High-Dimensional Robust Structure Learning of Ising Models on Sparse Random Graphs

Anonymous Author(s)
Affiliation
Address
email

Abstract

This paper considers structure learning of ferromagnetic Ising models Markov on sparse Erdős-Rényi random graphs with constant average degree $c > 0$. We propose simple, local and robust algorithms and analyze their performances in the regime of correlation decay, i.e., when $c \tanh J_{\max} < 1$ (where $J_{\max}$ is the maximum inverse temperature in the model). The algorithms are robust because (i) they do not depend upon the specific model parameters such as the average degree and (ii) they provide guaranteed performance for a large class of sparse n-node Erdős-Rényi random graphs. We prove that a structure learning algorithm based on a set of conditional mutual information tests is consistent in high-dimensions throughout the regime of correlation decay provided the number of samples scales as $\omega(\log n)$. A simpler algorithm based on correlation thresholding outputs a graph with a constant edit distance to the original graph when there is correlation decay, and the sample complexity is $\Omega(\log n)$. Under a more stringent condition on the inverse temperatures ($2 \tanh^2 J_{\max} < \tanh J_{\min}$), correlation thresholding is also shown to be consistent for structure learning. Finally, we show that $\Omega(c \log n)$ samples is in fact necessary for consistent reconstruction by any algorithm. Thus, we establish that consistent structure estimation is possible with almost order-optimal sample complexity throughout the regime of correlation decay.

1 Introduction

Many naturally occurring networks (e.g., online social networks) are well-modeled by random graphs [1]. The connectivity of these networks is usually unknown and has to be learned from a set of incomplete or noisy observations. For real-world datasets, the learning of such networks is typically performed in the so-called high-dimensional regime, where the size of the network is much larger than the number of observations. It is thus imperative to design robust algorithms to learn large random structures with low sample and computational complexities.

The interactions among the variables can be modeled via graphical models [2], where the distribution factorizes according to a graph. In this paper, we consider Ising models Markov on sparse Erdős-Rényi random graphs [3]. We propose robust algorithms for structure learning of such random graphical models. The algorithms are robust in the sense that they are *agnostic to* the parameters of the random graph realization such as its maximum degree. This is in stark contrast to previously proposed algorithms which assume that the underlying graph is *deterministic*. In fact, one of the algorithms proposed in this paper is also completely parameter-free; it does not require any knowledge of the parameters of the random graph distribution (e.g., the average node degree). This is likened to robust control where stabilizing controllers have to be designed bearing in mind that the underlying plant belongs to a potential large uncertainty class. We provide sufficient conditions for the success of such robust and local structure learning algorithms. It is known that structure learning of tree-structured graphical models is tractable [4] while structure learning of general Markov models is NP-hard [5]. However, it is known [3] that sparse random Ising models are "locally tree-like" (i.e., a typical node is not part of a short cycle w.h.p.). Given this, our work explores if and when consistent learning is tractable for Ising models Markov on random graphs.

The questions addressed in this paper are: Given i.i.d. observations from the Ising model on sparse random graphs, are there robust, efficient and preferably local algorithms which can recover the underlying graph structure? If so, under what conditions on the model can we guarantee these algorithms to be consistent? What is the sample complexity of these algorithms? Is there a lower bound or a converse result on the sample complexity for reconstructing random graphs by *any* algorithm? The answers to these questions provide insights into the complex interplay between the graph's structure and parameters, and the resulting learning performance in large-scale models.

While many algorithms have been proposed for structure learning (see [6–9] and references therein), most of these algorithms require the graph to have a bounded maximum degree. Hence, such algorithms are not directly applicable for random graphs which are sparse on average but may contain nodes with large degrees. To the best of our knowledge, this paper presents first class of algorithms that guarantee consistent reconstruction of random graphs with growing maximum degree assuming correlation decay [10]. The notion of correlation decay has been extensively studied by the physics, computer science and statistics communities, and is related to many phenomena such as fast mixing of Markov chain Monte Carlo (MCMC) algorithms [10]. This work establishes that correlation decay is sufficient to ensure consistent structure learning in ferromagnetic (attractive) Ising models on random graphs.

1.1 Summary of Contributions

Our main contributions in this work are threefold. Firstly, we establish that consistent structure learning is achieved by a set of conditional mutual information tests in the regime of correlation decay. This algorithm requires number of samples to scale as $m = \omega(\log n)$ and it has a computational complexity of $O(n^5)$. This local algorithm independently tests whether a node pair forms an edge in the graph. Moreover, the algorithm is parameter-free, meaning that it does not require the knowledge of any parameters such as the bounds on the inverse temperatures of the Ising model. Secondly, we also analyze a simpler algorithm based on thresholding of pairwise correlations. Under correlation decay, we prove that this algorithm is consistent under an additional condition of homogeneity, given by $2\tanh^2 J_{\max} < \tanh J_{\min}$. The number of samples required is $m = \Omega(\log n)$. Finally, we demonstrate a lower bound on the sample complexity required for consistent reconstruction with positive probability by *any* algorithm. We prove that $\Omega(c \log n)$ samples is required to ensure consistent learning of Erdős-Rényi random graphs, where c is the average degree.

Because of space constraints, we refer the reader to the arXiv preprint [11] for the proofs of the theorems.

2 Problem Formulation: Notation, Graphical Models and Consistency

Notation Let $G_n = (V_n, E_n)$ be a (labeled) undirected graph where $V_n = \{1, 2, \ldots, n\}$ is the vertex set and $E_n \subset \binom{V_n}{2}$ is the edge set. Let $G_n \sim \mathcal{G}(n, c/n)$ for some $c > 0$ denote a particular realization of an Erdős-Rényi random graph. Here, the appearance probability of each edge is $\frac{c}{n}$, and thus the average node degree is c.

Denote the pairwise correlation between variables X_i and X_j as $C(i,j) := \mathbb{E}[X_i X_j]$. Given m samples (x_i^m, x_j^m) drawn i.i.d. from $P(x_i, x_j)$, the *empirical correlation* between X_i and X_j is defined as $\widehat{C}_{i,j}^m := m^{-1} \sum_{k=1}^m x_{i,k} x_{j,k}$. For two discrete probability distributions P and Q on a common state space $\mathcal{X}$, the *Kullback-Leibler (KL) divergence* [12] is $D(P||Q) = \sum_{x \in \mathcal{X}} P(x) \log(P(x)/Q(x))$. The *mutual information* of $(X, Y) \sim P(x, y)$ is $I(X;Y) = D(P(x,y)||P(x)P(y))$. The conditional mutual information of (X, Y) given Z is $I(X;Y|Z) = \sum_z P(z) I(X;Y|Z=z)$. Empirical versions of information-theoretic quantities are defined by substituting the empirical distributions into the respective formulae. We denote the empirical conditional mutual information as $\widehat{I}(X;Y|Z)$.

Ising Models A *graphical model* [2] is a family of multivariate probability distributions that factorize in accordance to a particular undirected graph $G = (V, E)$. Each node in the graph $i \in V$ is associated to a random variable X_i. The set of edges $E \subset \binom{V}{2}$ represent the set of conditional independence relations among the random variables. We say that a vector of random variables $\mathbf{X} := (X_1, \ldots, X_n)$ with distribution joint P is Markov on the graph G if the local Markov property $P(x_i | x_{\mathcal{N}(i)}) = P(x_i | x_{V \setminus i})$ holds for all nodes $i \in V$.

For Ising models, each random variable X_i takes values in the set $\mathcal{X} = \{-1, +1\}$. When the underlying graph G_n is random, we let $P_{\mathbf{X}|G_n}(\mathbf{x}|G_n; \mathbf{J})$ denote the Ising model probability mass function associated with random vector $\mathbf{X} = (X_1, \ldots, X_n)$ conditioned on a particular graph realization G_n. The joint distribution is $P_{\mathbf{X}|G_n}(\mathbf{x}|G_n; \mathbf{J}) \propto \exp(\sum_{(i,j) \in G_n} J_{i,j} x_i x_j)$ where $J_{i,j}$ is known as the *inverse temperature* for node pair (i, j). Also, by Markovity,

$J_{i,j} = 0$ for $(i,j) \notin G_n$. We also assume that there exists $J_{\min}, J_{\max} \in \mathbb{R}$ and independent of n such that the inverse temperatures on the edges are uniformly bounded, i.e., $J_{i,j} \in [J_{\min}, J_{\max}]$, for all edges $(i,j) \in G_n$ and all n.

Consistency Conditioned on a particular graph realization G_n, let $\mathbf{x}^m := \{\mathbf{x}_1, \ldots, \mathbf{x}_m\}$ be m i.i.d. samples drawn from $P_{\mathbf{X}|G_n}(\cdot|G_n; \mathbf{J})$. Given $\mathbf{x}^m \in (\mathcal{X}^n)^m$, a *graph estimator* is a (deterministic or random) function $\widehat{G}_n^m : (\mathcal{X}^n)^m \to \mathcal{G}_n$, where $\mathcal{G}_n$ is the set of graphs with n nodes. We say that $\widehat{G}_n^m$ is *consistent in high-dimensions* with sample complexity $f(n)$ if $\mathbb{P}(\widehat{G}_n^m \neq G) \to 0$ as $n, m > f(n) \to \infty$.

3 Method and Guarantees

We propose two algorithms for reconstructing random graphs. The first algorithm is based on simple correlation thresholding while the second algorithm is based on conditional mutual information thresholding. We provide performance guarantees for both these algorithms and in particular, show that the conditional mutual information test leads to consistent structure learning throughout the regime of correlation decay.

Assumptions We make the following assumptions on the model:

- (A1) **High Dimensionality:** We consider the asymptotic setting where both the number of nodes n and the number of samples m go to infinity in accordance to a particular scaling law. We require the number of nodes $n \to \infty$ to exploit the locally tree-like property of random graphs [3].
- (A2) **Random Graph:** We consider Ising models which are Markov on a realization of the random graph $G_n \sim \mathcal{G}(n, \frac{c}{n})$, for some constant $c > 0$. The average degree of such models is c.
- (A3) **Non-singularity:** The minimum inverse temperature on the edges of the Ising model satisfies $J_{i,j} \geq J_{\min} > 0$, for all $(i,j) \in G_n$. In other words, the Ising model is strictly *ferromagnetic*.
- (A4) **Regime of Correlation Decay:** The average node degree c and the maximum inverse temperature $J_{\max}$ satisfy $c \tanh J_{\max} < 1$. Such models belong to the *uniqueness regime* and have correlations decay [13].

Assumption (A1) is concerned with the sample complexity. The sample size needs to be sufficiently large with respect to the number of variables in the model. We show in Theorem 3 that $m = \Omega(c \log n)$ is necessary for consistent structure reconstruction. Assumption (A3) is required so that there are no "weak" edges with low correlations. Such assumptions have been made previously in other learning contexts, e.g., [6]. The ferromagnetic assumption in (A3) and correlation decay assumption in (A4) allow us to obtain bounds on correlations and mutual information quantities.

3.1 Correlation Thresholding Algorithm

We first describe the correlation thresholding algorithm, which is provided in Algorithm 1. We use the notation $\mathsf{CT}(\{\widehat{C}_{i,j}^m\}; \delta)$ to denote the output edge set based on a set of empirical correlations $\{\widehat{C}_{i,j}^m\}_{i,j \in V}$ and a threshold $\delta > 0$. The threshold is chosen as follows: Assuming that $J_{\min}$ and $J_{\max}$ are known[1], define the constants $\zeta(J_{\min}, J_{\max}) := \min\{k \geq 2 : 2(\tanh J_{\max})^k < \tanh J_{\min}\}$ and $\eta(J_{\min}, J_{\max}) := \tanh J_{\min} - 2(\tanh J_{\max})^{\zeta}$. For the CT algorithm, we choose the threshold δ according to $\delta(J_{\min}, J_{\max}) = \eta(J_{\min}, J_{\max})/2$. In the following, $\mathbb{P}$ denotes the probability measure associated to both the randomness in the graph G_n and the samples $\mathbf{x}^m$.

Theorem 1 (Structural consistency of CT) *Assume (A1) – (A4). For structure learning of the Ising model on the random graph $G_n = (V_n, E_n) \sim \mathcal{G}(n, \frac{c}{n})$, if $\zeta(J_{\min}, J_{\max}) = 2$ (i.e., $2(\tanh J_{\max})^2 < \tanh J_{\min}$), then the CT algorithm is consistent for structure reconstruction for a.e. G_n, i.e.,*

$$\lim_{\substack{m,n \to \infty \\ m > M \log n}} \mathbb{P}\left(\mathsf{CT}(\{\widehat{C}_{i,j}^m\}; \delta) \neq G_n\right) = 0. \tag{1}$$

If $\zeta(J_{\min}, J_{\max}) > 2$, for a.e. G_n, the edit distance between G_n and $\mathsf{CT}(\{\widehat{C}_{i,j}^m\}; \delta)$ is finite with

$$\lim_{\substack{m,n \to \infty \\ m > M \log n}} \mathbb{P}\left(|\mathsf{CT}(\{\widehat{C}_{i,j}^m\}; \delta) \setminus G_n| > \omega(1)\right) = 0, \quad \textit{and} \quad \lim_{\substack{m,n \to \infty \\ m > M \log n}} \mathbb{P}\left(|G_n \setminus \mathsf{CT}(\{\widehat{C}_{i,j}^m\}; \delta)| \neq 0\right) = 0. \tag{2}$$

[1] Our next algorithm CMIT, removes the assumption that these parameters are known and does not require the knowledge of *any* model parameters.

Algorithm 1 Algorithm $\mathsf{CT}(\{\widehat{C}^m_{i,j}\};\delta)$ given empirical correlations $\widehat{C}^m_{i,j}$ and threshold δ.

Correlation Thresholding: For each $i,j \in V$, if $\widehat{C}^m_{i,j} > \delta$, add (i,j) to $\widehat{G}^m_n$.
Output: $\widehat{G}^m_n$.

Algorithm 2 Algorithm $\mathsf{CMIT}(\mathbf{x}^m;\xi_{n,m})$ given samples $\mathbf{x}^m$ and threshold $\xi_{n,m}$.

For each $(i,j) \in \widehat{G}^m_n$, if $\min_{S\subset V\setminus\{i,j\},|S|\leq 3} \widehat{I}(X_i;X_j|\mathbf{X}_S) > \xi_{n,m}$, then add (i,j) to $\widehat{G}^m_n$.
Output: $\widehat{G}^m_n$.

Eqn. (1) implies that correlation thresholding is consistent when $2\tanh^2 J_{\max} < \tanh J_{\min}$ which means that the inverse temperatures $J_{i,j}$ on different edges are nearly homogeneous and sufficiently weak. However, without this condition, eqn. (2) implies that assuming correlation-decay, this algorithm results in a constant edit distance wrt the original graph. Correlation thresholding requires only pairwise statistics and not any higher order statistics, and has a low computational complexity of $O(n^2)$. Moreover, the algorithm has logarithmic sample complexity. In Section 3.3, we provide a matching lower bound for random graph reconstruction. The homogeneity condition for consistency of correlation thresholding ($2\tanh^2 J_{\max} < \tanh J_{\min}$) is less stringent than that needed for general degree-bounded graphs [14], given by $2\Delta \tanh J_{\max} < 1$. Since the maximum degree $\Delta = \Theta(\frac{\log n}{\log\log n})$ for a.e. $G_n \sim \mathcal{G}(n, c/n)$ [3, Ex. 3.6], the condition in [14] implies that the maximum inverse temperature has to decay, which is rather restrictive.

3.2 Conditional Mutual Information Thresholding

The correlation thresholding algorithm CT is consistent for graph reconstruction under an additional condition ($2\tanh^2 J_{\max} < \tanh J_{\min}$). We now propose an algorithm, termed as conditional mutual information thresholding (CMIT) which is proven to be consistent for graph reconstruction assuming (A1) - (A4).

The procedure for CMIT is provided in Algorithm 2. Denote $\mathsf{CMIT}(\mathbf{x}^m;\xi_{n,m})$ as the output edge set from CMIT given $\mathbf{x}^m$ and a threshold $\xi_{n,m}$. The conditional mutual information test in the CMIT algorithm computes the empirical conditional mutual information for each node pair $(i,j) \in \binom{V}{2}$ and finds the conditioning set which achieves the minimum over all sets of cardinality 3. If the minimum exceeds $\xi_{n,m}$, the node pair is declared as an edge.

The threshold $\xi_{n,m}$ is chosen as a function n and m. In particular, we choose $\xi_{n,n}$ satisfying three conditions: (i) $\xi_{n,m} = o_n(1)$, (ii) $\xi_{n,m} = \omega_n((\log n)^{-\kappa})$ for $\kappa > 0$ and (iii) $\xi_{n,m} = \Omega(\frac{\log n}{m})$. For example, when $m = O(g_n \log n)$, for some sequence $g_n = \omega(1)$, we can choose $\xi_{n,m} = (\min(g_n, \log\log n))^{-1}$.

Note that there is dependence on both m and n, since we need to regularize for sample size and the size of the graph. In other words, with finite number of samples m, the empirical conditional mutual information quantities are noisy and the threshold $\xi_{n,m}$ takes this into account via its inverse dependence on m. Similarly, as n increases, we show that the conditional mutual information decays at a certain rate. Hence, the threshold $\xi_{n,m}$ also depends on the graph size n. Moreover, note that for all the conditions on $\xi_{n,m}$ to be satisfied, m should scale at least at a certain rate w.r.t. n.

Theorem 2 (Structural consistency of CMIT) *Assume (A1) – (A4). For structure learning of the Ising model on the random graph $G_n = (V_n, E_n) \sim \mathcal{G}(n, \frac{c}{n})$, CMIT is consistent for a.e. graph G_n, i.e.,*

$$\lim_{\substack{m,n\to\infty \\ m>Mg_n\log n}} \mathbb{P}\left[\mathsf{CMIT}\left(\{\mathbf{x}^m\};\xi_{n,m}\right) \neq G_n\right] = 0. \tag{3}$$

An outline of the proof of this theorem is provided in Section 4. The CMIT algorithm is structurally consistent and has low sample complexity in the correlation-decay regime. Unlike CT algorithm, the CMIT requires higher order statistics since it involves the computation of (empirical) conditional mutual information. The CMIT algorithm has sample complexity that is slightly worse than the logarithmic sample complexity of CT. The computational complexity is $O(n^5)$. Although the Ising model assumed here is symmetric (zero external magnetic fields), the results for CMIT are identical for general Ising models with bounded external magnetic fields. The algorithms proposed here require correlation decay [assumption (A4)]. It is an open question on whether consistent reconstruction of random graphs is feasible with polynomial computational complexity and low sample complexity when there is no correlation decay.

3.3 Lower Bound on Sample Complexity

We have so far proposed algorithms and provided performance guarantees for random graph reconstruction. We now provide a lower bound on sample complexity for random graph reconstruction by any algorithm. In the following result, c is allowed to depend on n and is thus more general than the previous results.

Theorem 3 (Lower bound on sample complexity) *Assume that $c \leq 0.5n$ and $G_n \sim \mathcal{G}(n, c/n)$. Then if $m \leq \varepsilon c \log n$ for sufficiently small $\varepsilon > 0$, we have $\lim_{n\to\infty} \mathbb{P}(\widehat{G}_n^m \neq G_n) = 1$ for any graph estimator $\widehat{G}_n^m$.*

Thus, $\Omega(c \log n)$ samples are *necessary* for structure recovery. Hence, larger the average degree, higher is the sample complexity. Intuitively this is because as c grows, the graph is denser and hence, we require more samples for learning. In information-theoretic terms, our result is a strong converse, since we show that the error probability of structure learning tends to one (instead of being merely bounded away from zero). In [9], it is shown that for graphs uniformly drawn from the class of graphs with maximum degree Δ, if $m < \varepsilon \Delta^k \log n$, there exists a graph for which any estimator fails with probability at least 0.5. These results cannot be applied here since the probability mass function is non-uniform for the class of random graphs. The result is not dependent on the inverse temperature parameters $J_{i,j}$ (for example, the model need not be ferromagnetic). In fact, the result is not restricted to Ising models, and it holds for *any* pairwise discrete Markov random field (i.e., $\mathcal{X}$ is a finite set). The result also does not require locally-tree like property and is valid for dense graphs as well, i.e., c can be any function of n satisfying $c < 0.5n$.

4 Discussion and Proof Steps

Correlation Thresholding The correlation thresholding algorithm CT is based on the intuition that the correlations of variables connected by an edge tend to be larger than the correlations of variables that are not connected by an edge. Note that this property is true in (nearly) homogeneous Ising models on trees but need not hold for general graphical models. The key aspect is to establish bounds correlation by using the properties of paths and cycles in random graphs.

The main steps in the proof of Theorem 1 are as follows. Firstly, we recap some simple known results for ferromagnetic Ising models, for example the so-called Griffith's inequality [15], which give a lower bound on correlations. We also recap the construction of the *self-avoiding walk tree* (SAW) for general graphs [16]. Secondly, bounds on correlation between any node pair in a general graph are derived as a function of correlation on the shortest path in the graph, the number of paths between the two nodes within a chosen distance and the size of the boundary at the chosen distance. This is again accomplished via the SAW tree construction. Thirdly, the correlation bounds are specialized for random graphs by analyzing the properties of paths and cycles in random graphs. For a tight bound, we need to choose an appropriate boundary, to have small number of short paths and far enough boundary for correlations to decay. We crucially use the property that in a.e. random graph, there are no overlapping cycles of length less than $o(\log n / \log\log n)$ w.h.p. Equivalently, there are at most two "short" paths between any two nodes in the random graph. The boundary is chosen such that only these paths are counted towards the upper bound on correlation. Finally, using the bounds on correlation, the node pairs whose correlation exceed the threshold δ can be characterized. The result is extended to empirical correlations obtained from i.i.d. samples by standard concentration inequalities.

Conditional Mutual Information Thresholding The conditional mutual information test is based on the following property: for a fixed graphical model $P(\mathbf{x})$ Markov on graph $G = (V, E)$, given any two non-adjacent nodes, we have $I(X_u; X_v|\mathbf{X}_S) = 0$ for all $(u, v) \notin E$, where $S \subset V$ separates u and v. On the other hand, for adjacent nodes we have $I(X_u; X_v|\mathbf{X}_S) > 0$ for all $(u, v) \in E$ and all $S \subset V$, for the Ising model under consideration with $0 < J_{\min}, J_{\max} < \infty$, as assumed earlier. This is, in general, not true for general graphical models.

These mutual information relationships can be used for graph reconstruction as follows: for any node pair u, v, find the minimum conditional mutual information by conditioning on all possible subsets. This is efficient only if the cardinalities of the separators S are bounded, otherwise this requires exponential time in n. Moreover, when only samples are available, computing empirical mutual information, even for a specific conditioning set, requires exponential computational and sample complexities in the size of the conditioning set [7]. Hence, the conditional mutual information criteria cannot be used directly used for efficient structure estimation.

Instead, we show that for Ising models on random graphs under correlation decay, there exist a sparse *approximate separator* for any non-adjacent node pair. Such an approximate separator is a subset of the exact separator, consisting only of nodes on "short paths". Under the correlation-decay condition, we show that the approximate-separator set suffices to distinguish true edges from non-adjacent node pairs. The key steps in the proof of Theorem 2 are as follows:

Firstly, we show using information-theoretic inequalities that the conditional mutual information between any two nodes on general graphs is related to total variation distance on the SAW tree [16]. Bounds on the conditional mutual information on the SAW tree are then derived on similar lines, as in the proof of Theorem 1. Secondly, asymptotically almost-sure (a.a.s.) bounds on the size of approximate-separator sets in $\mathcal{G}(n, \frac{c}{n})$ random graphs are derived. This argument is based on the a.a.s. number of overlapping short cycles in random graphs. Note that this is the reason we are able to constrain the conditioning set S for the mutual information in CMIT algorithm to be of cardinality 3.

Lower Bound on Random Graph Reconstruction The proof for the lower bound (converse) uses an information-theoretic covering-lemma like argument in [8, Theorem 1]. We show that the error probability for structure estimation is lower bounded by the likelihood of graphs occurring in the complement of the range of the estimator. Note that the size of the range of any estimator is at most $|\mathcal{X}|^{nm}$, where n is the size of the graph and m is the number of samples and $|\mathcal{X}| = 2$ for Ising models. We cover the range of the estimator by the collection of sparse random graphs with large likelihoods. Using standard bounds on the tails of binomial distributions, we obtain the desired result.

5 Conclusion

In this work, we considered structure learning of Ising models on sparse random graphs. We proved that our proposed algorithms are robust to the uncertainty in the underlying graphical structure. The algorithms are also consistent in high-dimensions and computationally tractable with almost order-optimal sample complexity throughout the regime of correlation decay. There are many other open questions at the intersection of statistical learning theory and statistical physics. Perhaps the most important one is the feasibility of high-dimensional structure learning of graphical models beyond the regime of correlation decay and whether simple robust algorithms can accomplish such a learning task.

References

[1] M. E. J. Newman, D. J. Watts, and S. H. Strogatz, "Random graph models of social networks," *Proc. Natl. Acad. Sci. USA*, vol. 99, pp. 2566–2572, 2002.

[2] S. Lauritzen, *Graphical models: Clarendon Press*. Clarendon Press, 1996.

[3] B. Bollobás, *Random Graphs*. Academic Press, 1985.

[4] C. Chow and C. Liu, "Approximating Discrete Probability Distributions with Dependence Trees," *IEEE Tran. on Information Theory*, vol. 14, no. 3, pp. 462–467, 1968.

[5] D. Karger and N. Srebro, "Learning markov networks: maximum bounded tree-width graphs," in *Symposium on Discrete Algorithms*, 2001, pp. 392–401.

[6] P. Ravikumar, M. Wainwright, and J. Lafferty, "High-dimensional Ising Model Selection Using l1-Regularized Logistic Regression," *Annals of Statistics*, 2008.

[7] A. Chechetka and C. Guestrin, "Efficient Principled Learning of Thin Junction Trees," in *Advances in Neural Information Processing Systems (NIPS)*, 2007.

[8] G. Bresler, E. Mossel, and A. Sly, "Reconstruction of Markov Random Fields from Samples: Some Observations and Algorithms," in *Intl. workshop APPROX Approximation, Randomization and Combinatorial Optimization*. Springer, 2008, pp. 343–356.

[9] N. Santhanam and M. Wainwright, "Information-theoretic Limits of High-dimensional Model Selection," in *International Symposium on Information Theory*, Toronto, Canada, July 2008.

[10] M. Mezard and A. Montanari, *Information, physics, and computation*. Oxford University Press, USA, 2009.

[11] A. Anandkumar, V. Y. F. Tan, and A. S. Willsky, "High-Dimensional Structure Learning of Ising Models on Sparse Random Graphs," *Preprint. Available on ArXiv:1011.0129*, Oct. 2010.

[12] T. Cover and J. Thomas, *Elements of Information Theory*. John Wiley & Sons, Inc., 1991.

[13] H. Georgii, *Gibbs Measures and Phase Transitions*. Walter de Gruyter, 1988.

[14] J. Bento and A. Montanari, "Which Graphical Models are Difficult to Learn?" in *Proc. of Neural Information Processing Systems (NIPS)*, Vancouver, Canada, Dec. 2009.

[15] R. Griffiths, "Correlations in Ising ferromagnets. III," *Communications in Mathematical Physics*, vol. 6, no. 2, pp. 121–127, 1967.

[16] D. Weitz, "Counting independent sets up to the tree threshold," in *Proc. of ACM symp. on Theory of computing*, 2006, pp. 140–149.