TU Delft

Delft University of Technology

Graph Topology Identification Based on Covariance Matching

Han, Yongsheng; Natali, Alberto; Leus, Geert

DOI
10.1109/ICASSP49660.2025.10888599

Publication date
2025

Document Version
Final published version

Published in
2025 IEEE International Conference on Acoustics, Speech, and Signal Processing, ICASSP 2025 - Proceedings

Citation (APA)
Han, Y., Natali, A., & Leus, G. (2025). Graph Topology Identification Based on Covariance Matching. In B. D. Rao, I. Trancoso, G. Sharma, & N. B. Mehta (Eds.), *2025 IEEE International Conference on Acoustics, Speech, and Signal Processing, ICASSP 2025 - Proceedings* (ICASSP, IEEE International Conference on Acoustics, Speech and Signal Processing - Proceedings). IEEE. https://doi.org/10.1109/ICASSP49660.2025.10888599

Important note
To cite this publication, please use the final published version (if applicable).
Please check the document version above.

Copyright
Other than for strictly personal use, it is not permitted to download, forward or distribute the text or part of it, without the consent of the author(s) and/or copyright holder(s), unless the work is under an open content license such as Creative Commons.

Takedown policy
Please contact us and provide details if you believe this document breaches copyrights.
We will remove access to the work immediately and investigate your claim.

This work is downloaded from Delft University of Technology.
For technical reasons the number of authors shown on this cover page is limited to a maximum of 10.

Green Open Access added to TU Delft Institutional Repository

'You share, we take care!' - Taverne project

https://www.openaccess.nl/en/you-share-we-take-care

Otherwise as indicated in the copyright section: the publisher is the copyright holder of this work and the author uses the Dutch legislation to make this work public.

ICASSP 2025 - 2025 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP) | 979-8-3503-6874-1/25/$31.00 ©2025 IEEE | DOI: 10.1109/ICASSP49660.2025.10888599

Graph Topology Identification Based on Covariance Matching

Yongsheng Han, Alberto Natali, Geert Leus
Faculty of Electrical Engineering, Mathematics and Computer Science
Delft University of Technology
Delft, The Netherlands
{y.han, a.natali, g.j.t.leus}@tudelft.nl

***Abstract*—This paper addresses graph topology identification for applications where the underlying structure of systems like brain and social networks is not directly observable. Traditional approaches based on signal matching and spectral templates have limitations, particularly in handling scale issues and sparsity assumptions. We introduce a novel covariance matching methodology that efficiently reconstructs the graph topology using observable data. For the structural equation model (SEM) using an undirected graph, we demonstrate that our method can converge to the correct result under relatively soft conditions. Furthermore, we extend our methodology to polynomial models and any known distribution of latent variables, broadening its applicability and utility in diverse graph-based systems.**

***Index Terms*—graph topology identification, covariance matching, structural equation model, polynomial model**

I. Introduction

Graph topology identification remains a critical issue in graph signal processing (GSP), where systems are modeled as networks, yet their actual underlying structure is often invisible. Examples of such systems include brain functional connectivity networks and social networks. In these applications, while the direct graph structure is not observable, nodal data is typically available. For instance, in academic networks [1], the advisor-advisee links may not be visible, yet we can analyze collaborative patterns to uncover these connections. Similarly, in brain networks [2], neural signals provide indirect clues about the connectivity. Therefore, the primary challenge in graph topology identification lies in deducing the hidden graph structure from these nodal observations, a task that is fundamental for analyzing and understanding the interactions within these networks.

The structural equation model (SEM) is a popular tool to link nodal data with the graph, and it has been frequently used in graph topology identification [3], [4], [5], [6]. A remarkable result was provided by [6], where it was demonstrated that for sparse directed acyclic graphs (DAGs), the graph can be uniquely determined when the unknown external inputs are Gaussian with equal variance. This foundational work spurred further developments, leading to more efficient algorithms as evidenced by [7] and [8]. However, for undirected graphs, methods using signal matching have performed poorly without the presence of exogenous variables, even with the introduction of sparsity constraints [9], [10]. On the other hand, spectral template-based approaches, such as the polynomial graphical lasso (PGL) algorithm, have shown potential in handling certain graphs [11], but the results typically differ from the true structure by a scale factor and require extensive restrictions (sparsity and sign) on the graph for the method to be effective.

To the best of our knowledge, no existing work has exhaustively addressed topology identification for undirected graphs using a SEM. This paper fills this gap by proposing a novel covariance matching-based method. We will prove that under relatively soft conditions, our proposed method consistently converges to the correct result without encountering the scale issue often associated with other approaches, and our method is more robust. Furthermore, we do not need to make any assumptions about sparsity. We will also extend our method to accommodate polynomial models and any known distribution of latent variables, broadening the applicability of our approach in graph topology identification.

*This work is partially supported by the NWO OTP GraSPA proposal #19497, and the EU HORIZON-CHIPS-JU-2023-2-RIA ShapeFuture project, under grant agreement No 101139996.

II. Preliminaries

In this section, some background information is presented that is required to explain the main contributions of this work.

Graph signal processing: We describe a graph as $\mathcal{G} = \{\mathcal{V}, \mathcal{E}, \mathbf{S}\}$, where $\mathcal{V} = \{1, \dots, N\}$ represents the set of vertices, $\mathcal{E} \subseteq \mathcal{V} \times \mathcal{V}$ denotes the set of edges, and $\mathbf{S} \in \mathbb{R}^{N \times N}$ is referred to as the graph shift operator (GSO). It is an $N \times N$ matrix that captures the local structure of the graph, whose entries S_{ij} are non-zero only if $i = j$ or if $(j, i) \in \mathcal{E}$ [9]. The adjacency matrix or the combinatorial graph Laplacian are matrices that can be used as GSO. Each node $i \in \mathcal{V}$ is associated with a scalar value x_i. By stacking these values into a vector $\mathbf{x} = [x_1, \dots, x_N]^\top \in \mathbb{R}^N$, we obtain what is known as a graph signal.

Structural equation model: Considering a simplified SEM excluding the influence of exogenous variables, the internal structure of a graph signal $\mathbf{x}$ can be expressed as

$$\mathbf{x} = \mathbf{S}\mathbf{x} + \mathbf{e}. \tag{1}$$

Here, the matrix $\mathbf{S}$ represents the GSO of the graph. Clearly, this GSO should have zero diagonal entries indicating no self-influence and hence can be interpreted as an adjacency matrix. Furthermore, the vector $\mathbf{e}$, assumed to follow a zero-mean white Gaussian distribution, i.e., $\mathbf{e} \sim \mathcal{N}(\mathbf{0}, \mathbf{I})$, accounts for the noise. This assumption is inspired by the findings in [6]. Therefore, the covariance of $\mathbf{x}$ can be derived as $\mathbb{E}[\mathbf{x}\mathbf{x}^\top] = \mathbb{E}[(\mathbf{I} - \mathbf{S})^{-1}\mathbf{e}\mathbf{e}^\top(\mathbf{I} - \mathbf{S})^{-\top}] = (\mathbf{I} - \mathbf{S})^{-1}(\mathbf{I} - \mathbf{S})^{-\top}$.

In case we consider multiple independent realizations of $\mathbf{e}$, which can be stacked in $\mathbf{E} = [\mathbf{e}_1, \mathbf{e}_2, \dots, \mathbf{e}_T]$, we obtain multiple independent realizations of $\mathbf{x}$, grouped in $\mathbf{X} = [\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_T]$, as

$$\mathbf{X} = \mathbf{S}\mathbf{X} + \mathbf{E}. \tag{2}$$

Polynomial model: The SEM described in (1) (or equivalently (2)) essentially takes the form $\mathbf{x} = \mathbf{H}\mathbf{e}$, where $\mathbf{H} = (\mathbf{I} - \mathbf{S})^{-1}$. This representation can be viewed as a special case of a polynomial relationship because the matrix $\mathbf{H} = (\mathbf{I} - \mathbf{S})^{-1}$ can usually be expanded into a power series of $\mathbf{S}$. This concept forms the basis of the polynomial model, which considers

$$\mathbf{x} = h(\mathbf{S})\mathbf{e}. \tag{3}$$

Authorized licensed use limited to: TU Delft Library. Downloaded on May 09,2025 at 07:29:37 UTC from IEEE Xplore. Restrictions apply.

In this model, $h(x)$ represents any polynomial of the form $h(x) = \sum_{l=0}^{L-1} h_l x^l$. We again assume that $\mathbf{e} \sim \mathcal{N}(\mathbf{0}, \mathbf{I})$.

III. Topology Identification for SEM

In this section, we focus on the SEM for undirected graphs, which we then later on will extend to any other known polynomial model or latent variable distribution. Note that we will use a hat notation to represent the optimization variable whereas we use a star to indicate the optimal solution.

A straightforward approach to estimate a symmetric hollow $\mathbf{S}$ from (2) is to minimize $\|\mathbf{X} - \hat{\mathbf{S}}\mathbf{X}\|_F^2$ using the constraints $\text{diag}(\hat{\mathbf{S}}) = \mathbf{0}$ and $\hat{\mathbf{S}} = \hat{\mathbf{S}}^\top$. However, it can be shown that this does not lead to the ground truth $\mathbf{S}$, even when the number of samples T grows to infinity and $\mathbf{e} \sim \mathcal{N}(\mathbf{0}, \mathbf{I})$. To prove this, observe that for an infinite number of observations, the objective function can be defined as $f(\hat{\mathbf{S}}) = \mathbb{E}\{\|\mathbf{x} - \hat{\mathbf{S}}\mathbf{x}\|_2^2\} = \text{tr}((\mathbf{I} - \mathbf{S})^{-2}(\mathbf{I} - \hat{\mathbf{S}})^2)$. Given that $\hat{\mathbf{S}}$ is a symmetric matrix, for the optimal solution we should obtain $\frac{df}{d\hat{\mathbf{S}}} = \frac{\partial f}{\partial \hat{\mathbf{S}}} + \frac{\partial f}{\partial \hat{\mathbf{S}}^\top} - \text{Diag}(\frac{\partial f}{\partial \hat{\mathbf{S}}}) = \mathbf{\Lambda}$ [12], where $\mathbf{\Lambda}$ is a diagonal matrix because non-zero off-diagonal elements would allow gradient descent to further minimize f [1]. However, setting $\hat{\mathbf{S}}$ to the true $\mathbf{S}$, it can be proven that $\frac{\partial f}{\partial \hat{\mathbf{S}}}|_{\hat{\mathbf{S}}=\mathbf{S}} = -2(\mathbf{I}-\mathbf{S})^{-1}$. As a result, $\frac{df}{d\hat{\mathbf{S}}}|_{\hat{\mathbf{S}}=\mathbf{S}} = \mathbf{\Lambda}$ is only possible when $\mathbf{S}$ is a diagonal matrix, which is not possible.

Instead, we focus on a covariance matching approach, where we match the sample covariance matrix $\mathbf{C}_\mathbf{x} = \mathbf{X}\mathbf{X}^\top/T$ to the theoretical covariance matrix $\mathbf{\Sigma}_\mathbf{x} = \mathbb{E}\{\mathbf{x}\mathbf{x}^\top\}$ expected from the model. Under the conditions $\mathbf{e} \sim \mathcal{N}(\mathbf{0}, \mathbf{I})$ and $\mathbf{S} = \mathbf{S}^\top$, the latter is given by

$$\mathbf{\Sigma}_\mathbf{x} = (\mathbf{I} - \mathbf{S})^{-1}(\mathbf{I} - \mathbf{S})^{-\top} = (\mathbf{I} - \mathbf{S})^{-2}. \tag{4}$$

Introducing $\mathbf{H} = (\mathbf{I} - \mathbf{S})^{-1}$ as earlier, we thus obtain $\mathbf{\Sigma}_\mathbf{x} = \mathbf{H}^2$. Estimating $\mathbf{H}$ instead of $\mathbf{S}$, our covariance matching problem can be formulated as

$$\begin{aligned} \mathbf{H}^* = \arg\min_{\hat{\mathbf{H}}} \quad & \|\hat{\mathbf{H}}^2 - \mathbf{C}_\mathbf{x}\|_F^2 \\ \text{subject to} \quad & \text{diag}(\hat{\mathbf{H}}^{-1}) = \mathbf{1}, \\ & \hat{\mathbf{H}} = \hat{\mathbf{H}}^\top. \end{aligned} \tag{5}$$

This problem is hard to solve though, so we further tune this into a more manageable form.

Let the eigenvalue decomposition (EVD) of the symmetric matrix variable $\hat{\mathbf{H}}$ be given by $\hat{\mathbf{H}} = \hat{\mathbf{U}}\text{diag}(\hat{\boldsymbol{\lambda}})\hat{\mathbf{U}}^T$, which leads to[2] $\hat{\mathbf{H}}^2 = \hat{\mathbf{U}}\text{diag}(\hat{\boldsymbol{\lambda}}^2)\hat{\mathbf{U}}^\top$. This allows us to replace the matrix variable $\hat{\mathbf{H}}$ by two new variables $\hat{\mathbf{U}}$ and $\hat{\boldsymbol{\lambda}}$. Consider now also the EVD of $\mathbf{C}_\mathbf{x}$, which is given by $\mathbf{C}_\mathbf{x} = \mathbf{U}_\mathbf{x}\text{diag}(\boldsymbol{\lambda}_\mathbf{x})\mathbf{U}_\mathbf{x}^\top$. Then we can simplify problem (5) by setting $\hat{\mathbf{U}} = \mathbf{U}_\mathbf{x}$ and restricting the problem to the single vector variable $\hat{\boldsymbol{\lambda}}$. Problem (5) can then be approximated as

$$\begin{aligned} \boldsymbol{\lambda}^* = \arg\min_{\hat{\boldsymbol{\lambda}}} \quad & \|\hat{\boldsymbol{\lambda}}^2 - \boldsymbol{\lambda}_\mathbf{x}\|_2^2 \\ \text{subject to} \quad & \text{diag}(\mathbf{U}_\mathbf{x}\text{diag}(\hat{\boldsymbol{\lambda}}^{-1})\mathbf{U}_\mathbf{x}^\top) = \mathbf{1}. \end{aligned} \tag{6}$$

Since this constraint is still hard to handle, we further approximate the problem by setting the objective to zero and turning the constraint into an objective.

[1]We adopt the definition of matrix differentiation for structured matrices from [12]. Specifically, for a scalar function $g(\mathbf{A})$, we define $\frac{dg}{dA_{ij}} = \sum_{kl} \frac{\partial g}{\partial A_{kl}} \frac{\partial A_{kl}}{\partial A_{ij}}$.

[2]All powers of vectors should be considered as element-wise.

Setting the objective to zero means that $\hat{\boldsymbol{\lambda}}^2 = \boldsymbol{\lambda}_\mathbf{x}$. This however introduces a sign ambiguity, which we can interpret as our new variable. So introducing $\hat{\mathbf{q}} \in \{-1, 1\}^{N\times 1}$ we can change the variable $\hat{\boldsymbol{\lambda}}$ into the binary variable $\hat{\mathbf{q}}$ by setting $\hat{\boldsymbol{\lambda}} = \text{diag}(\hat{\mathbf{q}})\boldsymbol{\lambda}_\mathbf{x}^{1/2}$, where $(\cdot)^{1/2}$ represents the positive square root.

Now turning the constraint in problem (6) into an objective function, we obtain

$$\begin{aligned} & \|\,\text{diag}(\mathbf{U}_\mathbf{x}\text{diag}(\hat{\boldsymbol{\lambda}}^{-1})\mathbf{U}_\mathbf{x}^\top) - \mathbf{1}\|_2^2 \\ &= \|\text{diag}(\mathbf{U}_\mathbf{x}\text{diag}(\hat{\mathbf{q}}^{-1})\text{diag}(\boldsymbol{\lambda}_\mathbf{x}^{-1/2})\mathbf{U}_\mathbf{x}^\top) - \mathbf{1}\|_2^2 \\ &= \|\text{diag}(\mathbf{U}_\mathbf{x}\text{diag}(\hat{\mathbf{q}})\text{diag}(\boldsymbol{\lambda}_\mathbf{x}^{-1/2})\mathbf{U}_\mathbf{x}^\top) - \mathbf{1}\|_2^2 \\ &= \|(\mathbf{U}_\mathbf{x} \odot \mathbf{U}_\mathbf{x})\text{diag}(\boldsymbol{\lambda}_\mathbf{x}^{-1/2})\hat{\mathbf{q}} - \mathbf{1}\|_2^2, \end{aligned} \tag{7}$$

where $\odot$ is the element-wise (Hadamard) product and where we have used $\hat{\mathbf{q}} = \hat{\mathbf{q}}^{-1}$. Hence, the objective now becomes a simple quadratic function in the binary variables $\hat{\mathbf{q}}$. Defining $\mathbf{W} = (\mathbf{U}_\mathbf{x} \odot \mathbf{U}_\mathbf{x})\text{diag}(\boldsymbol{\lambda}_\mathbf{x}^{-1/2})$ our proposed problem can finally be stated as the following binary least squares problem also known as an unconstrained binary quadratic programming (UBQP) problem:

$$\textbf{(P1)} \qquad \mathbf{q}^* = \arg\min_{\hat{\mathbf{q}}\in\{-1,1\}^{N\times 1}} \|\mathbf{W}\hat{\mathbf{q}} - \mathbf{1}\|_2^2. \tag{8}$$

For this problem, we can state the following identifiabily theorem.

Theorem 1: Let the EVD of the true GSO $\mathbf{S}$ be given by $\mathbf{S} = \mathbf{U}\,\text{diag}(\boldsymbol{\lambda})\mathbf{U}^\top$. Further assume $\hat{\mathbf{p}}$ is a binary variable and consider the equation

$$(\mathbf{U} \odot \mathbf{U})|\mathbf{I} - \text{diag}(\boldsymbol{\lambda})|\hat{\mathbf{p}} - \mathbf{1} = \mathbf{0}. \tag{9}$$

If this equation only has one binary solution $\mathbf{p}^*$, then the estimator $\mathbf{S}^*$, obtained from the solution of problem **(P1)**, i.e., $\mathbf{S}^* = \mathbf{I} - \mathbf{U}_\mathbf{x}\,\text{diag}(\mathbf{q}^*)\,\text{diag}(\boldsymbol{\lambda}_\mathbf{x}^{-1/2})\mathbf{U}_\mathbf{x}^\top$, will converge to the true $\mathbf{S}$ when the number of observations T goes to infinity.

Our proof sketch starts with observing that at $T = \infty$ we have $\mathbf{C}_\mathbf{x} = \mathbf{\Sigma}_\mathbf{x} = (\mathbf{I} - \mathbf{S})^{-2}$. As a result, the EVD of $\mathbf{C}_\mathbf{x}$ then is $\mathbf{C}_\mathbf{x} = \mathbf{U}(\mathbf{I} - \text{diag}(\boldsymbol{\lambda}))^{-2}\mathbf{U}^\top$, and thus $\mathbf{U}_\mathbf{x} = \mathbf{U}$ and $\boldsymbol{\lambda}_\mathbf{x}^{-1/2} = |\mathbf{1} - \boldsymbol{\lambda}|$. Hence, saying that (9) has a unique binary solution $\mathbf{p}^*$ is the same as saying that (8) has a unique solution $\mathbf{q}^*$ at $T = \infty$ and these solutions are then also the same.

Although this theorem may seem evident, it can be considered as a broadening of the theorem mentioned in [13], where $\text{rank}(\mathbf{U}\odot\mathbf{U}) = N - 1$ is required. Under this condition, our theorem holds trivially. However, our theorem has the potential to handle cases where $\text{rank}(\mathbf{U} \odot \mathbf{U}) < N - 1$ and in our experiments, we will verify this.

There are many ways to solve an UBQP. Here we consider the traditional semi-definite relaxation approach and we refer the reader to [14] for more details.

IV. Extension to Polynomial Model

In this section, we extend the SEM approach to the polynomial model. If $\mathbf{e} \sim \mathcal{N}(\mathbf{0}, \mathbf{I})$ and $\mathbf{S} = \mathbf{S}^\top$, then according to (3) we have

$$\mathbf{\Sigma}_\mathbf{x} = h(\mathbf{S})h^\top(\mathbf{S}) = h^2(\mathbf{S}) \tag{10}$$

Similar as in Section III, we will then try to match $\mathbf{C}_\mathbf{x}$ to $h^2(\mathbf{S})$. To do that, we basically follow the same ideas as for the SEM and start by defining the EVD of the matrix variable $\hat{\mathbf{S}}$ as $\hat{\mathbf{S}} = \hat{\mathbf{U}}\,\text{diag}(\hat{\boldsymbol{\lambda}})\hat{\mathbf{U}}^\top$. Further decomposing $\mathbf{C}_\mathbf{x}$ as $\mathbf{C}_\mathbf{x} = \mathbf{U}_\mathbf{x}\,\text{diag}(\boldsymbol{\lambda}_\mathbf{x})\mathbf{U}_\mathbf{x}^\top$, we can set

Authorized licensed use limited to: TU Delft Library. Downloaded on May 09,2025 at 07:29:37 UTC from IEEE Xplore. Restrictions apply.

$\hat{\mathbf{U}} = \mathbf{U_x}$. This allows us to match $h^2(\text{diag}(\hat{\boldsymbol{\lambda}}))$ to $\text{diag}(\boldsymbol{\lambda_x})$ under the hollow constraint $\text{diag}(\mathbf{U_x}\text{diag}(\hat{\boldsymbol{\lambda}})\mathbf{U_x}^\top) = (\mathbf{U_x} \odot \mathbf{U_x})\hat{\boldsymbol{\lambda}} = \mathbf{0}$. Switching again the objective and constraint, we finally obtain the problem

$$\begin{aligned} \boldsymbol{\lambda}^* = \arg\min_{\hat{\boldsymbol{\lambda}}} \quad & \|(\mathbf{U_x} \odot \mathbf{U_x})\hat{\boldsymbol{\lambda}}\|_2^2 \\ \text{subject to} \quad & h^2(\text{diag}(\hat{\boldsymbol{\lambda}})) - \text{diag}(\boldsymbol{\lambda_x}) = \mathbf{0}. \end{aligned} \tag{11}$$

The constraint basically represents a set of scalar polynomial constraints of the form $h^2(\hat{\lambda}_i) - \lambda_{i,\mathbf{x}} = 0$, $i = 1, 2, \ldots, N$, where $\hat{\lambda}_i$ ($\lambda_{i,\mathbf{x}}$) denotes the ith element of $\hat{\boldsymbol{\lambda}}$ ($\boldsymbol{\lambda_x}$). Denoting the roots of the ith scalar polynomial as $\mathcal{C}_i = \{c_i^1, c_i^2, \ldots, c_i^{p_i}\}$ we can replace $h^2(\hat{\lambda}_i) - \lambda_{i,\mathbf{x}} = 0$ by $\hat{\lambda}_i \in \mathcal{C}_i$. Our proposed problem can finally be stated as

$$\begin{aligned} \textbf{(P2)} \qquad \boldsymbol{\lambda}^* = \arg\min_{\hat{\boldsymbol{\lambda}}} \quad & \|(\mathbf{U_x} \odot \mathbf{U_x})\hat{\boldsymbol{\lambda}}\|_2^2 \\ \text{subject to} \quad & \hat{\lambda}_i \in \mathcal{C}_i, \ i = 1, 2, \ldots, N. \end{aligned} \tag{12}$$

Let us explore how **(P2)** specializes to **(P1)**. If we set $h(\mathbf{S}) = (\mathbf{I} - \mathbf{S})^{-1}$ in **(P2)**, then all solutions to the equation $h^2(\text{diag}(\hat{\boldsymbol{\lambda}})) - \text{diag}(\boldsymbol{\lambda_x}) = \mathbf{0}$ can be expressed as $\hat{\boldsymbol{\lambda}} = \text{diag}(\boldsymbol{\lambda_x}^{-1/2})\hat{\mathbf{q}} - \mathbf{1}$, where $\hat{\mathbf{q}} \in \{-1, 1\}^{N \times 1}$. Therefore, the objective of **(P2)** can be rewritten as $(\mathbf{U_x} \odot \mathbf{U_x})(\text{diag}(\boldsymbol{\lambda_x}^{-1/2})\hat{\mathbf{q}} - \mathbf{1})$, which, due to the property $(\mathbf{U_x} \odot \mathbf{U_x})\mathbf{1} = \mathbf{1}$, becomes identical to the problem defined in **(P1)**. Thus, **(P2)** can be viewed as an extension of **(P1)**.

Similar to the semi-definite relaxation approach used for the SEM problem, this problem can also be solved using convex relaxation approaches [15]. Additionally, solvers like Gurobi [16] that support integer programming can also be used to find a global minimum.

V. Extension to General Distribution

In this section, we go back to the regular SEM and explore extending the distribution of the latent variable $\mathbf{e}$. More specifically, consider $\mathbf{e}$ to be normally distributed as $\mathbf{e} \sim \mathcal{N}(\mathbf{0}, \boldsymbol{\Sigma_e})$, where $\boldsymbol{\Sigma_e}$ is known.

Estimating $\mathbf{S}$ then again boils down to estimating $\mathbf{H} = (\mathbf{I} - \mathbf{S})^{-1}$. At first sight, one could exploit the fact that $\boldsymbol{\Sigma_x} = \mathbf{H}\boldsymbol{\Sigma_e}\mathbf{H}^\top$ and match $\hat{\mathbf{H}}\boldsymbol{\Sigma_e}\hat{\mathbf{H}}^\top$ with $\mathbf{C_x}$. Solving this matching problem is challenging though. As an alternative, observe that $(\mathbf{H}\boldsymbol{\Sigma_e})^2 = \mathbf{H}\boldsymbol{\Sigma_e}\mathbf{H}\boldsymbol{\Sigma_e} = \boldsymbol{\Sigma_x}\boldsymbol{\Sigma_e}$. This allows us to match $(\hat{\mathbf{H}}\boldsymbol{\Sigma_e})^2$ with $\mathbf{C_x}\boldsymbol{\Sigma_e}$ which is similar to (5), where we matched $\hat{\mathbf{H}}^2$ with $\mathbf{C_x}$. As a result, we follow again the same steps.

First, we introduce two new variables $\hat{\mathbf{U}}$ and $\hat{\boldsymbol{\lambda}}$ by considering the EVD of $\hat{\mathbf{H}}\boldsymbol{\Sigma_e}$, i.e., $\hat{\mathbf{H}}\boldsymbol{\Sigma_e} = \hat{\mathbf{U}}\,\text{diag}(\hat{\boldsymbol{\lambda}})\hat{\mathbf{U}}^{-1}$. This obviously leads to $(\hat{\mathbf{H}}\boldsymbol{\Sigma_e})^2 = \hat{\mathbf{U}}\,\text{diag}(\hat{\boldsymbol{\lambda}})^2\hat{\mathbf{U}}^{-1}$. Computing the EVD of $\mathbf{C_x}\boldsymbol{\Sigma_e}$, we obtain[3] $\mathbf{C_x}\boldsymbol{\Sigma_e} = \mathbf{U_{xe}}\,\text{diag}(\boldsymbol{\lambda_{xe}})\mathbf{U_{xe}}^{-1}$. Setting now $\hat{\mathbf{U}} = \mathbf{U_{xe}}$ and replacing the matching problem by the constraint $\hat{\boldsymbol{\lambda}}^2 = \boldsymbol{\lambda_{xe}}$ introduces once again a sign ambiguity. More specifically, we can change the variable $\hat{\boldsymbol{\lambda}}$ by the binary variable $\hat{\mathbf{q}} \in \{-1, 1\}^{N \times 1}$ using $\hat{\boldsymbol{\lambda}} = \text{diag}(\hat{\mathbf{q}})\boldsymbol{\lambda_{xe}}^{1/2}$. Overall, this allows us to write $\hat{\mathbf{H}}$ as a function of $\hat{\mathbf{q}}$ through

$$\hat{\mathbf{H}} = \mathbf{U_{xe}}\,\text{diag}(\hat{\mathbf{q}})\text{diag}(\boldsymbol{\lambda_{xe}}^{1/2})\mathbf{U_{xe}}^{-1}\boldsymbol{\Sigma_e}^{-1}. \tag{13}$$

The inverse of $\hat{\mathbf{H}}$ is then given by

$$\hat{\mathbf{H}}^{-1} = \boldsymbol{\Sigma_e}\mathbf{U_{xe}}\text{diag}(\boldsymbol{\lambda_{xe}}^{-1/2})\,\text{diag}(\hat{\mathbf{q}})\mathbf{U_{xe}}^{-1} \tag{14}$$

[3]Note that we use the notation $\mathbf{U}$ primarily to align with the previous notation and it does not imply that $\mathbf{U}$ is unitary

and the diagonal of $\hat{\mathbf{H}}^{-1}$ is

$$\text{diag}(\hat{\mathbf{H}}^{-1}) = [(\boldsymbol{\Sigma_e}\mathbf{U_{xe}}) \odot \mathbf{U_{xe}}^{-\top}]\text{diag}(\boldsymbol{\lambda_{xe}}^{-1/2})\hat{\mathbf{q}}. \tag{15}$$

Finally, defining $\mathbf{W} = [(\boldsymbol{\Sigma_e}\mathbf{U_{xe}}) \odot \mathbf{U_{xe}}^{-\top}]\text{diag}(\boldsymbol{\lambda_{xe}}^{-1/2})$, the optimization problem simplifies to:

$$\textbf{(P3)} \qquad \min_{\hat{\mathbf{q}} \in \{1,1\}^{N \times 1}} \|\mathbf{W}\hat{\mathbf{q}} - \mathbf{1}\|_2^2, \tag{16}$$

which is again an UBQP that can be solved using semi-definite relaxation.

Comparing **(P1)** and **(P3)**, their formulations are almost identical. Furthermore, if we assume $\boldsymbol{\Sigma_e} = \mathbf{I}$, then $\mathbf{U_{xe}}$ in **(P3)** reduces to $\mathbf{U_x}$, and $\mathbf{U_{xe}}^{-\top}$ also reduces to $\mathbf{U_x}$. Moreover, $\boldsymbol{\lambda_{xe}}^{-1/2}$ reduces to $\boldsymbol{\lambda_x}^{-1/2}$. This extension of the SEM problem is quite elegant, as it scarcely alters the structure of the problem, and the complexity of solving the optimization problem remains the same.

VI. Experiments

Here we consider some experiments using simulated and real graphs. For the simulated graphs, we generate 100 realizations. We evaluate the normalized squared error, defined as $\text{NSE}(\mathbf{S}, \mathbf{S}^*) = \|\mathbf{S}^* - \mathbf{S}\|_F^2 / \|\mathbf{S}\|_F^2$, averaged over these 100 graphs across various sample sizes T ranging from 10^2 up to 10^6.

Comparison with other methods:

Here, we compare our approach (refered to as CovMatch) with SpecTemp [13] and with a trivial signal matching approach (SigMatch) [4] based on minimizing $\|\mathbf{X} - \hat{\mathbf{S}}\mathbf{X}\|_F^2$. Due to the sign constraints of SpecTemp, all graphs are assigned positive weights ranging from 0.1 to 2. At the same time, we adopt the simplest assumption that $\boldsymbol{\Sigma_e} = \mathbf{I}$. In the first scenario (labelled as simple), we deliberately generate graphs where $\text{rank}(\mathbf{U} \odot \mathbf{U}) = N - 1$. In the second scenario (labelled as hard), we only generate graphs with $\text{rank}(\mathbf{U} \odot \mathbf{U}) < N - 1$ to check the robustness of our method. In both scenarios, the number of nodes and edges are set to 20. For each scenario, we calculate the average NSE. Note that a singular value less than 5×10^{-4} is considered as a rank loss here.

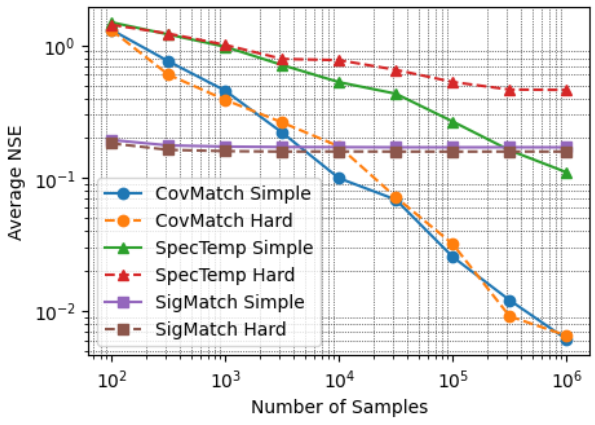
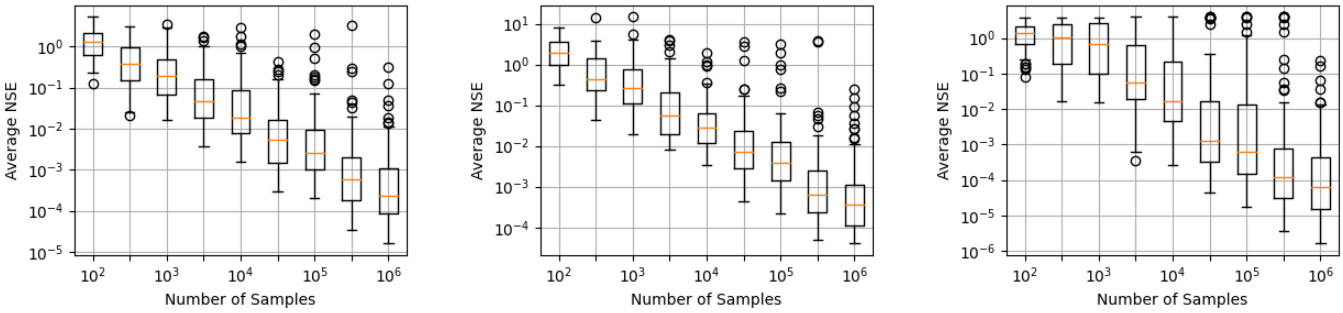


Fig. 1: Average NSE for different samples.

As shown in Fig. 1, it is evident that SpecTemp often fails due to a loss of rank. Conversely, our method, CovMatch, continues to perform well. This highlights the robustness and reliability of CovMatch. Further, the SigMatch approach never converges to the correct result, but it performs better than others with less observations.

Authorized licensed use limited to: TU Delft Library. Downloaded on May 09,2025 at 07:29:37 UTC from IEEE Xplore. Restrictions apply.



(a) SEM sparse graph (b) SEM fully connected graph (c) Polynomial model sparse graph

Fig. 2: Distribution of NSE across different graph configurations and methods. (a) SEM sparse graph, (b) SEM fully connected graph, and (c) polynomial model. As $T \to \infty$, the NSE values are all smaller than 2×10^{-5}.

This is because both SpecTemp and CovMatch highly rely on an accurate sample covariance, which requires many samples.

Besides the advantages mentioned above, our method can also handle negative edge weights and some very challenging graphs. For the next experiment, the edge weights of the 100 graph realizations are chosen within the range $[-2, -0.1] \cup [0.1, 2]$. For the covariance matrix of the latent variables $\mathbf{\Sigma_e}$, we begin by generating a random $N \times N$ matrix, $\mathbf{\Sigma}_{\text{half}}$, with each element uniformly distributed between $[-1, 1]$. The covariance matrix is then formed as $\mathbf{\Sigma_e} = \mathbf{\Sigma}_{\text{half}} \mathbf{\Sigma}_{\text{half}}^{\top}$, ensuring it is symmetric and positive semi-definite. For the SEM, we generate $\mathbf{\Sigma_e}$ according to this method, while for the polynomial model, we still adopt $\mathbf{\Sigma_e} = \mathbf{I}$.

SEM for sparse graphs: In this scenario, we employ a graph with 20 edges and 20 nodes, and intentionally create a challenging condition by generating graphs with a rank of $N - 3$. This condition may cause some methods to fail due to the presence of repeated eigenvalues in $\mathbf{C_x}$ [11]. However, our method remains effective under these constraints.

SEM for fully connected graphs: For the second configuration, we test our method on a fully connected graph and we ignore the rank constraint. Methods that rely on sparsity often fail in this scenario, but our approach continues to perform well.

Polynomial model: Here we consider graphs with 20 nodes and 40 edges while ignoring the rank constraint and we apply the polynomial model. Additionally, we define the polynomial function $h(x)$ as a third-order polynomial $h(x) = \sum_{i=0}^{3} h_i x^i$, where each coefficient h_i is randomly chosen from the interval $[-1, 1]$.

As demonstrated in Fig. 2, our methodology has been notably successful not only on two particularly challenging graph configurations for a SEM but also on nontrivial polynomial models. As the sample size increases, the average NSE generally decreases to significantly low levels. In ideal conditions, with an infinite sample size, our errors can approach zero, showcasing the robustness and effectiveness of our approach. These results also provide indirect confirmation of the correctness of Theorem 1.

Real data: We also compare our approach with network deconvolution (referred to as NetDeconv) [17], which similarly involves estimating $\mathbf{S}$ from $\mathbf{H}$. In this experiment, each node within the network corresponds to an amino acid residue, and the edges denote mutual information, reflecting co-variation among residues across multiple sequence alignments that include 2,000 to 72,000 sequences. Our objective is to deduce structural constraints among amino acid pairs to aid in predicting protein structures.

We employ a relatively straightforward polynomial $h(x) = \frac{1}{80}(x^3 + 2x^2 + 4x)$, whereas NetDeconv approximates $h(x) = \frac{x}{1-x}$. Terms such as "1wvn" shown in Figure 3 represent different protein labels. The figure illustrates that in various cases, our results outperform those of NetDeconv.

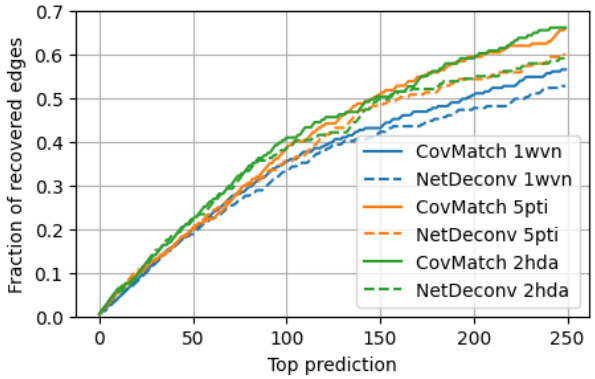


Fig. 3: Real contact edge recovery as a function of the number of edges considered.

VII. Conclusions

In this paper, we have introduced a promising topology identification methodology based on covariance matching, which is fundamentally based on reproducing the theoretical covariance model from the sample covariance matrix. This approach has significant potential due to its foundation in regenerating observable data characteristics. Focusing on the SEM as our primary area of study, we have described how we can simplify the problem into a UBQP which can be solved by semi-definite relaxation. For this UBQP we also provide a convergence proof to the true graph. Furthermore, we assert that our method can be extended to more complex scenarios, such as towards the polynomial model framework, and even to any known distribution of latent variables. We substantiate the efficacy and correctness of our approach through extensive experimental validation, demonstrating its robustness across a variety of settings.

Authorized licensed use limited to: TU Delft Library. Downloaded on May 09,2025 at 07:29:37 UTC from IEEE Xplore. Restrictions apply.

REFERENCES

[1] Jiaying Liu, Feng Xia, Lei Wang, Bo Xu, Xiangjie Kong, Hanghang Tong, and Irwin King, "Shifu2: A network representation learning based model for advisor-advisee relationship mining," *IEEE Transactions on Knowledge and Data Engineering*, vol. 33, no. 4, pp. 1763–1777, 2019.

[2] AR Mclntosh and Francisco Gonzalez-Lima, "Structural equation modeling and its application to network analysis in functional brain imaging," *Human brain mapping*, vol. 2, no. 1-2, pp. 2–22, 1994.

[3] Judea Pearl, "Graphs, causality, and structural equation models," *Sociological Methods & Research*, vol. 27, no. 2, pp. 226–284, 1998.

[4] Alberto Natali, Elvin Isufi, Mario Coutino, and Geert Leus, "Online graph learning from time-varying structural equation models," in *2021 55th Asilomar Conference on Signals, Systems, and Computers*, 2021, pp. 1579–1585.

[5] Juan Andrés Bazerque, Brian Baingana, and Georgios B Giannakis, "Identifiability of sparse structural equation models for directed and cyclic networks," in *2013 IEEE Global Conference on Signal and Information Processing*. IEEE, 2013, pp. 839–842.

[6] J. Peters and P. Bühlmann, "Identifiability of Gaussian structural equation models with equal error variances," *Biometrika*, vol. 101, no. 1, pp. 219–228, 11 2013.

[7] 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*, vol. 31, 2018.

[8] Seyed Saman Saboksayr, Gonzalo Mateos, and Mariano Tepper, "Block successive convex approximation for concomitant linear dag estimation," in *2024 IEEE 13rd Sensor Array and Multichannel Signal Processing Workshop (SAM)*. IEEE, 2024, pp. 1–5.

[9] Gonzalo Mateos, Santiago Segarra, Antonio G Marques, and Alejandro Ribeiro, "Connecting the dots: Identifying network structure via graph signal processing," *IEEE Signal Processing Magazine*, vol. 36, no. 3, pp. 16–43, 2019.

[10] Alberto Natali, Elvin Isufi, Mario Coutino, and Geert Leus, "Learning time-varying graphs from online data," *IEEE Open Journal of Signal Processing*, vol. 3, pp. 212–228, 2022.

[11] Andrei Buciulea, Jiaxi Ying, Antonio G. Marques, and Daniel P. Palomar, "Polynomial graphical lasso: Learning edges from gaussian graph-stationary signals," 2024.

[12] Kaare Brandt Petersen, Michael Syskind Pedersen, et al., "The matrix cookbook," *Technical University of Denmark*, vol. 7, no. 15, pp. 510, 2008.

[13] Santiago Segarra, Antonio G Marques, Gonzalo Mateos, and Alejandro Ribeiro, "Network topology inference from spectral templates," *IEEE Transactions on Signal and Information Processing over Networks*, vol. 3, no. 3, pp. 467–483, 2017.

[14] Zhi-Quan Luo, Wing-Kin Ma, Anthony Man-Cho So, Yinyu Ye, and Shuzhong Zhang, "Semidefinite relaxation of quadratic optimization problems," *IEEE Signal Processing Magazine*, vol. 27, no. 3, pp. 20–34, 2010.

[15] Christoph Buchheim and Angelika Wiegele, "Semidefinite relaxations for non-convex quadratic mixed-integer programming," *Mathematical Programming*, vol. 141, pp. 435–452, 2013.

[16] Gurobi Optimization, LLC, "Gurobi Optimizer Reference Manual," 2024.

[17] Soheil Feizi, Daniel Marbach, Muriel Médard, and Manolis Kellis, "Network deconvolution as a general method to distinguish direct dependencies in networks," *Nature biotechnology*, vol. 31, no. 8, pp. 726–733, 2013.

Authorized licensed use limited to: TU Delft Library. Downloaded on May 09,2025 at 07:29:37 UTC from IEEE Xplore. Restrictions apply.