Approximate Recovery Of Ising Models with Side Information

Saurabh Sihag and Ali Tajer
Rensselaer Polytechnic Institute
Troy, NY 12180

***Abstract*—This paper considers the problem of recovering the edge structures of two partially identical graphs in the class of Ising models. It is assumed that both graphs have the same number of nodes and a known *subset* of nodes have identical structures in both graphs. Therefore, inferring the structure of one graph can provide the side information that could be leveraged for inference related to the other graph. The objective is to recover the connectivity of both graphs under an approximate recovery criterion. The degree- and edge-bounded subclass of Ising models is considered and necessary conditions (information-theoretic) and sufficient conditions for the sample complexity to achieve a bounded probability of error are established. Furthermore, the scaling behavior of the sample complexity is analyzed in different regimes and specific regimes are identified for which the necessary and sufficient conditions coincide, thus, establishing the optimal sample complexity.**

I. Introduction

Graphical models have been used to model the conditional dependence among a set of random variables, in which the random variables are associated with the nodes of the graph and their interdependence is characterized by the edge structure among them [1] and [2]. Graph-based models have applications in a broad range of domains, e.g., computer vision [3], genetics [4]–[6], social networks [7], and power systems [8]. In this paper, we consider the problem of learning the edge structures of a pair of structurally similar graphs in the class of Ising models by utilizing the samples from their joint distributions.

Graphical models with partially similar structures arise in modeling the inference problems in biological networks [4], physical infrastructures [9], and behavioral analysis [10]. In such applications, the data is generated by multiple layered networks of information sources that are modeled by graphs, in which each layer shares some of its vertices and their respective data with the other layers. For instance, the relationships among a group of individuals that are active on multiple social networks (e.g., Twitter and Facebook) can be modeled by distinct, but potentially partially similar graphical models. In these scenarios, the data collected from different graphs have redundancy of information and inference about one model serves as side information for the other models.

Motivated by this premise, we analyze the sample complexity for joint model selection of a pair of partially similar graphs in the class of Ising models from an information-theoretic perspective. We consider an approximate recovery criterion, i.e., at most a fixed number of errors are tolerated in the estimated edge structures, and provide algorithm-independent information-theoretic bounds on the sample complexity that establish the statistical difficulty of the problem. Furthermore, we also analyze the performance of a maximum likelihood (ML) decoder to provide sufficient conditions on the sample complexity.

The problem of graph structure learning is NP-hard in general [11]. However, it becomes feasible under certain restrictions on the edge structure, e.g., sparsity [12]–[15]. Such restrictions can be accommodated by considering graph subclasses with restrictions on the maximum degree and number of edges. Sample complexity for structure learning of single graphs have been studied from an information-theoretic perspective in [16]–[20]. In [16] and [17], algorithm independent necessary conditions on the sample complexity for the exact structure recovery of graphs in different sub-classes of Ising models are established. In [18], structure learning for Ising models is investigated under the performance criterion of approximate recovery, i.e., at most a fixed number of errors can

This research was supported in part by the U. S. National Science Foundation under grants DMS-1737976 and ECCS-1933107.

be tolerated in the estimated graph structure. Necessary conditions for set-based graph model selection, that is, the graph estimator recovers a set of graphs that potentially contains the true graph, are characterized in [20].

Joint inference in multiple graph models has been investigated in [4], [5], [10], [21]–[29] for graphs that may share structural properties. In [5] and [21]–[24], optimization based methods are applied for joint inference of Gaussian graphical models. In [4] and [25], Bayesian frameworks are investigated for joint model inference.

Joint structure learning has been investigated from an information-theoretic perspecitve in [27]–[29]. Necessary conditions on structure recovery of partially similar graph models in path restricted subclass of Ising models and degree bounded subclass of Gaussian models are established in [27]. In [28], the problem of joint graph structure recovery is investigated under the exact recovery criterion. In [29], necessary and sufficient conditions on the approximate recovery of a degree bounded subclass of Ising models are established. In contrast to the studies in [27]–[29], we focus on the joint recovery of graphs in a degree- and edge-bounded subclass of Ising models under the approximate recovery criterion.

II. Graph Model

Consider two undirected graphs $\mathcal{G}_1 \triangleq (V, E_1)$ and $\mathcal{G}_2 \triangleq (V, E_2)$, where the graphs $\mathcal{G}_1$ and $\mathcal{G}_2$ are formed by the set of vertices $V \triangleq \{1, \ldots, p\}$ that are connected by the set of edges $E_1 \subseteq V \times V$ and $E_2 \subseteq V \times V$, respectively. An edge between a pair of nodes u and v in the graph $\mathcal{G}_i$ is denoted by $(u, v) \in E_i$. The set of neighbors of node u in $\mathcal{G}_i$ is denoted by $\mathcal{N}_i(u)$, and its degree is denoted by $d_u^i \triangleq |\mathcal{N}_i(u)|$. We use these two graphs to represent a pair of Ising graphical models.

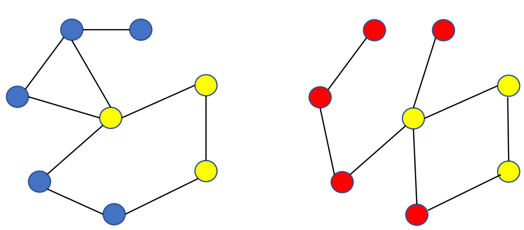

Fig. 1. Two graphs with partially similar structures. Yellow nodes in both graphs have the same internal edge structure.

In this paper, we assume that the graphs $\mathcal{G}_1$ and $\mathcal{G}_2$ are structurally identical within a pre-specified cluster of nodes $V_c \subseteq V$. An example of this setting is illustrated in Fig. 1.

Under the Ising model, each node $u \in V$ in the graph $\mathcal{G}_i$ is associated with a binary random variable $X_i^u \in \mathcal{X} \triangleq \{-1, 1\}$. The joint probability density function (pdf) of the random vector $\mathbf{X}_i \triangleq [X_i^1, \ldots X_i^p]$ is given by

$$f_i(\mathbf{X}_i) = \frac{1}{Z_i} \exp\left(\sum_{u,v \in V_i} \lambda_i^{uv} X_i^u X_i^v \right) , \qquad (1)$$

where

$$\lambda_i^{uv} \triangleq \begin{cases} \lambda, & \text{if } (u,v) \in E_i \\ 0, & \text{otherwise} \end{cases} , \qquad (2)$$

for $\lambda > 0$, and Z_i is the partition function given by

$$Z_i \triangleq \sum_{X_i \in \{-1,1\}^p} \exp\left(\sum_{u,v \in V} \lambda_i^{uv} X_i^u X_i^v \right) . \qquad (3)$$

Note that the parameter λ in (2) controls the dependence among the nodes in the graph. In [16], it is shown that recovering the graph structure from the data becomes more difficult as λ approaches 0 or grows to infinity.

III. Problem Formulation

In this section, we formalize the notation for structural similarity and the recovery criterion for structure learning.

A. Graph Similarity Models

Definition 1. *A pair of graphs $\mathcal{G}_1$ and $\mathcal{G}_2$ is said to be $\eta-$similar if both graphs have the same internal graphical structure within a cluster of nodes of size $\lfloor \eta p \rfloor$, for some $\eta \in (0, 1)$.*

For both $\mathcal{G}_1$ and $\mathcal{G}_2$, the edge structures between any pair of nodes with at least one node not in V_c are assumed to be structurally independent of each other. For ease in notations, we define $q \triangleq \lfloor \eta p \rfloor$ and $\bar{q} \triangleq p - \lfloor \eta p \rfloor$. We denote the family of Ising models by $\mathcal{I}$, and the family of $\eta-$similar pairs of Ising models by $\bar{\mathcal{I}}$. In this paper, we consider a restricted subclass of Ising models, given by $\bar{\mathcal{I}}^{d,k}$, that consists of η-similar pairs of graphs with each graph having at most k number of edges and at most $\lfloor \gamma k \rfloor$, for some $\gamma \in (0, 1)$, edges in the cluster with common structure, and each node in the graph having a degree of at most d. For convenience in notations, throughout the paper we use the shorthand $\bar{\mathcal{I}}$ to refer to $\bar{\mathcal{I}}^{d,k}$.

B. Recovery Criterion

For graph $\mathcal{G}_i$, we collect n graph samples that are generated according to the pdf f_i, for $i \in \{1,2\}$. We denote collection of n graph samples from $\mathcal{G}_i$ by $\mathbf{X}_i^n$. The objective is to jointly estimate the edge structures of $\mathcal{G}_1$ and $\mathcal{G}_2$. The graph decoder $\psi : \mathcal{X}^{n\times p} \times \mathcal{X}^{n\times p} \rightarrow \bar{\mathcal{I}}$ is the function that maps the collected data to the graphs in the class $\bar{\mathcal{I}}$.

We adopt an *approximate* graph recovery criteria, that is, at most a pre-specified number of erroneous decisions about the edges in the recovery of each graph are tolerated by the graph decoder. We define the maximal probability of error in the approximate graph recovery over the class $\bar{\mathcal{I}}$ as

$$\mathsf{P}(\bar{\mathcal{I}}, z) \triangleq \max_{\mathcal{G}_1,\mathcal{G}_2\in\bar{\mathcal{I}}} \mathbb{P}\left[\max_{i\in\{1,2\}}\{|E_i\Delta\hat{E}_i|\} > z\right] , \qquad (4)$$

where $|E_i\Delta\hat{E}_i|$ is the edit distance between E_i and its estimate $\hat{E}_i$ and is given by $|E_i\Delta\hat{E}_i| \triangleq |(E_i\backslash\hat{E}_i) \cup (\hat{E}_i\backslash E_i)|$. The parameter z is the pre-specified maximum distortion level that can be tolerated in the estimated graph structure with respect to the true graph structure. Note that $|E_i\Delta\hat{E}_i|$ represents the number of modifications to be made in the edge structure to transform E_i to $\hat{E}_i$.

IV. Main Results

In this section, we provide necessary and sufficient conditions on the sample complexity for approximate recovery of graph models in the class $\bar{\mathcal{I}}$.

A. Sufficient Conditions

The sufficient conditions for the approximate recovery of graph models are derived based on the large deviations analysis of an ML based graph decoder given by

$$(\hat{\mathcal{G}}_1, \hat{\mathcal{G}}_2) = \arg\max_{(\mathcal{G}_1,\mathcal{G}_2)\in\bar{\mathcal{I}}} f_{\mathcal{G}_1,\mathcal{G}_2}(\mathbf{X}_1^n, \mathbf{X}_2^n) , \qquad (5)$$

where $f_{\mathcal{G}_1,\mathcal{G}_2}(\mathbf{X}_1^n, \mathbf{X}_2^n)$ is the joint pdf of the data samples $\mathbf{X}_1^n$ and $\mathbf{X}_2^n$. The ML decoder in (5) is optimal for exact recovery criterion, i.e., when $z = 0$. For approximate recovery, we assume that the ML decoder does not declare an error if the estimated graph structures are within a distortion level of z with respect to the structures of the true graphs.

Theorem 1. *Consider a pair of $\eta-$similar graphs $\mathcal{G}_1$ and $\mathcal{G}_2$ in class $\bar{\mathcal{I}}$. If the sample size n satisfies*

$$n \geq r_1 \max\{C_1, 2C_2\} , \qquad (6)$$

where we have defined

$$r_1 \triangleq \frac{2d(3\exp(2\lambda d)+1)}{\sinh^2(\lambda/4)} , \qquad (7)$$

$$C_1 \triangleq \Big(\log(2\bar{q}d - z) + \log\left(\binom{\bar{q}}{2} + q\bar{q}\right) + \log\frac{4}{\delta} , \qquad (8)$$

$$C_2 \triangleq \left(2\log(qd - z) + \log\binom{q}{2} + \log\frac{2}{\delta}\right) , \qquad (9)$$

then there exists a graph decoder $\psi : \mathcal{X}^{n\times p}\times\mathcal{X}^{n\times p} \rightarrow \bar{\mathcal{I}}$ that achieves $\mathsf{P}(\bar{\mathcal{I}}, z) \leq \delta$.

Note that r_1C_1 dominates $2r_2C_2$ in the sample complexity if the size of the non-shared cluster of nodes is large, i.e., $\frac{\bar{q}}{q} \gg 1$,which illustrates the effect of structural similarity. We elaborate on the scaling behavior of the terms r_1C_1 and r_1C_2 under different regimes. In the following regimes, it is assumed that d and k are increasing with p.

1. $\lambda = \Theta(1)$: In this regime, both r_1C_1 and $2r_1C_2$ scale as $e^d d \log\frac{pd}{\delta}$.
2. $\lambda = O\left(\frac{1}{d}\right)$: In this regime, as $d \rightarrow \infty$, we have $\sinh(\lambda/4) = \Omega(\lambda)$. Therefore, the sufficient condition from Theorem 1 can be simplified to $n \geq c_1 \max\{d^2, \lambda^{-2}\} d \log p/\delta$, where c_1 is a positive constant. For a constant δ, the bound on the sample complexity has an asymptotic scaling behavior given by $\Omega(d^3 \log p)$.
3. $\lambda = \Theta(d)$: In this regime, the terms r_1C_1 and $2r_1C_2$ scale as $e^{\lambda d}\log p$ for a constant δ. Furthermore, if we have $\lambda d = \omega(\log(\log p))$, then the scaling behavior is simplified to $e^{\lambda d}$.

B. Necessary Conditions

Theorem 2 ($k \leq p/4$)**.** *Consider a pair of $\eta-$similar graphs $\mathcal{G}_1$ and $\mathcal{G}_2$ in the class $\bar{\mathcal{I}}$, such that, $d > 2$, $k = \omega(d^2)$ and $k \leq p/4$, and $\gamma \leq \min\left\{\frac{q}{2k}, \frac{2\binom{d}{2}q}{d+2}\right\}$. For*

TABLE I
SUMMARY OF THE MAIN RESULTS FOR RECOVERING ISING MODELS OF CLASS $\bar{\mathcal{I}}$.

Parameters	Approx. recovery ($z > 0$) (Necessary conditions)	Approx. recovery ($z > 0$) (Sufficient conditions)	Exact recovery ($z = 0$) (Necessary conditions)
$\lambda = O\left(\frac{1}{d}\right)$ $k = O(p)$ and $k = \Omega(d^2)$	$\Omega(d^2 \log p)$	$\Omega(d^3 \log p)$	$\Omega(d^2 \log p)$
$\lambda = O\left(\frac{1}{d}\right)$ $k = \Omega(p)$ and $k = O(p\sqrt{d})$	$\Omega(d^2)$	$\Omega(d^3 \log p)$	$\Omega(d^2 \log p)$
$\lambda = O\left(\frac{1}{d}\right)$ $k = \Omega(p\sqrt{d})$ and $k = O(\frac{pd}{2})$	$\Omega(\frac{d^3 p^2}{k^2})$	$\Omega(d^3 \log p)$	$\Omega(d^2 \log p)$
$\lambda = O\left(\frac{1}{p}\right)$ k, d fixed and $k \leq p/4$	$\Omega(p^2 \log p)$	$\Omega(p^2 \log p)$	$\Omega(p^2 \log p)$

any graph decoder $\psi : \mathcal{X}^{n\times p} \times \mathcal{X}^{n\times p} \to \bar{\mathcal{I}}$ that achieves

$$\mathsf{P}(\bar{\mathcal{I}}, z) \leq \delta , \tag{10}$$

for $z = \lfloor \theta k \rfloor$ for some $\theta \in (0, \frac{d-2}{4d})$, the sample size n should satisfy

$$n \geq \max\{D_1, D_2\}(1 - \delta - o(1)) , \tag{11}$$

where

$$D_1 \triangleq \frac{2(1-\gamma)\log(\bar{q}) + \gamma \log q - 2\theta \log p}{\lambda \tanh \lambda} , \tag{12}$$

$$D_2 \triangleq \frac{\exp(\lambda(d-2)/4)((1-\gamma/2)\log 2 - H_2(\frac{d}{d-2}\cdot 2\theta))}{3\lambda d^2} . \tag{13}$$

Note that D_1 and D_2 have different scaling behavior in λ, d, and p. In the following regimes, we assume that both d and k increase with p.

1. $\lambda = \Theta(1)$: In this regime, D_1 scales as $\log p$ and D_2 scales exponentially in d. From Theorem 2, we have $p = \omega(d^2)$. If d grows at a rate faster than $\log(\log p)$, the bound on sample complexity scales exponentially in d.
2. $\lambda = O(\frac{1}{d})$: In this regime, as $d \to \infty$, we have $\tanh \lambda = O(\lambda)$ and therefore, D_1 scales as $d^2 \log p$. On the other hand, we have $D_2 \to 0$ as $d \to \infty$. Therefore, the bound on sample complexity scales as $d^2 \log p$.
2. $\lambda = \Theta(d)$: In this regime, as $d \to \infty$, D_1 scales as $\frac{\log p}{d}$ and D_2 scales exponentially in d. Clearly, D_2 dominates the bound on sample complexity.

Theorem 3 ($k = \Omega(p)$)**.** *Consider a pair of $\eta-$similar graphs $\mathcal{G}_1$ and $\mathcal{G}_2$ in the class $\bar{\mathcal{I}}$, such that, $d > 2$, $k = \omega(d^2)$ and $k \leq \frac{1}{2}p(d'-1)$ for some $d' \leq d$, and $\gamma \leq \min\left\{\frac{(d'-1)q}{2k}, \frac{2\binom{d}{2}q}{d+2}\right\}$. For any graph decoder $\psi : \mathcal{X}^{n\times p} \times \mathcal{X}^{n\times p} \to \bar{\mathcal{I}}$ that achieves*

$$\mathsf{P}(\bar{\mathcal{I}}, z) \leq \delta , \tag{14}$$

for $z = \lfloor \theta k \rfloor$ for some $\theta \in (0, \frac{d-2}{4d})$, the sample size n should satisfy

$$n \geq \max\{D_2, D_3\}(1 - \delta - o(1)) , \tag{15}$$

where we have defined D_2 in (13) *and*

$$D_3 \triangleq \frac{(1-\gamma/2)\log 2 - h(\theta)}{\lambda \frac{e^{2\lambda}\cosh(2\lambda d') - 1}{e^{2\lambda}\cosh(2\lambda d') + 1}} . \tag{16}$$

We list different regimes under which D_2 and D_3 have distinct scaling behavior in λ, p, d and k. In the following regimes, we assume that d and k increase with p.

1. $\lambda = \Theta(1)$: In this regime, as $d \to \infty$, we have $D_3 = O(1)$ and D_2 scales exponentially in d. Therefore, D_2 dominates the bound on sample complexity.
1. $\lambda = O(\frac{1}{d})$: In this regime, $D_2 \to 0$ as $d \to \infty$. The analysis of D_3 shows that D_3 scales as $\Omega(d \min\{d, (d/d')^2\})$ and therefore, D_3 character-

izes the bound on the sample complexity. Furthermore, when $d' = O(\sqrt{d})$, k scales according to $p\sqrt{d}$ and also, $\Omega\left(d \min\left\{d, \left(\frac{d}{d'}^2\right)\right\}\right) = \Omega(d^2)$. When $k = \Theta(pd)$, we have $\Omega\left(d \min\left\{d, \left(\frac{d}{d'}^2\right)\right\}\right) = \Omega\left(\frac{d^3 p^2}{k^2}\right)$.

3. $\lambda = \Theta(d)$: In this regime D_3 scales as $1/d$ and D_2 scales exponentially in d^2. Therefore, the bound on scaling complexity scales exponentially in d^2.

We note that the necessary conditions and sufficient conditions on the sample complexity for approximate recovery in the class $\bar{\mathcal{I}}$ scale at the same rate asymptotically in a specific regime which is formalized in Corollary 1.

Corollary 1 (Optimal Sample Complexity)**.** *When the maximum degree d is fixed and satisfies $d > 2$ and the maximum number of edges k is fixed and satisfies $k \leq p/4$, and we have*

$$\gamma \leq \min\left\{\frac{q}{2k}, \frac{2\binom{d}{2}q}{d+2}\right\} \quad \textit{and} \quad \lambda = O\left(\frac{1}{p}\right), \quad (17)$$

the exact scaling behavior of the sample complexity for approximate recovery in the class $\bar{\mathcal{I}}$ is $\Omega(p^2 \log p)$, as the size of the graph p grows.

Table 1 summarizes the non-exponential scaling behavior of the necessary and sufficient conditions under different regimes. Since η is fixed, it does not feature in the scaling behavior of the sample complexity.

Next, we note that the extreme cases of $\eta = 0$ and $\eta = 1$ simplify to the problem of approximate recovery of single graphs analyzed in [18]. However, in general, the necessary conditions on the sample complexity for the approximate recovery of graphs in the class $\bar{\mathcal{I}}$ are different from the existing results for single graphs. This observation is formalized in the following Corollary.

Corollary 2 (Special cases for approximate recovery)**.** *The necessary conditions on the sample complexity for the approximate recovery of partially similar graphs in the sub-class $\bar{\mathcal{I}}$ in the extreme cases of $\eta = 0$ and $\eta = 1$ subsume the existing results for single graphs.*

V. Conclusion

In this paper, we have analyzed the problem of structure learning under an approximate recovery criterion in the presence of side information about the structure. This setting is posed naturally as joint recovery of a pair of Ising models with partially identical edge structures. Therefore, any inference about the structure of one graph serves as the side information in the structure recovery of the other graph. For the degree and edge bounded subclass of Ising models, we have established the algorithm independent necessary (information-theoretic) and sufficient conditions on the sample complexity for achieving a bounded probability of error under the approximate recovery criterion. We have also investigated the scaling behavior of these conditions under different regimes and identified a specific regime that renders the optimal sample complexity.

References

[1] S. L. Lauritzen, *Graphical models.* Clarendon Press, May 1996, vol. 17.

[2] J. Pearl, *Causality*, 2nd ed. Cambridge University Press, UK, 2009.

[3] C. S. Won and H. Derin, "Unsupervised segmentation of noisy and textured images using Markov random fields," *CVGIP: Graphical models and image processing*, vol. 54, no. 4, pp. 308–328, Jul. 1992.

[4] X. Chen, F. J. Slack, and H. Zhao, "Joint analysis of expression profiles from multiple cancers improves the identification of microRNA–gene interactions," *Bioinformatics*, vol. 29, no. 17, pp. 2137–2145, 2013.

[5] J. Fang, L. S. Dongdong, S. Charles, Z. Xu, V. D. Calhoun, and Y.-P. Wang, "Joint sparse canonical correlation analysis for detecting differential imaging genetics modules," *Bioinformatics*, vol. 32, no. 15, pp. 3480–3488, 2016.

[6] A. Dobra, C. Hans, B. Jones, J. R. Nevins, G. Yao, and M. West, "Sparse graphical models for exploring gene expression data," *Journal of Multivariate Analysis*, vol. 90, no. 1, pp. 196–212, 2004.

[7] Y. Jacob, L. Denoyer, and P. Gallinari, "Learning latent representations of nodes for classifying in heterogeneous social networks," in *Proc. ACM international Conference on Web Search and Data Mining*, New York, Feb. 2014, pp. 373–382.

[8] K. Dvijotham, M. Chertkov, P. V. Hentenryck, M. Vuffray, and S. Misra, "Graphical models for optimal power flow," *Constraints*, vol. 22, no. 1, pp. 24–49, 2017.

[9] M. Rabbat, R. Nowak, and M. Coates, "Network tomography and the identification of shared infrastructure," in *Proc. Asilomar Conference on Signals, Systems and Computers*, Pacific Grove, CA, Nov. 2002, pp. 34–38.

[10] J. Guo, J. Cheng, E. Levina, G. Michailidis, and J. Zhu, "Estimating heterogeneous graphical models for discrete data with an application to roll call voting," *The Annals of Applied Statistics*, vol. 9, no. 2, pp. 821 – 848, Jun. 2015.

[11] D. M. Chickering, "Learning Bayesian networks is NP-complete," *Learning from data*, vol. 112, pp. 121–130, 1996.

[12] M. Yuan and Y. Lin, “Model selection and estimation in the Gaussian graphical model,” *Biometrika*, vol. 94, no. 1, pp. 19–35, 2007.

[13] A. J. Rothman, P. J. Bickel, E. Levina, and J. Zhu, “Sparse permutation invariant covariance estimation,” *Electronic Journal of Statistics*, vol. 2, pp. 494–515, 2008.

[14] P. Ravikumar, M. J. Wainwright, and J. D. Lafferty, “High-dimensional Ising model selection using ℓ_1-regularized logistic regression,” *The Annals of Statistics*, vol. 38, no. 3, pp. 1287–1319, Jun. 2010.

[15] O. Banerjee, L. E. Ghaoui, and A. d’Aspremont, “Model selection through sparse maximum likelihood estimation for multivariate Gaussian or binary data,” *Journal of Machine Learning Research*, vol. 9, pp. 485–516, Jun. 2008.

[16] N. P. Santhanam and M. J. Wainwright, “Information-theoretic limits of selecting binary graphical models in high dimensions.” *IEEE Transactions on Information Theory*, vol. 58, no. 7, pp. 4117–4134, May 2012.

[17] R. Tandon, K. Shanmugam, P. K. Ravikumar, and A. G. Dimakis, “On the information theoretic limits of learning Ising models,” in *Proc. Advances in Neural Information Processing Systems*, Montreal, Canada, Dec. 2014, pp. 2303–2311.

[18] J. Scarlett and V. Cevher, “On the difficulty of selecting Ising models with approximate recovery,” *IEEE Transactions on Signal and Information Processing over Networks*, vol. 2, no. 4, pp. 625–638, Dec. 2016.

[19] W. Wang, M. J. Wainwright, and K. Ramchandran, “Information-theoretic bounds on model selection for Gaussian Markov random fields,” in *Proc. IEEE International Symposium on Information Theory*, Austin, Texas, Jun. 2010.

[20] D. Vats and J. M. Moura, “Necessary conditions for consistent set-based graphical model selection,” in *Proc. IEEE International Symposium on Information Theory*, Saint-Petersburg, Russia, Jul. 2011, pp. 303–307.

[21] J. Guo, E. Levina, G. Michailidis, and J. Zhu, “Joint estimation of multiple graphical models,” *Biometrika*, vol. 98, no. 1, pp. 1–15, 2011.

[22] P. Danaher, P. Wang, and D. M. Witten, “The joint graphical lasso for inverse covariance estimation across multiple classes.” *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, vol. 76, no. 2, pp. 373–397, Mar. 2014.

[23] K. Mohan, P. London, M. Fazel, D. Witten, and S.-I. Lee, “Node-based learning of multiple Gaussian graphical models,” *The Journal of Machine Learning Research*, vol. 15, no. 1, pp. 445–488, 2014.

[24] S. Yang, Z. Lu, X. Shen, P. Wonka, and J. Ye, “Fused multiple graphical lasso,” *SIAM Journal on Optimization*, vol. 25, no. 2, pp. 916–943, 2015.

[25] C. B. Peterson, F. C. Stingo, and M. Vannucci, “Bayesian inference of multiple Gaussian graphical models,” *Journal of the American Statistical Association*, vol. 110, no. 509, pp. 159–174, 2015.

[26] H. Qiu, F. Han, H. Liu, and B. Caffo, “Joint estimation of multiple graphical models from high-dimensional time series,” *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, vol. 78, no. 2, pp. 487–504, 2016.

[27] S. Sihag and A. Tajer, “Sample complexity of joint structure learning,” in *Proc. IEEE International Conference on Acoustics, Speech and Signal Processing*, Brighton, UK, May 2019, pp. 5292–5296.

[28] ——, “Structure learning of similar ising models: Information-theoretic bounds,” in *Proc. IEEE International Symposium on Information Theory*, Paris, France, Jun. 2019, pp. 1307–1311.

[29] ——, “Structure learning with side information: Sample complexity,” in *Proc. Advances in Neural Information Processing Systems*. Vancouver, Canada: Curran Associates, Inc., Dec. 2019, pp. 14 357–14 367.