Causal Impact Inference for Traffic Networks with Graph-Integrated Transfer Entropy

Junji Ye[a], Can Li[b,*], Fangni Zhang[a]

[a] Department of Industrial and Manufacturing Systems Engineering, The University of Hong Kong, Hong Kong, China

[b] The Key Laboratory of Road and Traffic Engineering of the Ministry of Education, Tongji University, Shanghai, China

yjj991210@gmail.com, lican@tongji.edu.cn, fnzhang@hku.hk

Extended abstract submitted for presentation at the Conference in Emerging Technologies in Transportation Systems (TRC-30)

September 02-03, 2024, Crete, Greece

April 28, 2024

Keywords: Causal Inference, Traffic Network, Transfer Entropy, Region-wise Impacts

1 INTRODUCTION

The study of causality on traffic networks has gained increasing attention in intelligent transportation systems, particularly in the context of prediction and structural issues. The causal graph plays a powerful role in congestion (Luan *et al.*, 2022) and accidents (Deng *et al.*, 2023) investigation, as it provides visual and mathematical representations of causal relations in traffic networks and reveals the propagation process of causal impacts.

Given the complexity of actual traffic network structures, previous studies have generated the causal graph solely based on information theory such as Lasso Granger Causality regression model (Li *et al.*, 2015) and transfer entropy (Chen *et al.*, 2021), resulting in limitations including imprecise mappings and unverifiable effects. To improve the precision of causal correlations, model-based approaches have been used to achieve trainable causal graph with continuity functions. However, these approaches primarily focus on detecting pair-wise mutual information, which may not adequately capture the complex relationships in real traffic networks. Furthermore, all previous methods discard the structural characteristics of the network topology graph, focusing only on the temporal correlations. Thus, architectural information on traffic topology is lost, hindering a more comprehensive understanding of the causal relations.

To address the limitations, we provide a new perspective to achieve the region-wise causal inference of complex traffic network with enhanced verifiability and interpretability. Specifically, the initial causal matrix constructed by trainable transfer entropy captures the region-wise causal effect by information aggregation within the designed causal-integrated Graph Convolutional Network (GCN). The improvement of traffic flow forecasting accuracy evaluated on real-world dataset using causal graph indicate the effectiveness of the proposed method in capturing causal relations. Furthermore, we perform comparisons between topology and causal graph, where the visualization of the asymmetric causal graph provides the understanding of traffic dynamics and underlying causal impacts.

2 METHODOLOGY

In the proposed model as shown in Figure 1, trainable transfer entropy is applied to build the initial causal matrix. Subsequently, this matrix is iteratively refined through a GCN-integrated ap-

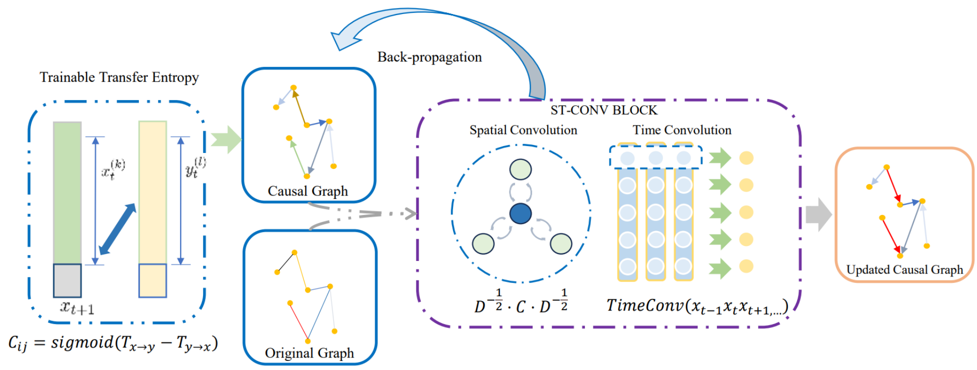


Figure 1 – *The overall framework of the proposed model*

proach, employing end-to-end training to enhance the precision of traffic flow forecasting thereby optimizing the matrix representation. To capture the region-wise causal impacts in traffic networks, the refined causal matrix is constructed through spatial convolution layers that aggregate spatiotemporal information from neighbors of each node as defined by the adjacency matrix.

2.1 Trainable Transfer Entropy Algorithm

Transfer Entropy is an extension of information entropy to quantify the transfer of information and reveal the impacts between nodes in the traffic topology. Combining the discrete information entropy $h(X) = -\sum_x p(x)\log p(x)$ and conditional probabilities, transfer entropy (Schreiber, 2000) at time step t is defined as:

$$T_{y\to x} = \sum_t p\left(x_{t+1}, X_{t\sim(t-k+1)}, Y_{t\sim(t-l+1)}\right) \log \frac{p\left(x_{t+1} \mid X_{t\sim(t-k+1)}, Y_{t\sim(t-l+1)}\right)}{p\left(x_{t+1} \mid X_{t\sim(t-k+1)}\right)} \tag{1}$$

where $Y_{t\sim(t-l+1)}$ and $X_{t\sim(t-k+1)}$ stand for the past l and k time steps' traffic flow of node Y and node X in the topology graph, respectively. $T_{y\to x}$ measures the impact of $Y_{t\sim(t-l+1)}$ to x_{t+1} in the presence of $X_{t\sim(t-k+1)}$. As joint probability provides a more comprehensive representation of the probability distribution, the conditional probability is transformed into its corresponding joint form, where Eq. (1) can be formulated as:

$$T_{y\to x} = \sum_t p\left(x_{t+1}, X_{t\sim(t-k+1)}, Y_{t\sim(t-l+1)}\right) \log \frac{p\left(x_{t+1}, X_{t\sim(t-k+1)}, Y_{t\sim(t-l+1)}\right) p\left(X_{t\sim(t-k+1)}\right)}{p\left(x_{t+1}, X_{t\sim(t-k+1)}\right) p\left(X_{t\sim(t-k+1)}, Y_{t\sim(t-l+1)}\right)} \tag{2}$$

To address the discrete nature of the node's value and inconsistent dimensions between two nodes, a novel way to estimate the joint probability density of x_t and $Y_{m\sim(m-l+1)}$ is proposed as:

$$\widehat{p}\left(x_t, Y_{t\sim(t-l+1)}\right) = \frac{1}{N}\sum_{m\neq t}^{N} Sigmoid\left(\lambda - \sqrt{(x_m - x_t)^2} - \sqrt{\sum_l (Y_{m\sim(m-l+1)} - Y_{t\sim(t-l+1)})^2}\right) \tag{3}$$

where m refers to all evaluative time steps. $\sqrt{(x_m - x_t)^2}$ and $\sqrt{\sum_l (Y_{m\sim(m-l+1)} - Y_{t\sim(t-l+1)})^2}$ represent $L2$ distance between all pairs across N time steps and λ is the threshold judging the possibility. $Sigmoid$ is an activation function that enables the update ability of λ via back-propagation and adds non-linear characteristics. If the $L2$ distance between the target $(x_t, Y_{t\sim(t-l+1)})$ and the current evaluative sample $(x_m, Y_{m\sim(m-l+1)})$ is less than the threshold λ, the target is potentially related to the evaluative sample. The final joint probability density is obtained by the sum of all evaluative pairs.

Thus, the transfer entropy $T_{y\to x}$ and $T_{x\to y}$ between any two nodes on the overall timeline are calculated to characterize the strength of directed causal impact based on the joint probability

Table 1 – ***PeMSD4*** *Comparison Result*

Graph Structure Type	MAE	RMSE	MAPE
Original Road Network	24.54	38.02	17.97%
causal Graph Three-Hop	24.38	37.81	17.87%
causal Graph Two-Hop	**21.01**	**33.03**	**14.44%**
causal Graph One-Hop	22.02	34.83	14.94%

density, where the nodes in the causal graph are defined as $\mathcal{C}_{xy} = \mathrm{S}igmoid(T_{y\to x} - T_{x\to y})$. According to the properties of the sigmoid function, $\mathcal{C}_{xy} > 0$ when $T_{y\to x} > T_{x\to y}$, which indicates the causal impact from y to x. To preserve the information on topology to a maximum extent and make comparisons between distinct causal graphs, we establish an initial causal graph among one-hop neighbors, two-hop neighbors and three-hop neighbors of nodes in the topology graph.

2.2 Causal-Integrated Graph Convolutional Network

Causal-Integrated Graph Convolutional Network (CIGCN) is proposed to analyze spatiotemporal correlations of traffic networks. CIGCN applies convolutional operations in spatial and temporal dimensions simultaneously, with emphasis on the attention mechanism. As shown in Figure 1, the spatial convolution is conducted on the causal graph derived from Subsection 2.1 instead of the original graph obtained from transportation topology. By integrating the threshold parameters of transfer entropy into overall parameters of GCN, the whole model can be end-to-end differentiable, allowing for continuous updates to the causal graph during training.

Following the principle of light graph convolution (He *et al.*, 2020), we transform the acquired causal matrix $C \in \mathbf{R}^{N\times N}$ into normalized form and define $\mathbf{E}_{T_r} \in \mathbf{R}^{N\times \mathbf{T}_r}$ as the historical flow of N nodes at time steps T_r. Then, the spatial convolution is defined as:

$$\mathbf{H}_{T_r} = \mathrm{Re}\,LU\left(\mathbf{D}^{-\frac{1}{2}}\mathcal{C}\mathbf{D}^{-\frac{1}{2}}\right)\mathbf{E}_{T_r} \tag{4}$$

where $\mathbf{D} \in \mathbf{R}^{\mathbf{N}\times\mathbf{N}}$ is the diagonal matrix and $\mathbf{H}_{T_r}$ represents the updated underlying features corresponding to the flow of N nodes over time steps T_r. $\mathrm{Re}\,LU$ is adopted as the activation function for negative outputs which is beneficial for noise removal. Based on the form of spatial convolution, temporal convolution can be formulated as:

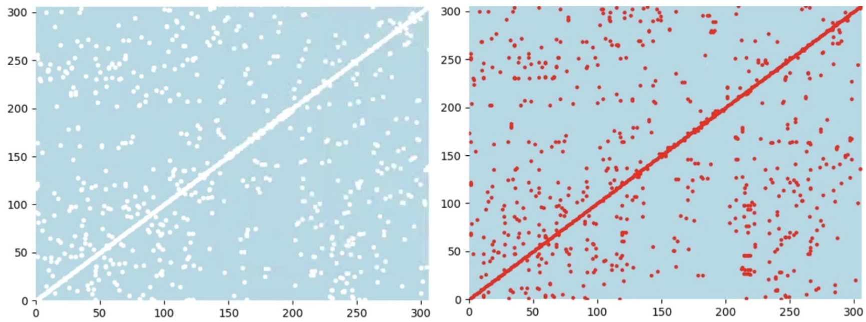
$$\hat{\mathbf{E}}_{\boldsymbol{T_r+1}} = \mathrm{Re}\,LU\left(\Phi * \left(\mathrm{Re}\,LU\left(\mathbf{D}^{-\frac{1}{2}}\mathcal{C}\mathbf{D}^{-\frac{1}{2}}\right)\mathbf{E}_{T_r}\right)\right) \tag{5}$$

where $\hat{\mathbf{E}}_{T_r+1} \in \mathbf{R}^{N\times 1}$ indicates the predicted result of next time step over historical time steps T_r, and Φ is the convolution kernel used to capture information over T_r time steps.

3 RESULTS

Dataset and experimental setting. The proposed model is evaluated on the real-world traffic dataset **PeMSD4**, which comprises real-time traffic flow data collected every 30 seconds from highways containing 307 sensors on 29 roads in the San Francisco Bay Area, California. The original dataset is aggregated into five-minute points, and missing values are filled by linear interpolation. The historical time steps of $(x_t^{(k)}, y_t^{(l)})$ is defined as 10. Additionally, the causality computation is conducted on one, two, and three-hop neighbors.

Experimental results and analysis. To validate the efficiency of the proposed causal graph, a comparative analysis is conducted between the causal graph and topology structure graph in the spatial convolution process defined as Eq. (4). The traffic forecasting results presented in Table 1 underscore the traffic forecasting improvement by the causal graph, which highlights the efficacy of incorporating the causal graph in enhancing the performance of CIGCN. Besides, the two-hop causal matrix achieves the best prediction performance. From a structural perspective, it aggregates two-hop neighbors with strong causality in the subsequent convolution process, making it less susceptible to the influence of noise. The three-hop causal matrix shows less satisfactory performance due to the excessive information from distant neighbors, which implies including too many neighbors in the analysis can have negative effects. Additionally, Figure 2 visualizes the topology graph and two-hop causal graph. In contrast to the symmetric topology matrix representing bidirectional relationships, the asymmetry of the derived causal matrix indicates the distinct and unidirectional correlations between each pair of nodes, which provides the foundation for understanding of traffic dynamics and analysis of underlying causal relations.



(a) *Topology matrix* (b) *Two-hop causal matrix*

Figure 2 – *Visualization of traffic topology graph and causal graph*

4 DISCUSSION

This study examines the potential of GCN-integrated end-to-end training methods based on trainable transfer entropy to derive region-wise causal graphs with strong interpretability of causal impacts. The superior traffic flow forecasting performance with the obtained causal graph illustrates that graph-structured learning is more controllable and interpretable to achieve causality detection, which can be wildly used for bottleneck identification as well as accidents and congestion prediction. Our future work will focus on further enhancing practical application capabilities of causal graph models from two aspects: 1) Complexity Optimization - fastening training speed and reducing time consumption caused by causal graph reconstruction. 2) Structure Analysis - Conducting structure mining through causal graphs in fields including bottleneck identification or excavation of vital network architecture, which further verifies the correctness of the causal graph.

References

Chen, Yuanyi, Yan, Pengquan, Zheng, Zengwei, & Chen, Dan. 2021. Identifying Traffic Bottleneck in Urban Road Networks via Causal Inference. *Security, Privacy, and Anonymity in Computation, Communication, and Storage*, 372–383.

Deng, Yongxin, Liu, Shiyan, & Zhou, Dong. 2023. Dependency Cluster Analysis of Urban Road Network Based on Percolation. *Transportation Research Part C: Emerging Technologies*, **154**, 104264.

He, Xiangnan, Deng, Kuan, Wang, Xiang, Li, Yan, Zhang, Yongdong, & Wang, Meng. 2020. Lightgcn: Simplifying and Powering Graph Convolution Network for Recommendation. *Proceedings of International ACM SIGIR Conference on Research and Development in Information Retrieval*, 639–648.

Li, Li, Su, Xiaonan, Wang, Yanwei, Lin, Yuetong, Li, Zhiheng, & Li, Yuebiao. 2015. Robust Causal Dependence Mining in Big Data Network and Its Application to Traffic Flow Predictions. *Transportation Research Part C: Emerging Technologies*, **58**, 292–307.

Luan, Sen, Ke, Ruimin, Huang, Zhou, & Ma, Xiaolei. 2022. Traffic Congestion Propagation Inference Using Dynamic Bayesian Graph Convolution Network. *Transportation Research Part C: Emerging Technologies*, **135**, 103526.

Schreiber, Thomas. 2000. Measuring Information Transfer. *Physical Review Letters*, **85**(2), 461.