Learning Graph Parameters from Linear Measurements: Fundamental Trade-offs and Application to Electric Grids

Tongxin Li, Lucien Werner, Steven H. Low

March 24, 2019

Abstract

We consider a specific graph learning task: reconstructing a symmetric matrix that represents an underlying graph using linear measurements. We study fundamental trade-offs between the number of measurements (sample complexity), the complexity of the graph class, and the probability of error by first deriving a necessary condition (fundamental limit) on the number of measurements. Then, by considering a two-stage recovery scheme, we give a sufficient condition for recovery. Furthermore, assuming the measurements are Gaussian IID, we prove upper and lower bounds on the (worst-case) sample complexity. In the special cases of the uniform distribution on trees with n nodes and the Erdős–Rényi (n, p) class, the fundamental trade-offs are tight up to multiplicative factors. Applying the Kirchhoff's matrix tree theorem, our results are extended to the scenario when part of the topology information is known a priori. In addition, we design and implement a polynomial-time (in n) algorithm based on the two-stage recovery scheme. Simulations for several canonical graph classes and IEEE power system test cases demonstrate the effectiveness of the proposed algorithm for accurate topology and parameter recovery.

Graph Learning, System Identification, Information Theory, Smart Grid, Sparse Recovery

1 Introduction

1.1 Background

Symmetric matrices are ubiquitous contstructs in graphical models with examples such as the (0, 1) adjacency matrix and the (generalized) Laplacian of an undirected graph. A major challenge in graph learning is inferring graph parameters embedded in those graph-based matrices from historical data or real-time measurements. In some problem settings, low sample complexity is important. One reason for this is that the graph parameters may change in a short time-scale, making the latter requirement of vital importance to guarantee that the recovery is accomplished with limited number of measurements. Real-time or nearly real-time graph algorithms based on temporal data require a frequent update on the underlying graph parameters. For example, power system applications, such as (real-time) optimal power flow [1, 2, 3], real-time contingency analysis [4] and frequency control [5] *etc*., require knowledge of underlying graph parameters.

In this work, we consider the case when the measurements and underlying matrix to be recovered can be represented as or approximated by a linear system. A *graph matrix* $\mathbf{Y}(G)$ with respect to an underlying graph G (see Definition 2.1) is defined as an $n \times n$ symmetric matrix with each nonzero (i, j)-th entry corresponding to an edge connecting node i and node j where $n \in \mathbb{N}_+$ is the number of nodes of the underlying *undirected* graph. The diagonal entries can be arbitrary. The measurements are summarized as two $m \times n$ $(1 \leq m \leq n)$

real or complex matrices $\mathbf{A}$ and $\mathbf{B}$ satisfying

$$\mathbf{A} = \mathbf{B}\mathbf{Y}(G) + \mathbf{Z} \tag{1}$$

where $\mathbf{Z}$ is additive noise.

We focus on the following problems:

- *Fundamental Trade-offs* What is the *minimum number m* of linear measurements required for reconstructing the *symmetric* matrix $\mathbf{Y}$? Is there an algorithm *asymptotically achieving* recovery with the minimum number of measurements? Can we characterize the sample complexity when the measurements are Gaussian IID[1]?

- *Applications to Electrical Grids* Do the theoretical guarantees on sample complexity result in a practical algorithm (in terms of both sample and computational complexity) for recovering electric grid topology and parameters?

1.2 Related Work

Information-theoretic tools have been widely applied to derive fundamental limits for learning graph structures. For a Markov random field (MRF) with bounded maximum degree, [6] derived necessary conditions on the number of samples for estimating the underlying graph structure using Fano's inequality (see [7]). For Ising models, [8] combined Fano's inequality with *typicality* to derive weak and strong converse. Similar techniques have also been applied to Gaussian graphical models [9] and Bayesian networks [10]. Fundamental limits for noisy compressed sensing have been extensively studied in [11] under an information theoretic framework.

Algorithms for learning sparse graphical model structures have a rich tradition in previous literature. For general MRFs, learning the underlying graph structures is known to be NP-hard [12]. However, in the case when the underlying graph is a tree, the classical Chow-Liu algorithm [13] offers an efficient approach to structure estimation. Recent results contribute to an extensive understanding of the Chow-Liu algorithm. The authors in [14] analyzed the error exponent and showed experimental results for chain graphs and star graphs. For pairwise binary MRFs with bounded maximum degree, [6] provides sufficient conditions for for correct graph selection. Similar achievability results for Ising models are in [8].

For applications in electric grids, based on a similar linear system as in (3), [15] uses regression to recover the symmetric graph parameters (which is the admittance matrix in the power network) where the matrix $\mathbf{B}$ is of full column rank, implying that at least $m = \Omega(n)$ measurements are necessary. Compressed sensing ([16, 17]), however suggests that recovering the graph matrix may take much fewer number of measurements by fully utilizing the sparsity of $\mathbf{Y}$. Some experimental results for recovering topology of a power network based on compressed sensing algorithms are reported in [18]. Nonetheless, in the worst case, some of the columns (or rows) of $\mathbf{Y}$ may be dense vectors consisting of many non-zeros, prohibiting us from applying compressed sensing algorithms to recover each of the columns (or rows) of $\mathbf{Y}$ separately. Moreover, the columns to be recovered may not share the same support set. Thus many distributed compressed sensing schemes (*cf.* [19]) are not directly applicable in this situation. This motivates us to handle the difficulty that for a randomly chosen graph, some of the columns (or rows) in the corresponding graph matrix may not be sparse by considering a new two-stage recovery scheme. Based on single-type measurements (either current or voltage), correlation analysis has been applied for topology identification [20, 21, 22]. Approximating the measurements as normal distributed random variables, [23] proposed an approach for topology identification

[1]This means the entries of the matrix $\mathbf{B}$ are IID normally distributed.

with limited measurements. A graphical learning-based approach was provided by [24]. Recently, data-driven methods were studied for parameter estimation [25].

1.3 Our Contributions

We demonstrate that the linear system in (1) can be used to learn the topology and parameters of a graph. Our framework can be applied to perform system identification in electrical grids by leveraging synchronous nodal current and voltage measurements obtained from phasor measurement units (PMUs).

The main results of this paper are summarized here.

1. *Fundamental Trade-offs*: In Theorem 3.1, we derive a general lower bound on the *probability of error* for topology identification (defined in (4)). In Section 3.2, we describe a simple two-stage recovery scheme combining ℓ_1-norm minimization with an additional step called *consistency-checking*. For any arbitrarily chosen distribution, we characterize it using the definition of (λ, K)-*sparsity* (see Definition 3.1) and argue that if a graph is drawn according to such a distribution, then the number of measurements required for exact recovery is bounded from above as in Theorem 3.2.

2. *(Worst-case) Sample Complexity*: We focus on the case when the matrix $\mathbf{B}$ has Gaussian IID entries in Section 4. Under this assumption, we provide upper and lower bounds on the worst-case sample complexity in Theorem 4.2. We show two applications of Theorem 4.2 for the uniform sampling of trees and the Erdős–Rényi (n, p) model in Corollary 4.1 and 4.2, respectively.

3. *(Heuristic) Algorithm*: Motivated by the two-stage recovery scheme, a heuristic algorithm with polynomial (in n) running-time is reported in Section 6, together with simulation results for power system test cases validating its performance in Section 7.

1.4 Outline of the Paper

The remaining content is organized as follows. In Section 2, we specify our models. In Section 3.1, we present the converse result as fundamental limits for recovery. The achievability is provided in 3.2. We present our main result as the worst-case sample complexity for Gaussian IID measurements in Section 4. A heuristic algorithm together with simulation results are reported in Section 6 and 7.

2 Model and Definitions

2.1 Notation

Let $\mathbb{F}$ denote a field that can either be the set of real numbers $\mathbb{R}$, or the set of complex numbers $\mathbb{C}$. The set of all symmetric $n \times n$ matrices whose entries are in $\mathbb{F}$ is denoted by $\mathbb{S}^{n\times n}$. The imaginary unit is denoted by j. Throughout the work, let $\log(\cdot)$ denote the binary logarithm with base 2 and let $\ln(\cdot)$ denote the natural logarithm with base e. We use $\mathbb{E}[\cdot]$ to denote the expectation of random variables if the underlying probability distribution is clear. The mutual information is denoted by $\mathbb{I}(\cdot)$. The (differential) entropy is denoted by $\mathbb{H}(\cdot)$ and in particular, we use $h(\cdot)$ for binary entropy. To distinguish random variables and their realizations, we follow the convention and denote the former by capital letters (*e.g.,* A) and the latter by lower case letters (*e.g.,* a). The symbol C is used to designate a constant.

Matrices are denoted in boldface (*e.g.,* $\mathbf{A}$, $\mathbf{B}$ and $\mathbf{Y}$). The i-th row, the j-th column and the (i, j)-th entry of a matrix $\mathbf{A}$ are denoted by $A^{(i)}$, A_j and $A_{i,j}$ respectively. For notational convenience, let $\mathcal{S}$ be a subset

of $\mathcal{N}$. Denote by $\mathcal{S}^{\mathrm{c}} := \mathcal{N}\backslash\mathcal{S}$ the complement of $\mathcal{S}$ and by $\mathbf{A}^{\mathcal{S}}$ a sub-matrix consisting of $|\mathcal{S}|$ columns of the matrix $\mathbf{A}$ whose indices are chosen from $\mathcal{S}$. The notation $\top$ denotes the transpose of a matrix, $\det(\cdot)$ calculates its determinant, and $\mathrm{supp}(\cdot)$ is its support. For the sake of notational simplicity, we use big O notation (o,ω,O,Ω,Θ) to quantify asymptotic behavior. Table 1 summarizes the notation used throughout the paper.

Table 1: List of commonly appeared notations, symbols and acronyms in this paper.

Graph Model	
n	Number of nodes
d_j	Degree of node j
$\mathcal{N}$	Set of n nodes
$\mathcal{E}$	Set of edges
G	Underlying graph
G	Candidacy set consisting of all possible graphs
$\mathsf{G}_{\mathrm{All}}$	Set of all graphs with n vertices
$\mathsf{T}_{\mathrm{All}}$	Set of all trees with n vertices
T_H	Set of all spanning trees of a sub-graph H
$\mathcal{G}$	Probability distribution of G in G
$\mathbf{Y}$	$n \times n$ symmetric matrix to be recovered
Measurements	
m	Number of measurements
$\mathcal{M}$	Set of measurement indexes
$\mathbf{A}, \mathbf{B}$	Matrices of Measurements
$A^{(i)}$	i-th row of the matrix $\mathbf{A}$
A_j	j-th column of the matrix $\mathbf{A}$
$A_{i,j}$	(i, j)-th entry of the matrix $\mathbf{A}$
$\mathbf{Z}$	Additive output noise
Other Nomenclature	
$\mathbb{R}$	Set of real numbers
$\mathbb{C}$	Set of complex numbers
$\mathbb{F}$	Either $\mathbb{R}$ or $\mathbb{C}$
j	Imaginary unit
supp	Support set of a vector

2.2 Graphical Model

Denote by $\mathcal{N} = \{1, \ldots, n\}$ a set of n nodes and consider an *undirected* graph $G = (\mathcal{N}, \mathcal{E})$ (with no self-loops) whose edge set $\mathcal{E} \subseteq \mathcal{N} \times \mathcal{N}$ contains the desired topology information. The degree of each node j is denoted by d_j. The connectivity between the nodes is unknown and our goal is to determine it by learning the associated *graph matrix* using linear measurements.

Definition 2.1 (Graph Matrix)**.** Provided with an underlying graph $G = (\mathcal{N}, \mathcal{E})$, a *symmetric* matrix $\mathbf{Y}(G) \in \mathbb{S}^{n\times n}$ is called a *graph matrix* if the following conditions hold:

$$Y(G)_{i,j} = \begin{cases} \neq 0 & \text{if } i \neq j \text{ and } (i,j) \in \mathcal{E} \\ 0 & \text{if } i \neq j \text{ and } (i,j) \notin \mathcal{E} \\ \text{arbitrary} & \text{otherwise} \end{cases}.$$

Remark 1. Our theorems can be generalized to recover a broader class of symmetric matrices, as long as the matrix to be recovered satisfies (1) Knowing $\mathbf{Y}(G) \in \mathbb{F}^{n\times n}$ gives the full knowledge of the topology of G; (2) The number of non-zero entries in a column of $\mathbf{Y}(G)$ has the same order as the degree of the corresponding

node, *i.e.,* there is a positive constant $C > 0$ such that $|\mathrm{supp}(Y_j)| = Cd_j$. for all $j \in \mathcal{N}$. To have a clear presentation, we consider specifically the case $C = 1$.

In this work, we employ a probabilistic model and assume that the graph G is chosen randomly from a *candidacy set* G, according to some distribution $\mathcal{G}$. Both the candidacy set G and distribution $\mathcal{G}$ are not known to the estimator.

Example 2.1. We exemplify some possible choices of the candidacy set and distribution:

(a) *(Mesh Network)* When G represents a transmission (mesh) power network and no prior information is available, the corresponding candidacy set $\mathsf{G}_{\mathrm{All}}$ consisting of all graphs with n nodes and G is selected uniformly at random from $\mathsf{G}_{\mathrm{All}}$. Moreover, $|\mathsf{G}_{\mathrm{All}}| = 2^{\binom{n}{2}}$ in this case.

(b) *(Radial Network)* When G represents a distribution (radial) power network and no other prior information available, then the corresponding candidacy set $\mathsf{T}_{\mathrm{All}}$ is a set containing all spanning trees of the complete graph with n buses and G is selected uniformly at random from $\mathsf{T}_{\mathrm{All}}$; the cardinality is $|\mathsf{T}_{\mathrm{All}}| = n^{n-2}$ followed by Cayley's formula.

(c) *(Radial Network with Prior Information)* When $G = (\mathcal{N}, \mathcal{E})$ represents a distribution (radial) power network, and we further know that some of the buses cannot be connected (which may be inferred from locational/geographical information), then the corresponding candidacy set T_H is a set of spanning trees of a sub-graph $H = (\mathcal{N}, \mathcal{E}_H)$ with n buses. An edge $e \notin \mathcal{E}_H$ if and only if we know $e \notin \mathcal{E}$. The size of T_H is given by Kirchhoff's matrix tree theorem (c.f. [26]). See Theorem 5.1.

(d) *(Erdős–Rényi (n, p) model)* In a more general setting, G can be a random graph chosen from an ensemble of graphs according to a certain distribution. When a graph G is sampled according to the Erdős–Rényi (n, p) model, each edge of G is connected IID with probability p. We denote the corresponding graph distribution for this case by $\mathcal{G}_{\mathrm{ER}}(n, p)$ for convenience.

The next section is devoted to describing available measurements.

2.3 Linear System of Measurements

Suppose the measurements are sampled discretely and indexed by the elements of the set $\{1, \ldots, m\}$. As a general framework, the measurements are collected in two matrices $\mathbf{A}$ and $\mathbf{B}$ and defined as follows.

Definition 2.2 (Generator and Measurement Matrices)**.** Let m be an integer with $1 \le m \le n$. The *generator matrix* $\mathbf{B}$ is an $m \times n$ *random* matrix and the *measurement matrix* $\mathbf{A}$ is an $m \times n$ matrix with entries selected from $\mathbb{F}$ that satisfy the linear system (1):

$$\mathbf{A} = \mathbf{B}\mathbf{Y}(G) + \mathbf{Z}$$

where $\mathbf{Y}(G) \in \mathbb{S}^{n \times n}$ is a graph matrix to be recovered, with an underlying graph G and $\mathbf{Z} \in \mathbb{F}^{m \times n}$ denotes the random *additive noise*. We call the the recovery *noiseless* if $\mathbf{Z} = \mathbf{0}$. Our goal is to resolve the matrix $\mathbf{Y}(G)$ based on given matrices $\mathbf{A}$ and $\mathbf{B}$.

2.4 Applications to Electrical Grids

Various applications fall into the framework in (1). Here we present two examples of the graph identification problem in power systems. The measurements are modeled as time series data obtained via nodal sensors at each node, *e.g.,* PMUs, smart switches, or smart meters.

2.4.1 Example 1: Nodal Current and Voltage Measurements

We assume data is obtained from a short time interval over which the unknown parameters in the network are *time-invariant*. $\mathbf{Y} \in \mathbb{C}^{n\times n}$ denotes the *nodal admittance matrix* of the network and is defined

$$Y_{i,j} := \begin{cases} -y_{i,j} & \text{if } i \neq j \\ y_i + \sum_{k\neq i} y_{i,k} & \text{if } i = j \end{cases} \tag{2}$$

where $y_{i,j} \in \mathbb{C}$ is the admittance of line $(i, j) \in \mathcal{E}$ and y_i is the self-admittance of bus i. Note that if two buses are not connected then $Y_{i,j} = 0$.

The corresponding generator and measurement matrices are formed by simultaneously measuring both currents (or equivalently, power injections) and voltages at each node and at each time step. For each $t = 1, \ldots, m$, the nodal current injections are collected in an n-dimensional random vector $I_t = (I_{t,1}, \ldots, I_{t,n})$. Concatenating the I_t into a matrix we get $\mathbf{I} := [I_1, I_2, \ldots, I_m]^\top \in \mathbb{C}^{m\times n}$. The generator matrix $\mathbf{V} := [V_1, V_2, \ldots, V_m]^\top \in \mathbb{C}^{m\times n}$ is constructed analogously. Each pair of measurement vectors (I_t, V_t) from $\mathbf{I}$ and $\mathbf{V}$ must satisfy Kirchhoff's and Ohm's laws,

$$I_t = \mathbf{Y}V_t, \quad t = 1, \ldots, m. \tag{3}$$

In matrix notation (3) is equivalent to $\mathbf{I} = \mathbf{VY}$, which is a noiseless version of the linear system defined in (1).

Compared with only obtaining one of the current, power injection or voltage measurements (for example, as in [20, 14, 21]), collecting simultaneous current-voltage pairs doubles the amount of data to be acquired and stored. There are benefits however. First, exploiting the physical law relating voltage and current not only enables us to identify the topology of a power network but also recover the parameters of the admittance matrix. Furthermore, dual-type measurements significantly reduce the sample complexity for topology recovery, compared with the results for single-type measurements.

2.4.2 Example 2: Nodal Power Injections and Phase Angles

Similar to the previous example, at each time $t = 1, \ldots, m$, denote by $P_{t,j}$ and $\theta_{t,j}$ the active nodal power injection and the phase of voltage at node j respectively. The matrices $\mathbf{P} \in \mathbb{R}^{m\times n}$ and $\boldsymbol{\theta} \in \mathbb{R}^{m\times n}$ are constructed in a similar way by concatenating the vectors $P_t = (P_{t,1}, \ldots, P_{t,n})$ and $\theta_t = (\theta_{t,1}, \ldots, \theta_{t,n})$. The matrix representation of the DC power flow model can be expressed as a linear system $\mathbf{P} = \boldsymbol{\theta}\mathbf{CSC}^\top$, which belongs to the general class represented in (1). Here, the diagonal matrix $\mathbf{S} \in \mathbb{R}^{|\mathcal{E}|\times|\mathcal{E}|}$ is the susceptence matrix whose e-th diagonal entry represents the susceptence on the e-th edge in $\mathcal{E}$ and $\mathbf{C} \in \{-1, 0, 1\}^{n\times|\mathcal{E}|}$ is the node-to-link incidence matrix of the graph. The vertex-edge incidence matrix[2] $\mathbf{C} \in \{-1, 0, 1\}^{n\times|\mathcal{E}|}$ is defined as

$$C_{j,e} := \begin{cases} 1, & \text{if bus } j \text{ is the source of } e \\ -1, & \text{if bus } j \text{ is the target of } e \\ 0, & \text{otherwise} \end{cases}.$$

Note that $\mathbf{CSC}^\top$ specifies both the network topology and the susceptences of power lines.

[2]Although the underlying network is a directed graph, when considering the fundamental limit for topology identification, we still refer to the recovery of an undirected graph G.

2.5 Probability of Error as the Recovery Metric

We define the error criteria considered in this paper. We refer to finding the edge set $\mathcal{E}$ of G via matrices $\mathbf{A}$ and $\mathbf{B}$ as the *topology identification problem* and recovering the graph matrix $\mathbf{Y}$ via matrices $\mathbf{A}$ and $\mathbf{B}$ as the *parameter reconstruction problem*.

Definition 2.3. Let f be a function or algorithm that returns an estimated graph matrix $\mathbf{X} = f(\mathbf{A}, \mathbf{B})$ given inputs $\mathbf{A}$ and $\mathbf{B}$. The *probability of error for topology identification* $\mathbb{P}_{\text{error}}^{\text{TI}}$ is defined to be the probability that the estimated edge set is not equal to the correct edge set:

$$\mathbb{P}_{\text{error}}^{\text{TI}} := \mathbb{P}\left(\text{supp}(\mathbf{X}) \neq \text{supp}\left(\mathbf{Y}(G)\right)\right) \tag{4}$$

where the probability is taken over the randomness in G, $\mathbf{B}$, and $\mathbf{Z}$. The *probability of error for (noiseless[3]) parameter reconstruction* $\mathbb{P}_{\text{error}}^{\text{PR}}$ is defined to be the probability that the estimate $\mathbf{X}$ is not equal to the original graph matrix $\mathbf{Y}(G)$:

$$\mathbb{P}_{\text{error}}^{\text{PR}} := \sup_{\mathbf{Y} \in \mathbb{Y}(G)} \mathbb{P}\left(\mathbf{X} \neq \mathbf{Y}(G)\right) \tag{5}$$

where $\mathbb{Y}(G)$ is the set of all graph matrices $Y(G)$ that are consistent with the underlying graph G and the probability is taken over the randomness in G and $\mathbf{B}$.

Remark 2. Note that for a fixed noiseless parameter reconstruction algorithm, we always have the corresponding $\mathbb{P}_{\text{error}}^{\text{PR}}$ greater than $\mathbb{P}_{\text{error}}^{\text{TI}}$. We use $\mathbb{P}_{\text{error}}^{\text{PR}}$ as the error metric in this work and refer it as the *probability of error* considered in the remainder of this paper.

3 Fundamental Trade-offs

We discuss fundamental trade-offs of the noiseless parameter recovery problem defined in Section 2.2 and 2.3. The converse result is summarized in Theorem 3.1 as an inequality involving the probability of error, the distributions of the underlying graph, generator matrix and noise. Next, in Section 3.2, we focus on a particular two-stage scheme, and show in Theorem 3.2 that under certain conditions, the probability of error is asymptotically zero (in n).

3.1 Converse

The following theorem states the fundamental limit.

Theorem 3.1 (Converse)**.** *The probability of error for topology identification* $\mathbb{P}_{\text{error}}^{\text{TI}}$ *is bounded from below as*

$$\mathbb{P}_{\text{error}}^{\text{TI}} \geq 1 - \frac{\mathbb{H}(\mathbf{B}) - \mathbb{H}(\mathbf{Z}) + \ln 2}{\mathbb{H}(\mathcal{G})} \tag{6}$$

where $\mathbb{H}(\mathbf{B})$*,* $\mathbb{H}(\mathbf{Z})$ *and* $\mathbb{H}(\mathcal{G})$ *are differential entropy (in base e) functions of the random variables* $\mathbf{B}$*,* $\mathbf{Z}$ *and probability distribution* $\mathcal{G}$*, respectively.*

Proof. The graph G is chosen from a discrete set G according to some probability distribution $\mathcal{G}$. As previously introduced, Fano's inequality [7] borrowed plays an important role in deriving fundamental limits.

[3]In this exploratory work, we assume the measurements are noiseless and algorithms seek to recover each entry of the graph matrix *exactly*. When the measurements are noisy, Theorem 3.1 provides general converse results as trade-offs between the number of measurement needed and the probability of error defined in (5).

We especially focus on its extended version. Similar generalizations appear in many places, *e.g.,* [11, 6] and [10].

Lemma 1 (Generalized Fano's inequality)**.** *Let G be a random graph and let $\mathbf{A}$ and $\mathbf{B}$ be measurement matrices defined in Section 2.2 and 2.3. Suppose the original graph G is selected from a nonempty candidacy set G according to a probability distribution $\mathcal{G}$. Let $\hat{G}$ denote the estimated graph. Then the conditional probability of error for estimating G from $\mathbf{B}$ given $\mathbf{A}$ is always bounded from below as*

$$\mathbb{P}\left(\hat{G} \neq G \middle| \mathbf{A}\right) \geq 1 - \frac{\mathbb{I}\left(G; \mathbf{B} \middle| \mathbf{A}\right) + \ln 2}{\mathbb{H}\left(\mathcal{G}\right)} \tag{7}$$

where the randomness is over the selections of the original graph G and the estimated graph $\hat{G}$.

In (7), the term $\mathbb{I}\left(G; \mathbf{B} \middle| \mathbf{A}\right)$ denotes the conditional mutual information (base e) between G and $\mathbf{B}$ conditioned on $\mathbf{A}$, which is defined as

$$\mathbb{I}\left(G; \mathbf{B} \middle| \mathbf{A}\right) := \sum_{G \in \mathsf{G}} \int_{\mathbf{B}} \int_{\mathbf{A}} p\left(\mathbf{A}, \mathbf{B}, G\right) \ln \frac{p\left(\mathbf{B}|\mathbf{A}, G\right)}{p\left(\mathbf{B}|\mathbf{A}\right)} \mathrm{d}\mathbf{A}\mathrm{d}\mathbf{B}$$

where the integrals are both taken over $\mathbb{F}^{n \times m}$. Furthermore, the conditional mutual information $\mathbb{I}\left(G; \mathbf{B}|\mathbf{A}\right)$ is bounded from above by the differential entropies of $\mathbf{B}$ and $\mathbf{A}$. It follows that

$$\begin{aligned}
\mathbb{I}\left(G; \mathbf{B}|\mathbf{A}\right) &= \mathbb{H}\left(\mathbf{B}|\mathbf{A}\right) - \mathbb{H}\left(\mathbf{B}|G, \mathbf{A}\right) && (8)\\
&\leq \mathbb{H}\left(\mathbf{B}|\mathbf{A}\right) - \mathbb{H}\left(\mathbf{B}|\mathbf{Y}, \mathbf{A}\right) && (9)\\
&= \mathbb{H}\left(\mathbf{B}|\mathbf{A}\right) - \mathbb{H}\left(\mathbf{Z}\right) && (10)\\
&\leq \mathbb{H}\left(\mathbf{B}\right) - \mathbb{H}\left(\mathbf{Z}\right). && (11)
\end{aligned}$$

Here, Eq. (8) follows from the definitions of mutual information and differential entropy. Moreover, knowing $\mathbf{Y}$, the graph G can be inferred. Thus, $\mathbb{H}\left(\mathbf{B}|G, \mathbf{A}\right) \geq \mathbb{H}\left(\mathbf{B}|\mathbf{Y}, \mathbf{A}\right)$ yields (9). Recalling the linear system in (1), we obtain (10). Furthermore, (11) holds since $\mathbb{H}\left(\mathbf{B}\right) \geq \mathbb{H}\left(\mathbf{B}|\mathbf{A}\right)$.

Plugging (11) into (7),

$$\begin{aligned}
\mathbb{P}_{\text{error}}^{\text{PR}} \geq \mathbb{P}_{\text{error}}^{\text{TI}} =& \mathbb{E}_{\mathbf{A}}\left[\mathbb{P}\left(\hat{G} \neq G \middle| \mathbf{A}\right)\right]\\
\geq& 1 - \frac{\mathbb{H}\left(\mathbf{B}\right) - \mathbb{H}\left(\mathbf{Z}\right) + \ln 2}{\mathbb{H}\left(\mathcal{G}\right)},
\end{aligned}$$

which yields the desired (6). □

3.2 Achievability

In this subsection, we consider the achievability for *noiseless* parameter reconstruction. The proofs rely on constructing a two-stage recovery scheme (Algorithm 1), which contains two steps – *column-retrieving* and *consistency-checking*. The worst-case running time of this scheme depends on the underlying distribution $\mathcal{G}$[4]. The scheme is presented as follows.

[4]Although for certain distributions, the computational complexity is not polynomial in n, the scheme still provides insights on the fundamental trade-offs between the number of samples and the probability of error for recovering graph matrices. Furthermore, motivated by the scheme, a polynomial-time heuristic algorithm is provided in Section 6 and experimental results are reported in Section 7.

Specification:

Data: Matrices of measurements $\mathbf{A}$ and $\mathbf{B}$
Result: Estimated graph matrix $\mathbf{X}$

Recovering columns independently:

for $j \in \mathcal{N}$ **do**
  Solve the following ℓ_1-minimization and obtain an optimal $\mathbf{X}$:

$$\text{minimize} \quad \left\|X_j\right\|_{\ell_1} \tag{12}$$
$$\text{subject to } \mathbf{B}X_j = A_j,$$
$$X_j \in \mathbb{F}^n. \tag{13}$$

end

Consistency-checking:

for $\mathcal{S} \subseteq \mathcal{N}$ *with* $|\mathcal{S}| = n - K$ **do**
  for $i, j \in \mathcal{S}$ **do**
    if $X_{i,j} \neq X_{j,i}$ **then**
      break;
    end
  end
  for $j \in \mathcal{S}^{\mathrm{c}}$ **do**
    Update $X_j^{\mathcal{S}^{\mathrm{c}}}$ by solving the linear system:

$$\mathbf{B}^{\mathcal{S}^{\mathrm{c}}} X_j^{\mathcal{S}^{\mathrm{c}}} = A_j - \mathbf{B}^{\mathcal{S}} X_j^{\mathcal{S}}. \tag{14}$$

  end
  return $\mathbf{X} = (X_1, \ldots, X_n)$;
end

Algorithm 1: A two-stage recovery scheme. The first stage focuses on solving each column of the matrix $\mathbf{Y}$ independently using ℓ_1-minimization. In the second stage, the recovery correctness of the first stage is further verified via *consistency-checking*, which utilizes the fact that the matrix to be recovered $\mathbf{Y}$ is *symmetric*.

3.2.1 Two-stage Recovery Scheme

Retrieving columns In the first stage, using ℓ_1-norm minimization, we recover each column of $\mathbf{Y}$ based on (1) (with no noise):

$$\text{minimize} \quad \left\|X_j\right\|_{\ell_1} \tag{15}$$
$$\text{subject to} \quad \mathbf{B}X_j = A_j, \tag{16}$$
$$X_j \in \mathbb{F}^n. \tag{17}$$

Let $X_j^{\mathcal{S}} := (X_{i,j})_{i \in \mathcal{S}}$ be a length-$|\mathcal{S}|$ column vector consisting of $|\mathcal{S}|$ coordinates in X_j, the j-th retrieved column. We do not restrict the methods for solving the ℓ_1-norm minimization in (15)-(17), as long as there is a unique solution for sparse columns with fewer than λ non-zeros (the parameter $\lambda > 0$ is defined in Definition 3.1 below).

Checking consistency In the second stage, we check for error in the decoded columns $X_1, \ldots, X_n$ using the symmetry property of the graph matrix $\mathbf{Y}$. Specifically, we fix a subset $\mathcal{S} \subseteq \mathcal{N}$ with a given size $|\mathcal{S}| = n - K$ for some integer $0 \leq K \leq n$. Then we check if $X_{i,j} = X_{j,i}$ for all $i, j \in \mathcal{S}$. If not, we choose a different set $\mathcal{S}$ of the same size. This procedure stops until either we find such a subset $\mathcal{S}$ of columns, or we go through all possible subsets without finding one. In the latter case, an error is declared and the recovery is unsuccessful. In the former case, we accept $X_j, j \in \mathcal{S}$, as correct. For each vector $X_j, j \notin \mathcal{S}$, we accept its entries $X_{i,j}, i \in \mathcal{S}$, as correct and use them to compute the other entries $X_{i,j}, i \notin \mathcal{S}$, of X_j using (16):

$$\mathbf{B}^{\mathcal{S}^c} X_j^{\mathcal{S}^c} = A_j - \mathbf{B}^{\mathcal{S}} X_j^{\mathcal{S}}, \quad j \in \mathcal{S}^c. \tag{18}$$

We combine $X_j^{\mathcal{S}}$ and $X_j^{\mathcal{S}^c}$ to obtain a new estimate X_j for each $j \in \mathcal{S}^c$. Together with the columns X_j, $j \in \mathcal{S}$, that we have accepted, they form the estimated graph matrix $\mathbf{X}$.

3.2.2 (λ, K)-sparse Distribution

We now analyze the sample complexity of the two-stage scheme. Let $d_j(G)$ denote the degree of node j in G. Denote by $\mathcal{N}_{\text{Large}}(\lambda)$ the set of nodes having degrees greater than the *threshold parameter* $0 \leq \lambda \leq n - 2$:

$$\mathcal{N}_{\text{Large}}(\lambda) := \Big\{ j \in \mathcal{N} : d_j(G) > \lambda \Big\} \tag{19}$$

and $\mathcal{N}_{\text{Small}}(\lambda) := \mathcal{N} \backslash \mathcal{N}_{\text{Large}}(\lambda)$. Making use of (19), we define the following set of graphs, with a *counting parameter* $0 \leq K \leq n$:

$$\mathsf{G}(\lambda, K) := \Big\{ G \in \mathsf{G} : \big|\mathcal{N}_{\text{Large}}(\lambda)\big| \leq K \Big\}.$$

The following definition characterizes graph distributions.

Definition 3.1 (Sparse Distribution)**.** A distribution $\mathcal{G}$ of graphs in G is said to be (λ, K)-*sparse* if assuming $G \sim \mathcal{G}$,

$$\lim_{n \to \infty} \mathbb{P}_{\mathcal{G}}\big(G \in \mathsf{G}(\lambda, K)\big) = 1. \tag{20}$$

The following lemmas provide examples of sparse distributions. Denote by $\mathcal{U}_{\mathsf{T}_{\text{All}}}$ the uniform distribution on the set T_{All} of all trees with n nodes.

Lemma 2. *For any $\lambda > 1$ and $K \geq \ln n$, the distribution $\mathcal{U}_{\mathsf{T}_{\text{All}}}$ is (λ, K)-sparse.*

Denote by $\mathcal{G}_{\text{ER}}(n, p)$ the graph distribution for the Erdős–Rényi (n, p) model.

Lemma 3. *For any $\lambda \geq 2nh(p)/(\ln 1/p)$ and K satisfying $nh(p) = \omega(\log(n/K))$, the distribution $\mathcal{G}_{\text{ER}}(n, p)$ is (λ, K)-sparse.*

The threshold and counting parameters for both examples are tight, as indicated in Corollary 4.1 and 4.2. The proofs of Lemma 2 and 3 are postponed to Appendix .2.

3.2.3 Analysis of the Scheme

We now present another of our main theorems, which makes use of the restricted isometry property (*cf.*, [16, 17]). Given a generator matrix $\mathbf{B}$, the corresponding *restricted isometry constant* denoted by σ_λ is the

smallest positive number with

$$C\,(1-\sigma_\lambda)\,||\mathbf{x}||_{\ell_2}^2 \le \left\|\mathbf{B}^{\mathcal{S}}\mathbf{x}\right\|_{\ell_2}^2 \le C\,(1+\sigma_\lambda)\,||\mathbf{x}||_{\ell_2}^2$$

for some constant $C > 0$ and for all subsets $\mathcal{S} \subseteq \mathcal{N}$ of size $|\mathcal{S}| \le \lambda$ and all $\mathbf{x} \in \mathbb{F}^{|\mathcal{S}|}$.

Denote by spark($\mathbf{B}$) the smallest number of columns in the matrix $\mathbf{B}$ that are linearly dependent (see [27] for the requirements on the spark of the generator matrix to guarantee desired recovery criteria). Consider the models defined in Section 2.2 and 2.3.

Theorem 3.2 (Achievability)**.** *Suppose the generator matrix* $\mathbf{B}$ *has restricted isometry constants* $\sigma_{3\lambda}$ *and* $\sigma_{4\lambda}$ *satisfying* $\sigma_{3\lambda} + 3\sigma_{4\lambda} < 2$ *and furthermore,* $m \ge \text{spark}(\mathbf{B}) > 2K$*. If* $\mathcal{G}$ *is a* (λ, K)*-sparse distribution satisfying (20), then the two-stage scheme always recovers a graph matrix* $\mathbf{Y}(G)$ *of* $G \sim \mathcal{G}$ *with a vanishing probability of error* $\lim_{n\to\infty} \mathbb{P}_{\text{error}}^{\text{PR}} = 0$*.*

Proof. First, the theory of compressed sensing (see [16, 17]) implies that if the generator matrix $\mathbf{B}$ has restricted isometry constants $\sigma_{3\lambda}$ and $\sigma_{4\lambda}$ satisfying $\sigma_{3\lambda} + 3\sigma_{4\lambda} < 2$, then all columns Y_j with $j \in \mathcal{N}_{\text{Small}}$ are correctly recovered using the minimization in (15)-(17). It remains to show that the consistency-check in our scheme works, which is summarized as the following lemma.

Lemma 4 (Consistency-check)**.** *Suppose the matrix* $\mathbf{B}$ *has restricted isometry constants* $\sigma_{3\lambda}$ *and* $\sigma_{4\lambda}$ *satisfying* $\sigma_{3\lambda} + 3\sigma_{4\lambda} < 2$*. Furthermore, suppose* $m \ge \text{spark}(\mathbf{B}) > 2K$*. If* $G \in \mathsf{G}(\lambda, K)$*, then the collection of columns* $\{X_j\}_{j\in\mathcal{S}}$ *passing the consistency-check such that* $X_{i,j} = X_{j,i}$ *for all* $i, j \in \mathcal{S}$*, are correctly decoded and together with (18), the two-stage scheme always returns the original (correct) graph matrix.*

The proof of Lemma 4 can be found in Appendix .1. Making use of Lemma 4, it follows that $\mathbb{P}_{\text{error}}^{\text{PR}} \le 1 - \mathbb{P}_{\mathcal{G}}(G \in \mathsf{G}(\lambda, K))$ provided $m \ge \text{spark}(\mathbf{B}) > 2K$. In agreement with the assumption that the graph distribution $\mathcal{G}$ is (λ, K)-sparse, (20) must be satisfied. Thus, the probability of error goes to zero as n goes to infinity. □

4 Gaussian IID Measurements

In this section, we consider a special regime when the measurements in the matrix $\mathbf{B}$ are Gaussian IID random variables. Utilizing the converse in Theorem 3.1 and the achievability in Theorem 3.2, the Gaussian IID assumption allows the derivation of explicit expressions of sample complexity as upper and lower bounds on the number of measurements m. Combining with the results in Lemma 2 and 3, we are able to show that for the corresponding lower and upper bounds match each other for graphs distributions $\mathcal{U}$ and $\mathcal{G}_{\text{ER}}(n, p)$ (with certain conditions on p and n).

For the convenience of presentation, in the remainder of the paper, we restrict that the measurements are chosen from $\mathbb{R}$, although the theorems can be generalized to the complex measurements. In realistic scenarios, for instance, a power network, besides the measurements collected from the nodes, nominal state values, *e.g.,* operating current and voltage measurements are known to the system designer a priori. Representing the nominal values at the nodes by $\overline{A} \in \mathbb{R}^n$ and $\overline{B} \in \mathbb{R}^n$ respectively, the measurements in $\mathbf{A}$

and $\mathbf{B}$ are centered around $m \times n$ matrices $\overline{\mathbf{A}}$ and $\overline{\mathbf{B}}$ defined as

$$\overline{\mathbf{A}} := \begin{bmatrix} \cdots & \overline{A} & \cdots \\ \cdots & \overline{A} & \cdots \\ & \vdots & \\ \cdots & \overline{A} & \cdots \end{bmatrix}, \quad \overline{\mathbf{B}} := \begin{bmatrix} \cdots & \overline{B} & \cdots \\ \cdots & \overline{B} & \cdots \\ & \vdots & \\ \cdots & \overline{B} & \cdots \end{bmatrix}.$$

The rows in $\mathbf{A}$ and $\mathbf{B}$ are the same, because the graph parameters are time-invariant, so are the nominal values. Without system fluctuations and noise, the nominal values satisfy the linear system in (1), *i.e.,*

$$\overline{\mathbf{A}} = \overline{\mathbf{B}}\mathbf{Y}. \tag{21}$$

Knowing $\overline{A}$ and $\overline{B}$ is not sufficient to infer the network parameters (the entries in the graph matrix $\mathbf{Y}$), since the rank of the matrix $\overline{B}$ is one. However, measurement fluctuations can be used to facilitate the recovery of $\mathbf{Y}$. The deviations from the nominal values are denoted by an additive perturbation matrices $\widetilde{\mathbf{A}}$ and $\widetilde{\mathbf{B}}$ such that $\mathbf{A} = \overline{\mathbf{A}} + \widetilde{\mathbf{A}}$. Similarly, $\mathbf{B} = \overline{\mathbf{B}} + \widetilde{\mathbf{B}}$ where $\widetilde{\mathbf{B}}$ is an $m \times n$ matrix consisting of additive perturbations. Thus, putting (3) and (21) together, the equations above imply that $\overline{\mathbf{A}} + \widetilde{\mathbf{A}} = \mathbf{B}\mathbf{Y} = \overline{\mathbf{B}}\mathbf{Y} + \widetilde{\mathbf{B}}\mathbf{Y}$ leading to $\widetilde{\mathbf{A}} = \widetilde{\mathbf{B}}\mathbf{Y}$ where the voltage and current perturbations can be extracted from the measurement matrices $\mathbf{B}$ and $\mathbf{A}$ and the known nominal matrices $\widetilde{\mathbf{B}}$ and $\widetilde{\mathbf{A}}$. In this exploratory work, we specifically consider the case when the additive perturbations $\widetilde{\mathbf{B}}$ is a matrix with Gaussian IID entries. Without loss of generality, we suppose the mean is zero and the variance is one. For simplicity, in the remainder of this paper, we slightly abuse the notation and replace the perturbations by matrices $\mathbf{B}$ (so that we assume that $\mathbf{B}$ is Gaussian IID), if the context is clear. Moreover, throughout this section, we focus on the case when the measurements are noiseless.

The next theorem implies that Gaussian IID random variables are not arbitrary selections. They are the most "informative" measurements in the sense that any measurement vector with fixed mean and covariance achieves the maximal entropy with normal distribution.

Theorem 4.1. *Suppose the row measurements of the generator matrix $\mathbf{B} \in \mathbb{R}^{n\times m}$ are identically distributed random vectors with zero mean and covariance $\mathbf{K} \in \mathbb{R}^{n\times n}$. The probability of error $\mathbb{P}_{\text{error}}^{\text{PR}}$ is bounded from below as*

$$\mathbb{P}_{\text{error}}^{\text{PR}} \geq 1 - \left[m \ln \left((2\pi e)^{2n} \det \mathbf{K} \right) + \ln 2 \right] / \mathbb{H}(\mathcal{G}) \tag{22}$$

for noiseless recovery where $\mathbb{H}(\mathcal{G})$ is the differential entropy (in base e) of the graph distribution $\mathcal{G}$.

Remark 3. It can be inferred from the theorem that the number of samples must be at least linear in n to ensure a small probability of error, the size of the graph, given that the graph, as a mesh network, is chosen uniformly at random from G_{All} (see Example 2.1 (a)). On the other hand, as corollaries, under the assumptions of a GIPM, $m = \Omega(\log n)$ is *necessary* for making the probability of error less or equal to $1/2$, if the graph is chosen uniformly at random from T_{All}; $m = \Omega(nh(p))$ is *necessary* if the graph is sampled according to $\mathcal{G}_{\text{ER}}(n, p)$, as in Examples 2.1 (b) and (c), respectively. The theorem can be generalized to complex measurements by adding additional multiplicative constants.

Proof. The proof is based on Theorem 3.1. The key fact used is that the entropy $\mathbb{H}(\mathbf{B})$ is maximized when $B^{(t)}$ is distributed normally with zero mean and covariance $\mathbf{K} \in \mathbb{R}^{n\times n}$, for all $t = 1, \ldots, m$,

$$\mathbb{H}(\mathbf{B}) \leq \sum_{t=1}^{m} \mathbb{H}\left(B^{(t)}\right) \leq \frac{1}{2} m \ln \left((2\pi e)^{2n} \det \mathbf{K} \right). \tag{23}$$

Substituting the above into Theorem 3.1 gives (22). □

4.1 Sample Complexity for Sparse Distributions

We consider the worst-case sample complexity for recovering graphs generated according to sparse distributions.

Theorem 4.2 (Worst-case Sample Complexity)**.** *Suppose that the generator matrix* $\mathbf{B}$ *has Gaussian IID entries with mean zero and variance one and assume* $\lambda < n^{-3/\lambda}(n-K)$*,* $K = o(n)$*. For any* (λ, K)*-sparse distribution, the two-stage scheme guarantees that* $\mathbb{P}_{\text{error}}^{\text{PR}} = o(1)$ *(in* n*) using* $m = O(\lambda \log(n/\lambda) + K)$ *measurements. Conversely, there exists a* (λ, K)*-sparse distribution such that the number of measurements must satisfy* $m = \Omega\left(\lambda \log(n/\lambda) + K/n^{-3/\lambda}\right)$ *to make the probability of error* $\mathbb{P}_{\text{error}}^{\text{PR}}$ *less than* $1/2$*.*

Proof. The first part is based on Theorem 3.2. Under the assumption of the generator matrix $\mathbf{B}$, using Gordon's escape-through-the-mesh theorem, Theorem 4.3 in [17] implies that for any columns Y_j with $j \in \mathcal{N}_{\text{Small}}$ are correctly recovered using the minimization in (15)-(17) with probability at least $1-2.5\exp(-(4/9)\lambda\log(n/\lambda))$, as long as the number of measurements satisfies $m \geq 48\lambda(3+2\log(n/\lambda))$, and $n/\lambda > 2, \lambda \geq 4$ (if $\lambda \leq 3$, the multiplicative constant increases but our theorem still holds). Similar results were first proved by Candes, *et al.* in [16] (see their Theorem 1.3). Therefore, applying the union bound, the probability that all the λ-sparse columns can be recovered simultaneously is at least $1 - 2.5n\exp(-(4/9)\lambda\log(n/\lambda))$. On the other hand, conditioned on that all the λ-sparse columns are recovered, Theorem 3.2 shows that $m \geq \text{spark}(\mathbf{B}) > 2K$ is sufficient for the two-stage scheme to succeed. Since each entry in $\mathbf{B}$ is an IID Gaussian random variable with zero mean and variance one, if $m \geq 48\lambda(3+2\log(n/\lambda)) + 2K$, with probability one that the spark of $\mathbf{B}$ is greater than $2K$, verifying the statement.

The converse follows directly from Theorem 4.1. Consider the uniform distribution $\mathcal{U}_{\mathsf{G}(\lambda,K)}$ on $\mathsf{G}(\lambda, K)$. Then $\mathbb{H}\left(\mathcal{U}_{\mathsf{G}(\lambda,K)}\right) = \ln|\mathsf{G}(\lambda,K)|$. Let $0 \leq \alpha, \beta \leq 1$ be parameters such that $\lambda < \beta(n-\alpha K)$. To bound the size of $\mathsf{G}(\lambda, K)$, we partition $\mathcal{N}$ into $\mathcal{N}_1$ and $\mathcal{N}_2$ with $|\mathcal{N}_1| = n - \alpha K$ and $|\mathcal{N}_2| = \alpha K$. First, we assume that the nodes in $\mathcal{N}_1$ form a $\lambda/2$-regular graph. For each node in $\mathcal{N}_2$, construct $\beta(n-\alpha K) \in \mathbb{N}_+$ edges and connect them to the other nodes in $\mathcal{N}$ with uniform probability. A graph constructed in this way always belongs to $\mathsf{G}(\lambda, K)$, unless the added edges create more than αK nodes with degrees larger than λ. Therefore, as $n \to \infty$,

$$|\mathsf{G}(\lambda,K)| \geq \rho \cdot \frac{e^{1/4}\binom{N-1}{\phi}^N \binom{\binom{N}{2}}{\phi N/2}}{\binom{N(N-1)}{\phi N}} \cdot \binom{n-1}{M}^{\alpha K} \tag{24}$$

where $N := n - \alpha K$, $M = \beta(n-\alpha K)$ and $\phi := \lambda/2$. The first term ρ denotes the fraction of the constructed graphs that are in $\mathsf{G}(\lambda, K)$. The second term in (24) counts the total number of ϕ-regular graphs [28], and the last term is the total number of graphs created by adding new edges for the nodes in $\mathcal{N}_2$. If $K = O(\lambda)$, there exists a constant $\alpha > 0$ small enough such that $\rho = 1$. If $\lambda = o(K)$, for any fixed node in $\mathcal{N}_1$, the probability that its degree is larger than λ is

$$\sum_{i=\phi+1}^{\alpha K}\binom{\alpha K}{i}\beta^i(1-\beta)^{\alpha K-i} \leq \sum_{i=\phi+1}^{\alpha K}\alpha K h\left(\frac{i}{\alpha K}\right)\beta^i \leq (\alpha K)^2\beta^{\phi+1}$$

where $h(i/\alpha K)$ is in base e. Take $\beta = n^{-3/\lambda}$ and $\alpha = 1/2$. The condition $\lambda < n^{-3/\lambda}(n-K)$ guarantees that

$\lambda < \beta(n - \alpha K)$. Letting $F(n) := 1/n$, we check that

$$(\alpha K)^2 \beta^{\phi+1} \leq \frac{1}{4n} \leq F(n) \cdot \left(1 - \frac{1}{F(n)}\right)^N \leq \frac{1}{en}.$$

Therefore, applying the Lovász local lemma, ρ can be bounded from below as $\rho \geq (1 - F(n))^N \geq 1/e$. Taking the logarithm,

$$\begin{aligned}
\mathbb{H}\left(\mathcal{U}_{\mathsf{G}(\lambda,K)}\right) &\geq \frac{(N-1)^2}{2} h(\varepsilon) - O(N \ln \lambda) \\
&\quad + \frac{K}{2}\left((n-1)h\left(\frac{M}{n-1}\right) - O(\ln n)\right) - O(1) \qquad (25) \\
&= \Omega\left(n^2 h(\varepsilon) + n^{1-3/\lambda} K\right) \qquad (26)
\end{aligned}$$

where $\varepsilon := \phi/(N-1) \leq 1/2$. In (25), we have used Stirling's approximation and the assumption that $K = o(n)$. Continuing from (26), since $2nh(\varepsilon) \geq \lambda \ln(n/\lambda)$, for sufficiently large n,

$$\mathbb{H}\left(\mathcal{U}_{\mathsf{G}(\lambda,K)}\right) = \Omega\left(n\lambda \log \frac{n}{\lambda} + n^{1-3/\lambda} K\right). \qquad (27)$$

Substituting (27) into (22) and noting that $\det(\mathbf{K}) = 1$, when $n \to \infty$, it must hold that

$$m = \Omega\left(\lambda \log(n/\lambda) + K/n^{-3/\lambda}\right)$$

to have $\mathbb{P}_{\text{error}}^{\text{PR}}$ smaller than $1/2$. □

4.1.1 Uniform Sampling of Trees

As one of the applications of Theorem 4.2, we characterize the sample complexity of the uniform sampling of trees.

Corollary 4.1. *Suppose that the generator matrix* $\mathbf{B}$ *has Gaussian IID entries with mean zero and variance one and assume* $G \sim \mathcal{U}_{\mathsf{T}_{\text{All}}}$. *There exists an algorithm that guarantees* $\lim_{n\to\infty} \mathbb{P}_{\text{error}}^{\text{PR}} = 0$ *using* $m = O(\log n)$ *measurements. Conversely, the number of measurements must satisfy* $m = \Omega(\log n)$ *to make the probability of error* $\mathbb{P}_{\text{error}}^{\text{PR}}$ *less than* $1/2$.

Sketch of Proof: The achievability follows from combining Theorem 4.2 and Lemma 2. Substituting $\mathbb{H}(\mathcal{U}_{\mathsf{T}_{\text{All}}}) = \Omega(n \log n)$ into (22) yields the desired result. □

4.1.2 Erdős–Rényi (n, p) model

Similarly, the following corollary is shown by recalling Lemma 3.

Corollary 4.2. *Assume* $G \sim \mathcal{G}_{\text{ER}}(n, p)$ *with* $1/n \leq p \leq 1 - 1/n$. *Under the same conditions in Corollary 4.1, there exists an algorithm that guarantees* $\lim_{n\to\infty} \mathbb{P}_{\text{error}}^{\text{PR}} = 0$ *using* $m = O(nh(p))$ *measurements. Conversely, the number of measurements must satisfy* $m = \Omega(nh(p))$ *to make the probability of error* $\mathbb{P}_{\text{error}}^{\text{PR}}$ *less than* $1/2$.

Sketch of Proof: Taking $K = nh(p)/\log n$ and $\lambda = 2nh(p)/(\ln 1/p)$, we check that $\lambda < n^{-3/\lambda}(n - K)$ and $K = o(n)$. The assumptions on $h(p)$ guarantee that $h(p) \geq (\log n)/n$, whence $nh(p) = \omega(\log(n/K))$. Theorem 4.2 implies that $m = O(nh(p))$ is sufficient for achieving a vanishing probability of error. For the second part of the corollary, substituting $\mathbb{H}(\mathcal{G}_{\text{ER}}(n, p)) = h(p)\binom{n}{2} = \Omega\left(n^2 h(p)\right)$ into (22) yields the desired result. □

5 Structure-based Parameter Recovery

Often in practice, some prior information of the graph topology is available. For example, in a power system, besides knowing that the power network is a radial network, if we can further infer from the locational/geographical information and assure that some of the nodes in $\mathcal{N}$ are *not* connected through a power line, then the size of the candidacy set G becomes smaller, allowing a potential improvement on sample complexity. Leveraging the Kirchhoff's theorem (*c.f.* [26]) stated below, our results are extended to practical situations.

Theorem 5.1 (Kirchhoff's Theorem)**.** *Let H be a connected graph with n labeled nodes. Then the number of spanning trees denoted by $\kappa(H)$ is given by the product of $1/n$ and all non-zero eigenvalues of the (unnormalized) Laplacian matrix of H:*

$$\kappa(H) = \frac{1}{n}\lambda_1\lambda_2\cdots\lambda_{n-1} = \det\left(L'_H\right) \tag{28}$$

where L'_H denotes the reduced Laplacian of H (cofactor) by deleting the first column and row from the Laplacian matrix L_H.

Therefore, if we know a priori that the topology to be recovered is a spanning tree lying in some known underlying graph H, then the size of the candidacy set G_H is given by $|\mathsf{G}_H| = \kappa(H)$. Let $\mathcal{U}_{\mathsf{G}_H}$ denote the uniform distribution on G_H. As a remark, when we have no additional information of the underlying graph and we only know G is a spanning tree, H becomes the complete graph with n nodes and $|\mathsf{G}| = n^{n-2}$.

Applying Kirchhoff's matrix tree theorem, the following corollary is obtained.

Corollary 5.1. *Under the assumption of the GIPM, if $G \sim \mathcal{U}_{\mathsf{G}_H}$, then the number of measurements m must satisfy*

$$m = \Omega\left(\frac{1}{n}\log\left(\prod_{j=1}^{n-1}\lambda_j\right)\right)$$

to make the probability of error $\mathbb{P}_e^{\mathrm{T}}$ less than $1/2$. Here $\lambda_1, \ldots, \lambda_{n-1}$ denote the non-zero eigenvalues of the Laplacian matrix of H.

Sketch of Proof: The proof follows along the same lines as those of Corollary 4.1 and 4.2. Putting $\kappa(H) = \lambda_1\lambda_2\cdots\lambda_{n-1}/n$ into (17) gives the bound. □

The next achievability follows straightforward by noting that the number of unknown entries in each j-th column of the graph matrix L_H is at most $\max_{j=1}^{n} \mathrm{diag}_j(L_H)$.

Corollary 5.2. *Under the assumption of the GIPM, if $G \sim \mathcal{U}_{\mathsf{G}_H}$, then the following upper bound on the number of measurements m is sufficient to achieve a vanishing probability of error $\mathbb{P}_{\mathrm{error}}^{\mathrm{PR}} = o(1)$:*

$$m = O\left(\log \max_{j=1}^{n} \mathrm{diag}_j(L_H)\right).$$

Here $\mathrm{diag}_j(L_H)$ denotes the j-th diagonal entry of the (unnormalized) Laplacian matrix of H.

6 Heuristic Algorithm

We present in this section an algorithm motivated by the consistency-checking step in the proof of achievability (see Section 3.2). Instead of checking the consistency of each subset of $\mathcal{N}$ consisting of $n - K$ nodes, as the

two-stage scheme does and which requires $O(n^K)$ operations, we compute an estimate X_j for each column of the graph matrix independently and then assign a score to each column based on its symemtric consistency with respect to the other columns in the matrix. The lower the score, the closer the estimate of the matrix column X_j is to the ground truth Y_j. Using a scoring function we rank the columns, select a subset of them to be "correct", and then eliminate this subset from the system. The size of the subset determines the number of iterations. Heuristically, this procedure results in a polynomial-time algorithm to compute an estimate $\mathbf{X}$ of the graph matrix $\mathbf{Y}$.

The algorithm proceeds in four steps.

6.0.1 Step 1. Initialization

Let matrices $\mathbf{A} \in \mathbb{R}^{m\times n}$ and $\mathbf{B} \in \mathbb{R}^{m\times n}$ be given and set the number of columns fixed in each iteration to be an integer s such that $1 \leq s \leq n$. For the first iteration, set $\mathcal{S}(0) \leftarrow \mathcal{N}$, $\mathbf{A}(0) \leftarrow \mathbf{A}$, and $\mathbf{B}(0) \leftarrow \mathbf{B}$.

For each iteration $r = 0, \ldots, \lceil n/s \rceil - 1$, we perform the remaining three stages. The system dimension is reduced by s after each iteration.

6.0.2 Step 2. Independent ℓ_1-minimization

For all $j \in \mathcal{S}(r)$, we solve the following ℓ_1-minimization:

$$\begin{aligned} X_j(r) = \underset{x\in\mathbb{F}^{n-sr}}{\arg\min} \quad & \|x\|_{\ell_1} \\ \text{subject to} \quad & \mathbf{B}(r)x = A_j(r), \\ & x \in \mathcal{X}_j(r) \end{aligned} \tag{29}$$

Constraint (29) is optional; the set $\mathcal{X}_j(r)$ may encode additional constraints on the form of x such as entry-wise positivity or negativity (*e.g.,* Section 7). The forms of reduced matrix $\mathbf{B}(r)$ and reduced vector $A_j(r)$ are specified in Step 4.

6.0.3 Step 3. Column scoring

We rank the *symmetric consistency* of the independently solved columns. For all $j \in \mathcal{S}(r)$, let

$$\mathsf{score}_j(r) := \sum_{i=1}^{n-sr} \left| X_{i,j}(r) - X_{j,i}(r) \right|$$

Note that if $\mathsf{score}_j(r) = 0$ then $X_j(r)$ and its partner symmetric row in $\mathbf{X}(r)$ are identical. Otherwise there will be some discrepancies between the entries and the sum will be positive. The subset of the $X_j(r)$ corresponding to the s smallest values of $\mathsf{score}_j(r)$ is deemed "correct". Call this subset of correct indices $\mathcal{S}'(r)$.

6.0.4 Step 4. System dimension reduction

Based on the assumption that s of the previously computed columns $X_j(r)$ are correct, the dimension of the linear system is reduced by s. We set $\mathcal{S}(r+1) \leftarrow \mathcal{S}(r)\backslash\mathcal{S}'(r)$. For all $i, j \in \mathcal{S}'(r)$, we fix

$$X_{i,j} = X_{i,j}(r),\ X_{j,i} = X_{i,j}(r) \tag{30}$$

The measurement matrices are reduced to

$$\begin{aligned}\mathbf{B}(r+1) &\leftarrow \underline{\mathbf{B}}^{\mathcal{S}(r+1)},\\ A_j(r+1) &\leftarrow \underline{A}_j(r) - \sum_{i\in\mathcal{S}'(r)} \underline{B}_i X_{i,j}.\end{aligned}$$

When $r \le n-m$, $\underline{\mathbf{B}}^{\mathcal{S}(r+1)} = \mathbf{B}^{\mathcal{S}(r+1)}$, $\underline{A}_j(r) = A_j(r)$ and $\underline{B}_i = B_i$. When $r > n-m$, to avoid making the reduced matrix $\mathbf{B}(r+1)$ over-determined, we set $\mathbf{B}(r+1)$ to be an $(n-r)\times(n-r)$ sub-matrix of $\mathbf{B}^{\mathcal{S}(r+1)}$ by selecting $n-r$ rows of $\mathbf{B}^{\mathcal{S}(r+1)}$ uniformly at random. A new length-$n-r$ vector $\underline{A}_j(r)$ is formed by selecting the corresponding entries from $A_j(r)$. Once the $\lceil n/s \rceil$ iterations complete, an estimate $\mathbf{X}$ is returned using (30). The algorithm requires at most $\lceil n/s \rceil$ iterations and in each iteration, the algorithm solves an ℓ_1-minimization and updates a linear system. Solving an ℓ_1-minimization can be done in polynomial time (*c.f.* [29]). Thus, the heuristic algorithm is a polynomial-time algorithm.

7 Simulations

Experimental results for the heuristic algorithm are given here for both synthetic data and IEEE standard power system test cases. The algorithm was implemented in Matlab; simulated power flow data was generated using Matpower 7.0 [30] and CVX 2.1 [31] with the Gurobi solver [32] was used to solve the sparse optimization subroutine.

7.1 Scalable Topologies and Error Criteria

We first demonstrate our results using synthetic data and two typical graph ensembles – stars and chains. For both topologies, the graph size was incremented from $n = 5$ to $n = 300$ and the number of samples required for accurate recovery of parameters and topology was recorded. For each simulation, we generated a complex-valued random admittance matrix $\mathbf{Y}$ as the ground truth. Both the real and imaginary parts of the line impedances of the network were selected uniformly and IID from $[-100, 100]$. A valid electrical admittance matrix was then constructed using these impedances. The real components of the entries of $\mathbf{B}$ were distributed IID according to $\mathcal{N}(1,1)$ and the imaginary components according to $\mathcal{N}(0,1)$. $\mathbf{A} = \mathbf{YB}$ gave the corresponding complex-valued measurement matrix.

Given data matrices $\mathbf{A}, \mathbf{B}$ the algorithm returned an estimate $\mathbf{X}$ of the ground truth $\mathbf{Y}$. If an entry of $\mathbf{X}$ had magnitude $|X_{i,j}| < \varepsilon$ (where $\varepsilon = 10^{-5}$ was the error threshold), then the entry was fixed to be 0. Following this, if $\operatorname{supp}(\mathbf{X}) = \operatorname{supp}(\mathbf{Y})$ then the topology recovery was deemed exact. The criterion for accurate parameter recovery was $|Y_{i,j} - X_{i,j}| < \varepsilon$ for all non-zero entries in both matrices. The number of samples m (averaged over repeated trials) required to meet these criteria was designated as the sample complexity for accurate recovery. The sample complexity trade-off displayed in Figure 1 shows approximately logarithmic dependence on graph size n for both ensembles.

7.2 IEEE Test Cases

We also validated the heuristic algorithm on 17 IEEE standard power system test cases ranging from 5 to 200 buses. The procedure for determining sample complexity for accurate recovery was the same as above, but the data generation was more involved.

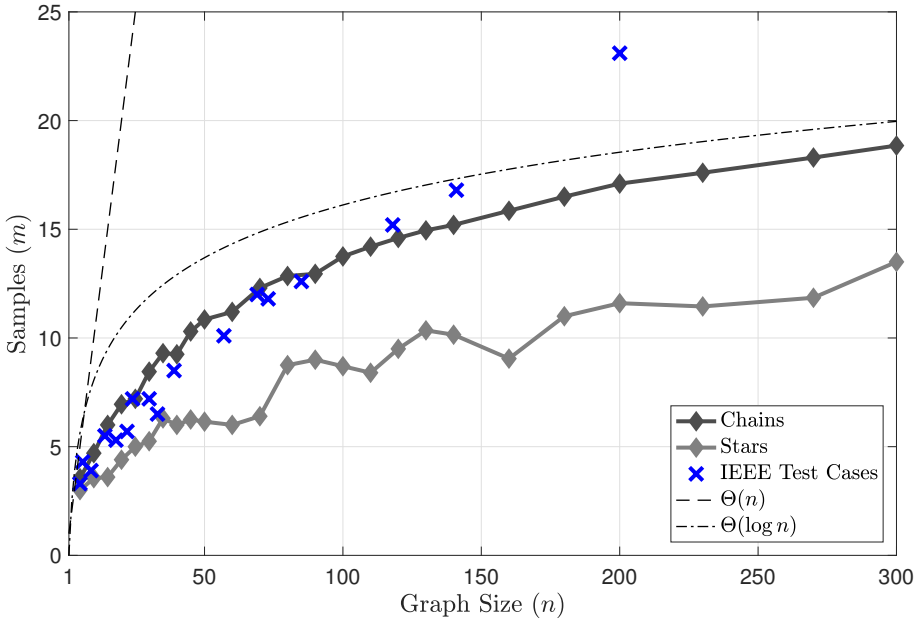

Figure 1: The number of samples required to accurately recover the nodal admittance matrix is shown on the vertical axis. Results were averaged over 20 independent simulations. Star and chain graphs were scaled in size between 5 and 300 nodes. IEEE test cases ranged from 5 to 200 buses. In the latter case, there were no assumptions on the random IID selection of the entries of $\mathbf{Y}$ (in contrast to the star/chain networks). Linear and logarithmic (in n) reference curves are plotted as dashed lines.

7.2.1 Power flow data generation

A sequence of time-varying loads was created by scaling the nominal load values in the test cases by a times series of Bonneville Power Administration's aggregate load on 02/08/2016, 6am to 12pm [33]. For each test case network, we performed the following steps to generate a set of measurements:

a) Interpolated the aggregate load profile to 6-second intervals, extracted a length-m random consecutive subsequence, and then scaled the real parts of bus power injections by the load factors in the subsequence.

b) Computed optimal power flow in Matpower for the network at each time step to determine bus voltage phasors.

c) Added a small amount of Gaussian random noise ($\sigma^2 = 0.001$) to the voltage measurements and generated corresponding current phasor measurements using the known admittance matrix.

7.2.2 Sample complexity for recovery of IEEE test cases

Figure 1 shows the sample complexity for accurate recovery of the IEEE test cases. The procedure and criteria for determining the necessary number of samples for accurate recovery of the admittance matrix were the same as for the synthetic data case. Unlike the previous setting, here we have no prior assumptions about the structure of the IEEE networks: networks have both mesh and radial topologies. However, because power system topologies are typically highly sparse, the heuristic algorithm was able to achieve accurate recovery with a comparable (logarithmic) dependence on graph size.

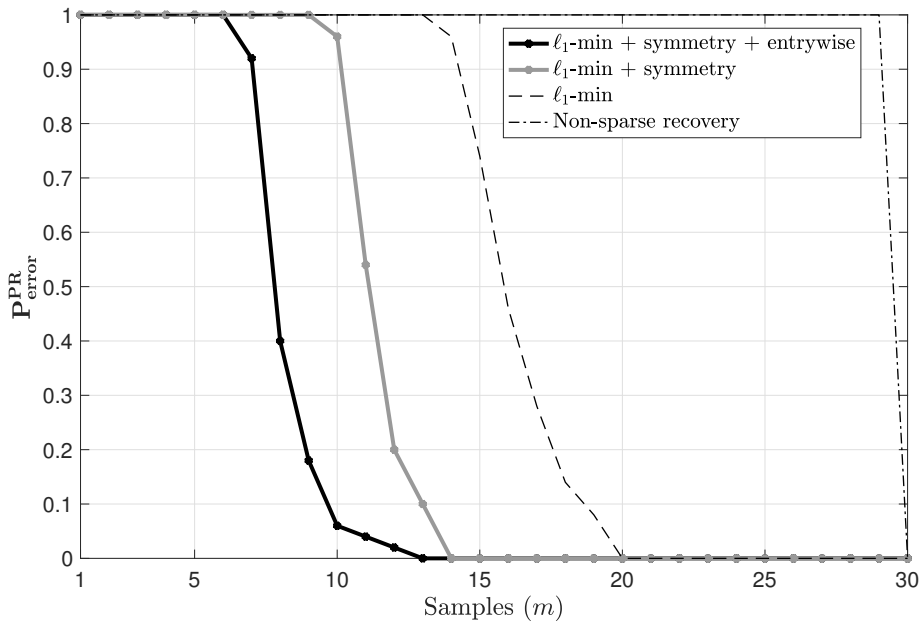

Figure 2: Probability of error for parameter recovery $\mathbb{P}^{\text{PR}}_{\text{error}}$ for the the IEEE 30-bus test case is displayed on the vertical axis. Probability is taken over 50 independent simulations. The horizontal axis shows the number of samples used to compute the estimate $\mathbf{X}$. The probability of error for independent recovery of each X_j via ℓ_1-norm minimization (double dashed line) and full rank non-sparse recovery (dot dashed line) are shown for reference. Adding the symmetry score function (second-to-left) improves over the naive column-wise scheme. Adding entry-wise positivity/negativity constraints on the entries of $\mathbf{X}$ (left-most curve) reduces sample complexity even further ($\approx 1/3$ samples needed compared to full rank recovery).

7.2.3 Influence of structure constraints on recovery

There are structural properties of the nodal admittance matrix for power systems—symmetry, sparsity, and entry-wise positivity/negativity—that we exploit in the heuristic algorithm to improve sample complexity for accurate recovery. The score function $\mathsf{score}_j(r)$ rewards symmetric consistency between columns in $\mathbf{X}$; the use of ℓ_1-minimization promotes sparsity in the recovered columns; and the constraint set $\mathcal{X}_j$ in (29) forces $\text{Re}(X_{i,j}) \leq 0,\ \text{Im}(X_{i,j}) \geq 0$ for $i \neq j$ and $\text{Re}(X_{i,j}) \geq 0$ for $i = j$. These entry-wise properties are commonly found in power system admittance matrices. In Figure 2 we show the results of an experiment on the IEEE 30-bus test case that quantify the effects of the structure constraints on the probability of error.

References

[1] J. A. Momoh, R. Adapa, and M. El-Hawary, "A review of selected optimal power flow literature to 1993. i. nonlinear and quadratic programming approaches," *IEEE transactions on power systems*, vol. 14, no. 1, pp. 96–104, 1999.

[2] S. H. Low, "Convex relaxation of optimal power flow—part i: Formulations and equivalence," *IEEE Transactions on Control of Network Systems*, vol. 1, no. 1, pp. 15–27, 2014.

[3] Y. Tang, K. Dvijotham, and S. Low, "Real-time optimal power flow," *IEEE Transactions on Smart Grid*, vol. 8, no. 6, pp. 2963–2973, 2017.

[4] A. Mittal, J. Hazra, N. Jain, V. Goyal, D. P. Seetharam, and Y. Sabharwal, "Real time contingency analysis for power grids," in *European Conference on Parallel Processing*. Springer, 2011, pp. 303–315.

[5] R. Horta, J. Espinosa, and J. Patiño, "Frequency and voltage control of a power system with information about grid topology," in *Automatic Control (CCAC), 2015 IEEE 2nd Colombian Conference on*. IEEE, 2015, pp. 1–6.

[6] N. P. Santhanam and M. J. Wainwright, "Information-theoretic limits of selecting binary graphical models in high dimensions." *IEEE Trans. Information Theory*, vol. 58, no. 7, pp. 4117–4134, 2012.

[7] R. M. Fano and D. Hawkins, "Transmission of information: A statistical theory of communications," *American Journal of Physics*, vol. 29, pp. 793–794, 1961.

[8] A. Anandkumar, V. Tan, and A. Willsky, "High dimensional structure learning of ising models on sparse random graphs," 2017.

[9] A. Anandkumar, V. Tan, and A. S. Willsky, "High-dimensional graphical model selection: tractable graph families and necessary conditions," in *Advances in Neural Information Processing Systems*, 2011, pp. 1863–1871.

[10] A. Ghoshal and J. Honorio, "Information-theoretic limits of bayesian network structure learning," *arXiv preprint arXiv:1601.07460*, 2016.

[11] S. Aeron, V. Saligrama, and M. Zhao, "Information theoretic bounds for compressed sensing," *IEEE Transactions on Information Theory*, vol. 56, no. 10, pp. 5111–5130, 2010.

[12] A. Bogdanov, E. Mossel, and S. Vadhan, "The complexity of distinguishing markov random fields," in *Approximation, Randomization and Combinatorial Optimization. Algorithms and Techniques*. Springer, 2008, pp. 331–342.

[13] C. Chow and C. Liu, "Approximating discrete probability distributions with dependence trees," *IEEE transactions on Information Theory*, vol. 14, no. 3, pp. 462–467, 1968.

[14] V. Y. Tan, A. Anandkumar, and A. S. Willsky, "Learning gaussian tree models: Analysis of error exponents and extremal structures," *IEEE Transactions on Signal Processing*, vol. 58, no. 5, pp. 2701–2714, 2010.

[15] Y. Yuan, O. Ardakanian, S. Low, and C. Tomlin, "On the inverse power flow problem," *arXiv preprint arXiv:1610.06631*, 2016.

[16] E. Candes, M. Rudelson, T. Tao, and R. Vershynin, "Error correction via linear programming," in *Foundations of Computer Science, 2005. FOCS 2005. 46th Annual IEEE Symposium on*. IEEE, 2005, pp. 668–681.

[17] M. Rudelson and R. Vershynin, "On sparse reconstruction from fourier and gaussian measurements," *Communications on Pure and Applied Mathematics*, vol. 61, no. 8, pp. 1025–1045, 2008.

[18] M. Babakmehr, M. G. Simões, M. B. Wakin, and F. Harirchi, "Compressive sensing-based topology identification for smart grids," *IEEE Transactions on Industrial Informatics*, vol. 12, no. 2, pp. 532–543, 2016.

[19] D. Baron, M. B. Wakin, M. F. Duarte, S. Sarvotham, and R. G. Baraniuk, "Distributed compressed sensing," 2005.

[20] V. Y. Tan and A. S. Willsky, "Sample complexity for topology estimation in networks of lti systems," *IFAC Proceedings Volumes*, vol. 44, no. 1, pp. 9079–9084, 2011.

[21] Y. Liao, Y. Weng, M. Wu, and R. Rajagopal, "Distribution grid topology reconstruction: An information theoretic approach," in *North American Power Symposium (NAPS), 2015*. IEEE, 2015, pp. 1–6.

[22] S. Bolognani, N. Bof, D. Michelotti, R. Muraro, and L. Schenato, "Identification of power distribution network topology via voltage correlation analysis," in *52nd IEEE Conference on Decision and Control*. IEEE, 2013, pp. 1659–1664.

[23] Y. Sharon, A. M. Annaswamy, A. L. Motto, and A. Chakraborty, "Topology identification in distribution network with limited measurements," in *2012 IEEE PES Innovative Smart Grid Technologies (ISGT)*. IEEE, 2012, pp. 1–6.

[24] D. Deka, S. Backhaus, and M. Chertkov, "Estimating distribution grid topologies: A graphical learning based approach," in *Power Systems Computation Conference (PSCC), 2016*. IEEE, 2016, pp. 1–7.

[25] J. Yu, Y. Weng, and R. Rajagopal, "Patopa: A data-driven parameter and topology joint estimation framework in distribution grids," *IEEE Transactions on Power Systems*, 2017.

[26] D. B. West *et al.*, *Introduction to graph theory*. Prentice hall Upper Saddle River, 2001, vol. 2.

[27] D. L. Donoho and M. Elad, "Optimally sparse representation in general (nonorthogonal) dictionaries via l1 minimization," *Proceedings of the National Academy of Sciences*, vol. 100, no. 5, pp. 2197–2202, 2003.

[28] A. Liebenau and N. Wormald, "Asymptotic enumeration of graphs by degree sequence, and the degree sequence of a random graph," *arXiv preprint arXiv:1702.08373*, 2017.

[29] D. Ge, X. Jiang, and Y. Ye, "A note on the complexity of lp minimization," *Mathematical programming*, vol. 129, no. 2, pp. 285–299, 2011.

[30] R. D. Zimmerman, C. E. Murillo-Sánchez, and R. J. Thomas, "Matpower: Steady-state operations, planning, and analysis tools for power systems research and education," *IEEE Transactions on power systems*, vol. 26, no. 1, pp. 12–19, 2011.

[31] CVX Research, Inc., "CVX: Matlab software for disciplined convex programming, version 2.0," http://cvxr.com/cvx, Aug. 2012.

[32] Gurobi Optimization, LLC, "Gurobi optimizer reference manual," 2018. [Online]. Available: http://www.gurobi.com

[33] Boneville Power Administration, "Bpa: Balancing authority load total wind generation," Oct. 2016. [Online]. Available: https://transmission.bpa.gov/ Business/Operations/Wind/

[34] H. Kajimoto, "An extension of the prüfer code and assembly of connected graphs from their blocks," *Graphs and Combinatorics*, vol. 19, no. 2, pp. 231–239, 2003.

.1 Proof of Lemma 4

Proof. Conditioned on $G \in \mathsf{G}(\lambda, K)$ and the assumption $\sigma_{3\lambda} + 3\sigma_{4\lambda} < 2$, there are no less than $n - K$ many columns correctly recovered. The consistency-checking verifies that if the collection of an arbitrary set of nodes $\mathcal{S}$ of cardinality $n - K$ satisfies the symmetry property as the true graph $\mathbf{Y}$ must obey. Therefore, any such set $\mathcal{S}$ with $|\mathcal{S}| = n - K$ must contain at least $n - 2K$ many corresponding indexes of the correctly recovered columns. Then if the consistency-checking fails, it is necessary that there exist two distinct length-n vectors Y' and Y^* in $\mathbb{F}^n$ such that Y^* is the minimizer of the ℓ_1-minimization (15)-(17) that differs from the correct answer Y', *i.e.,* $Y' \neq Y^*$ where $A = \mathbf{B}Y'$ and

$$\begin{aligned} Y^* = \arg\min_{Y} \ & ||Y||_{\ell_1} \\ \text{subject to } & A = \mathbf{B}Y \\ & Y \in \mathbb{F}^n \end{aligned}$$

for some $A \in \mathbb{F}^m$ and furthermore, the vectors Y' and Y^* can have at most $2K$ distinct coordinates,

$$|\mathrm{supp}\,(Y' - Y^*)| \leq 2K.$$

However, the constraints $\mathbf{B}Y' = A$ and $\mathbf{B}Y^* = A$ imply that $\mathbf{B}\,(Y' - Y^*) = 0$, contradicting to $\mathrm{spark}(\mathbf{B}) > 2K$. Therefore, $n - K$ many columns can be successfully recovered if the decoded solution passes the consistency-checking. Moreover, since $\mathrm{spark}(\mathbf{B}) > 2K$ and number of unknown coordinates in each length-K vector $X_j^{\mathcal{S}^c}$ (for $j = 1, \ldots, |\mathcal{S}^c|$) to be recovered is K, the solution of the system (18) is guaranteed to be unique. Thus, Algorithm 1 always recovers the correct columns $Y_1, \ldots, Y_N$ conditioned on $m \geq \mathrm{spark}(\mathbf{B}) > 2K$. □

.2 Proof of Lemma 2

Proof. Consider the following function

$$F(\mathcal{E}) = \sum_{j=1}^{n} f(d_j(G))$$

where $d_j(G)$ denotes the degree of the j-th node and

$$f(d_j(G)) := \begin{cases} 1 & \text{if } d_j(G) > \lambda \\ 0 & \text{otherwise} \end{cases}.$$

Applying the Markov's inequality,

$$\begin{aligned} \mathbb{P}\left(G \in \mathsf{T}_{\mathrm{All}}(\lambda, K)\right) &= \mathbb{P}_{\mathcal{U}_{\mathsf{T}_{\mathrm{All}}}}\left(F(\mathcal{E}) \geq K\right) \\ &\leq \frac{\mathbb{E}_{\mathcal{U}_{\mathsf{T}_{\mathrm{All}}}}\left[F(\mathcal{E})\right]}{K}. \end{aligned} \tag{31}$$

Continuing from (31), the expectation $\mathbb{E}_{\mathcal{U}_{\mathsf{T}_{\mathrm{All}}}}\left[F(\mathcal{E})\right]$ can be further expressed and bounded as

$$\begin{aligned} \mathbb{E}_{\mathcal{U}_{\mathsf{T}_{\mathrm{All}}}}\left[F(\mathcal{E})\right] &= \sum_{j=1}^{n} \mathbb{E}_{\mathcal{U}_{\mathsf{T}_{\mathrm{All}}}}\left[f(d_j(G))\right] \\ &= \sum_{j=1}^{n} \mathbb{P}_{\mathcal{U}_{\mathsf{T}_{\mathrm{All}}}}\left(d_j(G) > \lambda\right). \end{aligned} \tag{32}$$

Since G is chosen uniformly at random from $\mathsf{T}_{\mathrm{All}}$, it is equivalent to selecting its corresponding Prüfer sequence (by choosing $n-2$ integers independently and uniformly from the set $\mathcal{N}$, *c.f.* [34]) and the number of appearances of each $j \in \mathcal{N}$ equals to $d_j(G) - 1$. Therefore, for any fixed node $j \in \mathcal{N}$, the Chernoff bound implies that

$$\mathbb{P}_{\mathcal{U}_{\mathsf{T}_{\mathrm{All}}}}\left(d_j(G) > \lambda\right) \leq \exp\left(-(n-2)\mathbb{D}_{\mathrm{KL}}\left(\frac{\lambda}{n-2}\middle\|\frac{1}{n}\right)\right) \tag{33}$$

where $\mathbb{D}_{\mathrm{KL}}(\cdot||\cdot)$ is the Kullback–Leibler divergence and

$$\mathbb{D}_{\mathrm{KL}}\left(\frac{\lambda}{n-2}\middle\|\frac{1}{n}\right) \geq \frac{\lambda}{n-2}\ln n. \tag{34}$$

Therefore, substituting (34) this back into (33) and combining (31) and (32),

$$\mathbb{P}\left(G \in \mathsf{T}_{\mathrm{All}}(\lambda, K)\right) \leq \frac{n\exp(-\lambda \ln n)}{K}.$$

Setting $K = \ln n$, since $\lambda \geq 1$, the proof is complete. □

.3 Proof of Lemma 3

Proof. For any fixed node $j \in \mathcal{N}$, applying the Chernoff bound,

$$\mathbb{P}_{\mathcal{G}_{\mathrm{ER}}(n,p)}\left(d_j(G) > \lambda\right) \leq \exp\left(-n\mathbb{D}_{\mathrm{KL}}\left(\frac{\lambda}{n}\middle\|p\right)\right).$$

Continuing from (31), the expectation $\mathbb{E}_{\mathcal{G}_{\mathrm{ER}}(n,p}\left[F(\mathcal{E})\right]$ can be further expressed and bounded as

$$\mathbb{E}_{\mathcal{G}_{\mathrm{ER}}(n,p)}\left[F(\mathcal{E})\right] \leq n \cdot \exp\left(-n\mathbb{D}_{\mathrm{KL}}\left(\frac{\lambda}{n}\middle\|p\right)\right) \tag{35}$$

where the probability p satisfies $0 < p \leq \lambda/n < 1$. Note that

$$\mathbb{D}_{\mathrm{KL}}\left(\frac{\lambda}{n}\middle\|p\right) = \frac{\lambda}{n}\ln\frac{1}{p} + \left(1-\frac{\lambda}{n}\right)\ln\frac{1}{1-p} - h(p) \tag{36}$$

where the binary entropy $h(p)$ is in base e. Taking $\lambda = 2nh(p)/(\ln 1/p) \geq 2np$, substituting (36) into (35) leads to

$$\mathbb{E}_{\mathcal{G}_{\mathrm{ER}}(n,p)}\left[F(\mathcal{E})\right] \leq n\exp\left(-nh(p)\right).$$

Therefore, (31) gives

$$\mathbb{P}\left(G \in \mathsf{G}_{\mathrm{All}}(\lambda, K)\right) \leq \frac{n\exp\left(-nh(p)\right)}{K}.$$

Noticing that $nh(p) = \omega\left(\log(n/K)\right)$ completes the proof. □