Statistica Sinica Preprint No: SS-2025-0275	
Title	Efficient Learning of Sparse Differential Networks in Multi-Source Non-Paranormal Models
Manuscript ID	SS-2025-0275
URL	http://www.stat.sinica.edu.tw/statistica/
DOI	10.5705/ss.202025.0275
Complete List of Authors	Mojtaba Nikahd and Seyed Abolfazl Motahari
Corresponding Authors	Seyed Abolfazl Motahari
E-mails	motahari@sharif.edu
Notice: Accepted author version.	

EFFICIENT LEARNING OF SPARSE DIFFERENTIAL NETWORKS IN MULTI-SOURCE NON-PARANORMAL MODELS

Mojtaba Nikahd and Seyed Abolfazl Motahari

Sharif University of Technology

Abstract: This paper introduces an efficient method for learning sparse structural changes—known as the differential network—between two classes of non-paranormal graphical models. We consider the practically important setting where datasets are heterogeneous and originate from multiple sources, yet share a common latent Gaussian covariance structure. Among existing approaches for estimating the differential network, one prominent method is based on minimizing a lasso-penalized D-trace loss function. However, current implementations of this approach suffer from high computational costs and approximation errors. To address these limitations, we propose a solution-path algorithm for the lasso-penalized D-trace problem that builds on piecewise-linear path-following ideas and exploits the structure of the objective to efficiently compute exact solutions over a range of regularization parameters. Our method eliminates approximation error and substantially reduces computational cost. Further, we incorporate a data-integration mechanism to handle heterogeneous sources and derive non-asymptotic sample-complexity bounds that match those for homogeneous

data—demonstrating flexibility with no statistical efficiency loss. On synthetic data, our method significantly outperforms state-of-the-art techniques in both speed and estimation accuracy. Finally, we applied the method to two real-world datasets. In ovarian cancer drug-resistance data, it identified key genetic markers. In breast cancer subtype data, it identified genes that differentiate luminal A and basal-like tumors. Many of these genes are supported by the biomedical literature, demonstrating the practical utility of the method.

Key words and phrases: Differential network analysis, Graphical models, nonparanormal models, precision matrix estimation, structure learning.

1. Introduction

Underlying structures that govern the interactions between the components of a large system can sometimes be modeled by graphical models. Graphical models are used extensively in many scientific fields, including computational biology (Ni et al., 2018), social sciences (Amati et al., 2018), and financial sciences (Giudici, 2018) among others. In many applications, identifying structural changes between two states of a system is crucial. For instance, in computational biology, changes of regulatory mechanisms of normal cells and cancer cells reveal disease mechanisms (West et al., 2012). Therefore, efficient methods for detecting and analyzing structural changes in graphical models have garnered significant attention.

Various methods have been developed for identifying structural changes, a problem commonly known as differential network analysis. A comprehensive review of these methods is provided in (Shojaie, 2020). Gaussian Graphical Models (GGMs) and their extensions, including non-paranormal models which are the focus of this paper, are widely used for modeling dependencies among d random variables (Tavassolipour et al., 2018), particularly in biological networks (Xia and Li, 2019). In these models, structural inference is performed by identifying the sparsity pattern of the precision matrix (the inverse covariance matrix). Leveraging this property, several approaches have been proposed to estimate the differential network between two GGMs. These include (i) a naive approach that independently estimates the precision matrices for each group and computes their difference; (ii) joint estimation of both group structures in a unified framework (Mohan et al., 2014); and (iii) direct estimation of the differential network without requiring estimation of the individual precision matrices (Zhao et al., 2014; Yuan et al., 2017; Zhang et al., 2017; Tang et al., 2020). In this paper, we adopt the direct estimation approach, as it provides greater flexibility by focusing exclusively on the structural differences between the two models, without requiring additional assumptions on the full precision matrices.

Following the direct estimation approach, Zhao et al. (2014) introduced

an estimator for the differential network, shown to be consistent under mild conditions, assuming sparsity only in the differential network. However, this method is computationally expensive and requires an additional step to symmetrize the estimated differential matrix. To overcome these issues, subsequent works utilized D-trace loss functions to develop one-step symmetric estimators with weaker theoretical consistency conditions (Yuan et al., 2017; Wu et al., 2020). To solve this minimization problem, Yuan et al. (2017) developed an iterative algorithm based on the alternating direction method of multipliers (ADMM) (Boyd et al., 2011), with a per-iteration computational complexity of $\mathcal{O}(d^3)$. More recently, Wu et al. (2020) proposed a fused D-trace loss function for differential network estimation, using an iterative method with the same $\mathcal{O}(d^3)$ complexity per iteration. Their approach demonstrated improved numerical performance over Yuan et al. (2017), but remains computationally prohibitive in high-dimensional settings. More recently, Tang et al. (2020) considered the model from Yuan et al. (2017) under a special setting and proposed a new iterative approximation method. Their algorithm reduces the per-iteration complexity to $\mathcal{O}(nd^2)$, where n is the total number of samples across both categories. Nevertheless, existing methods still suffer from high computational costs, limiting their applicability in real-world problems.

A significant limitation of existing differential network estimation methods Yuan et al. (2017); Tang et al. (2020); Wu et al. (2020) is their pronounced sensitivity to the value of regularization parameters, making parameter selection a critical component of their practical use Tang et al. (2020). In practice, researchers typically rely on heuristic search procedures for tuning these parameters, which are computationally intensive and may produce extreme or unstable estimates (e.g., empty, overly dense, or redundant networks). Consequently, this reliance introduces substantial computational overhead and raises scalability challenges.

Beyond computational challenges, effectively utilizing all available datasets remains a key challenge in practical applications. Although integrating datasets from multiple sources is common, particularly in biological studies (Ou-Yang et al., 2021), such data often exhibit substantial heterogeneity due to differences in measurement techniques and experimental conditions (Wahlsten et al., 2003). Consequently, standard GGMs may be inadequate for modeling these data. Non-paranormal graphical models address this limitation by allowing unknown transformations in the data. These transformations preserve a shared underlying dependence structure while adapting to differences across datasets (Liu et al., 2009, 2012; Zhang et al., 2017). This flexibility makes non-paranormal models a promising framework for

differential network analysis.

Related lines of research have considered alternative modeling frameworks for capturing complex dependency structures in high-dimensional data. For example, copula-based methods provide a flexible approach for modeling non-Gaussian dependencies by separating marginal distributions from the dependence structure (Ma et al., 2012; Czado and Nagler, 2022). In addition, several works have explored structured graphical model estimation under assumptions on network topology. In particular, hub-based graphical models impose structured sparsity patterns to capture networks in which a small number of nodes have a large number of connections, a property frequently observed in biological systems (Mohan et al., 2014). These developments highlight the importance of flexible modeling strategies when analyzing high-dimensional dependency structures. The framework considered in this paper complements these approaches by focusing on efficient learning of sparse differential networks while allowing heterogeneous datasets to be integrated under a non-paranormal modeling assumption.

1.1 Our contribution

In this paper, we propose an efficient and practically useful framework for differential network estimation between two non-paranormal graphical

models in high-dimensional settings, where the differential network contains s nonzero entries for some $s \ll d$, without imposing sparsity assumptions on the individual network structures. Instead, sparsity is assumed only for the differential network. The proposed framework builds upon several existing components, including piecewise-linear solution-path methods (Rosset and Zhu, 2007), rank-based covariance estimation (Kruskal, 1958; Kendall, 1948), positive semi-definite correction techniques (Zhao et al., 2014) and the D-trace formulation for differential network estimation (Yuan et al., 2017). Our contribution lies in adapting and integrating these components to develop an efficient computational tool for differential network estimation under heterogeneous non-paranormal models. In the following sections, we elaborate on our contributions across four key dimensions: computational efficiency, statistical guarantees, accuracy and practical applicability. All of these contributions are supported by numerical experiments.

In terms of computational efficiency, we develop a D-trace-specific active-set path-following algorithm that builds upon piecewise-linear solution-path ideas (Rosset and Zhu, 2007). Unlike existing methods (Zhao et al., 2014; Yuan et al., 2017; Zhao et al., 2022) that rely on iterative optimization schemes such as ADMM or gradient-based algorithms, the proposed approach tracks the piecewise-linear regularization path via an active-set

strategy. Consequently, it computes exact solutions along the regularization path, rather than producing approximate solutions at fixed values of the regularization parameter. Each iteration of the proposed algorithm has computational complexity $\mathcal{O}(cd^2)$, where c limits the active set size. Over l iterations, the method yields l distinct solution-path points, resulting in total complexity $\mathcal{O}(lcd^2)$. This is lower than the $\mathcal{O}(lkd^3)$ complexity of ADMM-based methods (Yuan et al., 2017; Wu et al., 2020), and is also advantageous compared to Tang et al. (2020) in high-dimensional regimes where $s = \mathcal{O}(\log d) \ll n \ll d$.

In terms of statistical guarantees, our theoretical analysis builds upon the D-trace framework of Yuan et al. (2017). We consider a heterogeneous non-paranormal setting and employ a rank-based covariance estimator that aggregates information across datasets sharing a common latent covariance structure. We show that this estimator satisfies the conditions required by the D-trace framework, enabling the theoretical results of Yuan et al. (2017) to extend to the heterogeneous setting and yielding non-asymptotic error bounds and support recovery guarantees. The resulting sample complexity remains of the same order as that established in the homogeneous settings of Yuan et al. (2017) and Wu et al. (2020), indicating that accommodating heterogeneous datasets does not require additional sample complexity.

In terms of accuracy, unlike existing methods Yuan et al. (2017); Wu et al. (2020); Tang et al. (2020) that produce only approximate solutions, our approach is theoretically guaranteed to recover the exact solution, thereby eliminating approximation error. Specifically, for each value of the regularization parameter within a prescribed range, the optimization problem admits a unique solution, and our method identifies these solutions precisely. Consequently, the obtained estimator corresponds to the global minimizer of the optimization problem and is not susceptible to errors arising from incomplete convergence of iterative algorithms. We therefore refer to this estimator as the exact solution.

For practical applications, we evaluate the proposed method in two real biological studies: differential network analysis between drug-resistant and drug-sensitive cancer patients, and between two subtypes of breast cancer. In both settings, multiple gene expression datasets are integrated for each group, and the method is applied to identify key genetic differences. The analysis successfully recovers biologically meaningful genes, and the findings are further assessed through enrichment analyses.

The rest of this paper is organized as follows. Section 1.2 introduces the notations. In Section 2, we present our optimization algorithm and provide a theoretical analysis of its convergence and consistency. Section 3.1 reports

numerical results on simulated datasets, while Section 3.2 presents findings from real data experiments. Finally, Section 4 concludes the paper.

1.2 Notations

The set $\{1, \cdots, m\}$ is denoted by $[m]$. For a matrix $A \in \mathbb{R}^{d\times d}$, we denote the sub-matrix of A with row indices in $\mathcal{R} \subseteq [d]$ and column indices in $\mathcal{C} \subseteq [d]$ by $A_{\mathcal{R},\mathcal{C}}$. We refer to the ith row and the jth column of A by $A_{i,:}$ and $A_{:,j}$, respectively. Similarly, for a vector $\mathbf{x}$ and a set of indices $\mathcal{S}$, $\mathbf{x}_{\mathcal{S}}$ denotes a sub-vector of $\mathbf{x}$ whose components are in $\mathcal{S}$. Let $\|A\|_1 = \sum_{i,j=1}^{d} |A_{i,j}|$, $\|A\|_\infty = \max_{i,j} |A_{i,j}|$ and $\|A\|_{1,\infty} = \max_i \sum_j |A_{i,j}|$ denote the element-wise ℓ_1, max norm and $L_{1,\infty}$ of an arbitrary matrix A, respectively. For $A \in \mathbb{R}^{d\times d}$, $\mathrm{vec}\,(A)$ denotes the vectorized form of matrix A which is a $d^2 \times 1$ vector obtained by stacking the columns of A. In addition, the inner product of two matrices is defined as $\langle A, B\rangle = \mathrm{tr}\,(AB^\intercal)$ and $\langle A, A\rangle = \|A\|_F^2$. For a finite set $\mathcal{S}$, the cardinality of $\mathcal{S}$ is denoted by $|\mathcal{S}|$. Table 1 in Section S8 of the Supplementary Material summarizes all key notational symbols and their definitions used throughout this manuscript.

2. Material and Methods

In this section, we propose a method to learn the differential network between two structurally similar non-paranormal graphical models. To provide a clearer foundation for our approach, we first present a brief review of non-paranormal models and define a heterogeneous collection of datasets.

2.1 Non-paranormal graphical models

The non-paranormal graphical model (or the Gaussian copula distribution) is an extension of the GGM described as follows. A d-dimensional random vector $\mathbf{x} = [x_1, x_2, \ldots, x_d]^\intercal$ is said to follow a non-paranormal distribution, denoted by $\mathbf{x} \sim \mathrm{NPN}(\mathbf{0}, \Sigma, \mathbf{f})$, if there exists a set of univariate monotonically increasing functions $\mathbf{f} \triangleq \{f_j : \mathbb{R} \to \mathbb{R}\}_{j=1}^{d}$ such that the transformed random vector $[f_1(x_1), \cdots, f_d(x_d)]^\intercal$ follows a GGM with zero mean and covariance matrix $\Sigma \in \mathbb{R}^{d\times d} \succ 0$. The sparsity pattern of the precision matrix Σ^{-1} strictly encodes the conditional independence structure of the random vector $\mathbf{x}$, where variables x_i and x_j are conditionally independent if and only if $[\Sigma^{-1}]_{i,j} = 0$ (Liu et al., 2009). Consequently, estimating this sparsity pattern directly reveals the underlying graphical structure. Furthermore, the non-paranormal model serves as a reliable and theoretically grounded alternative to standard GGMs. It successfully captures non-Gaussian real-world

data while achieving the same near-optimal sample complexity for graph recovery and parameter estimation (Liu et al., 2012; Xue et al., 2012).

2.2 Heterogeneous collection of datasets

For a given covariance matrix Σ and a set of univariate monotonically increasing functions $\mathbf{f} = \{f_j\}_{j=1}^d$, a homogeneous dataset $\mathcal{D}$ is defined as a set of samples $\mathcal{D} \triangleq \{\mathbf{x}^{(1)}, \mathbf{x}^{(2)}, \ldots, \mathbf{x}^{(m)}\}$ drawn from a non-paranormal distribution with parameters $(\mathbf{0}, \Sigma, \mathbf{f})$. For brevity, whenever we mention a dataset, we are referring to a homogeneous dataset. In our setting, the set of functions $\mathbf{f}$ is unknown, which poses a significant challenge in real-world datasets. Notably, in gene expression data analysis, identifying the appropriate data transformation function is a crucial problem, as discussed by Zwiener et al. (2014). For $N \in \mathbb{N}$, a collection of N homogeneous datasets $\{\mathcal{D}_r\}_{r=1}^N$ is called heterogeneous if each $\mathcal{D}_r$ is generated from a non-paranormal model with the same covariance matrix Σ, but different datasets have distinct sets of functions and varying sample sizes.

2.3 Problem setting

In our problem, we are given two heterogeneous collections of datasets, $\{\mathcal{D}_r\}_{r=1}^N$ and $\{\mathcal{D}'_{r'}\}_{r'=1}^{N'}$, with unit-diagonal covariance matrices $\Sigma, \Sigma' \in$

$\mathbb{R}^{d\times d}$. Our goal is to identify the nonzero entries in the difference of the precision matrices, which encode structural changes between the two models. Specifically, let Δ^* denote the sparse matrix capturing this difference: $\Delta^* = \Sigma^{-1} - (\Sigma')^{-1} \in \mathbb{R}^{d\times d}$. Define the support of Δ^* as $\mathcal{A}^* = \big\{(i,j) : \Delta^*_{i,j} \neq 0\big\}$, with $s = |\mathcal{A}^*|$. We assume that Δ^* is sparse, i.e., $s \ll d$. Our objective is to recover $\mathcal{A}^*$ by leveraging information across all datasets.

2.4 Proposed algorithm

Our proposed algorithm is based on the lasso-penalized D-trace model introduced by Yuan et al. (2017). This method uses estimates of the true covariance matrices, denoted $\widehat{\Sigma}$ and $\widehat{\Sigma}'$. The estimation of the differential precision matrix is formulated as the following optimization problem:

$$\underset{\Delta\in\mathbb{R}^{d\times d}}{\arg\min}\ L_D(\Delta;\widehat{\Sigma},\widehat{\Sigma}') + \lambda\|\Delta\|_1, \tag{2.1}$$

where $\lambda > 0$ is a tuning parameter and $L_D(\Delta;\widehat{\Sigma},\widehat{\Sigma}')$ represents the D-trace loss function, defined as

$$L_D(\Delta;\widehat{\Sigma},\widehat{\Sigma}') \triangleq \frac{1}{4}\Big(\langle\widehat{\Sigma}\Delta,\Delta\widehat{\Sigma}'\rangle + \langle\widehat{\Sigma}'\Delta,\Delta\widehat{\Sigma}\rangle\Big) - \langle\Delta,\widehat{\Sigma}-\widehat{\Sigma}'\rangle. \tag{2.2}$$

In our approach, we adopt the same objective function but replace the covariance estimators with a more suitable alternative, which will be described in Section 2.5. By considering the subgradient optimality conditions

of problem (2.1), it can be shown that the problem admits a unique minimizer for all λ over a specific interval. We refer to the collection of these solutions as the solution path, viewed as a function of the tuning parameter λ and denoted by $\widehat{\Delta}(\lambda)$. Moreover, this solution path is piecewise linear and continuous. We develop an algorithm to characterize and efficiently compute the entire solution path. The formal statement of these properties is presented in the following theorem.

Theorem 1. *Suppose that $\widehat{\Sigma}$ and $\widehat{\Sigma}'$ are positive semi-definite. Then there exists a strictly decreasing sequence of regularization parameters $\infty = \lambda_0 > \lambda_1 > \lambda_2 > \cdots > \lambda_T \geq 0$ such that, for each $t \in \{0, 1, \ldots, T-1\}$ and any $\lambda \in [\lambda_{t+1}, \lambda_t]$, the optimization problem (2.1) admits a unique global minimizer at λ. This minimzer, denoted by $\widehat{\Delta}(\lambda)$, satisfies*

$$\begin{cases} \operatorname{vec}\left(\widehat{\Delta}(\lambda)\right)_{\mathcal{A}_t} = -\left(\Gamma_{\mathcal{A}_t,\mathcal{A}_t}\right)^{-1}\left(\mathbf{v}_{\mathcal{A}_t} + \lambda \mathbf{s}_t\right), \\ \operatorname{vec}\left(\widehat{\Delta}(\lambda)\right)_{\mathcal{A}_t^c} = 0. \end{cases} \tag{2.3}$$

The associated sequence of segment tuples $(\lambda_t, \mathcal{A}_t, \mathbf{s}_t)_{t=0}^{T}$ is obtained by applying Algorithm 1 to $\widehat{\Sigma}$ and $\widehat{\Sigma}'$. Here,

$$\Gamma = \frac{1}{2}\left\{\widehat{\Sigma} \otimes \widehat{\Sigma}' + \widehat{\Sigma}' \otimes \widehat{\Sigma}\right\}, \qquad \mathbf{v} = \operatorname{vec}\left(\widehat{\Sigma} - \widehat{\Sigma}'\right).$$

Remark 1. It is noteworthy that the Lasso-penalized D-trace loss is invariant under matrix transposition. Together with the uniqueness of the

solution for $\lambda \in (\lambda_T, \infty)$, this property implies that $\widehat{\Delta}(\lambda)$ is symmetric for all such values of λ.

Full details of the proof are provided in Section S3 of the Supplementary Material. For completeness, we briefly outline the main ideas here. The proof of Theorem 1 consists of three parts. First, the continuity of the solution path $\widehat{\Delta}(\lambda)$ is established, which follows directly from Theorem 5.2 of Still (2018). Second, the solution path is shown to be piecewise linear. This property is derived from the subgradient optimality conditions, which imply that the solution varies linearly within regions where the sparsity pattern is known and remains unchanged. Under this condition, the subgradient optimality conditions reduce to a linear system in λ. Finally, it is shown that Algorithm 1 correctly identifies the sequence of segment tuples $(\lambda_t, \mathcal{A}_t, \mathbf{s}_t)_{t=0}^{T}$, which fully characterizes the solution path over the interval (λ_T, ∞). This part of the proof proceeds by mathematical induction. At each iteration, the algorithm evaluates all potential changes in the sparsity pattern and determines the next segment tuple to ensure that optimality is maintained. By iteratively applying this procedure, the algorithm ensures that the subgradient optimality conditions are satisfied along the entire piecewise linear solution path.

Theorem 1 establishes that the estimator $\widehat{\Delta}(\lambda)$ follows a continuous

piecewise linear solution path. This property enables efficient characterization of the entire path through a sequence of segment tuples $(\lambda_t, \mathcal{A}_t, \mathbf{s}_t)_{t=0}^{T}$. Each tuple describes a linear segment of the path, where λ_t denotes a knot point marking a transition between adjacent segments, $\mathcal{A}_t$ denotes the active set (the nonzero entries within that segment), and $\mathbf{s}_t$ represents the corresponding sign pattern. Within each segment the sparsity structure of the estimated differential network remains unchanged, while transitions between segments correspond to changes in the active set. Consequently, identifying a new knot along the solution path is equivalent to identifying a new sparsity pattern underlying the inferred differential network.

Algorithm 1 sequentially identifies the segment tuples, thereby tracing the complete solution path. Once the sequence $(\lambda_t, \mathcal{A}_t, \mathbf{s}_t)_{t=0}^{T}$ is obtained, the unique minimizer of problem (2.1) can be computed exactly for any $\lambda \in (\lambda_T, \infty)$ using Equation (2.3) in Theorem 1, rather than approximated through iterative methods. Throughout this work we refer to such solutions as exact solutions. The algorithm incorporates termination criteria to ensure solution uniqueness and computational feasibility. Detailed derivations are provided in Section S3 of the Supplementary Material.

Because variables enter or leave the active set sequentially along the piecewise linear path, the active set size can be monitored during com-

putation. In practice, we impose an upper bound on the active-set size, controlled by parameter c, to limit the maximum explored sparsity level and thereby enhance computational efficiency in high-dimensional settings. The per-iteration computational complexity of the algorithm is $\mathcal{O}(cd^2 + c^3)$. Importantly, this upper bound does not impose a restrictive constraint, and the algorithm retains a resumable property. Further details including its termination criteria, the role of the active-set bound, sensitivity analyses with respect to this bound, computational complexity analysis, and the algorithm's resumable feature, are presented in Section S1 of the Supplementary Material.

2.5 Covariance matrix estimation

In the heterogeneous non-paranormal setting, each dataset may undergo unknown marginal transformations, which prevents direct estimation of the underlying Gaussian covariance matrix. A well-known property of non-paranormal distributions provides a way around this difficulty (Kruskal, 1958; Kendall, 1948): for any pair of variables k, l, the Pearson correlation and Kendall's tau correlation are linked by

$$\Sigma_{k,l} = \sin\left(\tfrac{\pi}{2}\tau_{k,l}\right). \tag{2.4}$$

Algorithm 1 Exact Solution Path with Active-Set Control

Require: Active-set bound $c > 0$; estimated covariance matrices $\widehat{\Sigma} \succeq 0$ and $\widehat{\Sigma}' \succeq 0$.

1: Initialize the iteration counter $t = 0$, the regularization parameter $\lambda_0 = \infty$, the active set $\mathcal{A} = \emptyset$, and the corresponding active signs $\mathbf{s}_{\mathcal{A}} = \emptyset$.

2: **while** $|\mathcal{A}| \leq c$ & $\Gamma_{\mathcal{A},\mathcal{A}}$ is invertible **do**

3: Compute the matrices $\Gamma_{:,\mathcal{A}}$, $\left(\Gamma_{\mathcal{A},\mathcal{A}}\right)^{-1}\mathbf{v}_{\mathcal{A}}$ and $\left(\Gamma_{\mathcal{A},\mathcal{A}}\right)^{-1}\mathbf{s}_{\mathcal{A}}$.

4: Identify the candidate knot associated with entry event:

$$\lambda^{\text{entry}} = \max_{\substack{u \in \mathcal{A}^c \\ s_u \in \{-1,1\}}} h(u, s_u) \cdot \mathbb{1}_{\{h(u,s_u) < \lambda_t\}},$$

where $h(u, s_u) = \frac{\Gamma_{u,\mathcal{A}}\left(\Gamma_{\mathcal{A},\mathcal{A}}\right)^{-1}\mathbf{v}_{\mathcal{A}} - V_u}{s_u - \Gamma_{u,\mathcal{A}}\left(\Gamma_{\mathcal{A},\mathcal{A}}\right)^{-1}\mathbf{s}_{\mathcal{A}}}$.

5: Identify the candidate knot associated with exit event:

$$\lambda^{\text{exit}} = \arg\max_{u \in \mathcal{A}} g(u) \cdot \mathbb{1}_{\{g(u) < \lambda_t\}},$$

where $g(u) = \frac{-\left[\left(\Gamma_{\mathcal{A},\mathcal{A}}\right)^{-1}\mathbf{v}_{\mathcal{A}}\right]_u}{\left[\left(\Gamma_{\mathcal{A},\mathcal{A}}\right)^{-1}\mathbf{s}_{\mathcal{A}}\right]_u}$.

6: Determine the next knot λ_{t+1} in the solution path by setting: $\lambda_{t+1} = \max\left\{\lambda^{\text{entry}}, \lambda^{\text{exit}}\right\}$.

7: **if** $\lambda^{\text{entry}} > \lambda^{\text{exit}}$ **then**

8: Apply the entry event: add all variables $u \in \mathcal{A}^c$ and their corresponding signs $s_u \in \{-1, 1\}$ that attain $\lambda^{\text{entry}} = h(u, s_u)$ to the active set $\mathcal{A}$ and the sign vector $\mathbf{s}_{\mathcal{A}}$, respectively.

9: **else**

10: Apply the exit event: remove all variables $u \in \mathcal{A}$ and their corresponding signs s_u that attain $\lambda^{\text{exit}} = g(u)$ from the active set $\mathcal{A}$ and the sign vector $\mathbf{s}_{\mathcal{A}}$, respectively.

11: **end if**

12: $\mathcal{A}_{t+1} = \mathcal{A}$ and $\mathbf{s}_{t+1} = \mathbf{s}_{\mathcal{A}}$.

13: Increment the iteration counter: $t = t + 1$.

14: **end while**

Ensure: Return the sequence of segment tuples $(\lambda_t, \mathcal{A}_t, \mathbf{s}_t)$.

Without loss of generality, we assume that the latent covariance matrices have unit diagonal. This assumption is equivalent to working with the corresponding Pearson correlation matrices. Under this normalization, the identity above permits consistent covariance estimation through rank-based methods without requiring knowledge of the marginal transformations.

Building on this property, we estimate Kendall's tau separately within each dataset and then aggregate these estimates across all heterogeneous datasets using sample-size weights. Applying the sine transformation to the aggregated Kendall's tau values yields an initial rank-based correlation estimator. However, this naive estimator is not guaranteed to be positive semi-definite (PSD), which is required to ensure the convexity of the D-trace loss and to obtain the exact global solution path described in Theorem 1.

To address this issue, we project the initial estimator onto the PSD cone using the method of Zhao et al. (2014), which produces a PSD matrix that is arbitrarily close (up to tolerance μ) to the original rank-based estimate. We denote the resulting covariance estimator by $\widehat{\Sigma}(\mu)$. The details of constructing the covariance matrix estimator is provided in Section S2 of the Supplementary Material. In the following, we formally present a concentration bound for the estimator $\widehat{\Sigma}(\mu)$ in lemma 1, with the proof provided in Section S4 of the Supplementary Material.

Lemma 1. *Let $\mu > 0$ and $\delta > 0$ be arbitrary positive constants. The estimator $\widehat{\Sigma}(\mu)$ satisfies the following concentration inequality:*

$$\mathbb{P}\left(\left\|\widehat{\Sigma}(\mu) - \Sigma\right\|_{\infty} \geq \delta\right) \leq d^2 \exp\left(\frac{-m\left(\max\left\{2\delta - \mu, 0\right\}\right)^2}{32\pi^2}\right).$$

In particular, when $\mu = \delta$, the bound simplifies to:

$$\mathbb{P}\left(\left\|\widehat{\Sigma}(\mu) - \Sigma\right\|_{\infty} \geq \mu\right) \leq d^2 \exp\left(\frac{-m\mu^2}{32\pi^2}\right).$$

2.6 Sample complexity analysis

In Section 2.5, we introduced a covariance matrix estimator. Applying this estimator to the datasets from the two groups yields the covariance matrix estimates $\widehat{\Sigma}(\mu)$ and $\widehat{\Sigma}'(\mu)$. Next, Algorithm 1 is applied to these matrices to compute the segment tuples. Finally, using Equation (2.3), we obtain the estimate of the differential network at a given value of the regularization parameter λ. This procedure constitutes our proposed method for differential network estimation under heterogeneous non-paranormal sampling.

In this section, we study the statistical properties of the proposed differential network estimator. Our analysis extends the framework of Yuan et al. (2017) along two major directions: we allow each group to contain multiple heterogeneous datasets, and we work under a non-paranormal graphical model rather than assuming sub-Gaussianity. To this end, we meticulously

design the sample covariance estimator to satisfy two essential properties. First, it ensures that the estimator is PSD, which is necessary for guaranteeing exact solution identification. Second, we establish a suitable concentration bound for this estimator, which is required to achieve the desired sample complexity rate in differential network analysis. Building on these ingredients, we adapt the proofs of Yuan et al. (2017) to our problem setting and obtain comparable non-asymptotic error bounds and support recovery guarantees, with the same order of sample complexity. To state these results, we introduce the following definitions:

$$\Gamma^* = \frac{1}{2}\left\{\Sigma \otimes \Sigma' + \Sigma' \otimes \Sigma\right\}, \alpha = 1 - \max_{e \in \mathcal{A}^{*c}} \left\|\Gamma^*_{e,\mathbf{A}^*}\left(\Gamma^*_{\mathcal{A}^*,\mathcal{A}^*}\right)^{-1}\right\|_1, A = \frac{\alpha}{4-\alpha},$$

$$\kappa_\Gamma = \left\|\left(\Gamma^*_{\mathcal{A}^*,\mathcal{A}^*}\right)^{-1}\right\|_{1,\infty}, \bar{\delta} = \min\left\{\sqrt{1+(6s\kappa_\Gamma)^{-1}} - 1,\ \sqrt{2+\tilde{M}^{-1}} - 1,\ A\right\},$$

$$\tilde{M} = \frac{24s\left(2s\kappa_\Gamma^2 + \kappa_\Gamma\right)}{\alpha}, M\left(\Delta\right) = \left\{\text{sign}\left(\Delta_{j,k}\right) : j,k = 1,\ldots,d\right\}.$$

Theorem 2. *Let $\mu \in \left(0, 2\bar{\delta}\right)$ and suppose the irrepresentability condition holds; that is,*

$$\max_{e \in \mathcal{A}^{*c}} \left\|\Gamma^*_{e,\mathbf{A}^*}\left(\Gamma^*_{\mathcal{A}^*,\mathcal{A}^*}\right)^{-1}\right\|_1 < 1.$$

Assume further that

$$\min\left\{m, m'\right\} > \frac{32\pi^2\eta\log d}{\left(2\bar{\delta} - \mu\right)^2}$$

and let the regularization parameter λ *be chosen as*

$$\lambda = \max\left\{\frac{2(\mu + G_A)}{A},\ \tilde{M}\left((G_B + (1+\frac{\mu}{2})(G_A+\mu) + \frac{\mu}{4}\right)\right\}$$

for some $\eta > 2$, *where* $\tilde{\delta} = \sqrt{8\pi^2\eta \log d}$, $G_A = \frac{\tilde{\delta}}{\sqrt{m}} + \frac{\tilde{\delta}}{\sqrt{m'}}$, *and* $G_B = \frac{\tilde{\delta}^2}{\sqrt{mm'}}$. *Then, with probability at least* $1 - 2/d^{\eta-2}$, *the support of* $\widehat{\Delta}(\lambda)$ *lies in the support of* Δ^*, *and the following error bound holds:*

$$\left\|\widehat{\Delta}(\lambda) - \Delta^*\right\|_F \le \sqrt{s}\left\|\widehat{\Delta}(\lambda) - \Delta^*\right\|_\infty \le C\left(\sqrt{\frac{\eta \log d}{\min\{m, m'\}}} + \frac{\mu}{4\pi\sqrt{2}}\right),$$

with

$$\begin{aligned} C =& 4\pi\sqrt{2s}\left(\left\{\kappa_\Gamma + 3s\kappa_\Gamma^2\left(A^2 + 2A\right)\right\}\left(1 + 2\bar{\lambda}\right) + 3s\kappa_\Gamma^2\left(A + 2\right)\right), \\ A =& \frac{\alpha}{4-\alpha}, \\ \bar{\lambda} =& \max\left\{\frac{1}{A}, \frac{6s\left(A+2\right)\left(2s\kappa_\Gamma^2 + \kappa_\Gamma\right)}{\alpha}\right\}. \end{aligned}$$

The proof is provided in Section S5 of the Supplementary Material.

Theorem 3. *Under the condition and notation in Theorem 2, if*

$$\min_{j,k:\Delta^*_{j,k}\neq 0}\left|\Delta^*_{j,k}\right| \ge 2C\left(\sqrt{\frac{\eta \log d}{\min\{m, m'\}}} + \frac{\mu}{4\pi\sqrt{2}}\right)$$

for some $\eta > 2$, *then* $M\left(\widehat{\Delta}\right) = M\left(\Delta^*\right)$ *with probability* $1 - 2/d^{\eta-2}$.

The proof is provided in Section S6 of the Supplementary Material.

3. Results

In this section, we present experimental results that demonstrate the efficacy of our proposed approach in terms of both speed and accuracy when applied to synthetic data, as well as its applicability in inferring differential structures from real-world data. All materials necessary to reproduce the experimental results, including code and data, are publicly available at `https://github.com/mojtabanikahd/SPD`.

3.1 Synthetic data

We conducted three sets of simulation studies to evaluate (i) overall performance relative to existing approaches, (ii) the exactness of optimization, and (iii) robustness in heterogeneous-data scenarios. Full experimental design, network-generation details, and evaluation metrics are provided in Section S9 of the Supplementary Material.

3.1.1 Experiment 1: performance comparison

In the first experiment, we evaluate the performance of the proposed method, solution path D-trace, in comparison with three representative approaches: (i) APGD D-trace (Yuan et al., 2017), which applies accelerated proximal gradient descent to obtain an approximate solution to the ℓ_1-penalized

D-trace loss minimization problem; (ii) CrossFDTL (Wu et al., 2020), which estimates the differential network via an alternative optimization formulation that directly targets structural changes; and (iii) the fused graphical lasso (FGL; Danaher et al., 2014), which jointly estimates two precision matrices using a fused lasso penalty.

Figure 1 illustrates a comparative analysis of the four methods in terms of precision-recall performance and computational efficiency. The precision-recall curves indicate that the proposed method consistently outperforms the APGD D-trace, CrossFDTL and FGL methods in terms of accuracy. The performance gap arises because the proposed method recovers the full sparse solution path exactly, whereas APGD D-trace misses many high-precision solutions due to incomplete path tracing and an insufficient number of iterations. Notably, both methods coincide at the λ values whose the corresponding APGD D-trace solutions are obtained using 50 iterations. In addition, the proposed method is much more computationally efficient, as it considers more tuning parameter values with higher accuracy and uses less computation time than all competing methods.

To assess robustness with respect to the structure of differential changes, we also conducted experiments in which structural differences were generated by perturbing a single node. The results of this analysis are presented

in Section S9.3 of the Supplementary Material and show similar performance patterns. Furthermore, additional details of this experiment are provided in Section S9.2 of the Supplementary Material.

3.1.2 Experiment 2: exact vs. approximate solutions

The second experiment examines whether the proposed method produces exact solutions, meaning solutions satisfying optimality conditions up to machine precision. It also assess the impact of approximation errors across values of the parameter λ. To this end, we evaluate four performance metrics: the objective gap, the Karush-Kuhn-Tucker (KKT) residual, the false selection rate, and the minimum computation time required to obtain a solution without false selections. Formal definitions of these metrics are provided in Section S9.4 of the Supplementary Material. In this experiment, we compare the proposed method with APGD solutions obtained after different numbers of iterations. The experimental setup follows that of the previous experiment, with two modifications: (1) to eliminate covariance estimation error, the true covariance matrices are provided as input to all methods, and (2) λ values are sampled from the interval $[0.001, 0.4]$.

Figure 2 depicts the results of this experiment. The left panel shows that the proposed method produces negligible KKT residuals for all exam-

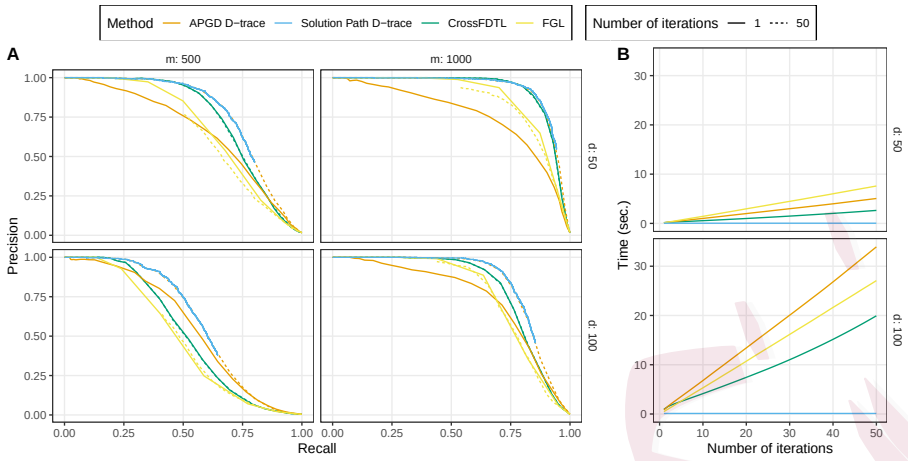

Figure 1: Performance comparison of four differential network estimation methods on synthetic data. Line color denotes the method. Data are generated from networks with $d \in \{50, 100\}$ nodes and $m = m' \in \{500, 1000\}$ samples per network. Results are averaged over 100 datasets. (**A**) Precision-recall curves for detecting differential edges, where line type indicates the iteration count for iterative methods. Our proposed method is shown with solid lines only. (**B**) Computational time for estimating the differential network across all regularization parameters. The x-axis shows the iteration counts of iterative methods and the y-axis shows runtime. The horizontal blue line indicates the average runtime of the proposed method.

ined values of λ, confirming that it consistently identifies exact solutions. In contrast, APGD exhibits considerable residuals that increase as λ decreases

and improve only with a greater number of iterations. The middle panel shows that the objective values produced by the APGD method converge toward the exact objective as the number of iterations increases. Consistent with the behavior of the KKT residuals, smaller values of λ require more iterations for APGD to attain a comparable objective gap. The right panel shows that approximation can lead to substantial errors. In particular, false selections—both false positives and false negatives—increase sharply as the regularization parameter decreases. These errors become smaller when more iterations are used, highlighting the trade-off between computation time and solution accuracy in iterative methods. To further compare computational resource usage, we record memory consumption and computation time for both methods. The corresponding results are reported in Section S9.5 of the Supplementary Material.

3.1.3 Experiment 3: heterogeneous data integration

The third experiment evaluates the effectiveness of the proposed method when applied to heterogeneous datasets. We compare its performance in this setting with two extreme scenarios: (i) the Single Dataset case, where the method is applied to one homogeneous dataset to illustrate the value of integrating heterogeneous data, and (ii) the Homogeneous Dataset case,

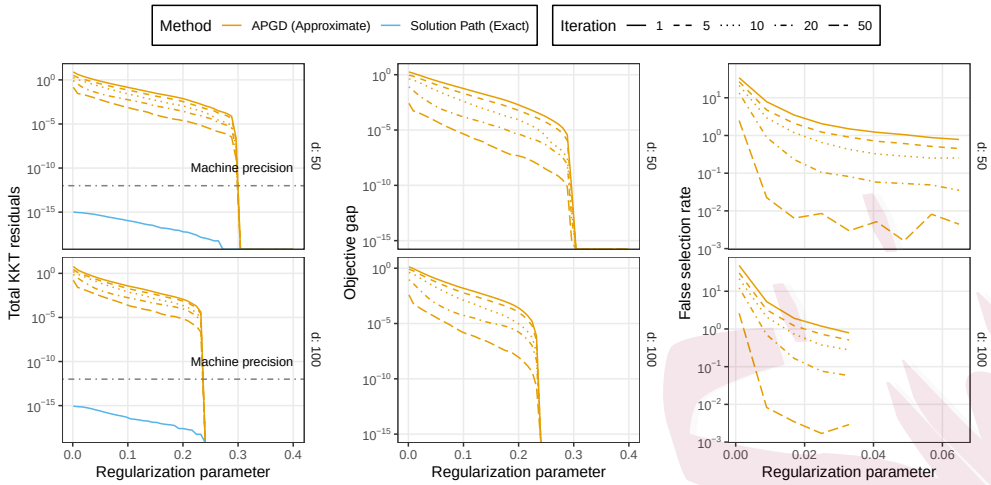

Figure 2: The panels compare our proposed solution path method (exact solver) with the APGD D-trace method evaluated at different iteration numbers, across a range of regularization parameters λ and problem dimensions $d \in \{50, 100\}$. The left panel reports the total KKT residuals, with the horizontal reference line indicating machine precision. The middle panel shows the objective gap between the APGD solutions and the exact solution. The right panel displays the false selection rate associated with approximation error. All vertical axes are shown on a logarithmic scale.

where a single homogeneous dataset with the same total sample size is used to assess the method's ability to exploit heterogeneity. More specifically, in the heterogeneous scenario, each group contains four distinct datasets,

each with 200 samples. In contrast, the Single Dataset case uses only one 200-sample dataset per group, and the Homogeneous Dataset case uses one 800-sample dataset per group. We apply the proposed method in all three settings and record its performance.

Figure 3 shows precision–recall curves averaged over 100 repetitions. The Heterogeneous Datasets setting yields an AUPRC of 0.867, compared with 0.871 for the Homogeneous Dataset setting and 0.419 for the Single Dataset setting, where AUPRC refer to area under the precision-recall curve. These results reveal a substantial loss in performance when only one dataset is used, while the heterogeneous setting achieves accuracy comparable to that of the homogeneous case. This indicates that the proposed method effectively leverages information across heterogeneous datasets.

3.2 Applications to real data

We further applied the proposed method to two large-scale gene-expression studies from The Cancer Genome Atlas (TCGA) (Network et al., 2011). These analyses demonstrate the method's ability to uncover biologically relevant structural changes between conditions. Complete workflow descriptions, parameter choices, and enrichment analyses appear in Section S10.1 of the Supplementary Material.

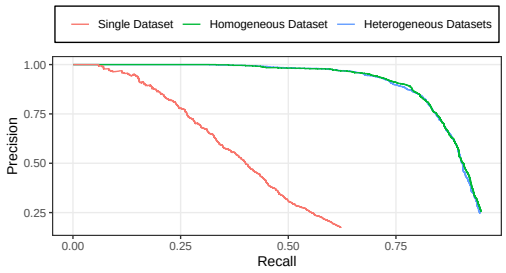

Figure 3: Precision-recall curves for evaluating the proposed method under three experimental scenarios: Single Dataset (one dataset per group, 200 samples), Homogeneous Dataset (one dataset per group, 800 samples), and Heterogeneous Datasets (four datasets per group, 200 samples in each dataset). Each curve is averaged over 100 repetitions, with line color indicating the scenario.

3.2.1 Ovarian cancer

We analyze gene expression data from ovarian cancer cases available in The Cancer Genome Atlas (TCGA) (Network et al., 2011), comprising two groups categorized by drug response according to the criterion proposed by Nabavi et al. (2016). Specifically, we select 242 platinum-sensitive and 98 platinum-resistant tumors. Genes expression for each case was measured using three different platforms, resulting in three datasets per group. To enhance computational efficiency and facilitate model interpretability, we restrict our analysis to 551 genes associated with relevant biological path-

ways, as proposed by Dilruba and Kalayda (2016) and Burris (2013).

Our method inferred a differential gene network comprising 77 edges among 127 genes. Notably, we identified several high-degree genes in the network, including CDKN1A, MAP2K1, MNAT1, and PRKCA, for which strong literature support exists. Detailed references and gene-specific evidence are provided in Section S10.2 of the Supplementary Material.

Pathway and regulon enrichment analyses highlight Gene-expression and transcription-related pathways, with significant involvement of the FOXM1 regulon. FOXM1 is a well-established transcription factor involved in cell cycle regulation, and recent studies have implicated it in cancer drug sensitivity (Koo et al., 2012). This result support the biological plausibility of the detected structural differences.

To further assess the biological relevance of our results, we compared the inferred differential network with a reference set of 912 genes previously linked to platinum resistance in cancer (Huang et al., 2021). We divided the genes in the identified differential network into two groups: (1) *non-isolated genes*, which participate in at least one interaction, and (2) *connector genes*, which have more than one interaction (degree greater than one). Using Fisher's exact test, we found that both groups were significantly enriched for platinum resistance–associated genes, with p-values of 0.01 for non-

isolated genes and 0.003 for connector genes. These results indicate that the genes identified by our method are strongly associated with known platinum resistance markers, providing additional support for the biological validity of the inferred differential network.

Details of these analyses, including the inferred differential network, the complete enrichment procedure and corresponding statistics, are provided in Section S10.2 of the Supplementary Material.

3.2.2 Breast cancer

Breast cancer remains one of the leading causes of cancer-related mortality among women worldwide and comprises four principal molecular subtypes: luminal A, luminal B, basal-like, and HER2-enriched (Network et al., 2012). These subtypes exhibit distinct genomic characteristics (Schnitt, 2010). In this study, we investigated differences in gene network between luminal A and basal-like molecular subtypes using TCGA Breast Invasive Carcinoma data (Network et al., 2012) by applying our proposed method. For each cancer subtypes, there is two datasets with names RNASeqGene and mRNAArray. Specifically, the RNA-seq dataset contains 247 luminal A samples and 99 basal-like samples, while the microarray dataset contains 264 luminal A samples and 106 basal-like samples. In this study, we restricted the

analysis to 145 genes from the breast cancer pathway (hsa05224) obtained from the Kyoto Encyclopedia of Genes and Genomes (KEGG) (Kanehisa and Goto, 2000).

We applied the proposed method to the breast cancer datasets following the same procedure described in Section S10.1 of the Supplementary Material. The analysis identified a differential network consisting of 17 genes connected by 14 edges. Notably, five of the six key genes previously implicated in subtype differentiation Tu et al. (2021) (including IGF1R, CDK6, ESR1, CCND1, and RB1) were recovered in our results.

To further evaluate biological relevance, we conducted over-representation analysis using the MSigDB C2 curated gene sets (Liberzon et al., 2011). The C2 collection comprises manually curated gene sets derived from pathway databases, published studies, and domain experts, and is widely used for hypothesis-driven functional interpretation of gene lists. Over-representation analysis revealed significant enrichment for the SMID_BREAST_CANCER_BASAL_UP signature, which consists of genes upregulated in basal-like breast cancer. Given that our analysis contrasts luminal A and basal-like tumors, the enrichment of a basal-upregulated signature provides independent biological validation that the inferred differential network captures meaningful subtype-specific molecular differences. Additional methodological details

and full network visualizations are presented in Section S10.3 of the Supplementary Material.

4. Conclusions

This paper studies sparse differential network estimation between two classes of non-paranormal models using multiple heterogeneous datasets, where each dataset follows a distinct distribution but shares a common covariance structure within each class. We formulate the problem via the lasso-penalized D-trace loss function and introduce a covariance estimation procedure that integrates information across heterogeneous sources. We show that the resulting solution path is piecewise linear and continuous with respect to the regularization parameter. Furthermore, we establish that the proposed method achieves the same order of sample complexity as the approaches of Yuan et al. (2017) and Wu et al. (2020), while operating in a distinct statistical setting.

Leveraging these properties, we develop an efficient method for computing the exact solution path, thereby improving both computational efficiency and differential network detection accuracy. Extensive experiments on synthetic data demonstrate these advantages. We further validate the method on real gene expression datasets. In ovarian cancer, it identifies

genes associated with drug resistance, and in breast cancer, it highlights genes that differentiate the luminal A and basal-like subtypes. Many of these findings are supported by independent biological evidence. The biological relevance of the results is additionally confirmed through systematic evaluations such as over-representation analysis.

Our framework also opens several promising directions for future research. In particular, an important next step is to develop extensions of the proposed methodology that provide rigorous false discovery rate (FDR) control when identifying differential edges. Incorporating principled FDR-controlling procedures would enhance the method's reliability in high-dimensional settings and offer stronger guarantees for downstream scientific interpretation.

Supplementary Materials

Extended experimental evaluations and rigorous proofs for all theoretical results are detailed in the online supplement.

References

Amati, V., A. Lomi, and A. Mira (2018). Social network modeling. *Annual Review of Statistics and Its Application 5*, 343–369.

REFERENCES

Boyd, S., N. Parikh, and E. Chu (2011). *Distributed optimization and statistical learning via the alternating direction method of multipliers*. Now Publishers Inc.

Burris, H. A. (2013). Overcoming acquired resistance to anticancer therapy: focus on the pi3k/akt/mtor pathway. *Cancer chemotherapy and pharmacology 71*(4), 829–842.

Czado, C. and T. Nagler (2022). Vine copula based modeling. *Annual Review of Statistics and Its Application 9*, 453–477.

Danaher, P., P. Wang, and D. M. Witten (2014). The joint graphical lasso for inverse covariance estimation across multiple classes. *Journal of the Royal Statistical Society. Series B, Statistical methodology 76*(2), 373.

Dilruba, S. and G. V. Kalayda (2016). Platinum-based drugs: past, present and future. *Cancer chemotherapy and pharmacology 77*(6), 1103–1124.

Giudici, P. (2018). Financial data science. *Statistics & Probability Letters 136*, 160–164.

Huang, D., S. R. Savage, A. P. Calinawan, C. Lin, B. Zhang, P. Wang, T. K. Starr, M. J. Birrer, and A. G. Paulovich (2021). A highly annotated database of genes associated with platinum resistance in cancer. *Oncogene 40*(46), 6395–6405.

Kanehisa, M. and S. Goto (2000). Kegg: kyoto encyclopedia of genes and genomes. *Nucleic acids research 28*(1), 27–30.

Kendall, M. G. (1948). *Rank correlation methods*. Griffin.

Koo, C.-Y., K. W. Muir, and E. W.-F. Lam (2012). Foxm1: From cancer initiation to pro-

gression and treatment. *Biochimica et Biophysica Acta (BBA) - Gene Regulatory Mechanisms 1819*(1), 28–37.

Kruskal, W. H. (1958). Ordinal measures of association. *Journal of the American Statistical Association 53*(284), 814–861.

Liberzon, A., A. Subramanian, R. Pinchback, H. Thorvaldsdóttir, P. Tamayo, and J. P. Mesirov (2011). Molecular signatures database (msigdb) 3.0. *Bioinformatics 27*(12), 1739–1740.

Liu, H., F. Han, M. Yuan, J. Lafferty, L. Wasserman, et al. (2012). High-dimensional semiparametric gaussian copula graphical models. *The Annals of Statistics 40*(4), 2293–2326.

Liu, H., J. Lafferty, and L. Wasserman (2009). The nonparanormal: Semiparametric estimation of high dimensional undirected graphs. *Journal of Machine Learning Research 10*(Oct), 2295–2328.

Ma, J., Z.-Q. Sun, S. Chen, and H.-H. Liu (2012). Dependence tree structure estimation via copula. *International Journal of Automation and Computing 9*(2), 113–121.

Mohan, K., P. London, M. Fazel, D. Witten, and S.-I. Lee (2014). Node-based learning of multiple gaussian graphical models. *The Journal of Machine Learning Research 15*(1), 445–488.

Nabavi, S., D. Schmolze, M. Maitituoheti, S. Malladi, and A. H. Beck (2016). Emdomics: a robust and powerful method for the identification of genes differentially expressed between heterogeneous classes. *Bioinformatics 32*(4), 533–541.

REFERENCES

Network, C. G. A. et al. (2012). Comprehensive molecular portraits of human breast tumors. *Nature 490*(7418), 61.

Network, C. G. A. R. et al. (2011). Integrated genomic analyses of ovarian carcinoma. *Nature 474*(7353), 609.

Ni, Y., P. Müller, L. Wei, and Y. Ji (2018). Bayesian graphical models for computational network biology. *BMC bioinformatics 19*(3), 59–69.

Ou-Yang, L., D. Cai, X.-F. Zhang, and H. Yan (2021). Wdne: an integrative graphical model for inferring differential networks from multi-platform gene expression data with missing values. *Briefings in Bioinformatics*.

Rosset, S. and J. Zhu (2007). Piecewise linear regularized solution paths. *The Annals of Statistics*, 1012–1030.

Schnitt, S. J. (2010). Classification and prognosis of invasive breast cancer: from morphology to molecular taxonomy. *Modern pathology 23*, S60–S64.

Shojaie, A. (2020). Differential network analysis: A statistical perspective. *Wiley Interdisciplinary Reviews: Computational Statistics*, e1508.

Still, G. (2018). Lectures on parametric optimization: An introduction. *Optimization Online*, 2.

Tang, Z., Z. Yu, and C. Wang (2020). A fast iterative algorithm for high-dimensional differential network. *Computational Statistics 35*(1), 95–109.

REFERENCES

Tavassolipour, M., S. A. Motahari, and M.-T. M. Shalmani (2018). Learning of tree-structured gaussian graphical models on distributed data under communication constraints. *IEEE Transactions on Signal Processing 67*(1), 17–28.

Tu, J.-J., L. Ou-Yang, Y. Zhu, H. Yan, H. Qin, and X.-F. Zhang (2021). Differential network analysis by simultaneously considering changes in gene interactions and gene expression. *Bioinformatics 37*(23), 4414–4423.

Wahlsten, D., P. Metten, T. J. Phillips, S. L. Boehm, S. Burkhart-Kasch, J. Dorow, S. Doerksen, C. Downing, J. Fogarty, K. Rodd-Henricks, et al. (2003). Different data from different labs: lessons from studies of gene–environment interaction. *Journal of neurobiology 54*(1), 283–311.

West, J., G. Bianconi, S. Severini, and A. E. Teschendorff (2012). Differential network entropy reveals cancer system hallmarks. *Scientific reports 2*(1), 1–8.

Wu, Y., T. Li, X. Liu, and L. Chen (2020). Differential network inference via the fused d-trace loss with cross variables. *Electronic Journal of Statistics 14*(1), 1269–1301.

Xia, Y. and L. Li (2019). Matrix graph hypothesis testing and application in brain connectivity alternation detection. *Statistica Sinica 29*(1), 303–328.

Xue, L., H. Zou, et al. (2012). Regularized rank-based estimation of high-dimensional nonparanormal graphical models. *The Annals of Statistics 40*(5), 2541–2571.

Yuan, H., R. Xi, C. Chen, and M. Deng (2017). Differential network analysis via lasso penalized d-trace loss. *Biometrika 104*(4), 755–770.

REFERENCES

Zhang, X.-F., L. Ou-Yang, and H. Yan (2017). Incorporating prior information into differential network analysis using non-paranormal graphical models. *Bioinformatics 33*(16), 2436–2445.

Zhao, B., Y. S. Wang, and M. Kolar (2022). Fudge: A method to estimate a functional differential graph in a high-dimensional setting. *Journal of Machine Learning Research 23*(82), 1–82.

Zhao, S. D., T. T. Cai, and H. Li (2014). Direct estimation of differential networks. *Biometrika 101*(2), 253–268.

Zhao, T., K. Roeder, and H. Liu (2014). Positive semidefinite rank-based correlation matrix estimation with application to semiparametric graph estimation. *Journal of Computational and Graphical Statistics 23*(4), 895–922.

Zwiener, I., B. Frisch, and H. Binder (2014). Transforming rna-seq data to improve the performance of prognostic gene signatures. *PloS one 9*(1), e85150.

Department of Computer Engineering, Sharif University of Technology

E-mail: nikahd@ce.sharif.edu

Mojtaba Nikahd, ORCID ID: `https://orcid.org/0009-0006-0311-8591`

Department of Computer Engineering, Sharif University of Technology

E-mail: motahari@sharif.edu

Seyed Abolfazl Motahari, ORCID ID: `https://orcid.org/0000-0002-7639-0362`