SUPPLEMENTARY MATERIAL

Supplementary Material to "Learning Loopy Graphical Models with Latent Variables: Efficient Methods and Guarantees"
(http://newport.eecs.uci.edu/anandkumar/).

1. Analysis of Ising Models.

Sufficient Conditions for Correlation Decay. We first recall the definition of correlation decay in graphical models. Given a graph $G = (W, E)$ and a distribution $P_{\mathbf{X}_W|G}$ Markov on it, and any subset $A \subset W$, let $P_{\mathbf{X}_A|G}$ denote the marginal distribution of variables in A. For some subgraph $F \subset G$, let $P_{\mathbf{X}_A|F}$ denote the marginal distribution on A obtained by setting the potentials of edges in $G \setminus F$ to zero. Thus, $P_{\mathbf{X}_A|F}$ is Markov on graph F. Let $\mathcal{N}[i; G] := \mathcal{N}(i; G) \cup i$ denote the closed neighborhood of node i in G. For any two sets $A_1, A_2 \subset W$, let $\operatorname{dist}(A_1, A_2) := \min_{i \in A_1, j \in A_2} \operatorname{dist}(i, j)$ denote the minimum graph distance[1]. Let $B_l(i)$ denote the set of nodes within graph distance l from node i and $\partial B_l(i)$ denote the boundary nodes, i.e., exactly at distance l from node i. Let $F_l(i; G) := G(B_l(i))$ denote the induced subgraph on $B_l(i)$. For any distributions P, Q, let $\|P - Q\|_1$ denote the ℓ_1 norm.

DEFINITION 1 (Correlation Decay). *A distribution $P_{\mathbf{X}_{W_m}|G}$ Markov on graph $G_m = (W_m, E_m)$ is said to exhibit correlation decay with a non-increasing rate function $\zeta(\cdot) > 0$ if for all $l, m \in \mathbb{N}$,*

$$\|P_{\mathbf{X}_A|G_m} - P_{\mathbf{X}_A|F_l(i;G_m)}\|_1 \leq \zeta(\operatorname{dist}(A, \partial B_l(i))), \quad \forall\, i \in W_m,\ A \subset B_l(i). \tag{1}$$

For Ising models, the regime of correlation decay can be explicitly established. Recall that $\Delta_{\max}$ is the maximum degree of graph G and the maximum absolute edge potential is $\theta_{\max}$.

LEMMA 1 (Correlation Decay in Ising Models). *The class of Ising models is in the regime of correlation decay when satisfies*

$$\alpha := \Delta_{\max} \tanh(\theta_{\max}) < 1. \tag{2}$$

The rate function $\zeta(\cdot)$ for correlation decay in (1) is given by

$$\zeta(l) = 2\alpha^l, \quad \forall l \in \mathbb{N}. \tag{3}$$

Proof: The above result on correlation decay is based on the concept of self-avoiding walk trees (SAW), which converts the conditional distributions of a general model to those on a tree model. See [3] for details. □

Distance Bounds. Recall that $\widehat{d}^n(i, j)$ denotes the estimated distance between nodes i and j using the empirical distribution $\widehat{P}^n_{\mathbf{X}_{i,j}}$ computed using n samples, i.e.,

$$\widehat{d}^n(i, j) := -\log|\det(\widehat{P}^n_{\mathbf{X}_{i,j}})|, \quad \forall i, j \in V. \tag{4}$$

Recall the notion of a local tree metric $\mathbf{d}_{\text{tree}}(V)$, computed by limiting the model to acyclic neighborhood subgraphs between the respective node pairs. Given a graph $G = (W, E)$, let $\operatorname{tree}(i, j; G) :=$

[1] We distinguish between the terms *graph distance* and *information distances*. The former refers to the number of edges on the shortest path connecting the two nodes on the (unweighted) graph, while the latter refers to the quantity in (4).

$G(B_l(i) \cup B_l(j))$, for $l = \lfloor g/2 \rfloor - 1$, denote the induced subgraph on $B_l(i) \cup B_l(j)$, where g is the girth of the graph. $B_l(i; G)$ denotes the set of nodes within graph distance l from i in G. When $l < g/2 - 1$ no cycles are encountered and thus the induced subgraph $\text{tree}(i, j; G)$ is acyclic. Let $P_{\mathbf{X}_{i,j} | \text{tree}(i,j)}$ denote the pairwise marginal distribution between i and j induced by considering only the subgraph $\text{tree}(i, j; G) \subset G$. Denote

$$d(i, j; \text{tree}) := -\log |\det P_{\mathbf{X}_{i,j} | \text{tree}(i,j)}|, d(i, j; G) := -\log |\det P_{\mathbf{X}_{i,j} | G}|. \tag{5}$$

Denote $\mathbf{d}_{\text{tree}}(V) := \{d(i, j; \text{tree}) : i, j \in V\}$ and $\mathbf{d}(V) := \{d(i, j; G) : i, j \in V\}$. We assume bounds

$$0 < d_{\min} \leq d(i, j; \text{tree}) \leq d_{\max} < \infty, \quad \forall (i, j) \in G, \quad \eta := \frac{d_{\max}}{d_{\min}}, \tag{6}$$

We now establish that the general distance bounds based on local tree approximation in (6) can be expressed in terms of potentials for the Ising model. Given an Ising model P with edge potentials $\boldsymbol{\theta} = \{\theta_{i,j}\}$ and node potentials $\boldsymbol{\phi} = \{\phi_i\}$, consider its attractive counterpart $\bar{P}$ with edge potentials $\bar{\boldsymbol{\theta}} := \{|\theta_{i,j}|\}$ and node potentials $\bar{\boldsymbol{\phi}} := \{|\phi_i|\}$. Define the following quantity:

$$\phi'_{\max} := \max_{i \in V} \operatorname{atanh}(\bar{\mathbb{E}}(X_i)), \tag{7}$$

where $\bar{\mathbb{E}}$ is the expectation with respect to the distribution $\bar{P}$. Finally let $P_{\mathbf{X}_{1,2}; \{\theta, \phi_1, \phi_2\}}$ denote the Ising model on two nodes $\{1, 2\}$ with edge potential θ and node potentials ϕ_1 and ϕ_2.

Lemma 2 (Distance Bounds and Bounds on Edge Potentials). *For an Ising model, the distance bounds in (6) are related to the bounds on edge potentials as follows:*

$$d_{\min} \geq -\log |\det P_{\mathbf{X}_{1,2}; \{\theta_{\max}, \phi'_{\max}, \phi'_{\max}\}}|, \tag{8}$$

$$d_{\max} \leq -\log |\det P_{\mathbf{X}_{1,2}; \{\theta_{\min}, 0, 0\}}|. \tag{9}$$

The proof follows. We first note an observation for tree models.

Lemma 3 (Marginal Distribution on a Tree). *For an Ising model $\{\boldsymbol{\theta}, \boldsymbol{\phi}\}$ on a tree T with edge potentials $\boldsymbol{\theta} = \{\theta_{i,j}\}$ and node potentials $\boldsymbol{\phi} = \{\phi_i\}$, the marginal distribution between any two neighbors $(i, j) \in T$ is given by an Ising model $\{\theta_{i,j}, \phi'_i, \phi'_j\}$ with edge potential $\theta_{i,j}$ and node potentials ϕ'_i and ϕ'_j given by*

$$\phi'_i = \operatorname{atanh}(\mathbb{E}[X_i; T_{-j}]), \quad \phi'_j = \operatorname{atanh}(\mathbb{E}[X_j; T_{-i}]), \tag{10}$$

where T_{-j} denotes the tree with node j removed.

Proof: A more general version of this result is proven in [7, Lemma 4.1]. □

We also use the following result for attractive models ($\theta_{i,j} > 0, \ \forall \, (i, j) \in G$).

Lemma 4 (Griffith's Second Inequality [9]). *For two attractive Ising models Markov on same graph $G = (W, E)$ with potentials $0 < \theta_{i,j} \leq \theta'_{i,j}$ for all $(i, j) \in G$, we have*

$$\mathbb{E}\left[\prod_{i \in U} X_i; \boldsymbol{\theta}\right] \leq \mathbb{E}\left[\prod_{i \in U} X_i; \boldsymbol{\theta}'\right], \quad \forall U \subset W. \tag{11}$$

In particular, this means that if the potentials of a model are increased, then the marginal expectation $\mathbb{E}[X_i]$ are also increased. This implies if some of the edge potentials are set to zero (meaning we take the model on a subgraph), $\mathbb{E}[X_i]$ is reduced.

Finally we note a simple expression for information distance in a symmetric Ising model (with zero node potentials) on two nodes.

Fact 1 (Symmetric Ising Model). *For a symmetric Ising model on two nodes* $\{1,2\}$ *with edge potential* θ *and zero node potentials, we have*

$$d(1,2) := -\log|\det P_{\mathbf{X}_{1,2};\{\theta,0,0\}}| = -\log|C_{1,2}| = -\log\tanh|\theta|, \tag{12}$$

where $C_{1,2} := \mathbb{E}[X_1X_2]$ *is the correlation between the two nodes.*

Proof: For a symmetric model, we have $P(X_i = x) = 0.5$ for $i = 1,2$ and $x \in \{-1,+1\}$. Similarly $P(X_1 = +|X_2 = -) = P(X_1 = -|X_2 = +)$. Using these facts, the distance $d(1,2) = -\log|C_{1,2}|$. The correlation simplifies to

$$C_{1,2} := \frac{1}{2}\left(\mathbb{E}[X_1|X_2 = +1] - \mathbb{E}[X_1|X_2 = -\right) = \tanh\theta. \tag{13}$$

□

From the above fact, assuming Lemma 2 holds, $d_{\max}$ for Ising models is given by

$$d_{\max} \leq -\log\tanh\theta_{\min}. \tag{14}$$

Proof of Lemma 2: For an Ising model $P_{\mathbf{X}_{1,2};\{\theta,\phi_1,\phi_2\}}$ on two nodes $\{1,2\}$ with edge potential θ and node potentials ϕ_1, ϕ_2, we have

$$\exp[-d(1,2;\{\theta,\phi_1,\phi_2\})] = |\det P| = \frac{\sinh(2|\theta|)}{2\left(e^{\theta}\cosh(\phi_1+\phi_2) + e^{-\theta}\cosh(\phi_1-\phi_2)\right)^2}. \tag{15}$$

Without loss of generality, consider an attractive model ($\theta > 0$). The above function is minimized with respect to $\{\phi_1,\phi_2\}$ when $\phi_1 = \phi_2 = 0$ since $\cosh(x) \geq \cosh(0) = 1$. Similarly it is maximized with respect to $\{\phi_1,\phi_2\}$ when $\phi_1 = \phi_2 = \phi'_{\max}$ for($\theta > 0$). We subsequently establish that it is the maximum allowed node potential. When $\phi_1 = \phi_2 = 0$, we can show that $\exp[-d(1,2)]$ is increasing in θ and thus, the minimum is attained when $\theta = \theta_{\min}$, and the maximum when $\theta = \theta_{\max}$.

From Lemma 3, the marginal distribution between two neighbors on a tree model is characterized. Only the node potentials at the two nodes are altered when the marginal distribution at the two nodes is considered. The (absolute) node potential at the two nodes is dominated by the attractive counterpart and cannot exceed $\phi'_{\max}$ in (10) from Griffith's property of attractive models in Lemma 4. □

Proof of Theorem 1 in [1]: Applying Theorem 2 in [1] to Ising models, we have structural consistency when $n = \Omega(\upsilon^{-2}\log p)$, where υ is given in $(B3)$. We have $\upsilon = \min(-0.5e^{-r}(e^{d_{\min}} - 1), \exp[-0.5d_{\max}(r/d_{\min}+2)])$. When r is chosen as $r = \delta(\eta+1)d_{\max} + \epsilon$, for some $\epsilon > 0$, we have that $\upsilon = \exp[-0.5d_{\max}(r/d_{\min}+2)]$, when $e^{-d_{\max}} < 1/3$. Using the fact that $e^{-d_{\max}} = \tanh\theta_{\min}$ from (14), we have that $\tanh\theta_{\min} < 1/3$ holds when the maximum degree $\Delta_{\max} > 3$ since the model is in the regime of correlation decay $(B3)$. Since we require minimum degree of three for identifiability of hidden nodes $(B1)$, this is satisfied, and we have the result.

For the special case when all the nodes are observed ($\delta = 1$), the sample complexity can be improved by selecting the parameter $r > d_{\max} + \epsilon$ for some $\epsilon > 0$, and only building local MSTs, and considering their union. In this case the sample complexity is given by $n = \Omega(e^{2r} \log p)$ which reduces to $n = \Omega(\theta_{\min}^{-2} \log p)$. □

2. Structural Consistency of LocalCLGrouping. We first establish that the LocalCLGrouping algorithm proposed in [1] recovers the unknown latent graph correctly when statistics corresponding to the tree limit are input. In Section 2.2, we then establish that distances based on exact statistics converge locally to their tree limit. Finally, we consider sample-based analysis in Section 2.3, and use standard concentration results, along the lines of [8, Section 6], and thereby proving Theorem 1 of [1].

2.1. *Correctness of* LocalCLGrouping *under Local Tree Metric* $\mathbf{d}_{\text{tree}}(V)$. We first establish that the LocalCLGrouping algorithm outputs the correct graph when a local tree metric $\mathbf{d}_{\text{tree}}(V)$, computed using acyclic neighborhood subgraphs according to (5), is input to the algorithm. In Section 2.2, we establish local convergence of sample distances $\mathbf{d}(V)$ to $\mathbf{d}_{\text{tree}}(V)$ under correlation decay and provide perturbation bounds.

Recall that $\mathbf{d}_{\text{tree}}(V) := \{d(i, j; \text{tree}) : i, j \in V\}$ is given by

$$d(i, j; \text{tree}) := -\log |\det P_{\mathbf{X}_{i,j}|\text{tree}(i,j)}|,$$

where $P_{\mathbf{X}_{i,j}|\text{tree}(i,j)}$ denotes the distribution at nodes i and j by limiting the model to the induced subgraph $\text{tree}(i, j)$. $\text{tree}(i, j) := G(B_l(i) \cup B_j(l))$ for $l = \lfloor g/2 \rfloor - 1$ and g is the girth of the graph. Since $\text{tree}(i, j)$ has no cycles, it immediately follows that $\mathbf{d}_{\text{tree}}(V)$ is a tree metric.

FACT 2 (Local Tree Metric). *The distances* $\mathbf{d}_{\text{tree}}(V)$ *form an additive tree metric.*

2.1.1. *Recap of* CLGrouping *for Learning Latent Trees.* We first recap the result from [6, Lemma 8] that relates a latent tree model with the minimum spanning tree over the observed nodes according to a tree metric. Note that in this case, $\mathbf{d}(V)$ coincides with $\mathbf{d}_{\text{tree}}(V)$. For every node $i \in W$ in the latent tree T, define a mapping $\text{Sg} : W \mapsto V$, termed as *surrogate mapping*, as follows:

$$\text{Sg}(i; \mathbf{d}) := \arg\min_{j \in V} d(i, j; T), \quad \forall, i \in W. \tag{16}$$

Thus, observed nodes V are their own surrogates while the hidden nodes H are mapped to the closest observed node according to metric $\mathbf{d}(V)$. See Fig.1 for an example.

PROPOSITION 1 (Relating Latent Tree and MST). *Given a latent tree $T = (W, E)$, set of observed nodes $V \subset W$ and a tree metric $\mathbf{d}(V)$, the minimum spanning tree* $\text{MST}(V; \mathbf{d})$ *over the observed nodes satisfies the following properties:*

1. *The* $\text{MST}(V; \mathbf{d})$ *is obtained from the latent tree T by merging each hidden node $h \in H$ with its surrogate* $\text{Sg}(h; \mathbf{d})$ *and viceversa.*
2. *Let ξ denote the maximum graph distance between a hidden node and its surrogate in the latent tree T and let δ denote the depth of tree T. We have*

$$\xi \leq \delta \frac{d_{\max;T}}{d_{\min;T}}, \tag{17}$$

where $d_{\min;T}$ and $d_{\max;T}$ are bounds on the distance in T.

(a) Latent Tree Model (b) Minimum Spanning Tree

FIG 1. *A latent tree model over $T = (W, E)$ and the corresponding minimum spanning tree* $\mathrm{MST}(V; \mathbf{d})$ *over the observed nodes $V \subset W$. The observed nodes are shaded while the hidden nodes are unshaded. The thick lines in Fig.1a represent the edge between a hidden node and its surrogate. See Lemma 1.*

2.1.2. *Union of Local MSTs under* LocalCLGrouping. Using the results of CLGrouping, we establish properties of the union of local minimum spanning trees for girth-constrained graphs under correlation decay. To this end, consider the choice of parameter r

$$\delta\,(\eta+1)\,d_{\max} < r < \frac{g}{4} d_{\min} - r, \tag{18}$$

which is implied by assumption $(B3)$ of [1].

Also define

$$r' := \lfloor \frac{r}{d_{\max;\mathrm{tree}}} \rfloor, \quad r'' := \lceil \frac{r}{d_{\min;\mathrm{tree}}} \rceil. \tag{19}$$

Recall that $B_r(i; \mathbf{d}_{\mathrm{tree}})$ denotes the set of observed nodes within distance r according to the metric $\mathbf{d}_{\mathrm{tree}}(V)$. Let $B_{r'}(i; G)$ denote the set of nodes (including hidden nodes) within graph distance r' from node $i \in V$ on graph G. By definition, $B_{r'}(i; G) \subset B_r(i; \mathbf{d}_{\mathrm{tree}}) \subset B_{r''}(i; G)$. In other words, the nodes in $B_r(i; \mathbf{d}_{\mathrm{tree}})$ have graph distance at least r' and at most r''. We have the following result.

LEMMA 5 (Properties of Union of Local MSTs under $\mathbf{d}_{\mathrm{tree}}(V)$). *The graph formed by the union of local minimum spanning trees ($G' := \cup_{i \in V} \mathrm{MST}(B_r(i); \mathbf{d}_{\mathrm{tree}})$) under* LocalCLGrouping *method using the distance metric $\mathbf{d}_{\mathrm{tree}}(V)$, when the parameter r is chosen according to (18), satisfies the following properties:*

1. *G' does not contain triangles.*
2. *G' is formed by contracting each hidden node $h \in H$ to its surrogate node $\mathrm{Sg}(h; \mathbf{d}_{\mathrm{tree}})$ (according to the distance metric (4)).*

Proof: The first result is easy to see. We have that for each edge $(i, j) \in G'$, $d(i, j; \mathrm{tree}) \leq r$ since the MSTs are formed on nodes within distance r. By contradiction, assume that a triangle exists between nodes $i, j, k \in V$ in G'. This implies that $d(i, j; \mathrm{tree}), d(j, k; \mathrm{tree}), d(k, i; \mathrm{tree}) \leq r$. For a triangle to exist, we require another node $l \in V$ such that $d(j, l; \mathrm{tree}), d(j, k; \mathrm{tree}), d(k, l; \mathrm{tree}) \leq r$. See Fig.2. Since the maximum graph distance between any two nodes i, j satisfying $d(i, j; \mathrm{tree}) \leq r$ is r'', we have that the maximum length of the cycle containing i, j, k, l is $4r''$. When $4r'' < g$ (which holds for r according to (18)), such a cycle cannot exist and such triangles cannot occur in G'.

For the second result, from Fact 2, the distances $\mathbf{d}_{\mathrm{tree}}(B_{r''}(i; G))$ form a tree metric when $2r'' < g$, where g is the girth of the graph G, which holds for the choice of r in (18). This implies that

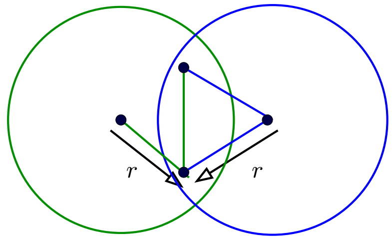

FIG 2. *Condition for existence for triangles in* ($G' := \cup_{i \in V} \mathrm{MST}(B_r(i); \mathbf{d}_{\mathrm{tree}})$).

Proposition 1 is applicable and the minimum spanning tree $\mathrm{MST}(B_r(v); \mathbf{d})$) is formed as a result of contraction of hidden nodes to their surrogates. Assumption ($B3$) in [1] implies that we require that the depth δ to satisfy

$$\frac{g}{4} d_{\min} > \delta\,(\eta + 1)\, d_{\max}. \tag{20}$$

When the parameter ξ in (17) satisfies $\xi + \delta < r'$, which is true under (20), then every hidden node has a surrogate within some local neighborhood $B_r(v)$ and forms a quartet with its surrogate node. This implies that every hidden node $h \in H$ contracts to its surrogate node in some local MST. □

Proof of Theorem 2 of [1] under $\mathbf{d}_{\mathrm{tree}}(V)$ *input:* We now show that the method LocalCLGrouping correctly recovers the graph G when tree-based distances $\mathbf{d}_{\mathrm{tree}}(V)$ are input under the assumptions of Theorem 2. From Lemma 5, we have that in the graph formed from the union of local MSTs ($G' := \cup_{v \in V} \mathrm{MST}(B_r(i); \mathbf{d}_{\mathrm{tree}})$), each hidden node is contracted to its surrogate node. The method LocalCLGrouping proceeds by reversing these contractions by considering neighborhoods on G' and constructing a local latent tree. Since there are no triangles in G', the construction of local latent trees are independent. From the correctness of CLGrouping developed in [6], the local latent trees are correct since the distance metric converges locally to a tree metric. Thus, the correctness of LocalCLGrouping under $\mathbf{d}_{\mathrm{tree}}(V)$ is proven. □

Proof of Theorem 2 of [1] with samples: Combining Lemma 8, , and Lemma 11. □

2.2. *Local Convergence to a Tree Metric.* We have so far analyzed the performance of LocalCLGrouping algorithm under tree-based distances $\mathbf{d}_{\mathrm{tree}}(V)$. We now relate the distances $\mathbf{d}(V)$ computed using exact pairwise statistics with $\mathbf{d}_{\mathrm{tree}}(V)$ under correlation decay according to (1). Let

$$d'_{\max}(l) := l d_{\max;\mathrm{tree}} - \log(1 - e^{l d_{\max}} |\mathcal{X}| \zeta(g/2 - l - 1)),$$

where $d_{\max;\mathrm{tree}}$ is the maximum $d(i, j; \mathrm{tree})$ for any two neighbors i, j on graph G according to (6).

PROPOSITION 2 (Local Convergence to a Tree Metric). *When a discrete graphical model satisfies correlation decay with rate* $\zeta(\cdot)$ *according to* (1)*, we have a.a.s., for nodes* $i, j \in W$ *with graph distance* l *in* G *and* $l < g/2 - 1$*,*

$$|\exp[-d(i, j; G)] - \exp[-d(i, j; \mathrm{tree})]| \le |\mathcal{X}| \zeta(g/2 - l - 1), \tag{21}$$

where g *is the girth of the graph, and* $|\mathcal{X}|$ *is the cardinality of the random variable at each node. Additionally, we have*

$$|d(i, j; G) - d(i, j; \mathrm{tree})| \le |\mathcal{X}| e^{d'_{\max}(l)} \zeta(g/2 - l - 1). \tag{22}$$

Proof: From the definition of correlation decay in (1), we have that

$$\|P_{\mathbf{X}_{i,j}|G} - P_{\mathbf{X}_{i,j}\,|\,\mathrm{tree}(i,j)}\|_1 \leq \zeta(g/2 - l - 1),$$

since $\mathrm{tree}(i,j;G) := G(B_{\lfloor g/2\rfloor-1}(i) \cup B_{\lfloor g/2\rfloor-1}(j))$ and $g/2 - l - 1$ is the distance from i and j to the nearest boundary.

From [4, Sec. 20], we have that for any $k \times k$ matrix A,

$$|\det(A+E) - \det(A)| \leq k \max\{\|A\|_q, \|A+E\|_q\}^{k-1}\|E\|_q. \tag{23}$$

Thus, we have that

$$|\det(P_{\mathbf{X}_{i,j}|G}) - \det(P_{\mathbf{X}_{i,j}\,|\,\mathrm{tree}(i,j)})| \leq |\mathcal{X}|\|P_{\mathbf{X}_{i,j}|G} - P_{\mathbf{X}_{i,j}\,|\,\mathrm{tree}(i,j)}\|_1 \leq |\mathcal{X}|\zeta(g/2 - l - 1).$$

From Lipschitz continuity, we have that

$$|d(i,j;G) - d(i,j;\mathrm{tree})| \leq e^{d'_{\max}(l)}|\det(P_{\mathbf{X}_{i,j}|G}) - \det(P_{\mathbf{X}_{i,j}\,|\,\mathrm{tree}(i,j)})|,$$

Let $d_{\max;G}(l)$ be the maximum $d(i,j;G)$ for any two nodes i,j at graph distance l, and similarly for $d_{\max;\mathrm{tree}}(l)$. Since no cycles are encountered in $\mathrm{tree}(i,j)$, $d(i,j;\mathrm{tree})$ is a tree metric and thus $d_{\max;\mathrm{tree}}(l) = l d_{\max;\mathrm{tree}}(1)$. For $d_{\max;G}(l)$, we note that

$$e^{-d_{\max;G}(l)} \geq e^{-l d_{\max;\mathrm{tree}}(1)} - |\mathcal{X}|\zeta(g/2 - l - 1).$$

□

Remark: When

$$e^{l d_{\max;\mathrm{tree}}}|\mathcal{X}|\zeta(g/2 - l - 1) = o(1), \tag{24}$$

then

$$|d(i,j;G) - d(i,j;\mathrm{tree})| \leq |\mathcal{X}|e^{l d_{\max;\mathrm{tree}} + o(1)}\zeta(g/2 - l - 1) = o(1), \tag{25}$$

2.3. *Sample-Based Analysis.*

2.3.1. *Concentration of Distance Estimates.*

We first derive the concentration bounds for distance estimates along the lines of from [8, 10]. Let $\widehat{\mathbf{d}}^n(V)$ be the estimated distances using n samples according to (4). We first recap the following result on empirical distribution [11, Thm. 2.1].

PROPOSITION 3 (Guarantees for General Empirical Distribution). *The following is true for the empirical distribution $\widehat{P}^n$, obtained using n i.i.d. samples from a discrete distribution P:*

$$\mathbb{P}[\|\widehat{P}^n - P\|_1 > \epsilon] \leq 2^k \exp[-n\epsilon^2/2], \tag{26}$$

where k is the dimension.

Given a graph G, let the graph distance between two nodes i and j under consideration on graph G be l. Recall that $|\mathcal{X}|$ is the dimension of the variable at each node.

LEMMA 6 (Concentration of Empirical Distances). *For empirical distance between node i and j at graph distance l, computed according to (4) using n samples, we have the following result:*

$$\mathbb{P}\left[|\exp[-\widehat{d}(i,j;G)] - \exp[-d(i,j;G)]| > \epsilon\right] \le 2^{|\mathcal{X}|} \exp\left[-\frac{n\epsilon^2}{2|\mathcal{X}|^2}\right]. \tag{27}$$

When $\epsilon > |\mathcal{X}|^2\zeta(g/2 - l - 1)$ and $l < g/2 - 1$, we additionally have that

$$\mathbb{P}\left[|\exp[-\widehat{d}(i,j;G)] - \exp[-d(i,j;\text{tree})]| > \epsilon\right] \le 2^{|\mathcal{X}|} \exp\left[-\frac{n}{2|\mathcal{X}|^2}\left(\epsilon - |\mathcal{X}|^2\zeta(\frac{g}{2} - l - 1)\right)^2\right]. \tag{28}$$

Proof: Along the lines of Proposition 2, using [4, Sec. 20], we have,

$$\mathbb{P}\left[|\det(\widehat{P}^n_{\mathbf{X}_{i,j}|G}) - \det(P_{\mathbf{X}_{i,j}|G})| > \epsilon\right] \le 2^{|\mathcal{X}|} \exp\left[-\frac{n\epsilon^2}{2|\mathcal{X}|^2}\right], \tag{29}$$

and thus (27) holds since $d(i,j;G) := -\log|\det(P_{\mathbf{X}_{i,j}|G})|$. Using Proposition 2, we also have (28). □

2.3.2. *Recap of Sample Analysis of* CLGrouping *for Learning Latent Trees.* We also recap the result [6] that the minimum spanning tree (MST) constructed over observed nodes under CLGrouping method is consistent when the underlying model is a latent tree.

Recall that $p := |V|$ is the number of observed nodes and n is the number of samples. Let η be the maximum graph distance (with respect to the latent tree T) between any two neighbors in $\text{MST}(V, \mathbf{d})$ and $d_{\min}, d_{\max}$ are distance bounds on the edges of the latent tree T.

LEMMA 7 (Consistency of MST using CLGrouping for Latent Trees). *Given a latent tree $T = (W, E)$ and observed node set $V \subset W$, the MST constructed by* CLGrouping *method using empirical distances $\widehat{\mathbf{d}}^n(V)$ does not coincide with the true MST based on exact distances $\mathbf{d}(V)$ with probability*

$$\mathbb{P}\left[\text{MST}(V;\widehat{\mathbf{d}}^n) \ne \text{MST}(V;\mathbf{d})\right] \le 2^{|\mathcal{X}|+1}p^3 \exp\left[-\frac{n}{8|\mathcal{X}|^2}e^{-2\eta d_{\max}}(1 - e^{-d_{\min}})^2\right].$$

Proof: From the property of the MST,

$$\begin{aligned}
\mathbb{P}\left[\text{MST}(V;\widehat{\mathbf{d}}^n) \ne \text{MST}(V;\mathbf{d})\right] &\overset{(a)}{=} \mathbb{P}\left[\bigcup_{\substack{(i,j)\in\text{MST}(V;\mathbf{d})\\(i,j)\in\text{Path}(u,v)}} \left(e^{-\widehat{d}(u,v)} > e^{-\widehat{d}(i,j)}\right)\right],\\
&\overset{(b)}{\le} p^3 \max_{\substack{(i,j)\in\text{MST}(V;\mathbf{d})\\(i,j)\in\text{Path}(u,v)}} \mathbb{P}\left[\epsilon_{u,v} - \epsilon_{i,j} > e^{-d(i,j)} - e^{-d(u,v)}\right]\\
&\overset{(c)}{\le} p^3 \max_{i,j,u,v\in V} \mathbb{P}\left[\epsilon_{u,v} - \epsilon_{i,j} > e^{-\eta d_{\max}}(1 - e^{-d_{\min}})\right]\\
&\overset{(d)}{\le} 2p^3 \max_{u,v\in V} \mathbb{P}\left[\epsilon_{u,v} > \frac{e^{-\eta d_{\max}}}{2}(1 - e^{-d_{\min}})\right],
\end{aligned}$$

where $\epsilon_{u,v} := \exp[-\widehat{d}(u,v)] - \exp[-d(u,v)]$ and similarly for $\epsilon_{i,j}$. Equality (a) is due to the property of the MST, inequality (b) is the union bound, inequality (c) is obtained by applying bounds on $d(i,j)$ and $d(u,v)$:

$$e^{-d(i,j)} - e^{-d(u,v)} > e^{-d(i,j)}(1 - e^{d(i,j)-d(u,v)}) > e^{-\eta d_{\max}}(1 - e^{-d_{\min}}),$$

since $d(i,j) \leq \eta d_{\max}$ for all $(i,j) \in \mathrm{MST}(V;\mathbf{d})$ and $d(i,j) - d(u,v) \leq -d_{\min}$ for all $(i,j) \in \mathrm{MST}(V;\mathbf{d})$ and $u,v \in V$ containing (i,j) on the path connecting them. Inequality (d) is obtained from the fact that $\epsilon_{u,v} - \epsilon_{i,j} \geq 2\max(\epsilon_{u,v}, \epsilon_{i,j})$ and applying the union bound. The final result is from (29) in Proposition 6. □

2.3.3. *Sample Analysis of Union of MSTs under* LocalCLGrouping. We now establish consistency under LocalCLGrouping algorithm using the above result and local convergence of the metric $\mathbf{d}(V)$ to tree-based metric $\mathbf{d}_{\text{tree}}(V)$, according to Proposition 2. Recall that $\widehat{\mathbf{d}}^n(V)$ denotes the estimates of the true distances $\mathbf{d}(V)$ according to graph G. Let $\mathbf{d}_{\text{tree}}(V)$ denote the distances by considering only acyclic neighborhood subgraphs, defined in Proposition 2. Given empirical distances $\widehat{\mathbf{d}}^n(V)$ and tree distances $\mathbf{d}_{\text{tree}}$ and parameter r according to (18), for each $i \in V$, let $\widehat{A}_i := B_r(i;\widehat{\mathbf{d}}^n)$. Define $\mathcal{L} := \mathbb{N} \cap (r/d_{\min}, g/2)$.

Lemma 8 (Union of Local MSTs under LocalCLGrouping: I). *Given a graphical model Markov on graph $G = (W,E)$ satisfying conditions of Theorem 2 in [1] with observed node set $V \subset W$, we have*

$$\begin{aligned}
&\mathbb{P}\left[\bigcup_{i\in V}\mathrm{MST}(\widehat{A}_i;\widehat{\mathbf{d}}^n) \neq \bigcup_{i\in V}\mathrm{MST}(\widehat{A}_i;\mathbf{d}_{\text{tree}})\right]\\
&\leq 2^{|\mathcal{X}|}p^3 \min_{l\in\mathcal{L}}\left(2p\exp\left[-\frac{n}{2|\mathcal{X}|^2}\left(0.5e^{-r}(e^{d_{\min}}-1) - |\mathcal{X}|^2\zeta(\frac{g}{2}-l-1)\right)^2\right]\right.\\
&\qquad\left.+\exp\left[-\frac{n}{2|\mathcal{X}|^2}\left(ld_{\min} - r - |\mathcal{X}|^2\zeta(\frac{g}{2}-l-1)\right)^2\right]\right).
\end{aligned} \tag{30}$$

Remark: In the high-dimensional regime, where $p \to \infty$, the first term dominates. Since $\zeta(\cdot)$ is monotonically decreasing, we can choose $l = \lceil r/d_{\min}\rceil + 1$. Roughly, we require $n = \Omega(e^r)$ when the other parameters are bounded, for the error probability to decay.

Proof: Along the lines of Lemma 7, for each $k \in V$, we have

$$\begin{aligned}
&\mathbb{P}\left[\mathrm{MST}(\widehat{A}_k;\widehat{\mathbf{d}}^n) \neq \mathrm{MST}(\widehat{A}_k;\mathbf{d}_{\text{tree}})\right]\\
=&\mathbb{P}\left[\bigcup_{\substack{(i,j)\in\mathrm{MST}(\widehat{A}_k;\mathbf{d}_{\text{tree}})\\(i,j)\in\mathrm{Path}(u,v)}}\left(e^{-\widehat{d}(u,v)} > e^{-\widehat{d}(i,j)}\right)\right],\\
\leq& p^3 \max_{\substack{(i,j)\in\mathrm{MST}(A_k;\mathbf{d}_{\text{tree}})\\(i,j)\in\mathrm{Path}(u,v)}} \mathbb{P}\left[\epsilon_{u,v} - \epsilon_{i,j} > e^{-d(i,j;\text{tree})} - e^{-d(u,v;\text{tree})}\right]\\
\overset{\text{(a)}}{\leq}& p^3 \max_{i,j,u,v\in V}\mathbb{P}\left[\epsilon_{u,v} - \epsilon_{i,j} > (e^{-r} - \epsilon_{i,j})(1-e^{-d_{\min}})\right]\\
\overset{\text{(b)}}{\leq}& 2p^3 \max_{u,v\in V}\mathbb{P}\left[\epsilon_{u,v} > \frac{e^{-r}}{2}(e^{d_{\min}}-1)\right],
\end{aligned}$$

where $\epsilon_{u,v} := \exp[-\widehat{d}(u,v;G)] - \exp[-d(u,v;\text{tree})]$ and similarly for $\epsilon_{i,j}$. Inequality (a) is obtained by applying bounds on $d(i,j)$ and $d(u,v)$:

$$e^{-d(i,j)} - e^{-d(u,v)} > e^{-d(i,j)}(1 - e^{d(i,j)-d(u,v)}) > (e^{-r} - \epsilon_{i,j})(1 - e^{-d_{\min}}),$$

since $\widehat{d}(i,j) \le r$ for all $(i,j) \in \text{MST}(V;\mathbf{d})$, $e^{-d(i,j)} > e^{-r} - \epsilon_{i,j}$ and $d(i,j) - d(u,v) \le -d_{\min}$ for all $(i,j) \in \text{MST}(V;\mathbf{d})$ and $u, v \in V$ containing (i,j) on the path connecting them. Inequality (b) is obtained from the fact that $\epsilon_{u,v} - e^{-d_{\min}}\epsilon_{i,j} \ge 2\max(\epsilon_{u,v}, e^{-d_{\min}}\epsilon_{i,j}) \ge 2e^{-d_{\min}}\max(\epsilon_{u,v}, \epsilon_{i,j})$ since $e^{-d_{\min}} < 1$.

Now define $\widehat{l}_{\max}$, the maximum graph distance between any two nodes in any $\widehat{A}_k$, i.e.,

$$\widehat{l}_{\max} := \max_k \left(\text{Diam}(\text{MST}(\widehat{A}_k))\right),$$

where $\text{Diam}(\cdot)$ is the diameter, in terms of graph distance on G. From (28) in Lemma 6, conditioned on $\{\widehat{l}_{\max} = l\}$ and union bound on $k \in V$,

$$\begin{aligned}&\mathbb{P}\left[\bigcup_{i\in V} \text{MST}(\widehat{A}_i; \widehat{\mathbf{d}}^n) \ne \bigcup_{i\in V} \text{MST}(\widehat{A}_i; \mathbf{d}_{\text{tree}}) \Big| \{\widehat{l}_{\max} = l\}\right] \\ &\le 2^{|\mathcal{X}|+1} p^4 \exp\left[-\frac{n}{2|\mathcal{X}|^2}\left(0.5e^{-r}(e^{d_{\min}} - 1) - |\mathcal{X}|^2\zeta(\frac{g}{2} - l - 1)\right)^2\right] \end{aligned} \tag{31}$$

We now derive characterize the event that $\{\widehat{l}_{\max} = l\}$. Note that $\widehat{l}_{\max} \le \widehat{d}'_{\max}/d_{\min}$, where

$$\widehat{d}'_{\max} := \max_{k\in V} \max_{i,j\in\widehat{A}_k} \widehat{d}(i,j). \tag{32}$$

Thus, we have

$$\begin{aligned}\mathbb{P}\left[\widehat{l}_{\max} > l\right] &\le \mathbb{P}\left[\bigcup_{\substack{k\in V\\ i,j\in\widehat{A}_k}} \widehat{d}(i,j) > l d_{\min}\right] \\ &\le 2^{|\mathcal{X}|} p^3 \exp\left[-\frac{n}{2|\mathcal{X}|^2}\left(l d_{\min} - r - |\mathcal{X}|^2\zeta(\frac{g}{2} - l - 1)\right)^2\right].\end{aligned}$$

□

We now provide conditions when $\bigcup_{i\in V} \text{MST}(\widehat{A}_i; \mathbf{d}_{\text{tree}})$ coincides with $\bigcup_{i\in V} \text{MST}(A_i; \mathbf{d}_{\text{tree}})$, where $A_i := B_r(i; \mathbf{d}_{\text{tree}})$.

Lemma 9 (Union of Local MSTs under LocalCLGrouping: II). *Given a graphical model Markov on graph $G = (W, E)$ satisfying conditions of Theorem 2 in [1] with observed node set $V \subset W$, we have*

$$\begin{aligned}&\mathbb{P}\left[\bigcup_{i\in V} \text{MST}(\widehat{A}_i; \mathbf{d}_{\text{tree}}) \ne \bigcup_{i\in V} \text{MST}(A_i; \mathbf{d}_{\text{tree}})\right] \\ &\le 2^{|\mathcal{X}|} p^3 \left(\exp\left[-\frac{n}{2|\mathcal{X}|^2}\left(\frac{g d_{\min}}{4} - r - |\mathcal{X}|^2\zeta(\frac{g}{2} - r/d_{\min} - 1)\right)^2\right]\right. \\ &\quad \left. + \exp\left[-\frac{n}{2|\mathcal{X}|^2}\left(r - \delta(\frac{d_{\max}}{d_{\min}} + 1)d_{\max} - |\mathcal{X}|^2\zeta(\frac{g}{2} - r/d_{\min} - 1)\right)^2\right]\right).\end{aligned}$$

Proof: Define

$$\widehat{l}_{\max} := \max_k \left(\mathrm{Diam}(\mathrm{MST}(\widehat{A}_k)) \right), \tag{33}$$

$$\widehat{l}_{\min} := \min_k \left(\mathrm{Diam}(\mathrm{MST}(\widehat{A}_k)) \right). \tag{34}$$

where $\mathrm{Diam}(\cdot)$ is the diameter, in terms of graph distance on G. Conditioned on the event $\{\widehat{l}_{\max} < \frac{g}{4}\} \cap \{\widehat{l}_{\min} > \xi + \delta\}$, the graph satisfies the properties listed in Lemma 5 and thus,

$$\mathbb{P}\left[\bigcup_{i \in V} \mathrm{MST}(\widehat{A}_i; \mathbf{d}_{\mathrm{tree}}) \neq \bigcup_{i \in V} \mathrm{MST}(A_i; \mathbf{d}_{\mathrm{tree}}) \Big| \{\widehat{l}_{\max} < \frac{g}{4}\} \cap \{\widehat{l}_{\min} > \xi + \delta\} \right] = 0. \tag{35}$$

Moreover,

$$\begin{aligned}
&\mathbb{P}\left[\{\widehat{l}_{\max} > \frac{g}{4}\} \cup \{\widehat{l}_{\min} < \xi + \delta\} \right] \\
&\leq 2^{|\mathcal{X}|} p^3 \left(\exp\left[-\frac{n}{2|\mathcal{X}|^2} \left(\frac{g d_{\min}}{4} - r - |\mathcal{X}|^2 \zeta(\frac{g}{2} - r/d_{\min} - 1) \right)^2 \right] \right. \\
&\left. + \exp\left[-\frac{n}{2|\mathcal{X}|^2} \left(r - (\xi + \delta) d_{\max} - |\mathcal{X}|^2 \zeta(\frac{g}{2} - r/d_{\min} - 1) \right)^2 \right] \right),
\end{aligned}$$

where $\xi := \delta d_{\max}/d_{\min}$ is the worst-case graph distance between a hidden node and its surrogate in G with respect to metric $\mathbf{d}_{\mathrm{tree}}$. This is because the worst-case distance in a quartet containing a hidden node and its surrogate is $(\xi + \delta) d_{\max}$. When the empirical version of this distance exceeds r, then we have a bad event. $\square$

2.3.4. *Analysis of the Recursive Grouping.* Recall that for each $i \in V$, let $\widehat{A}_i := B_r(i; \widehat{\mathbf{d}}^n)$ and $A_i := B_r(i; \mathbf{d}_{\mathrm{tree}})$. In $\mathsf{LocalCLGrouping}$, the recursive grouping procedure is run on subsets of nodes in each $\widehat{A}_i$. We first analyze the performance of quartet test.

Lemma 10 (Analysis of Quartet Test). *Given distance estimates $\widehat{\mathbf{d}}^n(\widehat{A}_i)$ over observed nodes in $\widehat{A}_i$, for each $i \in V$, $\mathsf{Quartet}(\widehat{\mathbf{d}}^n(\widehat{A}_i), \Lambda)$ returns the correct set of quartets (and no null results) with probability at least*

$$\begin{aligned}
&\mathbb{P}[\cup_{i \in V} \{\mathsf{Quartet}(\widehat{\mathbf{d}}^n(\widehat{A}_i), \Lambda) \neq \mathsf{Quartet}(\mathbf{d}_{\mathrm{tree}}(\widehat{A}_i), \Lambda)\}] \\
\leq &2^{|\mathcal{X}|} p^3 \left(p \exp\left[-\frac{n}{2|\mathcal{X}|^2} \left(\exp[-(r/d_{\min} + 2) d_{\max}/2] - |\mathcal{X}|^2 \zeta(\frac{g}{2} - r/d_{\min} - 1) \right)^2 \right] \right. \\
&\left. + \exp\left[-\frac{n}{2|\mathcal{X}|^2} \left(d_{\min} - |\mathcal{X}|^2 \zeta(\frac{g}{2} - r/d_{\min} - 1) \right)^2 \right] \right),
\end{aligned} \tag{36}$$

when Λ is chosen as

$$\Lambda = \exp[-(r/d_{\min} + 2) d_{\max}/2]. \tag{37}$$

Proof: For each quartet $Q(v_1 v_2 | v_3 v_4)$ under metric $\mathbf{d}_{\mathrm{tree}}(V)$ and $\mathcal{A} := \bigcup_{i=1}^4 v_i$, we have that

$$\mathbb{P}[\mathsf{Quartet}(\widehat{\mathbf{d}}^n(\mathcal{A}), \Lambda) \neq \mathsf{Quartet}(\mathbf{d}_{\mathrm{tree}}(\mathcal{A}), \Lambda) \Big| \bigcap_{a,b \in \mathcal{A}} \{|\widehat{d}^n(a,b) - d(a,b;\mathrm{tree})| < \Lambda\}] = 0, \tag{38}$$

and the test $\mathsf{Quartet}(\widehat{\mathbf{d}}^n(\mathcal{A}),\Lambda)$ does not return null when $\Lambda < \exp[-\max_{a,b\in\mathcal{A}} d(a,b;\text{tree})/2]$. Considering all sets $\widehat{A}_i$ for $i \in V$, we require $\Lambda < \exp[-\widehat{l}_{\max} d_{\max}/2]$ to not return null, where

$$\widehat{l}_{\max} := \max_k \left(\text{Diam}(\text{MST}(\widehat{A}_k))\right).$$

From Lemma 6, choosing $\Lambda = \exp[-(l+1)d_{\max}/2]$ we that

$$\begin{aligned}&\mathbb{P}\left[\cup_{i\in V}\{\mathsf{Quartet}(\widehat{\mathbf{d}}^n(\widehat{A}_i),\Lambda) \neq \mathsf{Quartet}(\mathbf{d}_{\text{tree}}(\widehat{A}_i),\Lambda)\middle|\{\widehat{l}_{\max} < l\}\}\right]\\ &\leq 2^{|\mathcal{X}|}p^4 \exp\left[-\frac{n}{2|\mathcal{X}|^2}\left(\exp[-(l+1)d_{\max}/2] - |\mathcal{X}|^2\zeta(\frac{g}{2}-l-1)\right)^2\right].\end{aligned}$$

Along the lines of analysis in Lemma 9, we have that

$$\mathbb{P}\left[\{\widehat{l}_{\max} > l\}\right] \leq 2^{|\mathcal{X}|}p^3 \exp\left[-\frac{n}{2|\mathcal{X}|^2}\left(l d_{\min} - r - |\mathcal{X}|^2\zeta(\frac{g}{2}-l-1)\right)^2\right].$$

Choosing l as $r/d_{\min}+1$, we have the result. □

This yields the following result on recursive grouping.

Lemma 11 (Results for Recursive Grouping). *The recursive grouping method* $\mathsf{RG}(\widehat{\mathbf{d}}(\widehat{A}_i),\Lambda,\tau)$ *returns the same tree as* $\mathsf{RG}(\mathbf{d}_{\text{tree}}(\widehat{A}_i),\Lambda,\tau)$ *with probability*

$$\begin{aligned}&\mathbb{P}[\cup_{i\in V}\{\mathsf{RG}(\widehat{\mathbf{d}}^n(\widehat{A}_i),\Lambda,\tau) \neq \mathsf{RG}(\mathbf{d}_{\text{tree}}(\widehat{A}_i),\Lambda,\tau)\}]\\ \leq& 2^{|\mathcal{X}|}p^3\left(p\exp\left[-\frac{n}{2|\mathcal{X}|^2}\left(\exp\left[-\frac{(\frac{r}{d_{\min}}+2)d_{\max}}{2}\right] - |\mathcal{X}|^2\zeta(\frac{g}{2}-\frac{r}{d_{\min}}-1)\right)^2\right]\right.\\ &+\exp\left[-\frac{n}{2|\mathcal{X}|^2}\left(d_{\min} - |\mathcal{X}|^2\zeta(\frac{g}{2}-\frac{r}{d_{\min}}-1)\right)^2\right]\\ &\left.+\exp\left[-\frac{n}{2|\mathcal{X}|^2}\left(\frac{d_{\min}}{2} - |\mathcal{X}|^2\zeta(\frac{g}{2}-1)\right)^2\right]\right),\end{aligned}\tag{39}$$

when Λ *is chosen as*

$$\Lambda = \exp[-(r/d_{\min}+2)d_{\max}/2].\tag{40}$$

and τ *is chosen as*

$$\tau = \frac{d_{\min}}{2} - |\mathcal{X}|^2\zeta(g/2-1).\tag{41}$$

Proof: Along the lines of analysis in [2], given the correct set of quartets, the recursive grouping procedure returns the correct tree structure when the nodes are merged correctly with threshold τ. It is easy to see that this happens with probability

$$2^{|\mathcal{X}|}p^2\exp\left[-\frac{n}{2|\mathcal{X}|^2}(d_{\min}/2 - |\mathcal{X}|^2\zeta(g/2-1))^2\right],$$

when the threshold is chosen as (41). □

Proof of Theorem 2 in [1]: From Lemma 8, Lemma 9 and Lemma 11. □

2.4. *Analysis Under Uniform Sampling.* *Proof of Lemma 1 in [1]:* Let $\mathcal{A}(e)$ denote the event that an hidden edge (with at least one hidden end point) has a representative quartet in which the end points are at most graph distance $l < g/2 - 1$. We have that

$$\mathbb{P}[\mathcal{A}^c(e)] \leq 4(1-\rho)^{(\Delta_{\min}-1)^{l-1}},$$

since there are at least $(\Delta_{\min} - 1)^{l-1}$ nodes in each of the four subtrees from which four observed nodes can be sampled and $\rho := p/m$ is the sampling probability. Taking the union bound, we have the probability that the depth δ is greater than $l < g/2 - 1$ as

$$\mathbb{P}[\delta > l] \leq 4m\Delta_{\max}(1-\rho)^{(\Delta_{\min}-1)^{l}}.$$

Thus, the result holds. □

3. Necessary Conditions for Graph Reconstruction. *Proof of Theorem 3 in [1]:* The proof is based on counting arguments along the lines of [5, Thm. 1]. For any deterministic estimator $\widehat{G}_m$, let $\mathcal{R} := \widehat{G}_m((\mathcal{X}^{m^\beta})^n)$ as the range of the estimator $\widehat{G}_m$, when the number of observed nodes is $|V| = m^\beta$ for $\beta \in (0, 1]$. Thus, we have $|\mathcal{R}| = |\mathcal{X}|^{nm^\beta}$.

For any fixed graph F_m and set of labeled nodes V, denote the set of graphs within graph distance ϵm as

$$\mathcal{D}(F_m; \epsilon m) := \{G_m : \mathrm{dist}(F_m, G_m; V) \leq \epsilon m\}.$$

We note that

$$|\mathcal{D}(F_m; \epsilon m)| \leq m! \binom{m^2}{\delta m} \leq m^{(2\epsilon+1)m} 3^{\epsilon m},$$

since we can permute the m vertices and change at most ϵm entries in the adjacency matrix $\mathbf{A}_F$ and we use the bound that $\binom{N}{k} \leq \frac{N^k}{k!} \leq N^k 3^k$.

Let $\mathcal{S}(\widehat{G}_m; \epsilon m)$ denote all the graphs which are within edit distance of ϵm of the graphs in range $\mathcal{R}$. We have that

$$|\mathcal{S}(\widehat{G}_m; \epsilon m)| \leq |\mathcal{X}|^{nm^\beta} m^{(2\epsilon+1)m} 3^{\epsilon m}.$$

Along the lines of [5, Thm. 1], we have that the probability of error should satisfy

$$\mathbb{P}[\mathrm{dist}(\widehat{G}_m, G_m; V) > \epsilon m] \geq 1 - \frac{|\mathcal{S}(\widehat{G}_m; \epsilon m)|}{|\mathcal{G}(m)|},$$

where $|\mathcal{G}(m)|$ is the number of graphs in the family under consideration.

From [3, Lemma 2], we have that for girth-constrained ensembles with girth g, minimum degree $\Delta_{\min}$, maximum degree $\Delta_{\max}$ and number of edges k, we have

$$m^k(m - g\Delta_{\max}^g)^k \leq |\mathcal{G}(m)| \leq m^k(m - \Delta_{\min}^g)^k, \tag{42}$$

and we have the result. □

References.

[1] Anandkumar, A. and Valluvan, R. (2012). Learning Loopy Graphical Models with Latent Variables: Efficient Methods and Guarantees. *ArXiv:1203.3887*.

[2] Anandkumar, A., Chaudhuri, K., Hsu, D., Kakade, S. M., Song, L. and Zhang, T. (2011). Spectral Methods for Learning Multivariate Latent Tree Structure. *Preprint, ArXiv 1107.1283*.

[3] Anandkumar, A., Tan, V. Y. F., Huang, F. and Willsky, A. S. (2012). High-dimensional structure learning of Ising models: local separation criterion. *The Annals of Statistics* **40** 1346–1375.

[4] Bhatia, R. (2007). *Perturbation Bounds for Matrix Eigenvalues (Classics in Applied Mathematics)*. Society for Industrial and Applied Mathematics.

[5] Bresler, G., Mossel, E. and Sly, A. (2008). Reconstruction of Markov random fields from samples: some observations and algorithms. In *Intl. workshop APPROX Approximation, Randomization and Combinatorial Optimization* 343–356. Springer.

[6] Choi, M. J., Tan, V. Y. F., Anandkumar, A. and Willsky, A. (2011). Learning latent tree graphical models. *J. of Machine Learning Research* **12** 1771–1812.

[7] Dembo, A. and Montanari, A. (2010). Ising Models on Locally Tree-like Graphs. *Annals of Applied Probability.*

[8] Erdös, P. L., Steel, M. A., Székely, L. A. and Warnow, T. J. (1999). A few logs suffice to build (almost) all trees: part II. *Theoretical Computer Science* **221** 77–118.

[9] Griffiths, R. B. (1967). Correlations in Ising ferromagnets. III. *Communications in Mathematical Physics* **6** 121–127.

[10] Mossel, E. (2007). Distorted metrics on trees and phylogenetic forests. *IEEE/ACM Transactions on Computational Biology and Bioinformatics* 108–116.

[11] Weissman, T., Ordentlich, E., Seroussi, G., Verdu, S. and Weinberger, M. L. (2003). Inequalities for the l_1 deviation of the empirical distribution Technical Report, Hewlett-Packard Labs.