A Proofs

A.1 Proof of Theorem 1 (Sufficient number of samples)

Lemmas 1, 2, 3 together with Lemma 4 give the required number of samples for exact structure recovery when observations from a hidden model are given. To analyze the error event we use the "Two Trees Lemma" of Bresler and Karzand [75, Lemmas 10.1, 10.2]. For two different spanning trees on same set of nodes. Informally, if two maximum spanning trees T, T′ have a pair of nodes connected in a different way then there exist at least one edge in $\mathcal{E}_{\mathrm{T}}$ which does not exist in $\mathcal{E}_{\mathrm{T}'}$ and vice versa.

Lemma 1. *Let $f = (w, \bar{w})$ be an edge such that $f \in \mathrm{T}$ and $f \notin \mathrm{T}_{\dagger}^{CL}$. Then there exists an edge $g \in \mathrm{T}_{\dagger}^{CL}$ and $g \notin \mathrm{T}$ such that $f \in \mathrm{path}_{\mathrm{T}}(u, \bar{u})$ and $g \in \mathrm{path}_{\mathrm{T}_{\dagger}^{CL}}(w, \bar{w})$ and the following holds under the error event $\mathrm{T} \neq \mathrm{T}_{\dagger}^{CL}$:*

$$\left(\sum_{i=1}^{n_\dagger} Z_{f,u,\bar{u}}^{(i)}\right)\left(\sum_{i=1}^{n_\dagger} M_{f,u,\bar{u}}^{(i)}\right) < 0, \tag{35}$$

where $Z_{f,u,\bar{u}} = Y_w Y_{\bar{w}} - Y_u Y_{\bar{u}}$ and $M_{f,u,\bar{u}} = Y_w Y_{\bar{w}} + Y_u Y_{\bar{u}}$.

Proof. Using the same argument as the noiseless case [75, Lemmas 9.6, 9.7] we see that $\left|\hat{\mu}_f^\dagger\right| \le \left|\hat{\mu}_g^\dagger\right|$ implies

$$\begin{aligned}
0 &\ge \left|\hat{\mu}_f^\dagger\right|^2 - \left|\hat{\mu}_g^\dagger\right|^2 \\
&= \left(\hat{\mu}_f^\dagger - \hat{\mu}_g^\dagger\right)\left(\hat{\mu}_f^\dagger + \hat{\mu}_g^\dagger\right) \\
&= \frac{1}{n_\dagger^2}\left(\sum_{i=1}^{n_\dagger} N_w^{(i)} X_w^{(i)} N_{\bar{w}}^{(i)} X_{\bar{w}}^{(i)} - N_u^{(i)} X_u^{(i)} N_{\bar{u}}^{(i)} X_{\bar{u}}^{(i)}\right) \\
&\quad \times \left(\sum_{i=1}^{n_\dagger} N_w^{(i)} X_w^{(i)} N_{\bar{w}}^{(i)} X_{\bar{w}}^{(i)} + N_w^{(i)} X_u^{(i)} N_{\bar{u}}^{(i)} X_{\bar{u}}^{(i)}\right) \\
&= \frac{1}{n_\dagger^2}\left(\sum_{i=1}^{n_\dagger} Z_{f,u,\bar{u}}^{(i)}\right)\left(\sum_{i=1}^{n_\dagger} M_{f,u,\bar{u}}^{(i)}\right).
\end{aligned} \tag{36}$$

□

Notice that the random variables $Z_{f,u,\bar{u}}^{(i)}$, $M_{f,u,\bar{u}}^{(i)}$ are functions of observations of the observable variables (noisy observations). These differ from the corresponding terms in the noiseless case and require a new analysis.

In Lemmas 2, 3, we derive two concentration of measure inequalities for the variables $Z_{f,u,\bar{u}}^{(i)}$ and $M_{f,u,\bar{u}}^{(i)}$. In fact, we have that the event E_Z in (28) as

$$E_Z \triangleq \bigcap_{(w,\bar{w})\in\mathcal{E},u,\bar{u}\in\mathcal{V}} E_Z^{(w,\bar{w}),u,\bar{u}}, \tag{37}$$

and

$$E_Z^{(w,\bar{w}),u,\bar{u}} \triangleq \left\{\left|\frac{1}{n_\dagger}\sum_{i=1}^{n_\dagger} Z_{e,u,\bar{u}}^{(i)} - \mathbb{E}\left[Z_{e,u,\bar{u}}\right]\right| \le \max\left\{8\epsilon_\dagger^2, 4\epsilon_\dagger\sqrt{1-\mu_A^\dagger}\right\}\right\}, \tag{38}$$

happens with probability at least $1 - \frac{\delta'}{2}$ and the event E_M, which is defined as

$$E_M \triangleq \bigcap_{(w,\bar{w})\in\mathcal{E},u,\bar{u}\in\mathcal{V}} E_M^{(w,\bar{w}),u,\bar{u}}, \tag{39}$$

and

$$E_M^{(w,\bar{w}),u,\bar{u}} \triangleq \left\{\left|\frac{1}{n_\dagger}\sum_{i=1}^{n_\dagger} M_{e,u,\bar{u}}^{(i)} - \mathbb{E}\left[M_{e,u,\bar{u}}\right]\right| \le \max\left\{8\epsilon_\dagger^2, 4\epsilon_\dagger\sqrt{1+\mu_A^\dagger}\right\}\right\}. \tag{40}$$

happens with probability at least $1 - \frac{\delta''}{2}$. The threshold variable $\epsilon_\dagger$ is a decreasing function of $n_\dagger$, both $\epsilon_\dagger$, μ_A, which are defined below. Finally, we apply union bound to guarantee that the event $E_Z \cup E_M$ happens with probability at least $1-\delta$, where $\frac{\delta'}{2} + \frac{\delta''}{2} \le 2\max\{\frac{\delta'}{2}, \frac{\delta''}{2}\} \triangleq \delta$. Then, we can apply the union bound over all pairs $w, \bar{w}, u, \bar{u}$ in Lemmas 2 and 3 and finally for the events E_Z and E_M.

Lemma 2. *For all pairs of vertices $u, \bar{u} \in V$ and edges $e = (w, \bar{w})$ in the path $\mathrm{path}_{\mathrm{T}}(u, \bar{u})$ from u to $\bar{u}$, given $n_\dagger$ samples $Z_{e,u,\bar{u}}^{(1)}, Z_{e,u,\bar{u}}^{(2)}, ..., Z_{e,u,\bar{u}}^{(n)}$ of $Z_{e,u,\bar{u}} = Y_w Y_{\bar{w}} - Y_u Y_{\bar{u}}$ we have*

$$\mathbb{P}\left(\left|\sum_{i=1}^{n_\dagger} Z_{e,u,\bar{u}}^{(i)} - n_\dagger \mathbb{E}\left[Z_{e,u,\bar{u}}\right]\right| \le n_\dagger \max\left\{8\epsilon_\dagger^2, 4\epsilon_\dagger\sqrt{1-\mu_A^\dagger}\right\}\right) \ge 1 - \frac{\delta}{2},$$

where $\epsilon_\dagger = \sqrt{2/n_\dagger \log\left(2p^2/\delta\right)}$ and $A = \mathrm{path}_{\mathrm{T}}(u, \bar{u}) \setminus \{e\}$.

Proof. The proof is based on Bernstein's inequality [78]. Expanding the definition of $Z_{e,u,\bar{u}}$,

$$\begin{aligned}
Z_{e,u,\bar{u}} &= X_w N_w X_{\bar{w}} N_{\bar{w}} - N_u X_u N_{\bar{u}} X_{\bar{u}} && (41) \\
&= N_w X_w N_{\bar{w}} X_{\bar{w}} \\
&\quad \times \left(1 - N_w X_w N_{\bar{w}} X_{\bar{w}} N_u X_u N_{\bar{u}} X_{\bar{u}}\right). && (42)
\end{aligned}$$

Then

$$\begin{aligned}\mathbb{E}\left[Z_{e,u,\bar{u}}\right] &= (1-2q)^2\,\mathbb{E}\left[X_w X_{\bar{w}} - X_u X_{\bar{u}}\right]\\ &= (1-2q)^2\,\mu_e\left(1-\mu_A\right)\end{aligned} \tag{43}$$

and

$$\begin{aligned}&\mathrm{Var}\left(Z_{e,u,\bar{u}}\right)\\ &= \mathbb{E}\left[\left(Z_{e,u,\bar{u}}\right)^2\right] - \mathbb{E}\left[\left(Z_{e,u,\bar{u}}\right)\right]^2\\ &= \mathbb{E}\left[\left(X_w N_w X_{\bar{w}} N_{\bar{w}} - N_{\bar{u}} X_u N_{\bar{u}} X_{\bar{u}}\right)^2\right]\\ &\qquad - \left[(1-2q)^2\,\mathbb{E}\left[X_w X_{\bar{w}} - X_u X_{\bar{u}}\right]\right]^2\\ &= \mathbb{E}\left[1+1-2X_w N_w X_{\bar{w}} N_{\bar{w}} N_u X_u N_{\bar{u}} X_{\bar{u}}\right]\\ &\qquad - (1-2q)^4\,\mathbb{E}\left[X_w X_{\bar{w}} - X_u X_{\bar{u}}\right]^2\\ &= 2 - 2\mathbb{E}\left[X_w N_w X_{\bar{w}} N_{\bar{w}} N_u X_u N_{\bar{u}} X_{\bar{u}}\right]\\ &\qquad - (1-2q)^4\,\mathbb{E}\left[X_w X_{\bar{w}} - X_u X_{\bar{u}}\right]^2\\ &= 2 - 2\left(1-2q\right)^4\mathbb{E}\left[X_w X_{\bar{w}} X_u X_{\bar{u}}\right]\\ &\qquad - (1-2q)^4\,\mathbb{E}\left[X_w X_{\bar{w}} - X_u X_{\bar{u}}\right]^2\\ &= 2 - 2\left(1-2q\right)^4\mu_A - (1-2q)^4\left(\mu_e\left(1-\mu_A\right)\right)^2\\ &= 2 - (1-2q)^4\left[2\mu_A + \mu_e^2\left(1-\mu_A\right)^2\right].\end{aligned} \tag{44}$$

Using the expressions for the mean and the variance, we apply Bernstein's inequality [78] for the noisy setting: for all $i \in [n_\dagger]$ we have $\left|Z_{e,u,\bar{u}}^{(i)} - \mathbb{E}\left[Z_{e,u,\bar{u}}\right]\right| \leq M$ almost surely, that is, for any $t > 0$

$$\begin{aligned}&\mathbb{P}\left[\left|\sum_{i=1}^{n_\dagger} Z_{e,u,\bar{u}}^{(i)} - n_\dagger \mathbb{E}\left[Z_{e,u,\bar{u}}\right]\right| \geq t\right]\\ &\quad \leq 2\exp\left(-\frac{t^2}{2n_\dagger\,\mathrm{Var}\left(Z_{e,u,\bar{u}}\right) + \frac{2}{3}Mt}\right)\end{aligned} \tag{45}$$

$$\begin{aligned}&\mathbb{P}\left[\left|\sum_{i=1}^{n_\dagger} Z_{e,u,\bar{u}}^{(i)} - n_\dagger \mathbb{E}\left[Z_{e,u,\bar{u}}\right]\right| \leq t\right]\\ &\quad \geq 1 - 2\exp\left(-\frac{t^2}{2n_\dagger\,\mathrm{Var}\left(Z_{e,u,\bar{u}}\right) + \frac{2}{3}Mt}\right).\end{aligned} \tag{46}$$

Set

$$\delta/2 = 2\exp\left(-\frac{t^2}{2n_\dagger\,\mathrm{Var}\left(Z_{e,u,\bar{u}}\right) + \frac{2}{3}Mt}\right), \tag{47}$$

then

$$\log\frac{4}{\delta} = \frac{t^2}{2n_\dagger\,\mathrm{Var}\left(Z_{e,u,\bar{u}}\right) + \frac{2}{3}Mt}. \tag{48}$$

By solving with respect to t, we have

$$\begin{aligned}t_{1,2} &= \frac{\frac{2}{3}M\log\frac{4}{\delta}}{2}\\ &\qquad \frac{\pm\sqrt{\left(\frac{2}{3}M\log\frac{4}{\delta}\right)^2 + 8n_\dagger\,\mathrm{Var}\left(Z_{e,u,\bar{u}}\right)\log\frac{4}{\delta}}}{2}\\ &= \frac{1}{3}M\log\frac{4}{\delta}\\ &\qquad \pm\sqrt{\left(\frac{1}{3}M\log\frac{4}{\delta}\right)^2 + 2n_\dagger\,\mathrm{Var}\left(Z_{e,u,\bar{u}}\right)\log\frac{4}{\delta}}.\end{aligned} \tag{49}$$

Since $t > 0$, we have

$$\begin{aligned}t &= \frac{1}{3}M\log\frac{4}{\delta}\\ &\qquad + \sqrt{\left(\frac{1}{3}M\log\frac{4}{\delta}\right)^2 + 2n_\dagger\,\mathrm{Var}\left(Z_{e,u,\bar{u}}\right)\log\frac{4}{\delta}}.\end{aligned} \tag{50}$$

Since $M = 4$,

$$\begin{aligned}t &= \frac{4}{3}\log\frac{4}{\delta}\\ &\qquad + \sqrt{\left(\frac{4}{3}\log\frac{4}{\delta}\right)^2 + 2n_\dagger\,\mathrm{Var}\left(Z_{e,u,\bar{u}}\right)\log\frac{4}{\delta}}.\end{aligned} \tag{51}$$

This makes the probability of the union of events

$$\bigcup_{u,\bar{u},w,\bar{w}:(w,\bar{w})\in \mathrm{path}_{\mathrm{T}}(u,\bar{u})}\left\{\left|\sum_{i=1}^{n_\dagger} Z_{e,u,\bar{u}}^{(i)} - n_\dagger \mathbb{E}\left[Z_{e,u,\bar{u}}\right]\right| \geq t\right\}$$

to be at most $\frac{\delta}{2p^3}$, then the union bound gives probability at most $\frac{\delta}{2}$. Also,

$$\begin{aligned}\mathrm{Var}\left(Z_{e,u,\bar{u}}\right) &= 2 - (1-2q)^4\left[2\mu_A + \mu_e^2\left(1-\mu_A\right)^2\right]\\ &= 2 - (1-2q)^4\,2\mu_A\\ &\qquad - (1-2q)^4\,\mu_e^2\left(1-\mu_A\right)^2\\ &\leq 2 - (1-2q)^4\,2\mu_A + 0\\ &= 2\left(1-(1-2q)^4\mu_A\right)\\ &= 2\left(1-\mu_A^\dagger\right).\end{aligned} \tag{52}$$

From (50) and (52)

$$t = \frac{4}{3}\log\frac{4p^3}{\delta} + \sqrt{\left(\frac{4}{3}\log\frac{4p^3}{\delta}\right)^2 + 4n_\dagger\left(1-\mu_A^\dagger\right)\log\frac{4p^3}{\delta}}$$
$$\leq \frac{8}{3}\log\frac{4p^3}{\delta} + \sqrt{4n_\dagger\left(1-\mu_A^\dagger\right)\log\frac{4p^3}{\delta}}, \text{ and}$$
$$t = n_\dagger\left(\frac{4}{3n_\dagger}\log\frac{4p^3}{\delta} + \sqrt{\left(\frac{4}{3n_\dagger}\log\frac{4p^3}{\delta}\right)^2 + \frac{4}{n_\dagger}\left(1-\mu_A^\dagger\right)\log\frac{4p^3}{\delta}}\right)$$
$$\leq n_\dagger\left(\frac{8}{3n_\dagger}\log\frac{4p^3}{\delta} + \sqrt{\frac{4}{n_\dagger}\left(1-\mu_A^\dagger\right)\log\frac{4p^3}{\delta}}\right). \quad (53)$$

Define $\epsilon_\dagger = \sqrt{\log(2p^2/\delta)\,2/n_\dagger}$ (as following the definiton by [28]), then we have

$$t \leq n_\dagger\left(4\epsilon_\dagger^2 + 2\epsilon_\dagger\sqrt{1-\mu_A^\dagger}\right) \leq n_\dagger \max\left\{8\epsilon_\dagger^2, 4\epsilon_\dagger\sqrt{1-\mu_A^\dagger}\right\}. \quad (54)$$

This completes the proof. □

Lemma 3 gives the concentration of measure bound for the event E_M defined in (39).

Lemma 3. *For all pairs of vertices $u,\bar{u} \in V$ and edges $e=(w,\bar{w})$ in the path $\text{path}_\text{T}(u,\bar{u})$ from u to $\bar{u}$, given $n_\dagger$ samples $M_{e,u,\bar{u}}^{(1)}, M_{e,u,\bar{u}}^{(2)}, ..., M_{e,u,\bar{u}}^{(n)}$ of $M_{e,u,\bar{u}} = Y_wY_{\bar{w}} + Y_uY_{\bar{u}}$, we have*

$$\mathbb{P}\left(\left|\sum_{i=1}^{n_\dagger} M_{e,u,\bar{u}}^{(i)} - n_\dagger\mathbb{E}\left[M_{e,u,\bar{u}}\right]\right| \leq n_\dagger\max\left\{8\epsilon_\dagger^2, 4\epsilon_\dagger\sqrt{1+\mu_A^\dagger}\right\}\right) \geq 1-\frac{\delta}{2},$$

$$\epsilon_\dagger \triangleq \sqrt{2/n_\dagger \log\left(2p^2/\delta\right)} \quad (55)$$

and

$$A \triangleq \text{path}_\text{T}(u,\bar{u}) \setminus \{e\}. \quad (56)$$

.

Proof.

$$\mathbb{E}\left[M_{e,u,\bar{u}}\right] = (1-2q)^2\,\mathbb{E}\left[X_wX_{\bar{w}} + X_uX_{\bar{u}}\right] = (1-2q)^2\,\mu_e\left(1+\mu_A\right). \quad (57)$$

$$\begin{aligned}
&\text{Var}\left(M_{e,u,\bar{u}}\right)\\
&= \mathbb{E}\left[\left(M_{e,u,\bar{u}}\right)^2\right] - \mathbb{E}\left[\left(M_{e,u,\bar{u}}\right)\right]^2\\
&= \mathbb{E}\left[\left(X_wN_wX_{\bar{w}}N_{\bar{w}} + N_uX_uN_{\bar{u}}X_{\bar{u}}\right)^2\right] - \left[(1-2q)^2\,\mathbb{E}\left[X_wX_{\bar{w}} + X_uX_{\bar{u}}\right]\right]^2\\
&= \mathbb{E}\left[1+1+2X_wN_wX_{\bar{w}}N_{\bar{w}}N_uX_uN_{\bar{u}}X_{\bar{u}}\right] - (1-2q)^4\,\mathbb{E}\left[X_wX_{\bar{w}} + X_uX_{\bar{u}}\right]^2\\
&= 2 + 2\mathbb{E}\left[X_wN_wX_{\bar{w}}N_{\bar{w}}N_uX_uN_{\bar{u}}X_{\bar{u}}\right] - (1-2q)^4\,\mathbb{E}\left[X_wX_{\bar{w}} + X_uX_{\bar{u}}\right]^2\\
&= 2 + 2(1-2q)^4\,\mathbb{E}\left[X_wX_{\bar{w}}X_uX_{\bar{u}}\right] - (1-2q)^4\,\mathbb{E}\left[X_wX_{\bar{w}} + X_uX_{\bar{u}}\right]^2\\
&= 2 + 2(1-2q)^4\,\mu_A - (1-2q)^4\left(\mu_e\left(1+\mu_A\right)\right)^2\\
&= 2 + (1-2q)^4\left[2\mu_A - \mu_e^2\left(1+\mu_A\right)^2\right]. \qquad (58)
\end{aligned}$$

By applying Bernstein's inequality, for any $t>0$

$$\mathbb{P}\left[\left|\sum_{i=1}^{n_\dagger} M_{e,u,\bar{u}}^{(i)} - n_\dagger\mathbb{E}\left[M_{e,u,\bar{u}}\right]\right| \geq t\right] \leq 2\exp\left(-\frac{t^2}{2n_\dagger\,\text{Var}\left(M_{e,u,\bar{u}}\right) + \frac{2}{3}Mt}\right),$$
$$\mathbb{P}\left[\left|\sum_{i=1}^{n_\dagger} M_{e,u,\bar{u}}^{(i)} - n_\dagger\mathbb{E}\left[M_{e,u,\bar{u}}\right]\right| \leq t\right] \geq 1-2\exp\left(-\frac{t^2}{2n_\dagger\,\text{Var}\left(M_{e,u,\bar{u}}\right) + \frac{2}{3}Mt}\right). \quad (59)$$

In the same way as done previously we get

$$t \leq n_\dagger\left(\frac{8}{3n_\dagger}\log\frac{4p^3}{\delta} + \sqrt{\frac{2}{n_\dagger}\text{Var}\left(M_{e,u,\bar{u}}\right)\log\frac{4p^3}{\delta}}\right) \quad (60)$$

and

$$\begin{aligned}
\text{Var}\left(M_{e,u,\bar{u}}\right) &= 2 + (1-2q)^4\left[2\mu_A - \mu_e^2\left(1+\mu_A\right)^2\right]\\
&\leq 2 + (1-2q)^4\,2\mu_A\\
&= 2\left(1+\mu_A^\dagger\right). \qquad (61)
\end{aligned}$$

By setting $\epsilon_\dagger = \sqrt{\log(2p^2/\delta)\,2/n_\dagger}$, we derive the following bound on t

$$t \leq n_\dagger\left(4\epsilon_\dagger^2 + 2\epsilon_\dagger\sqrt{1+\mu_A^\dagger}\right) \leq n_\dagger\max\left\{8\epsilon_\dagger^2, 4\epsilon_\dagger\sqrt{1+\mu_A^\dagger}\right\}. \quad (62)$$

□

In Lemma 4, we derive the set of strong edges for the hidden model. There is a threshold $\frac{\tau^\dagger}{(1-2q)^2} \geq \tau$ as in the case where there was no noise [75] and the threshold was τ. Also we find a lower bound for the necessary number of samples for exact structure recovery. In fact we have $n_\dagger \geq n$, as expected. By setting $q=0$ (then the probability to flip a bit equals to zero) we derive the exact expressions for the threshold τ and the sufficient number of samples defined in [75]. Under the event $\mathrm{E}_\dagger^{\mathrm{strong}}(\epsilon_\dagger)$ only the strong edges are guaranteed to exist in the estimated structure $\mathrm{T}_\dagger^{\mathrm{CL}}$.

Lemma 4. *Define the set of strong edges: $\left\{(i,j)\in\mathcal{E}_{\mathrm{T}} : |\tanh\theta_{ij}| \geq \frac{\tau^\dagger}{(1-2q)^2}\right\}$. Under the events defined in Lemmas 2 and 3 all the strong edges will be recovered from the Chow-Liu algorithm with probability at least $1-\delta$. That is,*

$$\mathbb{P}\left[\mathrm{E}_\dagger^{strong}(\epsilon_\dagger)\right] \geq 1-\delta = 1-2p^2\exp\left(-\frac{n_\dagger \epsilon_\dagger^2}{2}\right).$$

Proof. □

Lemma 1 gives

$$\left(\sum_{i=1}^{n_\dagger} Z_{f,u,\bar{u}}^{(i)}\right)\left(\sum_{i=1}^{n_\dagger} M_{f,u,\bar{u}}^{(i)}\right) < 0 \implies$$
$$\sum_{i=1}^{n_\dagger} Z_{f,u,\bar{u}}^{(i)} \leq 0 \text{ or } \sum_{i=1}^{n_\dagger} M_{f,u,\bar{u}}^{(i)} \leq 0 \implies$$

$$\left|\sum_{i=1}^{n_\dagger} Z_{f,u,\bar{u}}^{(i)} - n_\dagger\mathbb{E}\left[Z_{f,u,\bar{u}}^{(i)}\right]\right| \geq n_\dagger\mathbb{E}\left[Z_{f,u,\bar{u}}^{(i)}\right] \quad \text{or}$$
$$\left|\sum_{i=1}^{n_\dagger} Y_{f,u,\bar{u}}^{(i)} - n_\dagger\mathbb{E}\left[Y_{f,u,\bar{u}}^{(i)}\right]\right| \geq n_\dagger\mathbb{E}\left[M_{f,u,\bar{u}}^{(i)}\right]$$

$$\underset{(43),(57)}{\overset{\text{Lemmas 2 and 3}}{\implies}}$$

$$(1-2q)^2\mu_f(1-\mu_A) \leq \max\left\{8\epsilon_\dagger^2, 4\epsilon_\dagger\sqrt{1-\mu_A^\dagger}\right\} \quad \text{or}$$
$$(1-2q)^2\mu_f(1+\mu_A) \leq \max\left\{8\epsilon_\dagger^2, 4\epsilon_\dagger\sqrt{1+\mu_A^\dagger}\right\}$$

which implies that

$$\left|\mu_f^\dagger\right| \leq (1-\mu_A)^{-1}\max\left\{8\epsilon_\dagger^2, 4\epsilon_\dagger\sqrt{1-\mu_A^\dagger}\right\} \quad \text{or}$$
$$\left|\mu_f^\dagger\right| \leq (1+\mu_A)^{-1}\max\left\{8\epsilon_\dagger^2, 4\epsilon_\dagger\sqrt{1+\mu_A^\dagger}\right\}$$

and the lasts yields to

$$\left|\mu_f^\dagger\right| \leq \max\left\{\frac{8\epsilon_\dagger^2}{(1-\mu_A)}, \frac{8\epsilon_\dagger^2}{(1+\mu_A)}, \frac{4\epsilon_\dagger\sqrt{1-\mu_A^\dagger}}{(1-\mu_A)}, \frac{4\epsilon_\dagger\sqrt{1+\mu_A^\dagger}}{(1+\mu_A)}\right\} \implies$$
$$\left|\mu_f^\dagger\right| \leq \max\left\{\frac{8\epsilon_\dagger^2}{(1-\mu_A)}, \frac{4\epsilon_\dagger\sqrt{1-\mu_A^\dagger}}{(1-\mu_A)}\right\} \implies$$
$$\left|\mu_f^\dagger\right| \leq \frac{4\epsilon_\dagger\sqrt{1-\mu_A^\dagger}}{(1-\mu_A)}. \tag{63}$$

We get the last inequality for non trivial values of the bound $\frac{8\epsilon_\dagger^2}{(1-\mu_A^\dagger)} \leq 1$ and by using the following bound

$$\frac{8\epsilon_\dagger^2}{(1-\mu_A)} \leq \frac{16\epsilon_\dagger^2}{(1-\mu_A)} \leq \frac{4\epsilon_\dagger}{\sqrt{1-\mu_A}} = \frac{4\epsilon_\dagger\sqrt{1-\mu_A}}{(1-\mu_A)} \leq \frac{4\epsilon_\dagger\sqrt{1-\mu_A^\dagger}}{(1-\mu_A)}.$$

Finally, the function $f(\mu_A) = \frac{4\epsilon_\dagger\sqrt{1-\mu_A^\dagger}}{(1-\mu_A)} = \frac{4\epsilon_\dagger\sqrt{1-(1-2q)^2\mu_A}}{(1-\mu_A)}$ is increasing with respect to μ_A (for all $\mu_A \leq 1$) and $\mu_A \leq \tanh\beta < 1$, thus we have

$$\left|\mu_f^\dagger\right| \leq \frac{4\epsilon_\dagger\sqrt{1-\mu_A^\dagger}}{(1-\mu_A)} \tag{64}$$
$$\leq \frac{4\epsilon_\dagger\sqrt{1-(1-2q)^4\tanh\beta}}{(1-\tanh\beta)} \triangleq \tau^\dagger. \tag{65}$$

Notice that $\tau^\dagger > \tau = \frac{4\epsilon}{\sqrt{1-\tanh\beta}}$ when $n = n_\dagger$ (or $\epsilon = \epsilon_\dagger$).

The weakest edge should satisfy the following property to guarantee the correct recovery of the tree under the event $\mathrm{E}_\dagger^{\mathrm{strong}}(\epsilon_\dagger)$

$$\left|\mu_f^\dagger\right| \geq \tau^\dagger \implies$$
$$(1-2q)^2\tanh\alpha \geq \frac{4\epsilon_\dagger\sqrt{1-(1-2q)^4\tanh\beta}}{(1-\tanh\beta)} \implies$$
$$\tanh\alpha \geq \frac{4\epsilon_\dagger\sqrt{1-(1-2q)^4\tanh\beta}}{(1-2q)^2(1-\tanh\beta)}. \tag{66}$$

When there is no noise [75, Lemma 9.8], we can guarantee exact recover with high probability under the event $\mathrm{E}^{\mathrm{strong}}(\epsilon)$ and the assumption that the weakest edge satisfies the inequality

$$\tanh\alpha \geq \frac{4\epsilon}{\sqrt{1-\tanh\beta}}. \tag{67}$$

Notice that (67) can be obtained by (66) when $q = 0$ and $n = n_\dagger$. When $q > 0$ and $n = n_\dagger$ it is clear that the set of trees which can be recovered from noisy observations is a subset of the set of trees that can be recovered from the original observations. Also, we have

$$\epsilon \triangleq \sqrt{\frac{2\log\left(2p^2/\delta\right)}{n}} \implies n = \frac{2}{\epsilon^2}\log\left(2p^2/\delta\right) \quad \text{and}$$
$$\epsilon_\dagger \triangleq \sqrt{\frac{2\log\left(2p^2/\delta\right)}{n}} \implies n_\dagger = \frac{2}{\epsilon_\dagger^2}\log\left(2p^2/\delta\right). \quad (68)$$

By combining (66) with (68) we found the number of samples that we need to recover the tree with probality at $1-\delta$,

$$n_\dagger > \frac{32\left[1-(1-2q)^4\tanh\beta\right]}{(1-\tanh\beta)^2(1-2q)^4\tanh^2\alpha}\log\frac{2p^2}{\delta}. \quad (69)$$

On the other hand, when there is no noise [75] we need

$$n > \frac{32}{\tanh^2\alpha\,(1-\tanh\beta)}\log\frac{2p^2}{\delta}. \quad (70)$$

The last two inequalities give us how the number of samples scales as a function of the probability q

$$\frac{n_\dagger}{n} \geq \frac{1-(1-2q)^4\tanh\beta}{(1-\tanh\beta)(1-2q)^4}$$
$$= \frac{1}{2}\left[e^{2\beta}\left((1-2q)^{-4}-1\right)+1+(1-2q)^{-4}\right]. \quad (71)$$

From the above we can distinguish specific cases for values of q. For instance when $q \to \frac{1}{2}^-$ then we need $n_\dagger = \infty$ for exact structure recovery, when $q \to 0$ then we need at least n number of samples for exact structure recovery.

A.2 Proof of Theorem 2 (Necessary number of samples)

In this section, we use a strong data processing inequality together with a family of models (considered also by Bresler and Karzand [75]) to derive the proof of Theorem 2. Specifically, we combine the proofs of Theorem 3.2 by Bresler and Karzand [75, Lemma 8.1] and a strong data processing inequality result by Polyanskiy and Wu [29]. First, we consider the following variation of Fano's inequality [76].

Lemma 5. *[76, Corollary 2.6]: Assume that Θ is a family of $M+1$ distributions $\theta_0, \theta_1, \ldots, \theta_M$ such that $M \geq 2$. Let P_{θ_i} be the distribution of the variable X under the model θ_i, if*

$$\frac{1}{M+1}\sum_{i=1}^{M}\boldsymbol{D}_{KL}\left(P_{\theta_i}||P_{\theta_0}\right) \leq \gamma\log M, \quad (72)$$

for any $\gamma \in (0, \frac{1}{8})$, then for the probability of the error p_e the following inequality holds: $p_e \geq \frac{\log(M+1)-1}{\log(M)} - \gamma$. We restrict the values of γ to $(0, \frac{1}{8})$ because we are interested in the case where $p_e \geq \frac{1}{2}$, in general the above holds for all values of $\gamma \in (0,1)$, see [76, Corollary 2.6].

At this point we consider Bresler and Karzand's construction [75, section 8.1] of $M+1$ different Ising model distributions $\{P_{\theta^i} : i \in \{0,\ldots,M\}\}$. This family of structured distributions is chosen such that the recovery task is sufficiently hard. First, we define P_{θ^0} to be an Ising model distribution with underlying structure a chain with p nodes and parameters $\theta^0_{j,j+1} = \alpha$ when j is odd and $\theta^0_{j,j+1} = \beta$ when j is even. The rest of family is constructed as follows: the elements of each θ^i are equal to the elements of θ^0 apart from two elements $\theta^i_{i,i+1} = 0$ and $\theta^i_{i,i+2} = \alpha$ for each odd value of i. There are $(p+1)/2$ distributions in the constructed family. Bresler and Karzand evaluate the upper bound for the quantity $\boldsymbol{S}_{\mathrm{KL}}(P_{\theta^0}||P_{\theta^i})$ for all $i \in [M]$ under this family of distributions and we have [75, Section 8.1]:

$$\begin{aligned}\boldsymbol{S}_{\mathrm{KL}}&(P_{\theta^0}||P_{\theta^i})\\ &= 2\alpha\left(\tanh(\alpha)-\tanh(\alpha)\tanh(\beta)\right)\\ &\leq 4\alpha\tanh(\alpha)e^{-2\beta}.\end{aligned} \quad (73)$$

For each distribution P_{θ^i} and $i \in \{0,\ldots,M\}$ we consider the distribution of the noisy variable in the hidden model $P^\dagger_{\theta^i} \triangleq P_{\mathbf{Y}|\mathbf{X}} \circ P_{\theta^i}$ and we would like to find an upper bound for the quantities $\boldsymbol{S}_{\mathrm{KL}}(P^\dagger_{\theta^0}||P^\dagger_{\theta^i})$. To do this we a use a strong data processing inequality result [29] for any binary symmetric channel. The input random variable $\mathbf{X}$ is considered to have correlated elements while the noise variables N_i are i.i.d $\mathsf{Rademacher}(q)$ which is equivalent to the hidden model that we consider in this paper. In fact we have the following bound

$$\eta_{\mathrm{KL}} \leq 1-(4q(1-q))^p. \quad (74)$$

The quantity η_{KL} is defined as:

$$\eta_{\mathrm{KL}} \triangleq \sup_Q \sup_{P:0<\boldsymbol{D}_{\mathrm{KL}}(P||Q)<\infty} \frac{\boldsymbol{D}_{\mathrm{KL}}\left(P_{\mathbf{Y}|\mathbf{X}}\circ P||P_{\mathbf{Y}|\mathbf{X}}\circ Q\right)}{\boldsymbol{D}_{\mathrm{KL}}\left(P||Q\right)}, \quad (75)$$

where $P_{\mathbf{Y}|\mathbf{X}}$ is the distribution of the BSC and P, Q are any distributions of the input variable $\mathbf{X}$. Since (75) has the supremum over all possible distributions it covers any pair of distributions in the desired family

$\{P_{\theta^j} : j \in \{0, \ldots, M\}\}$ and we have

$$\frac{\boldsymbol{D}_{\text{KL}}(P^{\dagger}_{\theta^0}||P^{\dagger}_{\theta^i})}{\boldsymbol{D}_{\text{KL}}(P_{\theta^0}||P_{\theta^i})} \overset{(74),(75)}{\leq} 1-(4q(1-q))^p \implies \tag{76}$$

$$\boldsymbol{S}_{\text{KL}}(P^{\dagger}_{\theta^0}||P^{\dagger}_{\theta^i}) \leq [1-(4q(1-q))^p]\boldsymbol{S}_{\text{KL}}(P_{\theta^0}||P_{\theta^i}). \tag{77}$$

(73) and (77) give

$$\boldsymbol{S}_{\text{KL}}(P^{\dagger}_{\theta^0}||P^{\dagger}_{\theta^i}) \leq [1-(4q(1-q))^p]4\alpha^2 e^{-2\beta}. \tag{78}$$

Finally, from (78) and Lemma 5 we derive the bound of Theorem 2.

A.3 Proof of Theorem 3, (Sufficient number of samples for a noisy Gaussian model)

Let $\mathbf{X} = (X_1, X_2, \ldots, X_p)$ be a Gaussian random vector with distribution $\mathcal{N}(0, \Sigma)$. We assume that the Markov property holds such that the underlying graph is a tree $\mathrm{T} = (\mathcal{V}, \mathcal{E})$. Also assumption 2 holds;

$$\operatorname{Var}(X_i) = \mathbb{E}[X_i^2] = 1, \quad \forall i \in \mathcal{V}$$
$$0 < \rho_m \leq |\mathbb{E}[X_i X_j]| \leq \rho_M < 1, \quad \forall (i,j) \in \mathcal{E}.$$

We consider i.i.d. Gaussian noise $\boldsymbol{N} \sim \mathcal{N}(0, \sigma^2\mathcal{I})$. The noisy output variables of the hidden model are $\tilde{\mathbf{Y}} = \boldsymbol{X} + \boldsymbol{N} \sim \mathcal{N}(0, \Sigma + \sigma^2\mathcal{I})$. Then

$$\rho^{\dagger}_{i,j} \triangleq \frac{\mathbb{E}[\tilde{Y}_i\tilde{Y}_j]}{\sqrt{\mathbb{E}[(\tilde{Y}_i)^2]\mathbb{E}[(\tilde{Y}_j)^2]}} = \frac{\mathbb{E}[\tilde{Y}_i\tilde{Y}_j]}{\sqrt{1+\sigma^2}^2} \tag{79}$$

$$= \mathbb{E}\left[\frac{\tilde{Y}_i}{\sqrt{1+\sigma^2}}\frac{\tilde{Y}_j}{\sqrt{1+\sigma^2}}\right], \quad \forall i,j \in \mathcal{V}. \tag{80}$$

The random variables $Y_i \triangleq \tilde{Y}_i/\sqrt{1+\sigma^2}$ are normalized Gaussian with variance equal to 1. Instead of using $\tilde{\mathbf{Y}}$, we use the normalized variable $\mathbf{Y}$ with distribution

$$\mathbf{Y} \sim \mathcal{N}\left(0, \frac{\Sigma + \sigma^2\mathcal{I}}{1+\sigma^2}\right). \tag{81}$$

Then

$$\operatorname{Var}(Y_i) = \mathbb{E}[Y_i^2] = 1, \quad \forall i \in \mathcal{V}, \tag{82}$$

$$\frac{\rho_m}{1+\sigma^2} \leq |\mathbb{E}[Y_iY_j]| = \left|\frac{\mathbb{E}[X_iX_j]}{1+\sigma^2}\right| \leq \frac{\rho_M}{1+\sigma^2}, \quad \forall (i,j) \in \mathcal{E}. \tag{83}$$

(83) shows that noise makes the edges "weaker", since $1+\sigma^2 > 1$ and for $\sigma \to \infty$ we have $|\mathbb{E}[Y_iY_j]| \to 0$ which makes the structure learning task impossible.

The following Lemma provides upper bounds on the probabilities of the sufficient events.

Lemma 6. *Define*

$$f^{(1)}_{u,\tilde{u},e}\left(\mathbf{Y}^{1:n}\right) \triangleq \sum_{i=1}^{n} Z^{(i)}_{f,u,\bar{u}} = \sum_{i=1}^{n} Y^{(i)}_w Y^{(i)}_{\tilde{w}} - Y^{(i)}_u Y^{(i)}_{\tilde{u}}, \tag{84}$$

$$f^{(2)}_{u,\tilde{u},e}\left(\mathbf{Y}^{1:n}\right) \triangleq \sum_{i=1}^{n} \tilde{Z}^{(i)}_{f,u,\tilde{u}} = \sum_{i=1}^{n} Y^{(i)}_w Y^{(i)}_{\tilde{w}} + Y^{(i)}_u Y^{(i)}_{\tilde{u}}. \tag{85}$$

Then

$$\mathbb{P}\left[\bigcap_{u,\tilde{u},e}\left\{\left|f^{(1)}_{u,\tilde{u},e}\left(\mathbf{Y}^{1:n}\right) - \mathbb{E}[f^{(1)}_{u,\tilde{u},e}\left(\mathbf{Y}^{1:n}\right)]\right| \leq R\sqrt{\operatorname{Var}(f^{(1)}_{u,\tilde{u},e}\left(\mathbf{Y}^{1:n}\right))}\log^2\left(\frac{p^3e^2}{\delta}\right)\right\}\right] \geq 1-\frac{\delta}{2}.$$

and

$$\mathbb{P}\left[\bigcap_{u,\tilde{u},e}\left\{\left|f^{(2)}_{u,\tilde{u},e}\left(\mathbf{Y}^{1:n}\right) - \mathbb{E}[f^{(2)}_{u,\tilde{u},e}\left(\mathbf{Y}^{1:n}\right)]\right| \leq R\sqrt{\operatorname{Var}(f^{(2)}_{u,\tilde{u},e}\left(\mathbf{Y}^{1:n}\right))}\log^2\left(\frac{p^3e^2}{\delta}\right)\right\}\right] \geq 1-\frac{\delta}{2}.$$

and $R \in \mathbb{R}^+$.

Proof. We apply a concentration of measure Theorem by Schudy and Sviridenko [77, Theorem 1.10];

$$\mathbb{P}\left[\left|f\left(\mathbf{Y}^{1:n}\right) - \mathbb{E}[f\left(\mathbf{Y}^{1:n}\right)\right| \geq \lambda\right] \leq e^2 e^{-\left(\frac{\lambda}{R\sqrt{\operatorname{Var}(f(\mathbf{Y}^{1:n}))}}\right)^{1/q}}, \quad \forall \lambda > 0, \tag{86}$$

where $f\left(\mathbf{Y}^{1:n}\right) = f(Y_1, \ldots, Y_n)$ is a q degree polynomial and the random variables $Y_1, \ldots, Y_n$ are distributed according to a log-concave measure in $\mathbb{R}^n$ and they are not necessarily independent. In our case $f(\mathbf{Y}) = \sum_{i=1}^n Z^{(i)}_{f,u,\bar{u}}$ or $f(\mathbf{Y}) = \sum_{i=1}^n \tilde{Z}^{(i)}_{f,u,\tilde{u}}$ and we have $q = 2$. Then we choose the probability to be at least $\frac{\delta}{2p\binom{p}{2}}$, since we apply union bound for all pairs of nodes $u, \tilde{u}$ and edges $e = (w, \tilde{w}) \in \text{path}_{\mathrm{T}}(u, \tilde{u})$

$$\frac{\delta}{2p\binom{p}{2}} = e^2 e^{-\left(\frac{\lambda}{R\sqrt{\operatorname{Var}(f(\mathbf{Y}^{1:n}))}}\right)^{1/q}} \implies$$

$$\lambda = R\sqrt{\operatorname{Var}(f(\mathbf{Y}))}\log^2\left(\frac{2pe^2\binom{p}{2}}{\delta}\right) < R\sqrt{\operatorname{Var}(f(\mathbf{Y}))}\log^2\left(\frac{e^2p^3}{\delta}\right). \tag{87}$$

Then we have to calculate the $\mathrm{Var}(f(\mathbf{Y}))$ for both cases, when $f(\mathbf{Y}) = \sum_{i=1}^{n} Z^{(i)}_{f,u,\bar{u}}$ and $f(\mathbf{Y}) = \sum_{i=1}^{n} \tilde{Z}^{(i)}_{f,u,\tilde{u}}$. When $f(\mathbf{Y}) = \sum_{i=1}^{n} Z^{(i)}_{f,u,\bar{u}}$, by using Isserlis' theorem [79] we can express the higher order moments in terms of the covariates:

$$\mathbb{E}[Y_i^2 Y_j^2] = \mathbb{E}[Y_i^2]\mathbb{E}[Y_j^2] + 2\mathbb{E}^2[Y_i Y_j], \tag{88}$$

$$\begin{aligned}\mathbb{E}[Y_i Y_{\tilde{i}} Y_j Y_{\tilde{j}}] = \mathbb{E}[Y_i Y_{\tilde{i}}]\mathbb{E}[Y_j X_{\tilde{j}}] + \mathbb{E}[Y_i Y_j]\mathbb{E}[Y_{\tilde{i}} Y_{\tilde{j}}] \\ + \mathbb{E}[Y_i Y_{\tilde{j}}]\mathbb{E}[Y_{\tilde{i}} Y_j],\end{aligned} \tag{89}$$

and we have

$$\begin{aligned}
&\mathrm{Var}\left(f^{(1)}_{u,\tilde{u},e}\right) \\
&\overset{\text{i.i.d.}}{=} \sum_{i=1}^{n} \mathrm{Var}\left(Z^{(i)}_{f,u,\bar{u}}\right) \\
&= n\,\mathrm{Var}\left(Z_{f,u,\tilde{u}}\right) \\
&= n\left(\mathbb{E}[(Y_w Y_{\tilde{w}} - Y_u Y_{\tilde{u}})^2] - \mathbb{E}^2[Y_w Y_{\tilde{w}} - Y_u Y_{\tilde{u}}]\right) \\
&= n\Bigg(\mathbb{E}[Y_w^2 Y_{\tilde{w}}^2] + \mathbb{E}[Y_u^2 Y_{\tilde{u}}^2] - 2\mathbb{E}[Y_w Y_{\tilde{w}} Y_u Y_{\tilde{u}}] \\
&\qquad - \mathbb{E}^2[Y_w Y_{\tilde{w}}] - \mathbb{E}^2[Y_u Y_{\tilde{u}}] + 2\mathbb{E}[Y_w Y_{\tilde{w}}]\mathbb{E}[Y_u Y_{\tilde{u}}]\Bigg) \\
&= n\Big(2 + \mathbb{E}^2[Y_w Y_{\tilde{w}}] + \mathbb{E}^2[Y_u Y_{\tilde{u}}] \\
&\quad - 2\left(\mathbb{E}[Y_w Y_u]\mathbb{E}[Y_{\tilde{w}} Y_{\tilde{u}}] + \mathbb{E}[Y_w Y_{\tilde{u}}]\mathbb{E}[Y_{\tilde{w}} Y_u]\right)\Big) \qquad (90) \\
&\le n\left(6 + \mathbb{E}^2[Y_w Y_{\tilde{w}}] + \mathbb{E}^2[Y_u Y_{\tilde{u}}]\right) \\
&= n\left(6 + \frac{1}{(1+\sigma^2)^2}\mathbb{E}^2[X_w X_{\tilde{w}}] + \frac{1}{(1+\sigma^2)^2}\mathbb{E}^2[X_u X_{\tilde{u}}]\right) \\
&= 6n + \frac{n}{(1+\sigma^2)^2}\mathbb{E}^2[X_u X_{\tilde{u}}]\left(\prod_{e\in \mathrm{path}(w,\tilde{w})\setminus(u,\tilde{u})} \mu_e^2 + 1\right),
\end{aligned}$$

where (90) comes from (82), (88), (89) and the last comes from the correlation coefficient decay property.

In a similar way,

$$\begin{aligned}
&\mathrm{Var}\left(f^{(2)}_{u,\tilde{u},e}\right) \\
&\overset{\text{i.i.d.}}{=} \sum_{i=1}^{n} \mathrm{Var}\left(\tilde{Z}^{(i)}_{f,u,\tilde{u}}\right) \\
&= n\,\mathrm{Var}\left(\tilde{Z}_{f,u,\tilde{u}}\right) \\
&= n\left(\mathbb{E}[(Y_w Y_{\tilde{w}} + Y_u Y_{\tilde{u}})^2] - \mathbb{E}^2[Y_w Y_{\tilde{w}} + Y_u Y_{\tilde{u}}]\right) \\
&= n(\mathbb{E}[Y_w^2 Y_{\tilde{w}}^2] + \mathbb{E}[Y_u^2 Y_{\tilde{u}}^2] + 2\mathbb{E}[Y_w Y_{\tilde{w}} Y_u Y_{\tilde{u}}] \\
&\qquad - \mathbb{E}^2[Y_w Y_{\tilde{w}}] - \mathbb{E}^2[Y_u Y_{\tilde{u}}] - 2\mathbb{E}[Y_w Y_{\tilde{w}}]\mathbb{E}[Y_u Y_{\tilde{u}}]) \\
&= n(2 + \mathbb{E}^2[Y_w Y_{\tilde{w}}] + \mathbb{E}^2[Y_u Y_{\tilde{u}}] \\
&\qquad + 2\left(\mathbb{E}[Y_w Y_u]\mathbb{E}[Y_{\tilde{w}} Y_{\tilde{u}}] + \mathbb{E}[Y_w Y_{\tilde{u}}]\mathbb{E}[Y_{\tilde{w}} Y_u]\right)) \\
&\le n\left(6 + \mathbb{E}^2[Y_w Y_{\tilde{w}}] + \mathbb{E}^2[Y_u Y_{\tilde{u}}]\right) \\
&= n\left(6 + \frac{1}{(1+\sigma^2)^2}\mathbb{E}^2[X_w X_{\tilde{w}}] + \frac{1}{(1+\sigma^2)^2}\mathbb{E}^2[X_u X_{\tilde{u}}]\right) \\
&= 6n + \frac{6}{(1+\sigma^2)^2}\mathbb{E}^2[X_u X_{\tilde{u}}]\left(\prod_{e\in \mathrm{path}(w,\tilde{w})\setminus(u,\tilde{u})} \mu_e^2 + 1\right),
\end{aligned}$$

and we have

$$\begin{aligned}
\lambda &< R\sqrt{\mathrm{Var}\left(f\left(\mathbf{Y}^{1:n}\right)\right)}\log^2\left(\frac{e^2 p^3}{\delta}\right) \\
&\le R\sqrt{7n + \frac{n}{(1+\sigma^2)^2}\prod_{e\in \mathrm{path}(w,\tilde{w})\setminus(u,\tilde{u})} \mu_e^2}\,\log^2\left(\frac{e^2 p^3}{\delta}\right).
\end{aligned}$$

Similarly to the Ising model, we start by stating the condition for the error event; Let $f = (w, \bar{w})$ be an edge: $f \in \mathrm{T}$ and $f \notin \mathrm{T}^{\mathrm{CL}}_{\dagger}$ then $\exists\, g \in \mathrm{T}^{\mathrm{CL}}_{\dagger}$ and $g \notin \mathrm{T}$: $f \in \mathrm{path}_{\mathrm{T}}(u, \bar{u})$ and $g \in \mathrm{path}_{\mathrm{T}^{\mathrm{CL}}_{\dagger}}(w, \bar{w})$, then for the error event we have

$$\begin{aligned}
0 &\ge \left|\hat{\rho}^{\dagger}_f\right|^2 - \left|\hat{\rho}^{\dagger}_g\right|^2 \\
&= \left(\hat{\rho}^{\dagger}_f - \hat{\rho}^{\dagger}_g\right)\left(\hat{\rho}^{\dagger}_f + \hat{\rho}^{\dagger}_g\right) \\
&\overset{(82)}{=} \frac{1}{n^2}\left(\sum_{i=1}^{n} Y_w^{(i)} Y_{\bar{w}}^{(i)} - Y_u^{(i)} Y_{\bar{u}}^{(i)}\right) \\
&\qquad \times \left(\sum_{i=1}^{n} Y_w^{(i)} Y_{\bar{w}}^{(i)} + Y_u^{(i)} Y_{\bar{u}}^{(i)}\right) \\
&= \frac{1}{n^2}\left(\sum_{i=1}^{n} Z^{(i)}_{f,u,\bar{u}}\right)\left(\sum_{i=1}^{n} \tilde{Z}^{(i)}_{f,u,\tilde{u}}\right) \implies
\end{aligned}$$

$$\sum_{i=1}^{n} Z^{(i)}_{f,u,\bar{u}} \le 0 \quad \text{or}$$

$$\sum_{i=1}^{n} \tilde{Z}^{(i)}_{f,u,\tilde{u}} \le 0 \qquad \implies$$

$$\left|\sum_{i=1}^{n} Z_{f,u,\tilde{u}}^{(i)} - \mathbb{E}\left[\sum_{i=1}^{n} Z_{f,u,\tilde{u}}^{(i)}\right]\right| \geq \left|\mathbb{E}\left[\sum_{i=1}^{n} Z_{f,u,\tilde{u}}^{(i)}\right]\right| \quad \text{or}$$

$$\left|\sum_{i=1}^{n} \tilde{Z}_{f,u,\tilde{u}}^{(i)} - \mathbb{E}\left[\sum_{i=1}^{n} \tilde{Z}_{f,u,\tilde{u}}^{(i)}\right]\right| \geq \left|\mathbb{E}\left[\sum_{i=1}^{n} \tilde{Z}_{f,u,\tilde{u}}^{(i)}\right]\right| \quad (91)$$

From Lemma 6 with probability at least $1-\delta$ the following holds

$$R\sqrt{7n + \frac{n}{(1+\sigma^2)^2} \prod_{e\in \mathrm{path}(w,\tilde{w})\setminus(u,\tilde{u})} \mu_e^2 \log^2\left(\frac{e^2p^3}{\delta}\right)}$$
$$\geq R\sqrt{\mathrm{Var}(f_{u,\tilde{u},e}^{(1)}(\mathbf{Y}^{1:n}))\log^2\left(\frac{e^2p^3}{\delta}\right)}$$
$$\geq \left|\sum_{i=1}^{n} Z_{f,u,\tilde{u}}^{(i)} - \mathbb{E}\left[\sum_{i=1}^{n} Z_{f,u,\tilde{u}}^{(i)}\right]\right|,$$

$$R\sqrt{7n + \frac{n}{(1+\sigma^2)^2} \prod_{e\in \mathrm{path}(w,\tilde{w})\setminus(u,\tilde{u})} \mu_e^2 \log^2\left(\frac{e^2p^3}{\delta}\right)}$$
$$\geq R\sqrt{\mathrm{Var}(f_{u,\tilde{u},e}^{(2)}(\mathbf{Y}^{1:n}))\log^2\left(\frac{e^2p^3}{\delta}\right)}$$
$$\geq \left|\sum_{i=1}^{n} \tilde{Z}_{f,u,\tilde{u}}^{(i)} - \mathbb{E}\left[\sum_{i=1}^{n} \tilde{Z}_{f,u,\tilde{u}}^{(i)}\right]\right|. \quad (92)$$

We combine (91) and (92) and we have

$$R\sqrt{7n + \frac{n}{(1+\sigma^2)^2} \prod_{e\in \mathrm{path}(w,\tilde{w})\setminus(u,\tilde{u})} \mu_e^2 \log^2\left(\frac{e^2p^3}{\delta}\right)}$$
$$\geq \left|\mathbb{E}\left[\sum_{i=1}^{n} Z_{f,u,\tilde{u}}^{(i)}\right]\right|$$

or

$$R\sqrt{7n + \frac{n}{(1+\sigma^2)^2} \prod_{e\in \mathrm{path}(w,\tilde{w})\setminus(u,\tilde{u})} \mu_e^2 \log^2\left(\frac{e^2p^3}{\delta}\right)}$$
$$\geq \left|\mathbb{E}\left[\sum_{i=1}^{n} \tilde{Z}_{f,u,\tilde{u}}^{(i)}\right]\right| \implies$$

$$R\sqrt{7n + \frac{n}{(1+\sigma^2)^2} \prod_{e\in \mathrm{path}(w,\tilde{w})\setminus(u,\tilde{u})} \mu_e^2 \log^2\left(\frac{e^2p^3}{\delta}\right)}$$
$$\geq n\left|\frac{\mathbb{E}[X_u X_{\tilde{u}}]}{1+\sigma^2}\left(\prod_{e\in \mathrm{path}(w,\tilde{w})\setminus(u,\tilde{u})} \mu_e - 1\right)\right|$$

or

$$R\sqrt{7n + \frac{n}{(1+\sigma^2)^2} \prod_{e\in \mathrm{path}(w,\tilde{w})\setminus(u,\tilde{u})} \mu_e^2 \log^2\left(\frac{e^2p^3}{\delta}\right)}$$
$$\geq n\left|\frac{\mathbb{E}[X_u X_{\tilde{u}}]}{1+\sigma^2}\left(\prod_{e\in \mathrm{path}(w,\tilde{w})\setminus(u,\tilde{u})} \mu_e + 1\right)\right| \implies$$

$$\frac{R\sqrt{n\left(7 + \frac{1}{(1+\sigma^2)^2}\prod_{e\in \mathrm{path}(w,\tilde{w})\setminus(u,\tilde{u})} \mu_e^2\right)\log^2\left(\frac{e^2p^3}{\delta}\right)}}{1 - \prod_{e\in \mathrm{path}(w,\tilde{w})\setminus(u,\tilde{u})} \mu_e}$$
$$\geq n\left|\frac{\mathbb{E}[X_u X_{\tilde{u}}]}{1+\sigma^2}\right|$$

or

$$\frac{R\sqrt{n\left(7 + \frac{1}{(1+\sigma^2)^2}\prod_{e\in \mathrm{path}(w,\tilde{w})\setminus(u,\tilde{u})} \mu_e^2\right)\log^2\left(\frac{e^2p^3}{\delta}\right)}}{\prod_{e\in \mathrm{path}(w,\tilde{w})\setminus(u,\tilde{u})} \mu_e + 1}$$
$$\geq n\left|\frac{\mathbb{E}[X_u X_{\tilde{u}}]}{1+\sigma^2}\right|. \quad (93)$$

From (93) we find the sufficient condition for the weakest edge: for exact structure recovery we need ρ_m to be greater than the following term

$$\frac{R(1+\sigma^2)\sqrt{\left(7 + \frac{1}{(1+\sigma^2)^2}\prod_{e\in \mathrm{path}(w,\tilde{w})\setminus(u,\tilde{u})} \mu_e^2\right)}}{\sqrt{n}\left(1 - \prod_{e\in \mathrm{path}(w,\tilde{w})\setminus(u,\tilde{u})} \mu_e\right)}$$
$$\times \log^2\left(\frac{e^2p^3}{\delta}\right).$$

The function $f(x) = \frac{R\sqrt{(7+x)}\log^2\left(\frac{e^2p^3}{\delta}\right)}{\sqrt{n}(1-x)}$ is increasing for all $x \in [0,1)$ and $\left|\prod_{e\in \mathrm{path}(w,\tilde{w})\setminus(u,\tilde{u})} \mu_e^2\right| \leq \rho_M$. Thus, it is sufficient to have

$$\rho_m \geq \frac{R\sqrt{7(1+\sigma^2)^2 + \rho_M}\log^2\left(\frac{e^2p^3}{\delta}\right)}{\sqrt{n}\,(1-\rho_M)}$$

and the sufficient number of samples is given by:

$$n \geq \frac{R^2\left[7(1+\sigma^2)^2 + \rho_M\right]\log^4\left(\frac{e^2p^3}{\delta}\right)}{\rho_m^2\,(1-\rho_M)^2},$$

where R is a positive constant.

B Additional Experiments

We define the *distance between two tree structures* $\mathrm{T} = (\mathcal{V},\mathcal{E}), \mathrm{T}' = (\mathcal{V},\mathcal{E}')$ with identical node set and possibly different edge sets as

$$\mathcal{D}_{\mathcal{T}}(\mathrm{T},\mathrm{T}') \triangleq \frac{|\mathcal{E}\triangle\mathcal{E}'|}{2}, \quad (94)$$

where the symbol $\triangle$ denotes the symmetric difference between two sets. Note that $0 \leq \mathcal{D}_{\mathcal{T}}(\mathrm{T},\mathrm{T}') \leq \max\{|\mathcal{E}|,|\mathcal{E}'|\}$. The definition in (94) can be used as an alternative of (10) for evaluating the performance of the Chow-Liu algorithm. In particular, $\mathcal{D}_{\mathcal{T}}(\mathrm{T},\mathrm{T}^{\mathrm{CL}}_{\dagger})$

counts the number of incorrect edges for the estimated structure. Similar metrics can be found in the literature. A closely related one is *"false positive and false negative rates"*, which has been considered by Liu et.al [80].

Synthetic Data. To demonstrate the performance of the algorithm experimentally, we present the decay of the error based on the metric $\mathcal{D}_{\mathcal{T}}(\mathrm{T}, \mathrm{T}^{\mathrm{CL}}_{\dagger})$ (Figure 2) and the probability of the error event $\{\mathrm{T} \neq \mathrm{T}^{\mathrm{CL}}_{\dagger}\}$ (Figure 3), while the number of samples increases, for fixed values of the parameters α, β, p, while the crossover probability q varies between 0 and 1/2. Specifically, for the plots in Figures 2 and 3, we have chosen $\alpha = \operatorname{arctanh}(0.25)$, $\beta = \operatorname{arctanh}(0.75)$, $p = 100$. These results illustrate how noisy observations can significantly degrade performance unless we increase the sample size significantly. We consider synthetic Gaussian data for the plots of Figure 4. These show how the error $\mathcal{D}_{\mathcal{T}}(\mathrm{T}, \mathrm{T}^{\mathrm{CL}}_{\dagger})$ and the probability of the not exact recovery $\{\mathrm{T} \neq \mathrm{T}^{\mathrm{CL}}_{\dagger}\}$ varies as the number of observations increases and for different values of the signal to noise ration (SNR).

Real Data. We consider as observations the increase (spin up) or decrease (spin down) of the closing prices for 10 stocks. The estimated tree structure T^{CL} is found by applying Chow-Liu's algorithm, Figure 5. Noisy data are generated by flipping each observation with probability q. Then the structure $\mathrm{T}^{\mathrm{CL}}_{\dagger}$ is estimated by taking into consideration (semi-synthetic) noisy data. The error $\mathcal{D}_{\mathcal{T}}(\mathrm{T}^{\mathrm{CL}}, \mathrm{T}^{\mathrm{CL}}_{\dagger})$ is plotted as function of q in figure 5. Notice that for hidden model structure estimates, where $q \in (0, 1/2)$, we see that small noise levels lead to a modest increase in sample complexity for a target error probability, but as the channel gets worse, the sample complexity explodes.

C Connections with Differential Privacy

One way in which a hidden model can arise in in inference from data released under differential privacy. Suppose that data about individuals can be modeled as drawn from an Ising model: the j-th sample from the population has data $\mathbf{X}(j)$ drawn according to $\mathrm{p}(\cdot)$ representing p correlated features characterizing the individual. Because of privacy concerns, the analyst is only given access to $\mathbf{Y}(j)$, where each feature is randomly flipped with probability q. The noisy data guarantees *differential privacy* [81]: we can think of this process as a form of vectorized randomized response. More formally, the noisy samples guarantees

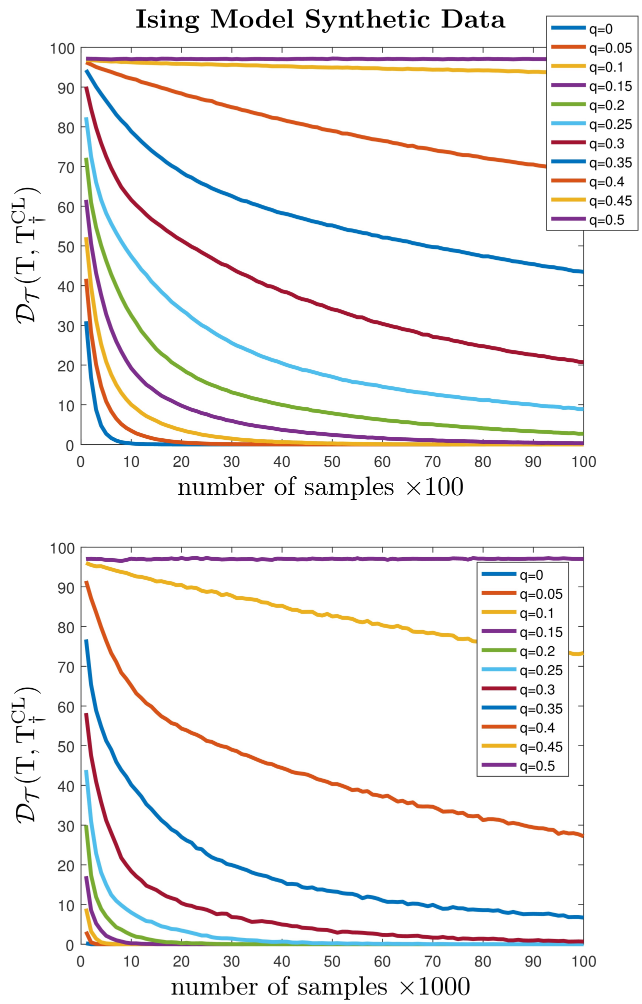


Figure 2: $\mathcal{D}_{\mathcal{T}}(\mathrm{T}, \mathrm{T}^{\mathrm{CL}}_{\dagger})$ as a function of number of samples. The upper graph is over 1000 independent runs and up to 10^4 independent samples, while the down over 100 independent runs and up to 10^5 independent samples.

ϵ-diffferential privacy if for all $\boldsymbol{c}, \boldsymbol{c'}, \boldsymbol{c''} \in \{0, 1\}^p$,

$$\frac{\mathbb{P}\left(\mathbf{Y} = \boldsymbol{c} | \mathbf{X} = \boldsymbol{c'}\right)}{\mathbb{P}\left(\mathbf{Y} = \boldsymbol{c} | \mathbf{X} = \boldsymbol{c''}\right)} \leq e^{\epsilon}. \tag{95}$$

For our choice of q,

$$\begin{aligned} \frac{\mathbb{P}\left(\mathbf{Y} = \boldsymbol{c} | \mathbf{X} = \boldsymbol{c'}\right)}{\mathbb{P}\left(\mathbf{Y} = \boldsymbol{c} | \mathbf{X} = \boldsymbol{c''}\right)} &= \frac{(1-q)^{p-\ell} q^{\ell}}{(1-q)^{p-\ell'} q^{\ell'}} \\ &= \left[\frac{1-q}{q}\right]^{\ell' - \ell}, \end{aligned} \tag{96}$$

where ℓ, ℓ' is the number of different elements of the pairs $\boldsymbol{c}, \boldsymbol{c'}$ and $\boldsymbol{c}, \boldsymbol{c''}$ respectively, for any $\boldsymbol{c}, \boldsymbol{c'}, \boldsymbol{c''} \in \{-1, +1\}^p$. Since $\ell, \ell \in \{1, 2, \dots, p\}$ and $q \in [0, 1/2]$,

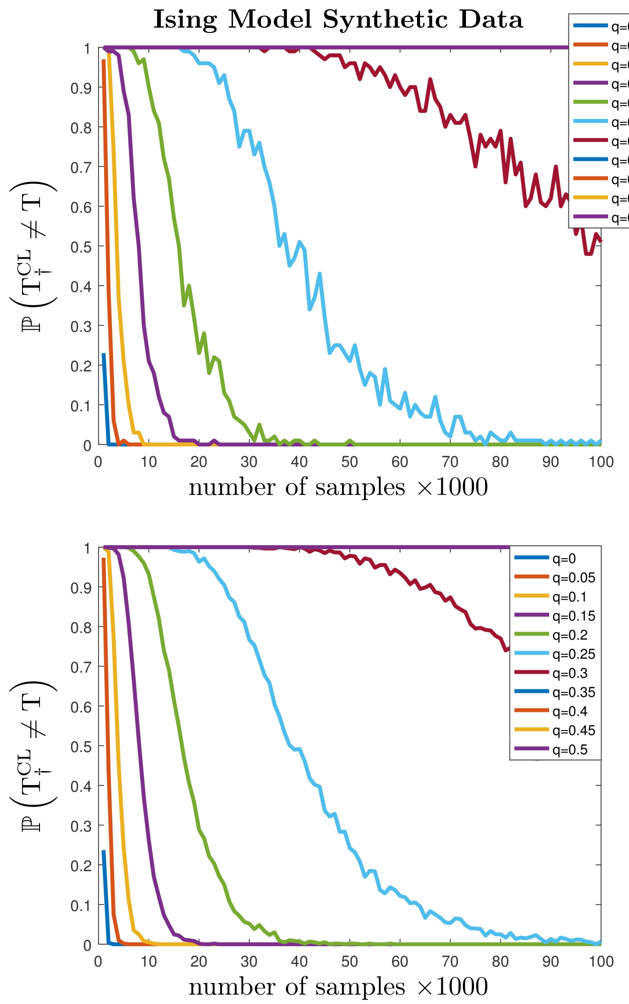


Figure 3: Estimating the probability of the error event, $\mathbb{P}\left(\mathrm{T}_{\dagger}^{\mathrm{CL}} \neq \mathrm{T}\right)$, as a function of number of samples. The upper graph is estimated through 100 independent runs while the right through 1000 independent runs.

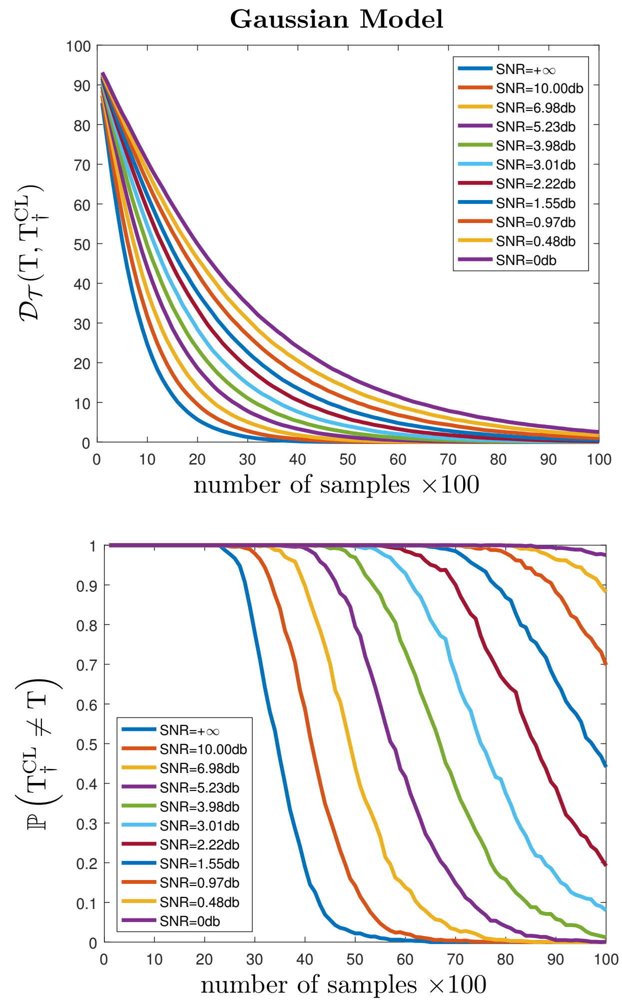


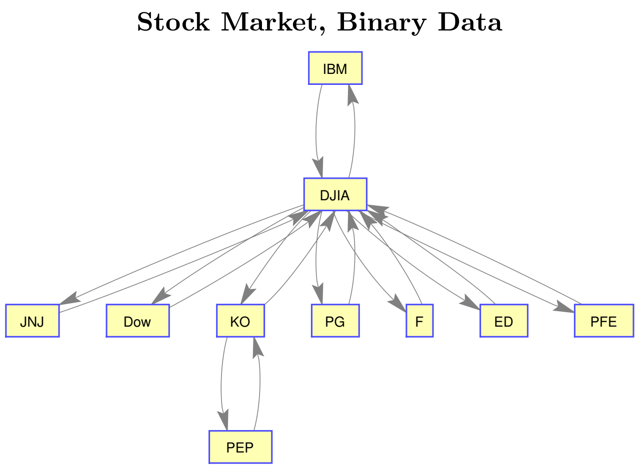
Figure 4: **Upper**: $\mathcal{D}_{\mathcal{T}}(\mathrm{T}, \mathrm{T}_{\dagger}^{\mathrm{CL}})$ as a function of number of samples. **Down**: Estimating the probability of the error event. Both simulations are through 1000 independent iterations.

we may write

$$\frac{\mathbb{P}(\mathbf{Y}=\boldsymbol{c}|\mathbf{X}=\boldsymbol{c'})}{\mathbb{P}(\mathbf{Y}=\boldsymbol{c}|\mathbf{X}=\boldsymbol{c''})} \leq \max_{\ell,\ell'} \left[\frac{1-q}{q}\right]^{\ell'-\ell} = \left[\frac{1-q}{q}\right]^{p}. \quad (97)$$

Thus for $\epsilon_o = p \log((1-q)/q)$ we guarantee ϵ_o-local differential privacy.

We can interpret the main result of this paper, in terms of differential privacy, as characterizing the tradeoff between privacy and sample complexity in inference from data protected by differential privacy. In this simplified mechanism, however, each individual data sample is perturbed, which is a form of *local* differential privacy or (alternatively) input perturbation. An interesting question would be the tradeoff in standard differential privacy, where the algorithm releases only the estimated tree.



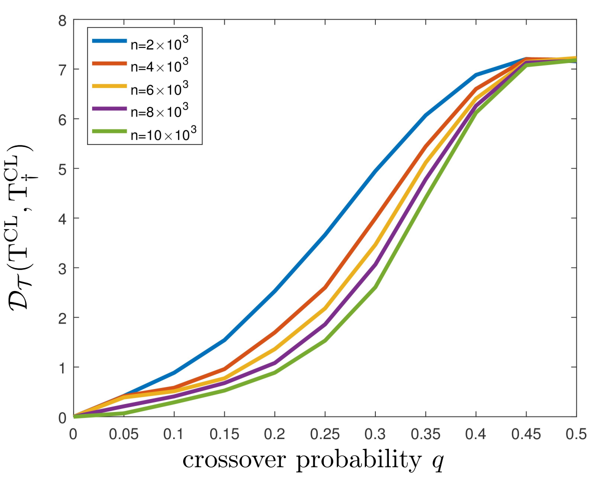


Figure 5: **Upper**: Tree structure estimate T^{CL} for the closing prices of 10 different stocks. 10^5 number of samples have been considered, data have been retrieved from `https://finance.yahoo.com`. **Down**: $\mathcal{D}_{\mathcal{T}}(\mathrm{T}^{\mathrm{CL}}, \mathrm{T}^{\mathrm{CL}}_{\dagger})$ as a function of q for different number of samples.