

Emergence-Inspired Multi-Granularity Causal Learning

Hanwen Luo^{1,2}, Guoxian Yu^{1,2,*}, Jun Wang², Yanyu Xu¹, Yongqing Zheng^{2,3}, Qingzhong Li^{2,3}

¹School of Software, Shandong University, Jinan, China

²SDU-NTU Joint Centre for AI Research, Shandong University, Jinan, China

³Dareway Software Co., Ltd., Jinan, China

hwluo@mail.sdu.edu.cn, {gxyu, kingjun, xu_yanyu, lqz}@sdu.edu.cn, zhengyongqing@dareway.com.cn

Abstract

Existing causal learning algorithms focus on micro-level causal discovery, confronting significant challenges in identifying the influence of macro systems, composed of micro-level variables, on other variables. This difficulty arises because the causal relationships in macro systems are often mediated through micro-level causal interactions, which can lead to erroneous causal discovery or omission when dispersed. To address this issue, we propose the Emergence-inspired Multi-granularity Causal learning (EMCausal) method. Inspired by the emerging phenomena of aggregating micro level variables into macro level representations, EMCausal introduces a progressive mapping encoder to simulate the process, thus capturing the causal relationships driven by these macro entities. Next, it introduces a causal consistency constraint to collaboratively reconstruct micro variables using macro-level representations, enabling the learning of a multi-granular causal structure. Experimental results on both synthetic and real datasets demonstrate that EMCausal can identify causal graphs under the influence of causal emergence, outperforming competitive baselines in term of accuracy and robustness.

Introduction

Causal discovery is integral to scientific research, as it focuses on uncovering the underlying mechanisms that drive observed phenomena. In the complex systems, distinguishing causal relationships from mere correlations is essential for the effective design of interventions, policy development, and the formulation of predictive models that can adapt to dynamic environments (Cui and Athey 2022; Cai et al. 2024).

Early causal discovery algorithms primarily rely on conditional independence tests (Spirtes et al. 2000; Colombo, Maathuis et al. 2014), which require large sample sizes. Alternatively, another line of works involve searching for the causal graph with the highest score using various search strategies (Chickering 2002b; Yang et al. 2023). Since Zheng et al. (2018) proposed the continuous optimization acyclicity constraint, researchers have further extended the continuous optimization of causal learning, including non-linear extensions (Zheng et al. 2020; Yu et al. 2019), op-

timization techniques (Lee et al. 2020), and robustness enhancements (He et al. 2021). These studies aim to improve the performance of causal learning algorithms when handling complex datasets, and increase their robustness and adaptability to noise. The exploration of multi-granularity data (Liang et al. 2024b) has emerged as a current research focus, presenting new challenges for application of causal learning methods in complex systems.

While micro structures focus on the direct causality between micro variables, we define the macro structures as the collective influence of certain micro variables impacting other micro ones. At the micro-level, it is possible to capture more detailed and comprehensive information. However, when the number of micro-level variables is large and the structure becomes overly complex, this complexity may lead to inaccuracies or omissions in causal discovery. In contrast, observing from a macro perspective allows certain causal relationships to manifest with stronger causal effects, thereby facilitating their identification. Contemporary causal learning algorithms predominantly focus on micro-level causal discovery, neglecting the potential insights that can be gained from learning causal structures at the macro-level.

Emergence phenomena manifesting as overall behaviors that cannot be directly inferred from the behavior of individual parts, exhibiting entirely new properties not present in the individuals (Tononi and Koch 2015; Epstein and Axtell 1996; Wei et al. 2022). It reveals that the collective influence of micro-level groups on other variables cannot simply be decomposed into the individual effects of each variable within the group. Inspired by the emergence phenomenon, we developed a multi-granularity causal discovery approach, EMCausal. Our method first aggregates micro-level variables into macro-level using a Progressive Mapping Encoder (PME), motivated by the concept of *supervenience* (Horgan 1993), wherein the complete observation of the micro system determines the state of the macro system then utilizes both the micro variables and macro representations to reconstruct the micro variables, simulating the process of causal emergence to extract a multi-granularity causal structure. Additionally, EMCausal imposes a micro-macro consistency constraint to ensure coherence in the multi-granularity causal structure. Upon completion of training, these multi-granularity causal structures can be derived directly from

*Corresponding author: Guoxian Yu.

Copyright © 2025, Association for the Advancement of Artificial Intelligence (www.aaai.org). All rights reserved.

the model parameters. The code of `EMCausal` is shared at <http://www.sdu-idea.cn/codes.php?name=EMCausal>.

Our main contributions can be outlined as follows:

- (i) We propose `EMCausal`, a multi-granularity causal discovery approach inspired by the emergence phenomenon. It introduces a Progressive Mapping Encoder to aggregate micro-level variables into macro-level representations and reconstructs micro variables from these representations by simulating causal emergence.
- (ii) `EMCausal` defines a micro-macro consistency constraint to ensure that the inferred causal structures at the macro-level align with those at the micro-level, thereby preserving the coherence of multi-granularity causal structure.
- (iii) Experimental results on synthetic multi-granularity dataset and real datasets demonstrate that `EMCausal` outperforms competitive baselines (Liang et al. 2024b; Yu et al. 2019; Ng et al. 2019; Chickering 2002b; Spirtes et al. 2000).

Related Work

Causal discovery infers causal structures from data, helping to identify the causal direction between variables. Existing causal discovery methods can be classified as constraint-based (Colombo, Maathuis et al. 2014; Vandenbroucke, Broadbent, and Pearce 2016; Peters et al. 2014; Spirtes, Meek, and Richardson 2013; Zhang, Zhou, and Guan 2018) and score-based (Chen et al. 2024; Ng et al. 2022; Shimizu et al. 2006; Zhang et al. 2023; Liang et al. 2023; Chickering 2002a) algorithms. The former infer causal structures by leveraging conditional independence information within the data. Nevertheless, the computation of conditional independence is highly demanding in terms of sample size and is computationally expensive, limiting its applicability to large-scale datasets. The latter category (Chickering 2002b; Zheng et al. 2020; Yu et al. 2019; Lu and Gao 2023; Wang et al. 2022; Liang et al. 2024a; Yang et al. 2024), on the other hand, defines a scoring function to evaluate the suitability of a given causal graph and employs search algorithms to seek the graph with the highest score. However, existing algorithms often overlook the impact of diluted macro causal effects at the micro-level on the discovery of causal graphs. This oversight can lead to erroneous causal inference in complex systems. Our innovation lies in addressing this gap, enabling more accurate identification of causal relationships across different granularity and improving the robustness of causal discovery in intricate environments.

Integrating the study of emergence with causal analysis making it possible to further dissect the relationship in complex systems, helping to explain the overall system behavior. Some studies consider the emergence phenomena with stronger causal effects, but do not discuss how to uncover these causal relationships (Klein et al. 2022; Breakspear 2017; Rubenstein et al. 2017). They utilized metrics like Effective Information (EI) to design strategies for reducing system uncertainty, and for exploring appropriate granularity to observe the system. In contrast to the previous studies, MgCSL (Liang et al. 2024b) reconstructs the micro variables to learn appropriate representations of macro variables, using path-product to uncover the causal relations among micro and macro variables. Compared with the MgCSL path

product calculation, `EMCausal` trains a PME with an attention matrix to mask the macro-micro causal structure, enhancing the certainty of the identified macro-micro causal relationship. In addition, `EMCausal` incorporates a micro-macro consistency constraint to account for the interactions between learned micro and macro structures and to avoid structural conflict.

Background

This section introduces the fundamental concepts relevant to this work.

Causal Graph: A causal graph is a DAG (Directed Acyclic Graph) used to represent causal relationships between variables. In a causal graph $\mathcal{G} = (\mathcal{V}, \mathcal{E})$, the set \mathcal{V} represents variables, and the edge set \mathcal{E} represents causal relationships between variables. If there is a directed edge $X \rightarrow Y$, it indicates that variables X is a direct cause of variables Y .

Data Generation Process: The data generation process involves generating observational data based on the structure and relationships in the causal graph. Suppose we have a causal graph G with variables $\mathcal{V} = \{V_1, V_2, \dots, V_d\}$ representing variables. To avoid confusion between nodes and variables, we will consistently use X_i . The data generation process can be expressed as:

$$X_i = f_i(Pa(X_i)) + e_i \quad (1)$$

where $Pa(X_i)$ denotes the parent set of X_i (i.e., the direct causes), f_i is a specific functional form, and e_i is the noise term. The noise affecting different variables is mutually independent.

Methodology

The multi-granularity causal structures can be described by $\mathcal{G}_M = (\mathcal{V}_M, \mathcal{E}_M)$. Unlike the causal graph of the micro-system, $\mathcal{V}_M = \{X_1, \dots, X_d, Y_1, \dots, Y_q\}$ incorporates variables of varying granularity, allowing \mathcal{G}_M to simultaneously reflect causal relationships at both micro and macro levels. Given the observed variables $\mathbf{X} = (X_1, X_2, \dots, X_d)$ with $\mathbf{X} \in \mathbb{R}^{n \times d}$, our goal is to discover the multi-granularity causal structure \mathcal{G}_M , which not only clarifies the causal relationships within the original micro-level system but also uncovers potential macro variables composed of micro variables.

Model Structure Overview

To obtain reliable multi-granularity causal structures from observational data \mathbf{X} , we design the PME to aggregate micro variables toward macro variables and achieve the micro-macro relationship discovery. Next, we perform cross-layer collaborative causal learning to learn multi-granularity causal structures by reconstructing micro variables. Our framework is illustrated in Figure 1.

Progressive Mapping Encoder Emergence phenomena manifesting as overall behaviors that cannot be directly inferred from the behavior of individual parts, which should

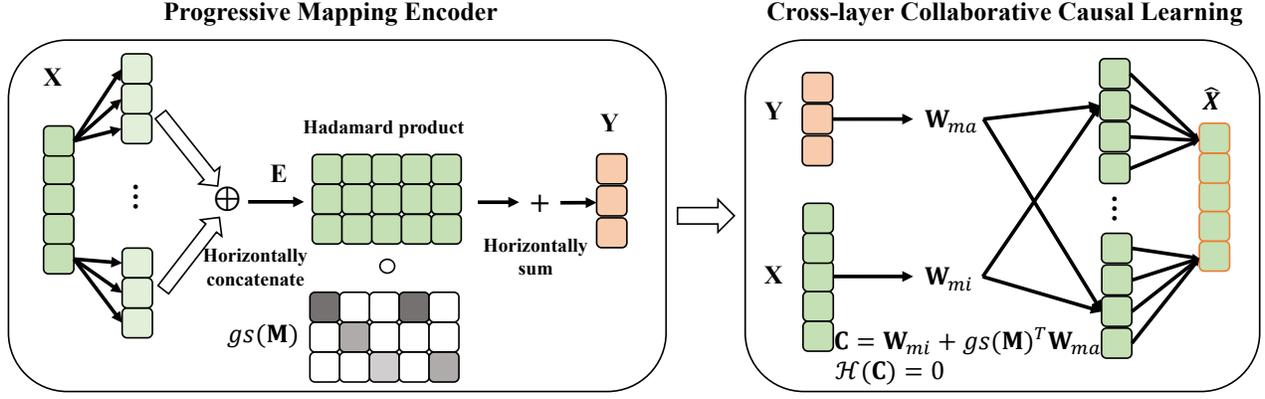


Figure 1: The conceptual framework of $EMCausal$. Given the micro-level input \mathbf{X} , the PME is used to aggregate micro-level variables into macro-level representations, and outputs the encoded results \mathbf{E} . As \mathbf{M} corresponds the relations from micro variables to macro representations, $EMCausal$ uses the Gumbel-Softmax operation $gs(\mathbf{M})$ to perform the Hadamard product with \mathbf{E} , and then sums the results to get the macro representation \mathbf{Y} . Under the micro-macro consistency constraint $\mathcal{H}(\mathbf{C}) = 0$, $EMCausal$ reconstructs origin micro-level variables \mathbf{X} based on \mathbf{X} and \mathbf{Y} with a MLP parameterized by \mathbf{W}_{ma} and \mathbf{W}_{mi} , where \mathbf{W}_{ma} encodes the relationship between macro representations and micro variables, \mathbf{W}_{mi} encodes the relationship among micro variables, and $\mathbf{C} = \mathbf{W}_{mi} + gs(\mathbf{M})^T \mathbf{W}_{ma}$ stores the reachable edges in the causal graph.

not be overlooked, as it may lead to erroneous causal discoveries. To address this gap, we introduce abstract macro variables using PME to discover collective influence of variables. Therefore, a robust macro-level representation is necessary, which relies not only on the correspondence between micro and macro variables but also on a thorough understanding of the macro variables generation process.

By employing the PME, we integrate an attention mechanism to capture the relationship between micro variables and abstract macro variables and iteratively adjust the model parameters during training, ultimately deriving a definitive cross-granularity variables relationship and a model for generating macro-level representations.

The PME framework comprises a stacked encoder coupled with a reweighting procedure which could be considered as an attention mechanism. Unlike MgCSL (Liang et al. 2024b) that directly obtains the macro-level variables representation via a reconstruction task, $EMCausal$ solely uses the encoder part to generate intermediate encoding states.

Suppose the number of micro variables and macro ones are d and q , respectively. Given the input \mathbf{X} , we first utilize each micro-level variable to predict various macro-level variables. $\mathbf{E}_i \in \mathbb{R}^{n \times q}$ denotes the encoding result using the i -th micro-level variable:

$$\mathbf{E}_i = \text{MLP}(\mathbf{X}_i; \theta_{enc}^i) \quad (2)$$

where θ_{enc}^i is the parameter of the i -th encoder of the stacked encoder.

To learn the correspondence between micro-level variables and macro-level one while aggregating micro-level variables, we then employ Gumbel-softmax (Jang, Gu, and Poole 2017), a technique that draws samples from a categorical distribution, to reweight all predictions. Denote the model internal parameters $\mathbf{M} \in \mathbb{R}^{d \times q}$, then we can obtain

the representation of macro variables $\mathbf{Y} \in \mathbb{R}^{n \times q}$ as follows:

$$\mathbf{Y} = \sum_i^d gs(\mathbf{M}_i) \circ \mathbf{E}_i \quad (3)$$

where \circ denotes the Hadamard product, \mathbf{M}_i is the i -th row of \mathbf{M} , and $gs(\mathbf{M}_i)$ refers to the outcome of Gumbel-Softmax. This operation facilitates sampling while preserving differentiability through the Softmax function, thereby enabling the gradient-based optimization of discrete variables.

For micro variable i , $\mathbf{M}_i = [M_{i,1}, M_{i,2}, \dots, M_{i,q}]$ corresponding to the probability of the i -th micro variable generating the j -th macro variable, ϵ_j is the gumbel noise for each variable:

$$\epsilon_j = -\log(-\log(M_{i,j})) \quad (4)$$

Next, we calculate the Gumbel-Softmax probability $p_{i,j}$ for corresponding the i -th micro variable with j -th macro one:

$$p_{i,j} = \frac{\exp((\log(M_{i,j}) + \epsilon_j)/\tau)}{\sum_j \exp((\log(M_{i,j}) + \epsilon_j)/\tau)} \quad (5)$$

where τ is the temperature parameter that controls the smoothness of the Softmax function.

By continuously training the parameters of PML, we can optimize them to approximate a binary matrix, facilitating the establishment of a threshold to identify the correspondence between micro-level and macro-level variables. With PME, macro state representations can be derived from the input and expressed mathematically as:

$$\mathbf{Y} = \text{PME}(\mathbf{X}) \quad (6)$$

The following section will detail how PME is integrated into subsequent processes and jointly trained within the reconstruction task.

Cross-Layer Collaborative Causal Learning To jointly learn micro and macro structures, we introduce a collaborative learning strategy that enables the simultaneous identification of causal relationships within both micro and macro systems. To avoid structural conflict between micro and macro structure, we develop a micro-macro consistency constraint mechanism to prevent erroneous causal structures.

Suppose $\mathbf{W}_{mi} \in \mathbb{R}^{d \times d}$ is the weight matrix connecting micro variables, $\mathbf{W}_{ma} \in \mathbb{R}^{q \times d}$ connecting macro variables to micro ones. Different from NOTEARS-MLP (Zheng et al. 2020) and its variants that only tune \mathbf{W}_{mi} at the micro-level, we use micro variables and macro representations to jointly reconstruct the original variable as follows:

$$\hat{\mathbf{X}} = \text{MLP}(\mathbf{X}\mathbf{W}_{mi} \oplus \text{PME}(\mathbf{X})\mathbf{W}_{ma}; \theta_{mlp}) \quad (7)$$

where θ_{mlp} is the parameter of the reconstruction model, \oplus is the concatenation operator. Then we can define the optimization target of reconstruction as:

$$\mathcal{L}_{re} = \mathcal{J}(\mathbf{X}, \hat{\mathbf{X}}) \quad (8)$$

where $\mathcal{J}(\mathbf{X}, \hat{\mathbf{X}}) = \|\mathbf{X} - \hat{\mathbf{X}}\|_2^2$ is the square loss. We omit the regularization term to simplify the presentation.

To avoid the representation of macro variables solely depending on the model’s initial parameters, causing biased results, we generally adjust the parameters of \mathbf{W}_{mi} first, and then learn the causal relationships at the macro-level after the convergence of micro-level causal structure. In the first optimization stage of EMcausal, our optimization objective is:

$$\begin{aligned} \min \quad & \mathcal{L}_{re} \\ \text{s.t.} \quad & \mathcal{H}(\mathbf{W}_{mi}) = 0 \end{aligned} \quad (9)$$

where $\mathcal{H}(\mathbf{W})$ is the acyclicity constraint,

$$\mathcal{H}(\mathbf{W}) = \text{tr}(e^{\mathbf{W} \circ \mathbf{W}}) - d = 0 \quad (10)$$

After learning the micro-level causal relations, EMcausal moves to the second optimization stage to learn causal relation at the macro-level. To ensure that the causal relationships at the macro level are effectively guided by those at the micro level during the optimizing process, we impose a *consistency constraint* on macro-level causal discovery to ensure the consistency between the micro-level causal structures and the underlying macro-level causality as follows:

$$\mathbf{C} = \mathbf{W}_{mi} + g_s(\mathbf{M})^T \mathbf{W}_{ma} \quad (11)$$

$\mathbf{C} \in \mathbb{R}^{d \times d}$ stores the reachable edges in the causal graph. We can formulate the optimization objective for this stage as follows:

$$\begin{aligned} \min \quad & \mathcal{L}_{re} \\ \text{s.t.} \quad & \mathcal{H}(\mathbf{C}) = \text{tr}(e^{\mathbf{C} \circ \mathbf{C}}) - d = 0 \end{aligned} \quad (12)$$

where the regularization term is also omitted to simplify the presentation. During this phase, we freeze the parameters of \mathbf{W}_{mi} , but optimize \mathbf{M} (induced by PME) and \mathbf{W}_{ma} to ensure that the causal relationships at the macro-level do not contradict with those at the micro-level.

Algorithm 1 outlines the main procedure of EMcausal. Lines 2-6 learns the micro-level causal relations between d variables, while lines 7-11 optimizes the causal relations at the macro-level.

Algorithm 1: EMcausal: Emergence-inspired Multi-granularity Causal Learning

Input: observed data \mathbf{X} and threshold δ

Output: Multi-granularity Causal Graph \mathcal{G}_M

- 1: Initialize the set of model parameters $\{\mathbf{M}, \mathbf{W}_{ma}, \theta_{enc}, \mathbf{W}_{mi}, \theta_{mlp}\}$
 - 2: \triangleright Learning on the micro-level graph
 - 3: **while** not arriving maximal iterations or triggering termination conditions **do**
 - 4: Fix $\mathbf{M}, \mathbf{W}_{ma}, \theta_{enc}$, and calculate \mathcal{L} by Eq. (9)
 - 5: Update $\mathbf{W}_{mi}, \theta_{mlp}$ to minimize \mathcal{L} in Eq. (9)
 - 6: **end while**
 - 7: \triangleright Learning on the macro-level graph
 - 8: **while** not arriving maximal iterations or triggering termination conditions **do**
 - 9: Fix \mathbf{W}_{mi} , and calculate \mathcal{L} by Eq. (12)
 - 10: Update $\mathbf{W}_{ma}, \theta_{mlp}, \theta_{enc}, \mathbf{M}$ to minimize \mathcal{L} in Eq. (12)
 - 11: **end while**
 - 12: Update \mathbf{C} by Eq. (11)
 - 13: $\mathcal{G}_M(i, j) = \mathbb{I}(\mathbf{C}_{ij} > \delta) \triangleright$ prune too low-weight edges
 - 14: **return** \mathcal{G}_M
-

Optimization

Both phases of our optimization (Eq. (9) and Eq. (12)) can be made by an augmented Lagrangian method as follows:

$$\min \quad \mathcal{L} = \mathcal{J}(\mathbf{X}, \hat{\mathbf{X}}) + \frac{\mu}{2} \mathcal{H}^2(\mathbf{W}) + \gamma \mathcal{H}(\mathbf{W}) \quad (13)$$

where μ is the penalty coefficient and γ is the Lagrange multiplier. The constrained problem in Eq. (13) can be converted into a sequence of unconstrained subproblems, which can be solved by gradient descent optimization strategy as follows:

$$\begin{aligned} \boldsymbol{\theta}^{(\kappa+1)} \leftarrow & \arg \min \{ \mathcal{L} = \mathcal{J}(\mathbf{X}, \hat{\mathbf{X}}) + \frac{\mu}{2} \mathcal{H}^2(\mathbf{W}) \\ & + \gamma \mathcal{H}(\mathbf{W}) \} \end{aligned} \quad (14)$$

$$\mu^{(\kappa+1)} \leftarrow \begin{cases} \eta \mu^{(\kappa)}, & \mathcal{H}(\mathbf{W}^{(\kappa+1)}) > \rho \mathcal{H}(\mathbf{W}^{(\kappa)}) \\ \mu^{(\kappa)}, & \text{otherwise} \end{cases} \quad (15)$$

$$\gamma^{(\kappa+1)} \leftarrow \gamma^{(\kappa)} + \mu \mathcal{H}(\mathbf{W}) \quad (16)$$

where for the first stage $\boldsymbol{\theta} \triangleq \{\theta_{mlp}, \mathbf{W}_{mi}\}$, as for the second stage $\boldsymbol{\theta} \triangleq \{\mathbf{M}, \theta_{mlp}, \theta_{enc}, \mathbf{W}_{ma}\}$, and $\boldsymbol{\theta}^{(\kappa)}$ is the iterative optimized $\boldsymbol{\theta}$ in the κ -th iteration. In this work, we apply the L-BFGS-B algorithm (Zhu et al. 1997) to solve the optimization problem. Although EMcausal incorporates several procedures, it introduces an acceptable computational complexity. The computational complexity of an iteration step of EMcausal is $O(ndmq + nd(d+q)k + d^2(d+q))$, where m and k are the dimensions of the hidden layer of PME and reconstruction model, respectively. As the computational complexity of one step of NOTEARS-MLP is $O(nd^2k + d^2k + d^3)$, our method introduces extra complexity by PME and macro variables number d , which are indispensable. While the number of macro variables q has the same magnitude as micro ones d , EMcausal holds an acceptable computational complexity.

d	method	Erdős-Rényi graph				Scale-Free graph			
		Precision(%) \uparrow	Recall(%) \uparrow	F1(%) \uparrow	SHD \downarrow	Precision(%) \uparrow	Recall(%) \uparrow	F1(%) \uparrow	SHD \downarrow
20	PC	50.89 \pm 6.61	36.04 \pm 9.54	41.91 \pm 8.47	26.77 \pm 4.47	53.63 \pm 6.93	45.00 \pm 10.29	48.70 \pm 8.57	23.80 \pm 3.16
	GES	45.06 \pm 5.81	38.70 \pm 7.37	41.38 \pm 6.11	31.40 \pm 4.55	46.61 \pm 10.88	43.21 \pm 8.80	44.51 \pm 9.05	27.67 \pm 5.41
	DAG-GNN	61.41 \pm 9.14	39.56 \pm 11.17	47.03 \pm 9.56	37.50 \pm 6.22	92.92\pm7.39	6.46 \pm 1.84	12.01 \pm 3.18	162.40 \pm 8.10
	GAE	73.56 \pm 30.06	13.04 \pm 9.11	21.39 \pm 12.95	41.60 \pm 4.88	80.33 \pm 17.06	24.64 \pm 15.73	34.47 \pm 19.27	34.67 \pm 7.23
	GraN-DAG	95.83\pm9.00	15.00 \pm 6.10	25.59 \pm 9.21	39.10 \pm 2.81	91.33 \pm 9.19	15.56 \pm 2.98	26.54 \pm 4.67	30.90 \pm 1.52
	MgCSL	33.42 \pm 6.49	60.87 \pm 9.11	42.96 \pm 7.01	49.00 \pm 11.49	28.40 \pm 7.68	52.50 \pm 18.35	36.25 \pm 9.78	45.60 \pm 8.39
	EMCausal	85.70 \pm 13.79	71.30\pm11.40	76.69\pm8.82	18.50\pm7.01	78.4 \pm 13.77	71.10\pm12.22	73.46\pm8.21	17.7\pm5.77
40	PC	44.55 \pm 5.83	19.82 \pm 3.24	27.41 \pm 4.13	140.75 \pm 11.30	43.30 \pm 7.16	19.59 \pm 3.31	26.96 \pm 4.50	138.36 \pm 7.05
	GES	43.58 \pm 6.32	29.87 \pm 4.36	35.41 \pm 4.99	160.00 \pm 12.68	45.89 \pm 10.32	36.94 \pm 6.75	40.88 \pm 8.14	109.70\pm14.80
	DAG-GNN	54.89 \pm 10.46	19.83 \pm 5.03	28.92 \pm 6.51	161.40 \pm 16.06	50.86 \pm 7.96	21.70 \pm 3.20	30.04 \pm 2.87	161.20 \pm 8.46
	GAE	59.21 \pm 22.46	3.24 \pm 2.56	5.92 \pm 4.32	170.80 \pm 7.52	60.42 \pm 14.91	4.70 \pm 2.41	8.60 \pm 4.18	160.60 \pm 3.63
	GraN-DAG	92.92\pm7.39	6.46 \pm 1.84	12.01 \pm 3.18	162.40 \pm 8.10	87.65\pm7.26	5.31 \pm 1.84	9.96 \pm 3.31	156.40 \pm 3.17
	MgCSL	27.93 \pm 2.91	38.70 \pm 4.98	32.29 \pm 2.66	220.70 \pm 20.51	41.16 \pm 9.71	31.71 \pm 4.95	31.71 \pm 4.95	228.20 \pm 23.28
	EMCausal	66.11 \pm 15.77	63.91\pm6.38	63.67\pm7.11	97.80\pm29.69	57.36 \pm 16.54	69.62\pm7.96	61.10\pm7.66	111.67 \pm 37.28

Table 1: Results on nonlinear models. The best result is highlighted in **bold font**. For the reported results, $\uparrow(\downarrow)$ means the larger(smaller) the score, the better the model performance is.

Experimental Results and Analysis

In this section, we conduct a series of experiments to study the effectiveness of EMCausal. We first compare the performance of various causal discovery algorithms using synthetic datasets and give a detailed analysis of the strengths and limitations of compared methods. We then investigate the impact of hidden macro variables to assess the ability of EMCausal to identify elusive causal relationships in the presence of emergent phenomena. Next, we analyze the consistency of causal relationships among micro and macro levels to further investigate the contribution factors within EMCausal. Finally, we conduct experiments on real-world datasets to testify EMCausal in practical applications.

We implement EMCausal using PyTorch 1.13 and conduct experiments on a server with the Intel(R) Xeon(R) Gold 6248R CPU, 512G memory, 8 NVIDIA GeForce RTX 3090 GPUs, and Ubuntu 22.04.

Performance Analysis

Experimental Setup We choose several representative causal discovery algorithms as our baselines, including classical PC (Spirtes et al. 2000), which is based on conditional independence tests, and score-based optimization methods such as GES (Chickering 2002b), DAG-GNN (Yu et al. 2019), GAE (Ng et al. 2019), and MgCSL (Liang et al. 2024b). The parameter configurations of these methods are summarized in Table 2.

To synthesize multi-granularity data, we set the number of micro variables that constitute each macro variable to 4 and 8 with the total number of micro variables as 20 or 40. The edge density is set to 2. We use randomly generated two layer MLP, whose hidden size is set to 100, to model the nonlinear relationships between variables. For each experiment, we sample 1000 data samples from the data generation process. To reduce randomness, we perform ten independent experiments and report the mean and standard deviation. We generate nonlinear multi-granularity causal graphs and corresponding data, and conduct experiments on this basis to study the effectiveness of EMCausal. Since contemporary

	parameter	value
PC	significance level α	0.05
GES	score function	BIC
DAG-GNN	learning rate	0.003
	VAE layers	[1,64,1,64,1]
	initial λ	0
	initial c	1
GAE	learning rate	0.001
	input units	1
	hidden layers	1
	hidden units	4
	λ	0
	initial α	0
	initial ρ	1
GraN-DAG	learning rate	0.001
	hidden layers	2
	hidden units	10
	PNS threshold	0.75
	initial λ	0
	initial μ	0.001
MgCSL	initial μ	0.001
	tolerance of \mathcal{H}	0.1
	maximum of μ	10^{16}
	ϵ	0.2
	SAE layers	[$d,0.75d,5,0.75d,d$]
	MLP layers	[$d+5,10,1$]
EMCausal	ϵ	0.3
	τ	0.2
	PME structure	[$d,0.75d,5$]
	MLP layers	[$d+5,10,1$]

Table 2: Parameter configuration of compared methods

algorithms (except MgCSL) are unable to directly uncover multi-granular causal relations, we map the true and predicted results of each algorithm to the micro-level graph for comparison.

We mainly use four metrics to quantify the performance, namely Precision, Recall, F1 and Structural Hamming Distance (SHD). The larger the value of the first three metrics, the better the performance is, the opposite applies for SHD.

EMCausal vs. Constraint-based methods: The results in Table 1 clearly show that our EMCausal significantly outperforms constraint-based algorithms across all metrics on synthetic datasets, with the most pronounced advantage observed in precision. This performance disparity is primarily due to the limitations of conditional independence tests in capturing nonlinear causal relationships, leading to a substantial number of missed causal directions. Furthermore, as the number of variables increases, this difference becomes even more evident for the reliability of conditional independence tests heavily rely on the quantity of samples, more variables need more samples.

EMCausal vs. Score-based methods: Compared to other score-based methods, such as GraN-DAG and DAG-GNN, EMCausal exhibits significant differences in both F1 score and SHD. While GraN-DAG and DAG-GNN gain a higher precision, their recall rates are notably lower. This discrepancy arises primarily because these score-based methods only report predicted edges with higher confidence, thereby reducing the number of predictions and having lower recall. As the causal effects from macro structures become dispersed across micro variables, the causal influence on each individual micro variable weakens, making it difficult for these score-based methods to capture a comprehensive causal structure, which in turn lowers their recall rates. Furthermore, the hidden noise of macro variables leads DAG-GNN to focus more on recovering causal relationships among micro-level variables decomposed within the same macro-level variables, rather than reflecting the causal effects driven by macro variables.

Our EMCausal is explicitly designed to encourage the discovery of causal relationships at the macro-level. Although this design may result in a slight decrease in precision, it significantly enhances recall. The observed increase in F1 score and the reduction in SHD further validate its effectiveness, showing its superior capability in capturing macro structures within complex systems, thereby revealing multi-granularity causal relationships.

EMCausal vs. MgCSL: Both EMCausal and MgCSL are multi-granularity causal discovery algorithms, warranting a comprehensive comparative analysis. Although both methods involve the process of identifying macro variables and joint causal discovery, they differ in identifying macro variables and the subsequent causal relationship discovery. Therefore, we standardize the comparison by evaluating Precision, Recall, SHD, and the Number of Non-Zero elements (NNZ) related to macro-level causal relationships within the multi-granular graph. This allows us to assess the performance difference in uncovering macro-level causal structures.

As shown in Figure 2, our EMCausal significantly outperforms MgCSL in macro-level causal discovery across all metrics. This indicates that EMCausal is more accurate and reliable in identifying macro-level causal structures. The Number of Non-Zero elements (NNZ) reflects the number of edges in the predicted causal graph. While delivering superior performance, EMCausal consistently exhibits higher NNZ values compared to MgCSL, suggesting that EMCausal can find out more true causal edges and more

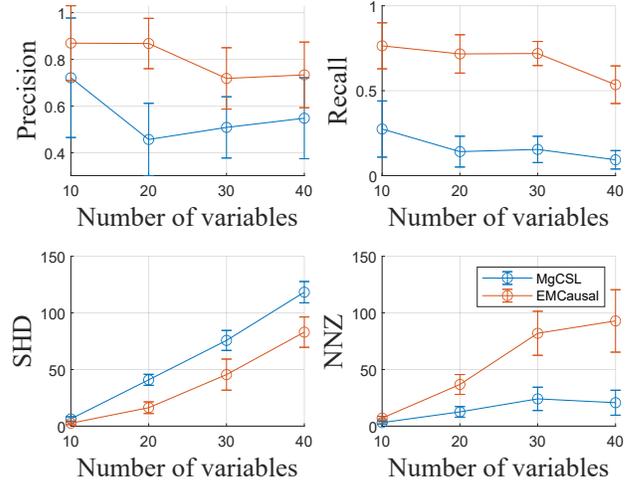


Figure 2: Precision, Recall, SHD, and NNZ of MgCSL and EMCausal on synthetic datasets with $q=4$ macro variables and $d \in \{10, 20, 30, 40\}$ micro-variables.

accurately identify causal relationships among micro and macro variables. In contrast, MgCSL misses many causal edges, thus has lower precision and recall values. This is because EMCausal defines a consistency constraint between macro-level and micro-level causal structures. Particularly, it ensures that the learning process of the macro-level causal graph is informed by the micro-level graph. This alignment guarantees the correctness of micro-level causal discovery and its effective guidance in constructing macro-level causal relationships, which MgCSL disregards this consistency and merely takes both macro and micro variables as a whole to learn the causal structure.

Impact of Macro-Level Variables

We further investigate the impact of hidden macro-level variables on various causal learning algorithms using synthetic multi-granular data. By fixing the number of micro variables to 20 and setting the edge density to 2, each macro variable is composed of 2 micro variables. We gradually increase the number of macro variables, and record the changes in Precision, Recall, F1 and SHD across different methods.

The results presented in Figure 3 demonstrate that the performance of nearly all algorithms declines as the number of macro variables decreases. This decline occurs because, in our experiments, a reduction in macro variables leads to an increase in causal relationships mediated by those macro variables. This observation aligns with our previous analysis. However, since EMCausal simulates the process of macro variables generation to consider the emergence-inspired causality, it generally manages a better robustness to this change.

Compared to conventional algorithms, our EMCausal can effectively resist with the impact caused by macro variables, maintain a relatively stable performance. This fact proves its advantages when dealing with complex macro systems.

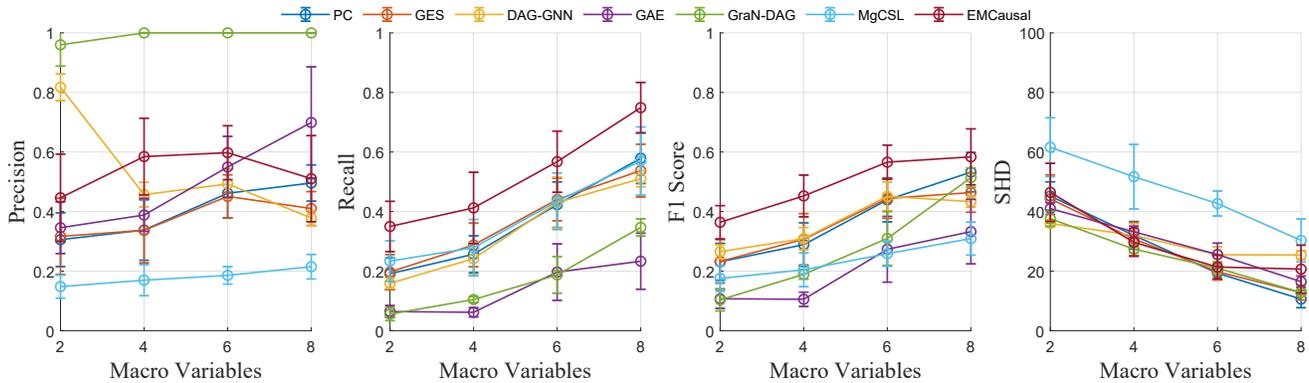


Figure 3: The performance of causal learning algorithms on synthetic multi-granularity data with the number of macro variables from 2 to 8.

Consistency of Causal Relationships

To verify that the macro variables identified by `EMCausal` encapsulate the common information of their corresponding micro variables, we conducted experiments on synthetic data. These experiments compared the consistency of causal relationships between micro and macro variables, aiming to demonstrate the coherence and effectiveness of the macro-level representations generated by `EMCausal`.

According to our criterion, if all micro variables associated with a macro variable point to another variable, the macro variable should exhibit such a causal relationship also. Based on this principle, we calculate the Precision and Recall and F1 score of the generated graphs, report the results in Table 3. The performance of both `EMCausal` and `MgCSL` shows some decline as the number of variables increases. However, `EMCausal` is generally much less affected, demonstrating its superior ability to maintain the consistency of causal structures and ground truths at both the micro and macro levels.

d	Method	Precision (%)	Recall (%)	F1 (%)
10	<code>MgCSL</code>	72.14±25.58	27.44±16.54	38.15±18.49
	<code>EMCausal</code>	86.96±16.01	76.41±13.54	79.84±10.28
20	<code>MgCSL</code>	45.71±15.48	14.13±9.08	21.12±11.98
	<code>EMCausal</code>	86.81±10.78	71.65±11.28	77.70±8.47
30	<code>MgCSL</code>	50.89±13.12	15.44±7.73	22.94±9.51
	<code>EMCausal</code>	71.86±13.14	71.89±7.09	70.93±6.47
40	<code>MgCSL</code>	54.81±17.34	9.32±5.43	15.54±8.50
	<code>EMCausal</code>	73.37±14.04	53.53±11.07	60.18±6.44

Table 3: `EMCausal` vs. `MgCSL` on causal consistency.

Result on the Sachs Dataset

To further validate the applicability of `EMCausal` in real-world scenarios, we conducted experiments on the Sachs dataset (Sachs et al. 2005). The Sachs dataset, comprising 853 samples, records the expression levels of 11 proteins and phosphorylated proteins measured from human T cells in the immune system. The results of our experiments are presented in Table 4.

Method	Precision(%) \uparrow	Recall(%) \uparrow	F1(%) \uparrow	SHD \downarrow
<code>PC</code>	42.86	<u>35.29</u>	<u>38.71</u>	11
<code>GES</code>	42.86	<u>35.29</u>	<u>38.71</u>	11
<code>DAG-GNN</code>	40.00	<u>35.29</u>	37.50	16
<code>GAE</code>	<u>50.00</u>	5.88	10.53	16
<code>GraN-DAG</code>	66.67	11.76	20.00	15
<code>MgCSL</code>	42.86	17.65	36.50	18
<code>EMCausal</code>	46.67	41.18	43.75	11

Table 4: Results on Sachs dataset. The best and the second results are highlighted in **bold font** and underline.

Our `EMCausal` achieves the highest recall and F1 score. Although `GraN-DAG` gives a superior precision, its overly conservative predictions resulted in a low recall value, leading to suboptimal SHD and F1 scores. In contrast, while `PC` and `GES` also achieve low SHD values similar to ours, their F1 scores do not demonstrate the same level of superiority. The results on the Sachs dataset further underscore `EMCausal`'s broad potential for real-world applications, highlighting its effectiveness in uncovering complex causal structures.

Conclusion

Inspired by the emergence, we propose a novel multi-granularity causal discovery algorithm, `EMCausal`. This approach integrates the discovery of micro-macro relationships with the aggregation of micro variables into macro-level representations. By reconstructing micro variables from these macro-level representations, `EMCausal` effectively learns multi-granularity causal structures. Experimental results demonstrate that `EMCausal` not only more accurately uncovers the underlying causal mechanisms within complex systems but also provides deeper insights into the behavior of these systems. This capability positions `EMCausal` as a powerful tool for analyzing intricate causal relationships, offering significant improvements over existing methods in both accuracy and interpretability.

Acknowledgments

This work is supported by National Key Research and Development Program of China (No. 2023YFF0725500), NSFC (62031003, 62272276 and 62432006), Shandong Provincial Natural Science Foundation (No. ZR2024JQ001) and Taishan Scholars Program (No. tsqn202306007).

References

- Breakspear, M. 2017. Dynamic models of large-scale brain activity. *Nat. Neurosci.*, 20(3): 340–352.
- Cai, R.; Zhu, Y.; Qiao, J.; Liang, Z.; Liu, F.; and Hao, Z. 2024. Where and How to Attack? A Causality-Inspired Recipe for Generating Counterfactual Adversarial Examples. In *AAAI*, 11132–11140.
- Chen, W.; Qiao, J.; Cai, R.; and Hao, Z. 2024. On the role of entropy-based loss for learning causal structure with continuous optimization. *TNNLS*, 99(1): 1–12.
- Chickering, D. M. 2002a. Learning equivalence classes of Bayesian-network structures. *JMLR*, 2: 445–498.
- Chickering, D. M. 2002b. Optimal structure identification with greedy search. *JMLR*, 3(11): 507–554.
- Colombo, D.; Maathuis, M. H.; et al. 2014. Order-independent constraint-based causal structure learning. *JMLR*, 15(1): 3741–3782.
- Cui, P.; and Athey, S. 2022. Stable learning establishes some common ground between causal inference and machine learning. *Nat. Mach. Intell.*, 4(2): 110–115.
- Epstein, J. M.; and Axtell, R. 1996. *Growing artificial societies: social science from the bottom up*. Brookings Institution Press.
- He, Y.; Cui, P.; Shen, Z.; Xu, R.; Liu, F.; and Jiang, Y. 2021. DARING: Differentiable causal discovery with residual independence. In *KDD*, 596–605.
- Horgan, T. 1993. From Supervenience to Superdupervenience: Meeting the Demands of a Material World. *Mind*, 102(408): 555–586.
- Jang, E.; Gu, S.; and Poole, B. 2017. Categorical Reparametrization with Gumble-Softmax. In *ICLR*.
- Klein, B.; Swain, A.; Byrum, T.; Scarpino, S. V.; and Fagan, W. F. 2022. Exploring noise, degeneracy and determinism in biological networks with the einet package. *Methods in Ecology and Evolution*, 13(4): 799–804.
- Lee, H.-C.; Danieletto, M.; Miotto, R.; Cherng, S. T.; and Dudley, J. T. 2020. Scaling structural learning with NO-BEARS to infer causal transcriptome networks. In *PSB*, 391–402.
- Liang, J.; Wang, J.; Yu, G.; Domeniconi, C.; Zhang, X.; and Guo, M. 2024a. Gradient-based local causal structure learning. *TCYB*, 54(1): 486–495.
- Liang, J.; Wang, J.; Yu, G.; Guo, W.; Domeniconi, C.; and Guo, M. 2023. Directed acyclic graph learning on attributed heterogeneous network. *TKDE*, 35(10): 10845–10856.
- Liang, J.; Wang, J.; Yu, G.; Xia, S.; and Wang, G. 2024b. Multi-Granularity Causal Structure Learning. In *AAAI*, 13727–13735.
- Lu, S.; and Gao, T. 2023. Meta-DAG: Meta causal discovery via bilevel optimization. In *ICASSP*, 1–5.
- Ng, I.; Zhu, S.; Chen, Z.; and Fang, Z. 2019. A graph autoencoder approach to causal structure learning. *arXiv preprint arXiv:1911.07420*.
- Ng, I.; Zhu, S.; Fang, Z.; Li, H.; Chen, Z.; and Wang, J. 2022. Masked gradient-based causal structure learning. In *SDM*, 424–432.
- Peters, J.; Mooij, J. M.; Janzing, D.; and Schölkopf, B. 2014. Causal Discovery with Continuous Additive Noise Models. *JMLR*, 15: 2009–2053.
- Rubenstein, P.; Weichwald, S.; Bongers, S.; Mooij, J.; Janzing, D.; Grosse-Wentrup, M.; and Schölkopf, B. 2017. Causal Consistency of Structural Equation Models. In *UAI*, 808–817.
- Sachs, K.; Perez, O.; Pe’er, D.; Lauffenburger, D. A.; and Nolan, G. P. 2005. Causal protein-signaling networks derived from multiparameter single-cell data. *Science*, 308(5721): 523–529.
- Shimizu, S.; Hoyer, P. O.; Hyvärinen, A.; Kerminen, A.; and Jordan, M. 2006. A linear non-Gaussian acyclic model for causal discovery. *JMLR*, 7(10): 2003–2030.
- Spirtes, P.; Glymour, C. N.; Scheines, R.; and Heckerman, D. 2000. *Causation, prediction, and search*. MIT Press.
- Spirtes, P. L.; Meek, C.; and Richardson, T. S. 2013. Causal inference in the presence of latent variables and selection bias. *arXiv preprint arXiv:1302.4983*.
- Tononi, G.; and Koch, C. 2015. Consciousness: here, there and everywhere? *Philosophical Transactions of the Royal Society B: Biological Sciences*, 370(1668): 20140167.
- Vandenbroucke, J. P.; Broadbent, A.; and Pearce, N. 2016. Causality and causal inference in epidemiology: the need for a pluralistic approach. *Int. J. of Epi.*, 45(6): 1776–1786.
- Wang, Y.; Zhang, A.; Wang, X.; Yuan, Y.; He, X.; and Chua, T.-S. 2022. Differentiable invariant causal discovery. *arXiv preprint arXiv:2205.15638*.
- Wei, J.; Tay, Y.; Bommasani, R.; Raffel, C.; Zoph, B.; Borgeaud, S.; Yogatama, D.; Bosma, M.; Zhou, D.; Metzler, D.; et al. 2022. Emergent Abilities of Large Language Models. In *TMLR*.
- Yang, D.; He, X.; Wang, J.; Yu, G.; Domeniconi, C.; and Zhang, J. 2024. Federated Causality Learning with Explainable Adaptive Optimization. In *AAAI*, 16308–16315.
- Yang, D.; Yu, G.; Wang, J.; Wu, Z.; and Guo, M. 2023. Reinforcement causal structure learning on order graph. In *AAAI*, 10737–10744.
- Yu, Y.; Chen, J.; Gao, T.; and Yu, M. 2019. DAG-GNN: DAG structure learning with graph neural networks. In *ICML*, 7154–7163.
- Zhang, A.; Liu, F.; Ma, W.; Cai, Z.; Wang, X.; and Chua, T.-s. 2023. Boosting differentiable causal discovery via adaptive sample reweighting. *arXiv preprint arXiv:2303.03187*.
- Zhang, H.; Zhou, S.; and Guan, J. 2018. Measuring conditional independence by independent residuals: Theoretical results and application in causal discovery. In *AAAI*, 2029–2036.

Zheng, X.; Aragam, B.; Ravikumar, P.; and Xing, E. P. 2018. DAGs with NO TEARS: continuous optimization for structure learning. In *NeurIPS*, 9492–9503.

Zheng, X.; Dan, C.; Aragam, B.; Ravikumar, P.; and Xing, E. 2020. Learning sparse nonparametric dags. In *AISTAT*, 3414–3425.

Zhu, C.; Byrd, R. H.; Lu, P.; and Nocedal, J. 1997. Algorithm 778: L-BFGS-B: Fortran subroutines for large-scale bound-constrained optimization. *ACM TOMS*, 23(4): 550–560.