Smoothing the Landscape: Causal Structure Learning via Diffusion Denoising Objectives

Hao Zhu* HAOZHU233@GMAIL.COM
Beth Israel Deaconess Medical Center
Harvard Medical School
Boston, MA, USA

Di Zhou DI.ZHOU@TUFTS.EDU
Department of Computer Science
Tufts University
Medford, MA, USA

Donna Slonim DONNA.SLONIM@TUFTS.EDU
Department of Computer Science
Tufts University
Medford, MA, USA

Editors: Bijan Mazaheri and Niels Richard Hansen

Abstract

Understanding causal dependencies in observational data is critical for informing decision-making. These relationships are often modeled as Bayesian Networks (BNs) and Directed Acyclic Graphs (DAGs). Existing methods, such as NOTEARS and DAG-GNN, often face issues with scalability and stability in high-dimensional data, especially when there is a feature-sample imbalance. Here, we show that the denoising score matching objective of diffusion models could smooth the gradients for faster, more stable convergence. We also propose an adaptive k-hop acyclicity constraint that improves runtime over existing solutions that require matrix inversion. We name this framework Denoising Diffusion Causal Discovery (DDCD). Unlike generative diffusion models, DDCD utilizes the reverse denoising process to infer a parameterized causal structure rather than to generate data. We demonstrate the competitive performance of DDCDs on synthetic benchmarking data. We also show that our methods are practically useful by conducting qualitative analyses on two real-world examples. Code is available at this url: https://github.com/haozhu233/ddcd.

Keywords: causal discovery, structural equation models, denoising models

1. Introduction

Learning network structures representing causal dependencies from high-dimensional observational data has long been a crucial goal. It plays an important role in multiple disciplines, including genetics, epidemiology, and economics (Pearl, 2009; Uhler et al., 2013; Spirtes and Zhang, 2016; Pingault et al., 2018; Glymour et al., 2019), and plays a growing role in healthcare research (Slonim, 2002; Chandak et al., 2023).

Such networks, where nodes represent feature variables and edges capture potential relationships, are often represented as Directed Acyclic Graphs (DAGs), which allow no cycles. There has been a great deal of prior work developing methods to address this problem. For example, the PC

* Work done while at the Department of Computer Science, Tufts University

algorithm (Spirtes and Glymour, 1991) is a constraint-based approach that iteratively tests conditional independence, GES (Chickering, 2002) searches for the causal structure maximizing a scoring function, and LiNGAM (Shimizu et al., 2012) relies on the structural equation model (SEM). While these early methods face scalability challenges, subsequent work has extended constraint-based and score-based approaches to high-dimensional settings, including methods for latent confounders (Ogarrio et al., 2016; Colombo et al., 2012; Pal et al., 2025; Ribeiro and Heider, 2025) and efficient Bayesian network structure learning (Zhu et al., 2021).

In the continuous optimization paradigm, NOTEARS (Zheng et al., 2018) introduced a continuous acyclicity score that can be solved by a regular numerical optimizer. Many other studies have since extended this formulation, focusing on scalability (Yu et al., 2019; Lee et al., 2019; Ng et al., 2020; Yu et al., 2021), convexity (Bello et al., 2022; Deng et al., 2023; Ng et al., 2024), sparsity control (Wei et al., 2020; Ng et al., 2020), extension to interventional data (Brouillard et al., 2020), and non-linearity (Yu et al., 2019; Ng et al., 2019; Zheng et al., 2020; Yang et al., 2021; Shen et al., 2022; Ng et al., 2022; Kalainathan et al., 2022; Lachapelle et al., 2019). Variations of this model have also been applied to a wide range of settings, such as time series (Pamfil et al., 2020; Sun et al., 2021; Shang et al., 2021) and gene networks (Shu et al., 2021; Agamah et al., 2022; Zhu and Slonim, 2024). There are also discussions of the application of such models to datasets with unequal variances and different data types (Reisach et al., 2021; Kaiser and Sipos, 2021; Ng et al., 2024). Methods that focus on topological ordering instead of a structural DAG constraint have also been explored (Sanchez et al., 2022).

We propose Denoising Diffusion Causal Discovery (DDCD), inspired by the forward diffusion perturbation and denoising score matching objective of DDPMs (Ho et al., 2020). Unlike generative DDPMs, DDCD repurposes the reverse process to learn structural parameters that best denoise the data. Our contributions are the following:

- We prove that the denoising score matching objective could be used for causal structural. We inference and show that it helps avoid sharp local minima and encourage smooth gradients.
- We propose an adaptive k-hop acyclicity constraint that transitions from local neighborhood to global enforcement, significantly reducing runtime while guaranteeing valid DAG recovery.
- We introduce a permutation-invariant batch sampling strategy that decouples optimization complexity from sample size, ensuring consistent convergence and scalability.
- We propose DDCD-Smooth, which addresses the "varsortability" problem in continuous causal discovery by normalizing features to equal scales. This prevents methods from exploiting variance differences as an artifact for identification, making the approach robust to heterogeneous feature scales in real-world data.

2. Background and Related Work

2.1. Problem Statement

Given a dataset $\boldsymbol{X} \in \mathbb{R}^{n \times d}$, with n samples and d feature variables, the objective is to learn a dependency graph $\mathcal{G}$ represented by the weighted adjacency matrix $\boldsymbol{W} \in \mathbb{R}^{d \times d}$. Such a graph is often defined as a Bayesian Network (BN), which permits no cycles.

To ensure the identifiability of $\mathcal{G}$ and the validity of our proposed denoising objective, we make the following standard assumptions common in causal discovery literature (Spirtes et al., 2000; Pearl, 2009): **Assumption 1 (Causal Sufficiency)**: there are no unobserved confounders influencing multiple observed variables. **Assumption 2 (No Selection Bias)**: the samples in $\boldsymbol{X}$ are drawn i.i.d. from the underlying distribution with no selection bias. **Assumption 3 (Acyclicity)**: the true causal structure $\mathcal{G}$ is a DAG. **Assumption 4 (Additive Noise Model)**: the data follows a structural equation model $x_j = f_j(PA_j) + z_j$ with additive noise. **Assumption 5 (Identifiability)**: The error terms z are jointly independent. For unique identifiability beyond the Markov Equivalence Class, we assume errors are either non-Gaussian (Shimizu et al., 2006) or Gaussian with equal variances (Peters and Bühlmann, 2014). **Assumption 6 (Faithfulness)**: The joint distribution P(X) is faithful to $\mathcal{G}$.

2.2. Structural Equation Models (SEMs) and Continuous DAG constraint

Structural Equation Models (SEMs) (Grace, 2006; Shipley, 2016; Kline, 2023) provide a framework to model variable dependencies. For a linear SEM, we simply assume that each variable is a linear combination of its parents with some noise. In its matrix multiplication form, we have

$$\boldsymbol{X} = \boldsymbol{X}\boldsymbol{W} + \boldsymbol{E}, \tag{1}$$

where $\boldsymbol{E} \in \mathbb{R}^{n\times d}$ captures the error terms. Based on this assumption, many existing SEMs aim to estimate matrix $\boldsymbol{W}$ such that the reconstruction error is minimized (van de Geer and Bühlmann, 2013; Loh and Bühlmann, 2014; Zheng et al., 2018). Since the adjacency matrix is often sparse, many methods choose to add either L1 or L2 regularization on $\boldsymbol{W}$ to encourage sparsity (Vowels et al., 2022). In this case, we have the following training objective,

$$\min_{W} \frac{1}{2n}\|\boldsymbol{X} - \boldsymbol{X}\boldsymbol{W}\|_F^2 + \lambda_1\|\boldsymbol{W}\|_1 + \lambda_2\|\boldsymbol{W}\|_2. \tag{2}$$

While traditional approaches often rely on combinatorial optimization, Zheng et al. (2018) proposed NOTEARS, which introduced a continuous score characterizing graph acyclicity:

$$h(\boldsymbol{W}) = \text{tr}(e^{\boldsymbol{W}\circ\boldsymbol{W}}) - d, \tag{3}$$

where $\circ$ is the Hadamard product, $e^{\boldsymbol{W}}$ is the matrix exponential of $\boldsymbol{W}$, and tr() is the trace of a matrix. Essentially, matrix $\boldsymbol{W}$ is a DAG if and only if $h(\boldsymbol{W}) = 0$. Since the function $h(\boldsymbol{W})$ has a simple and smooth gradient function, it can be used in many gradient-based continuous optimization algorithms. In terms of complexity, since the score function $h(\boldsymbol{W})$ requires the matrix exponential, the runtime of NOTEARS is at least $\mathcal{O}(d^3)$ (Arioli et al., 1996).

2.3. Denoising Diffusion Probabilistic Models

Denoising Diffusion Probabilistic Models (DDPMs) are a class of generative model that shows strong performance in modeling complex data distributions (Ho et al., 2020). A typical DDPM starts from a non-parameterized forward diffusion process. Given an unperturbed input $\boldsymbol{x}_0$, the forward process aims to generate a series of noisy samples $\boldsymbol{x}_0, \boldsymbol{x}_1, ..., \boldsymbol{x}_T$ over T steps, where $\boldsymbol{x}_T$ usually stands for pure noise. In each step, a small amount of Gaussian noise is gradually introduced following a diffusion schedule β as shown in Equation 4.

$$\boldsymbol{x}_t = \sqrt{1-\beta_t}\boldsymbol{x}_{t-1} + \sqrt{\beta_t}\boldsymbol{z}_{t-1} \tag{4}$$

Here, $\boldsymbol{z}_{t-1} \sim \mathcal{N}(0,1)$, so Equation 4 is essentially trying to reduce the means to $\mathbf{0}$ while increasing the variances to $\mathbf{1}$. With reparameterization, it can be rewritten as Equation 5:

$$\boldsymbol{x}_t = \sqrt{\overline{\alpha_t}}\boldsymbol{x}_0 + \sqrt{1-\overline{\alpha_t}}\boldsymbol{z}, \tag{5}$$

where $\overline{\alpha_t} = \prod_{i=0}^{t}(1-\beta_i)$ and $\boldsymbol{z} \sim \mathcal{N}(0,1)$, so the noisy data x_t can be generated in one step.

The actual modeling piece of DDPM is the reverse model, which is trained to predict the added noise $\boldsymbol{z}$ and to denoise the data. The choice of model for the reverse process depends on the input data. Recent studies have suggested similarity between diffusion models and a generalized form of variational autoencoder (VAE) (Kingma, 2013) with infinite latent spaces (Luo, 2022).

3. Methods

3.1. Denoising Objective for Linear SEMs & NOTEARS

Inspired by the denoising objective in DDPMs, we propose an alternative training objective for learning the adjacency matrix $\boldsymbol{W}$ in SEMs. We augment each sample in the same way as the forward diffusion process in Equation 5. Then, we will optimize a reverse model with the parameterized $\boldsymbol{W}$ to predict the added noise. Note that unlike standard diffusion models where the reverse model learns a time-dependent transition kernel, our reverse model focuses on the single optimal $\boldsymbol{W}$ that minimizes denoising error across all noise levels.

For linear SEMs, the reverse model is trying to minimize the following denoising objective:

$$\min_{W} \frac{1}{2n}\|(\boldsymbol{X_t} - \boldsymbol{X_t}\boldsymbol{W}) - \text{diag}(\sqrt{1-\overline{\alpha_t}})\boldsymbol{Z}(\boldsymbol{I}-\boldsymbol{W})\|_F^2 + \lambda_1\|\boldsymbol{W}\|_1 + \lambda_2\|\boldsymbol{W}\|_2. \tag{6}$$

Here, following diffusion literature, we use $t \in \{1,\dots,T\}$ to denote diffusion time steps, where T is the total number of diffusion steps. The original (unperturbed) data is $\boldsymbol{X}_0$, and $\boldsymbol{X_t}$ denotes data perturbed at diffusion step $\boldsymbol{t}$, which is a vector of diffusion time steps for all samples. The cumulative noise schedule is $\overline{\alpha}_t = \prod_{i=1}^{t}(1-\beta_i)$, where β_i is the noise added at step i. The random noise matrix is $\boldsymbol{Z} \sim \mathcal{N}(0,\mathbf{I})$. $\text{diag}(\boldsymbol{v})$ is the diagonal operator that converts vector $\boldsymbol{v}$ to a diagonal matrix.

Theorem 1 (Objective Equivalence) *For linear SEMs, minimizing the denoising objective in Eq. 6 is equivalent to minimizing the standard SEM reconstruction loss in Eq. 2. Therefore, under Assumption 1-6 in section 2.1, the denoising objective can be used for causal structure learning.*

Proof Consider the case when each sample in $\boldsymbol{X}$ is perturbed using the forward diffusion process in 5. The perturbed observational data could be written as in Equation 7. Here we use $\boldsymbol{t}$ to denote the different diffusion time steps for all samples in $\boldsymbol{X}$ and the diag() operator scales each row of $\boldsymbol{X}_0$ and $\boldsymbol{Z}$ accordingly based on the diffusion schedule.

$$\boldsymbol{X_t} = \text{diag}(\sqrt{\overline{\alpha_t}})\boldsymbol{X_0} + \text{diag}(\sqrt{1-\overline{\alpha_t}})\boldsymbol{Z}, \tag{7}$$

With Equations 7 and 1, we can develop the following equations.

$$\boldsymbol{X_t W} = \text{diag}(\sqrt{\overline{\alpha_t}})\boldsymbol{X_0 W} + \text{diag}(\sqrt{1-\overline{\alpha_t}})\boldsymbol{Z W}, \tag{8}$$

$$\boldsymbol{X_t} - \boldsymbol{X_t W} = \text{diag}(\sqrt{\overline{\alpha_t}})(\boldsymbol{X_0} - \boldsymbol{X_0 W}) + \text{diag}(\sqrt{1-\overline{\alpha_t}})(\boldsymbol{Z} - \boldsymbol{Z W}), \tag{9}$$

$$\text{diag}(\sqrt{\overline{\alpha_t}})(\boldsymbol{X_0} - \boldsymbol{X_0 W}) = (\boldsymbol{X_t} - \boldsymbol{X_t W}) - \text{diag}(\sqrt{1-\overline{\alpha_t}})\boldsymbol{Z}(\boldsymbol{I} - \boldsymbol{W}). \tag{10}$$

Recall that in Equation 2, the main objective is to minimize $\boldsymbol{X_0} - \boldsymbol{X_0 W}$. The proof above shows that minimizing $\boldsymbol{X_0} - \boldsymbol{X_0 W}$ is equivalent to minimizing the right-hand side, which measures the distance between $(\boldsymbol{X_t} - \boldsymbol{X_t W})$ and $\text{diag}(\sqrt{1-\overline{\alpha_t}})\boldsymbol{Z}(\boldsymbol{I} - \boldsymbol{W})$. Therefore, the denoising objective in Equation 6 is algebraically equivalent to the standard SEM reconstruction objective.

■

Notes on Gradient Smoothing: Beyond objective equivalence, the denoising objective helps adjust the optimization dynamics. By minimizing loss over perturbed data $\boldsymbol{X}_t$, we are in fact performing Randomized Smoothing (Duchi et al., 2012; Nesterov and Spokoiny, 2017). This places an upper bound on the curvature of the loss landscape (the Lipschitz constant of the gradient) and effectively smooths the sharp local minima common in high-dimensional SEMs (Keskar et al., 2016). Consequently, the optimizer avoids volatile jumps and converges more reliably, even with limited samples. To illustrate this point, we implemented a toy model called NOTEARS-Denoising with the denoising diffusion process, while everything else is the same as in the original implementation of NOTEARS-Linear.

3.2. Adaptive k-hop Acyclicity Constraint with gradient clipping

In practice, the calculation of NOTEARS' DAG constraint involves the matrix exponential, so its runtime is at least $\mathcal{O}(d^3)$. One potential optimization here is to enforce a "k-hop acyclicity constraint" that only checks the acyclicity score within local neighborhoods of k-hops. By keeping a running sum, we replace the matrix exponential with just matrix multiplication, reducing the runtime and theoretical complexity. The formula for the k-hop acyclicity constraint is in Equation 11, where γ is a scaling factor. A detailed discussion of this is provided in Appendix A.1.

$$h(\boldsymbol{W}, k, \gamma) = \sum_{j=1}^{k+1} \frac{1}{j!\gamma^{2j}} \text{tr}((\gamma \boldsymbol{W} \circ \gamma \boldsymbol{W})^j) \tag{11}$$

However, fixing k to a small constant value introduces a theoretical risk of longer-range cycles. To eliminate this risk without sacrificing training efficiency, we introduce an *Adaptive Acyclicity Curriculum*. Motivated by the observation that early optimization steps on the adjacency matrix $\boldsymbol{W}$ are noisy, we suggest that enforcing global acyclicity at the start of training is computationally wasteful. Instead, we dynamically schedule k during optimization. Let τ be the current training iteration and N_{iter} be the total number of iterations.[1] We define k_τ as:

$$k_\tau = \begin{cases} 3 & \text{if } \tau < 0.4 N_{\text{iter}} \quad \text{(Local Alignment)} \\ 10 & \text{if } 0.4 N_{\text{iter}} \le \tau < 0.9 N_{\text{iter}} \quad \text{(Structure Refinement)} \\ d & \text{if } \tau \ge 0.9 N_{\text{iter}} \quad \text{(Global Guarantee)} \end{cases} \tag{12}$$

1. We use τ and N_{iter} for training iterations to distinguish from diffusion time steps t and T_{diff}.

This schedule proceeds in three phases: **1. Local Alignment:** In the early phase, we restrict constraints to a small neighborhood ($k = 3$). This prevents tight feedback loops (e.g., $A \to B \to A$) while allowing the optimizer to rapidly traverse the loss landscape. **2. Structure Refinement:** As the adjacency matrix sparsifies, we increase k to detect medium-range cycles. **3. Global Guarantee:** In the final phase, we set k equal to the number of feature variables d. This is mathematically equivalent to the full NOTEARS constraint (Zheng et al., 2018), guaranteeing a valid DAG, but it runs quickly because W is usually sparse at this stage.
This strategy ensures our final output is a theoretically-valid DAG, while maintaining the practical computational efficiency of the k-hop approximation for the vast majority of the training process.

3.3. Permutation-Invariant Batch Sampling

Existing causal discovery algorithms often operate on the full data matrix $\boldsymbol{X} \in \mathbb{R}^{n \times d}$. To improve scalability and stability across varying dataset sizes, we employ a fixed-size bootstrap sampling strategy from RegDiffusion (Zhu and Slonim, 2024). This approach is theoretically grounded in the 'Deep Sets' perspective (Zaheer et al., 2017). Since causal structure learning is permutation invariant in terms of samples, we could treat data as an exchangeable set. This allows us to use fixed-size mini-batches with resampling as uniform Monte Carlo estimators of the underlying population. At the same time, we also decouple algorithm runtime from the total sample size n and gain more stable and consistent behavior on datasets of different sizes.

3.4. DDCD Models

With the new denoising objective in mind, we propose a set of 3 DDCD models: DDCD Linear, DDCD Nonlinear, and DDCD Smooth. All three models, as illustrated in Figure 1, are trained with the denoising objective, the diffusion training framework, and the k-hop acyclicity constraint.

3.4.1. DDCD LINEAR

The only trainable parameters in DDCD-Linear are values in the adjacency matrix $\boldsymbol{W}$. The input of the model includes the perturbed $\boldsymbol{X_t}$, the diffusion variance schedule $\sqrt{1-\overline{\alpha_t}}$, and the noise term $\boldsymbol{Z}$. All three inputs are generated during the forward process given $\boldsymbol{X_0}$. The model is trained with both L1 and L2 regularization on the adjacency matrix together with the k-hop acyclicity constraint. The model is optimized using the Adam optimizer with increasing weights on the DAG constraint. We will explain this in detail in section 3.5.

3.4.2. DDCD NONLINEAR

Applying denoising diffusion to non-linear SEMs presents a challenge in that the added noise $\boldsymbol{Z}$ cannot be easily separated from $\boldsymbol{X}$ in the original feature space. To address this, we adopt a Generalized SEM formulation. Here we assume the observed data $\boldsymbol{X}$ is the result of a invertible element-wise nonlinear transformation of latent variable $\boldsymbol{Y}$, which interact via a linear causal mechanism.

$$\boldsymbol{Y} = \boldsymbol{Y}\boldsymbol{W} + \boldsymbol{Z}_{SEM}, \quad \boldsymbol{X} = f_2(\boldsymbol{Y}) + \boldsymbol{E} \tag{13}$$

where $\boldsymbol{Z}_{SEM}$ is the noise in the latent causal system on $\boldsymbol{Y}$, $\boldsymbol{W}$ is the weighted adjacency matrix encoding the causal DAG, and f_2 is a nonlinear mapping (decoder). To recover this structure, we introduce an inference network f_1 (encoder) such that $\boldsymbol{Y} = f_1(\boldsymbol{X})$. By jointly learning f_1, f_2 and

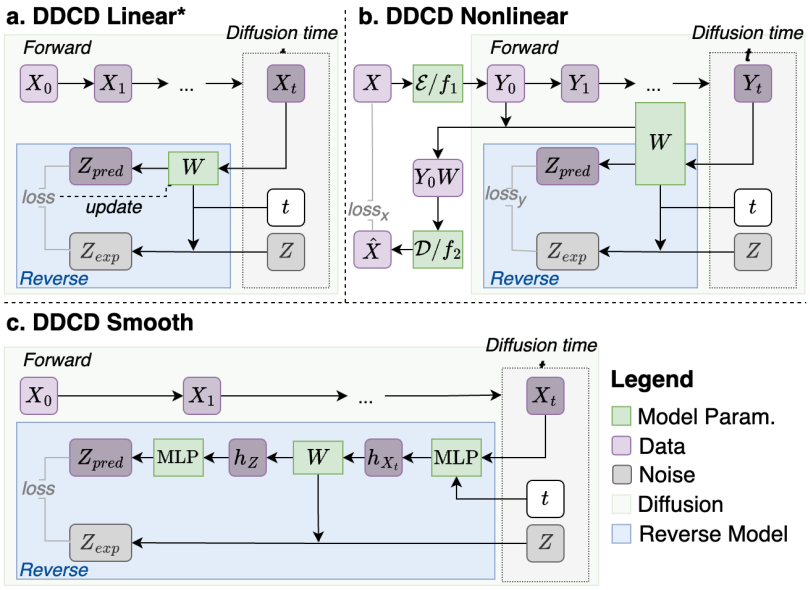

Figure 1: Model architectures of proposed models in this paper. Note that the reverse path does not generate new samples and the only focus is to learn W via the denoising objective.

$\boldsymbol{W}$, we aim to recover the latent causal space $\boldsymbol{Y}$ where linear dependencies hold. The structural equation becomes:

$$\boldsymbol{X} = f_2(f_1(\boldsymbol{X})\boldsymbol{W}) + \boldsymbol{E}, \text{ or } \boldsymbol{X} = f_2(\boldsymbol{Y}\boldsymbol{W}) + \boldsymbol{E}_1 \tag{14}$$

This formulation essentially treats the causal discovery problem as learning an autoencoder where the latent space is constrained by a structural prior $\boldsymbol{W}$, which is identical to the work proposed in DAG-GNN (Yu et al., 2019), GAE (Ng et al., 2019), and recent robust variants for zero-inflated data (Zhu and Slonim, 2025).

The total training objective combines the nonlinear reconstruction loss, the latent diffusion denoising loss, and the DAG constraints:

$$\min_{W,f_1,f_2} \underbrace{\frac{1}{2n}\|X - f_2(f_1(X)W)\|_F^2}_{\text{SEM Reconstruction}} + \underbrace{\|(Y_t - Y_tW) - \text{diag}(\sqrt{1-\bar{\alpha}_t})Z(I-W)\|_F^2}_{\text{Latent Denoising}} + \mathcal{R}(W). \tag{15}$$

where $\mathcal{R}(W)$ includes the sparsity and k-hop acyclicity constraints. By anchoring the latent representation Y to a denoising objective, we encourage the model to learn a representation where the noise Z is additive and Gaussian, aligning with the identifiability conditions for linear causal models.

This architecture, as shown in Figure 1b, is very similar to a Latent Diffusion Model (LDM) (Rombach et al., 2022), where noise is added to the latent representation learned by an autoencoder. We can treat $\boldsymbol{Y}$ as the unobserved latent variable, and the learned nonlinear transformation functions f_1 and f_2 can be viewed as the encoder and decoder. Furthermore, the learned adjacency matrix $\boldsymbol{W}$ for the SEM could be considered as a form of attention or graph neural network.

3.4.3. DDCD SMOOTH

While continuous optimization algorithms excel on synthetic SEMs, recent studies (Reisach et al., 2021; Kaiser and Sipos, 2021) show that their performance often degrades on real-world datasets, where feature scales vary. Reisach et al. (2021) demonstrated that these methods can exploit "varsortability" and infer directionality strictly from increasing variance instead of identifying true causal mechanisms. This is problematic because 1. it violates the equal-variance assumption implicit in standard SEM formulations; 2. real-world data rarely exhibits the same patterns as synthetic data (Kaiser and Sipos, 2021). To address this limitation, we propose DDCD-Smooth. Instead of learning exact edge weights, DDCD-Smooth learns a normalized adjacency matrix where expected edge weights are scaled to $\frac{1}{d}$. We achieve this by employing an MLP with **Tanh activation to normalize all features to the range of** $[-1, 1]$. This normalization serves two purposes: (1) it prevents the method from exploiting variance differences for spurious identification, and (2) it makes the method robust to heterogeneous feature scales common in real-world data. An illustration of the architecture of this model is provided in Figure 1c. In appendix A.2, we provide a proof showing that with normalized adjacency, we can directly estimate the noise $\boldsymbol{Z}$ when the graph is large.

3.5. Optimization

In this experiment, the models are optimized using the Adam optimizer (Kingma, 2014) since it has fewer restrictions. Distinct from NOTEARS, we replace the dual-ascending augmented Lagrangian formulation with a linear penalty schedule for the acyclicity constraint (See Supplement Section A.6 for our justification and empirical evidence).

4. Results

4.1. Denoising Objective Smooths the Optimization Landscape

We first compare NOTEARS-Denoising against NOTEARS-Linear to isolate the impact of the denoising objective on optimization. Except for the noise perturbation process and the denoising objective, both models are identical and are optimized using dual-ascending augmented Lagrangian with the L-BFGS-B optimizer. We evaluate performance on a synthetic Scale-Free graph (20 nodes, degree 10) and analyze gradient behavior and convergence speed.

Note that NOTEARS-Linear (Zheng et al., 2018) converges slowly when the number of samples is very limited (Figure 2a). This strange behavior occurs because the reconstruction loss landscape contains sharp local minima that produce volatile gradients during optimization. Recent work has also identified instabilities in the acyclicity constraint itself (Nazaret et al., 2023), though these manifest differently (vanishing constraint gradients vs. volatile loss gradients). Our denoising objective addresses the loss landscape component: it takes many steps for the L-BFGS-B optimizer to converge in each iteration (as shown in Figure 2b and c). In contrast, the denoising objective inherently smooths these gradients, allowing the optimizer to bypass sharp minima and converge in a fraction of the steps. Additional comparisons are provided in Appendix section A.5 and Figure 6.

4.2. DDCD Demonstrates Robust Scalability and Accuracy on Synthetic Benchmarks

We evaluated DDCD models across a range of synthetic Scale-Free (SF) and Erdős-Rényi (ER) graphs (d = 20 - 5,000, degree = 10 - 500) and we compared the performances against a set of

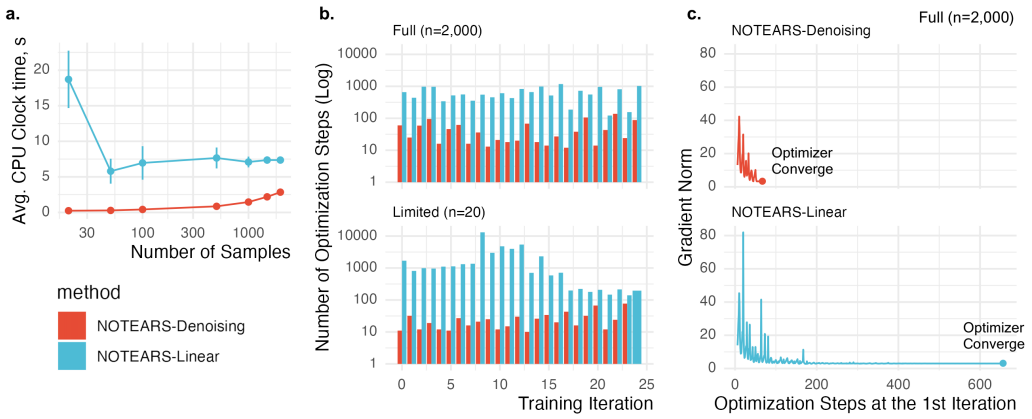

Figure 2: **The denoising objective accelerates convergence by smoothing gradients.** (a) Average runtime vs. sample size (10 runs). (b) L-BFGS-B steps to convergence; NOTEARS-Denoising requires significantly fewer iterations. (c) Gradient norm trajectories during the first iteration on full data ($n = 2,000$); the denoising objective yields a stable descent compared to the volatile gradients of the linear baseline. (First 5 steps omitted for clarity).

continuous optimization baselines, including NOTEARS (Zheng et al., 2018), NOTEARS-MLP, DAG-GNN (Yu et al., 2019), GOLEM (Ng et al., 2020), GAE (Ng et al., 2019), and DAGMA (Bello et al., 2022). Baseline methods use default hyperparameters from their original papers. For DDCD, we selected hyperparameters via grid search (Supplementary Figure 8): $T = 5000$, linear noise schedule, start noise=10^{-4}, end noise=0.02. The synthetic data were generated via additive noise models (ANMs) with Gaussian noise, using a linear and seven distinct nonlinear functions, consist of smooth ($\sin, \cos, \tanh$), non-smooth (ReLU), bounded (Sigmoid, $\tanh$), and unbounded (polynomial). Performance was assessed via Structural Hamming Distance (SHD), True Positive Rate (TPR), False Discovery Rate (FDR), False Positive Rate (FPR), and runtime (see Appendix A.4 for metric definitions and a note on identifiability).

Linear Structure Recovery. On linear benchmarks, the performance of DDCD-Linear (converged at 5,000 steps) matches or exceeds state-of-the-art algorithms such as DAGMA and GOLEM (Figure 3b) with an exception on the small ER graphs, while only consuming a fraction of time (Figure 3c). In terms of runtime, all DDCD models manage to converge within roughly 10 seconds, while other algorithms sometimes take 20 minutes to converge. Visual inspection (Figure 3a) also confirms that DDCD-Linear yield near perfect structure recovery when n is large (SHD = 9, n=2,000) while recovering most of the structure when n is small (SHD=233, n=20).

Nonlinear Disentanglement. On nonlinear benchmarks, DDCD-Nonlinear (converged at 1,000 steps) shows superior performance on structure recovery (Figure 3f) and inference speed. Visual inspection (Figure 3d) shows that it was able to recover most of the structures (SHD=107), even though it is a bit challenging to figure out the signs correctly. One unique feature of DDCD-Nonlinear is that it could learn a precise approximation of the underlying generation mechanism; as shown in (Figure 3e), the recovered function (blue dots) tightly fits the ground truth cosine transformation ($y = \cos(x + 1)$), which further validates the model's capacity for joint structure and function learning.

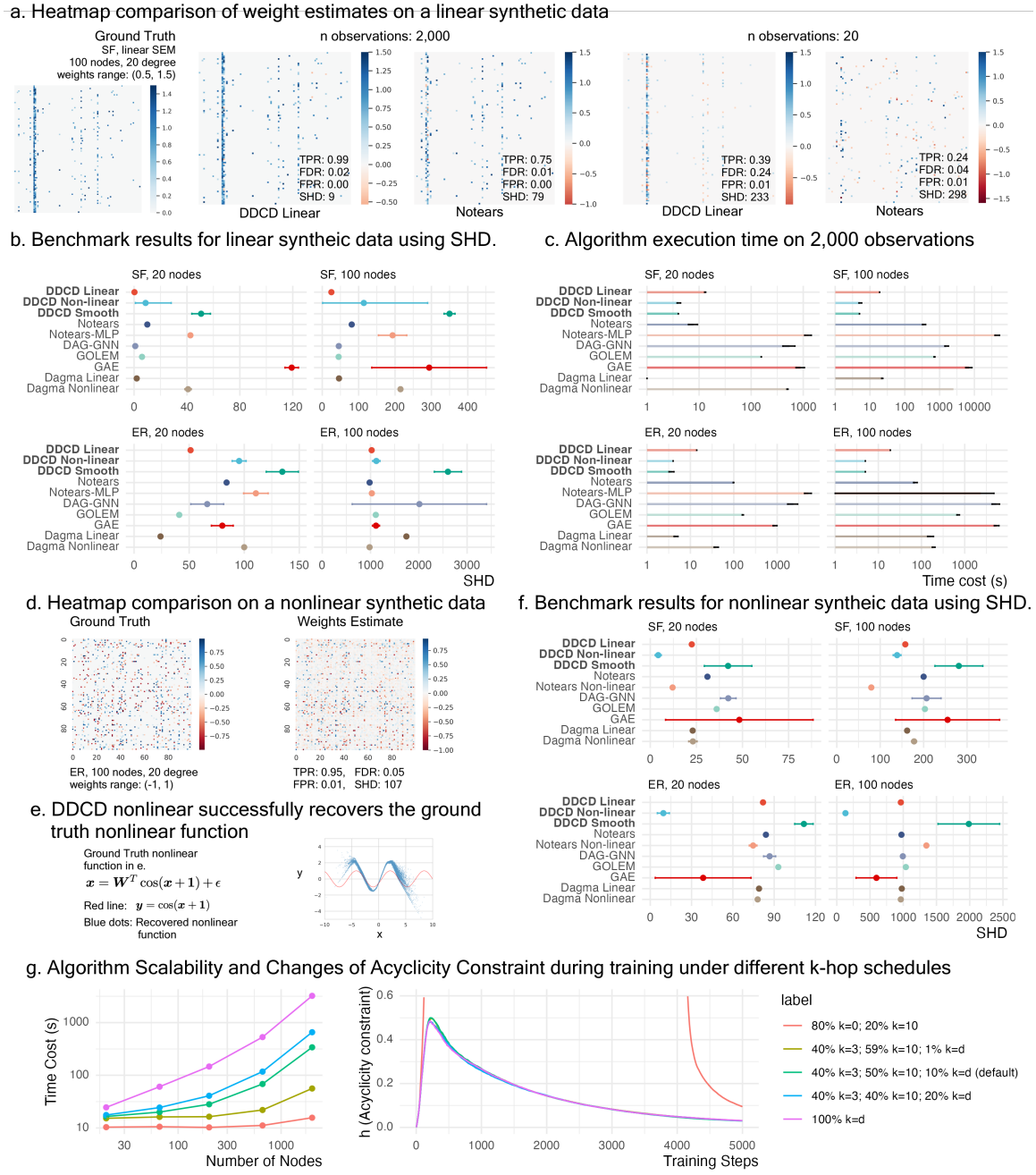

Figure 3: DDCD Linear and Nonlinear models demonstrate robust scalability and accuracy on synthetic benchmarks.

Computational Scalability. One major advantage of the DDCD framework is its runtime efficiency. When the size of the graph increases from 20 to 100, we see dramatic increases in runtime for standard algorithms (e.g., NOTEARS: 7 seconds to 6 minutes for a SF graph); the runtime of

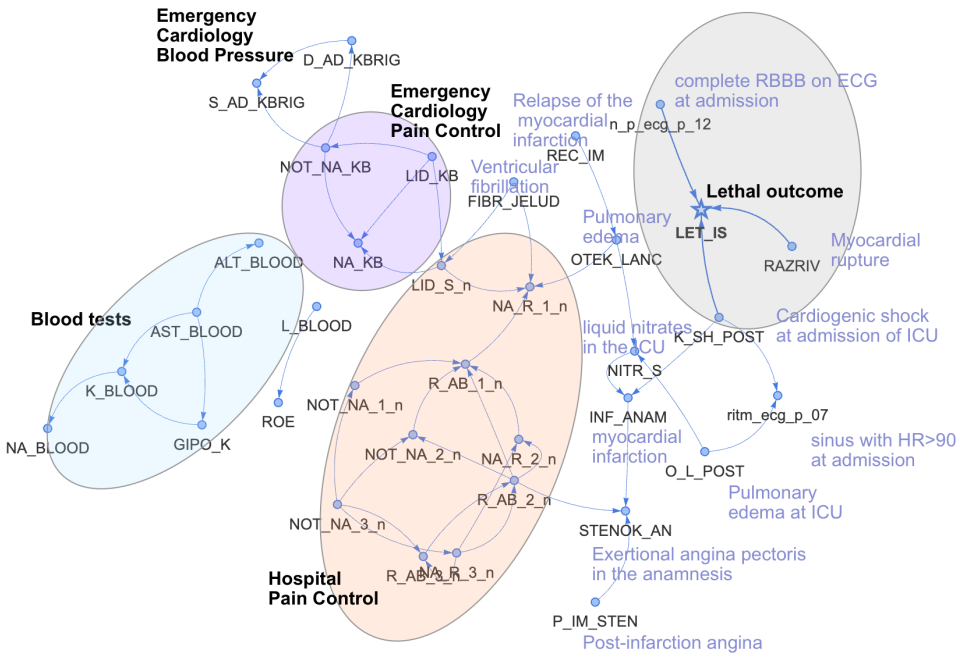

Figure 4: Inferred Causal Network around Lethal Outcome in Myocardial Infarction

DDCD-Linear increased from 13 seconds to 19 seconds (Figure 3c). The same patterns persist in ER graphs and under nonlinear mechanisms. Figure 3g shows the scale of time costs as we increase the size of the graph (the number of nodes increases from 20 to 2,000). With DDCD-Linear + the default k-hop schedule (40% k=3; 50% k=10; 10% Full DAG), it only takes 5.7 minutes to finish 5,000 steps of inference on a graph with 2,000 nodes on a modern GPU (NVIDIA H-100).

Adaptive k-hop acyclicity constraint helps resolve DAG violation. To understand the impact of the k-hop schedule on the training dynamics, we studied the time costs and the changes of acyclicity constraint under a few different schedules and showed the results in Figure 3g. Applying different k-hop schedules significantly affects algorithm speed. For example, using global DAG checks throughout training (100% k=d) requires 53.7 minutes for a 2,000 node graph, while our default schedule (40% k=3; 50% k=10; 10% k=d) completes in 5.7 minutes—a nearly 90% reduction. Importantly, the acyclicity constraint h follows near-identical trajectories in both cases, confirming that the k-hop schedule achieves the same acyclicity guarantees with substantially lower computational cost. Detailed analysis of training dynamics under various k-hop schedules is provided in Supplement §A.7.

4.3. Qualitative Analysis on Real Data

In this section, we assess the performance of DDCD Smooth on the real-world myocardial infarction dataset. This ran in 9 seconds on an NVIDIA H100 GPU. Due to the lack of ground truth networks for real data, we visualize regions of the inferred networks and assess quality via domain knowledge.

After training, we extracted all edges with weights above the cut-off threshold (0.2) in the inferred normalized adjacency matrix. This cut-off threshold was set by inspecting the distribution of edge weights and selecting the right tail. As an example, we examine a graph of the 2-hop neighbors around the Lethal Outcome (LET_IS) node (Figure 4). We identify with ovals several meaningful node clusters in this graph, including lethal outcome with its primary causes, critical conditions and interventions, hospital pain control, emergency cardiology pain control, and blood test results, purely based on the topological relationships among the nodes. The edge weights for a replicate of this network are provided in Appendix A.13. In comparison, as shown in Appendix A.14 and Figure 13, networks generated by NOTEARS appear less relevant.

In the network generated by DDCD Smooth, the three most important direct causes of lethal outcome include myocardial rupture, cardiogenic shock, and complete Right Bundle Branch Block (RBBB) on ECG at admission; all of these are known to be conditions with a poor prognosis. Cardiogenic shock (K_SH_POST) is further shown to cause "sinus ECG rhythm with heart rate > 90" (ritm_ecg_p_07), consistent with a high heart rate (tachycardia) being a symptom of cardiogenic shock (Kosaraju et al., 2025). Pulmonary edema (OTEK_LANC) is shown to cause the use of liquid nitrates in the ICU (NITR_S); this is indeed a common practice for rapidly managing pulmonary edema (Purvey and Allen, 2017). Other inferred edges include that NSAID drugs used by the emergency team (NOT_NA_KB) cause blood pressure to increase (Aljadhey et al., 2012), and that relapsing pain in the 2nd hospitalization period causes NSAID use in the same period.

There are also some node pairs for which plausible edges are inferred, but in an implausible direction. For example, in the lower right of the figure, "post-infarction angina" (chest pain after the heart attack causing the current hospital admission) is shown to cause "exertional angina pectoris in the anamnesis" (e.g., a reported history of chest pain after exercise), when the former clearly occurs after the latter. Still, many of the directed edges appear consistent with known causal relationships. Another similar analysis on real data on aging is included in Appendix A.12.

5. Discussion and Conclusion

In this work, we showed that we could leverage the denoising score-matching objectives of diffusion models to stabilize and accelerate causal structure learning. We demonstrated both theoretically and empirically that the denoising objective acts as a strong regularizer that smooths the loss landscape for faster and better convergence. It allowed the optimizer to avoid sharp local minima, particularly when the number of samples is limited. Furthermore, the introduction of the adaptive k-hop acyclicity curriculum provides a scalable solution to the computationally expensive matrix exponential. In our experiment, it helped reduce the training time by 90% while maintaining global acyclicity guarantees. With modern GPU support, the inference on a 2,000 node graph only took 5.7 minutes and the performance is on par with state of the art methods.

The DDCD-Smooth model also demonstrated strong practical performance on real-world clinical data. By learning a normalized adjacency matrix, the model extracted biologically and medically plausible causal clusters from the myocardial infarction and aging datasets. While there are inherent limitations of inferring directionality from static, cross-sectional data, the clarity of the clusters produced by DDCD-Smooth significantly outperformed the hub-and-spoke artifacts seen in previous methods. Ultimately, DDCD demonstrates that the principles of diffusion modeling can be repurposed to provide a stable, scalable, and explainable path for discovering dependencies in complex tabular data.

References

Francis E Agamah, Jumamurat R Bayjanov, Anna Niehues, Kelechi F Njoku, Michelle Skelton, Gaston K Mazandu, Thomas HA Ederveen, Nicola Mulder, Emile R Chimusa, and Peter AC 't Hoen. Computational approaches for network-based integrative multi-omics analysis. *Frontiers in Molecular Biosciences*, 9:967205, 2022.

Hisham Aljadhey, Wanzhu Tu, Richard A Hansen, Susan J Blalock, D Craig Brater, and Michael D Murray. Comparative effects of non-steroidal anti-inflammatory drugs (nsaids) on blood pressure in patients with hypertension. *BMC cardiovascular disorders*, 12:1–10, 2012.

M Arioli, B Codenotti, and C Fassino. The padé method for computing the matrix exponential. *Linear algebra and its applications*, 240:111–130, 1996.

Kevin Bello, Bryon Aragam, and Pradeep Ravikumar. Dagma: Learning dags via m-matrices and a log-determinant acyclicity characterization. *Advances in Neural Information Processing Systems*, 35:8226–8239, 2022.

Christopher R Bowie and Philip D Harvey. Administration and interpretation of the trail making test. *Nature protocols*, 1(5):2277–2281, 2006.

Philippe Brouillard, Sébastien Lachapelle, Alexandre Lacoste, Simon Lacoste-Julien, and Alexandre Drouin. Differentiable causal discovery from interventional data. In *Advances in Neural Information Processing Systems*, volume 33, pages 21865–21877, 2020.

Payal Chandak, Kexin Huang, and Marinka Zitnik. Building a knowledge graph to enable precision medicine. *Nature Scientific Data*, 2023. doi: https://doi.org/10.1038/s41597-023-01960-3. URL https://www.nature.com/articles/s41597-023-01960-3.

David Maxwell Chickering. Optimal structure identification with greedy search. *Journal of machine learning research*, 3(Nov):507–554, 2002.

Diego Colombo, Marloes H Maathuis, Markus Kalisch, and Thomas S Richardson. Learning high-dimensional directed acyclic graphs with latent and selection variables. *The Annals of Statistics*, pages 294–321, 2012.

Chang Deng, Kevin Bello, Bryon Aragam, and Pradeep Kumar Ravikumar. Optimizing notears objectives via topological swaps. In *International Conference on Machine Learning*, pages 7563–7595. PMLR, 2023.

John C Duchi, Peter L Bartlett, and Martin J Wainwright. Randomized smoothing for stochastic optimization. *SIAM Journal on Optimization*, 22(2):674–701, 2012.

Clark Glymour, Kun Zhang, and Peter Spirtes. Review of causal discovery methods based on graphical models. *Frontiers in genetics*, 10:524, 2019.

Daniel J Goble and Harsimran S Baweja. Postural sway normative data across the adult lifespan: Results from 6280 individuals on the balance tracking system balance test. *Geriatrics & gerontology international*, 18(8):1225–1229, 2018.

James B Grace. *Structural equation modeling and natural systems*. Cambridge University Press, 2006.

Jonathan Ho, Ajay Jain, and Pieter Abbeel. Denoising diffusion probabilistic models. *Advances in neural information processing systems*, 33:6840–6851, 2020.

Marcus Kaiser and Maksim Sipos. Unsuitability of notears for causal graph discovery. *arXiv preprint arXiv:2104.05441*, 2021.

Diviyan Kalainathan, Olivier Goudet, Isabelle Guyon, David Lopez-Paz, and Michèle Sebag. Structural agnostic modeling: Adversarial learning of causal graphs. *Journal of Machine Learning Research*, 23(219):1–62, 2022.

Nitish Shirish Keskar, Dheevatsa Mudigere, Jorge Nocedal, Mikhail Smelyanskiy, and Ping Tak Peter Tang. On large-batch training for deep learning: Generalization gap and sharp minima. *arXiv preprint arXiv:1609.04836*, 2016.

Diederik P Kingma. Auto-encoding variational bayes. *arXiv preprint arXiv:1312.6114*, 2013.

Diederik P Kingma. Adam: A method for stochastic optimization. *arXiv preprint arXiv:1412.6980*, 2014.

Thomas N Kipf and Max Welling. Semi-supervised classification with graph convolutional networks. *arXiv preprint arXiv:1609.02907*, 2016.

Rex B Kline. *Principles and practice of structural equation modeling*. Guilford publications, 2023.

A Kosaraju, VS Pendela, and O Hai. *Cardiogenic shock*. StatPearls Publishing, 2025. URL https://www.ncbi.nlm.nih.gov/books/NBK482255/.

Sébastien Lachapelle, Philippe Brouillard, Tristan Deleu, and Simon Lacoste-Julien. Gradient-based neural dag learning. *arXiv preprint arXiv:1906.02226*, 2019.

Hao-Chih Lee, Matteo Danieletto, Riccardo Miotto, Sarah T Cherng, and Joel T Dudley. Scaling structural learning with no-bears to infer causal transcriptome networks. In *Pacific Symposium on Biocomputing 2020*, pages 391–402. World Scientific, 2019.

Po-Ling Loh and Peter Bühlmann. High-dimensional learning of linear causal networks via inverse covariance estimation. *The Journal of Machine Learning Research*, 15(1):3065–3105, 2014.

Calvin Luo. Understanding diffusion models: A unified perspective. *arXiv preprint arXiv:2208.11970*, 2022.

Achille Nazaret, Justin Hong, Elham Azizi, and David Blei. Stable differentiable causal discovery. *arXiv preprint arXiv:2311.10263*, 2023.

Yurii Nesterov and Vladimir Spokoiny. Random gradient-free minimization of convex functions. *Foundations of Computational Mathematics*, 17(2):527–566, 2017.

Ignavier Ng, Shengyu Zhu, Zhitang Chen, and Zhuangyan Fang. A graph autoencoder approach to causal structure learning. *arXiv preprint arXiv:1911.07420*, 2019.

Ignavier Ng, AmirEmad Ghassami, and Kun Zhang. On the role of sparsity and dag constraints for learning linear dags. *Advances in Neural Information Processing Systems*, 33:17943–17954, 2020.

Ignavier Ng, Shengyu Zhu, Zhuangyan Fang, Haoyang Li, Zhitang Chen, and Jun Wang. Masked gradient-based causal structure learning. In *Proceedings of the 2022 SIAM International Conference on Data Mining (SDM)*, pages 424–432. SIAM, 2022.

Ignavier Ng, Biwei Huang, and Kun Zhang. Structure learning with continuous optimization: A sober look and beyond. In *Causal Learning and Reasoning*, pages 71–105. PMLR, 2024.

Juan Miguel Ogarrio, Peter Spirtes, and Joe Ramsey. A hybrid causal search algorithm for latent variable models. In *Conference on probabilistic graphical models*, pages 368–379. PMLR, 2016.

Samhita Pal, Dhrubajyoti Ghosh, and Shu Yang. Penalized FCI for causal structure learning in a sparse DAG for biomarker discovery in Parkinson's disease. *arXiv preprint arXiv:2507.00173*, 2025.

Roxana Pamfil, Nisara Sriwattanaworachai, Shaan Desai, Philip Pilgerstorfer, Konstantinos Georgatzis, Paul Beaumont, and Bryon Aragam. Dynotears: Structure learning from time-series data. In *International Conference on Artificial Intelligence and Statistics*, pages 1595–1605. Pmlr, 2020.

Judea Pearl. *Causality: Models, Reasoning and Inference*. Cambridge University Press, USA, 2nd edition, 2009. ISBN 052189560X.

Geraldine L Pellecchia. Postural sway increases with attentional demands of concurrent cognitive task. *Gait & posture*, 18(1):29–34, 2003.

Jonas Peters and Peter Bühlmann. Identifiability of gaussian structural equation models with equal error variances. *Biometrika*, 101(1):219–228, 2014.

Jonas Peters, Joris M Mooij, Dominik Janzing, and Bernhard Schölkopf. Causal discovery with continuous additive noise models. *The Journal of Machine Learning Research*, 15(1):2009–2053, 2014.

Jean-Baptiste Pingault, Paul F O'reilly, Tabea Schoeler, George B Ploubidis, Frühling Rijsdijk, and Frank Dudbridge. Using genetic data to strengthen causal inference in observational research. *Nature Reviews Genetics*, 19(9):566–580, 2018.

M Purvey and G Allen. Managing acute pulmonary oedema. *Aust Prescr.*, 40(2):59–63, Apr 2017. doi: doi:10.18773/austprescr.2017.012.

Alexander Reisach, Christof Seiler, and Sebastian Weichwald. Beware of the simulated dag! causal discovery benchmarks may be easy to game. *Advances in Neural Information Processing Systems*, 34:27772–27784, 2021.

Adèle H Ribeiro and Dominik Heider. dcFCI: Robust causal discovery under latent confounding, unfaithfulness, and mixed data. *arXiv preprint arXiv:2505.06542*, 2025.

Robin Rombach, Andreas Blattmann, Dominik Lorenz, Patrick Esser, and Björn Ommer. High-resolution image synthesis with latent diffusion models. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 10684–10695, 2022.

Pedro Sanchez, Xiao Liu, Alison Q O'Neil, and Sotirios A Tsaftaris. Diffusion models for causal discovery via topological ordering. *arXiv preprint arXiv:2210.06201*, 2022.

Chao Shang, Jie Chen, and Jinbo Bi. Discrete graph structure learning for forecasting multiple time series. *arXiv preprint arXiv:2101.06861*, 2021.

Xinwei Shen, Furui Liu, Hanze Dong, Qing Lian, Zhitang Chen, and Tong Zhang. Weakly supervised disentangled generative causal representation learning. *Journal of Machine Learning Research*, 23(241):1–55, 2022.

Shohei Shimizu, Patrik O Hoyer, Aapo Hyvärinen, Antti Kerminen, and Michael Jordan. A linear non-gaussian acyclic model for causal discovery. *Journal of Machine Learning Research*, 7(10), 2006.

Shohei Shimizu, Aapo Hyvarinen, Yutaka Kano, and Patrik O Hoyer. Discovery of non-gaussian linear causal models using ica. *arXiv preprint arXiv:1207.1413*, 2012.

K Shinozaki, T Honda, K Yamaji, E Nishijima, I Ichi, and D. Yamane. Impaired ApoB secretion triggers enhanced secretion of apoe to maintain triglyceride homeostasis in hepatoma cells. *J Lipid Res. Epub*, April 2025. doi: doi:10.1016/j.jlr.2025.100795.

Bill Shipley. *Cause and correlation in biology: A user's guide to path analysis, structural equations and causal inference with R*. Cambridge university press, 2016.

Hantao Shu, Jingtian Zhou, Qiuyu Lian, Han Li, Dan Zhao, Jianyang Zeng, and Jianzhu Ma. Modeling gene regulatory networks using neural network architectures. *Nature Computational Science*, 1(7):491–501, 2021.

Donna K Slonim. From patterns to pathways: gene expression data analysis comes of age. *Nature Genetics*, 32(4):502–508, 2002.

Peter Spirtes and Clark Glymour. An algorithm for fast recovery of sparse causal graphs. *Social science computer review*, 9(1):62–72, 1991.

Peter Spirtes and Kun Zhang. Causal discovery and inference: concepts and recent methodological advances. In *Applied informatics*, volume 3, pages 1–28. Springer, 2016.

Peter Spirtes, Clark N Glymour, and Richard Scheines. *Causation, prediction, and search*. MIT press, 2000.

Xiangyu Sun, Oliver Schulte, Guiliang Liu, and Pascal Poupart. Nts-notears: Learning nonparametric dbns with prior knowledge. *arXiv preprint arXiv:2109.04286*, 2021.

Tom N Tombaugh. Trail making test a and b: normative data stratified by age and education. *Archives of clinical neuropsychology*, 19(2):203–214, 2004.

Tom N Tombaugh and Nancy J McIntyre. The mini-mental state examination: a comprehensive review. *Journal of the American Geriatrics Society*, 40(9):922–935, 1992.

Caroline Uhler, Garvesh Raskutti, Peter Bühlmann, and Bin Yu. Geometry of the faithfulness assumption in causal inference. *The Annals of Statistics*, pages 436–463, 2013.

Sara van de Geer and Peter Bühlmann. ℓ_0-penalized maximum likelihood for sparse directed acyclic graphs. *The Annals of Statistics*, 41(2):536 – 567, 2013. doi: 10.1214/13-AOS1085. URL https://doi.org/10.1214/13-AOS1085.

Matthew J Vowels, Necati Cihan Camgoz, and Richard Bowden. D'ya like dags? a survey on structure learning and causal discovery. *ACM Computing Surveys*, 55(4):1–36, 2022.

Dennis Wei, Tian Gao, and Yue Yu. Dags with no fears: A closer look at continuous optimization for learning bayesian networks. *Advances in Neural Information Processing Systems*, 33:3895–3906, 2020.

Mengyue Yang, Furui Liu, Zhitang Chen, Xinwei Shen, Jianye Hao, and Jun Wang. Causalvae: Disentangled representation learning via neural structural causal models. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 9593–9602, 2021.

Yue Yu, Jie Chen, Tian Gao, and Mo Yu. DAG-GNN: Dag structure learning with graph neural networks. In *International conference on machine learning*, pages 7154–7163. PMLR, 2019.

Yue Yu, Tian Gao, Naiyu Yin, and Qiang Ji. Dags with no curl: An efficient dag structure learning approach. In *International Conference on Machine Learning*, pages 12156–12166. Pmlr, 2021.

Manzil Zaheer, Satwik Kottur, Siamak Ravanbakhsh, Barnabas Poczos, Russ R Salakhutdinov, and Alexander J Smola. Deep sets. *Advances in neural information processing systems*, 30, 2017.

Xun Zheng, Bryon Aragam, Pradeep K Ravikumar, and Eric P Xing. DAGS with NO TEARS: Continuous optimization for structure learning. *Advances in neural information processing systems*, 31, 2018.

Xun Zheng, Chen Dan, Bryon Aragam, Pradeep Ravikumar, and Eric Xing. Learning sparse nonparametric dags. In *International Conference on Artificial Intelligence and Statistics*, pages 3414–3425. Pmlr, 2020.

Junhong Zhou, Daniel Habtemariam, Ikechukwu Iloputaife, Lewis A Lipsitz, and Brad Manor. The complexity of standing postural sway associates with future falls in community-dwelling older adults: the mobilize boston study. *Scientific reports*, 7(1):2924, 2017.

Hao Zhu. lightgraph, 2025. URL https://github.com/haozhu233/lightgraph.

Hao Zhu and Donna K Slonim. From noise to knowledge: probabilistic diffusion-based neural inference of gene regulatory networks. *J Comput Biol*, 31(11):1087–1103, 2024.

Hao Zhu and Donna K Slonim. Improved gene regulatory network inference from single cell data with dropout augmentation. *PLOS Computational Biology*, 21(10):e1013603, 2025.

Rong Zhu, Andreas Pfadler, Ziniu Wu, Yuxing Han, Xiaoke Yang, Feng Ye, Zhenping Qian, Jingren Zhou, and Bin Cui. Efficient and scalable structure learning for bayesian networks: Algorithms and applications. In *2021 IEEE 37th International Conference on Data Engineering (ICDE)*, pages 2613–2624. IEEE, 2021.

Appendix A. Appendix

A.1. Explanation of the k-hop acyclicity constraint

In this section, we explain how to derive the proposed k-hop acyclicity constraint in Equation 11 from the NOTEARS DAG constraint in Equation 3.

The NOTEARS DAG constraint is in the form of $h_{\text{NOTEARS}}(\boldsymbol{W}) = \text{tr}(e^{\boldsymbol{W}\circ\boldsymbol{W}}) - d$, where $\circ$ is the Hadamard product, $e^{\boldsymbol{W}}$ is the matrix exponential of $\boldsymbol{W}$, and tr() is the trace of a matrix. In this case, matrix exponential is the sum of a weighted power series as shown below.

$$e^{\boldsymbol{W}} = \sum_{j=0}^{\infty} \frac{1}{j!}\boldsymbol{W}^j. \tag{16}$$

The trace of the summed matrix and the sum of all the traces are equivalent. At the same time, since $\boldsymbol{W}^0$ is simply the identity matrix, whose trace equals to the value of d, we can rewrite the NOTEARS DAG function in the following form:

$$h_{\text{NOTEARS}}(\boldsymbol{W}) = \sum_{j=0}^{\infty} \frac{1}{j!}\text{tr}((\boldsymbol{W}\circ\boldsymbol{W})^j) - d = \sum_{j=1}^{\infty} \frac{1}{j!}\text{tr}((\boldsymbol{W}\circ\boldsymbol{W})^j). \tag{17}$$

To account for cycles within k hops, we compute:

$$h_{\text{k-hop}}(\boldsymbol{W}, k) = \sum_{j=1}^{k+1} \frac{1}{j!}\text{tr}((\boldsymbol{W}\circ\boldsymbol{W})^j) \tag{18}$$

As mentioned in the main text, in the case when values in the weighted adjacency matrix are tiny, it might be helpful to multiply the values in the adjacency matrix by a constant multiplier γ and then remove it after the trace calculation. Then the equation can be expressed in the following form:

$$h_{\text{k-hop}}(\boldsymbol{W}, k, \gamma) = \sum_{j=1}^{k+1} \frac{1}{j!\gamma^{2j}}\text{tr}((\gamma\boldsymbol{W}\circ\gamma\boldsymbol{W})^j) \tag{19}$$

If we keep a running product for $j!$, γ^{2j}, and $(\gamma\boldsymbol{W}\circ\gamma\boldsymbol{W})^j$, we can keep the complexity within $\mathcal{O}(d^2)$.

A.2. Special Case in DDCD Smooth

In DDCD Smooth, all the inputs are transformed into the range of -1 to 1 through MLP and the Tanh activation function. We expected to learn a normalized adjacency matrix $\hat{\boldsymbol{W}}$, where the expected value is $\frac{1}{d}$. This normalized adjacency matrix would be conceptually similar to the normalized adjacency matrix in graph convolution (Kipf and Welling, 2016). Under these assumptions, Theorem 2 shows that we can predict the added noise $\boldsymbol{Z}$ directly.

Theorem 2 *With a normalized adjacency matrix, we can directly infer the added noise* $\boldsymbol{Z}$.

Proof Starting from Equation 8, let's pick a random sample $\boldsymbol{x}$ and perturb that with a Gaussian noise vector $\boldsymbol{z} \in \mathcal{N}(0, 1)$ to build $\boldsymbol{x}_t$.

$$\boldsymbol{W}^T\boldsymbol{x}_t = \sqrt{\overline{\alpha_t}}\boldsymbol{W}^T\boldsymbol{x}_0 + \sqrt{1-\overline{\alpha_t}}\boldsymbol{W}^T\boldsymbol{z}, \tag{20}$$

We can in fact write each element in $\boldsymbol{W}^T\boldsymbol{z}$ as a form of weighted Gaussian mixtures. Since all values in $\boldsymbol{z}$ are standard Gaussian noise with a mean of 0 and variance of 1, the weighted sum of such a mixture will also be centered at 0. Given that the expected value of edge weight in $\boldsymbol{W}$ is $\frac{1}{d-1}$ and there are $d-1$ entries, the expected value for the entire variance is $\sum_{i=0}^{d} \frac{1-\overline{\alpha_t}}{d^2} = \frac{1-\overline{\alpha_t}}{d}$. When d is large, this variance of $\boldsymbol{W}^T\boldsymbol{x}_t$ will be much smaller than the variance term in $\boldsymbol{x}_t$, which is $1-\overline{\alpha_t}$. When d is really large and the diffusion coeffient is small, we can therefore use $\boldsymbol{W}^T\boldsymbol{x}_t$ to approximate the unperturbed $\boldsymbol{x}_0$. This argument is very similar to the Central Limit Theorem, but on noise. As a result, $\boldsymbol{W}^T\boldsymbol{x}_t - \boldsymbol{x}_t$ will give us a close estimate of the added noise z.

■

A.3. Performance on Large Graphs

Here, we include two sample weight estimates on larger graphs with 1,000 nodes. The main structures of the graphs are recovered (Figure 5).

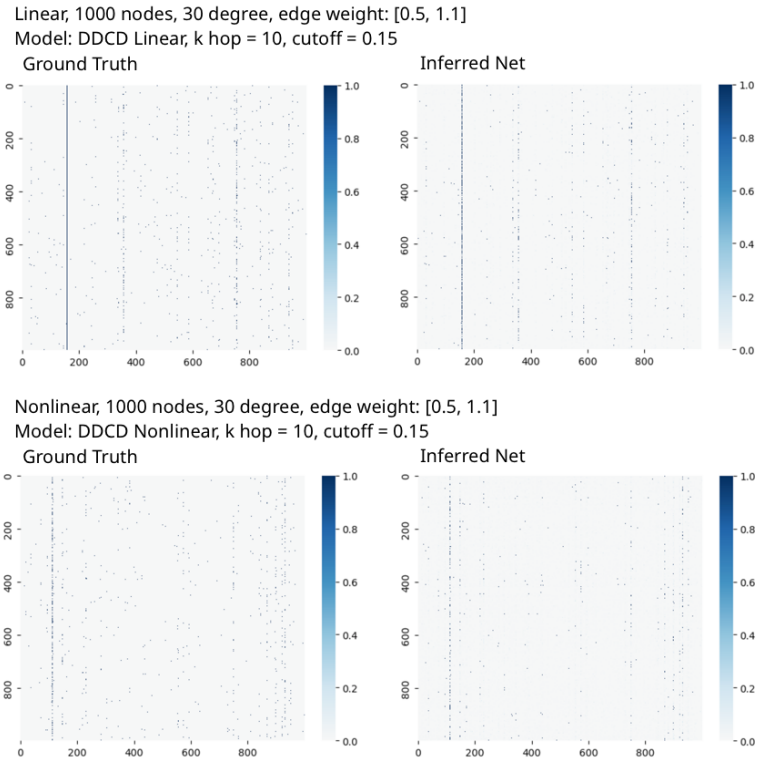

Figure 5: Example Weight Estimates on Graphs with 1,000 nodes. Number of samples in both cases is 2,000.

A.4. Metrics

Since the inferred graphs are directed graphs, we use the same evaluation methods used in NOTEARS. Since we are not generating non-directed edge predictions at all, here is a simplified description of the metrics that we are using.

1. True Positive Rate (TPR) is defined as

$$\text{TPR} = \frac{\text{True Positive}}{\text{Condition Positive}} \tag{21}$$

   True positive is the number of cases when the predicted association exists in the condition in the correct direction. Condition positive is the total number of true edges in the ground truth graph.

2. False Discovery Rate (FDR) is defined as

$$\text{FDR} = \frac{\text{False Positive} + \text{Reverse}}{\text{Prediction Positive}} \tag{22}$$

   False positive is the number of cases when the predicted association does not exist in the condition. Reverse is the number of cases when the predicted association exists in the condition but in the opposite direction. Prediction Positive is the total number of positive predictions in the inferred graph.

3. False Positive Rate (FPR) is defined as

$$\text{FPR} = \frac{\text{False Positive} + \text{Reverse}}{\text{Condition Negative}} \tag{23}$$

   Condition negative is the total number of edges that do not exist.

4. Structural Hamming Distance (SHD) is a measure used to quantify the difference between two directed acyclic graphs (DAGs). It counts the number of operations, including adding an edge, removing an edge, and reversing an edge, required to transform one graph into another. Here, the SHD is the sum of reversed positive predictions, false positive predictions regardless of direction, and false negative predictions regardless of direction.

Note on Identifiability and SHD: We report SHD against the ground-truth DAG. For linear models with Gaussian noise, the causal structure is generally identifiable only up to the Markov Equivalence Class (MEC), meaning some edge reversals may not be true errors. Following standard practice in continuous DAG learning (Zheng et al., 2018; Ng et al., 2020; Bello et al., 2022), we report SHD against the ground-truth DAG to enable direct comparison with prior work. For nonlinear models with additive noise, identifiability of the full DAG follows from ANM assumptions (Peters et al., 2014), even with Gaussian noise, because the nonlinearity breaks the symmetry that causes equivalence classes in the linear case.

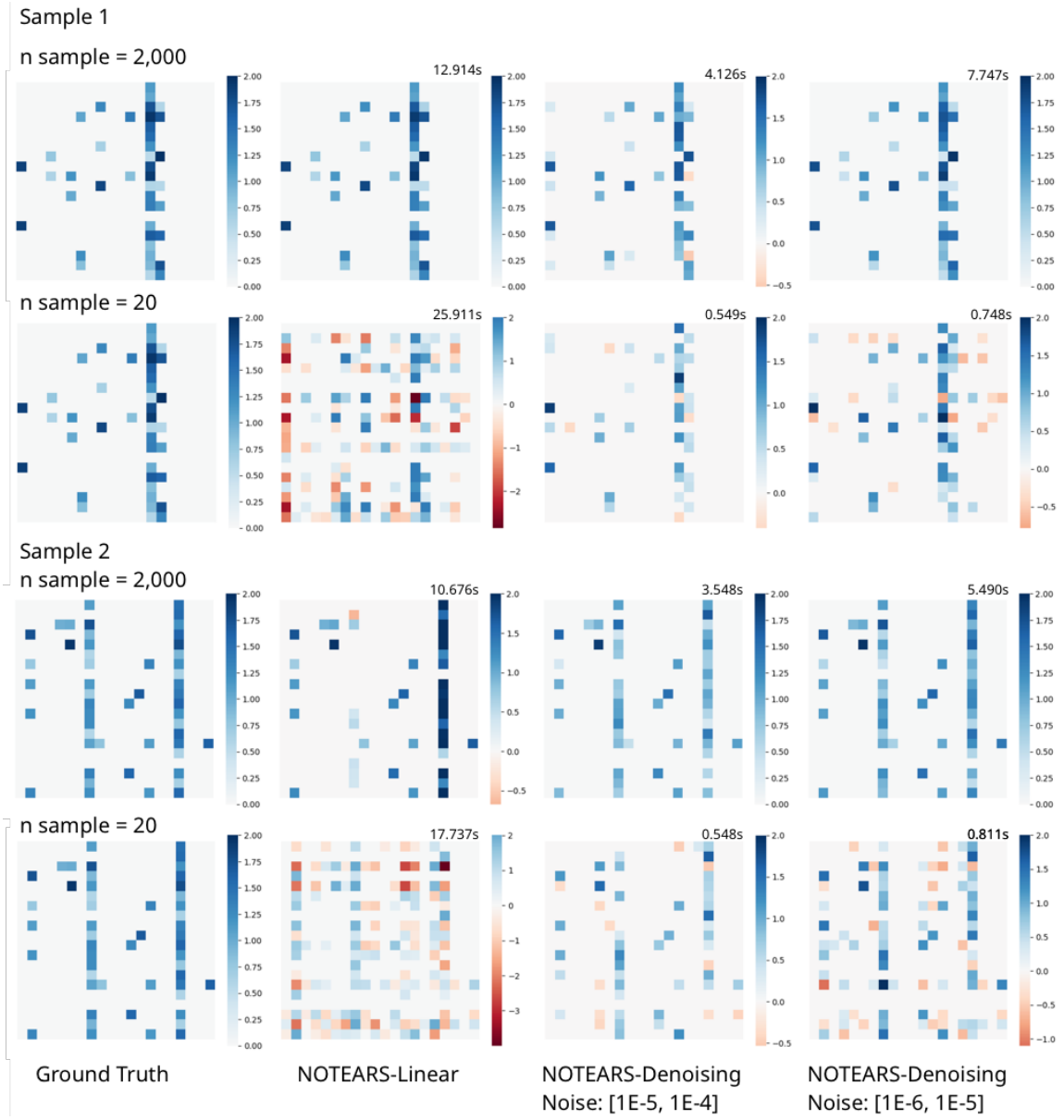

Figure 6: Comparison of some results from NOTEARS-Linear and NOTEARS-Denoising. Execution times are displayed on the top-right corner of each figure.

A.5. Inferred examples from NOTEARS-Denoising

In Figure 6 we show some additional comparisons between results from NOTEARS-Linear and NOTEARS-Denoising. Overall, as reported in the main paper, when the number of samples is limited (2nd and 4th rows), the results from NOTEARS-Linear (2nd column) may include a lot of noise. This could be resolved by using the denoising diffusion objective (3rd and 4th columns). When the number of samples is sufficient (1st and 3rd rows), in most cases, NOTEARS-Linear will

generate very good inference but in some cases, such as Sample 2 in rows 3 and 4, it may end up in a local minimum. On the other hand, using the denoising objective does come with a cost. When the added noise is not small enough, it may introduce some small noisy values in the inferred matrix. This could be resolved by adding smaller noises instead (column 4) but smaller noises will also increase the runtime.

A.6. Linear multiplier outperforms Augmented Lagrangian in optimization

Another experiment design we tested involves replacing the dual-ascending augmented Lagrangian optimization with a simple linear multiplier using training epoch steps. Here We justify this simplification for the following reasons, along with empirical evidence: **1. Smoother Optimization and Better Convergence:** As shown in Figure 2, the denoising objective effectively smooths the loss gradient. This stability allows the optimizer to converge effectively using a monotonic linear schedule. **2. Bounded Edge Weights:** In models like DDCD-Smooth, feature normalization restricts edge weights to a small range (typically $|w_{ij}| < 1$). The values of the quadratic penalty term of the augmented Lagrangian $\rho\|h(\boldsymbol{W})\|^2$ become extremely small and cannot be used as a penalty.

In our empirical experiment, we found that while the Augmented Lagrangian approach maintains a very low FDR and FPR, its TPR degrades significantly as the graph node count increases. In contrast, the linear multiplier maintains high discovery power with a TPR above 0.95 and maintains a substantially lower Structural Hamming Distance (SHD) across all scales. Furthermore, the linear approach offers a more efficient optimization path, consistently reducing total computation time. This experiment clearly shows that at least for our diffusion setup, linear multiplier is a better choice compared with the Augmented Lagrangian.

Table 1: Optimization Strategy Comparison: Augmented Lagrangian vs. Linear Multiplier

Optimizer	Nodes	SHD	TPR	FPR	FDR	Time(s)
Augmented Lagrangian	20	32.5±3.0	0.45±0.03	**0.01±0.01**	**0.04±0.02**	15.8±8.2
	50	135.8±11.7	0.23±0.04	**0.00±0.00**	**0.10±0.08**	13.8±0.8
	100	328.5±18.1	0.13±0.01	**0.00±0.00**	**0.15±0.03**	27.4±8.1
Linear Multiplier	20	**10.7±5.9**	**0.96±0.04**	0.07±0.04	0.14±0.07	**8.9±1.7**
	50	**46.2±12.5**	**0.97±0.01**	0.04±0.01	0.21±0.05	**12.8±2.1**
	100	**139.3±16.2**	**0.96±0.01**	0.03±0.00	0.26±0.03	**15.9±3.3**

A.7. Detailed analysis of the k-hop schedule Training Dynamics

In this section, we provide detailed analysis of how different k-hop schedules affect training dynamics and the acyclicity constraint. This experiment was done on a 100-node SF graph with linear synthetic data, and we used DDCD-Linear to illustrate the point. As shown in the figure, applying different k-hop schedules can significantly affect the speed of the algorithm. For example, in the case where we applied global DAG check all the time (100% k=d), the entire training process for a 2,000 node graph would take 53.7 minutes while if we used the default schedule (40% k=3; 50% k=10; 10% k=d), we could reduce the time to 5.7 minutes, nearly 90% reduction. At the same time, the movements of acyclicity constraint h during training are near identical in these two cases.

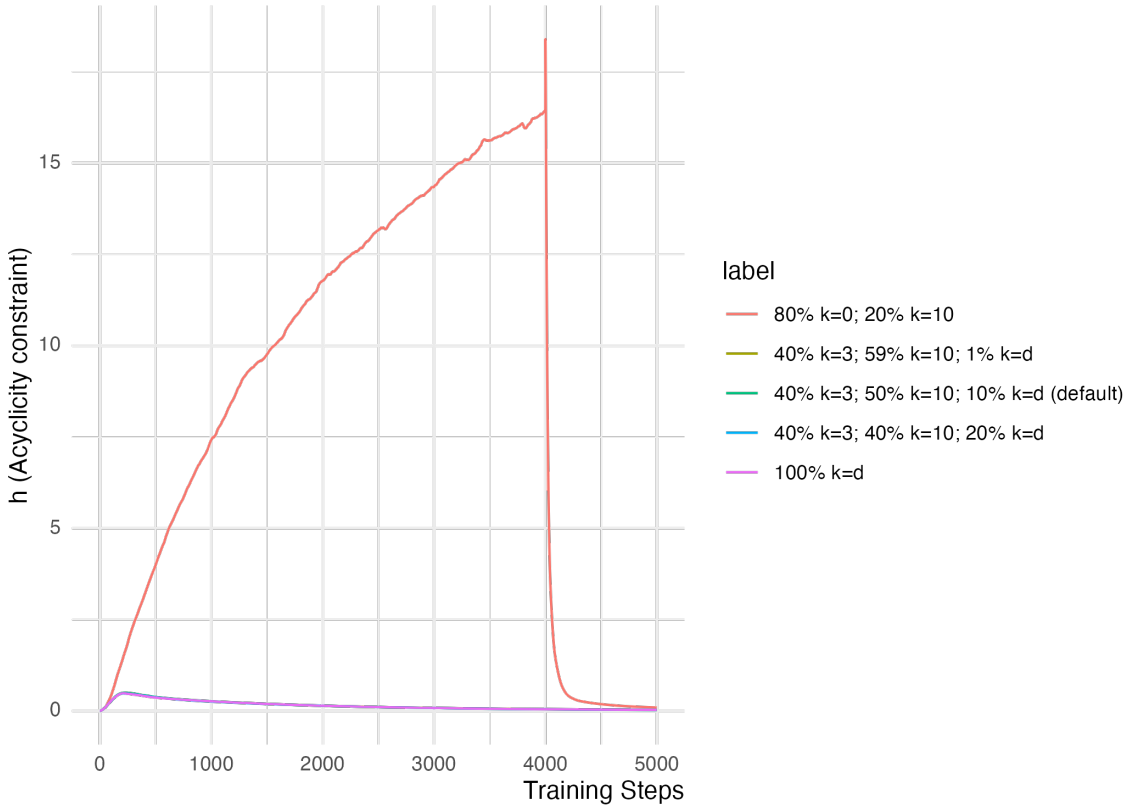

Figure 7: Change of DAG constraint h under different k-hop schedules

We also conducted an experiment where we applied no DAG constraint at the first 80% of training steps and $k = 10$ for the last 20%, the acyclicity score h skyrocketed at the beginning but dropped quickly after we put on the constraint (after step 4,000) (Figure 7). In the cases where we did apply some acyclicity constraint at the beginning, the computed h still rapidly increased from 0.0 to 0.5 in all cases, because at this stage, the SEM reconstruction loss is much larger than the acyclicity constrain and is the main factor that drives the direction of the optimizer. After this initial stage, the training starts to balance the power of SEM reconstruction and DAG constraint. Based on these observations, we concluded that 1. the sharp increase of h could be prevented by applying a local acyclicity constraint; 2. spending resources on computing the exact value of global DAG constraint is not necessary for early stage training; 3. using a k-hop acyclicity schedule is as effective as full DAG calculation while consuming a fraction of time.

A.8. Hyperparameter Search

To study the sensitivity of DDCD to diffusion noise parameters, we conducted a comprehensive grid search. All experiments were performed on synthetic SF graphs with 50 nodes, degree 10, and 2,000 samples using linear SEMs with Gaussian noise.

Here we explored:

- Noise Schedule: Linear, Cosine, Power

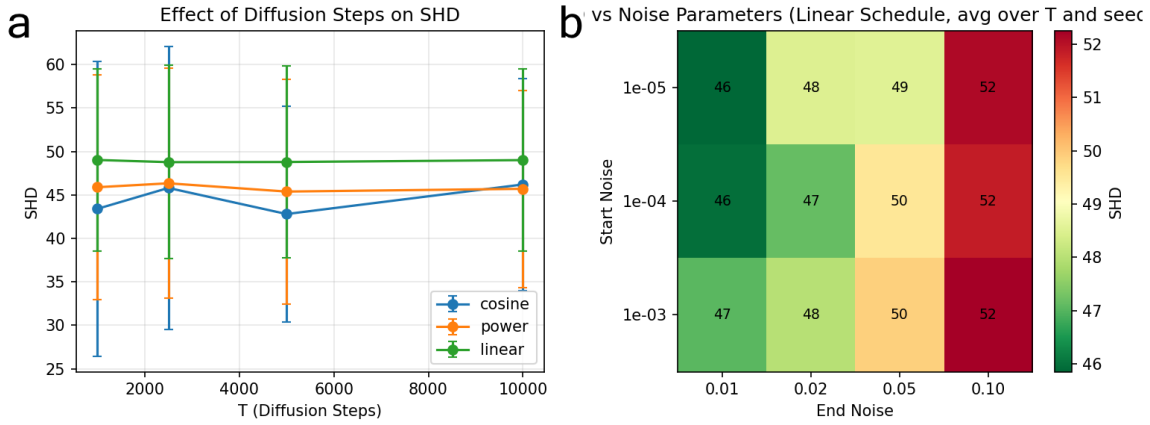

Figure 8: Effect of Diffusion timestep and noise schedule

- Diffusion timesteps T: 1000, 2500, 5000, 10000
- Start noise: 1e-5, 1e-4, 1e-3
- End noise: 0.01, 0.02, 0.05, 0.1

Based on our experiment, all three schedules achieved comparable SHD performances as shown in Figure 8a. The cosine schedule seems to yield overall better performance, but it's not statistically significant. We also observe similar performances across Diffusion Steps T, suggesting that DDCD is robust across noise schedules and diffusion steps.

We also did hyperparameter search in terms of the start noise and end noise and the results are included in Figure 8b. The best performing combination happens when Start Noise is 1e-4 or 1e-5 and End Noise is 1e-2. Note that, in our other experiments, we followed standard Diffusion literature and use $T = 5,000$, Start Noise = 1e-4 and End Noise = 0.02. This is not the optim setup but yield similar performance.

A.9. Directionality Analysis

We investigated whether the gradient smoothing provided by the denoising objective affects directionality. To isolate this effect, we decomposed the SHD metric into false positives, false negatives, and edge reversals. In both low-sample and high-sample conditions, DDCD persistently reduces directional errors compared to NOTEARS (Figure 9). These findings indicate that the denoising objective is, in fact, helpful for the correctness of directionality.

A.10. Full results for main benchmark experiments (Figure 3)

Table 2: Main benchmark results (SHD, TPR, FDR, FPR across all methods)

N	Data	Graph	Method	SHD	TPR	FDR	FPR	Time(s)
2000	linear	SF-20	DAG-GNN	0.9±1.0	0.99±0.01	0.02±0.02	0.01±0.01	548.8
2000	linear	SF-20	DDCD Linear	**0.3±0.9**	**1.00±0.01**	0.01±0.02	**0.00±0.01**	13.1
2000	linear	SF-20	DDCD Non-linear	9.0±18.3	0.87±0.25	0.07±0.19	0.02±0.04	4.1

N	Data	Graph	Method	SHD	TPR	FDR	FPR	Time(s)
2000	linear	SF-20	DDCD Smooth	50.7±7.0	0.17±0.16	0.73±0.18	0.17±0.06	4.0
2000	linear	SF-20	GAE	139.3±3.2	0.60±0.03	0.86±0.00	1.43±0.10	890.5
2000	linear	SF-20	GOLEM	6.0±0.0	0.92±0.00	0.08±0.00	0.03±0.00	156.4
2000	linear	SF-20	Notears	7.0±0.0	0.88±0.00	0.06±0.00	0.02±0.00	7.8
2000	linear	SF-20	Notears Non-linear	42.6±1.1	0.48±0.01	0.51±0.01	0.19±0.01	1247.4
2000	linear	SF-20	PC	27.0±0.0	0.60±0.00	0.21±0.00	0.06±0.00	4.6
2000	linear	SF-20	DAGMA Linear	2.0±0.0	0.96±0.00	**0.00±0.00**	**0.00±0.00**	**1.0**
2000	linear	SF-20	DAGMA Non-linear	40.1±2.1	0.37±0.04	0.37±0.04	0.08±0.01	581.9
2000	linear	SF-20	GES	74.0±0.0	0.73±0.00	0.65±0.00	0.51±0.00	19.9
2000	linear	SF-100	DAG-GNN	45.4±4.4	0.91±0.01	0.05±0.01	**0.00±0.00**	1607.6
2000	linear	SF-100	DDCD Linear	**25.1±4.3**	**0.98±0.00**	0.07±0.01	**0.00±0.00**	19.0
2000	linear	SF-100	DDCD Non-linear	101.7±122.9	0.77±0.24	0.13±0.21	0.01±0.01	**5.3**
2000	linear	SF-100	DDCD Smooth	349.8±15.0	0.05±0.03	0.87±0.06	0.02±0.00	5.0
2000	linear	SF-100	GAE	527.4±535.4	0.45±0.30	0.49±0.29	0.09±0.14	7076.6
2000	linear	SF-100	GOLEM	45.0±0.0	0.87±0.00	**0.01±0.00**	**0.00±0.00**	714.2
2000	linear	SF-100	Notears	71.0±0.0	0.81±0.00	0.04±0.00	**0.00±0.00**	360.7
2000	linear	SF-100	Notears Non-linear	193.3±38.8	0.58±0.11	0.31±0.06	0.02±0.00	45896.1
2000	linear	SF-100	PC	294.0±0.0	0.28±0.00	0.52±0.00	0.02±0.00	52.8
2000	linear	SF-100	DAGMA Linear	46.0±0.0	0.86±0.00	**0.01±0.00**	**0.00±0.00**	22.6
2000	linear	SF-100	DAGMA Non-linear	213.6±5.0	0.34±0.01	0.08±0.02	**0.00±0.00**	2385.2
2000	linear	SF-100	GES	98.0±0.0	0.94±0.00	0.24±0.00	0.02±0.00	14992.6
2000	linear	ER-20	DAG-GNN	66.4±14.9	0.58±0.21	0.31±0.05	0.29±0.10	2360.7
2000	linear	ER-20	DDCD Linear	51.3±1.1	**0.91±0.01**	0.32±0.00	0.48±0.01	14.0
2000	linear	ER-20	DDCD Non-linear	99.2±6.7	0.35±0.10	0.51±0.06	0.38±0.07	4.0
2000	linear	ER-20	DDCD Smooth	134.6±14.6	0.42±0.19	0.75±0.08	1.30±0.48	**3.7**
2000	linear	ER-20	GAE	79.5±11.0	0.39±0.14	0.34±0.09	0.22±0.10	898.7
2000	linear	ER-20	GOLEM	41.0±0.0	0.71±0.00	0.18±0.00	0.18±0.00	161.6
2000	linear	ER-20	Notears	83.0±0.0	0.42±0.00	0.39±0.00	0.30±0.00	98.7
2000	linear	ER-20	Notears Non-linear	110.6±11.1	0.12±0.10	0.66±0.25	0.41±0.36	5204.3
2000	linear	ER-20	PC	104.0±0.0	0.09±0.00	0.80±0.00	0.39±0.00	1.7
2000	linear	ER-20	DAGMA Linear	**24.0±0.0**	0.86±0.00	0.13±0.00	0.14±0.00	4.7
2000	linear	ER-20	DAGMA Non-linear	100.0±0.0	0.00±0.00	**0.00±0.00**	**0.00±0.00**	39.7
2000	linear	ER-20	GES	120.0±0.0	0.57±0.00	0.66±0.00	1.21±0.00	187.8
2000	linear	ER-100	DAG-GNN	2012.4±1391.4	0.26±0.33	0.81±0.07	0.45±0.70	5350.8
2000	linear	ER-100	DDCD Linear	1023.5±2.3	0.02±0.00	0.75±0.01	0.02±0.00	19.0
2000	linear	ER-100	DDCD Non-linear	1058.0±46.4	0.03±0.02	0.71±0.25	0.03±0.02	**5.0**
2000	linear	ER-100	DDCD Smooth	2603.0±281.4	0.30±0.06	0.90±0.01	0.68±0.14	**5.0**
2000	linear	ER-100	GAE	1189.7±131.1	0.07±0.04	0.79±0.03	0.07±0.04	5537.6
2000	linear	ER-100	GOLEM	1109.0±0.0	0.03±0.00	0.87±0.00	0.04±0.00	707.5
2000	linear	ER-100	Notears	983.0±0.0	0.01±0.00	0.57±0.00	**0.00±0.00**	72.6
2000	linear	ER-100	Notears Non-linear	1027.2±37.2	0.00±0.00	0.92±0.06	0.02±0.01	2356.0
2000	linear	ER-100	PC	1094.0±0.0	0.06±0.00	0.80±0.00	0.06±0.00	112.6
2000	linear	ER-100	DAGMA Linear	1742.6±0.5	**0.34±0.00**	0.77±0.00	0.28±0.00	163.9
2000	linear	ER-100	DAGMA Non-linear	**980.0±0.0**	0.00±0.00	**0.00±0.00**	**0.00±0.00**	191.7
2000	linear	ER-100	GES	-	-	-	-	¿12hr
2000	Non-linear 1	SF-20	DAG-GNN	42.3±4.2	0.25±0.07	0.19±0.12	0.02±0.02	507.1
2000	Non-linear 1	SF-20	DDCD Linear	22.5±1.1	0.59±0.02	0.06±0.00	0.01±0.00	14.0
2000	Non-linear 1	SF-20	DDCD Non-linear	**8.4±1.0**	**0.84±0.02**	**0.00±0.00**	**0.00±0.00**	4.0
2000	Non-linear 1	SF-20	DDCD Smooth	42.2±12.9	0.33±0.23	0.60±0.25	0.18±0.09	4.0
2000	Non-linear 1	SF-20	GAE	90.4±44.9	0.78±0.35	0.68±0.23	0.76±0.48	987.5
2000	Non-linear 1	SF-20	GOLEM	36.0±0.0	0.46±0.00	0.27±0.00	0.07±0.00	169.5
2000	Non-linear 1	SF-20	Notears	21.0±0.0	0.62±0.00	0.06±0.00	0.01±0.00	2.0
2000	Non-linear 1	SF-20	Notears Non-linear	12.1±0.3	0.79±0.01	0.03±0.01	0.01±0.00	93.9

N	Data	Graph	Method	SHD	TPR	FDR	FPR	Time(s)
2000	Non-linear 1	SF-20	PC	27.0±0.0	0.58±0.00	0.25±0.00	0.07±0.00	2.1
2000	Non-linear 1	SF-20	DAGMA Linear	23.0±0.0	0.58±0.00	0.06±0.00	0.01±0.00	**0.0**
2000	Non-linear 1	SF-20	DAGMA Non-linear	23.2±2.1	0.55±0.04	0.02±0.02	0.01±0.00	360.5
2000	Non-linear 1	SF-20	GES	48.0±0.0	0.54±0.00	0.59±0.00	0.29±0.00	11.1
2000	Non-linear 1	SF-100	DAG-GNN	206.8±33.3	0.38±0.11	0.08±0.03	**0.00±0.00**	1784.4
2000	Non-linear 1	SF-100	DDCD Linear	157.8±3.1	0.54±0.01	0.08±0.00	**0.00±0.00**	18.5
2000	Non-linear 1	SF-100	DDCD Non-linear	153.0±12.1	0.56±0.04	0.09±0.02	**0.00±0.00**	**5.0**
2000	Non-linear 1	SF-100	DDCD Smooth	281.1±55.1	0.38±0.27	0.58±0.21	0.03±0.01	**5.0**
2000	Non-linear 1	SF-100	GAE	275.8±194.1	0.44±0.33	0.31±0.28	0.02±0.03	3872.4
2000	Non-linear 1	SF-100	GOLEM	203.0±0.0	0.40±0.00	0.09±0.00	**0.00±0.00**	680.6
2000	Non-linear 1	SF-100	Notears	154.0±0.0	0.55±0.00	0.07±0.00	**0.00±0.00**	42.3
2000	Non-linear 1	SF-100	Notears Non-linear	**79.5±5.1**	**0.83±0.01**	0.10±0.01	0.01±0.00	11903.1
2000	Non-linear 1	SF-100	DAGMA Linear	162.0±0.0	0.52±0.00	0.06±0.00	**0.00±0.00**	11.0
2000	Non-linear 1	SF-100	DAGMA Non-linear	176.7±3.4	0.45±0.01	**0.02±0.00**	**0.00±0.00**	1263.5
2000	Non-linear 1	SF-100	GES	316.0±0.0	0.50±0.00	0.58±0.00	0.05±0.00	21959.0
2000	Non-linear 1	ER-20	DAG-GNN	86.7±4.3	0.12±0.05	0.22±0.11	0.04±0.02	885.5
2000	Non-linear 1	ER-20	DDCD Linear	81.9±1.2	0.18±0.01	0.23±0.05	0.05±0.02	14.0
2000	Non-linear 1	ER-20	DDCD Non-linear	**23.1±4.3**	0.76±0.05	**0.00±0.01**	**0.00±0.01**	**4.0**
2000	Non-linear 1	ER-20	DDCD Smooth	111.6±6.7	0.30±0.06	0.71±0.05	0.75±0.15	**4.0**
2000	Non-linear 1	ER-20	GAE	34.5±32.2	**0.86±0.27**	0.24±0.21	0.25±0.15	1019.8
2000	Non-linear 1	ER-20	GOLEM	93.0±0.0	0.12±0.00	0.48±0.00	0.12±0.00	292.4
2000	Non-linear 1	ER-20	Notears	79.0±0.0	0.25±0.00	0.31±0.00	0.12±0.00	3.0
2000	Non-linear 1	ER-20	Notears Non-linear	74.7±2.8	0.45±0.03	0.41±0.03	0.31±0.02	415.7
2000	Non-linear 1	ER-20	PC	105.0±0.0	0.19±0.00	0.71±0.00	0.48±0.00	6.1
2000	Non-linear 1	ER-20	DAGMA Linear	79.0±0.0	0.20±0.00	0.17±0.00	0.04±0.00	**0.0**
2000	Non-linear 1	ER-20	DAGMA Non-linear	76.7±1.5	0.28±0.01	0.29±0.02	0.12±0.01	1890.7
2000	Non-linear 1	ER-20	GES	120.0±0.0	0.46±0.00	0.68±0.00	1.00±0.00	42.5
2000	Non-linear 1	ER-100	DAG-GNN	993.7±20.4	0.03±0.02	0.05±0.06	**0.00±0.00**	3035.9
2000	Non-linear 1	ER-100	DDCD Linear	963.7±5.5	0.11±0.00	0.34±0.02	0.01±0.00	19.3
2000	Non-linear 1	ER-100	DDCD Non-linear	**177.6±28.7**	**0.83±0.03**	**0.00±0.00**	**0.00±0.00**	**5.0**
2000	Non-linear 1	ER-100	DDCD Smooth	1986.3±463.1	0.20±0.11	0.87±0.03	0.36±0.17	**5.0**
2000	Non-linear 1	ER-100	GAE	676.8±473.4	0.68±0.30	0.23±0.28	0.09±0.12	8688.7
2000	Non-linear 1	ER-100	GOLEM	1039.0±0.0	0.01±0.00	0.71±0.00	0.01±0.00	1090.3
2000	Non-linear 1	ER-100	Notears	1000.0±0.0	0.16±0.00	0.48±0.00	0.04±0.00	711.7
2000	Non-linear 1	ER-100	Notears Non-linear	1348.1±26.7	0.42±0.01	0.65±0.01	0.20±0.01	37706.2
2000	Non-linear 1	ER-100	PC	1177.0±0.0	0.14±0.00	0.73±0.00	0.10±0.00	2864.1
2000	Non-linear 1	ER-100	DAGMA Linear	978.0±0.0	0.10±0.00	0.37±0.00	0.01±0.00	13.0
2000	Non-linear 1	ER-100	DAGMA Non-linear	962.3±3.2	0.07±0.00	0.27±0.02	0.01±0.00	1508.5
2000	Non-linear 1	ER-100	GES	-	-	-	-	¿12hr
20	linear	SF-20	DAG-GNN	63.2±8.3	**0.84±0.03**	0.56±0.04	0.41±0.06	15.4
20	linear	SF-20	DDCD Linear	68.3±1.6	**0.84±0.01**	0.58±0.01	0.44±0.01	14.0
20	linear	SF-20	DDCD Non-linear	**31.8±5.4**	0.70±0.15	**0.31±0.05**	0.12±0.04	4.1
20	linear	SF-20	DDCD Smooth	98.4±10.9	0.17±0.09	0.89±0.05	0.56±0.11	3.9
20	linear	SF-20	GAE	67.5±18.1	0.15±0.12	0.73±0.29	0.20±0.14	454.2
20	linear	SF-20	GOLEM	77.0±0.0	0.56±0.00	0.66±0.00	0.41±0.00	106.8
20	linear	SF-20	Notears	57.0±0.0	0.50±0.00	0.56±0.00	0.24±0.00	25.2
20	linear	SF-20	Notears Non-linear	108.6±3.7	0.49±0.03	0.78±0.01	0.65±0.02	168.4
20	linear	SF-20	PC	57.0±0.0	0.10±0.00	0.75±0.00	**0.11±0.00**	**0.0**
20	linear	SF-20	DAGMA Linear	80.0±0.0	0.58±0.00	0.66±0.00	0.43±0.00	1.0
20	linear	SF-20	DAGMA Non-linear	105.3±9.1	0.34±0.04	0.81±0.02	0.54±0.08	189.6
20	linear	SF-20	GES	66.2±5.4	0.32±0.08	0.70±0.05	0.28±0.00	5.3
20	linear	SF-100	DAG-GNN	729.3±175.7	**0.51±0.06**	0.77±0.05	0.12±0.04	186.0

N	Data	Graph	Method	SHD	TPR	FDR	FPR	Time(s)
20	linear	SF-100	DDCD Linear	519.6±74.2	0.47±0.17	0.63±0.22	0.08±0.03	19.0
20	linear	SF-100	DDCD Non-linear	**290.0±30.3**	0.14±0.16	**0.13±0.15**	**0.00±0.00**	5.1
20	linear	SF-100	DDCD Smooth	876.0±208.1	0.12±0.06	0.94±0.02	0.14±0.05	5.0
20	linear	SF-100	GAE	294.1±27.8	0.24±0.10	0.35±0.14	0.01±0.01	1018.2
20	linear	SF-100	GOLEM	436.0±0.0	0.20±0.00	0.74±0.00	0.04±0.00	580.8
20	linear	SF-100	Notears	450.0±0.0	0.27±0.00	0.72±0.00	0.05±0.00	2718.5
20	linear	SF-100	Notears Non-linear	606.0±25.7	0.13±0.02	0.90±0.02	0.08±0.01	17922.8
20	linear	SF-100	PC	362.0±0.0	0.06±0.00	0.80±0.00	0.02±0.00	**3.0**
20	linear	SF-100	DAGMA Linear	579.1±8.6	0.25±0.01	0.81±0.01	0.08±0.00	72.7
20	linear	SF-100	DAGMA Non-linear	540.3±99.6	0.13±0.01	0.84±0.10	0.06±0.02	461.1
20	linear	SF-100	GES	-	-	-	-	¿12hr
20	linear	ER-20	DAG-GNN	**64.1±5.2**	0.85±0.05	**0.37±0.02**	0.56±0.04	31.1
20	linear	ER-20	DDCD Linear	65.0±1.5	**0.90±0.01**	0.39±0.01	0.63±0.02	14.0
20	linear	ER-20	DDCD Non-linear	92.8±4.0	0.15±0.06	**0.33±0.07**	0.09±0.05	4.0
20	linear	ER-20	DDCD Smooth	134.3±11.9	0.25±0.16	0.81±0.08	1.07±0.34	4.0
20	linear	ER-20	GAE	90.3±5.5	0.29±0.16	0.40±0.04	0.22±0.12	353.7
20	linear	ER-20	GOLEM	81.0±0.0	0.65±0.00	0.42±0.00	0.52±0.00	112.7
20	linear	ER-20	Notears	85.7±0.9	0.37±0.01	0.41±0.01	0.29±0.00	46.3
20	linear	ER-20	Notears Non-linear	106.5±6.3	0.49±0.10	0.58±0.05	0.75±0.12	697.1
20	linear	ER-20	PC	101.0±0.0	0.09±0.00	0.65±0.00	0.19±0.00	**0.0**
20	linear	ER-20	DAGMA Linear	71.8±0.6	0.69±0.00	0.40±0.00	0.51±0.01	5.8
20	linear	ER-20	DAGMA Non-linear	100.0±0.0	0.00±0.00	**0.00±0.00**	**0.00±0.00**	10.7
20	linear	ER-20	GES	110.0±9.7	0.20±0.08	0.70±0.11	0.51±0.00	14.6
20	linear	ER-100	DAG-GNN	1245.9±247.7	0.09±0.07	0.79±0.03	0.09±0.08	183.6
20	linear	ER-100	DDCD Linear	1014.0±1.2	0.02±0.00	0.74±0.00	**0.01±0.00**	19.8
20	linear	ER-100	DDCD Non-linear	1056.7±47.7	0.03±0.02	**0.61±0.32**	0.03±0.02	**5.0**
20	linear	ER-100	DDCD Smooth	2266.4±325.5	0.21±0.06	0.91±0.01	0.49±0.13	**5.0**
20	linear	ER-100	GAE	1121.8±84.3	0.05±0.03	0.72±0.26	0.05±0.03	2100.3
20	linear	ER-100	GOLEM	1117.0±0.0	0.02±0.00	0.89±0.00	0.04±0.00	590.7
20	linear	ER-100	Notears	**994.0±0.0**	0.01±0.00	0.68±0.00	**0.01±0.00**	45.2
20	linear	ER-100	Notears Non-linear	1001.8±10.6	0.00±0.00	0.97±0.07	**0.01±0.00**	224.1
20	linear	ER-100	PC	1013.0±0.0	0.03±0.00	0.72±0.00	0.02±0.00	6.9
20	linear	ER-100	DAGMA Linear	1767.3±0.9	**0.33±0.00**	0.77±0.00	0.28±0.00	90.7
20	linear	ER-100	DAGMA Non-linear	980.0±0.0	0.00±0.00	**0.00±0.00**	**0.00±0.00**	40.7
20	linear	ER-100	GES	-	-	-	-	¿12hr
20	Non-linear 1	SF-20	DAG-GNN	76.4±14.1	0.48±0.09	0.65±0.07	0.36±0.12	14.8
20	Non-linear 1	SF-20	DDCD Linear	74.1±3.9	**0.69±0.02**	0.62±0.02	0.43±0.03	14.0
20	Non-linear 1	SF-20	DDCD Non-linear	53.4±5.8	0.33±0.05	**0.51±0.08**	0.14±0.05	4.0
20	Non-linear 1	SF-20	DDCD Smooth	84.3±10.7	0.26±0.14	0.82±0.10	0.44±0.09	4.0
20	Non-linear 1	SF-20	GAE	88.9±35.0	0.26±0.21	0.82±0.15	0.46±0.46	406.0
20	Non-linear 1	SF-20	GOLEM	85.0±0.0	0.60±0.00	0.68±0.00	0.47±0.00	109.2
20	Non-linear 1	SF-20	Notears	53.0±0.0	0.46±0.00	0.56±0.00	0.22±0.00	9.2
20	Non-linear 1	SF-20	Notears Non-linear	102.4±4.1	0.64±0.05	0.74±0.02	0.68±0.03	26.3
20	Non-linear 1	SF-20	PC	**45.0±0.0**	0.19±0.00	**0.29±0.00**	**0.03±0.00**	**0.0**
20	Non-linear 1	SF-20	DAGMA Linear	72.0±0.0	0.56±0.00	0.65±0.00	0.38±0.00	1.0
20	Non-linear 1	SF-20	DAGMA Non-linear	111.4±4.4	0.53±0.06	0.78±0.02	0.72±0.02	239.1
20	Non-linear 1	SF-20	GES	66.2±8.7	0.19±0.04	0.75±0.06	0.22±0.00	5.9
20	Non-linear 1	SF-100	DAG-GNN	798.6±149.3	**0.36±0.08**	0.84±0.03	0.13±0.04	185.6
20	Non-linear 1	SF-100	DDCD Linear	686.4±10.8	0.06±0.00	0.96±0.00	0.08±0.00	19.0
20	Non-linear 1	SF-100	DDCD Non-linear	502.6±96.6	0.07±0.05	0.90±0.08	0.04±0.02	5.0
20	Non-linear 1	SF-100	DDCD Smooth	1254.0±135.8	0.19±0.14	0.94±0.04	0.22±0.02	5.0
20	Non-linear 1	SF-100	GAE	421.1±154.6	0.20±0.08	**0.66±0.17**	0.04±0.04	1123.3
20	Non-linear 1	SF-100	GOLEM	400.0±0.0	0.01±0.00	0.96±0.00	**0.02±0.00**	553.0

N	Data	Graph	Method	SHD	TPR	FDR	FPR	Time(s)
20	Non-linear 1	SF-100	Notears	542.0±0.0	0.13±0.00	0.87±0.00	0.06±0.00	570.4
20	Non-linear 1	SF-100	Notears Non-linear	764.6±14.1	0.13±0.01	0.93±0.00	0.11±0.00	23053.8
20	Non-linear 1	SF-100	PC	**382.0±0.0**	0.04±0.00	0.88±0.00	**0.02±0.00**	**1.0**
20	Non-linear 1	SF-100	DAGMA Linear	633.2±2.3	0.13±0.01	0.90±0.00	0.08±0.00	42.6
20	Non-linear 1	SF-100	DAGMA Non-linear	385.0±7.7	0.07±0.01	0.81±0.02	**0.02±0.00**	503.2
20	Non-linear 1	SF-100	GES	-	-	-	-	¿12hr
20	Non-linear 1	ER-20	DAG-GNN	99.6±4.2	0.39±0.06	0.57±0.03	0.54±0.09	34.6
20	Non-linear 1	ER-20	DDCD Linear	96.8±3.8	**0.53±0.03**	0.55±0.02	0.67±0.03	13.8
20	Non-linear 1	ER-20	DDCD Non-linear	**86.0±8.7**	0.24±0.08	**0.42±0.10**	0.19±0.08	4.0
20	Non-linear 1	ER-20	DDCD Smooth	111.1±6.1	0.27±0.08	0.72±0.06	0.71±0.16	4.0
20	Non-linear 1	ER-20	GAE	105.0±6.6	0.23±0.12	0.67±0.06	0.45±0.18	713.1
20	Non-linear 1	ER-20	GOLEM	112.0±0.0	0.36±0.00	0.63±0.00	0.64±0.00	272.4
20	Non-linear 1	ER-20	Notears	93.0±0.0	0.35±0.00	0.56±0.00	0.46±0.00	3.6
20	Non-linear 1	ER-20	Notears Non-linear	110.4±8.5	0.48±0.06	0.65±0.05	0.89±0.07	94.9
20	Non-linear 1	ER-20	PC	98.0±0.0	0.05±0.00	0.71±0.00	**0.13±0.00**	**0.0**
20	Non-linear 1	ER-20	DAGMA Linear	105.0±0.0	0.36±0.00	0.61±0.00	0.59±0.00	1.0
20	Non-linear 1	ER-20	DAGMA Non-linear	111.9±10.8	0.43±0.09	0.67±0.06	0.88±0.07	183.2
20	Non-linear 1	ER-20	GES	106.8±4.1	0.12±0.04	0.74±0.05	0.36±0.00	6.9
20	Non-linear 1	ER-100	DAG-GNN	1451.2±160.1	**0.15±0.05**	0.80±0.01	0.15±0.05	350.3
20	Non-linear 1	ER-100	DDCD Linear	1273.9±8.3	0.11±0.00	0.77±0.01	0.09±0.00	19.2
20	Non-linear 1	ER-100	DDCD Non-linear	1151.2±38.2	0.06±0.01	**0.76±0.03**	0.05±0.01	5.2
20	Non-linear 1	ER-100	DDCD Smooth	2021.6±141.4	0.12±0.04	0.92±0.03	0.34±0.04	5.0
20	Non-linear 1	ER-100	GAE	1231.2±67.6	0.07±0.03	0.81±0.05	0.08±0.02	1934.0
20	Non-linear 1	ER-100	GOLEM	1273.0±0.0	0.07±0.00	0.82±0.00	0.08±0.00	750.7
20	Non-linear 1	ER-100	Notears	1336.6±5.8	0.10±0.00	0.82±0.00	0.11±0.00	1425.9
20	Non-linear 1	ER-100	Notears Non-linear	1352.8±44.8	0.09±0.01	0.83±0.02	0.12±0.01	10844.1
20	Non-linear 1	ER-100	PC	**1072.0±0.0**	0.01±0.00	0.86±0.00	**0.02±0.00**	**1.0**
20	Non-linear 1	ER-100	DAGMA Linear	1363.7±2.4	0.09±0.00	0.84±0.00	0.11±0.00	55.7
20	Non-linear 1	ER-100	DAGMA Non-linear	1103.2±53.7	0.03±0.01	0.78±0.04	0.03±0.02	459.5
20	Non-linear 1	ER-100	GES	-	-	-	-	¿12hr

A.11. Non-linear Diversity Experiments

To address concerns about testing DDCD in a limited range of nonlinear functions, we conducted a comprehensive evaluation across seven distinct non-linear functions as shown in Figure 10. These functions consist of smooth ($\sin, \cos, \tanh$), non-smooth (ReLU), bounded (Sigmoid, $\tanh$), and unbounded (polynomial). Similar to the main experiment, we generated both SF and ER graphs at different sizes. Here we compared the performance among the Non-linear and Linear models from DDCD, DAGMA, and GES.

Based on our experiment, DDCD Non-Linear consistently achieves the lowest error rates (SHD) across most non-linear functions. Specifically, in the periodic (Types 1 & 2) and activation-function-based relationships (Types 5, 6, & 7), DDCD Non-Linear significantly outperforming linear baselines and often surpassing the state-of-the-art DAGMA Non-Linear (orange). However, quadratic and sigmoid relationships (Types 3 & 4) prove difficult for DDCD Non-linear but they seem to be challenging for all the other methods as well.

In terms of runtime, DDCD Non-Linear (blue) present sharp advantage compared with other methods, especially on large graphs. While DAGMA Non-Linear and GES exhibit exponential scaling and consume thousands of seconds on graph with 100 nodes, DDCD Non-Linear can finish within 20 seconds (on CPU). Note that we don't have data points from some GES runs because they run out of time (3 hrs).

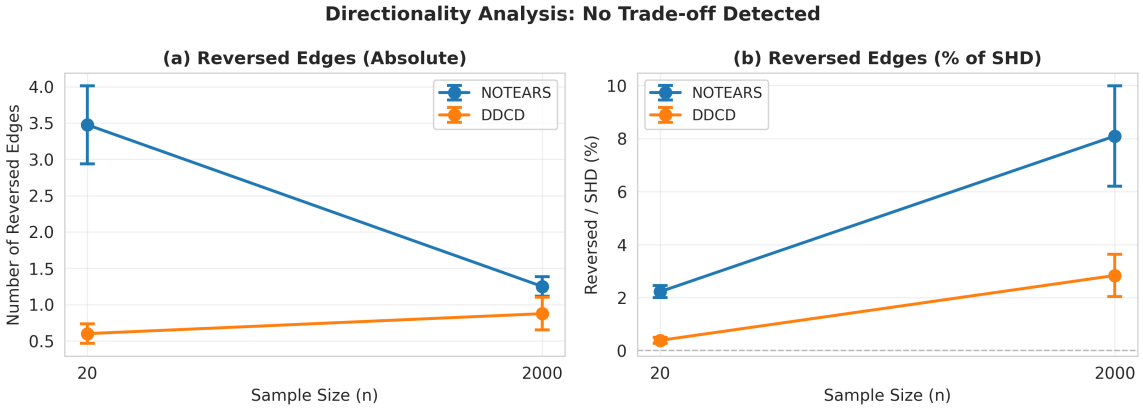

Figure 9: Reversed Edge Analysis

A.12. Case Study 2: Physical Function in Aging Population

In the aging dataset, the average age was 70.6 $\pm$ 7.5 years old. Within the cohort, 53.8% of the participants were women; 60.6% had hypertension, and 30.4% had diabetes. The causal network was also inferred using DDCD smooth and a network was extracted with a threshold at 0.1. The network visualization was done using python package lightgraph (Zhu, 2025).

In Figure 12, we identified several correlated node clusters, which we circled and summarized in the legend. These clusters suggest that inferred edges capture similar functions; many specific causal relationships in the clusters make sense. For example, education (level) is inferred as a cause of performance scores on several cognitive tests, including the Trail Making Tests (TMT) (Bowie and Harvey, 2006; Tombaugh, 2004), Mini-Mental State Examination (MMSE) (Tombaugh and McIntyre, 1992), and the performance on a postural sway task (Zhou et al., 2017). While the latter does not initially appear to be related to education level, postural sway is known to increase with increased cognitive load (Pellecchia, 2003), suggesting an indirect causal relationship captured here. Current norms for postural sway tests do not reflect education levels (Goble and Baweja, 2018), but our findings suggest that investigating this possibility might be valuable.

In the diabetes cluster, APOB and cholesterol (CHOL) appear as root causes. APOB secretion affects production of APOE (Shinozaki et al., 2025), which binds to Triglycerides (TG), which contribute to Glucose Level (GLU), which is among the diagnostic criteria for diabetes. Between the clusters of hypertension and physical strength, we see nodes representing Albumin (ALB), Globulin (GLB) and Total Protein (TP). ALB and GLB are two main types of blood proteins, so the directions of these arrows are logical as well. As in the MI data set above, the directions of a few inferred edges are reversed. For example, physical activity was inferred as the cause of sex/gender.

A.13. Inferred weights of the MI network

Here we show a table containing the weights of an inferred network by DDCD Smooth for the MI experiment, sorted by edge weights, similar to that in Figure 4. Most of the edges here seem explicable, reflecting standard medical understanding of correlations if not causality. We do not claim that they represent novel findings, only that the existence of a directed "causal" edge in this graph is generally logical. Therefore, investigating any edges whose explanations are *not* readily apparent in similar data sets may lead to new findings in the future.

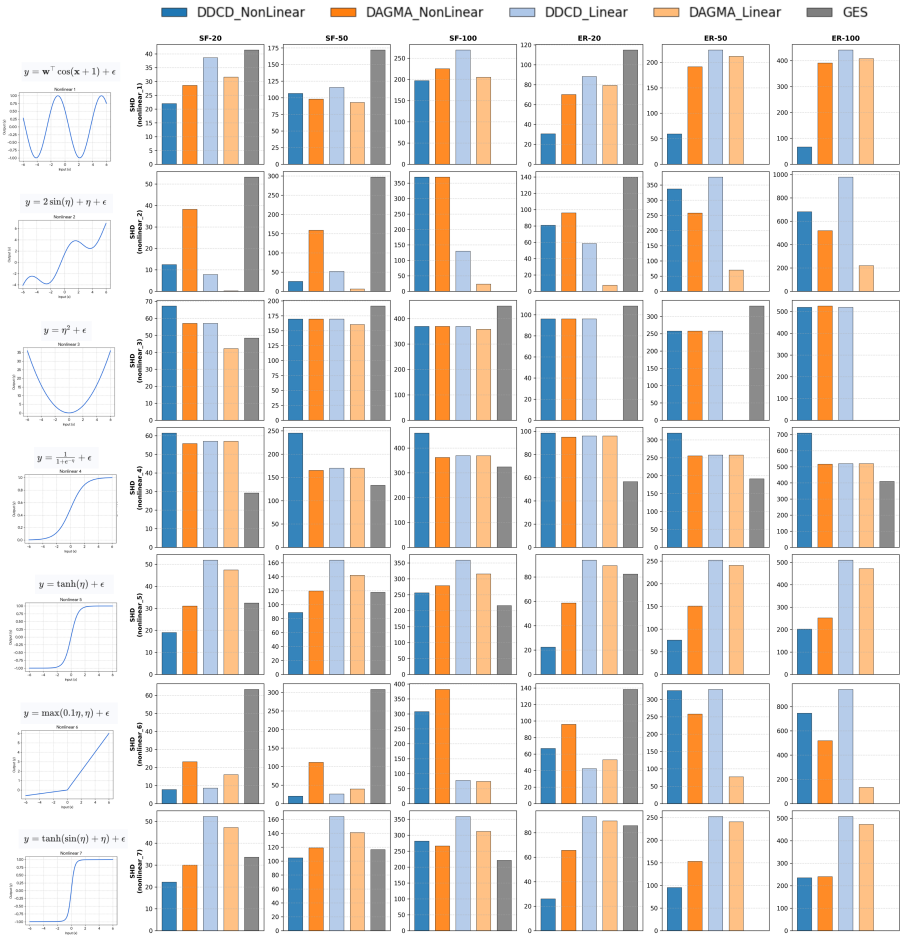

Figure 10: SHD Results from the Non-linear Diversity Experiments

Table 3: DDCD Smooth MI Edges

Source	Source Label	Target	Target Label	Weight
LID_KB	Use of lidocaine by the Emergency Cardiology Team	NA_KB	Use of opioid drugs by the Emergency Cardiology Team	0.66
R_AB_2_n	Relapse of the pain in the second day of the hospital period	NA_R_2_n	Use of opioid drugs in the ICU in the second day of the hospital period	0.58
R_AB_1_n	Relapse of the pain in the first hours of the hospital period	NA_R_1_n	Use of opioid drugs in the ICU in the first hours of the hospital period	0.52
R_AB_3_n	Relapse of the pain in the third day of the hospital period	NA_R_3_n	Use of opioid drugs in the ICU in the third day of the hospital period	0.50

Source	Source Label	Target	Target Label	Weight
NOT_NA_2_n	Use of NSAIDs in the ICU in the second day of the hospital period	R_AB_2_n	Relapse of the pain in the second day of the hospital period	0.49
RAZRIV	Myocardial rupture	LET_IS	Lethal outcome	0.48
NA_KB	Use of opioid drugs by the Emergency Cardiology Team	NOT_NA_KB	Use of NSAIDs by the Emergency Cardiology Team	0.47
NOT_NA_2_n	Use of NSAIDs in the ICU in the second day of the hospital period	NOT_NA_3_n	Use of NSAIDs in the ICU in the third day of the hospital period	0.47
NOT_NA_3_n	Use of NSAIDs in the ICU in the third day of the hospital period	R_AB_3_n	Relapse of the pain in the third day of the hospital period	0.46
LID_KB	Use of lidocaine by the Emergency Cardiology Team	NOT_NA_KB	Use of NSAIDs by the Emergency Cardiology Team	0.44
K_SH_POST	Cardiogenic shock at the time of admission to intensive care unit	LET_IS	Lethal outcome	0.39
ASP_S_n	Use of acetylsalicylic acid in the ICU	GEPAR_S_n	Use of a anticoagulants (heparin) in the ICU	0.38
FIBR_JELUD	Ventricular fibrillation	LID_S_n	Use of lidocaine in the ICU	0.32
n_r_ecg_p_03	Premature ventricular contractions on ECG at the time of admission to hospital	LID_S_n	Use of lidocaine in the ICU	0.32
ZSN	Chronic heart failure	ZSN_A	Presence of chronic Heart failure (HF) in the anamnesis	0.31
n_r_ecg_p_05	Paroxysms of atrial fibrillation on ECG at the time of admission to hospital	FIBR_PREDS	Atrial fibrillation	0.29
n_r_ecg_p_04	Frequent premature ventricular contractions on ECG at the time of admission to hospital	LID_S_n	Use of lidocaine in the ICU	0.27
R_AB_1_n	Relapse of the pain in the first hours of the hospital period	NOT_NA_1_n	Use of NSAIDs in the ICU in the first hours of the hospital period	0.26
R_AB_3_n	Relapse of the pain in the third day of the hospital period	R_AB_2_n	Relapse of the pain in the second day of the hospital period	0.25
NOT_NA_2_n	Use of NSAIDs in the ICU in the second day of the hospital period	R_AB_1_n	Relapse of the pain in the first hours of the hospital period	0.23
NITR_S	Use of liquid nitrates in the ICU	OTEK_LANC	Pulmonary edema	0.22

Source	Source Label	Target	Target Label	Weight
NOT_NA_2_n	Use of NSAIDs in the ICU in the second day of the hospital period	NOT_NA_1_n	Use of NSAIDs in the ICU in the first hours of the hospital period	0.21
LID_S_n	Use of lidocaine in the ICU	NA_KB	Use of opioid drugs by the Emergency Cardiology Team	0.19
R_AB_3_n	Relapse of the pain in the third day of the hospital period	R_AB_1_n	Relapse of the pain in the first hours of the hospital period	0.19
NOT_NA_3_n	Use of NSAIDs in the ICU in the third day of the hospital period	NOT_NA_1_n	Use of NSAIDs in the ICU in the first hours of the hospital period	0.19
zab_leg_01	Chronic bronchitis in the anamnesis	ZSN	Chronic heart failure	0.19
NA_R_2_n	Use of opioid drugs in the ICU in the second day of the hospital period	R_AB_1_n	Relapse of the pain in the first hours of the hospital period	0.19
REC_IM	Relapse of the myocardial infarction	R_AB_3_n	Relapse of the pain in the third day of the hospital period	0.18
n_p_ecg_p_12	Complete RBBB on ECG at the time of admission to hospital	LET_IS	Lethal outcome	0.18
FIBR_JELUD	Ventricular fibrillation	GIPO_K	Hypokalemia (¡ 4 mmol/L)	0.18
NOT_NA_3_n	Use of NSAIDs in the ICU in the third day of the hospital period	NA_R_3_n	Use of opioid drugs in the ICU in the third day of the hospital period	0.18
LET_IS	Lethal outcome	ASP_S_n	Use of acetylsalicylic acid in the ICU	-0.20
TRENT_S_n	Use of Trental in the ICU	ASP_S_n	Use of acetylsalicylic acid in the ICU	-0.29

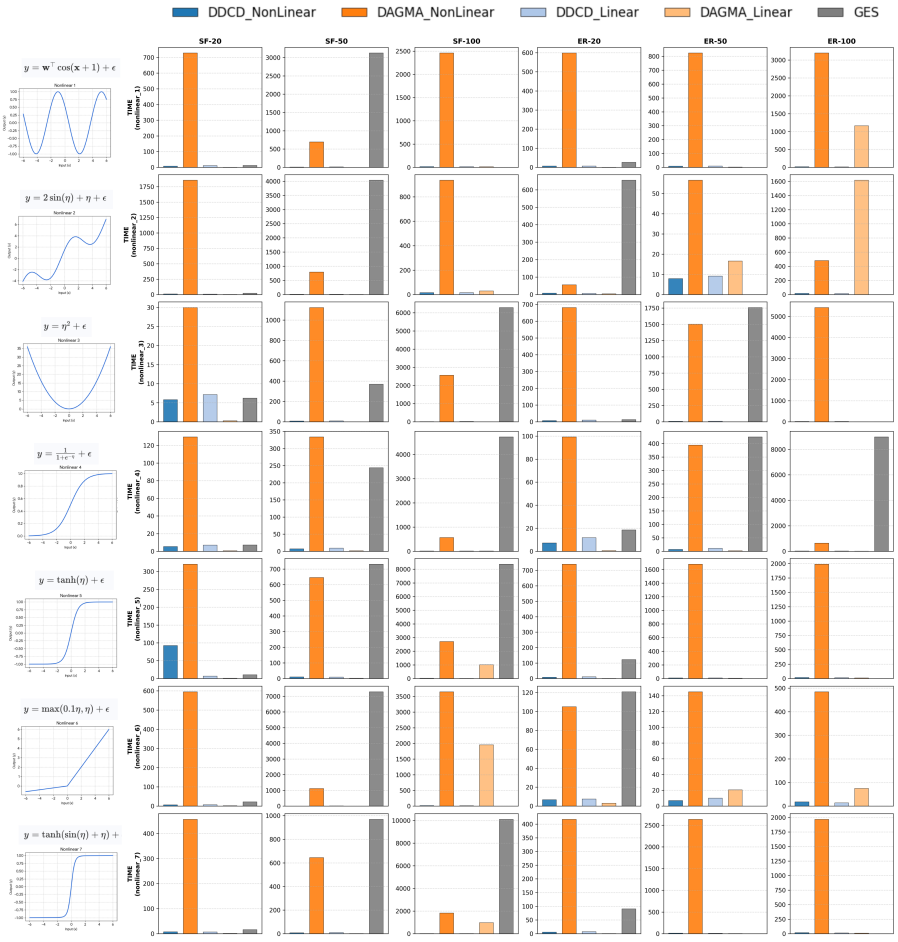

Figure 11: Time Costs from the Non-linear Diversity Experiments

A.14. Inferred MI network with NOTEARS

Next, we show an example of the inferred network by NOTEARS, using recommended parameters and focused on the same 2-hop neighborhood around the ”lethal outcome” node, on the exact same MI dataset. While some edges make sense, most of the edges seem to be clustered around a few hub targets and overall seem less “causal.” Patient age is one of the main target nodes, supposedly caused by other clinical variables. Compare the DDCD results in Figure 4 to the NOTEARS results on the same data shown in Figure 13 and whose edges are detailed in the table below.

This excessively hub-and-spoke structure, with fewer separate clusters of connected nodes, in our experience reflects higher numbers of spurious connections. The intuition behind this claim is that multiple causes of one outcome (and thus edges connected to hubs) will also, if real, be detected as causally related to one another. While we have removed short cycles from these models, there are enough related causal terms to form directed causal clusters, as highlighted in the ovals in Figure 4, rather than forming hubs with many unrelated spoke nodes, as we see with NOTEARS.

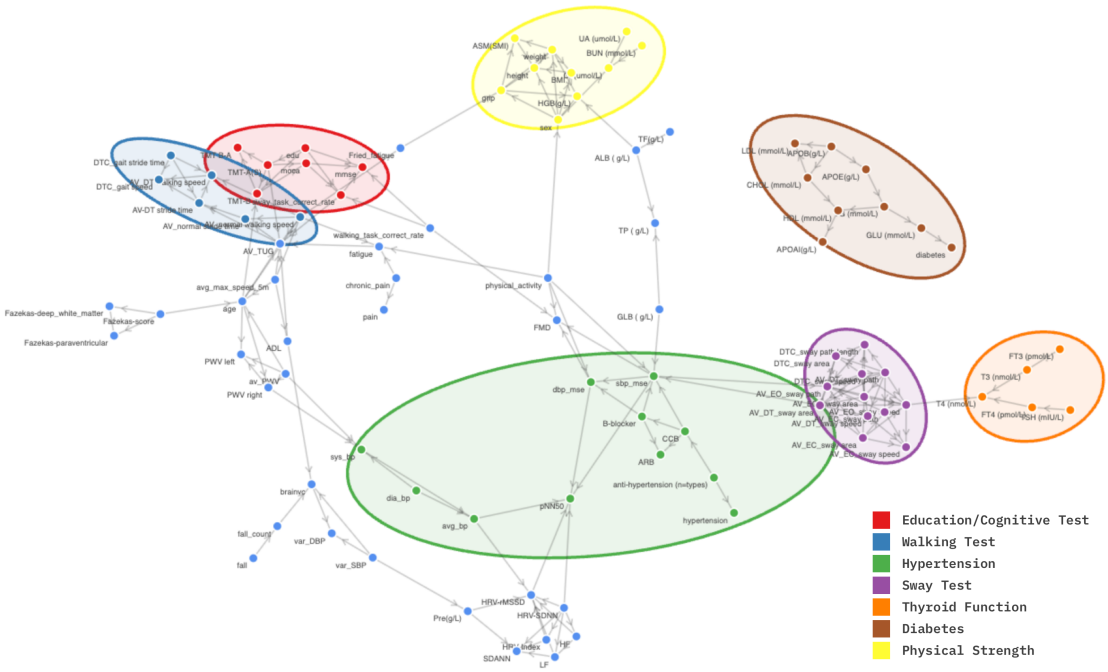

Figure 12: Inferred Causal Network in an Aging Study.

Table 4: NOTEARS MI Edges

Source	Source Label	Target	Target Label	Weight
LET_IS	Lethal outcome	AGE	Age of patient.	1
LET_IS	Lethal outcome	S_AD_ORIT	Systolic blood pressure according to intensive care unit	1
LET_IS	Lethal outcome	D_AD_ORIT	Diastolic blood pressure according to intensive care unit	1
FK_STENOK	Functional class (FC) of angina pectoris in the last year.	NA_BLOOD	Serum sodium content	1
MP_TP_POST	Paroxysms of atrial fibrillation at the time of admission to intensive care unit, (or at a pre-hospital stage)	NA_BLOOD	Serum sodium content	1
nr_11	Observing of arrhythmia in the anamnesis	NA_BLOOD	Serum sodium content	1
ZSN_A	Presence of chronic Heart failure (HF) in the anamnesis	NA_BLOOD	Serum sodium content	1
OTEK_LANC	Pulmonary edema	NA_BLOOD	Serum sodium content	1
K_BLOOD	Serum potassium content	NA_BLOOD	Serum sodium content	1

Source	Source Label	Target	Target Label	Weight
FIBR_PREDS	Atrial fibrillation	NA_BLOOD	Serum sodium content	1
TRENT_S_n	Use of Trental in the ICU	NA_BLOOD	Serum sodium content	1
ASP_S_n	Use of acetylsalicylic acid in the ICU	NA_BLOOD	Serum sodium content	1
NOT_NA_1_n	Use of NSAIDs in the ICU in the first hours of the hospital period	NA_BLOOD	Serum sodium content	1
GIPO_K	Hypokalemia (¡ 4 mmol/L)	NA_BLOOD	Serum sodium content	1
NA_R_3_n	Use of opioid drugs in the ICU in the third day of the hospital period	NA_BLOOD	Serum sodium content	1
NA_R_1_n	Use of opioid drugs in the ICU in the first hours of the hospital period	NA_BLOOD	Serum sodium content	1
NITR_S	Use of liquid nitrates in the ICU	NA_BLOOD	Serum sodium content	1
NA_KB	Use of opioid drugs by the Emergency Cardiology Team	NA_BLOOD	Serum sodium content	1
ritm_ecg_p_07	ECG rhythm at the time of admission to hospital: sinus with a heart rate above 90 (tachycardia)	NA_BLOOD	Serum sodium content	1
R_AB_1_n	Relapse of the pain in the first hours of the hospital period	NA_BLOOD	Serum sodium content	1
ALT_BLOOD	Serum AlAT content (ALT_BLOOD)	NA_BLOOD	Serum sodium content	1
GIPER_NA	Increase of sodium in serum (more than 150 mmol/L)	NA_BLOOD	Serum sodium content	1
SEX	Male	NA_BLOOD	Serum sodium content	1
R_AB_3_n	Relapse of the pain in the third day of the hospital period	ROE	ESR (Erythrocyte sedimentation rate)	1
R_AB_3_n	Relapse of the pain in the third day of the hospital period	S_AD_ORIT	Systolic blood pressure according to intensive care unit	1
R_AB_3_n	Relapse of the pain in the third day of the hospital period	D_AD_KBRIG	Diastolic blood pressure according to Emergency Cardiology Team	1
R_AB_3_n	Relapse of the pain in the third day of the hospital period	D_AD_ORIT	Diastolic blood pressure according to intensive care unit	1
IBS_NASL	Heredity on CHD	AGE	Age of patient.	1
zab_leg_02	Obstructive chronic bronchitis in the anamnesis	AGE	Age of patient.	1

Source	Source Label	Target	Target Label	Weight
zab_leg_01	Chronic bronchitis in the anamnesis	AGE	Age of patient.	1
GB	Presence of an essential hypertension	AGE	Age of patient.	1
fibr_ter_03	Fibrinolytic therapy by Celiasum 3m IU	AGE	Age of patient.	1
SEX	Male	AGE	Age of patient.	1
MP_TP_POST	Paroxysms of atrial fibrillation at the time of admission to intensive care unit, (or at a pre-hospital stage)	AGE	Age of patient.	1
lat_im	Presence of a lateral myocardial infarction (left ventricular)	AGE	Age of patient.	1
ritm_ecg_p_01	ECG rhythm at the time of admission to hospital: sinus (with a heart rate 60-90)	AGE	Age of patient.	1
n_r_ecg_p_01	Premature atrial contractions on ECG at the time of admission to hospital	AGE	Age of patient.	1
ALT_BLOOD	Serum AlAT content (ALT_BLOOD)	AGE	Age of patient.	1
NA_R_1_n	Use of opioid drugs in the ICU in the first hours of the hospital period	AGE	Age of patient.	1
B_BLOK_S_n	Use of beta-blockers in the ICU	AGE	Age of patient.	1
GEPAR_S_n	Use of a anticoagulants (heparin) in the ICU	AGE	Age of patient.	1
FIBR_PREDS	Atrial fibrillation	AGE	Age of patient.	1
OTEK_LANC	Pulmonary edema	AGE	Age of patient.	1
RAZRIV	Myocardial rupture	AGE	Age of patient.	1
ZSN	Chronic heart failure	AGE	Age of patient.	1
REC_IM	Relapse of the myocardial infarction	AGE	Age of patient.	1
LET_IS	Lethal outcome	AGE	Age of patient.	1
NITR_S	Use of liquid nitrates in the ICU	NA_BLOOD	Serum sodium content	1
GEPAR_S_n	Use of a anticoagulants (heparin) in the ICU	L_BLOOD	White blood cell count	1
IBS_POST	Coronary heart disease (CHD) in recent weeks, days before admission to hospital	L_BLOOD	White blood cell count	1
ritm_ecg_p_01	ECG rhythm at the time of admission to hospital: sinus (with a heart rate 60-90)	L_BLOOD	White blood cell count	1

Source	Source Label	Target	Target Label	Weight
L_BLOOD	White blood cell count	ROE	ESR (Erythrocyte sedimentation rate)	1
NA_KB	Use of opioid drugs by the Emergency Cardiology Team	D_AD_KBRIG	Diastolic blood pressure according to Emergency Cardiology Team	1
NA_KB	Use of opioid drugs by the Emergency Cardiology Team	ROE	ESR (Erythrocyte sedimentation rate)	1
NA_KB	Use of opioid drugs by the Emergency Cardiology Team	TIME_B_S	Time elapsed from the beginning of the attack of CHD to the hospital	1
NA_KB	Use of opioid drugs by the Emergency Cardiology Team	NA_BLOOD	Serum sodium content	1
NA_KB	Use of opioid drugs by the Emergency Cardiology Team	D_AD_ORIT	Diastolic blood pressure according to intensive care unit	1
ALT_BLOOD	Serum AlAT content (ALT_BLOOD)	AGE	Age of patient.	1
ALT_BLOOD	Serum AlAT content (ALT_BLOOD)	D_AD_ORIT	Diastolic blood pressure according to intensive care unit	1
ALT_BLOOD	Serum AlAT content (ALT_BLOOD)	NA_BLOOD	Serum sodium content	1
ALT_BLOOD	Serum AlAT content (ALT_BLOOD)	D_AD_KBRIG	Diastolic blood pressure according to Emergency Cardiology Team	1
ALT_BLOOD	Serum AlAT content (ALT_BLOOD)	S_AD_ORIT	Systolic blood pressure according to intensive care unit	1
SIM_GIPERT	Symptomatic hypertension	S_AD_ORIT	Systolic blood pressure according to intensive care unit	1
INF_ANAM	Quantity of myocardial infarctions in the anamnesis.	S_AD_ORIT	Systolic blood pressure according to intensive care unit	1
GB	Presence of an essential hypertension	S_AD_ORIT	Systolic blood pressure according to intensive care unit	1
SEX	Male	S_AD_ORIT	Systolic blood pressure according to intensive care unit	1
D_AD_ORIT	Diastolic blood pressure according to intensive care unit	S_AD_ORIT	Systolic blood pressure according to intensive care unit	1
K_SH_POST	Cardiogenic shock at the time of admission to intensive care unit	S_AD_ORIT	Systolic blood pressure according to intensive care unit	1

Source	Source Label	Target	Target Label	Weight
ant_im	Presence of an anterior myocardial infarction (left ventricular)	S_AD_ORIT	Systolic blood pressure according to intensive care unit	1
ritm_ecg_p_01	ECG rhythm at the time of admission to hospital: sinus (with a heart rate 60-90)	S_AD_ORIT	Systolic blood pressure according to intensive care unit	1
ritm_ecg_p_07	ECG rhythm at the time of admission to hospital: sinus with a heart rate above 90 (tachycardia)	S_AD_ORIT	Systolic blood pressure according to intensive care unit	1
n_r_ecg_p_05	Paroxysms of atrial fibrillation on ECG at the time of admission to hospital	S_AD_ORIT	Systolic blood pressure according to intensive care unit	1
P_IM_STEN	Post-infarction angina	S_AD_ORIT	Systolic blood pressure according to intensive care unit	1
K_BLOOD	Serum potassium content	S_AD_ORIT	Systolic blood pressure according to intensive care unit	1
ALT_BLOOD	Serum AlAT content (ALT_BLOOD)	S_AD_ORIT	Systolic blood pressure according to intensive care unit	1
R_AB_3_n	Relapse of the pain in the third day of the hospital period	S_AD_ORIT	Systolic blood pressure according to intensive care unit	1
NOT_NA_KB	Use of NSAIDs by the Emergency Cardiology Team	S_AD_ORIT	Systolic blood pressure according to intensive care unit	1
LID_S_n	Use of lidocaine in the ICU	S_AD_ORIT	Systolic blood pressure according to intensive care unit	1
ANT_CA_S_n	Use of calcium channel blockers in the ICU	S_AD_ORIT	Systolic blood pressure according to intensive care unit	1
TRENT_S_n	Use of Trental in the ICU	S_AD_ORIT	Systolic blood pressure according to intensive care unit	1
RAZRIV	Myocardial rupture	S_AD_ORIT	Systolic blood pressure according to intensive care unit	1
LET_IS	Lethal outcome	S_AD_ORIT	Systolic blood pressure according to intensive care unit	1
IBS_POST	Coronary heart disease (CHD) in recent weeks, days before admission to hospital	ROE	ESR (Erythrocyte sedimentation rate)	1

Source	Source Label	Target	Target Label	Weight
GB	Presence of an essential hypertension	ROE	ESR (Erythrocyte sedimentation rate)	1
INF_ANAM	Quantity of myocardial infarctions in the anamnesis.	ROE	ESR (Erythrocyte sedimentation rate)	1
L_BLOOD	White blood cell count	ROE	ESR (Erythrocyte sedimentation rate)	1
NA_R_1_n	Use of opioid drugs in the ICU in the first hours of the hospital period	ROE	ESR (Erythrocyte sedimentation rate)	1
endocr_01	Diabetes mellitus in the anamnesis	ROE	ESR (Erythrocyte sedimentation rate)	1
ritm_ecg_p_07	ECG rhythm at the time of admission to hospital: sinus with a heart rate above 90 (tachycardia)	ROE	ESR (Erythrocyte sedimentation rate)	1
ASP_S_n	Use of acetylsalicylic acid in the ICU	ROE	ESR (Erythrocyte sedimentation rate)	1
GEPAR_S_n	Use of an anticoagulants (heparin) in the ICU	ROE	ESR (Erythrocyte sedimentation rate)	1
R_AB_1_n	Relapse of the pain in the first hours of the hospital period	ROE	ESR (Erythrocyte sedimentation rate)	1
R_AB_2_n	Relapse of the pain in the second day of the hospital period	ROE	ESR (Erythrocyte sedimentation rate)	1
R_AB_3_n	Relapse of the pain in the third day of the hospital period	ROE	ESR (Erythrocyte sedimentation rate)	1
NA_KB	Use of opioid drugs by the Emergency Cardiology Team	ROE	ESR (Erythrocyte sedimentation rate)	1
LID_KB	Use of lidocaine by the Emergency Cardiology Team	ROE	ESR (Erythrocyte sedimentation rate)	1
SEX	Male	ROE	ESR (Erythrocyte sedimentation rate)	1
NA_R_2_n	Use of opioid drugs in the ICU in the second day of the hospital period	D_AD_ORIT	Diastolic blood pressure according to intensive care unit	1
PREDS_TAH	Supraventricular tachycardia	D_AD_ORIT	Diastolic blood pressure according to intensive care unit	1
FIBR_JELUD	Ventricular fibrillation	D_AD_ORIT	Diastolic blood pressure according to intensive care unit	1
NA_R_3_n	Use of opioid drugs in the ICU in the third day of the hospital period	D_AD_ORIT	Diastolic blood pressure according to intensive care unit	1

Source	Source Label	Target	Target Label	Weight
ritm_ecg_p_07	ECG rhythm at the time of admission to hospital: sinus with a heart rate above 90 (tachycardia)	D_AD_ORIT	Diastolic blood pressure according to intensive care unit	1
n_p_ecg_p_06	Third-degree AV block on ECG at the time of admission to hospital	D_AD_ORIT	Diastolic blood pressure according to intensive care unit	1
NOT_NA_KB	Use of NSAIDs by the Emergency Cardiology Team	D_AD_ORIT	Diastolic blood pressure according to intensive care unit	1
NA_KB	Use of opioid drugs by the Emergency Cardiology Team	D_AD_ORIT	Diastolic blood pressure according to intensive care unit	1
R_AB_3_n	Relapse of the pain in the third day of the hospital period	D_AD_ORIT	Diastolic blood pressure according to intensive care unit	1
R_AB_1_n	Relapse of the pain in the first hours of the hospital period	D_AD_ORIT	Diastolic blood pressure according to intensive care unit	1
KFK_BLOOD	Serum CPK content	D_AD_ORIT	Diastolic blood pressure according to intensive care unit	1
ALT_BLOOD	Serum AlAT content (ALT_BLOOD)	D_AD_ORIT	Diastolic blood pressure according to intensive care unit	1
K_BLOOD	Serum potassium content	D_AD_ORIT	Diastolic blood pressure according to intensive care unit	1
GIPO_K	Hypokalemia (¡ 4 mmol/L)	D_AD_ORIT	Diastolic blood pressure according to intensive care unit	1
fibr_ter_06	Fibrinolytic therapy by Celiasum 500k	D_AD_ORIT	Diastolic blood pressure according to intensive care unit	1
P_IM_STEN	Post-infarction angina	D_AD_ORIT	Diastolic blood pressure according to intensive care unit	1
fibr_ter_01	Fibrinolytic therapy by Celiasum 750k	D_AD_ORIT	Diastolic blood pressure according to intensive care unit	1
n_p_ecg_p_12	Complete RBBB on ECG at the time of admission to hospital	D_AD_ORIT	Diastolic blood pressure according to intensive care unit	1
n_p_ecg_p_07	LBBB (anterior branch) on ECG at the time of admission to hospital	D_AD_ORIT	Diastolic blood pressure according to intensive care unit	1
LET_IS	Lethal outcome	D_AD_ORIT	Diastolic blood pressure according to intensive care unit	1

Source	Source Label	Target	Target Label	Weight
R_AB_1_n	Relapse of the pain in the first hours of the hospital period	ROE	ESR (Erythrocyte sedimentation rate)	1
R_AB_1_n	Relapse of the pain in the first hours of the hospital period	NA_BLOOD	Serum sodium content	1
R_AB_1_n	Relapse of the pain in the first hours of the hospital period	D_AD_ORIT	Diastolic blood pressure according to intensive care unit	1
AST_BLOOD	Serum AsAT content	D_AD_KBRIG	Diastolic blood pressure according to Emergency Cardiology Team	1
R_AB_2_n	Relapse of the pain in the second day of the hospital period	ROE	ESR (Erythrocyte sedimentation rate)	1
R_AB_2_n	Relapse of the pain in the second day of the hospital period	D_AD_KBRIG	Diastolic blood pressure according to Emergency Cardiology Team	1
NA_KB	Use of opioid drugs by the Emergency Cardiology Team	TIME_B_S	Time elapsed from the beginning of the attack of CHD to the hospital	1
NA_R_1_n	Use of opioid drugs in the ICU in the first hours of the hospital period	TIME_B_S	Time elapsed from the beginning of the attack of CHD to the hospital	1
K_BLOOD	Serum potassium content	S_AD_ORIT	Systolic blood pressure according to intensive care unit	1
K_BLOOD	Serum potassium content	NA_BLOOD	Serum sodium content	1
K_BLOOD	Serum potassium content	D_AD_KBRIG	Diastolic blood pressure according to Emergency Cardiology Team	1
K_BLOOD	Serum potassium content	D_AD_ORIT	Diastolic blood pressure according to intensive care unit	1
NOT_NA_KB	Use of NSAIDs by the Emergency Cardiology Team	S_AD_ORIT	Systolic blood pressure according to intensive care unit	1
NOT_NA_KB	Use of NSAIDs by the Emergency Cardiology Team	S_AD_KBRIG	Systolic blood pressure according to Emergency Cardiology Team	1
NOT_NA_KB	Use of NSAIDs by the Emergency Cardiology Team	D_AD_KBRIG	Diastolic blood pressure according to Emergency Cardiology Team	1
NOT_NA_KB	Use of NSAIDs by the Emergency Cardiology Team	D_AD_ORIT	Diastolic blood pressure according to intensive care unit	1
LID_KB	Use of lidocaine by the Emergency Cardiology Team	S_AD_KBRIG	Systolic blood pressure according to Emergency Cardiology Team	1

Source	Source Label	Target	Target Label	Weight
LID_KB	Use of lidocaine by the Emergency Cardiology Team	D_AD_KBRIG	Diastolic blood pressure according to Emergency Cardiology Team	1
LID_KB	Use of lidocaine by the Emergency Cardiology Team	ROE	ESR (Erythrocyte sedimentation rate)	1
DRESSLER	Dressler syndrome	D_AD_ORIT	Diastolic blood pressure according to intensive care unit	1
KFK_BLOOD	Serum CPK content	D_AD_ORIT	Diastolic blood pressure according to intensive care unit	1
NA_R_1_n	Use of opioid drugs in the ICU in the first hours of the hospital period	D_AD_KBRIG	Diastolic blood pressure according to Emergency Cardiology Team	1
NA_R_2_n	Use of opioid drugs in the ICU in the second day of the hospital period	D_AD_KBRIG	Diastolic blood pressure according to Emergency Cardiology Team	1
zab_leg_01	Chronic bronchitis in the anamnesis	D_AD_KBRIG	Diastolic blood pressure according to Emergency Cardiology Team	1
ZSN_A	Presence of chronic Heart failure (HF) in the anamnesis	S_AD_KBRIG	Systolic blood pressure according to Emergency Cardiology Team	1
endocr_01	Diabetes mellitus in the anamnesis	D_AD_KBRIG	Diastolic blood pressure according to Emergency Cardiology Team	1
RAZRIV	Myocardial rupture	S_AD_KBRIG	Systolic blood pressure according to Emergency Cardiology Team	1
ZSN	Chronic heart failure	D_AD_KBRIG	Diastolic blood pressure according to Emergency Cardiology Team	1
IBS_NASL	Heredity on CHD	D_AD_KBRIG	Diastolic blood pressure according to Emergency Cardiology Team	1
B_BLOK_S_n	Use of beta-blockers in the ICU	S_AD_KBRIG	Systolic blood pressure according to Emergency Cardiology Team	1
B_BLOK_S_n	Use of beta-blockers in the ICU	D_AD_KBRIG	Diastolic blood pressure according to Emergency Cardiology Team	1
lat_im	Presence of a lateral myocardial infarction (left ventricular)	ant_im	Presence of an anterior myocardial infarction (left ventricular)	1
K_SH_POST	Cardiogenic shock at the time of admission to intensive care unit	S_AD_KBRIG	Systolic blood pressure according to Emergency Cardiology Team	1

Source	Source Label	Target	Target Label	Weight
TRENT_S_n	Use of Trental in the ICU	D_AD_KBRIG	Diastolic blood pressure according to Emergency Cardiology Team	1
ritm_ecg_p_07	ECG rhythm at the time of admission to hospital: sinus with a heart rate above 90 (tachycardia)	D_AD_KBRIG	Diastolic blood pressure according to Emergency Cardiology Team	1
GB	Presence of an essential hypertension	D_AD_KBRIG	Diastolic blood pressure according to Emergency Cardiology Team	1
ASP_S_n	Use of acetylsalicylic acid in the ICU	D_AD_KBRIG	Diastolic blood pressure according to Emergency Cardiology Team	1
GIPO_K	Hypokalemia (¡ 4 mmol/L)	D_AD_KBRIG	Diastolic blood pressure according to Emergency Cardiology Team	1
P_IM_STEN	Post-infarction angina	D_AD_KBRIG	Diastolic blood pressure according to Emergency Cardiology Team	1
lat_im	Presence of a lateral myocardial infarction (left ventricular)	D_AD_KBRIG	Diastolic blood pressure according to Emergency Cardiology Team	1
lat_im	Presence of a lateral myocardial infarction (left ventricular)	ant_im	Presence of an anterior myocardial infarction (left ventricular)	1
FK_STENOK	Functional class (FC) of angina pectoris in the last year.	S_AD_KBRIG	Systolic blood pressure according to Emergency Cardiology Team	1
FK_STENOK	Functional class (FC) of angina pectoris in the last year.	D_AD_KBRIG	Diastolic blood pressure according to Emergency Cardiology Team	1
MP_TP_POST	Paroxysms of atrial fibrillation at the time of admission to intensive care unit, (or at a pre-hospital stage)	D_AD_KBRIG	Diastolic blood pressure according to Emergency Cardiology Team	1
FIBR_PREDS	Atrial fibrillation	D_AD_KBRIG	Diastolic blood pressure according to Emergency Cardiology Team	1
GEPAR_S_n	Use of an anticoagulants (heparin) in the ICU	S_AD_KBRIG	Systolic blood pressure according to Emergency Cardiology Team	1
K_SH_POST	Cardiogenic shock at the time of admission to intensive care unit	S_AD_KBRIG	Systolic blood pressure according to Emergency Cardiology Team	1
ZSN_A	Presence of chronic Heart failure (HF) in the anamnesis	S_AD_KBRIG	Systolic blood pressure according to Emergency Cardiology Team	1

Source	Source Label	Target	Target Label	Weight
B_BLOK_S_n	Use of beta-blockers in the ICU	S_AD_KBRIG	Systolic blood pressure according to Emergency Cardiology Team	1
D_AD_KBRIG	Diastolic blood pressure according to Emergency Cardiology Team	S_AD_KBRIG	Systolic blood pressure according to Emergency Cardiology Team	1
RAZRIV	Myocardial rupture	S_AD_KBRIG	Systolic blood pressure according to Emergency Cardiology Team	1
FK_STENOK	Functional class (FC) of angina pectoris in the last year.	S_AD_KBRIG	Systolic blood pressure according to Emergency Cardiology Team	1
GEPAR_S_n	Use of an anticoagulants (heparin) in the ICU	S_AD_KBRIG	Systolic blood pressure according to Emergency Cardiology Team	1
IBS_POST	Coronary heart disease (CHD) in recent weeks, days before admission to hospital	D_AD_KBRIG	Diastolic blood pressure according to Emergency Cardiology Team	1
FIBR_PREDS	Atrial fibrillation	D_AD_KBRIG	Diastolic blood pressure according to Emergency Cardiology Team	1
NA_R_2_n	Use of opioid drugs in the ICU in the second day of the hospital period	D_AD_KBRIG	Diastolic blood pressure according to Emergency Cardiology Team	1
ASP_S_n	Use of acetylsalicylic acid in the ICU	D_AD_KBRIG	Diastolic blood pressure according to Emergency Cardiology Team	1
NA_R_1_n	Use of opioid drugs in the ICU in the first hours of the hospital period	D_AD_KBRIG	Diastolic blood pressure according to Emergency Cardiology Team	1
MP_TP_POST	Paroxysms of atrial fibrillation at the time of admission to intensive care unit, (or at a pre-hospital stage)	D_AD_KBRIG	Diastolic blood pressure according to Emergency Cardiology Team	1
B_BLOK_S_n	Use of beta-blockers in the ICU	D_AD_KBRIG	Diastolic blood pressure according to Emergency Cardiology Team	1
P_IM_STEN	Post-infarction angina	D_AD_KBRIG	Diastolic blood pressure according to Emergency Cardiology Team	1
ritm_ecg_p_07	ECG rhythm at the time of admission to hospital: sinus with a heart rate above 90 (tachycardia)	D_AD_KBRIG	Diastolic blood pressure according to Emergency Cardiology Team	1
lat_im	Presence of a lateral myocardial infarction (left ventricular)	D_AD_KBRIG	Diastolic blood pressure according to Emergency Cardiology Team	1

Source	Source Label	Target	Target Label	Weight
D_AD_KBRIG	Diastolic blood pressure according to Emergency Cardiology Team	S_AD_KBRIG	Systolic blood pressure according to Emergency Cardiology Team	1

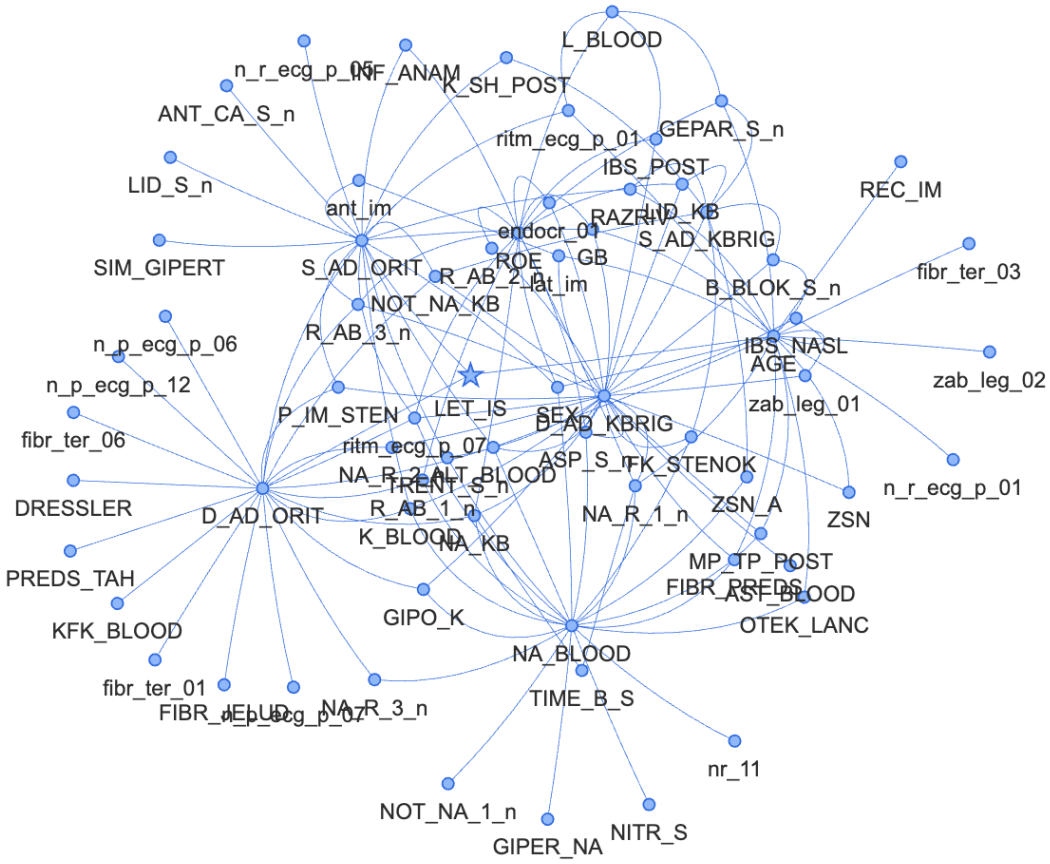

Figure 13: Inferred Causal Network around Lethal Outcome in Myocardial Infarction by NOTEARS.