From Knowledge to Causality: Self-Supervised Representation Learning for Granger Causal Discovery in Groups of Time Series

Extended Abstract

Bo Liu
Peking University
Beijing, China
liubo2022@stu.pku.edu.cn

Hongyan Li*
Peking University
Beijing, China
leehy@pku.edu.cn

Shenda Hong*
Peking University
Beijing, China
hongshenda@pku.edu.cn

ABSTRACT

Causal inference among groups of time series is crucial for understanding complex systems like brain networks and climate dynamics. Existing methods often rely on simple aggregation to represent groups, leading to significant information loss and suboptimal causal discovery. We propose CausalKGR, a novel framework that learns expressive latent representations for Granger causal discovery. By introducing a Knowledge-Conditional Attention mechanism, CausalKGR distills temporal knowledge into a compact latent space, capturing both intra-group dynamics and inter-group interactions. Extensive experiments on synthetic and real-world datasets demonstrate that CausalKGR significantly outperforms state-of-the-art baselines in accuracy and interpretability.

KEYWORDS

Granger Causal Discovery; Group of Representation Learning

ACM Reference Format:
Bo Liu, Hongyan Li, and Shenda Hong. 2026. From Knowledge to Causality: Self-Supervised Representation Learning for Granger Causal Discovery in Groups of Time Series: Extended Abstract. In *Proc. of the 25th International Conference on Autonomous Agents and Multiagent Systems (AAMAS 2026), Paphos, Cyprus, May 25 – 29, 2026*, IFAAMAS, 3 pages. https://doi.org/10.65109/CZZR3069

1 INTRODUCTION

Understanding causal relationships among groups of time series (TS) is critical in complex systems such as geography [18], genetics [17], and biology [24]. In many real-world applications, causal interactions occur at the "group" granularity rather than between individual time series. For instance, in neuroscience, fMRI-derived Regions of Interest represent functionally homogeneous voxels [2, 11], and analyzing collective behavior across these regions offers more interpretable insights than micro-level graphs [22].

However, effectively capturing these group dynamics remains a challenge. Existing methods generally fall into two categories: condensing micro-level causal graphs (which is computationally expensive) or treating groups as atomic units using vectorized operations. These approaches suffer from two critical limitations leading to information bottlenecks:

1. Group Knowledge Loss. A group consisting of p_i time series is often reduced to a single sequence using linear aggregation methods like mean aggregation [9] or Principal Component Analysis (PCA) [25]. These unsupervised techniques frequently discard crucial group-specific information and struggle to capture complex, nonlinear intra-group dependencies.

2. Causal Knowledge Loss. In the context of Granger causality, relationships are inferred via predictive modeling within a specific look-back window. Critical causal features that fall outside this window are often excluded, compromising the analysis [4, 19]. Conversely, including all prior temporal points to capture these features can prohibitively increase computational costs and obscure the most relevant causal signals [6, 10].

To address these challenges, we propose CausalKGR, a novel framework that integrates group representation learning with Granger causal discovery. Unlike previous methods, CausalKGR does not simply compress data; it distills temporal knowledge into a high-dimensional latent space optimized for reconstructing the system's dynamics. We introduce a Knowledge-Conditional Attention mechanism to enhance the expression of key causal features. CausalKGR ensembles two rounds of self-supervised learning—encoding local intra-group dynamics and fusing global inter-group interactions—to generate representations that serve as compact, expressive conditions for identifying interpretable causal graphs.

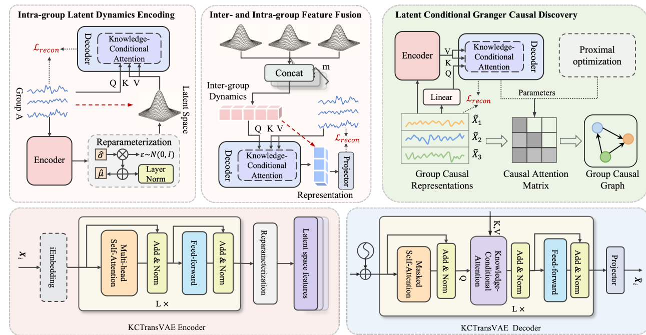

Figure 1: Illustration of the network architecture of the CausalKGR.

2 PROBLEM DEFINITION

Consider m groups of multivariate time series, where the i-th group contains p_i observed time series: $X_i = (X_{i,1}, \ldots, X_{i,p_i}) \in \mathbb{R}^{T\times p_i}$.

*Corresponding Authors.

Proc. of the 25th International Conference on Autonomous Agents and Multiagent Systems (AAMAS 2026), C. Amato, L. Dennis, V. Mascardi, J. Thangarajah (eds.), May 25 – 29, 2026, Paphos, Cyprus. © 2026 International Foundation for Autonomous Agents and Multiagent Systems (www.ifaamas.org). https://doi.org/10.65109/CZZR3069

The objective is twofold: (1) Learn a high-level representation $\tilde{X}_i$ for each group. (2) Infer a directed group-level Granger causal graph $G = (V, E)$ among these representations [3].

Group j is said to Granger-cause group i if past of $\tilde{X}_j$ provide significant predictive information about $\tilde{X}_i$ beyond other groups.

3 THE CAUSALKGR FRAMEWORK

CausalKGR consists of three key components: local dynamics encoding, global interactions fusion, and knowledge-conditional Granger causal discovery, as illustrated in Fig. 1.

3.1 Local Dynamics Encoding

To capture the specific dynamics within each group, we employ a Transformer-based Variational Autoencoder. The encoder first embeds the p_i TS of group i into temporal tokens and maps them to a latent-space distribution, sampling latent variables z via the reparameterization trick to encode intra-group dynamics.

A critical innovation in the decoding phase is the Knowledge-Conditional Attention (KCA) mechanism. Unlike standard self-attention [20], KCA explicitly utilizes the learned latent variables as a "knowledge condition." Specifically, the Keys ($\widetilde{K}$) and Values ($\widetilde{V}$) are derived from the latent variables z, while the Queries (Q) are projected from the input embeddings [8, 13]. This forces the model to reconstruct the group dynamics based heavily on the condensed latent knowledge: $\text{KCA}_{local}(Q, \widetilde{K}, \widetilde{V}) = \text{Attention}(Q, \mathbf{z}_K^\top, \mathbf{z}_V)$. $Q \in \mathbb{R}^{T \times d}$ are the original input embeddings participating in self-attention; $\widetilde{K} = \mathbf{z}_K, \widetilde{V} = \mathbf{z}_V \in \mathbb{R}^{1 \times d}$ are sampled from the latent space of each corresponding group.

3.2 Global Interactions Fusion

Solely encoding local dynamics is insufficient for capturing complex system-wide behaviors. To address this, we introduce a *Global Interactions Fusion* module that integrates inter-group dependencies by aggregating latent features from all m groups. Specifically, we sample latent variables $\mathbf{z}^i$ from the m local latent spaces to construct global context matrices $\mathbf{Z}_K = (\mathbf{z}_K^1, \cdots, \mathbf{z}_K^m)$ and $\mathbf{Z}_V = (\mathbf{z}_V^1, \cdots, \mathbf{z}_V^m) \in \mathbb{R}^{m \times d}$. These global features are fused with the local group features via a global KCA layer. In this formulation, the attention mechanism concatenates the global latent context with the local representations to guide the reconstruction: $\text{KCA}_{global}(Q, K, V) = \text{Attention}(Q, [\mathbf{Z}_K^\top; K], [\mathbf{Z}_V; V])$. $[\cdot;\cdot]$ denotes row-wise concatenation. The network is optimized via a composite loss function $\mathcal{L}_{global}$ that combines reconstruction error (MSE and L_1 norm), a Fast Fourier Transform [15] loss to capture frequency-domain discrepancies, and KL divergence for latent space regularization [12].

3.3 Knowledge-Conditional Granger Causal Discovery

Once expressive representations are learned, we infer Granger causality by modeling the temporal evolution of group representations[cite: 116]. We construct a nonlinear attention-based autoregressive predictor g_i for each group i: $\tilde{X}_i^t = g_i(\tilde{X}_1^{<t}, \ldots, \tilde{X}_m^{<t}) + e_i^t$. To ensure the resulting causal graph is sparse and interpretable, we impose a Group Lasso penalty on the input weight matrices $\mathbf{W}^0$ of the attention mechanism [19]. The training objective $\mathcal{L}_{CausalKGR}$ minimizes the prediction error regularized by the column-wise L_2 norms of the input weights: $\mathcal{L}_{CausalKGR} = \sum_t \|\tilde{X}_i^t - \hat{X}_i^t\|_2^2 + \lambda \sum_{j=1}^m \|\mathbf{W}_{:j}^0\|_2$. We optimize this objective via proximal gradient descent [16]. If the proximal operator shrinks the column norm $\|\mathbf{W}_{:j}^0\|_2$ to zero, group j is deemed Granger non-causal to group i. Notably, CausalKGR is a general neural framework that can be applied directly to standard multivariate time-series causal discovery by treating individual series as single-variable groups.

4 EXPERIMENTS

We benchmarked CausalKGR against four state-of-the-art baselines: GC (Standard Granger Causality) [5], 2GVCI [21], gCDMI [1], and HGCRM [3].

Synthetic Data Performance. On both the linear VAR and nonlinear Lorenz-96 datasets, CausalKGR consistently achieved the highest AUROC and AUPRC scores. Notably, in the chaotic Lorenz-96 system, CausalKGR maintained robust performance (AUROC> 0.90) as the number of groups increased, whereas baseline methods like HGCRM degraded significantly (AUROC $\approx$ 0.78). This demonstrates CausalKGR's superior ability to capture complex, nonlinear inter-group interactions that traditional linear mappings fail.

Real-World Application: fMRI. We applied CausalKGR to a resting-state fMRI dataset to decipher effective connectivity among 7 key Regions of Interest (ROIs) [14]. CausalKGR successfully identified the Posterior Cingulate Cortex (PCC) as a central functional hub exerting causal influence on the Angular Gyrus (AG) and Middle Temporal Gyrus (MTG). These findings align with established neuroscience literature [23]. In contrast, the baseline HGCRM failed to detect the MTG $\rightarrow$ AG pathway, a connection previously validated in clinical studies of schizophrenia [7].

Ablation & Representation Analysis. To validate our architectural contributions, we conducted extensive ablation studies. Results confirm that the GIF module is critical; removing it led to a marked performance drop across all datasets, underscoring the need to model system-wide joint distributions. Furthermore, we evaluated the quality of learned representations by reconstructing the original time series. CausalKGR's representations preserved more temporal dynamics and distributional features than standard aggregation methods such as Mean, PCA, and MCCA [3], demonstrating that CausalKGR effectively prevents group knowledge loss.

5 CONCLUSION

We present CausalKGR, a unified framework for group-level Granger causal discovery that integrates self-supervised representation learning with sparse causal inference. By leveraging knowledge-conditional attention and proximal optimization, CausalKGR effectively captures intra- and inter-group dynamics while producing interpretable causal graphs. Extensive experiments demonstrate its superiority on both synthetic and real-world datasets, making it a practical tool for high-dimensional grouped time-series causal analysis.

6 ACKNOWLEDGEMENTS

This work is supported by the National Natural Science Foundation of China (No.62172018, No.62102008) and CCF-Zhipu Large Model Innovation Fund (CCF-Zhipu202414).

REFERENCES

[1] Wasim Ahmad, Maha Shadaydeh, and Joachim Denzler. 2024. Deep Learning-based Group Causal Inference in Multivariate Time-series. In *AAAI Workshop on AI for Time-series (AAAI-WS)*. arXiv:2401.08386 [cs.LG] https://arxiv.org/abs/2401.08386

[2] Federico Battiston, Giulia Cencetti, Iacopo Iacopini, Vito Latora, Maxime Lucas, Alice Patania, Jean-Gabriel Young, and Giovanni Petri. 2020. Networks beyond pairwise interactions: structure and dynamics. *Physics Reports* (Aug 2020), 1–92. https://doi.org/10.1016/j.physrep.2020.05.004

[3] Ruijia Cai, Yinchuan Wu, Xuan Huang, et al. 2024. Granger causal representation learning for groups of time series. *Science China Information Sciences* 67 (2024), 152103. https://doi.org/10.1007/s11432-021-3724-0

[4] Yuxiao Cheng, Runzhao Yang, Tingxiong Xiao, Zongren Li, Jinli Suo, Kunlun He, and Qionghai Dai. 2023. CUTS: Neural Causal Discovery from Irregular Time-Series Data. In *The Eleventh International Conference on Learning Representations, ICLR 2023, Kigali, Rwanda, May 1-5, 2023*. OpenReview.net. https://openreview.net/forum?id=UG8bQcD3Emv

[5] C. W. J. Granger. 1969. Investigating Causal Relations by Econometric Models and Cross-Spectral Methods. *Econometrica* 37, 3 (1969), 424–438. https://doi.org/10.2307/1912791

[6] Karol Gregor, Ivo Danihelka, Andriy Mnih, Charles Blundell, and Daan Wierstra. 2014. Deep autoregressive networks. In *International Conference on Machine Learning*. PMLR, 1242–1250.

[7] Wenbin Guo, Feng Liu, Changqing Xiao, Miaoyu Yu, Zhikun Zhang, Jianrong Liu, Jian Zhang, and Jingping Zhao. 2015. Increased Causal Connectivity Related to Anatomical Alterations as Potential Endophenotypes for Schizophrenia. *Medicine* 94, 42 (October 2015), e1493. https://doi.org/10.1097/MD.0000000000001493

[8] Irina Higgins, Loic Matthey, Arka Pal, Christopher Burgess, Xavier Glorot, Matthew Botvinick, Shakir Mohamed, and Alexander Lerchner. 2017. beta-VAE: Learning Basic Visual Concepts with a Constrained Variational Framework. *International Conference on Learning Representations,International Conference on Learning Representations* (Apr 2017).

[9] Biwei Huang, Kun Zhang, Ruben Sanchez-Romero, Joseph Ramsey, Madelyn Glymour, and Clark Glymour. 2019. Diagnosis of Autism Spectrum Disorder by Causal Influence Strength Learned from Resting-State fMRI Data. *ArXiv* abs/1902.10073 (2019). https://api.semanticscholar.org/CorpusID:67855944

[10] Binyan Jiang, Jialiang Li, and Qiwei Yao. 2023. Autoregressive networks. *Journal of Machine Learning Research* 24, 227 (2023), 1–69.

[11] Genta Kikuchi and Shohei Shimizu. 2023. Structure Learning for Groups of Variables in Nonlinear Time-Series Data with Location-Scale Noise. In *Proceedings of the 2023 Causal Analysis Workshop Series (Proceedings of Machine Learning Research, Vol. 223)*, Erich Kummerfeld, Sisi Ma, Eric Rawls, and Bryan Andrews (Eds.). PMLR, 20–39. https://proceedings.mlr.press/v223/kikuchi23a.html

[12] Hongming Li, Shujian Yu, and Jose Principe. 2023. Causal Recurrent Variational Autoencoder for Medical Time Series Generation. *Proceedings of the AAAI Conference on Artificial Intelligence* 37, 7 (Jun. 2023), 8562–8570. https://doi.org/10.1609/aaai.v37i7.26031

[13] Yong Liu, Tengge Hu, Haoran Zhang, Haixu Wu, Shiyu Wang, Lintao Ma, and Mingsheng Long. 2024. iTransformer: Inverted Transformers Are Effective for Time Series Forecasting. In *The Twelfth International Conference on Learning Representations*. https://openreview.net/forum?id=JePfAI8fah

[14] Kate Brody Nooner, Stanley J. Colcombe, Russell H. Tobe, Maarten Mennes, Melissa M. Benedict, Alexis L. Moreno, Laura J. Panek, Shaquanna Brown, Stephen T. Zavitz, Qingyang Li, Sharad Sikka, David Gutman, Saroja Bangaru, Rochelle Tziona Schlachter, Stephanie M. Kamiel, Ayesha R. Anwar, Caitlin M. Hinz, Michelle S. Kaplan, Anna B. Rachlin, Samantha Adelsberg, Brian Cheung, Ranjit Khanuja, Chaogan Yan, Cameron C. Craddock, Vincent Calhoun, William Courtney, Margaret King, Dylan Wood, Christine L. Cox, A. M. Clare Kelly, Adriana Di Martino, Eva Petkova, Philip T. Reiss, Nancy Duan, Dawn Thomsen, Bharat Biswal, Barbara Coffey, Matthew J. Hoptman, Daniel C. Javitt, Nunzio Pomara, John J. Sidtis, Harold S. Koplewicz, Francisco Xavier Castellanos, Bennett L. Leventhal, and Michael P. Milham. 2012. The NKI-Rockland Sample: A Model for Accelerating the Pace of Discovery Science in Psychiatry. *Frontiers in Neuroscience* 6 (Oct 2012). https://doi.org/10.3389/fnins.2012.00152

[15] H. J. Nussbaumer. 1982. The Fast Fourier Transform. In *The Fast Fourier Transform*. Springer Berlin Heidelberg, Berlin, Heidelberg, 80–111. https://doi.org/10.1007/978-3-642-81897-4_4

[16] Neal Parikh, Stephen Boyd, et al. 2014. Proximal algorithms. *Foundations and Trends in Optimization* 1, 3 (2014), 127–239.

[17] Rohit Singh, Alexander P Wu, and Bonnie Berger. 2022. Granger causal inference on DAGs identifies genomic loci regulating transcription. In *International Conference on Learning Representations*.

[18] Gideon Stein, Maha Shadaydeh, Jan Blunk, Niklas Penzel, and Joachim Denzler. 2025. CausalRivers–Scaling up benchmarking of causal discovery for real-world time-series. In *International Conference on Learning Representations*.

[19] Alex Tank, Ian Covert, Nicholas Foti, Ali Shojaie, and Emily B Fox. 2022. Neural granger causality. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 44, 8 (2022), 4267–4279.

[20] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, L. Kaiser, and I. Polosukhin. 2017. Attention is all you need. In *Advances in Neural Information Processing Systems*. 5998–6008.

[21] Jonas Wahl, Urmi Ninad, and Jakob Runge. 2023. Vector Causal Inference between Two Groups of Variables. *Proceedings of the AAAI Conference on Artificial Intelligence* 37, 10 (Jun. 2023), 12305–12312. https://doi.org/10.1609/aaai.v37i10.26450

[22] Jonas Wahl, Urmi Ninad, and Jakob Runge. 2024. Foundations of causal discovery on groups of variables. *Journal of Causal Inference* 12, 1 (2024), 20230041. https://doi.org/doi:10.1515/jci-2023-0041

[23] Hao Yan, Yu Zhang, Huasheng Chen, Yuhui Wang, and Yijun Liu. 2013. Altered Effective Connectivity of the Default Mode Network in Resting-State Amnestic Type Mild Cognitive Impairment. *Journal of the International Neuropsychological Society* 19, 4 (April 2013), 400–409. https://doi.org/10.1017/S1355617712001580 Epub 2013 Feb 21.

[24] Yue Yu, Xuan Kan, Hejie Cui, Ran Xu, Yujia Zheng, Xiangchen Song, Yanqiao Zhu, Kun Zhang, Razieh Nabi, Ying Guo, et al. 2023. Deep dag learning of effective brain connectivity for fmri analysis. In *2023 IEEE 20th International Symposium on Biomedical Imaging (ISBI)*. IEEE, 1–5.

[25] Zhenyu Zhou, Yonghong Chen, Mingzhou Ding, Paul Wright, Zuhong Lu, and Yijun Liu. 2009. Analyzing brain networks with PCA and conditional Granger causality. *Human Brain Mapping* 30, 7 (2009), 2197–2206. https://doi.org/10.1002/hbm.20661 arXiv:https://onlinelibrary.wiley.com/doi/pdf/10.1002/hbm.20661