Structure Learning of H-colorings

Antonio Blanca * ABLANCA@CC.GATECH.EDU
Zongchen Chen * CHENZONGCHEN@GATECH.EDU
Eric Vigoda * VIGODA@GATECH.EDU
Georgia Institute of Technology, USA

Daniel Štefankovič † STEFANKO@CS.ROCHESTER.EDU
University of Rochester, USA.

Abstract

We study the structure learning problem for H-colorings, an important class of Markov random fields that capture key combinatorial structures on graphs, including proper colorings and independent sets, as well as spin systems from statistical physics. The learning problem is as follows: for a fixed (and known) constraint graph H with q colors and an unknown graph $G = (V, E)$ with n vertices, given uniformly random H-colorings of G, how many samples are required to learn the edges of the unknown graph G? We give a characterization of H for which the problem is identifiable for every G, i.e., we can learn G with an infinite number of samples. We also show that there are identifiable constraint graphs for which one cannot hope to learn every graph G efficiently.

We focus particular attention on the case of proper vertex q-colorings of graphs of maximum degree d where intriguing connections to statistical physics phase transitions appear. We prove that in the tree uniqueness region (i.e., when $q > d$) the problem is identifiable and we can learn G in $\mathsf{poly}(d, q) \times O(n^2 \log n)$ time. In contrast for soft-constraint systems, such as the Ising model, the best possible running time is exponential in d. In the tree non-uniqueness region (i.e., when $q \leq d$) we prove that the problem is not identifiable and thus G cannot be learned. Moreover, when $q < d - \sqrt{d} + \Theta(1)$ we prove that even learning an equivalent graph (any graph with the same set of H-colorings) is computationally hard—sample complexity is exponential in n in the worst case. We further explore the connection between the efficiency/hardness of the structure learning problem and the uniqueness/non-uniqueness phase transition for general H-colorings and prove that under a well-known condition in statistical physics, known as the Dobrushin uniqueness condition, we can learn G in $\mathsf{poly}(d, q) \times O(n^2 \log n)$ time.

Keywords: structure learning, identifiability, Markov random fields, H-colorings, spin systems.

1. Introduction

Structure learning is a general framework for supervised learning where, instead of learning labels or real numbers as in classification or regression, the task is to learn a more complex structure such as a graph. A myriad of fundamental learning problems can be studied in this framework. Notably, structure learning for Markov random fields (undirected graphical models), where the goal is to recover the underlying graph from random samples, has found important applications in diverse fields, including the study of phylogeny [29], gene expression [35], protein interactions [38], neuroscience [42], image processing [40] and sociology [20].

* Research supported in part by NSF grants CCF-1617306 and CCF-1563838.
† Research supported in part by NSF grant CCF-1318374.

Our goal in this paper is to understand when is structure learning for Markov random fields possible in polynomial time. We mostly focus on the task of *exact recovery*, where a learning algorithm is said to succeed only when it outputs exactly the hidden graph. In applications, exact recovery of graphical models is often of interest since the true graph structure contains valuable information about the dependencies in the model. Consequently, it has been very well-studied; see, e.g., [12, 14, 46, 13, 33, 2, 30, 39, 5, 7, 3, 47, 26, 32]. While the typical setting in these works are *soft-constraint models*, our focus here are models with *hard constraints*. Specifically, we consider the structure learning problem and the closely related question of statistical identifiability in the general setting of H-colorings, an important class of Markov random fields that include all hard-constraint models and capture key combinatorial structures on graphs, including proper colorings and independent sets.

Given an undirected, connected constraint graph $H = (V(H), E(H))$, with vertices $V(H) = \{1, \dots, q\}$ referred to as colors (or spins), an H-coloring of a graph $G = (V, E)$ is an assignment of colors $\{1, \dots, q\}$ to the vertices of G such that adjacent vertices of G receive adjacent colors in H. That is, an H-coloring σ is a mapping $\sigma : V \to V(H)$ such that if $\{v, w\} \in E$, then $\{\sigma(v), \sigma(w)\} \in E(H)$. If such an assignment is possible we say that G is H-colorable. The constraint graph H is allowed to have self-loops, but not parallel edges, and every $\{i, j\}$ such that $\{i, j\} \notin E(H)$ is called a *hard constraint*.

When, for example, H is the complete graph on q vertices with no self-loops, denoted K_q, neighboring vertices of G must be assigned different colors, and thus the proper q-colorings of G are precisely its H-colorings. If, on the other hand, H is the graph with two vertices $V(H) = \{0, 1\}$ and two edges $E(H) = \{\{0, 0\}, \{0, 1\}\}$, then the subset of vertices assigned color 1 in any H-coloring of G form an independent set. Hence, in this case, there is one-to-one correspondence between the independent sets and the H-colorings of G. Spin systems without hard constraints, i.e., *soft-constraint systems*, correspond to the constraint graph with all possible edges; that is, $H = K_q^+$ which is the complete graph K_q with a self-loop at every vertex. In this case, all $q^{|V|}$ labelings of G are valid H-colorings.

We consider structure learning and statistical identifiability for H-colorings with at least one hard constraint; that is, $H \neq K_q^+$. (Note that the missing edge could be a self-loop.) H-colorings are well-studied in several other contexts. For the decision problem (for a fixed H is a graph H-colorable?), a dichotomy result [27, 28, 10, 43] has been established characterizing for which H the problem is either in P or is NP-complete. The corresponding dichotomy conjecture for the directed case (directed graph homomorphisms) has also received considerable attention [22]; see also the recent works [11, 49] and the references therein. The complexity of the exact counting version of the problem was characterized by Dyer and Greenhill [17], and the complexity of approximate counting/sampling was studied in [25, 18, 23].

For an H-colorable graph G, let Ω_G^H be the set of all possible H-colorings of G and let π_G^H denote the uniform measure over Ω_G^H. (Typically the constraint graph H will be fixed and thus we will drop the dependence on H in our notation.) Some of our results extend to the more general setting of a weighted constraint graph H and a weighted graph G, where π_G^H is the corresponding Gibbs distribution; see Appendix E for a precise definition.

For statistical identifiability our goal is to characterize the cases when every graph is learnable with an infinite number of samples. For structure learning our goal is to efficiently learn the graph G from samples drawn independently from π_G. More formally, let H be a fixed constraint graph and let $\mathcal{G}$ be a family of H-colorable graphs. Suppose that we are given L samples $\sigma^{(1)}, \sigma^{(2)}, \dots, \sigma^{(L)}$

drawn independently from the distribution π_G^H where $G \in \mathcal{G}$. A *structure learning algorithm* for the constraint graph H and the graph family $\mathcal{G}$ takes as input the sample sequence $\sigma^{(1)}, \sigma^{(2)}, \ldots, \sigma^{(L)}$ and outputs an estimator $\hat{G} \in \mathcal{G}$ such that $\Pr[G = \hat{G}] \geq 1 - \varepsilon$ where $\varepsilon > 0$ is a prescribed error (failure probability).

As mentioned earlier, structure learning has been well-studied for soft-constraint models where $H = K_q^+$. In the context of the Ising model, the most well-known and widely studied soft-constraint system, for an unknown graph G with n vertices, maximum degree d and maximum interaction strength β (which corresponds to the inverse temperature in the homogenous model), Bresler [3] presented an algorithm to learn G in $O(n^2 \log n) \times \exp(\exp(O(\beta d^2)))$ time. A different algorithm was provided by Vuffray et al. [47] with running time $O(n^4 \log n) \times \exp(O(\beta d))$. Recently, Bresler's algorithm was extended to arbitrary Markov random fields by Hamilton et al. [26], and a new approach was presented by Klivans and Meka [32] which achieves nearly-optimal running time of $O(n^2 \log n) \times \exp(O(\beta d))$. Both of these general results [26, 32] are for the case of soft-constraint models and do not apply to the setting of hard-constraint systems. We shall see that, while the algorithm in [32] achieves (optimal) single exponential dependence on d for general soft-constraint systems, the structure learning problem for hard-constraint systems is quite different. Indeed, some hard-constraint systems are not statistically identifiable (and thus the unknown graph G cannot be learned); others allow very efficient structure learning algorithms with $\mathsf{poly}(n, d, q)$ running time; while in others any structure learning algorithm requires exponentially (in n) many samples.

For hard-constraint systems, the structure learning problem was previously studied by Bresler, Gamarnik, and Shah [6] for independent sets (more generally, for the hard-core model where the independent sets are weighted by their size and a model parameter $\lambda > 0$). They achieve nearly-optimal running time of $O(n^2 \log n) \times \exp(O(d\lambda))$. For our positive results we generalize the structure learning algorithm in [6].

Finally, we remark that while our results aim to learn the underlying graph G exactly (i.e., exact recovery), in some cases, we also consider the problem of learning an equivalent graph G' such that $\pi_G = \pi_{G'}$ (equivalent-structure learning). The corresponding approximation problem of finding a graph G' such that $\pi_{G'}$ is close to π_G in some notion of distance, such as total variation distance or Kullback-Leibler divergence (see, e.g., [1, 4]), is apparently much simpler for hard-constraint systems; see Appendix D.

1.1. Results

We first address the statistical identifiability problem for general H-colorings.

Definition 1 *A constraint graph H is said to be identifiable with respect to a family of H-colorable graphs $\mathcal{G}$ if for any two distinct graphs $G_1, G_2 \in \mathcal{G}$ we have $\pi_{G_1} \neq \pi_{G_2}$ (or equivalently $\Omega_{G_1} \neq \Omega_{G_2}$). In particular, when $\mathcal{G}$ is the set of all finite H-colorable graphs we say that H is identifiable.*

To characterize identifiability we introduce a supergraph G_{ij} of H. This supergraph will not be used as a constraint graph, but rather we will consider the H-colorings of G_{ij} in our characterization theorem. Consider an edge $\{i, j\} \in E(H)$. We construct G_{ij} by starting from H and duplicating the colors i and j. These new copies, denoted i' and j', have the same neighbors as the original colors i and j, respectively, except for the one edge $\{i', j'\}$ which is not included; see Figure 1 for an illustration of this supergraph and Definition 6 for a formal definition.

For a constraint graph H we say that a pair of colors $i, j \in V(H)$ are compatible (resp., incompatible) if $\{i, j\} \in E(H)$ (resp., $\{i, j\} \notin E(H)$). Also, if σ is an H-coloring of a graph $G = (V, E)$, we use $\sigma(v)$ to denote the color of $v \in V$ in σ. Our characterization theorem considers whether in every H-coloring of G_{ij} the new vertices i', j' receive compatible colors.

Theorem 2 *Let $H \neq K_q^+$ be an arbitrary constraint graph. If H has at least one self-loop, then H is identifiable. Otherwise, H is identifiable if and only if for each $\{i, j\} \in E(H)$ there exists an H-coloring σ of G_{ij} such that $\sigma(i')$ and $\sigma(j')$ are incompatible colors in H.*

The intuition for the role of the graph G_{ij} in our characterization theorem is that if we added the edge $\{i', j'\}$ to G_{ij} to form the graph $G'_{ij} = G_{ij} \cup \{i', j'\}$, then in every H-coloring of G'_{ij} the vertices i', j' receive compatible colors. If the same property holds for G_{ij}, i.e., in every H-coloring of G_{ij} the vertices i', j' receive compatible colors, then the edge $\{i', j'\}$ plays no role and the pair of graphs G_{ij} and G'_{ij} have the same set of H-colorings. That is, G_{ij} and G'_{ij} are indistinguishable, and so H is not identifiable.

Let $\mathcal{G}(n, d)$ be the family of n-vertex graphs of maximum degree at most d. Our next result shows that for some identifiable (with respect to $\mathcal{G}(n, d)$) constraint graphs, one cannot hope to learn the underlying graph G efficiently. We remark that this is not the case for soft-constraint models, where one can always learn G in time $O(n^2 \log n) \times \exp(O(d\beta))$ [32].

Theorem 3 *There exists an identifiable constraint graph H and a constant $c > 0$ such that, for all $n \geq 8$, any structure learning algorithm for H and the graph family $\mathcal{G}(n, 7)$ that succeeds with probability at least $\exp(-cn)$ requires at least $\exp(cn)$ samples.*

While this theorem shows that there is no efficient learning algorithm for all identifiable models, for some relevant models structure learning can be done efficiently.

We focus first on the case of proper q-colorings where $H = K_q$. In general, the colorings problem is not identifiable. However, if we consider identifiability with respect to the graph family $\mathcal{G}(n, d)$ we get a richer picture. We prove that when $q > d$ the q-colorings problem is identifiable, whereas when $q \leq d$ it is not. This phase transition for identifiability/non-identifiability at $q = d$ is quite intriguing since it coincides with the statistical physics uniqueness/non-uniqueness phase transition of the Gibbs distribution on infinite d-regular trees.

Uniqueness on an infinite d-regular tree T may be defined as follows. Let T_ℓ be a finite d-regular tree of height ℓ rooted at vertex r and consider the uniform measure μ_ℓ over the proper q-colorings of T_ℓ. We say that the Gibbs measure on T is unique if, as $\ell \to \infty$, the configuration in the leaves of T_ℓ has no influence on the color of r. Formally, the Gibbs measure on T is unique iff for every color $c \in \{1, \ldots, q\}$ and every configuration τ on the leaves of T_ℓ, we have $\mu_\ell(r = c|\tau) \to 1/q$ as $\ell \to \infty$. That is, the conditional distribution at the root given τ is the uniform measure over $\{1, \ldots, q\}$ as $\ell \to \infty$. In [31, 8] it was shown that there is a unique Gibbs measure iff $q > d$. The uniqueness/non-uniqueness phase transition on the infinite d-regular tree is known to be closely connected to the efficiency/hardness of other fundamental computational problems, such as sampling and counting; see, e.g., [48, 44, 45, 24, 34].

In the identifiable region $q > d$ we present an efficient structure learning algorithm with $O(n^2 \log n) \times \mathsf{poly}(d, q)$ running time. When $q \leq d$, where we cannot hope to learn the hidden graph G since there are pairs of graphs with the same set of H-colorings, we may be interested in a learning algorithm that outputs a graph G' that is equivalent to the unknown graph G, in the

sense that $\Omega_{G'} = \Omega_G$. We say that an algorithm is an *equivalent-structure learning algorithm* for a fixed constraint graph H and a graph family $\mathcal{G}$ if for every $G \in \mathcal{G}$, with probability at least $1 - \varepsilon$, the algorithm outputs $G' \in \mathcal{G}$ such that $\Omega_G = \Omega_{G'}$.

In the q-coloring setting, it turns out that when $q \leq d - \sqrt{d} + \Theta(1)$ there is a family of exponentially many n-vertex graphs with different sets of q-colorings that only differ on an exponentially small (in n) fraction of their q-colorings. Consequently, any equivalent-structure learning algorithm requires exponential many samples to distinguish among these graphs.

Our results for statistical identifiability, structure learning and equivalent-structure learning for proper q-colorings are stated in the following theorem.

Theorem 4 *Consider the q-colorings problem $H = K_q$. The following hold for all d.*

1. ***Efficient learning for** $q > d$**:** For all $q > d$, $n \geq 1$ and any $G \in \mathcal{G}(n, d)$, there is a structure learning algorithm that given $L = O(qd^3 \log(\frac{n^2}{\varepsilon}))$ independent samples from π_G outputs G with probability at least $1 - \varepsilon$ and has running time $O(Ln^2)$.*

2. ***Non-identifiability for** $q \leq d$**:** For all $q \leq d$ and $n \geq q + 2$ there exist q-colorable graphs $G_1, G_2 \in \mathcal{G}(n, d)$ such that $G_1 \neq G_2$ and $\pi_{G_1} = \pi_{G_2}$.*

3. ***Lower bound for** $q < d - \sqrt{d} + \Theta(1)$**:** For all $q < d - \sqrt{d} + \Theta(1)$ there exists a constant $c > 0$ such that any equivalent-structure learning algorithm for the family of graphs $\mathcal{G}(n, d)$ that succeeds with probability at least $\exp(-cn)$ requires at least $\exp(cn)$ samples, provided n is sufficiently large.*

As mentioned earlier, the phase transition for identifiability/non-identifiability at $q = d$ coincides with the statistical physics uniqueness/non-uniqueness phase transition. The lower bound when $q < d - \sqrt{d} + \Theta(1)$ is also quite curious since it coincides exactly with the threshold for polynomial-time/NP-completeness for the decision problem [21, 37]. In fact, the graph used in the proof of Part 3 of Theorem 4 is inspired by the graph used in the NP-completeness proof in [21] and the graph theoretic result in [37].

Since our results for the structure learning problem in the setting of q-colorings suggest an intimate connection between the efficiency/hardness of the learning problem and the uniqueness/non-uniqueness phase transition of the Gibbs distribution, we further explore this connection for general H-colorings. A sufficient condition for uniqueness on general graphs is the *Dobrushin uniqueness condition* [16], which is a standard tool in statistical physics for establishing uniqueness of the Gibbs distribution.

Dobrushin's condition considers the so-called influence matrix R where the entry R_{vw} measures the worst-case influence of vertex w on v. In particular, consider all pairs of color assignments τ, τ_w that differ only at vertex w. R_{vw} is the maximum (over these pairs τ, τ_w) of the difference in total variation distance of the marginal distribution at v conditional on the color assignment $\tau(V \setminus v)$ versus $\tau_w(V \setminus v)$. Observe that $R_{vw} = 0$ for non-adjacent pairs v, w. The Dobrushin uniqueness condition holds if the maximum row sum in R is strictly less than 1, so that the total influence on a vertex of its neighborhood is less than 1; see Definition 16 for a precise definition.

We prove that if the Dobrushin uniqueness condition holds, then we can learn the underlying n-vertex graph in $\mathsf{poly}(n, q)$ time.

Theorem 5 *Let $H \neq K_q^+$ be an arbitrary constraint graph. Suppose G is such that π_G satisfies the Dobrushin uniqueness condition. Then, there is a structure learning algorithm that given $L =$*

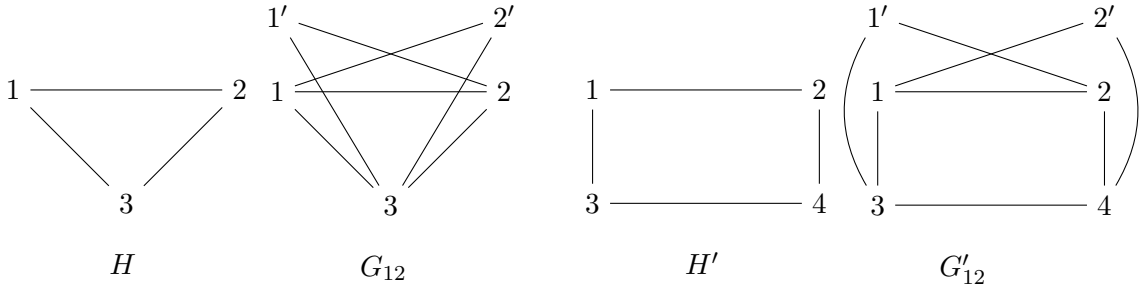

Figure 1: Two constraint graphs H and H' with corresponding supergraphs G_{12} and G'_{12}.

$O(q^2 \log(\frac{n^2}{\varepsilon}))$ *independent samples from* π_G *outputs the graph* G *with probability at least* $1 - \varepsilon$ *and has running time* $O(Ln^2)$.

The above theorem provides an efficient structure learning algorithm under a fairly strong model assumption. For soft-constraint systems, under a similar but weaker (i.e., easier to satisfy) condition, Bresler, Mossel and Sly [5] give a structure learning algorithm with running time *exponential* in the maximum degree d of the graph G. We provide a structure learning algorithm for hard-constraint systems with a similar running time that works under an even weaker assumption. Our algorithm works for *permissive systems*, which are a class of hard-constraint models widely studied in statistical physics; see, e.g., [19, 36, 15]. The precise definition, as well as the running time and sample complexity of our algorithm, are provided in Appendix E.3.

The rest of the paper is organized as follows. In Section 2 we prove our characterization theorem (Theorem 2) and our learning lower bound for identifiable models (Theorem 3). In Section 3, we prove our results for colorings (Theorem 4). In particular, in Section 3.1 we introduce a general structure learning algorithm STRUCTLEARN-H which will be the basis of all our algorithmic result. Our poly(n, d, q)-time algorithm under the Dobrushin uniqueness condition (Theorem 5) is established in Section 4. In Appendix D we consider the approximation problem of learning a graph G such that π_G is close in total variation distance to the true distribution. Finally, the case of weighted H and G is considered in Appendix E.

2. Identifiability

As discussed in the introduction, given a constraint graph H, it is possible that $\pi_{G_1} = \pi_{G_2}$ for two distinct H-colorable graphs G_1 and G_2; i.e., the structure learning problem is not identifiable. In this section we prove Theorem 2 from the introduction that characterizes the identifiable constraint graphs.

Let $\mathcal{G}$ be a family of H-colorable graphs. Recall that a constraint graph H is *identifiable with respect to* $\mathcal{G}$ if for any two distinct graphs $G_1, G_2 \in \mathcal{G}$ we have $\pi_{G_1} \neq \pi_{G_2}$ or equivalently $\Omega_{G_1} \neq \Omega_{G_2}$. In particular, when $\mathcal{G}$ is the set of all finite H-colorable graphs we say that H is identifiable; see Definition 1.

To characterize identifiability of H-colorings we previously introduced the supergraph G_{ij} of H obtained by duplicating the colors i and j. The new copies of i and j, denoted i' and j', have the same neighbors as i and j, respectively, except for the one edge $\{i', j'\}$ that is not present in G_{ij}—see Figure 1 for examples of the graph G_{ij}. We provide next the formal definition of G_{ij}.

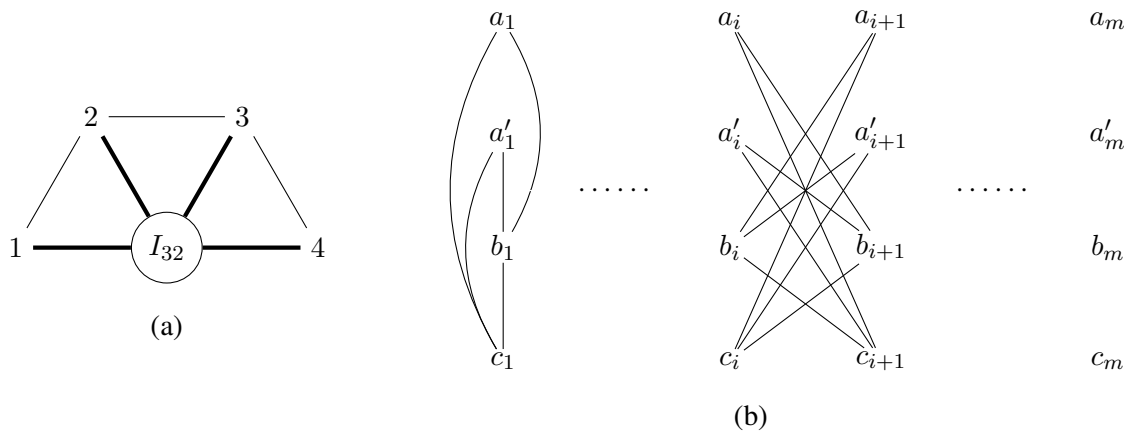

Figure 2: (a) The constraint graph F; each thick edge corresponds to 32 edges, one incident to each vertex of I_{32}. (b) The graph G_m showing the edges in $E_{i,i+1}$ and the connections between the vertices a_1, a'_1, b_1, c_1.

Definition 6 *Let $H = (V(H), E(H))$ be an arbitrary constraint graph with no self-loops. For each $\{i, j\} \in E(H)$, we define the graph $G_{ij} = (V(G_{ij}), E(G_{ij}))$ as follows:*

1. *$V(G_{ij}) = V(H) \cup \{i', j'\}$ where i' and j' are two new colors;*
2. *If $\{a, b\} \in E(H)$, then the edge $\{a, b\}$ is also in $E(G_{ij})$;*
3. *For each $k \in V(G_{ij}) \setminus \{i', j'\}$, the edge $\{i', k\}$ is in G_{ij} if and only if the edge $\{i, k\}$ is in H, and similarly $\{j', k\} \in E(G_{ij})$ if and only if $\{j, k\} \in E(H)$;*

Our characterization of identifiability, i.e., Theorem 2 from the introduction, is established by the next two lemmas. Lemma 7 deals with the case of constraint graphs with at least one self-loop, while Lemma 8 considers constraint graphs with no self-loops.

Lemma 7 *If $H \neq K_q^+$ has at least one self-loop, then H is identifiable.*

Lemma 8 *If $H \neq K_q^+$ has no self-loops, H is identifiable if and only if for each $\{i, j\} \in E(H)$ there exists an H-coloring of G_{ij} where i' and j' receive incompatible colors.*

The proofs of Lemmas 7 and 8 are provided in Appendix A.

2.1. Learning lower bounds for identifiable models

In this section we prove Theorem 3 from the introduction. In particular, we provide a constraint graph F and a family of F-colorable n-vertex graphs of maximum degree 7 such that the number of samples from π_G^F required to learn any graph in this family, even with success probability $\exp(-O(n))$, is exponential in n.

We define the constraint graph $F = (V(F), E(F))$ first, which consists of an independent set of size 32, denoted I_{32}, and four additional vertices $\{1, 2, 3, 4\}$. Every vertex in the independent set I_{32} is connected to these four vertices and also $\{1, 2\}, \{2, 3\}, \{3, 4\} \in E(F)$; see Figure 2(a).

Lemma 9 *F is identifiable.*

We define next a family $\mathcal{G}_n$ of n-vertex graphs of maximum degree 7 such that every graph in the family have almost the same set of F-colorings. Every graph in $\mathcal{G}_n$ will be a supergraph of the graph $G_m = (V_m, E_m)$, whose vertex set is given by $V_m = \{a_i, a'_i, b_i, c_i : 1 \leq i \leq m\}$ with $m = n/4$. (For clarity we assume first that 4 divides n; we later explain how to adjust the definition of $\mathcal{G}_n$ when n is not divisible by 4.) For each $2 \leq i \leq m$, $\{a_i, a'_i, b_i, c_i\}$ is an independent set. The edges with both endpoints in $\{a_1, a'_1, b_1, c_1\}$ are:

$$E_{1,1} := \{\{a_1, b_1\},\ \{a_1, c_1\},\ \{a'_1, b_1\},\ \{a'_1, c_1\}, \{b_1, c_1\}\}.$$

The edges between the independent sets $\{a_i, a'_i, b_i, c_i\}$ and $\{a_{i+1}, a'_{i+1}, b_{i+1}, c_{i+1}\}$ for $1 \leq i < m$ are:

$$\begin{aligned} E_{i,i+1} := \big\{&\{a_i, b_{i+1}\}, \{a_i, c_{i+1}\}, \{a'_i, b_{i+1}\}, \{a'_i, c_{i+1}\}, \{b_i, a_{i+1}\}, \\ &\{b_i, a'_{i+1}\}, \{b_i, c_{i+1}\}, \{c_i, a_{i+1}\}, \{c_i, a'_{i+1}\}, \{c_i, b_{i+1}\}\big\}. \end{aligned}$$

We then let $E_m = \left(\bigcup_{i=1}^{m-1} E_{i,i+1}\right) \cup E_{1,1}$; see Figure 2(b).

Now, let

$$M = \{\{a_i, b_{i+2}\} : i = 1, 2, \ldots, m-2\} \tag{1}$$

and let $E^{(1)}, E^{(2)}, \ldots, E^{(t)}$ be all the possible subsets of M; hence $t = 2^{m-2}$. We define $\mathcal{G}_n$ as:

$$\mathcal{G}_n = \{G^{(1)} = (V_m, E_m \cup E^{(1)}), G^{(2)} = (V_m, E_m \cup E^{(2)}), \ldots, G^{(t)} = (V_m, E_m \cup E^{(t)})\}.$$

Since the maximum degree of G_m is 6, every graph in $\mathcal{G}_n$ has maximum degree 7. Moreover, if we let $A_m = \{a_i, a'_i : 1 \leq i \leq m\}$, $B_m = \{b_i : 1 \leq i \leq m\}$ and $C_m = \{c_i : 1 \leq i \leq m\}$, then it is clear from our construction that (A_m, B_m, C_m) is a tripartition for every graph in $\mathcal{G}_n$. An immediate consequence of this is that every graph in $\mathcal{G}_n$ has an F-coloring that assigns, for example, color 2 to every vertex in A_m, color 3 to every vertex in B_m and a color from the independent set I_{32} to every vertex in C_m. Therefore, all graphs in the family $\mathcal{G}_n$ are F-colorable.

The next theorem shows that structure learning is hard for F and $\mathcal{G}_n$.

Theorem 10 *Let $m \in \mathbb{N}^+$ such that $m \geq 2$ and let $n = 4m$. Then, any structure learning algorithm for the constraint graph F and the family of graphs $\mathcal{G}_n$ that succeeds with probability at least $2^{-(m-3)}$ requires at least 2^{m+1} samples.*

Before proving this theorem we state two key facts that will be used in its proof. In particular, Fact 11 shows that actually (A_m, B_m, C_m) is the unique tripartition of G_m, and Fact 12 gives a lower bound for the number of samples required by a structure learning algorithm to guarantee a prescribed success probability.

Fact 11 *For any $m \in \mathbb{N}^+$ the graph G_m is tripartite and has a unique tripartition (A_m, B_m, C_m), where $A_m = \{a_i, a'_i : 1 \leq i \leq m\}$, $B_m = \{b_i : 1 \leq i \leq m\}$ and $C_m = \{c_i : 1 \leq i \leq m\}$.*

Fact 12 *Let H be an arbitrary constraint graph. Suppose $\hat{\mathcal{G}} = \{\hat{G}_1, \hat{G}_2, \ldots, \hat{G}_r\}$ is a family of r distinct H-colorable graphs such that H is identifiable with respect to $\hat{\mathcal{G}}$. Assume also that $\hat{G}_1$ is a subgraph of $\hat{G}_i$ for all $2 \leq i \leq r$ and that $\hat{G}_r$ is a supergraph of $\hat{G}_i$ for all $1 \leq i \leq r-1$. Let*

$$\eta = 1 - \frac{|\Omega_{\hat{G}_r}|}{|\Omega_{\hat{G}_1}|}.$$

If there exists a structure learning algorithm for H and $\hat{\mathcal{G}}$ such that for any $G^* \in \hat{\mathcal{G}}$, given L independent samples from $\pi^H_{G^*}$ as input, it outputs G^* with probability at least $1/r + \alpha$ with $\alpha > 0$, then $L \geq \alpha/\eta$.

We observe that when $\alpha = 0$, the structure learning algorithm that outputs a graph from $\hat{\mathcal{G}}$ uniformly at random has success probability $1/r$ without requiring any samples. We are now ready to prove Theorem 10.

Proof of Theorem 10 Let $m \geq 2$, $n = 4m$ and for ease of notation let $\mathcal{G} = \mathcal{G}_n$. Let $G^{(1)} = G_m$ and $G^{(t)} = G_m \cup M$ where M is defined in (1) and $t = 2^{m-2}$. Hence, the graph $G^{(1)}$ (resp., $G^{(t)}$) is a subgraph (resp., supergraph) of every other graph in $\mathcal{G}$. Moreover, all graphs in $\mathcal{G}$ are distinct and F is identifiable by Lemma 9. Hence, to apply Fact 12 all we need is a lower bound for

$$\eta = 1 - |\Omega_{G^{(t)}}|/|\Omega_{G^{(1)}}|.$$

By Fact 11, G_m has a unique tripartition (A_m, B_m, C_m). Since the constraint graph F has a tripartition $\{\{1,3\},\{2,4\},I_{32}\}$, every F-coloring of G_m induces the same tripartition of G_m. That is to say, in every F-coloring of G_m one of the sets A_m, B_m or C_m is colored with colors $\{1,3\}$, another is colored with $\{2,4\}$ and the third one is colored using colors from the independent set I_{32} of F. Then, the number of F-colorings of G_m such that A_m is colored with colors from I_{32} is $K \cdot 32^{2m}$, where K is the number of F-colorings of $B_m \cup C_m$ given a fixed F-coloring of A_m that only uses colors from I_{32}. Observe that K is the same for every F-coloring of A_m, and $K \geq 2$ since we can always color B_m with color 2 and C_m with 3, or color B_m with 3 and C_m with 2. On the other hand, the number of F-colorings where B_m receives colors from I_{32} is at most $2 \cdot 32^m \cdot 2^m \cdot 2^{2m} = 2 \cdot 32^m \cdot 8^m$, and similarly for C_m. Hence, the probability that in a uniformly random F-coloring of G_m, A_m is colored with colors from the independent set I_{32} is at least

$$\frac{K \cdot 32^{2m}}{K \cdot 32^{2m} + 4 \cdot 32^m \cdot 8^m} = 1 - \frac{4 \cdot 8^m}{K \cdot 32^m + 4 \cdot 8^m} \geq 1 - \frac{1}{2^{2m-1}}.$$

Let σ be an F-coloring of G_m such that A_m is colored with colors from I_{32}. Since colors from I_{32} are compatible with any other color, the pair of colors $\sigma(a_i), \sigma(b_{i+2})$ are compatible for any $1 \leq i \leq m-2$. Therefore, σ is also a valid F-coloring of $G^{(t)}$ and thus $\sigma \in \Omega_{G^{(t)}}$. We then deduce that

$$1 - \eta = \frac{|\Omega_{G^{(t)}}|}{|\Omega_{G^{(1)}}|} \geq 1 - \frac{1}{2^{2m-1}}.$$

Since $|\mathcal{G}| = 2^{m-2}$, it follows from Fact 12 that the number of samples required to learn a graph in $\mathcal{G}$ with success probability $2^{-(m-3)}$ is at least

$$L \geq \frac{2^{-(m-3)} - 2^{-(m-2)}}{\eta} \geq 2^{m+1}.$$ ■

The proofs of Lemma 9, Fact 11 and Fact 12 are provided in Appendix A.1.

2.2. Proof of Theorem 3

To conclude this section, we provide the proof of Theorem 3 from the introduction, which follows straightforwardly from Theorem 10.

Proof of Theorem 3 Let $n \geq 8$. If 4 divides n, then Theorem 10 implies there exists a constant $c > 0$ such that any structure learning algorithm for the constraint graph F and the graph family $\mathcal{G}_n$ with success probability at least $\exp(-cn))$ requires at least $\exp(cn)$ samples. Since $\mathcal{G}_n \subseteq \mathcal{G}(n, 7)$, the result follows.

The same ideas carry over without significant modification to the case when 4 does not divide n. For example, suppose that $n = 4m + 1$ for some $m \geq 2$. For $G \in \mathcal{G}_{4(m+1)}$, let $\hat{G}$ be the subgraph of G induced by $V_{m+1} \setminus \{a'_{m+1}, b_{m+1}, c_{m+1}\}$. We define $\mathcal{G}_n$ as the family of the subgraphs $\hat{G}$ for each $G \in \mathcal{G}_{4(m+1)}$. When $n = 4m + 2$ or $n = 4m + 3$, we consider instead the graph families of the subgraphs induced by $V_{m+1} \setminus \{b_{m+1}, c_{m+1}\}$ and $V_{m+1} \setminus \{c_{m+1}\}$, respectively. The argument in the proof of Theorem 10 carries over to these graph families straightforwardly. Since in every case $\mathcal{G}_n \subseteq \mathcal{G}(n, 7)$ the result follows. ■

3. Learning proper q-colorings

In this section we consider statistical identifiability, structure learning and equivalent-structure learning for proper q-colorings, where $H = K_q$ and π_G is the uniform distribution over the proper q-colorings of the graph G. In particular, we prove Theorem 4 from the introduction: Part 1 follows from Lemmas 14 and 13 in Section 3.2; Part 2 from Theorem 15 in Section 3.3; and we establish Part 3 of the theorem in Appendix B.1.

3.1. A structure learning algorithm

In this subsection we introduce a general structure learning algorithm for any constraint graph H with at least one hard constraint; i.e., $H \neq K_q^+$. In Section 3.2, we analyze its running time and sample complexity for proper colorings. Later in Sections 4 and E.3, we consider more general settings where this algorithm is also efficient.

Fix $H \neq K_q^+$ and suppose $\{i, j\} \notin E(H)$. Given independent samples $\sigma^{(1)}, \ldots, \sigma^{(L)}$ from $\pi_G = \pi_G^H$ for some unknown graph $G = (V, E)$, the algorithm checks for every pair of vertices $u, v \in V$ whether there is at least one sample $\sigma^{(k)}$ such that $\sigma_u^{(k)} = i$ and $\sigma_v^{(k)} = j$. If this is the case, then the edge $\{u, v\}$ does not belong to E. Otherwise, the algorithm adds the edge $\{u, v\}$ to the estimator $\hat{E}$ of E. This structure learning algorithm, which we call STRUCTLEARN-H, has running time $O(Ln^2)$ and was used before in [6] for the hard-core model. The effectiveness of STRUCTLEARN-H depends crucially on how likely are nonadjacent vertices to receive colors i and j. For $v \in V$, let X_v be the random variable for the color of v under π_G.

Lemma 13 *Let $H \neq K_q^+$ and $\{i, j\} \notin E(H)$. Suppose $\Pr[X_u = i, X_v = j] \geq \delta$ for all $\{u, v\} \notin E$ and some $\delta > 0$. Let $\hat{G} = (V, \hat{E})$ be the output of the algorithm* STRUCTLEARN-H. *Then, for all $\varepsilon \in (0, 1)$, $\Pr[E = \hat{E}] \geq 1 - \varepsilon$ provided $L \geq 8\delta^{-1} \log(\frac{n^2}{2\varepsilon})$.*

Proof Suppose $\{u, v\} \notin E$ and let Z_{uv} be the number of samples where vertices u and v are assigned colors i and j, respectively. Since $\mathrm{E}[Z_{uv}] \geq \delta L$, a Chernoff bound implies

$$\Pr[Z_{uv} = 0] \leq \Pr\left[Z_{uv} \leq \frac{\delta L}{2}\right] \leq \exp\left(\frac{-\delta L}{8}\right) \leq \frac{2\varepsilon}{n^2}.$$

The result follows from a union bound over the edges. ■

3.2. Efficient structure learning when $q \geq d+1$

In this subsection we prove Part 1 of Theorem 4. We show that for proper q-colorings with $q \geq d+1$ and any graph in $\mathcal{G}(n,d)$, the structure learning algorithm STRUCTLEARN-H (see Section 3.1) requires $O(qd^3 \log(n/\varepsilon))$ samples to succeed with probability at least $1-\varepsilon$ and has running time $O(qd^3n^2 \log(n/\varepsilon))$. This can be deduced immediately from the next lemma and Lemma 13, since $H = K_q \neq K_q^+$ in this setting.

Lemma 14 *Suppose that $q \geq d+1$ and let $\{u,v\} \notin E$. Then $\Pr[X_u = X_v] \geq \frac{1}{q(d+1)^3}$.*

Proof Let $u, v \in V$ be such that $\{u,v\} \notin E$, and let ∂u, ∂v denote the neighborhoods of u and v, respectively, which may overlap. Let G' be G with the vertices u, v removed (the edges adjacent to u and v are removed as well). We will generate a uniformly random coloring of G using rejection sampling as follows. Pick a uniformly random coloring of G', a uniformly random color c_1 for u, and a uniformly random color c_2 for v. If the resulting coloring is valid for G then accept, otherwise reject. Since in every round each coloring has the same probability of being picked, the generated coloring is a uniformly random coloring of G.

Let $A(c,s,t)$ be the set of colorings of G' where color c appears exactly s times in ∂u and exactly t times in ∂v. Given a coloring in $A(c,s,t)$ we produce a coloring in $A(c,0,0)$ as follows. List the positions where c occurs in $\partial u \cup \partial v$ and then re-color the vertices in the order of the list. Note that every vertex w in $\partial u \cup \partial v$ has at least 2 colors that do not occur in its neighborhood since in G' the vertex w has degree at most $d-1$ (recall that we removed u and v from G to obtain G'). This maps at most d^{s+t} colorings from $A(c,s,t)$ to a coloring in $A(c,0,0)$ (given the list of positions we can recover the original coloring; there are at most d^{s+t} lists where we first list the vertices in ∂u and then vertices in $\partial v \setminus \partial u$). Hence we have

$$|A(c,0,0)| \geq \frac{|A(c,s,t)|}{d^{s+t}}. \tag{2}$$

Let $A(c, \leq j, \leq k) := \sum_{s \leq j, t \leq k} A(c,s,t)$ and let μ be the uniform distribution over the colorings of G'. We claim that in any coloring there exists at least $q-d$ colors that satisfy the following: c occurs at most once in ∂u and at most once in ∂v. Indeed, adding over all colors the number of occurrences in ∂u and the number of occurrences in ∂v we can get at most $2d$; thus at most d colors can occur at least twice in ∂u or at least twice in ∂v. Thus

$$\mu(A(1,\leq 1,\leq 1)) + \mu(A(2,\leq 1,\leq 1)) + \cdots + \mu(A(q,\leq 1,\leq 1)) \geq q-d,$$

and by symmetry $\mu(A(1,\leq 1,\leq 1)) \geq \frac{q-d}{q} \geq \frac{1}{d+1}$. Since by definition $\mu(A(1,\leq 1,\leq 1)) = \sum_{i,j \in \{0,1\}} \mu(A(1,i,j))$, from (2) we get

$$\mu(A(1,0,0)) \geq \frac{1}{(d+1)^3}.$$

Therefore with probability at least $1/(d+1)^3$ color 1 does not occur in $\partial u \cup \partial v$ in a random coloring of G'. With probability $1/q^2$ we propose $c_1 = c_2 = 1$ in the rejection sampling procedure and hence with probability at least $\frac{1}{q^2(d+1)^3}$ our process accepts and produces a coloring of G where u, v both receive color 1. Thus in a uniformly random coloring of G vertices u, v receive color 1 with probability at least $\frac{1}{q^2(d+1)^3}$, and by symmetry the probability that they receive the same color is at least $\frac{1}{q(d+1)^3}$, as claimed. ■

3.3. Identifiability for proper q-colorings

In this subsection we prove Part 2 of Theorem 4. We show that when $q \leq d$ there exist two distinct graphs $G, G' \in \mathcal{G}(n,d)$ with the same set of q-colorings (i.e., $\pi_G = \pi_{G'}$).

Theorem 15 *Let $n, q, d \in \mathbb{N}^+$ such that $q \leq d$ and $n \geq q+2$. Then, the structure learning problem for q-colorings is not identifiable with respect to the family of graphs $\mathcal{G}(n,d)$.*

Proof Let $G = (V, E)$ be a graph with

$$V = \{c_1, \ldots, c_{q-1}, u, v, w_0, \ldots, w_{n-q-2}\},$$

where $\{c_1, \ldots, c_{q-1}, u, v\}$ is a clique of size $q+1$ except for the one edge $\{u, v\}$ that is not in E, and $\{w_0, \ldots, w_{n-q+2}\}$ is a simple path from w_0 to w_{n-q-2}. G has one additional edge connecting v and w_0. Then, in every q-coloring of G the vertices u and v receive the same color, and so u and w_0 are assigned distinct colors. Hence, the graph G and the graph $G' = (V, E \cup \{u, w_0\})$ have the same set of q-colorings. Since both G and G' are n-vertex graphs of maximum degree at most $q \leq d$, the structure learning for q-colorings is not identifiable with respect to $\mathcal{G}(n,d)$. ■

4. Learning H-colorings in Dobrushin uniqueness

As mentioned in the introduction, our results in Section 3 for statistical identifiability, structure learning and equivalent-structure learning for proper colorings reveal a tight connection between the computational hardness of these problems and the uniqueness/non-uniqueness phase transition.

In this section we explore this connection in a more general setting. For this we define the *Dobrushin uniqueness condition*, which is a standard tool in statistical physics for establishing uniqueness of the Gibbs distribution in infinite graphs.

Definition 16 *Let H be an arbitrary constraint graph and let $G = (V, E)$ be an H-colorable graph. For $w \in V$, let*

$$S_w := \{(\tau, \tau_w) : \tau, \tau_w \in \{1, \ldots, q\}^{|V|} \text{ and } \tau(z) = \tau_w(z)\ \forall z \neq w\}.$$

For $v, w \in V$, let

$$R_{vw} := \max_{(\tau,\tau_w)\in S_w} \|\pi_v(\cdot \mid \tau(\partial v)) - \pi_v(\cdot \mid \tau_w(\partial v))\|_{\mathrm{TV}},$$

where $\pi_v(\cdot|\tau(\partial v))$ and $\pi_v(\cdot|\tau_w(\partial v))$ are the conditional distributions at v given the respective assignments τ and τ_w on the neighbors of v. Let $\alpha := \max_{v\in V} \sum_{w\in\partial v} R_{vw}$. When $\alpha < 1$, π_G is said to satisfy the Dobrushin uniqueness condition.

We note that the Dobrushin uniqueness condition (typically) concerns soft-constraint systems on infinite graphs. The definition we use here for hard-constraint models in finite graphs appeared in [9]; see also [41]. The Dobrushin uniqueness condition implies the following key property.

Lemma 17 *Let $H \neq K_q^+$ be an arbitrary constraint graph and suppose $\{i, j\} \notin E(H)$. Let $G = (V, E)$ be a graph such that π_G satisfies the Dobrushin uniqueness condition. Then, for all $\{u, v\} \notin E$, $\Pr[X_u = i, X_v = j] \geq \frac{(1-\alpha)^2}{q^2}$.*

Lemmas 17 and 13 imply that the STRUCTLEARN-H algorithm requires $L = O(q^2 \log(\frac{n^2}{\varepsilon}))$ independent samples to succeed with probability at least $1 - \varepsilon$ and has running time is $O(Ln^2)$. This establishes Theorem 5 from the introduction. The proof of Lemma 17 is provided in Appendix C.

References

[1] Pieter Abbeel, Daphne Koller, and Andrew Y. Ng. Learning factor graphs in polynomial time and sample complexity. *Journal of Machine Learning Research*, 7(Aug):1743–1788, 2006.

[2] Anima Anandkumar, Daniel J. Hsu, Furong Huang, and Sham M. Kakade. Learning mixtures of tree graphical models. In *Advances in Neural Information Processing Systems (NIPS)*, pages 1052–1060, 2012.

[3] Guy Bresler. Efficiently learning Ising models on arbitrary graphs. In *Proceedings of the Forty-seventh Annual ACM Symposium on Theory of Computing (STOC)*, pages 771–782. ACM, 2015.

[4] Guy Bresler and Mina Karzand. Learning a tree-structured Ising model in order to make predictions. *ArXiv preprint*, 2016. https://arxiv.org/pdf/1604.06749.pdf.

[5] Guy Bresler, Elchanan Mossel, and Allan Sly. Reconstruction of Markov random fields from samples: some observations and algorithms. *SIAM Journal on Computing*, 42(2):563–578, 2013.

[6] Guy Bresler, David Gamarnik, and Devavrat Shah. Structure learning of antiferromagnetic Ising models. In *Advances in Neural Information Processing Systems (NIPS)*, pages 2852–2860, 2014.

[7] Guy Bresler, David Gamarnik, and Devavrat Shah. Hardness of parameter estimation in graphical models. In *Advances in Neural Information Processing Systems (NIPS)*, pages 1062–1070, 2014.

[8] Graham R. Brightwell and Peter Winkler. Random colorings of a Cayley tree. *Contemporary Combinatorics*, 10:247–276, 2002.

[9] Russ Bubley and Martin Dyer. Path coupling, Dobrushin uniqueness, and approximate counting. *Research Report Series–University of Leeds, School of Computer Studies*, 1997.

[10] Andrei A. Bulatov. H-coloring dichotomy revisited. *Theoretical Computer Science*, 349(1): 31–39, 2005.

[11] Andrei A. Bulatov. A dichotomy theorem for nonuniform CSPs. In *Proceedings of the 58th Annual Symposium on Foundations of Computer Science (FOCS)*, pages 319–330. IEEE, 2017.

[12] C. K. Chow and Cong Liu. Approximating discrete probability distributions with dependence trees. *IEEE Transactions on Information Theory*, 14(3):462–467, 1968.

[13] Imre Csiszár and Zsolt Talata. Consistent estimation of the basic neighborhood of Markov random fields. *The Annals of Statistics*, 34(1):123–145, 2006.

[14] Sanjoy Dasgupta. Learning polytrees. In *Proceedings of the Fifteenth Conference on Uncertainty in Artificial Intelligence (UAI)*, pages 134–141. Morgan Kaufmann Publishers Inc., 1999.

[15] Amir Dembo, Andrea Montanari, and Nike Sun. Factor models on locally tree-like graphs. *The Annals of Probability*, 41(6):4162–4213, 2013.

[16] Roland L. Dobrushin. The description of a random field by means of conditional probabilities and conditions of its regularity. *Theory of Probability & its Applications*, 13(2):197–224, 1968.

[17] Martin Dyer and Catherine Greenhill. The complexity of counting graph homomorphisms. *Random Structures and Algorithms*, 17(3-4):260–289, 2000.

[18] Martin Dyer, Leslie Ann Goldberg, and Mark Jerrum. Counting and sampling H-colourings. *Information and Computation*, 189(1):1–16, 2004.

[19] Martin Dyer, Alistair Sinclair, Eric Vigoda, and Dror Weitz. Mixing in time and space for lattice spin systems: A combinatorial view. *Random Structures & Algorithms*, 24(4):461–479, 2004.

[20] Nathan Eagle, Alex Sandy Pentland, and David Lazer. Inferring friendship network structure by using mobile phone data. *Proceedings of the National Academy of Sciences*, 106(36): 15274–15278, 2009.

[21] Thomas Emden-Weinert, Stefan Hougardy, and Bernd Kreuter. Uniquely colourable graphs and the hardness of colouring graphs of large girth. *Combinatorics, Probability and Computing*, 7(4):375–386, 1998.

[22] Tomás Feder and Moshe Y. Vardi. Monotone monadic SNP and constraint satisfaction. In *Proceedings of the Twenty-fifth Annual ACM Symposium on Theory of Computing (STOC)*, pages 612–622. ACM, 1993.

[23] Andreas Galanis, Leslie Ann Goldberg, and Mark Jerrum. Approximately counting H-colorings is #BIS-Hard. *SIAM Journal on Computing*, 45(3):680–711, 2016.

[24] Andreas Galanis, Daniel Štefankovič, and Eric Vigoda. Inapproximability of the partition function for the antiferromagnetic Ising and hard-core models. *Combinatorics, Probability and Computing*, 25(4):500–559, 2016.

[25] Leslie Ann Goldberg, Steven Kelk, and Mike Paterson. The complexity of choosing an H-colouring (nearly) uniformly at random. In *Proceedings of the Thirty-fourth Annual ACM Symposium on Theory of Computing (STOC)*, pages 53–62. ACM, 2002.

[26] Linus Hamilton, Frederic Koehler, and Ankur Moitra. Information theoretic properties of Markov random fields, and their algorithmic applications. In *Advances in Neural Information Processing Systems (NIPS)*, pages 2460–2469, 2017.

[27] Pavol Hell and Jaroslav Nešetřil. On the complexity of H-coloring. *Journal of Combinatorial Theory, Series B*, 48(1):92–110, 1990.

[28] Pavol Hell and Jaroslav Nešetřil. Graphs and homomorphisms. *Oxford Lecture Series in Mathematics and its Applications*, 2004.

[29] John P. Huelsenbeck, Fredrik Ronquist, Rasmus Nielsen, and Jonathan P. Bollback. Bayesian inference of phylogeny and its impact on evolutionary biology. *Science*, 294(5550):2310–2314, 2001.

[30] Ali Jalali, Pradeep Ravikumar, Vishvas Vasuki, and Sujay Sanghavi. On learning discrete graphical models using group-sparse regularization. In *Proceedings of the Fourteenth International Conference on Artificial Intelligence and Statistics*, pages 378–387, 2011.

[31] Johan Jonasson. Uniqueness of uniform random colorings of regular trees. *Statistics & Probability Letters*, 57(3):243–248, 2002.

[32] Adam Klivans and Raghu Meka. Learning graphical models using multiplicative weights. In *Proceedings of the 58th Annual Symposium on Foundations of Computer Science (FOCS)*, pages 343–354. IEEE, 2017.

[33] Su-In Lee, Varun Ganapathi, and Daphne Koller. Efficient structure learning of Markov networks using l_1-regularization. In *Advances in Neural Information Processing Systems (NIPS)*, pages 817–824, 2007.

[34] Liang Li, Pinyan Lu, and Yitong Yin. Correlation decay up to uniqueness in spin systems. In *Proceedings of the Twenty-fourth Annual ACM-SIAM Symposium on Discrete Algorithms (SODA)*, pages 67–84. SIAM, 2013.

[35] Daniel Marbach, James C. Costello, Robert Küffner, Nicci Vega, Robert J. Prill, Diogo M. Camacho, Kyle R. Allison, Manolis Kellis, James J. Collins, and Gustavo Stolovitzky. Wisdom of crowds for robust gene network inference. *Nature Methods*, 9(8):796–804, 2012.

[36] Fabio Martinelli, Alistair Sinclair, and Dror Weitz. Fast mixing for independent sets, colorings, and other models on trees. *Random Structures & Algorithms*, 31(2):134–172, 2007.

[37] Michael Molloy and Bruce Reed. Colouring graphs when the number of colours is nearly the maximum degree. In *Proceedings of the Thirty-third Annual ACM Symposium on Theory of Computing (STOC)*, pages 462–470. ACM, 2001.

[38] Faruck Morcos, Andrea Pagnani, Bryan Lunt, Arianna Bertolino, Debora S. Marks, Chris Sander, Riccardo Zecchina, José N. Onuchic, Terence Hwa, and Martin Weigt. Direct-coupling analysis of residue coevolution captures native contacts across many protein families. *Proceedings of the National Academy of Sciences*, 108(49):E1293–E1301, 2011.

[39] Pradeep Ravikumar, Martin J. Wainwright, and John D. Lafferty. High-dimensional Ising model selection using ℓ_1-regularized logistic regression. *The Annals of Statistics*, 38(3):1287–1319, 2010.

[40] Stefan Roth and Michael J. Black. Fields of experts: A framework for learning image priors. In *Proceedings of the 2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR)*, volume 2, pages 860–867. IEEE, 2005.

[41] Jesús Salas and Alan D. Sokal. Absence of phase transition for antiferromagnetic Potts models via the Dobrushin uniqueness theorem. *Journal of Statistical Physics*, 86(3-4):551–579, 1997.

[42] Elad Schneidman, Michael J. Berry II, Ronen Segev, and William Bialek. Weak pairwise correlations imply strongly correlated network states in a neural population. *Nature*, 440 (7087):1007, 2006.

[43] Mark H. Siggers. A new proof of the H-coloring dichotomy. *SIAM Journal on Discrete Mathematics*, 23(4):2204–2210, 2010.

[44] Allan Sly. Computational transition at the uniqueness threshold. In *Proceedings of the 51st Annual Symposium on Foundations of Computer Science (FOCS)*, pages 287–296. IEEE, 2010.

[45] Allan Sly and Nike Sun. The computational hardness of counting in two-spin models on d-regular graphs. In *Proceedings of the 53rd Annual Symposium on Foundations of Computer Science (FOCS)*, pages 361–369. IEEE, 2012.

[46] Nathan Srebro. Maximum likelihood bounded tree-width Markov networks. In *Proceedings of the Seventeenth Conference on Uncertainty in Artificial Intelligence (UAI)*, pages 504–511. Morgan Kaufmann Publishers Inc., 2001.

[47] Marc Vuffray, Sidhant Misra, Andrey Lokhov, and Michael Chertkov. Interaction screening: Efficient and sample-optimal learning of Ising models. In *Advances in Neural Information Processing Systems (NIPS)*, pages 2595–2603, 2016.

[48] Dror Weitz. Counting independent sets up to the tree threshold. In *Proceedings of the Thirty-eighth Annual ACM Symposium on Theory of Computing (STOC)*, pages 140–149. ACM, 2006.

[49] Dmitriy Zhuk. The proof of CSP dichotomy conjecture. In *Proceedings of the 58th Annual Symposium on Foundations of Computer Science (FOCS)*, pages 331–342. IEEE, 2017.

Appendix A. Proofs missing from Section 2

In this appendix we provide the proofs omitted in Section 2. The following lemma provides a useful alternative definition of identifiability.

Lemma 18 *A constraint graph $H \neq K_q^+$ is identifiable if and only if for any H-colorable graph $G = (V, E)$ and any two nonadjacent vertices $u, v \in V$ there is an H-coloring of G where u and v are assigned incompatible colors.*

Proof For the forward direction we prove the contrapositive. Let $G = (V, E)$ be an H-colorable graph and suppose there exists two nonadjacent vertices $u, v \in V$ such that in every H-coloring σ of G these vertices receive compatible colors. Then, the graph G and the graph $G' = (V, E \cup \{u, v\})$ have the same set of H-colorings. Hence, $\pi_G = \pi_{G'}$ and so H is not identifiable.

For the reverse direction, suppose that for every H-colorable graph $G = (V, E)$ and every pair of nonadjacent vertices $u, v \in V$ there exists an H-coloring of G such that u and v are assigned incompatible colors. Suppose also that for a pair of H-colorable graphs $G_1 = (V, E_1)$ and $G_2 = (V, E_2)$, we have $\pi_{G_1} = \pi_{G_2}$ (or equivalently that $\Omega_{G_1} = \Omega_{G_2}$). We show that $G_1 = G_2$. First consider $\{u, v\} \notin E_1$. Then, there exists an H-coloring $\tau \in \Omega_{G_1}$ where u and v receive incompatible colors. Since also $\tau \in \Omega_{G_2}$, $\{u, v\} \notin E_2$. Similarly, if $\{u, v\} \notin E_2$, then $\{u, v\} \notin E_1$. Thus, $G_1 = G_2$ and so H is identifiable. ■

Proof of Lemma 7 The proof is divided into two cases corresponding to whether all vertices of H have self-loops or not.

Case 1: *At least one but not all vertices of H have self-loops.* Let U be the set of vertices that have self-loops and let $W = V(H) \backslash U$ be the set of vertices that do not. By assumption both U and W are not empty. Moreover, U and W are connected because by assumption H is connected. Thus, there exist $i \in U$ and $j \in W$ such that $\{i, i\}, \{i, j\} \in E(H)$ and $\{j, j\} \notin E(H)$. We use this gadget to show that for any H-colorable graph $G = (V, E)$ and any two nonadjacent vertices $u, v \in V$ of G, there exists an H-coloring σ of G where u and v are assigned incompatible colors. Then, by Lemma 18, H is identifiable. The H-coloring σ is defined as follows: $\sigma(w) = i$ for all $w \neq u, v$ and $\sigma(u) = \sigma(v) = j$. Since $\{i, i\}, \{i, j\} \in E(H)$, σ is a valid H-coloring of G. Moreover, since $\{j, j\} \notin E(H)$, u and v receive incompatible colors and the result follows.

Case 2: *All vertices of H have self-loops*. Observe first that if H is connected, $H \neq K_q^+$ and every vertex in H has a self-loop, then there exist $i, j, k \in V(H)$ such that

$$\{i, j\}, \{j, k\}, \{i, i\}, \{j, j\}, \{k, k\} \in E(H) \quad \text{and} \quad \{i, k\} \notin E(H).$$

We use this gadget to show for any H-colorable graph $G = (V, E)$ and any pair of nonadjacent vertices $u, v \in V$ there is an H-coloring σ of G such that $(\sigma(u), \sigma(v)) \notin E(H)$. Lemma 18 then implies that H is identifiable. The H-coloring σ is given by: $\sigma(w) = j$ for all $w \neq u, v$, $\sigma(u) = i$ and $\sigma(v) = k$. Since color j is compatible with colors i, j and k in H, σ is a valid H-coloring of G. Moreover, u and v receive the incompatible colors i and k and so the result follows. ■

Proof of Lemma 8 Assume first that $H \neq K_q^+$ is identifiable and has no self-loops. For every $\{i,j\} \in E(H)$, G_{ij} is clearly H-colorable (simply assign color $k \in V(H)$ to the corresponding vertex in G_{ij}, color i to i' and color j to j'). Hence, Lemma 18 implies that there exists an H-coloring of G_{ij} where i' and j' receive incompatible colors. This proves the forward direction of the lemma.

For the reverse direction, suppose that for every $\{i,j\} \in E(H)$ there exists an H-coloring of G_{ij} where i' and j' are assigned incompatible colors. Let $G = (V,E)$ be an arbitrary H-colorable graph. We show that for every pair of nonadjacent vertices $u, v \in V$ in G there exists an H-coloring of G where u and v receive incompatible colors. It then follows from Lemma 18 that H is identifiable.

Let σ be an H-coloring of G and let us assume that $\sigma(u)$ and $\sigma(v)$ are compatible colors. (If $\sigma(u)$ and $\sigma(v)$ are incompatible colors in H, there is nothing to prove.) We use σ to construct an H-coloring σ' where u and v receive incompatible colors. Let $a = \sigma(u)$ and $b = \sigma(v)$. By assumption, there exists an H-coloring τ of G_{ab} where the corresponding copies of a and b, a' and b', receive incompatible colors. Define the H-coloring σ' of G as follows:

$$\sigma'(w) = \tau(\sigma(w)), \quad \forall w \neq u, v; \quad \sigma'(u) = \tau(a'); \quad \sigma'(v) = \tau(b').$$

It is straightforward to check that σ' is a proper H-coloring of G. Since u and v receive incompatible colors in σ' (i.e., $\tau(a')$ and $\tau(b')$), the proof is complete. ■

A.1. Proofs missing from Section 2.1

Proof of Lemma 9 Let $G = (V,E)$ be an F-colorable graph. Since F is tripartite with a unique tripartition $\{\{1,3\},\{2,4\},I_{32}\}$, then so is G. Let $\{V_1,V_2,V_3\}$ be a tripartition of G and let u, v be any two nonadjacent vertices of G. We show that there is always an F-coloring of G where u and v receive incompatible colors. The result then follows from Lemma 18.

If u and v belong to the same V_i, then by coloring all the vertices V_1 with color 1, all the vertices of V_2 with color 2 and all the vertices in V_3 with any color c from I_{32}, we have a coloring of G where u and v receive the same color. Since F has no self-loops u and v are assigned incompatible colors.

If u and v belong to different V_i's, suppose without loss of generality that $u \in V_1$ and $v \in V_2$. Consider the following F-coloring σ of G where c is any color from I_{32}:

$$\sigma(w) = \begin{cases} 1 & \text{if } w = u; \\ 4 & \text{if } w = v; \\ 3 & \text{if } w \in V_1 \backslash \{u\}; \\ 2 & \text{if } w \in V_2 \backslash \{v\}; \\ c & \text{if } w \in V_3. \end{cases}$$

In σ, u and v receive the incompatible colors 1 and 4. Thus, we have shown that it is always possible to color nonadjacent vertices of G with incompatible colors and the result follows immediately from Lemma 18. ■

Proof of Fact 11 We prove this fact by induction. G_1 has exactly one tripartition $(\{a_1,a_1'\},b_1,c_1)$. Suppose inductively that $(A_{m-1},B_{m-1},C_{m-1})$ is the only tripartition of G_{m-1}. Since by definition $\{a_i,b_{i-1}\},\{a_i,c_{i-1}\} \in E_m$, a_m belongs to A_m in any tripartition of G_m. Similar statements hold for a_m', b_m and c_m as well. Therefore, (A_m,B_m,C_m) is the unique tripartition of G_m. ■

Proof of Fact 12 Let $\mathcal{A}$ be any (possibly randomized) structure learning algorithm that given L independent samples $\Gamma = (\sigma^{(1)}, \ldots, \sigma^{(L)}) \in \Omega_{G^*}^L$ from an unknown distribution π_{G^*} for some $G^* \in \hat{\mathcal{G}}$, outputs a graph $\mathcal{A}(\Gamma)$ in $\hat{\mathcal{G}}$. For any $G^* \in \hat{\mathcal{G}}$, the probability that $\mathcal{A}$ learns the graph correctly given L independent samples from π_{G^*} is

$$\Pr[\mathcal{A}(\Gamma) = G^*] = \sum_{x \in \Omega_{G^*}^L} \Pr_{\pi_{G^*}}[\Gamma = x] \Pr[\mathcal{A}(x) = G^*].$$

Since $\hat{G}_r$ is a supergraph of G^*, we have $\Omega_{\hat{G}_r} \subseteq \Omega_{G^*}$. Let $\mathcal{T}$ be the set of all sample sequences $\sigma^{(1)}, \ldots, \sigma^{(L)}$ such that $\sigma^{(i)} \notin \Omega_{\hat{G}_r}$ for at least one i; namely, $\mathcal{T} = \Omega_{G^*}^L \backslash \Omega_{\hat{G}_r}^L$. Note that $|\mathcal{T}| = |\Omega_{G^*}|^L - |\Omega_{\hat{G}_r}|^L$. Then,

$$\begin{aligned}\Pr[A(\Gamma) = G^*] &= \sum_{x \in \Omega_{\hat{G}_r}^L} \frac{1}{|\Omega_{G^*}|^L} \cdot \Pr[\mathcal{A}(x) = G^*] + \sum_{x \in \mathcal{T}} \frac{1}{|\Omega_{G^*}|^L} \cdot \Pr[\mathcal{A}(x) = G^*] \\ &\leq \frac{1}{|\Omega_{\hat{G}_r}|^L} \sum_{x \in \Omega_{\hat{G}_r}^L} \Pr[\mathcal{A}(x) = G^*] + \frac{|\Omega_{G^*}|^L - |\Omega_{\hat{G}_r}|^L}{|\Omega_{G^*}|^L}.\end{aligned}$$

Since $\hat{G}_1$ is a subgraph of G^*, we have $\Omega_{G^*} \subseteq \Omega_{\hat{G}_1}$. Thus,

$$\frac{|\Omega_{G^*}|^L - |\Omega_{\hat{G}_r}|^L}{|\Omega_{G^*}|^L} = 1 - \frac{|\Omega_{\hat{G}_r}|^L}{|\Omega_{G^*}|^L} \leq 1 - \frac{|\Omega_{\hat{G}_r}|^L}{|\Omega_{\hat{G}_1}|^L} = 1 - (1-\eta)^L \leq L\eta.$$

Suppose the structure learning algorithm $\mathcal{A}$ has success probability at least $1/r + \alpha$ for any $G^* \in \hat{\mathcal{G}}$; that is,

$$\Pr[\mathcal{A}(\Gamma) = G^*] \geq \frac{1}{r} + \alpha, \quad \forall G^* \in \hat{\mathcal{G}}.$$

Then,

$$\frac{1}{r} + \alpha \leq \frac{1}{|\Omega_{\hat{G}_r}|^L} \sum_{x \in \Omega_{\hat{G}_r}^L} \Pr[\mathcal{A}(x) = G^*] + L\eta, \quad \forall G^* \in \hat{\mathcal{G}}.$$

Since $\sum_{G^* \in \hat{\mathcal{G}}} \Pr[\mathcal{A}(x) = G^*] = 1$ for any fixed sample sequence x, summing up over $\hat{\mathcal{G}}$ we get

$$1 + r\alpha \leq \frac{1}{|\Omega_{\hat{G}_r}|^L} \sum_{x \in \Omega_{\hat{G}_r}^L} \sum_{G^* \in \hat{\mathcal{G}}} \Pr[\mathcal{A}(x) = G^*] + rL\eta = 1 + rL\eta.$$

Hence, $L \geq \alpha/\eta$ as claimed. ■

Appendix B. Lower bounds for proper colorings

B.1. Strong lower bound when $q < d - \sqrt{d} + \Theta(1)$

In this subsection we prove Part 3 of Theorem 4, establishing a strong learning lower bound for proper colorings when $q < d - \sqrt{d} + \Theta(1)$.

As previously defined, an equivalent-structure learning algorithm for a graph family $\mathcal{G}$ finds a graph $\hat{G} \in \mathcal{G}$ such that $\Omega_G = \Omega_{\hat{G}}$, where $G \in \mathcal{G}$ is the actual hidden graph. We exhibit a family of graphs of maximum degree $q + \sqrt{q} + \Theta(1)$ such that every graph in the family has almost the same set of q-colorings. This makes equivalent-structure learning hard in this family. We use this fact to prove Part 3 of Theorem 4.

We define first a graph $G_{m,t} = (V_{m,t}, E_{m,t})$ and every graph in our graph family will be a supergraph of $G_{m,t}$. For any $m, t \in \mathbb{N}^+$ with $t < q$, the graph $G_{m,t}$ is defined as follows. Let $C_1, \ldots, C_m$ and $C'_1, \ldots, C'_m$ be cliques of size $q-1$, and let $I_1, \ldots, I_m$ and $I'_1, \ldots, I'_m$ be independent sets of size t. Moreover, let $s_1, \ldots, s_m$ be m additional vertices. Then,

$$V_{m,t} = \bigcup_{i=1}^{m} \{V(C_i), V(C'_i), V(I_i), V(I'_i), s_i\},$$

where $V(C_i), V(C'_i), V(I_i), V(I'_i)$ are the vertices of C_i, C'_i, I_i, I'_i, respectively. In addition to the edges in the cliques C_i and C'_i for $1 \leq i \leq m$, $E_{m,t}$ contains the following edges:

1. For $1 \leq i \leq m$, there is a complete bipartite graph between C_i and I_i. That is, for $u \in C_i$ and $v \in I_i$, $\{u, v\} \in E_{m,t}$.
2. For $2 \leq i \leq m$, C_i is partitioned into t almost-equally-sized disjoint subsets $C_{i,1}, \ldots, C_{i,t}$ of size either $\lfloor (q-1)/t \rfloor$ or $\lceil (q-1)/t \rceil$. Then, the j-th vertex of I_{i-1} is connected to every vertex in $C_{i,j}$.
3. Edges between $C'_1, \ldots, C'_m$ and $I'_1, \ldots, I'_m$ are defined in exactly the same manner.
4. For $1 \leq i \leq m$, the vertex s_i is adjacent to exactly one vertex in I_i and to exactly one in I'_i;

see Figure 3 for an illustration of the graph $G_{m,t}$. The key fact about the graph $G_{m,t}$ that allows us to construct a graph family with the desired properties is the following.

Lemma 19 *Let $q, m, t \in \mathbb{N}^+$ and $q \geq 3$. In every q-coloring of $G_{m,t}$ all cliques $C_1, \ldots, C_m$ are colored by the same set of $q-1$ colors, and all independent sets $I_1, \ldots, I_m$ are colored with the remaining color. The same holds for $C'_1, \ldots, C'_m$ and $I'_1, \ldots, I'_m$.*

Lemma 19 implies that every q-coloring of $G_{m,t}$ is determined by the colors of I_1, I'_1 and those of the vertices $s_1, \ldots, s_m$.

We define next the family of graphs $\mathcal{G}_{m,t}$. All the graphs in this family are distinct and are supergraphs of $G_{m,t}$. For each $1 \leq i \leq m$ choose (and fix) a pair of vertices $x_i \in C_i$ and $y_i \in I'_i \setminus \partial s_i$. (Recall that ∂s_i denotes the neighborhood of s_i.) Let

$$M = \{\{x_i, y_i\} : 1 \leq i \leq m\}. \tag{3}$$

Let $E^{(1)}, \ldots, E^{(l)}$ be all the subsets of M; hence, $l = 2^m$. We let

$$\mathcal{G}_{m,t} = \{G^{(j)} = (V_{m,t}, E_{m,t} \cup E^{(j)}) : 1 \leq j \leq l\}.$$

The graphs in $\mathcal{G}_{m,t}$ satisfy the following.

Fact 20 *If $G = (V, E) \in \mathcal{G}_{m,t}$, then $|V| = (2q + 2t - 1)m$ and the maximum degree of G is at most*

$$\max\left\{q + t, q + \left\lceil (q-1)/t \right\rceil + 1\right\}.$$

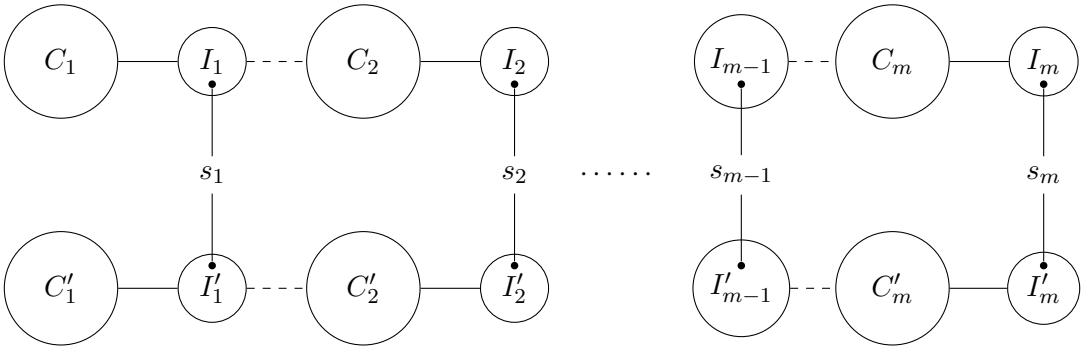

Figure 3: The graph $G_{m,t}$. Each of $C_1, \ldots, C_m, C'_1, \ldots, C'_m$ is a clique of size $q-1$ and each of $I_1, \ldots, I_m, I'_1, \ldots, I'_m$ is an independent set of size $t < q$. Solid lines between two clusters mean that every vertex from one cluster is adjacent to every vertex in the other cluster. Dashed lines between I_{i-1} and C_i mean that every vertex in I_{i-1} is adjacent to roughly $(q-1)/t$ vertices in C_i with no two vertices in I_{i-1} sharing a common neighbor in C_i.

Using Lemma 19 and ideas similar to those in the proof of Theorem 10, we can show that both structure and equivalent-structure learning are computationally hard in $\mathcal{G}_{m,t}$ (sample complexity is exponential in m). This immediately implies that *structure learning* is also hard in $\mathcal{G}(n,d)$ provided d is large enough so that $\mathcal{G}_{m,t} \subseteq \mathcal{G}(n,d)$. However, this does not necessarily imply that *equivalent-structure learning* is hard for $\mathcal{G}(n,d)$ which is our goal. For this, we introduce instead a larger graph family $\mathcal{F}_{m,t}$ that contains $\mathcal{G}_{m,t}$. Suppose d is an integer such that

$$d \geq \max\left\{q+t, q+\left\lceil (q-1)/t \right\rceil + 1\right\},$$

the maximum degree of any graph in $\mathcal{G}_{m,t}$; see Fact 20. The graph family $\mathcal{F}_{m,t}$ contains all the graphs in $\mathcal{G}(n,d)$ that have the same set of colorings as some graph in $\mathcal{G}_{m,t}$. Namely,

$$\mathcal{F}_{m,t} = \{G \in \mathcal{G}(n,d) : \Omega_G = \Omega_{G'}, G' \in \mathcal{G}_{m,t}\},$$

and $\mathcal{G}_{m,t} \subseteq \mathcal{F}_{m,t} \subseteq \mathcal{G}(n,d)$ by definition.

The next theorem establishes a lower bound for any equivalent-structure learning algorithm for $\mathcal{F}_{m,t}$. We will later see that this lower bound applies also to the graph family $\mathcal{G}(n,d)$, which establishes Part 3 of Theorem 4.

Theorem 21 *Let $q, m \in \mathbb{N}^+$, $q \geq 3$ and $t = \lceil \sqrt{q-1} \rceil$. Any equivalent-structure learning algorithm for proper vertex q-colorings and the graph family $\mathcal{F}_{m,t}$ that succeeds with probability at least $q \cdot \exp\left[-m/(2(q-1))\right]$ requires at least $\exp\left[m/(2(q-1))\right]$ samples.*

The following generalization of Fact 12 will be used in the proof of Theorem 21.

Fact 22 *Let H be an arbitrary constraint graph and let $\mathcal{F}_1, \ldots, \mathcal{F}_r$ be r families of distinct H-colorable graphs. Suppose that for all $1 \leq i \leq r$ every graph in the in $\mathcal{F}_i$ has the same set of H-colorings $\Omega_{\mathcal{F}_i}$. Assume also that $\Omega_{\mathcal{F}_i} \neq \Omega_{\mathcal{F}_j}$ for $i \neq j$ and that $\Omega_{\mathcal{F}_r} \subseteq \Omega_{\mathcal{F}_i} \subseteq \Omega_{\mathcal{F}_1}$ for all $1 \leq i \leq r$. Let $\hat{\mathcal{F}} = \mathcal{F}_1 \cup \cdots \cup \mathcal{F}_r$ and let*

$$\eta = 1 - \frac{|\Omega_{\mathcal{F}_r}|}{|\Omega_{\mathcal{F}_1}|}.$$

If there exists an equivalent-structure learning algorithm for H and $\hat{\mathcal{F}}$, such that for any $G^* \in \hat{\mathcal{F}}$, given L independent samples from $\pi^H_{G^*}$ as input, it outputs a graph G satisfying $\Omega_G = \Omega_{G^*}$ with probability at least $1/r + \alpha$ with $\alpha > 0$, then $L \geq \alpha/\eta$.

Note that Fact 12 corresponds to the special case where each family $\mathcal{F}_i$ contains a single graph, and thus it follows immediately from Fact 22. We are now ready to prove Theorem 21.

Proof of Theorem 21 Let $G^{(1)} = G_{m,t}$ and $G^{(l)} = G_{m,t} \cup M$ where M is as in (3) and $l = 2^m$. Let $\mathcal{F}_j$ be the class of graphs that contains all the graphs in $\mathcal{F}_{m,t}$ that has the same set of colorings as $G^{(j)}$. (Recall that $\mathcal{F}_{m,t}$ is the set of graphs in $\mathcal{G}(n,d)$ that have the same set of colorings as some graph in $\mathcal{G}_{m,t}$.) Let $\Omega_{\mathcal{F}_j} = \Omega_{G^{(j)}}$ for all j. Note that $\Omega_{\mathcal{F}_l} \subseteq \Omega_{\mathcal{F}_j} \subseteq \Omega_{\mathcal{F}_1}$ for all $1 \leq j \leq l$. Let

$$\eta = 1 - \frac{|\Omega_{\mathcal{F}_l}|}{|\Omega_{\mathcal{F}_1}|}.$$

We establish a lower bound η and then apply Fact 22 to prove the theorem.

By Lemma 19 each q-coloring of $G_{m,t}$ (i.e., of $G^{(1)}$) is determined by the colors of the independent sets I_1, I_1' and of the vertices $s_1, \ldots, s_m$. Then, the number of q-colorings of $G^{(1)}$ where all vertices in I_1 and I_1' receive the same color is equal to $q(q-1)^m[(q-1)!]^{2m}$, since there are q choices for the color of I_1 and I_1', $q-1$ choices for the color of each s_i, and $(q-1)!$ colorings for each C_i and C_i'. Similarly, the number of colorings where I_1 and I_1' receive distinct colors is equal to $q(q-1)(q-2)^m[(q-1)!]^{2m}$. Thus, the probability that in a uniform random q-coloring of $G^{(1)}$ the vertices in I_1 and I_1' have the same color is

$$\begin{aligned}
\frac{q(q-1)^m[(q-1)!]^{2m}}{q(q-1)^m[(q-1)!]^{2m} + q(q-1)(q-2)^m[(q-1)!]^{2m}} &= 1 - \frac{(q-2)^m}{(q-1)^{m-1} + (q-2)^m} \\
&\geq 1 - (q-1)\Big(\frac{q-2}{q-1}\Big)^m \\
&\geq 1 - (q-1)\,\mathrm{e}^{-\frac{m}{q-1}}.
\end{aligned}$$

Let σ be a q-coloring of $G^{(1)}$ where I_1 and I_1' receive the same color. Then, Lemma 19 implies that $\sigma(x_i) \neq \sigma(y_i)$ for all $1 \leq i \leq m$, since C_i and I_i' have distinct colors. (Recall that $x_i \in C_i$ and $y_i \in I_i' \backslash \partial s_i$ are the vertices used to define the graph family $\mathcal{G}_{m,t}$). Therefore, σ is a proper q-coloring of $G^{(l)}$; hence $\sigma \in \Omega_{\mathcal{F}_l}$ and

$$\frac{|\Omega_{\mathcal{F}_l}|}{|\Omega_{\mathcal{F}_1}|} = \Pr_{\pi_{\mathcal{F}_1}}[\sigma \in \Omega_{\mathcal{F}_l}] \geq \Pr_{\pi_{\mathcal{F}_1}}[\sigma(I_1) = \sigma(I_1')] \geq 1 - (q-1)\,\mathrm{e}^{-\frac{m}{q-1}}.$$

(Note that $\pi_{\mathcal{F}_1} = \pi_{G^{(1)}} = \pi_{G_{m,t}}$.) Then,

$$\eta = 1 - \frac{|\Omega_{\mathcal{F}_l}|}{|\Omega_{\mathcal{F}_1}|} \leq (q-1)\mathrm{e}^{-\frac{m}{q-1}}. \tag{4}$$

Every graph in $\mathcal{G}_{m,t}$ is a distinct supergraph of $G_{m,t}$ and $|\mathcal{G}_{m,t}| = 2^m$. Moreover, for any $G = (V,E) \in \mathcal{G}_{m,t}$ and any $1 \leq i \leq m$ such that $\{x_i, y_i\} \notin E$, there are q-colorings of G where x_i and y_i are assigned the same color. Consequently, for any $G, G' \in \mathcal{G}_{m,t}$, we have $\Omega_G \neq \Omega_{G'}$ whenever $G \neq G'$. Then, $\Omega_{\mathcal{F}_i} \neq \Omega_{\mathcal{F}_j}$ for any $i \neq j$, and by definition all the graphs in $\mathcal{F}_i$ are distinct

for each i. Therefore, Fact 22 implies that to equivalently learn any $G \in \mathcal{F}_{m,t}$ with probability at least $2^{-m} + \alpha$, the number of random samples needed is $L \geq \alpha/\eta$. Setting

$$\alpha = q\,\mathrm{e}^{-\frac{m}{2(q-1)}} - 2^{-m} > 0,$$

we get that to equivalently learn a graph $G \in \mathcal{F}_{m,t}$ with success probability at least $q\,\mathrm{e}^{-\frac{m}{2(q-1)}}$, we require

$$L \geq \frac{q\,\mathrm{e}^{-\frac{m}{2(q-1)}} - 2^{-m}}{(q-1)\,\mathrm{e}^{-\frac{m}{q-1}}} \geq \frac{q\,\mathrm{e}^{-\frac{m}{2(q-1)}} - \mathrm{e}^{-\frac{m}{2(q-1)}}}{(q-1)\,\mathrm{e}^{-\frac{m}{q-1}}} \geq \mathrm{e}^{\frac{m}{2(q-1)}}$$

where the second inequality follows from $2 \geq \mathrm{e}^{\frac{1}{2(q-1)}}$ when $q \geq 3$. ■

The following corollary of Theorem 21 corresponds to Part 3 of Theorem 4.

Corollary 23 *Let $q, n, d \in \mathbb{N}^+$ such that $3 \leq q < d - \sqrt{d} + \Theta(1)$ and $n \geq 2q + 2\lceil\sqrt{q-1}\,\rceil - 1$. Then, there exists a constant $c > 0$ such that any equivalent-structure learning algorithm for q-colorings and the graph family $\mathcal{G}(n,d)$ that succeeds with probability at least $\exp(-cn)$ requires at least $\exp(cn)$ samples.*

Proof Let $k = 2q + 2\lceil\sqrt{q-1}\,\rceil - 1$. If k divides n, then take $m = n/k$. By Fact 20, every graph in $\mathcal{G}_{m,t}$ has $n = mk$ vertices and maximum degree $q + \lceil\sqrt{q-1}\,\rceil + 1$ and so $\mathcal{G}_{m,t} \subseteq \mathcal{F}_{m,t} \subseteq \mathcal{G}(n,d)$ provided $d \geq q + \lceil\sqrt{q-1}\,\rceil + 1$. Theorem 21 implies that there exists $c = c(q) > 0$ such that any equivalent-structure learning algorithm for $\mathcal{F}_{m,t}$ with success probability at least $\exp(-cn)$ requires $\exp(cn)$ samples. By definition, the set of q-colorings of any graph in $\mathcal{G}(n,d)\backslash\mathcal{F}_{m,t}$ is distinct from the set of q-colorings of any graph in $\mathcal{F}_{m,t}$. Since also $\mathcal{F}_{m,t} \subseteq \mathcal{G}(n,d)$, equivalent-structure learning in $\mathcal{G}(n,d)$ is strictly harder than in $\mathcal{F}_{m,t}$. Specially, any equivalent-structure learning algorithm for $\mathcal{G}(n,d)$ with success probability at least $\exp(-cn)$ requires at least $\exp(cn)$ samples. Note that $d \geq q + \lceil\sqrt{q-1}\,\rceil + 1$ implies $q < d - \sqrt{d} + \Theta(1)$.

The result follows in similar fashion when k does not divide n, but we are required to modify slightly the graph families $\mathcal{G}_{m,t}$ and $\mathcal{F}_{m,t}$. Suppose $n = km + r$ where $1 \leq r \leq k-1$ and let $W = \{w_0, \ldots, w_{r-1}\}$ be a simple path. For every $G \in \mathcal{G}_{m,t}$, add W and the edge $\{s_m, w_0\}$ to G to obtain a graph $\hat{G}$. Let $\hat{\mathcal{G}}_{m,t}$ be the resulting graph family. Every graph in $\hat{\mathcal{G}}_{m,t}$ has exactly n vertices and maximum degree $q + \lceil\sqrt{q-1}\,\rceil + 1$. Moreover, every q-coloring of $G \in \mathcal{G}_{m,t}$ corresponds to exactly $(q-1)^r$ colorings of $\hat{G} \in \hat{\mathcal{G}}_{m,t}$. Define $\hat{\mathcal{F}}_{m,t}$ as before; i.e., $\hat{\mathcal{F}}_{m,t}$ is the set of all graphs in $\mathcal{G}(n,d)$ that have the same set of colorings as some graph in $\hat{\mathcal{G}}_{m,t}$. The argument in the proof of Theorem 21 and Fact 22 imply that any equivalent-structure learning algorithm for $\hat{\mathcal{F}}_{m,t}$ with success probability at least $\exp(-cn)$ requires $\exp(cn)$ independent samples, where $c = c(q) > 0$ is a suitable constant. Since we have $\hat{\mathcal{G}}_{m,t} \subseteq \hat{\mathcal{F}}_{m,t} \subseteq \mathcal{G}(n,d)$ for $d \geq q + \lceil\sqrt{q-1}\,\rceil + 1$, the result follows. ■

We conclude this section with the proofs of Lemma 19, Fact 20 and Fact 22.

Proof of Lemma 19 Let σ be a q-coloring of $G_{m,t}$. For $1 \leq i < m$, since every vertex in I_i is adjacent to every vertex in C_i, all vertices in I_i have the same color in σ, which is the one color not used by C_i. Moreover, every vertex in C_{i+1} is adjacent to a vertex of I_i, and so C_{i+1} is colored with the same set of $q-1$ colors as C_i in σ. Then, every clique $C_1, \ldots, C_m$ is colored with the same set of $q-1$ colors and every independent set $I_1, \ldots, I_m$ is colored with the one remaining color in σ. The same holds for $C'_1, \ldots, C'_m$ and $I'_1, \ldots, I'_m$ as well. ■

Proof of Fact 20 The number of vertices in $G_{m,t}$ is $m(|C_1| + |C_1'| + |I_1| + |I_1'| + 1) = (2q + 2t - 1)m$. The degree of the vertices in the cliques C_i or C_i' is at most $q - 2 + t + 1 = q - 1 + t$. Moreover, the degree of the vertices in the independent sets I_i or I_i' is at most $q-1+\lceil(q-1)/t\rceil+1 = q + \lceil(q-1)/t\rceil$. Thus, the maximum degree of $G_{m,t}$ is no more than

$$\max\left\{q + t - 1, q + \left\lceil (q-1)/t \right\rceil\right\}.$$

Therefore, the maximum degree of any graph in $\mathcal{G}_{m,t}$ is at most

$$\max\left\{q + t, q + \left\lceil (q-1)/t \right\rceil + 1\right\}.$$ ■

Proof of Fact 22 Let $\mathcal{A}$ be any (possibly randomized) equivalent-structure learning algorithm that, given L independent samples $\Gamma = (\sigma^{(1)}, \ldots, \sigma^{(L)}) \in \Omega_{\mathcal{F}_i}^L$ from an unknown distribution $\pi_{\mathcal{F}_i} = \pi_{G^*}$ for some $G^* \in \mathcal{F}_i$, outputs a graph $\mathcal{A}(\Gamma)$ in $\hat{\mathcal{F}}$. For any G^*, the probability that $\mathcal{A}$ equivalently learns the graph given L independent samples from $\pi_{\mathcal{F}_i}$ is

$$\Pr[\mathcal{A}(\Gamma) \in \mathcal{F}_i] = \sum_{x \in \Omega_{\mathcal{F}_i}^L} \Pr[\Gamma = x] \Pr[\mathcal{A}(x) \in \mathcal{F}_i].$$

Recall that by assumption $\Omega_{\mathcal{F}_r} \subseteq \Omega_{\mathcal{F}_i}$. Let $\mathcal{T}$ be the set of all sample sequences $\sigma^{(1)}, \ldots, \sigma^{(L)}$ such that $\sigma^{(j)} \notin \Omega_{\mathcal{F}_r}$ for at least one j; namely, $\mathcal{T} = \Omega_{\mathcal{F}_i}^L \backslash \Omega_{\mathcal{F}_r}^L$. Note that

$$|\mathcal{T}| = |\Omega_{\mathcal{F}_i}|^L - |\Omega_{\mathcal{F}_r}|^L.$$

Then,

$$\begin{aligned}\Pr[A(\Gamma) \in \mathcal{F}_i] &= \sum_{x \in \Omega_{\mathcal{F}_r}^L} \frac{1}{|\Omega_{\mathcal{F}_i}|^L} \cdot \Pr[\mathcal{A}(x) \in \mathcal{F}_i] + \sum_{x \in \mathcal{T}} \frac{1}{|\Omega_{\mathcal{F}_i}|^L} \cdot \Pr[\mathcal{A}(x) \in \mathcal{F}_i] \\ &\leq \frac{1}{|\Omega_{\mathcal{F}_r}|^L} \sum_{x \in \Omega_{\mathcal{F}_r}^L} \Pr[\mathcal{A}(x) \in \mathcal{F}_i] + \frac{|\Omega_{\mathcal{F}_i}|^L - |\Omega_{\mathcal{F}_r}|^L}{|\Omega_{\mathcal{F}_i}|^L}.\end{aligned}$$

Since $\Omega_{\mathcal{F}_i} \subseteq \Omega_{\mathcal{F}_1}$, we get

$$\frac{|\Omega_{\mathcal{F}_i}|^L - |\Omega_{\mathcal{F}_r}|^L}{|\Omega_{\mathcal{F}_i}|^L} = 1 - \frac{|\Omega_{\mathcal{F}_r}|^L}{|\Omega_{\mathcal{F}_i}|^L} \leq 1 - \frac{|\Omega_{\mathcal{F}_r}|^L}{|\Omega_{\mathcal{F}_1}|^L} = 1 - (1-\eta)^L \leq L\eta.$$

Suppose the equivalent-structure learning algorithm $\mathcal{A}$ has success probability at least $1/r + \alpha$, then

$$\Pr[\mathcal{A}(\Gamma) \in \mathcal{F}_i] \geq \frac{1}{r} + \alpha.$$

Hence,

$$\frac{1}{r} + \alpha \leq \frac{1}{|\Omega_{\mathcal{F}_r}|^L} \sum_{x \in \Omega_{\mathcal{F}_r}^L} \Pr[\mathcal{A}(x) \in \mathcal{F}_i] + L\eta.$$

Since $\sum_{i=1}^r \Pr[\mathcal{A}(x) \in \mathcal{F}_i] = 1$ for any fixed sample sequence x, summing up over i we get

$$1 + r\alpha \leq \frac{1}{|\Omega_{\mathcal{F}_r}|^L} \sum_{x \in \Omega_{\mathcal{F}_r}^L} \sum_{i=1}^{r} \Pr[\mathcal{A}(x) \in \mathcal{F}_i] + rL\eta = 1 + rL\eta.$$

Hence, $L \geq \alpha/\eta$ as claimed. ■

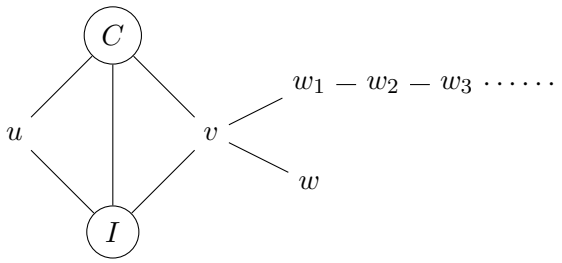

Figure 4: The graph G where $C = K_{q-3}$ and I is an independent set of size $d - q + 1$.

B.2. General lower bound for proper q-colorings

When $d - \sqrt{d} + \Theta(1) \leq q \leq d$ structure learning for q-colorings is not identifiable, and the strong lower bound from Part 3 of Theorem 15 (i.e., Corollary 23) does not apply either. In this subsection we establish a weaker but more general lower bound for proper colorings that applies in this regime. Specifically, we provide a family of graphs $\mathcal{F} \subseteq \mathcal{G}(n, d)$ such that the number of random q-colorings required to learn any graph in $\mathcal{F}$ with success probability at least $1/2$ is $\exp(\Omega(d - q))$.

Theorem 24 *Let $d, q, n \in \mathbb{N}^+$ such that $3 \leq q < d$ and $n \geq d + 2$. Then, any equivalent-structure learning algorithm for $\mathcal{G}(n, d)$ with success probability at least $1/2$ requires at least $\exp(\Omega(d - q))$ samples.*

Let $d, q, n \in \mathbb{N}^+$ such that $3 \leq q < d$ and $n \geq d + 2$. Let C be a clique of size $q - 3$, let I be an independent set of size $d - q + 1$ and let $W = \{w_1, \ldots, w_{n-d-1}\}$ be a simple path. Also, let u, v, w be three additional vertices that are not in C, I or W. Define the graph $G = (V, E)$ such that

$$V = V(C) \cup V(I) \cup \{w_1, \ldots, w_{n-d-1}\} \cup \{u, v, w\},$$

where $V(C)$ and $V(I)$ are the vertices in C and I, respectively. In addition to the edges in C and W, G has the following edges:

1. every vertex in C is adjacent to every vertex in I;
2. u and v are adjacent to every vertex in C and I;
3. w, w_1 are adjacent to v;

see Figure 4.

Let

$$\mathcal{G} = \{G_1 = G, G_2 = (V, E \cup \{uw\}), G_3 = (V, E \cup \{uw_1\}), G_4 = (V, E \cup \{uw, uw_1\})\}.$$

Note that every graph in $\mathcal{G}$ is an n-vertex graph of maximum degree at most d and so $\mathcal{G} \subseteq \mathcal{G}(n, d)$. Furthermore, for $1 \leq i \leq 4$ let $\mathcal{F}_i$ be the family of all graphs in $\mathcal{G}(n, d)$ that have the same set of q-colorings as G_i and let $\mathcal{F} = \bigcup_{i=1}^{4} \mathcal{F}_i$. The following theorem immediately implies Theorem 24.

Theorem 25 *Let $d, q, n \in \mathbb{N}^+$ such that $3 \leq q < d$ and $n \geq d+2$. Then the number of independent random q-colorings required to learn any graph in $\mathcal{F}$ with probability at least $1/2$ is $\exp(\Omega(d - q))$.*

Proof If u and v receive distinct colors in a q-coloring of G, then the clique C will be colored by $q-3$ of the $q-2$ colors not used by u and v and the number of available colors for every vertex in I is only 1. Thus, the number of colorings of G where u and v receive distinct colors is $q!(q-1)^{n-d}$. Otherwise, if u and v receive the same color in a coloring of G, then C will use $q-3$ of the $q-1$ available colors and every vertex in I has 2 available color choices. Hence, the number of such colorings is $q!(q-1)^{n-d} \cdot 2^{d-q}$. Any q-coloring of G where u and v receive the same color is also a proper q-coloring of G_4. Therefore, we get

$$\frac{|\Omega_{\mathcal{F}_4}|}{|\Omega_{\mathcal{F}_1}|} = \frac{|\Omega_{G_4}|}{|\Omega_G|} \geq \frac{q!(q-1)^{n-d} \cdot 2^{d-q}}{q!(q-1)^{n-d} + q!(q-1)^{n-d} \cdot 2^{d-q}} = 1 - \mathrm{e}^{-\Omega(d-q)} .$$

Thus, it follows from Fact 22 that any equivalent-structure learning algorithm for $\mathcal{F}$ that succeeds with probability at least $1/2$ requires

$$L \geq \Big(\frac{1}{2} - \frac{1}{4}\Big)\Big(1 - \frac{|\Omega_{\mathcal{F}_4}|}{|\Omega_{\mathcal{F}_1}|}\Big)^{-1} = \mathrm{e}^{\Omega(d-q)}$$

samples. ■

Theorem 24 follows immediately from Theorem 25 and the fact that $\mathcal{F} \subseteq \mathcal{G}(n,d)$.

Appendix C. Proofs missing from Section 4

Proof of Lemma 17 For $u \in V$ we show first that $\Pr[X_u = i] \geq (1-\alpha)/q$. If σ is an H-coloring of G sampled according to π_G, we may update the color of any vertex $w \in V$ by choosing a new color for w uniformly at random among the available colors for w given $\sigma(V \setminus w)$. The resulting H-coloring after this update has distribution π_G.

Suppose σ_0 is an H-coloring of G sampled according to π_G, and let τ_0 be the color assignment that agrees with σ_0 everywhere except possibly at u, where we set $\tau_0(u) = i$. (Note that τ is not necessarily a valid H-coloring.)

Let $\partial u = \{v_1, \ldots, v_l\}$. We update the configuration in v_1, then in v_2 and so on, in both σ_0 and τ_0; then we update the color of u. Let σ_k and τ_k be the configuration after updating v_k in σ_{k-1} and τ_{k-1}, respectively. The color of v_k in both σ_{k-1} and τ_{k-1} is updated using the optimal coupling ν_k between the distributions $\pi_{v_k}(\cdot|\sigma_{k-1})$ and $\pi_{v_k}(\cdot|\tau_{k-1})$ as follows. Sample (a_k, b_k) from ν_k and let $\sigma_k(V \setminus v_k) = \sigma_{k-1}(V \setminus v_k)$, $\sigma_k(v_k) = a_k$, $\tau_k(V \setminus v_k) = \tau_{k-1}(V \setminus v_k)$ and $\tau(v_k) = b_k$. After updating $\partial u = \{v_1, \ldots, v_l\}$ in this manner, σ_l has law π_G. Moreover,

$$\begin{aligned}
\Pr[\sigma_l \neq \tau_l] &\leq \Pr_{\nu_l}[\sigma_l \neq \tau_l | \sigma_{l-1} = \tau_{l-1}] + \Pr[\sigma_{l-1} \neq \tau_{l-1}] \\
&\leq \sum_{k=1}^{l} \Pr_{\nu_k}[\sigma_k \neq \tau_k | \sigma_{k-1} = \tau_{k-1}] \\
&= \sum_{k=1}^{l} \|\pi_{v_k}(\cdot|\tau_{k-1}) - \pi_{v_k}(\cdot|\tau_{k-1})\|_{\mathrm{TV}} \\
&\leq \alpha,
\end{aligned}$$

where the last inequality follows from the definition of the Dobrushin condition. Hence, with probability at least $1-\alpha$, $\sigma_l = \tau_l$. If this is the case, then color i is compatible with $\sigma_l(V \setminus u)$ and thus when u is updated it receives color i with probability at least $1/q$. Thus, we get

$$\Pr[X_u = i] \geq \frac{(1-\alpha)}{q}.$$

Finally, let $v \in V$ such that $v \notin \partial u$. Using the procedure described above to update the configuration in $\partial u \cup u$, and then in $\partial v \cup v$ we obtain

$$\Pr[X_u = i, X_v = j] \geq \frac{(1-\alpha)^2}{q^2}.$$ ■

Appendix D. Approximate-structure learning of H-colorings

In addition to structure learning (exact recovery of the hidden graph G) and equivalent-structure learning (learning a graph with the same set of H-colorings), we may consider the corresponding approximation problem of finding a graph $\hat{G}$ such that $\pi_{\hat{G}}$ is close to π_G in some notion of distance, such as total variation distance or Kullback-Leibler divergence. Apparently, this task is much simpler for hard-constraint systems.

In this section we consider this approximation variant of structure learning for hard-constraint systems with respect to total variation distance. Specifically, given L samples $\sigma^{(1)}, \ldots, \sigma^{(L)}$ from π_G, we consider the problem of finding a graph $\hat{G}$ such that

$$\left\|\pi_G - \pi_{\hat{G}}\right\|_{\text{TV}} < \gamma,$$

where $\gamma > 0$ is a desired precision.

Theorem 26 *Suppose $H \neq K_q^+$ and let $\hat{G}$ be the output of the* STRUCTLEARN-H *algorithm. Then, for all $\varepsilon \in (0,1)$ and $\gamma \in (0,1)$,*

$$\Pr\left[\left\|\pi_G - \pi_{\hat{G}}\right\|_{\text{TV}} < \gamma\right] \geq 1 - \varepsilon$$

provided $L \geq 4\gamma^{-1} n^2 \log(\frac{n^2}{2\varepsilon})$.

Recall that the running time of STRUCTLEARN-H is $O(Ln^2)$, so from Theorem 26 we get an algorithm for approximate structure learning with running time $O(\gamma^{-1} n^4 \log(\frac{n}{\varepsilon}))$.

Proof of Theorem 26 Let $\hat{G} = (V(\hat{G}), E(\hat{G}))$. Recall that $\{u,v\} \notin E(\hat{G})$ if and only if u, v receive incompatible colors in one of the samples $\sigma^{(1)}, \ldots, \sigma^{(L)}$ from π_G. Hence, $\hat{G}$ is a supergraph of G and so $\Omega_{\hat{G}} \subseteq \Omega_G$. Moreover,

$$\left\|\pi_G - \pi_{\hat{G}}\right\|_{\text{TV}} = \sum_{\sigma \in \Omega_G \setminus \Omega_{\hat{G}}} \frac{1}{|\Omega_G|} = \frac{|\Omega_G| - |\Omega_{\hat{G}}|}{|\Omega_G|} = \Pr\left[\sigma \notin \Omega_{\hat{G}}\right],$$

assuming σ is an H-coloring of G chosen uniformly at random (i.e., σ is drawn from π_G). If we let $\Gamma = E(\hat{G}) \setminus E(G)$, then

$$\Pr\left[\sigma \notin \Omega_{\hat{G}}\right] = \Pr\left[\exists \{u,v\} \in \Gamma : \{\sigma_u, \sigma_v\} \notin E(H)\right] \leq \sum_{\{u,v\} \in \Gamma} \Pr\left[\{\sigma_u, \sigma_v\} \notin E(H)\right] \quad (5)$$

by a union bound.

Now, for $\gamma > 0$ let

$$M_\gamma = \left\{ \{u,v\} \notin E(G) : \Pr\left[\{\sigma_u, \sigma_v\} \notin E(H)\right] \geq \frac{2\gamma}{n^2} \right\}.$$

Let Z_{uv} be the number of samples $\sigma^{(1)}, \ldots, \sigma^{(L)}$ where vertices u and v receive incompatible colors. A Chernoff bound implies that, for any $\{u,v\} \in M_\gamma$,

$$\Pr[Z_{uv} = 0] \leq \Pr\left[Z_{uv} \leq \frac{\gamma L}{n^2}\right] \leq \exp\left(\frac{-\gamma L}{4n^2}\right) \leq \frac{2\varepsilon}{n^2}.$$

A union bound then implies that with probability at least $1-\varepsilon$ all edges in M_γ are not in $E(\hat{G})$. Hence, with probability at least $1-\varepsilon$, all edges in $E(\hat{G})$ satisfy:

$$\Pr\left[\{\sigma_u, \sigma_v\} \notin E(H)\right] < \frac{2\gamma}{n^2}.$$

Plugging this bound into (5), we get

$$\Pr\left[\left\|\pi_G - \pi_{\hat{G}}\right\|_{\text{TV}} < \gamma\right] \geq 1 - \varepsilon$$

as desired. ■

Appendix E. Learning weighted H-colorings

In this section we consider the more general setting of weighted H-colorings. We restrict our attention to constraint graphs with at least one hard constraint, which corresponds to spin systems with hard constraints.

E.1. Spin systems with hard constraints

Let $G = (V, E, \theta)$ be an undirected weighted graph with weights given by the function $\theta : E \cup V \to \mathbb{R}^+$. (For definiteness we only consider a positive weight function θ.) A *spin system* on the graph G consists of a set of *spins* $[q] = \{1, \ldots, q\}$, a symmetric *edge potential* $J : [q] \times [q] \to \mathbb{R} \cup \{-\infty\}$ and a *vertex potential* $h : [q] \to \mathbb{R}$. A *configuration* $\sigma : V \to [q]$ of the system is an assignment of spins to the vertices of G. Each configuration $\sigma \in [q]^V$ is assigned probability

$$\pi_G(\sigma) = \frac{1}{Z_G} \exp\left(\sum_{(u,v)\in E} \theta(u,v) J(\sigma_u, \sigma_v) \; + \; \sum_{u \in V} \theta(u) h(\sigma_u) \right), \tag{6}$$

where Z_G is the normalizing constant called the *partition function*. If $J(i,j) = -\infty$ for some $i, j \in [q]$, then $\{i,j\}$ is a hard constraint; otherwise i and j are compatible.

Unweighted H-colorings, which were considered in Sections 2, 3 and 4, correspond to the special case where $\theta = 1, h = 0$ and

$$J(i,j) = \begin{cases} 1 & \text{if } (i,j) \in E(H); \\ -\infty & \text{if } (i,j) \notin E(H). \end{cases}$$

In this section we consider the structure learning problem for a class of models known as *permissive systems*. This is a widely used notion in statistical physics for spin systems with hard constraints; see, e.g., [19, 36, 15]. There are several different notions in the literature, but we consider here the weakest one (i.e., the easiest to satisfy). Roughly, the condition says that for any boundary condition there is always a valid configuration for the interior.

Definition 27 *A spin system is called permissive if for any $A \subseteq V$ and any valid configuration τ on $V \setminus A$, there is at least one valid configuration σ on A such that $\pi(\sigma|\tau) > 0$.*

Independent sets, and more generally the hard-core model, are examples of permissive models since we can assign spin 0 (unoccupied) to the vertices in A.

E.2. Structure learning for spin systems with hard constraints

We first formalize the notion of structure learning for the setting of weighted constraint graphs. Suppose we know the number of spins q, the edge potential $J \in \mathbb{R}^{q\times q}$ and the vertex potential $h \in \mathbb{R}^q$ of a spin system $\mathcal{S}$. Consider the family of graphs

$$\begin{aligned}\mathcal{G}(n,d,\alpha,\beta,\gamma) = \{G=(V,E,\theta) : {} & |V| = n,\\ & \Delta(G) \le d,\\ & \alpha \le |\theta(u,v)| \le \beta \text{ for all } \{u,v\} \in E,\\ & |\theta(v)| \le \gamma \text{ for all } v \in V\},\end{aligned}$$

where $\Delta(G)$ denotes the maximum degree of the graph G. Suppose that we are given L independent samples $\sigma^{(1)}, \sigma^{(2)}, \ldots, \sigma^{(L)}$ from the distribution π_G where $G \in \mathcal{G}$. A *structure learning algorithm* for the spin system $\mathcal{S}$ and the family $\mathcal{G}(n,d,\alpha,\beta,\gamma)$ takes as input the sample sequence $\sigma^{(1)}, \sigma^{(2)}, \ldots, \sigma^{(L)}$ and outputs an estimator $\hat{G} \in \mathcal{G}(n,d,\alpha,\beta,\gamma)$ such that $\Pr[G = \hat{G}] \ge 1-\varepsilon$, where $\varepsilon > 0$ is a prescribed failure probability.

E.3. Learning permissive spin systems

In this section we analyze the running time and sample complexity of the STRUCTLEARN-H algorithm for permissive spin systems.

Let $\hat{\gamma} = \gamma \cdot \max_{i\in[q]} |h(i)|$ and $\hat{\beta} = \beta \cdot \max_{i,j\in[q]} |J(i,j)|$. Recall that for $v \in V$, X_v denotes the random variable for the color of v under π_G. We show that for permissive systems, the running time of STRUCTLEARN-H is polynomial in the size of the graph, but depends exponentially on $\hat{\gamma}$, $\hat{\beta}$ and its maximum degree.

Theorem 28 *Let $G = (V,E) \in \mathcal{G}(n,d,\alpha,\beta,\gamma)$ and suppose that $\mathcal{S}$ is a permissive spin system with at least one hard constraint. Then, if the structure learning algorithm receives as input*

$$L \ge 8q^{2(d+1)}\mathrm{e}^{4(2\hat{\beta}d^2+\hat{\gamma})}\log\left(\frac{n^2}{2\varepsilon}\right)$$

independent samples from π_G, it outputs the graph G with probability at least $1-\varepsilon$ and has running time $O(Ln^2)$.

Theorem 28 yields a structure learning algorithm for the hard-core model for all $\lambda > 0$; thus it generalizes the algorithmic result of [6]. We observe also that the running time of our algorithm for permissive systems is comparable to the running time of the optimal structure learning algorithms for soft-constraint systems in [32].

Theorem 28 is a direct corollary of the following lemma and Lemma 13.

Lemma 29 *Suppose that $\mathcal{S}$ is a permissive spin system with at least one hard constraint $\{i,j\} \in [q] \times [q]$ on a graph $G = (V,E) \in \mathcal{G}(n,d,\alpha,\beta,\gamma)$. Then, for all $\{u,v\} \notin E$*

$$\Pr[X_u = i, X_v = j] \geq \frac{1}{q^{2(d+1)}\mathrm{e}^{4(2\hat{\beta}d^2+\hat{\gamma})}}.$$

In the proof of Lemma 29 we use the following fact.

Fact 30 *Let $R \subseteq V$ and let τ be a configuration on ∂R. If $\Omega^\tau(R) \neq \emptyset$ is the set of valid configurations on R given τ, then for any $\sigma \in \Omega^\tau(R)$*

$$\Pr[X_R = \sigma \mid X_{\partial R} = \tau] \geq \frac{1}{q^{|R|}\mathrm{e}^{2(\hat{\beta}d|R|+\hat{\gamma})}}.$$

We are now ready to prove Lemma 29.

Proof of Lemma 29 For any $A \subseteq V$ and any spin configuration σ of A, with a slight abuse of notation we use $\{\sigma\}$ for the event $\{X_A = \sigma\}$.

Let $u, v \in V$ such that $\{u,v\} \notin E$ and let N_1 and N_2 be the set of vertices at distances one and two, respectively, from $\{u,v\}$; i.e., $N_1 = \partial u \cup \partial v$ and $N_2 = \{w \in \partial N_1 : w \neq u, w \neq v\}$. Let Ω_1 and Ω_2 be the set of valid configurations for N_1 and N_2, respectively. Then,

$$\Pr[X_u = i, X_v = j] \geq \min_{\tau_2 \in \Omega_2} \Pr[X_u = i, X_v = j \mid \tau_2]. \tag{7}$$

Since the spin system is permissive, for any $\tau_2 \in \Omega_2$ there exists $\tau_1 \in \Omega_1$ such that

$$\Pr[\tau_1 \mid \tau_2, X_u = i, X_v = j] > 0.$$

Then,

$$\Pr[X_u = i, X_v = j \mid \tau_2] \geq \Pr[X_u = i, X_v = j \mid \tau_1] \Pr[\tau_1 \mid \tau_2] \geq \frac{1}{q^2\mathrm{e}^{2(2\hat{\beta}d+\hat{\gamma})}} \Pr[\tau_1 \mid \tau_2], \tag{8}$$

by Fact 30. Now,

$$\Pr[\tau_1 \mid \tau_2] = \sum_{a,b \in [q]} \Pr[\tau_1 \mid X_u = a, X_v = b, \tau_2] \Pr[X_u = a, X_v = b \mid \tau_2].$$

Since $|N_1| \leq 2d$, by Fact 30, $\Pr[\tau_1 \mid X_u = a, X_v = b, \tau_2] \geq \frac{1}{q^{2d}\mathrm{e}^{2(2\hat{\beta}d^2+\hat{\gamma})}}$. Together with (7) and (8) this implies

$$\Pr[X_u = i, X_v = j] \geq \frac{1}{q^{2(d+1)}\mathrm{e}^{4(2\hat{\beta}d^2+\hat{\gamma})}}.$$ ■

Remark 31 *A simplified version of this argument can be used to show that in a permissive H-coloring, for any hard constraint $\{i,j\} \notin E(H)$, $\Pr[X_u = i, X_v = j] \geq 1/q^{2d}$. From this we obtain a structure learning algorithm for permissive H-colorings with running time $O(q^{2d}n^2 \log n)$ via Lemma 13.*

We conclude this section with the proof of Fact 30.

Proof of Fact 30 For $\sigma \in \Omega^\tau(R)$ let

$$w(\sigma) = \exp\left[\sum_{u\in R}\sum_{v\in\partial u\cap R}\theta(u,v)J(\sigma_u,\sigma_v) + \sum_{u\in R}\sum_{v\in\partial u\cap\partial R}\theta(u,v)J(\sigma_u,\tau_v) + \theta(u)h(\sigma_u)\right].$$

Then,

$$\Pr[X_R = \sigma \mid X_{\partial R} = \tau] = \frac{w(\sigma)}{Z_{R,\tau}},$$

with $Z_{R,\tau} = \sum_{\sigma'\in\Omega^\tau(R)} w(\sigma')$. Observe that for all $\sigma' \in \Omega^\tau(R)$, $\mathrm{e}^{-\hat{\beta}d|R|-\hat{\gamma}} \le w(\sigma') \le \mathrm{e}^{\hat{\beta}d|R|+\hat{\gamma}}$. Hence, $Z_{R,\tau} \le q^{|R|}\mathrm{e}^{\hat{\beta}d|R|+\hat{\gamma}}$ and

$$\Pr[X_R = \sigma \mid X_{\partial R} = \tau] \ge \frac{1}{q^{|R|}\mathrm{e}^{2(\hat{\beta}d|R|+\hat{\gamma})}}.$$

■

E.4. Identifiability for weighted H-colorings

We prove next an analog of our characterization theorem (Theorem 2) for identifiability of weighted H-colorings. The edge potential J corresponds to the weighted adjacency matrix of a weighted constraint graph $H^J = (V(H^J), E(H^J))$, where $V(H^J) = \{1,\dots,q\}$, $\{i,j\} \notin E(H^J)$ iff $J(i,j) = -\infty$, and the weight of $\{i,j\} \in E(H^J)$ is $J(i,j)$. As before, we say that a graph G is H^J-colorable if there is a valid H^J-coloring for G. If $\{i,j\} \notin E(H^J)$ we call $\{i,j\}$ a hard constraint. The notion of identifiability extends to the weighted setting as follows.

Definition 32 *A weighted constraint graph H^J is said to be identifiable with respect to a family of H^J-colorable graphs $\mathcal{G}$ if for any two distinct graphs $G_1, G_2 \in \mathcal{G}$ we have $\pi_{G_1} \neq \pi_{G_2}$. In particular, when $\mathcal{G}$ is the set of all finite H^J-colorable graphs we say that H^J is identifiable.*

(Definition 1 is the analog definition in the unweighted setting.)

In our characterization theorem we consider the supergraphs G_{ij}'s introduced in the unweighted setting; see Definition 6 and Figure 1.

Theorem 33 *Let H^J be a weighted constraint graph with at least one hard constraint. If H^J has a self-loop, then H^J is identifiable. Otherwise H^J is identifiable if and only if for each $\{i,j\} \in E(H^J)$ there exists an H^J-coloring σ of G_{ij} such that*

$$J(\sigma_i,\sigma_j) + J(\sigma_{i'},\sigma_{j'}) \neq J(\sigma_{i'},\sigma_j) + J(\sigma_i,\sigma_{j'}).$$

(Recall that i' and j' are the copies of the vertices i and j in G_{ij}.)

Proof For clarity, we shall assume in this proof that the underlying graph $G = (V,E)$ is unweighted and that there is no external field; i.e., $\theta = 1$ and $h = 0$. The same proof generalizes to spin systems on weighted graphs with external field.

Henceforth we use H for H^J to simplify the notation. We consider first the case when H has no self-loops. For the forward direction we consider the contrapositive. Suppose that there exists $\{i,j\} \in E(H)$ such that for every proper H-coloring σ of G_{ij} we have

$$J(\sigma_i,\sigma_j) + J(\sigma_{i'},\sigma_{j'}) = J(\sigma_{i'},\sigma_j) + J(\sigma_i,\sigma_{j'}).$$

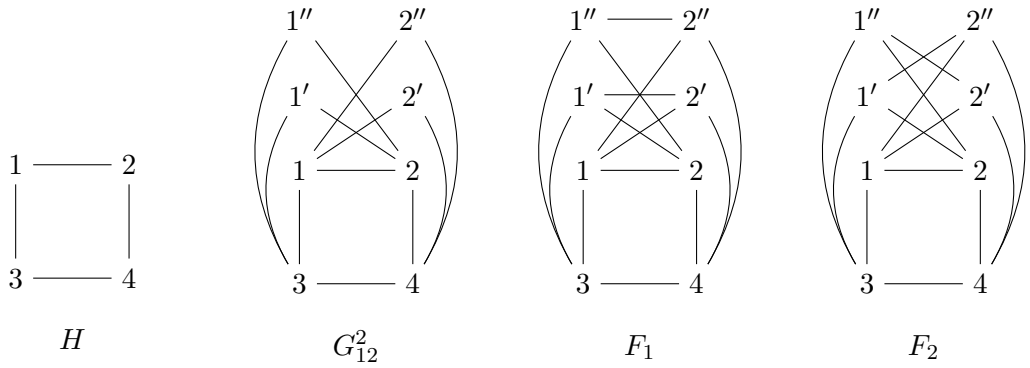

Figure 5: A constraint graph H, its supergraph G_{12}^2 and the graphs F_1 and F_2.

Under this assumption we construct two distinct H-colorable graphs F_1, F_2 such that $\pi_{F_1} = \pi_{F_2}$; this implies that H is not identifiable, which would complete the proof of the forward direction. For this, for each $\{i, j\} \in E(H)$ let us define the supergraph G_{ij}^2 of H that is the result of creating two copies i', i'' of vertex i and two copies j', j'' of vertex j, with no edges between i', i'', j', j''. Formally, for each $\{i, j\} \in E(H)$, we define the graph $G_{ij}^2 = (V(G_{ij}^2), E(G_{ij}^2))$ as follows:

1. $V(G_{ij}) = V(H) \cup \{i', i'', j', j''\}$ where i', i'', j', j'' are four new colors;
2. If $\{a, b\} \in E(H)$, then the edge $\{a, b\}$ is also in $E(G_{ij}^2)$;
3. For each $k \in V(G_{ij}^2) \setminus \{i', i'', j', j''\}$, the edges $\{i', k\}$ and $\{i'', k\}$ are in G_{ij}^2 if and only if the edge $\{i, k\}$ is in H, and similarly $\{j', k\}, \{j'', k\} \in E(G_{ij}^2)$ if and only if $\{j, k\} \in E(H)$;

see Figure 5 for an example.

Let σ be an H-coloring of G_{ij}^2. Since the subgraphs induced by $V(G_{ij}^2) \setminus \{i^*, j^*\}$ with $i^* \in \{i', i''\}$ and $j^* = \{j', j''\}$ are all isomorphic to G_{ij}, our assumption implies

$$J(\sigma_i, \sigma_j) + J(\sigma_{i'}, \sigma_{j'}) = J(\sigma_{i'}, \sigma_j) + J(\sigma_i, \sigma_{j'}), \tag{9}$$
$$J(\sigma_i, \sigma_j) + J(\sigma_{i''}, \sigma_{j''}) = J(\sigma_{i''}, \sigma_j) + J(\sigma_i, \sigma_{j''}), \tag{10}$$
$$J(\sigma_i, \sigma_j) + J(\sigma_{i''}, \sigma_{j'}) = J(\sigma_{i''}, \sigma_j) + J(\sigma_i, \sigma_{j'}), \tag{11}$$
$$J(\sigma_i, \sigma_j) + J(\sigma_{i'}, \sigma_{j''}) = J(\sigma_{i'}, \sigma_j) + J(\sigma_i, \sigma_{j''}). \tag{12}$$

Since the sum of the right-hand sides of (9) and (10) is equal to the sum of the right-hand sides of (11) and (12), we get

$$J(\sigma_{i'}, \sigma_{j'}) + J(\sigma_{i''}, \sigma_{j''}) = J(\sigma_{i''}, \sigma_{j'}) + J(\sigma_{i'}, \sigma_{j''}). \tag{13}$$

Now, let

$$F_1 = (V(G_{ij}^2), E(G_{ij}^2) \cup \{\{i', j'\}, \{i'', j''\}\}),$$
$$F_2 = (V(G_{ij}^2), E(G_{ij}^2) \cup \{\{i'', j'\}, \{i', j''\}\});$$

see Figure 5. Then, using (6), for any H-coloring σ of G_{ij}^2 we have

$$\frac{\pi_{F_1}(\sigma)}{\pi_{F_2}(\sigma)} = \frac{Z_{F_1}^{-1} \exp\left(J(\sigma_{i'}, \sigma_{j'}) + J(\sigma_{i''}, \sigma_{j''})\right)}{Z_{F_2}^{-1} \exp\left(J(\sigma_{i''}, \sigma_{j'}) + J(\sigma_{i'}, \sigma_{j''})\right)} = \frac{Z_{F_2}}{Z_{F_1}},$$

which is a constant independent of σ. By (13), π_{F_1} and π_{F_2} have the same support. Moreover, any H-coloring of F_1 or F_2 is also an H-coloring of G_{ij}^2. Hence, we conclude that $\pi_{F_1} = \pi_{F_2}$, implying H is not identifiable. This completes the proof of the forward direction.

For the reverse direction suppose that for all $\{i, j\} \in E(H)$ there exists an H-coloring σ of G_{ij} where

$$J(\sigma_i, \sigma_j) + J(\sigma_{i'}, \sigma_{j'}) \neq J(\sigma_{i'}, \sigma_j) + J(\sigma_i, \sigma_{j'}).$$

Consider two H-colorable graphs $G_1 = (V, E_1)$ and $G_2 = (V, E_2)$ such that $\pi_{G_1} = \pi_{G_2}$. We show that for any $u, v \in V$, $\{u, v\} \in E_1$ iff $\{u, v\} \in E_2$ and thus $G_1 = G_2$. This implies that H is identifiable.

Let $u, v \in V$ and let τ be an H-coloring of G_1. Since $\pi_{G_1} = \pi_{G_2}$, then τ is also an H-coloring of G_2. Suppose $i = \tau(u)$ and $j = \tau(v)$. If i and j are not compatible, then $\{u, v\} \notin E_1$ and $\{u, v\} \notin E_2$. Thus, let us assume i and j are compatible. Let σ be an H-coloring of G_{ij} such that

$$J(a, b) + J(a', b') \neq J(a', b) + J(a, b'),$$

where $a = \sigma_i, b = \sigma_j, a' = \sigma_{i'}, b' = \sigma_{j'}$; we know such an H-coloring exists by assumption.

Now consider the conditional distribution $\pi^{\tau}_{G_1,uv}$ on the vertices u and v of G_1 given the configuration $\tau(V \setminus \{u, v\})$. Then,

$$\begin{aligned}
p_1(G_1) &:= \pi^{\tau}_{G_1,uv}(X_u = a, X_v = b) = \frac{1}{Z_{\text{cond}}(G_1)} \exp[h_a + h_b + \mathbb{1}(\{u, v\} \in E_1) J(a, b)] \\
p_2(G_1) &:= \pi^{\tau}_{G_1,uv}(X_u = a', X_v = b') = \frac{1}{Z_{\text{cond}}(G_1)} \exp[h_{a'} + h_{b'} + \mathbb{1}(\{u, v\} \in E_1) J(a', b')] \\
p_3(G_1) &:= \pi^{\tau}_{G_1,uv}(X_u = a', X_v = b) = \frac{1}{Z_{\text{cond}}(G_1)} \exp[h_{a'} + h_b + \mathbb{1}(\{u, v\} \in E_1) J(a', b)] \\
p_4(G_1) &:= \pi^{\tau}_{G_1,uv}(X_u = a, X_v = b') = \frac{1}{Z_{\text{cond}}(G_1)} \exp[h_a + h_{b'} + \mathbb{1}(\{u, v\} \in E_1) J(a, b')],
\end{aligned}$$

where $Z_{\text{cond}}(G_1)$ is the normalizing factor for $\pi^{\tau}_{G_1,uv}$,

$$\begin{aligned}
h_a &= \sum_{w \in \partial u} J(a, \tau(w)), \\
h_b &= \sum_{w \in \partial v} J(b, \tau(w))
\end{aligned}$$

and $h_{a'}, h_{b'}$ are defined in similar manner. This gives

$$\frac{p_1(G_1) p_2(G_1)}{p_3(G_1) p_4(G_1)} = \exp\left[\mathbb{1}(\{u, v\} \in E_1)(J(a, b) + J(a', b') - J(a', b) - J(a, b'))\right]$$

Since by assumption $J(a, b) + J(a', b') - J(a', b) - J(a, b') \neq 0$, $\{u, v\} \in E(G_1)$ if and only if $p_1(G_1)p_2(G_1) \neq p_3(G_1)p_4(G_1)$. Moreover, $\pi_{G_1} = \pi_{G_2}$ and thus $p_k(G_1) = p_k(G_2)$ for $k \in \{1, 2, 3, 4\}$. Hence,

$$\frac{p_1(G_1) p_2(G_1)}{p_3(G_1) p_4(G_1)} = \frac{p_1(G_2) p_2(G_2)}{p_3(G_2) p_4(G_2)}.$$

This implies that $\{u, v\} \in E_1$ iff $\{u, v\} \in E_2$ and so $G_1 = G_2$. This completes the proof of the reverse direction when H does not have self-loops. When H has at least one self-loop then using

the argument in the proof of Lemma 7, which generalizes straightforwardly to the weighted setting, we get that H is identifiable. ■