methods is their inability to systematically link environmental and driver behavior factors with the spatiotemporal dynamics of traffic breakdowns. For instance, while studies highlight precursors such as road geometry or the braking of a lead vehicle in a platoon [3, 4], input data is often simplified into tabular formats. This simplification discards critical spatial and temporal relationships.

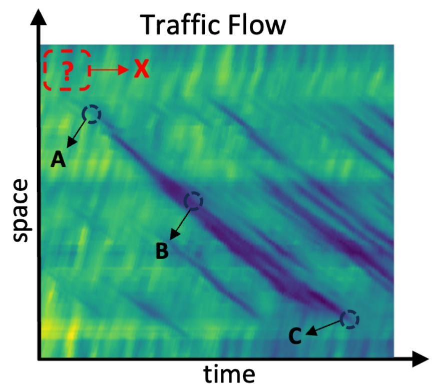


Figure 1: Illustration of traffic breakdowns. A traffic breakdown contains phases of trigger&formation (A), propagation (B) and dissipation (C). Our goal is to discover the potential traffic breakdown precursors from region X, which is the downstream area antecedent to the breakdown trigger.

Fortunately, recent advancements in spatiotemporal graph neural networks (ST-GNNs) shed light on addressing these gaps by effectively capturing both spatial and temporal dependencies [5, 6]. To enhance the interpretability of GNNs, techniques such as Shapley values [7, 8] provide robust frameworks for feature attribution in complex models [9]. Inspired by these techniques, this paper proposes applying ST-GNNs combined with Shapley values to uncover precursors of traffic breakdowns. Specifically, the contributions of this paper are summarized as follows: (1) we extend GNN-based Shapley explanation

*Corresponding author.
Email address: yanbing.wang@asu.edu (Yanbing Wang)

methods to a spatiotemporal setting, enabling a richer interpretation of traffic dynamics by incorporating both spatial and temporal factors, (2) our proposed ST-GNN Shapley method identifies traffic breakdown precursors by bridging the gap between black-box neural network predictions and human-understandable causes, and (3) we validate our approach through a real-world experiment on I-24, demonstrating its effectiveness and robustness in capturing traffic breakdown precursors. Our findings indicate that the lane adjacent to the merging area is the primary source of traffic fluctuations and congestion.

2. Problem Statement

Illustrated in Fig. 1, our goal is to identify the potential traffic breakdown precursors in the spatiotemporal domain. To identify the areas and characteristics of the traffic breakdowns, we discretize the road into spatiotemporal cell of fixed spatial and temporal intervals. The road geometry can be represented as an undirected graph $G = (V, E, A)$. Here, $V = \{v_i\}_{i=1}^{N}$ represents the set of nodes with $N = |V|$ indicating the number of nodes. The set of edges is denoted by E, and $A \in \mathbb{R}^{N\times N}$ is the adjacency matrix, a square matrix capturing the relationships between the nodes in the graph. In this matrix A, each row and column is associated with a node, and the elements A_{ij} specify the existence of edges between the nodes. The features on each node include macroscopic ones like flow, density and speed in each cell, as well as microscopic ones like the trajectory of the vehicles. In this study, we will focus on using the macroscopic features and leave the microscopic ones in the future.

3. Methodology

In this section, we introduce the proposed method, which follows a two-step approach. First, we employ a spatiotemporal prediction model to capture traffic dynamics patterns. Second, we develop an explanation model to interpret the predictions and identify precursors of traffic breakdowns. Fig. 2 illustrates the structure of the proposed framework. In the upper part, features are input into a spatiotemporal graph neural network for prediction. In the lower part, a mask generator produces a graph mask, which is then used to create "what-if" subgraphs by masking node or edge features of the original graph. These subgraphs are fed into a Shapley-based generator, which outputs the explanation ϕ, identifying the spatiotemporal indices of traffic breakdowns.

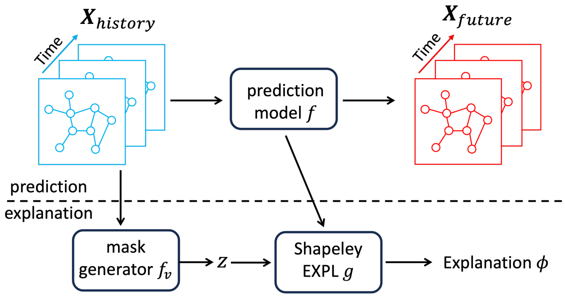


Figure 2: Flowchart of the proposed framework.

3.1. Mask Generator

We begin by developing a mask generator to create discrete masks for spatial and temporal nodes, denoted as $M_S \in \{0,1\}^S$ and $M_T \in \{0,1\}^T$, respectively. For the instance v being explained, our goal is to examine the combined effect of a subset of features and its neighbors on the prediction $f_v(X, A)$. The mask generator identifies the subset to be analyzed. A value of 1 indicates that the variable (node or feature) is included, while 0 indicates it is excluded. First, the mask generator randomly samples from all possible 2^{S+T-1} pairs of masks M_S and M_T, representing all potential coalitions of spatiotemporal nodes (excluding v itself in explanations). Let z be the random variable for selected nodes, $z = (M_S \| M_T)$. Then, the "what-if" subgraphs $\tilde{X}$ are generated by masking out the unselected features, $\tilde{X} = X \odot z$. These subgraphs $\tilde{X}$ will be used to generate explanations.

3.2. Spatiotemporal Prediction Model

There are many choices of the spatiotemporal prediction model f. In fact, our proposed Shapley based methods are model-agnostic and can be used to explain ST prediction models. In this paper we adopt the spatiotemporal graph convolutional network (STGCN) as the neural backbone example. The reason of using STGCN is because of its wide application and generalization.

In STGCN, we first compute the adjacency matrix A of the road topology. The normalized Laplacian matrix is defined as $L = I - D^{-1/2}AD^{-1/2}$, where I denotes the identity matrix and D is the diagonal degree matrix, with each diagonal element D_{ii} being the sum of the elements in the i^{th} row of A. To approximate the graph convolution operator $*_G$, we utilize K-order Chebyshev polynomials, represented by:

$$g_\theta *_G x = g_\theta(L)x = \sum_{k=0}^{K-1} \theta_k \left(T_k(\tilde{L})\right) x,$$

Here, $\theta \in \mathbb{R}^K$ is the vector of polynomial coefficients, and the scaled Laplacian $\tilde{L}$ is calculated as $\tilde{L} = \frac{2}{\lambda_{\max}} L - I$, where $\lambda_{\max}$ is the largest eigenvalue of L. The Chebyshev polynomials $T_k(x)$ are defined recursively by $T_k(x) = 2xT_{k-1}(x) - T_{k-2}(x)$, starting from $T_0(x) = 1$ and $T_1(x) = x$. This K-order Chebyshev polynomial approximation enables each node to be updated based on information from its K neighboring nodes.

After the graph convolution captures the spatial relationships among neighboring nodes, we apply a standard convolution layer in the temporal dimension to refine the node signals. This temporal convolution layer integrates data across neighboring time slices, updating the node signals accordingly. Finally, a 1×D convolution followed by a nonlinear neural network layer is used to produce the final prediction $\hat{X}_{(t-\tau+1):t}$. The mean squared error (MSE) is employed as the loss function:

$$\mathcal{L} = \frac{||X_{(t-\tau+1):t} - \hat{X}_{(t-\tau+1):t}||^2}{N\tau},$$

where $||\cdot||^2$ denotes the $\mathbb{L}_2$ norm.

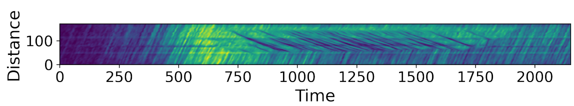
3.3. Explanation Generator

In this section, we construct a surrogate model g based on the dataset $D = \{(z, f(z))\}$ and use it to generate explanations. Formally, an explanation ϕ for f is selected from a set of potential explanations known as the interpretable domain Ω. The explanation ϕ is the result of the optimization process:

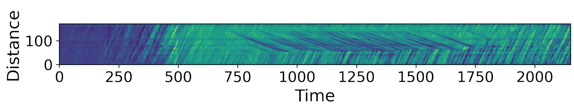
$$\phi = \arg\min_{g\in\Omega} L_f(g), \tag{3.1}$$

where the loss function assesses the quality of each explanation. The choice of Ω plays a crucial role in determining the type and quality of the explanation produced. In this paper, we broadly define Ω as the set of interpretable models, specifically focusing on Weighted Linear Regression (WLR) models. Additionally, we incorporate Shapley values to quantify the contribution of each feature or node in the spatiotemporal graph to the prediction outcome. By computing Shapley values, we ensure a fair attribution of importance among different input variables, providing a principled way to interpret model decisions. For more details on the integration of Shapley values, please refer to [10].

4. Experiment Results



(a) Lane 1 true flow



(b) Lane 1 predicted flow

Figure 3: Comparison of true and predicted flow and density across four lanes for the I-24 dataset. The distance unit is 0.1 miles and the time unit is 10 seconds.

We use macroscopic traffic data (flow, speed, and density) from I-24 [11], obtained via radar detection systems (RDS) and interpolated using the Adaptive Smoothing Method (ASM) [12]. The Tennessee Department of Transportation provides the RDS data, collected through a highway surveillance system. Aggregated at 30-second intervals, it includes lane-by-lane occupancy, vehicle counts, and speed. ASM reconstructs smooth, continuous estimates of traffic states by dynamically adjusting smoothing parameters based on local conditions [13]. The analyzed I-24 segment consists of four main lanes and on/off ramps. Due to data availability, we focus on the four main lanes. The space is divided into 0.1-mile cells and 10-second intervals, forming 684 nodes.

The prediction results are shown in Fig. 3, demonstrating the model's ability to approximate traffic behavior with high accuracy. The predicted flow patterns closely match the ground truth, capturing key variations in traffic dynamics across different lanes. This confirms that the spatiotemporal graph neural network effectively learns the underlying traffic patterns.

The explanation results are presented in Fig. 4, illustrating the spatial distribution of Shapley values among neighboring nodes, with the red node representing the one under investigation. The results indicate that the most influential neighbors are not necessarily the closest ones, underscoring the impact of road geometry on traffic breakdowns. Specifically, we observe that lane adjacency and merging zones play a significant role in congestion propagation, emphasizing the need to account for spatial dependencies when analyzing traffic breakdown precursors.

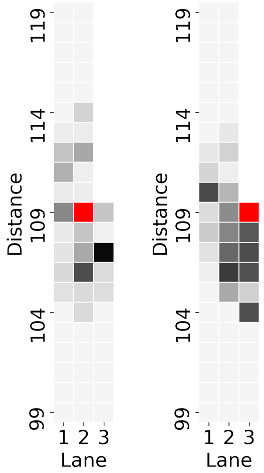


Figure 4: Spatial distributions of Shapley values among different neighbors for various SAGs. The two subfigures correspond to different investigated nodes, highlighted in red—Lane 2 in (a) and Lane 3 in (b), respectively.

5. Conclusion

In this paper, we propose a novel framework combining spatiotemporal graph neural networks (ST-GNNs) with Shapley value-based explanations to identify precursors of traffic breakdowns. Our approach extends existing GNN-based Shapley explanation methods to a spatiotemporal setting, enabling a more interpretable analysis of traffic dynamics. By applying our method to the I-24 dataset, we demonstrate its capability to uncover key contributors to traffic breakdowns, such as lane adjacency effects and abrupt braking patterns.

Our results highlight that the lane adjacent to the merging area is a major source of traffic fluctuations, often leading to congestion. Additionally, the explanation results reveal that the most influential neighboring nodes are not always the closest ones, underscoring the importance of road topology in traffic breakdown formation. Future work includes extending our model to incorporate microscopic vehicle trajectory data, improving real-time inference capabilities, and exploring broader applications in intelligent traffic management and congestion mitigation strategies.

References

[1] M. Filipovska and H. Mahmassani. Traffic flow breakdown prediction using machine learning approaches. *Transportation Research Record Journal of the Transportation Research Board*, 2674:560–570, 2020.

[2] Y. Zhao and J. Dong. Prediction of freeway traffic breakdown using artificial neural networks. *Algorithms*, 16:298, 2023.

[3] M. Mu, J. Zhang, C. Wang, J. Zhang, and C. Yang. String stability and platoon safety analysis of a new car-following model considering a stabilization strategy. *Ieee Access*, 9:111336–111345, 2021.

[4] J. Tian, B. Jia, S. Ma, C. Zhu, R. Jiang, and Y. Ding. Cellular automaton model with dynamical 2d speed-gap relation. *Transportation Science*, 51:807–822, 2017.

[5] Xiaoyang Wang, Yao Ma, Yiqi Wang, Wei Jin, Xin Wang, Jiliang Tang, Caiyan Jia, and Jian Yu. Traffic flow prediction via spatial temporal graph neural network. In *Proceedings of the web conference 2020*, pages 1082–1092, 2020.

[6] Hao Peng, Hongfei Wang, Bowen Du, Md Zakirul Alam Bhuiyan, Hongyuan Ma, Jianwei Liu, Lihong Wang, Zeyu Yang, Linfeng Du, Senzhang Wang, et al. Spatial temporal incidence dynamic graph neural networks for traffic flow forecasting. *Information Sciences*, 521:277–290, 2020.

[7] Eyal Winter. The shapley value. *Handbook of game theory with economic applications*, 3:2025–2054, 2002.

[8] Sergiu Hart. Shapley value. In *Game theory*, pages 210–216. Springer, 1989.

[9] Mukund Sundararajan and Amir Najmi. The many shapley values for model explanation. In *International conference on machine learning*, pages 9269–9278. PMLR, 2020.

[10] Alexandre Duval and Fragkiskos D Malliaros. Graphsvx: Shapley value explanations for graph neural networks. In *Machine Learning and Knowledge Discovery in Databases. Research Track: European Conference, ECML PKDD 2021, Bilbao, Spain, September 13–17, 2021, Proceedings, Part II 21*, pages 302–318. Springer, 2021.

[11] Derek Gloudemans, Yanbing Wang, Junyi Ji, Gergely Zachar, William Barbour, Eric Hall, Meredith Cebelak, Lee Smith, and Daniel B Work. I-24 motion: An instrument for freeway traffic science. *Transportation Research Part C: Emerging Technologies*, 155:104311, 2023.

[12] Martin Treiber, Arne Kesting, and R Eddie Wilson. Reconstructing the traffic state by fusion of heterogeneous data. *Computer-Aided Civil and Infrastructure Engineering*, 26(6):408–419, 2011.

[13] Junyi Ji, Yanbing Wang, Derek Gloudemans, Gergely Zachár, William Barbour, and Daniel B Work. Virtual trajectories for i–24 motion: Data and tools. In *2024 Forum for Innovative Sustainable Transportation Systems (FISTS)*, pages 1–6. IEEE, 2024.