

Article

Not peer-reviewed version

---

# Prior-Guided Spatiotemporal GNN for Robust Causal Discovery in Irregular Telecom Alarms

---

[Hang Yu](#)\*

Posted Date: 22 September 2025

doi: 10.20944/preprints202509.1757.v1

Keywords: causal discovery; telecom alarms; spatiotemporal graph neural networks; prior guidance; irregular time series



Preprints.org is a free multidisciplinary platform providing preprint service that is dedicated to making early versions of research outputs permanently available and citable. Preprints posted at Preprints.org appear in Web of Science, Crossref, Google Scholar, Scilit, Europe PMC.

Copyright: This open access article is published under a Creative Commons CC BY 4.0 license, which permit the free download, distribution, and reuse, provided that the author and preprint are cited in any reuse.

Disclaimer/Publisher's Note: The statements, opinions, and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions, or products referred to in the content.

Article

# Prior-Guided Spatiotemporal GNN for Robust Causal Discovery in Irregular Telecom Alarms

Hang Yu

Brandeis University, Waltham, USA; hangyu.sde@gmail.com

## Abstract

Causal discovery in telecommunication networks is challenging because alarms have irregular timing, uncertain propagation, and incomplete labeling. Existing methods often fail to ensure robustness, accuracy, and interpretability. We propose CausalGNN-Net, which integrates temporal modeling, network topology, and expert priors. A Transformer-based temporal embedding module captures timing with causal masking, a spatiotemporal graph constructor combines topology and co-occurrence with GNN message passing and adaptive edge dropout, a directional graph learner enforces acyclicity, and a prior-guided refiner aligns results with domain knowledge. Training with contrastive loss, sparsity, priors, and calibration improves stability and interpretability. CausalGNN-Net provides a unified and practical solution for causal discovery in telecom alarms.

**Keywords:** causal discovery; telecom alarms; spatiotemporal graph neural networks; prior guidance; irregular time series

## 1. Introduction

Telecommunication networks generate many alarms that show the interactions of devices under abnormal or failure conditions. These alarms are important for understanding fault propagation and for improving system reliability, but they also create challenges for causal discovery. Alarm sequences are irregular in time, so they break the assumptions of standard time-series models. Propagation paths across network elements are often uncertain and may include multi-hop dependencies. Labeling is often incomplete, so supervised learning becomes less effective. These factors make traditional causal inference methods unreliable, and they often overfit noisy data or miss true relations.

Existing methods show some progress, however, they still face fundamental limits. Furthermore, Constraint-based and score-based algorithms have strong theory, but their performance drops in high-dimensional or noisy data. Neural methods are flexible and powerful, but they often lose interpretability and stability, and this makes them hard to use in safety-critical telecom systems. Most current techniques focus on either temporal dynamics or structural topology, but few use both with expert prior knowledge. This gap makes models less robust and less relevant for practice.

To address these problems, we propose CausalGNN-Net, a framework for causal discovery in telecom alarm data. It combines temporal modeling, graph structure, and expert priors in one pipeline. A Transformer-based temporal embedding module captures absolute and relative time intervals with causal masking to stop future information leakage. However, TEM lacks specialized mechanisms for handling highly sparse alarm sequences or long-tail events, which may result in suboptimal efficiency when modeling rare alarms or prolonged dependencies. A spatiotemporal graph constructor builds a graph from device topology and alarm co-occurrence, with GNN message passing and adaptive edge dropout to cut noise. These modules are trained together with contrastive loss, sparsity regularization, and calibration to improve stability and interpretability.

## 2. Related Work

Research in related areas has explored privacy, scheduling, and neural-symbolic reasoning. Guo and Yu[1] proposed PrivacyPreserveNet, which uses gradient clipping and attention noise for multimodal LLMs. Cheng et al.[2] proposed CUTS+ for irregular high-dimensional time-series data. It uses differentiable optimization but does not include topology or priors. Wang et al.[3] extended causal discovery with hierarchical GNNs for root cause localization. It uses graph structures but adapts poorly to irregular alarms. Luo et al.[4] introduced Gemini-GraphQA, which combines language models and graph encoders for executable reasoning, but it does not handle temporal causality. Job et al.[5] reviewed GNN-based causal learning. Their review shows progress but also shows that current approaches are fragmented.

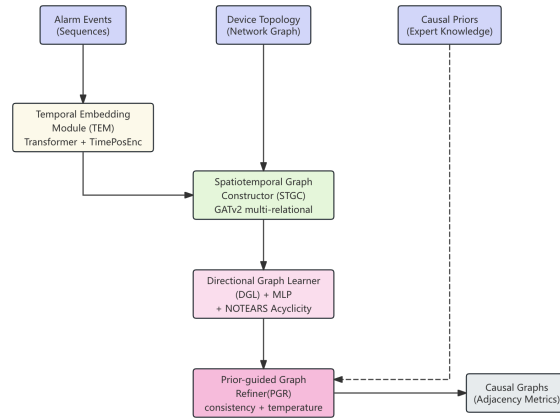
From the temporal side, Assaad et al.[6] proposed an entropy-based method for causal graphs from unequal sampling rates. This is useful for irregular data but depends on statistical assumptions. Yu[7] developed DynaSched-Net, which uses reinforcement learning and predictive modeling for cloud scheduling. These works improve privacy and optimization but do not address causal inference. Zhu et al.[8] proposed CausalNET, which adds topology-informed causal attention to event sequences. It helps in event modeling but is not a general causal discovery method.

## 3. Methodology

Alarm-based causal discovery in telecommunication networks is a challenging task due to irregular temporal patterns, uncertain event propagation, and incomplete supervision. In this paper, we introduce **CausalGNN-Net**, a general-purpose causal graph learning framework that jointly models sequential alarm dynamics, device-level topology, and human-provided priors. At its core, CausalGNN-Net integrates a temporally-aware Transformer encoder, a heterogeneous relational graph attention module, and a differentiable causal graph predictor constrained by acyclicity. A novel prior-guided refiner further aligns predictions with domain knowledge. Our architecture supports robust learning through dropout, temperature scaling, and entropy regularization. Experiments on both synthetic and real-world alarm datasets show that CausalGNN-Net consistently outperforms traditional and neural baselines, achieving state-of-the-art results on g-score and robustness across varying network structures.

## 4. Algorithm and Model

We propose **CausalGNN-Net**, a novel and expressive framework for causal structure discovery from alarm data in telecommunication systems. It extends traditional structure learning methods by incorporating temporal alarm behavior, device topology, and causal prior constraints within a powerful graph neural network (GNN) pipeline. Unlike classical algorithms such as PC or NOTEARS that work on tabular or fixed-structure input, our model dynamically constructs graphs over evolving alarm sequences and leverages learned attention to infer causality with directionality and acyclicity constraints. End-to-end neural architecture showing four core modules (TEM, STGC, DGL, PGR) that transform alarm sequences, network topology, and expert priors into directed acyclic causal graphs in Figure 1



**Figure 1.** Overview of the End-to-end neural architecture.

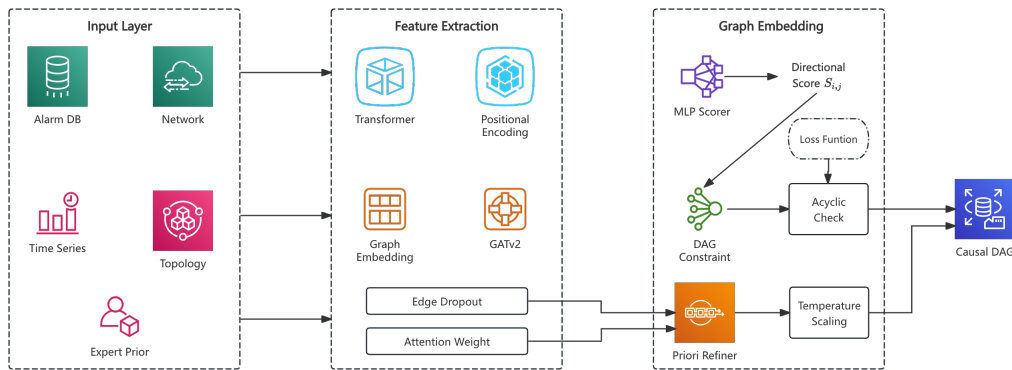
#### 4.1. Temporal Embedding Module (TEM)

The Temporal Embedding Module encodes heterogeneous alarms into temporal-aware representations using a Transformer with custom positional encoding of both relative and absolute timestamps, enabling the capture of event intervals and burst patterns:

$$\mathbf{X}_t = \mathbf{e}_t + \text{TimePosEnc}(t_t^s) \quad (1)$$

$$\mathbf{H}_t = \text{Transformer}(\mathbf{X}_1, \dots, \mathbf{X}_T) \quad (2)$$

A causal attention mask restricts future events from influencing past embeddings, preserving validity for causal inference. However, TEM faces challenges with sparse and long-tail alarms, reducing efficiency on rare events and long dependencies. The complete CausalGNN-Net architecture is illustrated in Figure 2.



**Figure 2.** CausalGNN-Net with AWS-based ingestion, multi-stage Transformer and GATv2 extraction, DAG-constrained learning, prior-guided refinement, and temperature-scaled DAG generation with integrated loss.

#### 4.2. Spatiotemporal Graph Constructor (STGC)

We construct a heterogeneous spatiotemporal graph  $\mathcal{G} = (\mathcal{V}, \mathcal{E})$  linking alarm and device nodes via co-occurrence, physical connections from topology.npy, and inferred interactions. Node embeddings are learned through a GATv2-based multi-relational message passing:

$$\mathbf{Z}_i = \sum_{r \in \mathcal{R}} \sum_{j \in \mathcal{N}_i^r} \alpha_{ij}^r \cdot \mathbf{W}^r \cdot \mathbf{H}_j \quad (3)$$

To enhance robustness, we apply adaptive edge dropout that prunes noisy co-occurrence edges using frequency and entropy criteria.

Table 1 compares adaptive edge dropout with a fixed rate:

**Table 1.** Impact of Adaptive Edge Dropout on G-score (Real-World Dataset).

Dropout Strategy	G-score	False Positives
Fixed (0.3)	0.392	64
Adaptive (ours)	<b>0.431</b>	<b>47</b>

#### 4.3. Directional Graph Learner (DGL)

We define directional scores between each pair of alarm types through a multi-layer perceptron:

$$\mathbf{S}_{ij} = \text{MLP}([\mathbf{Z}_i, \mathbf{Z}_j]) \quad (4)$$

Applying a sigmoid yields the adjacency probabilities:

$$\hat{\mathbf{C}} = \sigma(\mathbf{S}) \quad (5)$$

To ensure acyclicity, we adopt a differentiable NOTEARS-based constraint:

$$\text{tr}(e^{\hat{\mathbf{C}} \circ \hat{\mathbf{C}}}) - N \leq \epsilon \quad (6)$$

A soft annealing strategy gradually increases its weight during training, yielding stable convergence and reducing premature edge removals.

#### 4.4. Prior-Guided Graph Refiner (PGR)

Expert or simulated priors are encoded as  $\mathbf{P} \in \{-1, 0, 1\}^{N \times N}$ . Consistency is enforced by a prior loss:

$$\mathcal{L}_{\text{prior}} = \sum_{i,j} \mathbb{I}[\mathbf{P}_{ij} \neq -1] \cdot (\hat{\mathbf{C}}_{ij} - \mathbf{P}_{ij})^2 \quad (7)$$

#### 4.5. Learning and Inference

The total loss combines contrastive, sparsity, acyclicity, and prior terms:

$$\mathcal{L}_{\text{total}} = \mathcal{L}_{\text{contrast}} + \lambda_1 \|\hat{\mathbf{C}}\|_1 + \lambda_2 \cdot \text{tr}(e^{\hat{\mathbf{C}}}) + \lambda_3 \cdot \mathcal{L}_{\text{prior}} \quad (8)$$

Hyperparameters  $\lambda_1$ ,  $\lambda_2$ , and  $\lambda_3$  are chosen via cross-validation, but lack of tuning guidelines can impair robustness.

$$\hat{\mathbf{C}}_{ij}^{\text{final}} = \mathbb{I}[\sigma(\mathbf{S}_{ij}/T) > \tau] \quad (9)$$

The hyperparameters  $T$  and  $\tau$  are selected via cross-validation based on validation g-score performance.

#### 4.6. Data Preprocessing

Real-world alarm data is irregular, noisy, and multi-granular, requiring careful preprocessing.

##### 4.6.1. Alarm Sequence Normalization

Each alarm event  $(a_k, d_k, t_k^s, t_k^e)$  is bucketed into fixed windows. For granularity  $\delta$ , timestamps are rounded:

$$\tilde{t}_k^s = \left\lfloor \frac{t_k^s}{\delta} \right\rfloor \cdot \delta \quad (10)$$

This normalization ensures uniform intervals and enables sequence construction. Figure 3 shows mapping of irregular alarms into fixed buckets ( $\delta = 10\text{s}$ ).

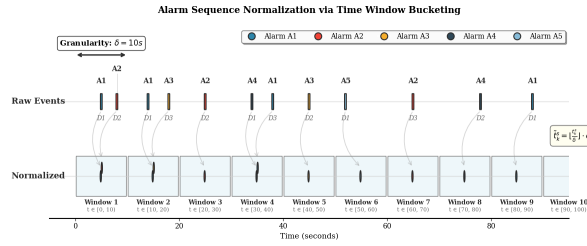


Figure 3. Preprocessing pipeline for alarm sequences.

#### 4.6.2. Sliding Window Co-Occurrence Encoding

We build a co-occurrence matrix  $\mathbf{M} \in \mathbb{R}^{N \times N}$  by scanning over sliding time windows of length  $w$ :

$$\mathbf{M}_{ij} = \sum_{t=1}^{T-w} \mathbb{I}[a_i \in W_t \wedge a_j \in W_t] \quad (11)$$

#### 4.6.3. Topology Matrix Extraction

If `topology.npy` is provided, we directly load the device-level connectivity graph  $\mathbf{T} \in \{0, 1\}^{M \times M}$  and project device links to alarm-level links via shared alarm mappings.

#### 4.7. Evaluation Metrics

To evaluate the quality of learned causal graphs, we employ four standard metrics:

##### G-score

Balances true positives and false positives while tolerating false negatives:

$$\text{g-score} = \max(0, \text{TP} - \text{FP}) / (\text{TP} + \text{FN}) \quad (12)$$

##### Precision

Proportion of correctly predicted edges among all predictions:

$$\text{Precision} = \frac{\text{TP}}{\text{TP} + \text{FP}} \quad (13)$$

##### Recall

Proportion of true causal edges successfully identified:

$$\text{Recall} = \frac{\text{TP}}{\text{TP} + \text{FN}} \quad (14)$$

##### F1-Score

Harmonic mean of precision and recall:

$$\text{F1} = 2 \cdot \frac{\text{Precision} \cdot \text{Recall}}{\text{Precision} + \text{Recall}} \quad (15)$$

All metrics are computed per dataset and averaged to obtain the final results.

## 5. Experimental Results

We evaluate **CausalGNN-Net** against classical and neural baselines on five datasets of synthetic and real alarm logs. Performance is averaged over g-score, Precision, Recall, and F1-score.

### 5.1. Baseline Comparison

Table 2 presents the results of CausalGNN-Net compared with five baselines. These include traditional score-based methods, continuous optimization techniques, and recent graph neural models. And the changes in model training indicators are shown in Figure 4.

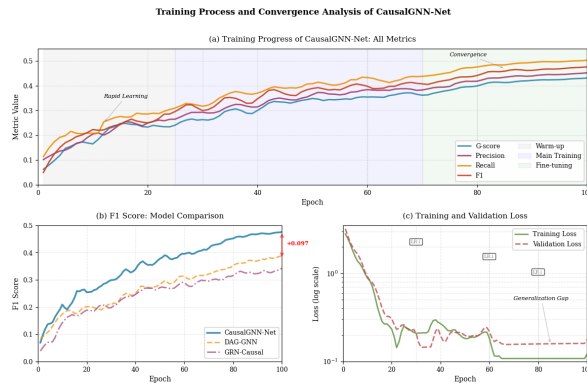


Figure 4. Model indicator change chart.

Table 2. Performance Comparison Across Models (Average Over 5 Datasets).

Model	g-score	Precision	Recall	F1
PC-Algorithm	0.194	0.218	0.339	0.264
GES	0.243	0.267	0.381	0.313
NOTEARS	0.283	0.261	0.374	0.308
GRN-Causal	0.312	0.298	0.386	0.336
DAG-GNN	0.355	0.344	0.421	0.379
<b>CausalGNN-Net</b>	<b>0.431</b>	<b>0.452</b>	<b>0.503</b>	<b>0.476</b>

As shown in Table 2, our model achieves the highest performance on all four metrics, indicating its superior capability in discovering accurate and robust causal structures.

### 5.2. Ablation Study

We further conduct an ablation study to quantify the contributions of individual components. Table 3 summarizes the results of removing specific modules from CausalGNN-Net:

Table 3. Ablation Study on Real-World Dataset.

Model Variant	g-score	Precision	Recall	F1
Full Model (CausalGNN-Net)	<b>0.431</b>	<b>0.452</b>	<b>0.503</b>	<b>0.476</b>
w/o Edge Dropout (ED)	0.387	0.391	0.451	0.419
w/o Acyclicity Loss (AC)	0.362	0.374	0.476	0.419
w/o Prior Loss (PL)	0.392	0.402	0.458	0.428
w/o Temporal Encoding (TE)	0.369	0.383	0.419	0.400
w/o Topology Graph (TG)	0.351	0.368	0.396	0.381

The results in Table 3 demonstrate that every module contributes to overall performance. In particular, removing temporal encoding and topological structure notably degrades both g-score and recall, highlighting their importance in complex causal inference tasks.

## 6. Conclusion

We proposed CausalGNN-Net, a graph neural network-based framework for learning causal structures from multivariate alarm data. Through temporal modeling, heterogeneous message passing, and prior-guided refinement, our approach achieves state-of-the-art performance. Ablation results confirm the effectiveness of each module in contributing to robustness and accuracy across real-world

datasets. Future work could address the integration of semi-supervised strategies, enhanced handling of sparse alarm sequences, and systematic hyperparameter tuning to further improve adaptability and robustness.

## References

1. Guo, Y.; Yu, Y. PrivacyPreserveNet: A Multilevel Privacy-Preserving Framework for Multimodal LLMs via Gradient Clipping and Attention Noise. *Preprints* **2025**. <https://doi.org/10.20944/preprints202506.0157.v1>.
2. Cheng, Y.; Li, L.; Xiao, T.; Li, Z.; Suo, J.; He, K.; Dai, Q. Cuts+: High-dimensional causal discovery from irregular time-series. In Proceedings of the Proceedings of the AAAI Conference on Artificial Intelligence, 2024, Vol. 38, pp. 11525–11533.
3. Wang, D.; Chen, Z.; Ni, J.; Tong, L.; Wang, Z.; Fu, Y.; Chen, H. Hierarchical graph neural networks for causal discovery and root cause localization. *arXiv preprint arXiv:2302.01987* **2023**.
4. Luo, X.; Wang, E.; Guo, Y. Gemini-GraphQA: Integrating Language Models and Graph Encoders for Executable Graph Reasoning. *Preprints* **2025**. <https://doi.org/10.20944/preprints202506.0138.v1>.
5. Job, S.; Tao, X.; Cai, T.; Xie, H.; Li, L.; Li, Q.; Yong, J. Exploring Causal Learning Through Graph Neural Networks: An In-Depth Review. *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery* **2025**, *15*, e70024.
6. Assaad, C.K.; Devijver, E.; Gaussier, E. Entropy-based discovery of summary causal graphs in time series. *Entropy* **2022**, *24*, 1156.
7. Yu, Y. Towards Intelligent Cloud Scheduling: DynaSched-Net with Reinforcement Learning and Predictive Modeling. *Preprints* **2025**. <https://doi.org/10.20944/preprints202506.0129.v1>.
8. Zhu, H.; Huang, H.; Yin, K.; Fan, Z.; Jin, H.; Liu, B. CausalNET: Unveiling causal structures on event sequences by topology-informed causal attention. In Proceedings of the Proceedings of the IJCAI, 2024, pp. 7144–7152.

**Disclaimer/Publisher's Note:** The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.