On the Difficulty of Learning Power Law Graphical Models : Proofs

Rashish Tandon
Department of Computer Science
University of Texas at Austin
TX, USA
Email: rashish@cs.utexas.edu

Pradeep Ravikumar
Department of Computer Science
University of Texas at Austin
TX, USA
Email: pradeepr@cs.utexas.edu

I. Proofs

Proof of Lemma 1. We have

$$p = \sum_{i=1}^{\Delta} \left\lfloor \frac{e^\alpha}{i^\beta} \right\rfloor \geq (e^\alpha - 1) + \sum_{i=2}^{\Delta} \left\lfloor \frac{e^\alpha}{i^\beta} \right\rfloor \geq e^\alpha \tag{1}$$

Also,

$$\begin{aligned} p &\leq \sum_{i=1}^{\Delta} \frac{e^\alpha}{i^\beta} \leq e^\alpha + e^\alpha \int_1^\Delta \frac{1}{i^\beta} \\ &= e^\alpha + e^\alpha \left(\frac{1}{\beta - 1} - \frac{1}{(\beta-1)\Delta^{\beta-1}} \right) \text{ (for } \beta > 1) \\ &\leq \frac{e^\alpha \beta}{\beta - 1} \end{aligned} \tag{2}$$

Similarly,

$$m = \frac{1}{2} \sum_{i=1}^{\Delta} i \left\lfloor \frac{e^\alpha}{i^\beta} \right\rfloor \geq \frac{e^\alpha}{2} \tag{3}$$

and,

$$m \leq \frac{1}{2} \sum_{i=1}^{\Delta} \frac{e^\alpha}{i^{\beta-1}} \leq \frac{1}{2} \frac{e^\alpha(\beta - 1)}{\beta - 2} \text{ (for } \beta > 2) \tag{4}$$

□

Proof of Lemma 2. For an (α, β)-sequence D, we have

$$M_2(D) = \sum_{i=1}^{p} d_i^2 = \sum_{i=1}^{\Delta} i^2 y_i = \sum_{i=1}^{\Delta} i^2 \left\lfloor \frac{e^\alpha}{i^\beta} \right\rfloor \tag{5}$$

Now, as in Lemma 1, we get

$$M_2(D) \geq e^\alpha \tag{6}$$

and,

$$M_2(D) \leq \sum_{i=1}^{\Delta} \frac{e^\alpha}{i^{\beta-2}} \leq \frac{e^\alpha(\beta - 2)}{\beta - 3} \text{ (for } \beta > 3) \tag{7}$$

□

Proof of Proposition 1. From the discussion, we have

$$|\mathcal{S}_{D,\alpha,\beta}| = P(W = 1) \frac{(2m)!}{m! 2^m} \frac{1}{\prod_{i=1}^{p} (d_i)!} \tag{8}$$

Now,

$$\begin{aligned} \log \left(\frac{2m!}{m! 2^m} \right) &\geq \log \left(\frac{m^m}{2^m} \right) \\ &\geq m \log m - m \\ &\geq \frac{e^\alpha}{2} \log \left(\frac{e^\alpha}{2} \right) - \frac{e^\alpha}{2} \left(\frac{\beta - 1}{\beta - 2} \right) \\ &= \frac{e^\alpha \alpha}{2} - \frac{e^\alpha}{2} \left(\frac{2\beta - 3}{\beta - 2} \right) \end{aligned} \tag{9}$$

and,

$$\begin{aligned} \log \left(\prod_{i=1}^{p} d_i! \right) &= \sum_{i=1}^{p} \log(d_i!) \\ &= \sum_{i=2}^{\Delta} y_i \log(i!) \\ &\leq \sum_{i=2}^{\Delta} \frac{e^\alpha}{i^\beta} \log(i!) \\ &\leq \sum_{i=2}^{\Delta} \frac{e^\alpha}{i^\beta} i \log i \\ &\leq e^\alpha \int_1^\Delta \frac{\log t}{t^{\beta-1}} dt \\ &\leq \frac{e^\alpha}{(\beta - 2)^2} \end{aligned} \tag{10}$$

Combining these with $P(W = 1) = e^{-O\left(\frac{(\beta-2)^2}{(\beta-3)^2}\right)}$ (for $\beta > 3$), we get the stated result. □

Proof of Theorem 1. Following the approach of [1], for $G \sim \mathcal{G}_\mathcal{U}(\alpha, \beta, D)$, we have

$$\begin{aligned} P(\phi(\mathbf{X}) \neq G) &= \sum_{g \in \mathcal{S}_{D,\alpha,\beta}} P(\phi(\mathbf{X}) \neq g \,|\, G = g) P(G = g) \\ &= \sum_{g \in \mathcal{S}_{D,\alpha,\beta} \cap \mathcal{R}(\phi)} P(\phi(\mathbf{X}) \neq g \,|\, G = g) P(G = g) \\ &\quad + \sum_{g \in \mathcal{S}_{D,\alpha,\beta} \cap \mathcal{R}(\phi)^c} P(\phi(\mathbf{X}) \neq g \,|\, G = g) P(G = g) \end{aligned} \tag{11}$$

where $\mathcal{R}(\phi)$ is the range of the estimator ϕ. So, $P(\phi(\mathbf{X}) \neq g \,|\, G = g) = 1$ for $g \in \mathcal{R}(\phi)^c$. Thus,

$$\begin{aligned}
P(\phi(\mathbf{X}) \neq G) &\geq \sum_{g \in \mathcal{S}_{D,\alpha,\beta} \cap \mathcal{R}(\phi)^c} P(\phi(\mathbf{X}) \neq g \,|\, G = g) P(G = g) \\
&= \sum_{g \in \mathcal{S}_{D,\alpha,\beta} \cap \mathcal{R}(\phi)^c} P(G = g) \\
&= 1 - \sum_{g \in \mathcal{S}_{D,\alpha,\beta} \cap \mathcal{R}(\phi)} P(G = g) \\
&= 1 - \frac{|\mathcal{S}_{D,\alpha,\beta} \cap \mathcal{R}(\phi)|}{|\mathcal{S}_{D,\alpha,\beta}|} \\
&\geq 1 - \frac{|\mathcal{X}|^{np}}{|\mathcal{S}_{D,\alpha,\beta}|}
\end{aligned} \tag{12}$$

where we used $|\mathcal{S}_{D,\alpha,\beta} \cap \mathcal{R}(\phi)| \leq |\mathcal{R}(\phi)| \leq |\mathcal{X}|^{np}$ for the last inequality.

Now, for any constant δ s.t. $0 < \delta < 1$, if $n \leq \delta \frac{\log|\mathcal{S}_{D,\alpha,\beta}|}{p \log|\mathcal{X}|}$, we would have $P(\phi(\mathbf{X}) \neq G) \to 1$ as $p \to \infty$. Then, plugging in the lower bound on $\log|\mathcal{S}_{D,\alpha,\beta}|$ from Proposition 1, we get the stated bound for n. $\square$

Proof of Proposition 2. We see that

$$\begin{aligned}
\log|\mathcal{D}| &= \log p! - \sum_{i=1}^{\Delta} \log y_i! \\
&= \Omega\left(p \log p - \sum_{i=1}^{\Delta} y_i \log y_i\right) \\
&= \Omega\left(\sum_{i=1}^{\Delta} y_i \log \frac{p}{y_i}\right) \\
&= \Omega\left(\sum_{i=1}^{\Delta} y_i \log \frac{p}{e^\alpha} i^\beta\right) \\
&= \Omega\left(p \log \frac{p}{e^\alpha} + \beta \sum_{i=1}^{\Delta} y_i \log i\right)
\end{aligned} \tag{13}$$

$\square$

Now, $\frac{p}{e^\alpha} \geq 1 + \frac{1}{2^\beta}$ and

$$\begin{aligned}
\sum_{i=1}^{\Delta} y_i \log i &\geq e^\alpha \int_1^{\Delta} \left(\frac{1}{i^\beta} - 1\right) \log i \\
&= \Omega\left(\frac{e^\alpha}{(\beta-1)^2}\right) = \Omega\left(\frac{p}{\beta(\beta-1)}\right)
\end{aligned} \tag{14}$$

Thus, $\log|\mathcal{D}| = \Omega(\log p)$ and so, considering all (α, β)-sequences does not influence the asymptotic lower bound for $|\mathcal{S}_{\alpha,\beta,D}|$. Thus, we may use the same lower bound for the case of $|\mathcal{S}_{\alpha,\beta}|$.

Proof of Theorem 2. The proof of Theorem 2 follows the same argument as Theorem 1, with $G \sim \mathcal{G}_{\mathcal{U}}(\alpha, \beta)$ (instead of $\mathcal{G}_{\mathcal{U}}(\alpha, \beta, D)$. So, we would have,

$$P(\phi(\mathbf{X}) \neq G) \geq 1 - \frac{|\mathcal{X}|^{np}}{|\mathcal{S}_{\alpha,\beta}|} \tag{15}$$

Thus, for any constant δ s.t. $0 < \delta < 1$, if $n \leq \delta \frac{\log|\mathcal{S}_{D,\alpha,\beta}|}{p \log|\mathcal{X}|}$, we would have $P(\phi(\mathbf{X}) \neq G) \to 1$ as $p \to \infty$. Now, plugging in the lower bound on $\log|\mathcal{S}_{\alpha,\beta}|$ from Proposition 2, we get the stated bound for n. $\square$

Proof of Theorem 3. For $G \sim \mathcal{G}_{\mathcal{U}}(\alpha, \beta, D)$, Fano's Lemma [2] gives us

$$\begin{aligned}
1 + P(\phi(\mathbf{X}) \neq G) \log|\mathcal{S}_{D,\alpha,\beta}| &\geq H(G \,|\, \mathbf{X}) \\
&= H(G) - I(G; \mathbf{X}) \\
&\geq \log|\mathcal{S}_{D,\alpha,\beta}| - np \log|\mathcal{X}|
\end{aligned} \tag{16}$$

where we have used $I(G; \mathbf{X}) \leq H(\mathbf{X}) \leq \log(|\mathcal{X}|^{np})$ in the last inequality. Thus,

$$n \geq \frac{\log|\mathcal{S}_{D,\alpha,\beta}|}{p \log|\mathcal{X}|}(1 - P(\phi(\mathbf{X}) \neq G)) \tag{17}$$

Now, using Proposition 1 and $P(\phi(\mathbf{X}) \neq G) \leq \delta$ gives us the result. $\square$

Proof of Proposition 3. Let $G \sim \mathcal{G}_{CL}(p, \overline{w}, \beta)$ and $G = (V, E)$. Then G may be viewed as a vector of random variables, $G = (G_{11}, G_{12}, \ldots, G_{ij}, \ldots)$, where

$$G_{ij} = \begin{cases} 1 & \text{if } (i,j) \in E \\ 0 & \text{otherwise} \end{cases} \tag{18}$$

All G_{ij}'s are independent and for any $i, j \in V$, $P(G_{ij} = 1) = \frac{w_i w_j}{\rho}$. So,

$$\begin{aligned}
H(G) &= \sum_{i=1}^{p} \sum_{j=i+1}^{p} H(G_{ij}) \\
&\geq -\sum_{i=1}^{p} \sum_{j=i+1}^{p} \frac{w_i w_j}{\rho} \log\left(\frac{w_i w_j}{\rho}\right) \\
&= -\sum_{i=1}^{p} \sum_{j=i+1}^{p} \frac{\alpha^2}{\rho} (ij)^{-\frac{1}{\beta-1}} \log\left(\frac{\alpha^2}{\rho} (ij)^{-\frac{1}{\beta-1}}\right) \\
&= \frac{\alpha^2}{\rho} \log\left(\frac{\rho}{\alpha^2}\right) \sum_{i=1}^{p} \sum_{j=i+1}^{p} (ij)^{-\frac{1}{\beta-1}} \\
&\quad + \frac{\alpha^2}{\rho(\beta-1)} \sum_{i=1}^{p} \sum_{j=i+1}^{p} (ij)^{-\frac{1}{\beta-1}} \log(ij)
\end{aligned} \tag{19}$$

Now, we have,

$$\sum_{i=1}^{p}\sum_{j=i+1}^{p}(ij)^{-\frac{1}{\beta-1}} \geq \int_{1}^{p+1}\int_{i+1}^{p+1}(ij)^{-\frac{1}{\beta-1}}djdi$$
$$\geq \int_{1}^{p}\int_{i+1}^{p+1}(ij)^{-\frac{1}{\beta-1}}djdi$$
$$\geq \frac{\beta-1}{\beta-2}\int_{1}^{p} i^{-\frac{1}{\beta-1}}\left[p^{\frac{\beta-2}{\beta-1}}-(i+1)^{\frac{\beta-2}{\beta-1}}\right]$$
$$= \Omega\left(\left(\frac{\beta-1}{\beta-2}\right)^{2} p^{\frac{2(\beta-2)}{(\beta-1)}}\right) \tag{20}$$

and,

$$\sum_{i=1}^{p}\sum_{j=i+1}^{p}(ij)^{-\frac{1}{\beta-1}}\log(ij) \geq \sum_{i=e^{\frac{\beta-1}{2}}}^{p}\sum_{j=i+1}^{p}(ij)^{-\frac{1}{\beta-1}}\log(ij)$$
$$\geq \int_{e^{\frac{\beta-1}{2}}}^{p}\int_{i+1}^{p+1}(ij)^{-\frac{1}{\beta-1}}\log(ij)$$
$$= \Omega\left(\left(\frac{\beta-1}{\beta-2}\right)^{2} p^{\frac{2(\beta-2)}{(\beta-1)}}\log p\right) \tag{21}$$

where the asymptotic bounds are obtained by computing the integral.

Also,

$$\begin{aligned}\rho &= \alpha\sum_{k=1}^{p}k^{-\frac{1}{\beta-1}} \geq \alpha\int_{1}^{p+1}k^{-\frac{1}{\beta-1}}dk\\ &\geq \alpha\int_{1}^{p}k^{-\frac{1}{\beta-1}}dk = \overline{w}\left(p-p^{\frac{1}{\beta-1}}\right)\\ &= \Omega(\overline{w}p)\end{aligned} \tag{22}$$

and,

$$\begin{aligned}\rho &= \alpha\sum_{k=1}^{p}k^{-\frac{1}{\beta-1}} \leq \alpha+\alpha\int_{2}^{p+1}(k-1)^{-\frac{1}{\beta-1}}dk\\ &= \overline{w}\left(p-\frac{1}{\beta-1}p^{\frac{1}{\beta-1}}\right)\\ &= O(\overline{w}p)\end{aligned} \tag{23}$$

So, $\rho = \Theta(\overline{w}p)$.

Combining all of these bounds in the expression of $H(G)$ from Eq. 19 and substituting the value of α for the Chung-Lu model gives us the stated result. □

Proof of Theorem 4. We have $G \sim \mathcal{G}_{CL}(p,\overline{w},\beta)$. Let $P_e = P(\phi(\mathbf{X}) \neq G)$ and M be the set of all simple graphs on p nodes. Then, $|M| = 2^{\binom{p}{2}}$. Using Fano's Lemma [2], we have

$$\begin{aligned}1+P_e\log(|M|) &\geq H(G|\mathbf{X})\\ &\geq H(G)-np\log|\mathcal{X}|\end{aligned} \tag{24}$$

using the same argument as in Theorem 3. Thus,

$$\begin{aligned}n &\geq \frac{H(G)-P_e\log(|M|)}{p\log|\mathcal{X}|}-\frac{1}{p\log|\mathcal{X}|}\\ &= \frac{H(G)-P_ep^2}{p\log|\mathcal{X}|}-\frac{1}{p\log|\mathcal{X}|}\\ &\geq \frac{H(G)}{p\log|\mathcal{X}|}-\frac{1}{\log|\mathcal{X}|}-\frac{1}{p\log|\mathcal{X}|}\\ &= \frac{H(G)}{p\log|\mathcal{X}|}-O(1)\\ \Rightarrow n &= \Omega\left(\frac{H(G)}{p\log|\mathcal{X}|}\right)\end{aligned} \tag{25}$$

Now, substituting the lower bound on $H(G)$ from Proposition 3 gives us the desired result. □

References

[1] G. Bresler, E. Mossel, and A. Sly, "Reconstruction of markov random fields from samples: Some observations and algorithms," in *APPROX*. Springer-Verlag, 2008, pp. 343–356.

[2] T. M. Cover and J. A. Thomas, *Elements of Information Theory (Wiley Series in Telecommunications and Signal Processing)*. Wiley-Interscience, 2006.