Connected Causal Graphs for Real-World Science

Amine M'Charrak[1,*] **Thomas Lukasiewicz**[1,2], **Michael Bronstein**[1,3]
Abbavaram Gowtham Reddy[4], **Krikamol Muandet**[4]
[1]University of Oxford, [2]Vienna University of Technology, [3]AITHYRA,
[4]CISPA Helmholtz Center for Information Security

Abstract

In scientific practice, variables are rarely measured at random: they are chosen because experts expect them to be causally relevant and part of the same underlying causal system. This implies that realistic causal graphs should be *sparse*, reflecting simple mechanisms, yet also *connected*, since no variable is truly isolated. Existing continuous optimisation methods for learning directed acyclic graphs (DAGs) enforce sparsity and acyclicity but often produce fragmented structures, contradicting this basic property of scientific data. We address this gap by introducing a spectral regulariser based on *algebraic connectivity*, the Fiedler eigenvalue of the graph Laplacian. The penalty is differentiable, inexpensive to compute, and model-agnostic, and can be added to any learner that outputs a weighted adjacency. We demonstrate its effectiveness in two representative frameworks—GOLEM (likelihood-based, linear Gaussian) and a graph autoencoder (nonlinear encoder–decoder)—without altering their optimisation routines. Across synthetic benchmarks of sparse weakly connected DAGs and Erdős–Rényi DAGs with up to 200 nodes, the regulariser consistently improves global graph structure, yielding larger components and fewer isolated nodes, while preserving or improving edge-level recovery (higher F1, lower SHD and SID). These results establish algebraic connectivity as a principled and practical tool for causal discovery, aligning learned graphs with the way scientific data are collected and offering a simple drop-in enhancement to existing methods.

1 Introduction

In many areas of science such as biology, chemistry, or medicine, data collection is rarely carried out in a neutral way. As a process of measurement is often expensive and time-consuming, researchers typically decide in advance which variables are worth collecting. This decision is not arbitrary: variables are chosen because experts believe they are likely to be causally relevant. As a result, scientific datasets usually reflect a strong bias towards relevance: the measured variables are expected to participate in the same underlying process rather than being unrelated or isolated. We refer to this as a *principle of modularity*[1]. At the same time, the mechanisms behind these processes are often sparse, with only a limited number of direct interactions. Taken together, these considerations suggest that realistic causal graphs for scientific data should be both *sparse* and *connected*.

This poses a difficulty for current causal discovery methods. Modern continuous optimisation approaches to learning directed acyclic graphs (DAGs) (e.g. Zheng et al. [2018], Ng et al. [2020]) are designed to enforce sparsity, which is desirable for interpretation and often necessary for identifiability. However, strong sparsity penalties tend to fragment the learned graphs, leading to multiple

*Corresponding author: amine.mcharrak@cs.ox.ac.uk

[1]Here we use "modularity" to mean that variables are curated to belong to a single causal system; this is distinct from the notion of modularity in network science, which refers to the presence of community structure.

39th Conference on Neural Information Processing Systems (NeurIPS 2025) Workshop: New Perspectives in Advancing Graph Machine Learning.

disconnected components or even isolated nodes. While mathematically valid, such graphs conflict with the way scientific data is collected: if variables were hand-picked by experts for their importance, we do not expect any of them to be completely unrelated to the rest.

The aim of this paper is to address this gap. We postulate that causal discovery methods should allow for sparse structures, yet simultaneously reflect the expectation that all measured variables belong to a single weakly connected system. To achieve this, we draw on ideas from spectral graph theory, and in particular the notion of algebraic connectivity given by the Fiedler eigenvalue of the graph Laplacian [Fiedler, 1975]. We incorporate a connectivity penalty based on the Fiedler eigenvalue into continuous DAG learning objectives. This addition biases the optimisation towards solutions that are not only sparse, but also more likely to form a single connected component. Even when full connectivity is not achieved, the penalty increases the size of components and reduces the occurrence of isolated nodes.

To the best of our knowledge, this is the first attempt to combine algebraic connectivity with continuous optimisation-based causal discovery. The method is simple, computationally light, and easy to integrate into existing frameworks. We view this as a step towards causal discovery that better reflects the way scientific data is generated, making these methods more useful for applied domains such as biology, chemistry, and medicine.

2 Related Work

Causal graph structure learning. We focus on the setting where all common causes of the measured variables are observed (causal sufficiency). The task is to recover a directed acyclic graph (DAG) from observational data by optimising a score that balances data fit with structural constraints. These constraints must rule out cycles and typically include sparsity-inducing penalties to avoid spurious edges. Classical score- and constraint-based approaches remain widely used, but they rely on discrete search and often struggle to scale. A recent line of work has instead formulated DAG learning as a smooth optimisation problem, making it possible to apply gradient-based methods.

Zheng et al. [2018] were the first to introduce this approach: they proposed a differentiable characterisation of acyclicity and optimised a continuous objective that jointly enforces data fit, sparsity, and acyclicity. This formulation removes the need for combinatorial search and opened the door to scalable continuous methods for structure learning. Building on this idea, Ng et al. [2020] analysed the role of sparsity and acyclicity specifically in the linear Gaussian case and proposed a likelihood-based objective where both structural parameters and the adjacency matrix are estimated jointly. Their method, GOLEM, is computationally simple and provides a strong baseline for score-based continuous learning. In parallel, Ng et al. [2019] developed a graph autoencoder approach in which dependencies between variables are parameterised by an encoder–decoder architecture. The model is trained to reconstruct the data under sparsity and acyclicity regularisation, providing a flexible framework that can naturally capture non-linear effects. In this paper, we adopt both the GOLEM and graph autoencoder settings as representative baselines to demonstrate that our proposed connectivity regularisation is broadly applicable and can be integrated without altering the overall learning framework.

Spectral graph theory and algebraic connectivity. A separate line of work in spectral graph theory connects graph properties to the eigenvalues of matrices such as the Laplacian. A central concept is *algebraic connectivity*, introduced by Fiedler [1975], which is defined as the second smallest eigenvalue of the Laplacian. This value quantifies how well a graph is connected: a positive eigenvalue indicates that the graph forms a single connected component, whereas a value of zero implies disconnectedness. Since its introduction, algebraic connectivity has become a standard tool for reasoning about graph robustness, bottlenecks, and connectedness. Several recent contributions have applied this concept in machine learning. Tam and Dunson [2020] proposed *Fiedler regularisation* for training sparse neural networks. They represent the nonzero weights of a network as edges in a graph and add a penalty based on the Fiedler eigenvalue, in addition to standard sparsity penalties. This regularisation discourages pruning strategies that break the network into many isolated subgraphs, and instead promotes architectures that remain structurally coherent. Their results show that explicitly encouraging algebraic connectivity stabilises training and improves generalisation compared with sparsity alone. In a different domain, He et al. [2022] studied global ranking from noisy pairwise comparisons. They modelled comparisons as a directed graph with items as nodes and edges as observed outcomes, and introduced *GNNRank*, a graph neural network designed to exploit this structure. The method propagates information along directed edges and leverages spectral principles to ensure that the learned embeddings preserve consistent global rankings. This illustrates how graph-theoretic and spectral methods can be integrated into neural architectures even outside traditional graph recovery tasks.

Positioning. Taken together, these strands of work highlight two complementary directions: continuous optimisation frameworks that enable scalable causal discovery, and the use of spectral connectivity measures to regularise graph learning in machine learning. To our knowledge, these directions have not yet been combined to address a fundamental property of scientific datasets: variables are deliberately chosen and therefore expected to belong to a single weakly connected system. Our contribution is to bridge this gap by incorporating algebraic connectivity into continuous causal discovery. This allows learned graphs to remain sparse, while reducing the likelihood of fragmentation into disconnected components, thereby aligning causal discovery methods more closely with the way scientific data are generated in practice.

3 Proposed Approach and Methodology

High-level goal. We assume data are generated from a linear Gaussian structural equation model (SEM) with equal noise variances [Peters and Bühlmann, 2014]. Given n i.i.d. samples $X \in \mathbb{R}^{n \times d}$ of d observed variables, the task is to recover the weighted adjacency matrix $B \in \mathbb{R}^{d \times d}$, where $B_{ij} \neq 0$ encodes a directed edge $i \to j$. Our target class is the set of weighted DAGs. In addition to acyclicity, we favour solutions whose undirected skeleton is connected.

Notation. $B \in \mathbb{R}^{d \times d}$ denotes the weighted directed adjacency (SEM coefficients), where $B_{ij} \neq 0$ encodes a directed edge $i \to j$. For a matrix M, $\|M\|_1$ is the entrywise ℓ_1 norm, $M \circ N$ is the Hadamard product, and $\exp(M)$ is the matrix exponential.

To measure connectivity, we construct an undirected smooth skeleton W from B:

$$W_{ij} \;=\; \sigma\!\left(\tfrac{|B_{ij}|}{\tau}\right) \;\vee\; \sigma\!\left(\tfrac{|B_{ji}|}{\tau}\right), \qquad \sigma(t) = \tfrac{1}{1+e^{-t}},$$

followed by symmetrisation $W \leftarrow \frac{1}{2}(W + W^\top)$. Here $\tau > 0$ controls the sharpness; as $\tau \downarrow 0$, W approaches the binary skeleton $\mathbf{1}\{\, B_{ij} \neq 0 \text{ or } B_{ji} \neq 0 \,\}$.

From W we form the degree matrix $D = \operatorname{diag}(W\mathbf{1})$ and the graph Laplacian $L = D - W$. Let $0 = \lambda_1(L) \leq \cdots \leq \lambda_d(L)$ be its eigenvalues; the second-smallest eigenvalue $\lambda_2(L)$, the *Fiedler value*, is positive if and only if the skeleton is connected.

3.1 Soft vs. hard acyclicity: formulation choice

There are two main strategies for enforcing acyclicity. *Hard* formulations impose $h_{\mathrm{DAG}}(B) = 0$ as an equality constraint and solve a constrained optimisation problem, typically with an augmented Lagrangian, as in NOTEARS [Zheng et al., 2018]. *Soft* formulations incorporate acyclicity as a penalty with finite weight, yielding an unconstrained problem that can be solved with gradient-based methods; this is the approach in GOLEM [Ng et al., 2020].

We adopt the soft route for three reasons: (i) it avoids maintaining dual variables and tuning feasibility schedules; (ii) it allows additional penalties such as connectivity to be added naturally; and (iii) under the linear Gaussian SEM with equal variances, the soft formulation is consistent and recovers the correct DAG asymptotically [Peters and Bühlmann, 2014, Ng et al., 2020]. A hard variant is possible but not required for our aims.

3.2 Likelihood-based score and acyclicity surrogate

Our score function is the likelihood for the linear Gaussian SEM with equal error variances:

$$\mathcal{L}_{\mathrm{EV}}(B; X) = \frac{d}{2} \log\left(\sum_{j=1}^{d} \sum_{k=1}^{n} \big(x_j^{(k)} - B_{:j}^\top x^{(k)}\big)^2 \right) \;-\; \log\big|\det(I - B)\big|, \tag{1}$$

where $B_{:j}$ is column j. The log-determinant term couples the regressions across variables and prevents solutions that trivially assign each variable as parent of the others (i.e. two-cycles).

To ensure acyclicity, we use the smooth surrogate from Zheng et al. [2018]:

$$h_{\mathrm{DAG}}(B) \;=\; \mathrm{tr}\big(\exp(B \circ B)\big) \;-\; d, \tag{2}$$

which satisfies $h_{\mathrm{DAG}}(B) \geq 0$ with equality iff B encodes a DAG.

Remark 1 (Properties)**.** *(i) If B is strictly upper–triangular under some permutation, then $\log|\det(I-B)| = 0$ and $h_{\mathrm{DAG}}(B) = 0$. (ii) The gradient of h_{DAG} is everywhere defined and efficiently computable by automatic differentiation.*

3.3 Connectedness via algebraic connectivity

We measure connectivity using the Fiedler value $\lambda_2(L)$ of the Laplacian $L = D - W$, where W is the smooth skeleton defined above. By classical spectral graph theory, $\lambda_2(L) > 0$ if and only if the skeleton is connected. Equivalently, the directed graph defined by B is weakly connected. To encourage connectivity, we penalise shortfalls below a fixed target $\epsilon > 0$:

$$\mathrm{pen}_{\mathrm{conn}}(B;\epsilon) \;=\; \phi\big(\epsilon - \lambda_2(L)\big), \qquad \phi(t) = \log(1+e^t) \;\; \text{(softplus)}. \tag{3}$$

Remark 2 (Differentiability)**.** *Because W is a smooth function of B (via the logistic approximation), the mapping $B \mapsto \lambda_2(L)$ is differentiable almost everywhere. If u is the Fiedler eigenvector of L corresponding to $\lambda_2(L)$, then*

$$\frac{\partial \lambda_2}{\partial W_{ij}} = -(u_i - u_j)^2,$$

and gradients propagate back to B through the smooth skeleton construction. This makes the connectivity penalty fully compatible with modern autodiff frameworks.

3.4 Full objective and optimisation

Our final training objective is the unconstrained optimisation problem

$$\min_{B \in \mathbb{R}^{d \times d}} \;\; \mathcal{J}(B) = \underbrace{\mathcal{L}_{\mathrm{EV}}(B;X)}_{\text{likelihood}} + \lambda_1 \|B\|_1 \;+\; \lambda_2\, h_{\mathrm{DAG}}(B) \;+\; \lambda_3\, \mathrm{pen}_{\mathrm{conn}}(B;\epsilon). \tag{4}$$

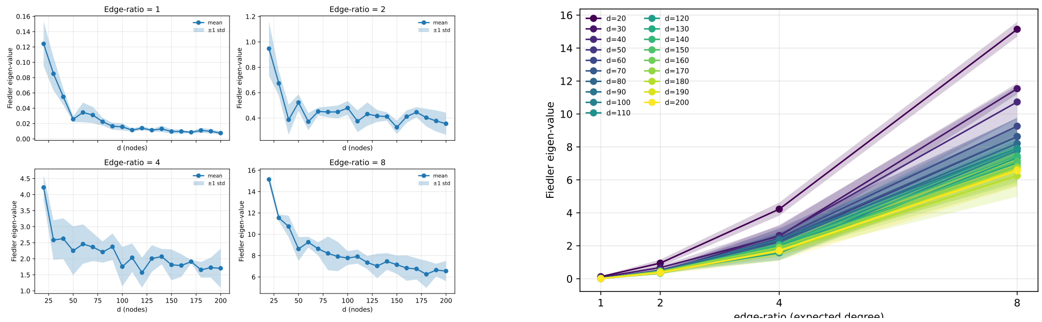
We optimise (4) directly with first–order methods (Adam) and automatic differentiation. Unlike constrained approaches, our method does not require projecting iterates onto the set of DAGs; acyclicity is enforced softly through $h_{\mathrm{DAG}}(B)$, and connectivity through $\lambda_2(L)$.

Per iteration, the main computational costs are: (i) evaluating the log-determinant and the matrix exponential (both $\mathcal{O}(d^3)$ but GPU-friendly), and (ii) computing $\lambda_2(L)$ and its eigenvector for the smooth skeleton W.

Contrast with hard constraints. If the acyclicity condition is imposed exactly as $h_{\mathrm{DAG}}(B) = 0$, as in NOTEARS [Zheng et al., 2018], the problem becomes a constrained optimisation task solved with an augmented Lagrangian. This requires a dual ascent schedule and, in practice, large penalty parameters that can cause ill-conditioning. Our approach avoids this by treating acyclicity and connectivity as soft penalties within a single smooth scalar objective. The inclusion of the log-determinant term in the likelihood ensures that the optimisation does not favour mutual regressions that form two-cycles: if both B_{ij} and B_{ji} are large, then $|\det(I-B)|$ shrinks, raising the objective and discouraging such cyclic structures.

3.5 Practical details

Post–processing. After optimisation, we apply a threshold $\omega > 0$ to $|B_{ij}|$ to remove edges with very small weights. If cycles remain after thresholding, we iteratively prune the lowest-weight cycle–inducing edges until a DAG is obtained. Because the acyclicity penalty already discourages cycles, such edges typically have negligible magnitude.



(a) **Random—undirected**: Fiedler eigenvalue vs. graph size d (mean $\pm$ std over 5 seeds).

(b) **Graph—undirected**: Fiedler eigenvalue vs. edge–to–node ratio r (mean $\pm$ std over 5 seeds).

Figure 1: Algebraic connectivity (λ_2) in undirected graphs: (a) scaling with number of nodes d; (b) dependence on edge–to–node ratio r. Shaded regions show ± 1 std across 5 random seeds.

When a hard constraint may be preferable. If an exactly acyclic solution is strictly required (e.g. feasibility to numerical precision), one can enforce $h_{\mathrm{DAG}}(B) = 0$ using an augmented Lagrangian [Zheng et al., 2018]. This, however, introduces extra hyperparameters and often leads to numerical instability near feasibility, whereas our soft formulation avoids these issues while still converging to acyclic solutions in practice.

Regularisation Parameters. Because our experiments span graphs of different sizes and densities, we use the standard choices of (λ_1, λ_2) reported in the literature rather than dataset-specific tuning. The weight λ_3 is set by forward search, gradually increasing its value until the learned graph shows a measurable deviation from the baseline without the connectivity term. The connectivity threshold $\epsilon > 0$ is chosen such that $\lambda_2(L) < \epsilon$ for the initial skeleton, ensuring that the connectivity penalty is active from the start of optimisation.

4 Experiments

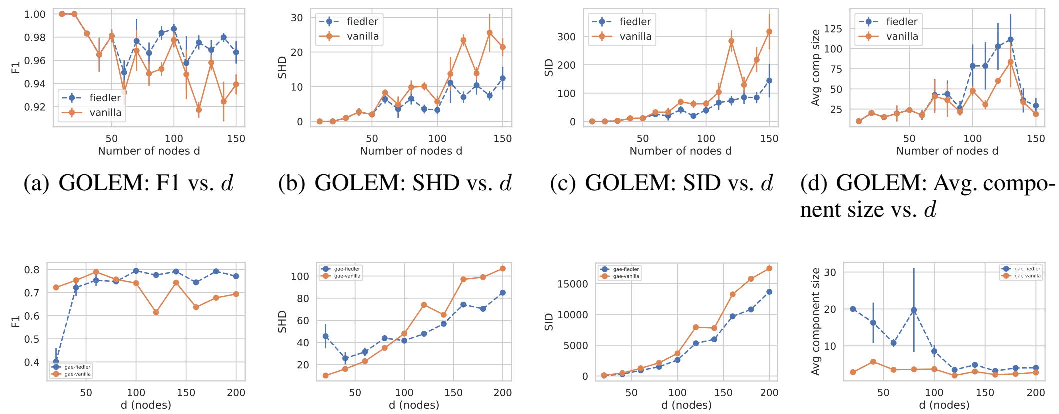
Graph and dataset generation. We evaluate our methods on two synthetic graph families. *(i) Sparse, weakly connected DAGs:* for each number of nodes $d \in \{20, 30, \ldots, 200\}$ we draw a random topological order and first construct a recursive spanning tree to ensure weak connectivity. We then add additional forward edges (respecting the topological order) until a prescribed edge–to–node ratio r is reached, with a minimum of $d - 1$ edges. The result is a binary adjacency matrix $B \in \{0, 1\}^{d \times d}$ that is acyclic and weakly connected. Each present edge is assigned a weight $W_{ij} \sim \mathrm{Unif}[0.5, 2.0]$, independently across edges (optionally with random signs). To generate data, we simulate $n = 1000$ samples from a linear Gaussian structural equation model (SEM) in topological order:

$$X_{\cdot j} \; = \; X_{\cdot \mathrm{pa}(j)} W_{\mathrm{pa}(j), j} + \varepsilon_{\cdot j}, \qquad \varepsilon_{\cdot j} \sim \mathcal{N}(0, 1),$$

where $\mathrm{pa}(j)$ denotes the set of parents of node j.

(ii) Erdős–Rényi DAGs (ERk): for each d and $k \in \{1, 2, 4, 8\}$ we generate a random topological order and include each forward edge independently with probability chosen such that the expected in-degree is k, i.e. the expected total number of edges is approximately kd. We use the same weight assignment and SEM simulator as above.

To better understand the behaviour of algebraic connectivity in our synthetic datasets, we provide visualisations in Figure 1. Panel 1(a) shows that for a fixed edge–to–node ratio, the Fiedler eigenvalue systematically decreases as the number of nodes d grows. This reflects the fact that larger sparse graphs are easier to disconnect and thus have lower algebraic connectivity. Panel 1(b) shows the complementary effect of varying the expected edge–ratio: as the ratio increases, the Fiedler eigenvalue increases, indicating that denser graphs are more robustly connected. Together, these plots provide intuition for why sparse large-scale graphs tend to fragment under standard learning objectives, and why explicitly regularising algebraic connectivity is necessary to counteract this effect.



(a) GOLEM: F1 vs. d

(b) GOLEM: SHD vs. d

(c) GOLEM: SID vs. d

(d) GOLEM: Avg. component size vs. d

(e) GAE: F1 vs. d

(f) GAE: SHD vs. d

(g) GAE: SID vs. d

(h) GAE: Avg. component size vs. d

Figure 2: Scalability with increasing d on sparse DAGs (edge–to–node ratio ≈ 1). Orange lines: vanilla baselines (GOLEM or GAE). Blue lines: our Fiedler-regularised variants.

Further visualisations provided in the Supplementary Material illustrate the relationship between algebraic connectivity and the edge–to–node ratio, and how this behaviour varies with the number of nodes.

Evaluation metrics. We evaluate the learned graphs using both standard structure-recovery and connectivity-aware metrics. Specifically, we report Structural Hamming Distance (SHD), Structural Intervention Distance (SID), and F1 score on edges as well as the average component size. These metrics capture not only the accuracy of causal discovery but also whether the recovered graph satisfies the scientific prior that observed variables belong to a single or a few connected systems.

Experimental Results for Causal Graph Structure Learning with Fiedler Eigenvalue Penalty Figure 2 compares vanilla GOLEM and GAE (orange) with their Fiedler-regularised counterparts (blue) on sparse DAGs with approximately d edges. Each panel reports one of four evaluation measures as a function of the number of nodes d: F1 score, Structural Hamming Distance (SHD), Structural Intervention Distance (SID), and the average size of weakly connected components in the learned graph. For a fair comparison, both variants are trained under identical hyper-parameters and optimisation settings, with the sole difference being the inclusion of the Fiedler penalty.

Two consistent patterns emerge. First, the addition of the Fiedler penalty improves the weak connectivity of the learned graphs, whereas the vanilla methods frequently yield estimates that fragment into multiple small components, the regularised variants return graphs with substantially larger components, often close to full weak connectivity. Second, recovery performance is not only preserved but frequently improved: F1 scores are higher and SHD/SID values lower for the Fiedler variants, with the performance gap widening as d increases.

These findings confirm that our regularisation is both effective and model-agnostic. It can be integrated into different continuous frameworks (here, a likelihood-based method and an autoencoder-based method) without changing their optimisation behaviour, while consistently enhancing both the structural accuracy and the global connectivity of the learned causal graphs.

5 Conclusion

We introduced a spectral regularisation for continuous causal discovery that encourages weak connectivity through the Fiedler eigenvalue of the Laplacian. This addresses a basic but often overlooked property of real scientific datasets: variables are rarely isolated, and by design reflect a bias towards relevance, a *principle of modularity* whereby measured variables are expected to belong to the same causal system. Our approach is model-agnostic, integrates seamlessly into existing optimisation frameworks, and improves global graph structure without compromising edge-level accuracy.

These results highlight the importance of connectivity-aware constraints for making causal discovery methods more realistic and reliable in scientific applications. Looking ahead, we view the principle of modularity as a foundation for developing causal discovery methods that better align statistical optimisation with the structural properties of real-world scientific data.

Acknowledgments and Disclosure of Funding

We thank the reviewers for their valuable feedback. Amine M'Charrak gratefully acknowledges support from the Evangelisches Studienwerk e.V. Villigst through a doctoral fellowship. Michael Bronstein is partially supported by the EPSRC Turing AI World-Leading Research Fellowship No. EP/X040062/1 and EPSRC AI Hub No. EP/Y028872/1.

References

Miroslav Fiedler. A property of eigenvectors of nonnegative symmetric matrices and its application to graph theory. *Czechoslovak Mathematical Journal*, 25:619–633, 1975. URL `https://api.semanticscholar.org/CorpusID:123006817`.

Yixuan He, Quan Gan, David Wipf, Gesine D Reinert, Junchi Yan, and Mihai Cucuringu. Gnnrank: Learning global rankings from pairwise comparisons via directed graph neural networks. In *International conference on machine learning*, pages 8581–8612. PMLR, 2022.

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.

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

Edric Tam and David Dunson. Fiedler regularization: Learning neural networks with graph sparsity. In *International Conference on Machine Learning*, pages 9346–9355. PMLR, 2020.

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.

A Additional Visualisations on Graph Properties

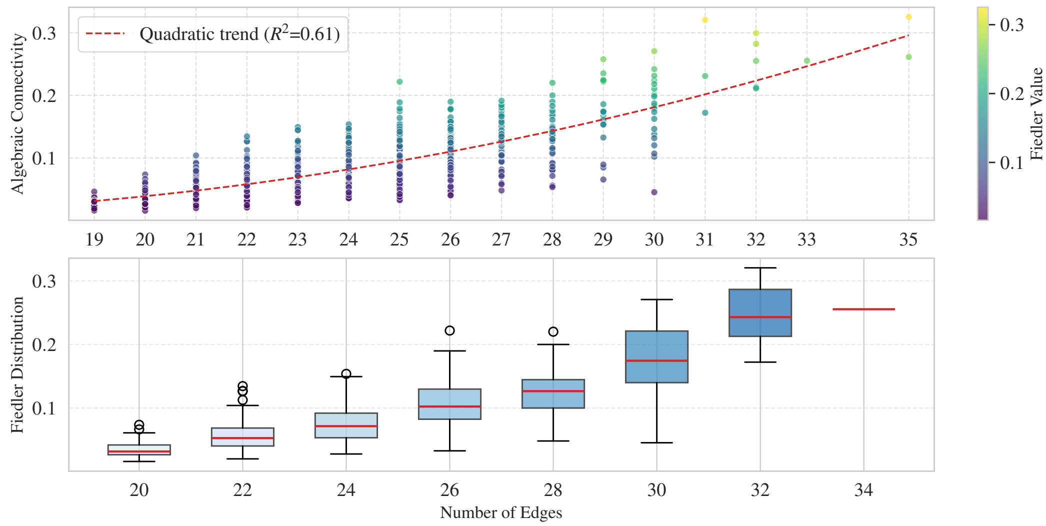


Figure 3: Relationship between Fiedler Eigenvalue and Graph Sparsity.

Table 1: Synthetic Graph Statistics

d	Edges	Edge-to-d Ratio
30	30	1.00
40	40	1.00
50	54	1.08
60	67	1.12
70	82	1.17
80	101	1.26
90	111	1.23
100	130	1.30
110	142	1.29
120	161	1.34
130	174	1.34
140	188	1.34
150	195	1.30