

The Sample Complexity of Differential Analysis for Networks that Obey Conservation Laws*

Jiajun Cheng*, Anirudh Rayas*, Rajasekhar Anguluri[†], Gautam Dasarathy*

*Arizona State University, [†]University of Maryland, Baltimore County

Abstract—Networked systems that obey conservation laws are common in many domains such as power grids, biological systems, and social networks. These systems are described by so-called balance equations that link injected flows and node potentials, ensuring that the flow at each node is balanced. For example, electric networks follow Kirchhoff's laws, while social networks model group consensus. Understanding the structure of these networks based on node potential data has become an important research topic. In this work, we focus on the problem of differential network analysis for systems that obey conservation laws. That is, instead of the structure of a network, we focus on estimating the structural differences between two networks from their node potential data. We propose a method that uses a high-dimensional estimator to directly identify these structural changes. We provide theoretical guarantees and test our method on both synthetic networks and benchmark power network data to validate its performance. The results show that our method works well but also highlight some gaps between the theoretical guarantees and experimental outcomes. Addressing these gaps is an important step for improving future methods.

Index Terms—differential network analysis, structure learning.

I. INTRODUCTION

Consider a graph $\mathcal{G} = ([p], E)$ and let $X \in \mathbb{R}^p$ be a p -dimensional vector of *injected flows* at the vertices $[p]$. Let $Y \in \mathbb{R}^p$ be the vector of the *vertex potentials*. The flows and potentials are said to satisfy a *conservation law* with respect to the graph $\mathcal{G}$ if they obey the relationship $Y = (B^*)^{-1}X$, where B^* is an invertible Laplacian matrix associated with G [1]. At each vertex, the flows directly counteract the injections. Such an equation is called a *balance equation* between potential and inject flows. For instance, electric networks obey Kirchhoff's laws and social networks reflect a consensus of views. In many real-world applications, learning the unknown structure of a network is essential for understanding and managing complex systems. This task often involves inferring the structure of graphs that obey conservation laws based on node potentials, a topic that has garnered considerable attention in recent years (see, for instance, [1, 4, 6]).

Network systems in many practical applications are not static; their structure evolves over time. Thus, one often aims to understand the difference in structures of two networks. This problem of *differential network analysis* has become crucial for several tasks across a wide range of fields, including biomedicine, social networks, and cyber-physical systems

[2, 3, 5, 7]. Given the importance of this problem, we are motivated to study the estimation of the *difference matrix* $\Delta^* = B_2^* - B_1^*$ between the structures of two networks that obey conservation laws, where B_1^* and B_2^* are invertible Laplacian matrix that represent the structure of the corresponding networks. In [12], we proposed an estimator that directly estimates the *difference matrix* based on node potential data. Additionally, we establish the convexity of the estimator and demonstrate its superior performance in experiments. In this paper, we mainly focus on the statistical guarantees of this estimator. Building on the *primal-dual witness* framework (see, e.g., [10]), we quantify the sample requirement necessary for accurately recovering the true sparse network changes. Specifically, we guarantee that an element-wise ℓ_∞ error bound of order $\mathcal{O}(\sqrt{\frac{p \log(p)}{n}})$ holds with high probability, with a sample complexity satisfying $\Omega(pd^2 \log(p))$, where d represents the maximum degree of any row or column in Δ^* . Finally, we note a gap between our theorem and the results obtained in the experiments. The challenge arise from the *square root perturbation bounds*, which we discuss later in the paper. Addressing this issue could lead to more accurate statistical guarantees. Given the importance of this problem, we are motivated to study the estimation of the *difference matrix* $\Delta^* = B_2^* - B_1^*$ between the structures of two networks that obey conservation laws, where B_1^* and B_2^* are invertible Laplacian matrix that represent the structure of the corresponding networks. In [12], we proposed an estimator that directly estimates the *difference matrix* based on node potential data. Additionally, we establish the convexity of the estimator and demonstrate its superior performance in experiments. In this paper, we mainly focus on the statistical guarantees of this estimator. Building on the *primal-dual witness* framework (see, e.g., [10]) we quantify the sample requirement necessary for accurately recovering the true sparse network changes. Specifically, we guarantee that an element-wise ℓ_∞ error bound of order $\mathcal{O}(\sqrt{\frac{p \log(p)}{n}})$ holds with high probability, with a sample complexity satisfying $\Omega(pd^2 \log(p))$, where d represents the maximum degree of any row or column in Δ^* . Finally, we note a gap between our theorem and the scaling law suggested by experiments. The difference arises from our sub-optimal *square root perturbation bounds*, which we discuss later in the paper. Improving on this could lead to more accurate (and significantly better) statistical guarantees.

This work was supported in part by the National Science Foundation (NSF) under the grants NSF CCF-2048223, NSF CNS-2003111, NSF CCF-2007688, and the National Institutes of Health (NIH) under the grant 1R01GM140468-01

II. PRELIMINARIES AND BACKGROUND

Consider two graphs, $\mathcal{G}_1 = ([p], E_1)$ and $\mathcal{G}_2 = ([p], E_2)$, with identical node sets $[p]$ but different edge sets E_1 and E_2 . These graphs are associated with corresponding (invertible) Laplacian matrices B_1^* and B_2^* , respectively. Let $X_1 \sim \mathcal{N}(0, \Sigma_{X_1})$ and $X_2 \sim \mathcal{N}(0, \Sigma_{X_2})$ be random injection vectors associated with each graph, where the covariances Σ_{X_1} and Σ_{X_2} are assumed to be known. These injection vectors can be interpreted as electric flows governed by Kirchhoff's law or as traffic flows in a transportation network. The goal is to estimate the *difference matrix* Δ^* by observing node potentials $Y_i = (B_i^*)^{-1}X_i$ for $i \in \{1, 2\}$. First, notice that $Y_i \sim \mathcal{N}(0, \Theta_i^{*-1})$, where $\Theta_i^* = B_i^* \Sigma_{X_i}^{-1} B_i^*$, $i \in \{1, 2\}$. Letting $M_{X_i} \succ 0$ denote the unique square root [8] of Σ_{X_i} . We next need to develop an expression for Δ^* as a function of Θ_1^* and Θ_2^* , we can define $\tilde{Y}_i = M_{X_i} Y_i$ and set $\tilde{\Theta}_i^* = (\text{Cov}[\tilde{Y}_i])^{-1}$, this expression follows by direct substitution. Now notice that with these definitions, we have $\Delta^* = M_{X_2}(\tilde{\Theta}_2^*)^{\frac{1}{2}} M_{X_2} - M_{X_1}(\tilde{\Theta}_1^*)^{\frac{1}{2}} M_{X_1}$. In [12], we use this insight to design a regularized estimator

$$\hat{\Delta} \in \underset{\Delta \in \mathbb{R}^{p \times p}}{\text{argmin}} L(\Delta) + \lambda_n \|\Delta\|_{1, \text{off}}, \quad (1)$$

where $\lambda_n \geq 0$, $L(\Delta) = \frac{1}{4} (\langle \hat{\Psi}_1 \Delta, \hat{\Psi}_2 \rangle + \langle \hat{\Psi}_2 \Delta, \hat{\Psi}_1 \rangle) - \langle \Delta, \hat{\Psi}_1 - \hat{\Psi}_2 \rangle$, and $\hat{\Psi}_i = M_{X_i}^{-1} \tilde{S}_i^{\frac{1}{2}} M_{X_i}^{-1}$ which is an estimate of $\tilde{\Psi}_i = M_{X_i}^{-1}(\tilde{\Theta}_i^*)^{-\frac{1}{2}} M_{X_i}^{-1}$; $\tilde{S}_i^{\frac{1}{2}} \cdot \|\Delta\|_{1, \text{off}}$ represents the ℓ_1 -norm applied to the off-diagonal elements of Δ . This estimator is an ℓ_1 regularized variant of the D-trace loss [9]. While the above problem setup assumes Gaussian injection vectors, our methods hold for more general distributions. Toward characterizing the performance in these cases, we begin by defining tail conditions that dictate the achievable sample complexity of the problem.

III. TAIL CONDITIONS

In this section, we introduce the tail conditions that form the foundation of our analysis. The estimator in (1) depends on the sample covariance $\tilde{S}_i$ as an approximation of the (unknown) true covariance $\tilde{\Theta}_i^{*-1}$. To ensure consistency and to quantify data requirements, it is necessary to establish bounds on the difference $\tilde{S}_i - \tilde{\Theta}_i^{*-1}$. Specifically, we define the following:

Definition 1 (Tail Conditions, [10]). The random vector Y satisfies the tail condition $\mathcal{T}(f, v_*)$ if there exists a constant $v_* \in (0, \infty)$ and a function $f : \mathbb{N} \times (0, \infty) \rightarrow (0, \infty)$ such that for any $(i, j) \in [p] \times [p]$:

$$\mathbb{P} \left[|(\tilde{S}_i)_{kl} - (\tilde{\Theta}_i^{*-1})_{kl}| \geq \delta \right] \leq \frac{1}{f(n, \delta)}$$

Where $f(n, \delta)$ is monotonically increasing in n (for fixed n) or δ (for fixed n).

Both the exponential tail $f(n, \delta) = \exp(cn\delta^a)$ and polynomial tail $f(n, \delta) = cn^m \delta^{2m}$, where $m \in \mathbb{N}$, $c, a > 0$,

satisfy the monotonicity in Definition 1. To analyze sample complexity, we define the inverse functions:

$$n_f(\delta, p^\eta) := \max\{n \mid f(n, \delta) \leq p^\eta\}, \\ \delta_f(n, p^\eta) := \max\{\delta \mid f(n, \delta) \leq p^\eta\}.$$

Furthermore, if $n > n_f(\delta, p^\eta)$, then $\delta \geq \delta_f(n, p^\eta)$. Due to space constraints, we only state the result for the sub-Gaussian case, a generalization to the Gaussian case.

Definition 2 (Sub-Gaussian random variable). A mean-zero random vector $Z \in \mathbb{R}^p$ is sub-Gaussian if there exists $\sigma > 0$ such that

$$\mathbb{E}[\exp(tZ)] \leq \exp\left(\frac{\sigma^2 t^2}{2}\right),$$

For sub-Gaussian random vectors, the exponential-type tail bound is given by:

$$v_* = \left(\max_i (\tilde{\Theta}^*)_{ii}^{-1} \cdot 8(1 + 4\sigma^2)\right)^{-1}, \\ f(n, \delta) = \frac{1}{4} \exp(-c_* n \delta^2),$$

where $c_* = \left(128(1 + 4\sigma^2)^2 \max_i (\tilde{\Theta}^*)_{ii}^{-1}\right)^{-1}$. The inverse functions are:

$$\delta_f(p^\eta, n) = \sqrt{\frac{\log(4/p^\eta)}{c_* n}}, \quad n_f(p^\eta, \delta) = \frac{\log(4/p^\eta)}{c_* \delta^2}.$$

Beyond sub-Gaussian distributions, we also provide guarantees for cases where the injection vectors X_1 and X_2 follow other distributions, including those with polynomial tails in [13].

IV. MAIN RESULT

Our results provide a theoretical analysis of the performance of the estimator (1) when Y_1 and Y_2 follow a sub-Gaussian distribution. Our analysis shows that the estimator (1) reliably captures the sparsity structure of Δ^* and closely approximates Δ^* with high probability, provided the sample size n satisfies $n = \Omega(pd^2 \log p)$. Since our approach is an ℓ_1 -regularized variant of the D-trace loss function, the main results may appear similar to those in [7]. However, we address a novel challenge, referred to as the *Square Root Perturbation Bounds*, which leads to new theorems beyond the scope of previous work and opens valuable directions for future research.

A. The irrepresentability condition

We assume the true network difference Δ^* is sparse, that is, let $S = \{(i, j) : \Delta_{i,j}^* \neq 0\}$ be the support of Δ^* and $s = |S|$, $s < p$. We use $\|A\|_\infty = \max_{i,j} |A_{ij}|$ to denote the element wise norm and in addition, we also define $\|A\|_1 = \|\text{vec}(A)\|_1$ and $\|A\|_{1,\infty} \triangleq \max_{i=1,\dots,p} \sum_{j=1}^p |A_{ij}|$. Let d denote the maximum node degree in Δ^* , and $\kappa_\Gamma = \|\Gamma_{S,S}^*\|_{1,\infty}$. We also assume that $\max\{\|\tilde{\Psi}_1\|_\infty, \|\tilde{\Psi}_2\|_\infty\} \leq M$. Denote $\hat{\Gamma} = \frac{\tilde{\Psi}_1 \otimes \tilde{\Psi}_2 + \tilde{\Psi}_2 \otimes \tilde{\Psi}_1}{2}$ and $\Gamma^* = \frac{\tilde{\Psi}_1 \otimes \tilde{\Psi}_2 + \tilde{\Psi}_2 \otimes \tilde{\Psi}_1}{2}$. This is also the Hessian matrix with respect to Δ of the D-trace loss function that we will define it later. We can consider that Γ could be used to describe the relationships between two sets of variables

in this case could be for understanding the overall interaction between the two sets. Given two sets T_1 and T_2 , the submatrix of Γ with rows indexed by T_1 and columns indexed by T_2 is denoted by $\Gamma_{T_1 T_2}$. This submatrix is composed of elements $(A_{j,l} B_{k,m})$ for each (j,k) .

Assumption 1. We assume the following irrepresentability condition

$$\|\Gamma_{S^c, S}^* (\Gamma_{S, S}^*)^{-1}\|_1 \leq 1 - \alpha. \quad (2)$$

Where there are some $\alpha \in (0, 1]$

The Hessian $\Gamma_{(j,k),(l,m)}^*$ also shows the covariance of the random variable linked to each edge of the graph. Therefore, the above assumption imposes control on the influences that non-edge terms (indexed by S^c), can have on the edge-based terms (indexed by S).

B. Convergence Rate

Theorem 1. Let $\tilde{Y}_1$ and $\tilde{Y}_2$ be the node potential vector. Suppose that $\tilde{Y}_1$ and $\tilde{Y}_2$ are sub-Gaussian with parameter σ_1 and σ_2 , respectively. Under the irrepresentability condition (1) for some $\eta > 2$ and with a sample size for both $\tilde{Y}_1$ and $\tilde{Y}_2$ that is lower bounded as $n \geq p\tilde{C}_0(\eta \log(p) + \log(4))$, Then, with probability larger than $1 - 2/p^{\eta-2}$, for some $\eta > 2$:

- (a) $\hat{\Delta}$ recovers the sparsity structure of Δ^* ; that is, $\hat{\Delta}_{S^c} = 0$.
- (b) $\hat{\Delta}$ satisfies the element-wise ℓ_∞ bound $\|\hat{\Delta} - \Delta^*\|_\infty \leq \tilde{C}_1 \sqrt{p} \left[\frac{\eta \log(p) + \log(4)}{n} \right]^{\frac{1}{2}}$.

Here, $\tilde{C}_0$ and $\tilde{C}_1$ are constants that depend on κ_Γ , M , and α . We assume they remain constant as functions of n , p , and d . Specifically, the asymptotic behavior of $\tilde{C}_0$ can be expressed as $\mathcal{O}(d^2)$ (see [13] for their definitions and detailed proofs).

Thus, we obtain an element-wise ℓ_∞ bound of $\|\hat{\Delta} - \Delta^*\|_\infty \in \mathcal{O}(\sqrt{p \log(p)/n})$, which holds with high probability for a sample size $n = \Omega(pd^2 \log(p))$. For the other quantities involved in the theorem, κ_Γ and M characterize the size of the Hessian $(\Gamma^*)^{-1}$. Finally, both $\tilde{C}_0$ and $\tilde{C}_1$ also depend on the parameter α introduced in Assumption 1.

V. PROOF OF THEOREM 1

Our proof builds on the *Primal-Dual Witness* framework established in [10]. The method involves constructing a *primal-dual witness* pair $(\tilde{\Delta}, \tilde{Z})$ that satisfies the optimality conditions of our problem (1). If the construction succeeds, then $\tilde{\Delta} = \hat{\Delta}$. Thus, the main direction of our proof is to demonstrate that this construction succeeds with high probability.

A. Primal-Dual witness method

We first define the restrictive problem:

$$\tilde{\Delta} = \arg \min_{\Delta \in \mathbb{R}^{p \times p}, \Delta_{S^c} = 0} L(\Delta) + \lambda_n \|\Delta\|_{1, \text{off}}, \quad (3)$$

where $\tilde{\Delta}$ is the solution to the restrictive problem. We note that the sub-differential of $\|\cdot\|_{1, \text{off}}$ with respect of Δ contains matrices $Z \in \mathbb{R}^{p \times p}$ such that

$$Z_{ij} = \begin{cases} 0, & \text{if } i = j \\ \text{sign}(\Delta_{ij}), & \text{if } i \neq j \text{ and } \Delta_{ij} \neq 0 \\ \in [-1, +1], & \text{if } i \neq j \text{ and } \Delta_{ij} = 0 \end{cases} \quad (4)$$

From the directional derivative of (3), we obtain the equality,

$$\left\{ \frac{1}{2} (\hat{\Psi}_1 \tilde{\Delta} \hat{\Psi}_2 + \hat{\Psi}_2 \tilde{\Delta} \hat{\Psi}_1) - \hat{\Psi}_1 + \hat{\Psi}_2 + \lambda_n Z \right\}_S = 0. \quad (5)$$

We choose $\tilde{Z}_S$ as a member of the sub differential of $\|\cdot\|_{1, \text{off}}$ with respect of Δ . To ensure that constructed matrices $(\tilde{\Delta}, \tilde{Z})$ satisfy the optimality condition of (5). We set $\tilde{Z}_{S^c}$ as

$$\tilde{Z}_{S^c} = -\frac{1}{\lambda_n} \left\{ \frac{1}{2} (\hat{\Psi}_1 \tilde{\Delta} \hat{\Psi}_2 + \hat{\Psi}_2 \tilde{\Delta} \hat{\Psi}_1) - \hat{\Psi}_1 + \hat{\Psi}_2 \right\}_{S^c}.$$

Finally, we need to verify the *strict dual feasibility*

$$|\tilde{Z}_{ij}| < 1, \text{ For all } (i, j) \in S^c.$$

B. Supporting Lemmas

In this section, we present a sequence of lemmas that establish the successful construction of the *Primal-Dual Witness* pair $(\tilde{\Delta}, \tilde{Z})$. First, we define the *sufficient conditions for strict dual feasibility*, which guarantee that *strict dual feasibility* holds. Finally, we demonstrate that with the assumption 1, an appropriate choice of λ_n and a sample size bounded as specified in Theorem 1, these *sufficient conditions for strict dual feasibility* are satisfied with high probability.

1) *Sufficient condition for strict dual feasibility*: Due to space limitations, we do not state and explain all the detailed conditions required to verify strict dual feasibility. Readers are referred to Lemma 1 and Lemma 2, along with their proofs, in [13] for a comprehensive understanding of how these lemmas verify *Strict Dual Feasibility* and facilitate the successful construction of primal-dual witnesses.

2) *Square Root Perturbation Bounds*: The estimator (1) use $\hat{\Psi}_i$ as proxy for the true $\tilde{\Psi}_i$. The first step is to obtain bounds on the differences $\hat{\Psi}_i - \tilde{\Psi}_i$, which can be written as $M_{X_i}^{-1}(\tilde{S}_i^{\frac{1}{2}} - (\tilde{\Theta}_i^*)^{-\frac{1}{2}})M_{X_i}^{-1}$. To bound the differences $\tilde{S}_i^{\frac{1}{2}} - (\tilde{\Theta}_i^*)^{-\frac{1}{2}}$, we need to identify alternative expressions that bound the behavior of the square root function.

Lemma 1. (Generalized Power-Stormer Inequalities [11]) Let A and B be positive semidefinite matrices, and let $n \geq 1$. If f is an operator monotone function, then

$$\|A^{\frac{1}{n}} - B^{\frac{1}{n}}\|_{op} \leq \|A - B\|_{op}^{\frac{1}{n}}.$$

From Lemma 1, we apply this result to the case where the square root function, $f(X) = \sqrt{X}$, is an operator monotone function. Using this, we derive the following chain of inequalities:

$$\begin{aligned} \|\sqrt{\tilde{S}_i} - \sqrt{\Theta_i^*}^{-1}\|_\infty &\leq \sqrt{\|S_i - \Theta_i^*^{-1}\|_{op}} \\ &\leq \sqrt{p} \|S_i - \Theta_i^*^{-1}\|_\infty. \end{aligned}$$

Instead of direct calculation on the $\tilde{S}_i^{\frac{1}{2}} - (\tilde{\Theta}_i^*)^{-\frac{1}{2}}$, now we can use $\sqrt{p} \|\tilde{S}_i - \Theta_i^*^{-1}\|_\infty$ to have bounds on the differences $\tilde{S}_i - (\tilde{\Theta}_i^*)^{-1}$. The final result of the Theorem 1 is based on the the control of noise $\hat{\Psi}_i - \tilde{\Psi}_i$. We have inequality:

$$\begin{aligned} \|\hat{\Psi}_i - \tilde{\Psi}_i\|_\infty &= \|M_{X_i}^{-1}\|_\infty (\tilde{S}_i^{\frac{1}{2}} - (\tilde{\Theta}_i^*)^{-\frac{1}{2}}) \|M_{X_i}^{-1}\|_\infty \\ &\leq \|M_{X_i}^{-1}\|_\infty \sqrt{p} \|\tilde{S}_i - (\tilde{\Theta}_i^*)^{-1}\|_\infty \|M_{X_i}^{-1}\|_\infty \end{aligned} \quad (6)$$

Lemma 2 (Control of noise term). *Let*

$$\tilde{\delta}_{f_i} = \tilde{\delta}_{f_i}(n, p^\eta) = \|M_{X_i}^{-1}\|_\infty \sqrt{p} \delta_{f_i}(n, p^\eta) \|M_{X_i}^{-1}\|_\infty. \quad (7)$$

For some $\eta > 2$, we have

$$\mathbb{P}\left(\|\tilde{\Psi}_i - \tilde{\Psi}_i\|_\infty \leq \tilde{\delta}_{f_i}(n, p^\eta)\right) \geq 1 - \frac{1}{p^{\eta-2}}.$$

Proof. For any $0 < \delta < 1/v_*$ and $r \geq 1$, if $n > n_f(\delta, r)$, we have $f(n, \delta) > r$, and thus $\delta_f(n, r) < \delta$ since $f(n, \delta)$ is monotonically increasing in δ . Therefore, $\mathbb{P}\left\{\left|(\tilde{S}_i)_{kl} - (\tilde{\Theta}_i^*)_{kl}^{-1}\right| \geq \delta_f(n, r)\right\} \leq r^{-1}$, where we used $1/f(n, \delta_f(n, r)) = r^{-1}$. Applying the union bound across all entries gives $\mathbb{P}\{\|(\tilde{S}_i) - (\tilde{\Theta}_i^*)^{-1}\|_\infty < \delta_f(n, r)\} > 1 - p^2 r^{-1}$. Finally, applying inequalities (6) and setting $r = p^\tau$, the result follows. $\square$

Note that by applying the union bound across both matrices, we obtain a combined probability bound of $1 - \frac{2}{p^{\eta-2}}$.

3) *Central Lemma:* We now need to offer guarantees for support recovery and bounds for the infinity norm of $\hat{\Delta}$.

Lemma 3. *Let $\hat{\Delta}$ be the unique solution of the problem (1) with*

$$\lambda_n = \max \left\{ \frac{2(4 - \alpha)(\tilde{\delta}_{f_1} + \tilde{\delta}_{f_2})}{\alpha}, \frac{24dM(\kappa_\Gamma + dM^2\kappa_\Gamma^2)(\tilde{\delta}_{f_1}\tilde{\delta}_{f_2} + M\tilde{\delta}_{f_2} + M\tilde{\delta}_{f_1})}{\alpha} \right\}.$$

For a distribution satisfying the assumption 1 and the tail condition in Definition 2, if the sample size is lower bounded as

$$n > n_f \left(\min \left\{ -M + \sqrt{M^2 + (6d\kappa_\Gamma)^{-1}}, -M + \sqrt{M^2 + \frac{\alpha}{24d(\kappa_\Gamma + dM^2\kappa_\Gamma^2)}}, \frac{\alpha M}{4 - \alpha} \right\}, p^\eta \right)$$

then with probability greater than $1 - \frac{1}{p^{\eta-2}}$, the $\hat{\Delta}$ recovers the sparsity structure of B^* , and satisfies the ℓ_∞ bound

$$\|\hat{\Delta} - \Delta^*\|_\infty \leq (\tilde{\delta}_{f_1} + \tilde{\delta}_{f_2} + \lambda_n)\kappa_\Gamma + 3d\kappa_\Gamma^2(\tilde{\delta}_{f_1} + \tilde{\delta}_{f_2} + \lambda_n + 2M) + (\tilde{\delta}_{f_1}\tilde{\delta}_{f_2} + M\tilde{\delta}_{f_2} + M\tilde{\delta}_{f_1}) \quad (8)$$

Proof sketch: We show that the Primal-Dual witness construction succeeds (see Proof of Lemma 1 and Lemma 2 in [13]). This is equivalent to demonstrating that the *Sufficient Conditions for Strict Dual Feasibility* hold with the required probability. Specifically, we have shown in Lemma 2 that event $\|\tilde{\Psi}_i - \tilde{\Psi}_i\|_\infty \leq \tilde{\delta}_{f_i}(n, p^\eta)$ satisfies $\mathbb{P}(\|\tilde{\Psi}_i - \tilde{\Psi}_i\|_\infty \leq \tilde{\delta}_{f_i}(n, p^\eta)) \geq 1 - \frac{1}{p^{\eta-2}}$, and conditioned on this event, the *Sufficient Conditions for Strict Dual Feasibility* holds under the assumption 1, the sample complexity and choice of λ_n stated in this Lemma, see Appendix A.1 in [13] for full details.

The results in Theorem 1 are derived by combining the sequence of lemmas. Next, we substitute the inverse functions of the *Sub-Gaussian* distribution, defined in Definition 2, into (7) to obtain $\tilde{\delta}_{f_1}$ and $\tilde{\delta}_{f_2}$. These values are then substituted into (8). For further details, see Appendix A.2 in [13].

VI. EXPERIMENTAL RESULTS AND CONCLUSION

It should be noted that while the sample complexity scales as $pd^2 \log p$, one might have hoped it would scale as $d^2 \log p$. In fact, this is what the experimental results in [12] appear to suggest. In this section, we perform additional simulations on various networks to evaluate the performance of the proposed estimator. The experiments include four synthetic random networks of size $p = 30$ and three benchmark networks. We assess the edge recovery performance using the F-score.

$$\text{F-score} = \frac{2\text{TP}}{2\text{TP} + \text{FP} + \text{FN}} \in [0, 1],$$

where TP (true positives) is the number of correctly detected edges, FP (false positives) is the number of non-existent edges that were detected, and FN (false negatives) is the number of actual edges that were not detected. We also evaluate the worst-case error using $\|\hat{\Delta} - \Delta^*\|_\infty$.

A. Synthetic Networks

The synthetic networks examined in Figure 1 include Erdős-Rényi graphs, characterized by randomly connected nodes; Small-World graphs, generated using the Watts-Strogatz model; Scale-Free graphs, constructed through the Barabási-Albert model; and structured Grid Graphs. For each network, an adjacency matrix A is constructed, and the corresponding Laplacian matrix $L = D - A$ is computed, where D is the degree matrix. The Laplacian is then regularized as $B_1 = L + I_p$, ensuring positive definiteness, where I_p denotes the p -dimensional identity matrix.

To model changes in edge structure, a difference matrix Δ^* is applied such that $\Delta^* + B_1 = B_2$. The support of Δ^* is determined by a binary sparsity pattern, with a sparsity factor of 0.90. This means 90% of the entries in Δ^* are zeros, ensuring the *difference matrix* is highly sparse. Sample sizes are scaled with the network properties according to $N \propto d^2 \log p$, where d is the maximum degree of the network.

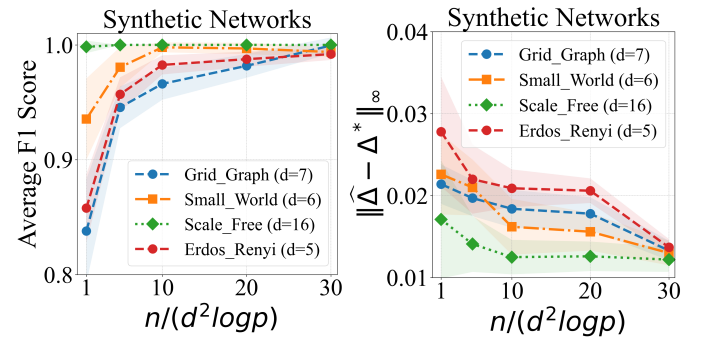


Fig. 1. Estimation performance for synthetic networks. Solid curves represent the mean performance, while shaded areas indicate standard deviations.

B. Benchmark Networks

In Figure 2, we also evaluate the performance on three benchmark networks: the power distribution network, the water network, and the brain network. Each network is represented by a corresponding adjacency matrices, B_1 . Similar

to Synthetic Experiment, $B_1 = B_1 + I_p$, and $\Delta^* + B_1 = B_2$, where support of Δ^* is based on a sparsity factor of 0.90.

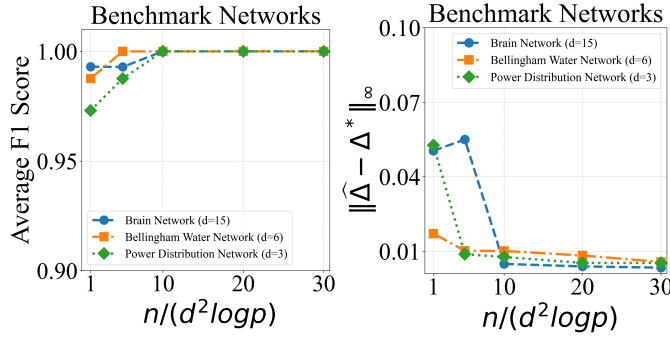


Fig. 2. Estimation performance for Benchmark Networks

We focus on three distinct networks in this analysis: **Power Distribution Network:** The IEEE 33-bus power distribution system consists of 33 buses and 32 branches (edges), with a maximum degree of $d = 3$. The raw dataset is publicly accessible at ¹. **Water Distribution Network:** The ground truth adjacency matrix, featuring 121 nodes and 162 edges with a maximum degree of $d = 6$, is created by processing the raw dataset described in [15] using the WNTR simulator². The dataset is available at ³. **Brain Network:** This study utilizes a publicly available benchmark connectivity matrix⁴, with details on its construction outlined in [14]. The ground truth adjacency matrix, denoted as A , is a 90×90 matrix, where each row and column represent a unique region of interest (ROI) in the brain.

C. Conclusion

We plot the performance of the estimator as it improves with the rescaled sample size $n/d^2 \log(p)$ instead of $\mathcal{O}(pd^2 \log(p))$. This scaling is sublinear in the dimension p , so that for sub-Gaussian random variables, the method can succeed in $n < p$ case. Although all simulation results demonstrate that the estimator improves with the rescaled sample size $n/d^2 \log(p)$, our theoretical analysis has only established a sample complexity of $\mathcal{O}(pd^2 \log(p))$. Closing this gap between theory and simulation remains a key objective for our future research. The primary challenge in achieving a tighter bound lies in deriving a concentration inequality for the square root of a matrix.

REFERENCES

- [1] A. Rayas, R. Anguluri, and G. Dasarathy, "Learning the Structure of Large Networked Systems Obeying Conservation Laws," in *Advances in Neural Information Processing Systems*, Curran Associates, Inc., 2022, vol. 35, pp. 14637–14650.
- [2] G. Dasarathy, P. Shah, and R.G. Baraniuk, "Sketched covariance testing: A compression-statistics tradeoff," in *2017 51st Asilomar Conference on Signals, Systems, and Computers*, IEEE, 2017, pp. 676–680.
- [3] S. Na, M. Kolar, and O. Koyejo, "Estimating differential latent variable graphical models with applications to brain connectivity," in *Biometrika*, vol. 108, no. 2, 2021, pp. 425–442.
- [4] D. Deka, S. Talukdar, M. Chertkov, and M.V. Salapaka, "Graphical models in meshed distribution grids: Topology estimation, change detection & limitations," in *IEEE Transactions on Smart Grid*, vol. 11, no. 5, 2020, pp. 4299–4310.
- [5] A. Shojaie, "Differential network analysis: A statistical perspective," in *Wiley Interdisciplinary Reviews: Computational Statistics*, vol. 13, no. 2, 2021.
- [6] R. Anguluri, G. Dasarathy, O. Kosut, and L. Sankar, "Grid Topology Identification With Hidden Nodes via Structured Norm Minimization," in *IEEE Control Systems Letters*, vol. 6, 2021, pp. 1244–1249.
- [7] H. Yuan, R. Xi, C. Chen, and M. Deng, "Differential network analysis via lasso penalized D-trace loss," in *Biometrika*, Oxford University Press, 2017, vol. 104, no. 4, pp. 755–770.
- [8] R. Bhatia, "Positive definite matrices," Princeton University Press, 2009.
- [9] T. Zhang and H. Zou, "Sparse precision matrix estimation via lasso penalized D-trace loss," in *Biometrika*, vol. 101, no. 1, 2014, pp. 103–120.
- [10] P. Ravikumar, M.J. Wainwright, G. Raskutti, and B. Yu, "High-dimensional covariance estimation by minimizing ℓ_1 -penalized log-determinant divergence," in *Electronic Journal of Statistics*, vol. 22, 2011, pp. 935–980.
- [11] J. Phillips, "On the uniform continuity of operator functions and generalized Powers-Stormer inequalities," Technical Report DM-442-IR, 1987. [Online]. Available: <http://hdl.handle.net/1828/1506>.
- [12] A. Rayas, R. Anguluri, J. Cheng, and G. Dasarathy, "Differential Analysis for Networks Obeying Conservation Laws," in *ICASSP 2023-2023 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, IEEE, 2023, pp. 1–5.
- [13] J. Cheng, A. Rayas, R. Anguluri, and G. Dasarathy, "The Sample Complexity of Differential Analysis for Networks that Obey Conservation Laws," available online: <https://shorturl.at/abTrR>, 2024.
- [14] Škoch, Antonín, Reháč Bučková, Barbora, Mareš, Jan, Tintěra, Jaroslav, Sanda, Pavel, Jajcay, Lucia, Horáček, Jiří, Španiel, Filip, and Hlinka, Jaroslav. "Human brain structural connectivity matrices—ready for modelling." *Scientific Data*, vol. 9, no. 1, 2022. Nature Publishing Group UK London.
- [15] Hernandez, Erika, Hoagland, Steven, and Ormsbee, Lindell. "Water distribution database for research applications." In *World Environmental and Water Resources Congress*, pp. 465–474, 2016.

¹<https://www.mathworks.com/matlabcentral/fileexchange/73127-ieee-33-bus-system>

²<https://github.com/USEPA/WNTR>

³<https://www.uky.edu/WDST/index.html>

⁴<https://osf.io/yw5vf/>