Learning Graphs with a Few Hubs - Supplementary

Rashish Tandon, Pradeep Ravikumar {RASHISH, PRADEEPR}@CS.UTEXAS.EDU

Department of Computer Science
The University of Texas at Austin, USA

1. Proof of Corollary 1

Proof. For any $t \in \mathcal{N}^*_{\text{sub}}(r)$, we have

$$\widehat{\mathcal{N}}_\lambda(r; D) = \mathcal{N}^*_{\text{sub}}(r) \Rightarrow t \in \widehat{\mathcal{N}}_\lambda(r; D). \tag{1}$$

For any $t \notin \mathcal{N}^*_{\text{sub}}(r)$, we have

$$t \in \widehat{\mathcal{N}}_\lambda(r; D) \Rightarrow \widehat{\mathcal{N}}_\lambda(r; D) \neq \mathcal{N}^*_{\text{sub}}(r). \tag{2}$$

Thus,

$$\begin{aligned} &\mathbb{P}(t \in \widehat{\mathcal{N}}_\lambda(r; D)) \geq \mathbb{P}(\widehat{\mathcal{N}}_\lambda(r; D) = \mathcal{N}^*_{\text{sub}}(r)) \quad \text{if } t \in \mathcal{N}^*_{\text{sub}}(r) \text{ and,} \\ &\mathbb{P}(t \in \widehat{\mathcal{N}}_\lambda(r; D)) \leq \mathbb{P}(\widehat{\mathcal{N}}_\lambda(r; D) \neq \mathcal{N}^*_{\text{sub}}(r)) \quad \text{if } t \notin \mathcal{N}^*_{\text{sub}}(r). \end{aligned} \tag{3}$$

Now, using the result of Theorem 1 proves the corollary. □

2. Proof of Proposition 1

Proof. The proof of this proposition is similar to Theorem 4.1 in (Liu et al., 2010). First note that,

$$\mathbb{E}\left[\widetilde{p}_{r,b,\lambda}(t; D)\right] = \frac{1}{\binom{n}{b}} \sum_{D_b \in S_b(D)} \mathbb{E}\left[F^t_{\lambda,r}(D_b)\right] = \frac{1}{\binom{n}{b}} \sum_{D_b \in S_b(D)} \mathbb{P}\left(t \in \widehat{\mathcal{N}}_{b,\lambda}(r; D_b)\right), \tag{4}$$

where the expectation and probability are taken over the samples D being drawn i.i.d. For any fixed set of b indices, drawing n samples i.i.d. and then choosing the b samples corresponding to the fixed indices is equivalent to drawing b samples i.i.d. Thus, for any $D_b \in S_b(D)$, we have $\mathbb{P}\left(t \in \widehat{\mathcal{N}}_{b,\lambda}(r; D_b)\right) = p_{r,b,\lambda}(t)$, which implies

$$\mathbb{E}\left[\widetilde{p}_{r,b,\lambda}(t; D)\right] = p_{r,b,\lambda}(t). \tag{5}$$

Using Hoeffding's inequality for a U-statistics (Serfling, 1981), we can concentrate $\widetilde{p}_{r,b,\lambda}(t; D)$ around its expectation as

$$\mathbb{P}\left(\left|\widetilde{p}_{r,b,\lambda}(t; D) - p_{r,b,\lambda}(t)\right| > \frac{\epsilon}{2}\right) \leq 2\exp\left(-\frac{n\epsilon^2}{2b}\right). \tag{6}$$

Now, consider $\widetilde{p}_{r,b,\lambda}(t; D)$ for a fixed set of samples D. We can think of $\widetilde{p}_{r,b,\lambda}(t; D)$ as the expected value of a random variable on a uniform distribution over subsets of size b *i.e.* imagine we have a random variable Y which can take values $F^t_{\lambda,r}(D_b)$ for $D_b \in S_b(D)$, and

$$\mathbb{P}\left(Y = F^t_{\lambda,r}(D_b)\right) = \frac{1}{\binom{n}{b}}, \tag{7}$$

so that $\widetilde{p}_{r,b,\lambda}(t;D) = \mathbb{E}[Y]$. Then, $\widehat{p}_{r,b,\lambda}(t;D)$ is an estimate of $\mathbb{E}[Y]$, computed by averaging N values of Y, chosen independently and uniformly randomly. Using McDiarmid's inequality (McDiarmid, 1989), we can therefore concentrate $\widehat{p}_{r,b,\lambda}(t;D)$ around $\widetilde{p}_{r,b,\lambda}(t;D)$ as

$$\begin{aligned}
&\mathbb{P}\left(|\widehat{p}_{r,b,\lambda}(t;D) - \widetilde{p}_{r,b,\lambda}(t;D)| > \frac{\epsilon}{2} \,\middle|\, D\right) \leq 2\exp\left(-\frac{N\epsilon^2}{2}\right), \\
\Rightarrow\ &\mathbb{P}\left(|\widehat{p}_{r,b,\lambda}(t;D) - \widetilde{p}_{r,b,\lambda}(t;D)| > \frac{\epsilon}{2}\right) \leq 2\exp\left(-\frac{N\epsilon^2}{2}\right),
\end{aligned} \tag{8}$$

where we obtain the second inequality by integrating D out, since the RHS does not depend on D.

Combining Equation (6) and (8), we get

$$\mathbb{P}\Big(|\widehat{p}_{r,b,\lambda}(t;D) - p_{r,b,\lambda}(t)| > \epsilon\Big) \leq 2\exp\left(-\frac{n\epsilon^2}{2b}\right) + 2\exp\left(-\frac{N\epsilon^2}{2}\right). \tag{9}$$

For, $N \geq \lceil \frac{n}{b} \rceil$, this becomes

$$\mathbb{P}\Big(|\widehat{p}_{r,b,\lambda}(t;D) - p_{r,b,\lambda}(t)| > \epsilon\Big) \leq 4\exp\left(-\frac{n\epsilon^2}{2b}\right). \tag{10}$$

Now, by the union bound,

$$\begin{aligned}
\mathbb{P}\Big(\exists\, t \in V \setminus r \text{ s.t. } |\widehat{p}_{r,b,\lambda}(t;D) - p_{r,b,\lambda}(t)| > \epsilon\Big) &\leq 4(p-1)\exp\left(-\frac{n\epsilon^2}{2b}\right) \\
&\leq 4p\exp\left(-\frac{n\epsilon^2}{2b}\right)
\end{aligned} \tag{11}$$

Finally, observe that $\exists t' \in V \setminus r$ s.t.

$$\begin{aligned}
|\widehat{\mathcal{M}}_{r,b,\lambda}(D) - \mathcal{M}_{r,b,\lambda}| &= \left| \max_{t_1 \in V\setminus r} \widehat{p}_{r,b,\lambda}(t_1;D)\,(1-\widehat{p}_{r,b,\lambda}(t_1;D)) - \max_{t_2 \in V\setminus r} p_{r,b,\lambda}(t_2)\,(1-p_{r,b,\lambda}(t_2)) \right| \\
&\leq \left| \widehat{p}_{r,b,\lambda}(t';D)\,(1-\widehat{p}_{r,b,\lambda}(t';D)) - p_{r,b,\lambda}(t')\,(1-p_{r,b,\lambda}(t')) \right| \\
&\leq \left| \widehat{p}_{r,b,\lambda}(t';D) - p_{r,b,\lambda}(t') \right| + \left| (\widehat{p}_{r,b,\lambda}(t';D) - p_{r,b,\lambda}(t'))\,(\widehat{p}_{r,b,\lambda}(t';D) + p_{r,b,\lambda}(t')) \right| \\
&\leq 3|\widehat{p}_{r,b,\lambda}(t';D) - p_{r,b,\lambda}(t')|
\end{aligned} \tag{12}$$

An instance of the t' used in the above set of inequations can be one of t_1^* or t_2^*, corresponding to the optimal for $\left(\arg\max_{t_1 \in V\setminus r} \widehat{p}_{r,b,\lambda}(t_1;D)\,(1-\widehat{p}_{r,b,\lambda}(t_1;D))\right)$ and $\left(\arg\max_{t_2 \in V\setminus r} p_{r,b,\lambda}(t_2)\,(1-p_{r,b,\lambda}(t_2))\right)$ respectively.

Thus,

$$|\widehat{\mathcal{M}}_{r,b,\lambda}(D) - \mathcal{M}_{r,b,\lambda}| > \epsilon \Rightarrow \exists t' \in V \setminus r \text{ s.t. } |\widehat{p}_{r,b,\lambda}(t';D) - p_{r,b,\lambda}(t')| > \epsilon/3 \tag{13}$$

Using the result of Equation (10) now proves the lemma. □

3. Proof of Proposition 2

Proof. Consider any $t \in V \setminus r$. From Assumption 1, we know that

$$\begin{aligned}
\forall\, \lambda \in [0, \lambda_{\min}(t)), &\quad p_{r,b,\lambda}(t) > (1 - 2\exp(-c\log p)) \text{ and,} \\
\forall\, \lambda \in [\lambda_{\min}(t), \lambda_{\max}(t)], &\quad 2\exp(-c\log p) \leq p_{r,b,\lambda}(t) \leq (1 - 2\exp(-c\log p)).
\end{aligned} \tag{14}$$

This implies that

$$\begin{aligned}
\forall\, \lambda \in [0, \lambda_{\min}(t)), &\quad p_{r,b,\lambda}(t)\,(1 - p_{r,b,\lambda}(t)) < \gamma \text{ and,} \\
\forall\, \lambda \in [\lambda_{\min}(t), \lambda_{\max}(t)], &\quad p_{r,b,\lambda}(t)\,(1 - p_{r,b,\lambda}(t)) \geq \gamma.
\end{aligned} \tag{15}$$

Suppose we pick $\lambda'_l = \min_{t \in V \setminus r} \lambda_{\min}(t)$. Then for all $\lambda < \lambda'_l$, $\mathcal{M}_{r,b,\lambda} < \gamma$, and at λ'_l, $\mathcal{M}_{r,b,\lambda'_l} \geq \gamma$. This means that λ'_l is the solution to $\inf\{\lambda \geq 0 : \mathcal{M}_{r,b,\lambda} \geq \gamma\}$. Thus, $\lambda_l = \inf\{\lambda \geq 0 : \mathcal{M}_{r,b,\lambda} \geq \gamma\}$ exists and

$$\lambda_l = \lambda'_l = \min_{t \in V \setminus r} \lambda_{\min}(t). \tag{16}$$

To prove the existence of λ_u, we first have the following claim, the proof of which is described in Subsection 3.1.

Claim 1. *For any node $r \in V$, there exists a regularization parameter λ_s $(0 \leq \lambda_s \leq 1)$ s.t. for all $\lambda > \lambda_s$, $p_{r,b,\lambda}(t) = 0 \,\forall\, t \in V \setminus r$, and as a consequence, $\mathcal{M}_{r,b,\lambda} = 0$.*

Now, observe that $\mathcal{M}_{r,b,\lambda}$ is a continuous function of λ, since $\mathcal{M}_{r,b,\lambda} = \max_{t \in V \setminus r} p_{r,b,\lambda}(t)\,(1 - p_{r,b,\lambda}(t))$ is just a maximum of continuous functions.

So, $\mathcal{M}_{r,b,\lambda_l} \geq \gamma$, $\mathcal{M}_{r,b,\lambda_s} = 0$ (from Claim 1) and the continuity of $\mathcal{M}_{r,b,\lambda}$, together imply that $\lambda_u = \inf\{\lambda > \lambda_l : \mathcal{M}_{r,b,\lambda} < \gamma\}$ exists. Also, we have $\lambda_u \leq \lambda_s$.

Finally, (b) is a consequence of the continuity of $p_{r,b,\lambda}(t)$. From (16), we know that $\lambda_l = \min_{t \in V \setminus r} \lambda_{\min}(t)$. Therefore, at $t' = \arg\min_{t \in V \setminus r} \lambda_{\min}(t)$ we have

$$p_{r,b,\lambda_l}(t') = 1 - 2\exp(-c \log p). \tag{17}$$

Note that equality occurs due to continuity of $p_{r,b,\lambda}(t)$. At λ_u, since $\mathcal{M}_{r,b,\lambda_u} < \gamma$, we must have either $p_{r,b,\lambda_u}(t') > 1 - 2\exp(-c \log p)$ or $p_{r,b,\lambda}(t') < 2\exp(-c \log p)$. This means that either $\lambda_u < \lambda_{\min}(t')$ or $\lambda_u > \lambda_{\max}(t')$. However, since $\lambda_u > \lambda_l = \lambda_{\min}(t')$, we cannot have the former. Thus, $p_{r,b,\lambda_u}(t') < 2\exp(-c \log p)$.

So, to summarize,

$$\begin{aligned} &\text{At } \lambda_l,\ p_{r,b,\lambda_l}(t') = 1 - 2\exp(-c \log p) \text{ and} \\ &\text{at } \lambda_u,\ p_{r,b,\lambda_u}(t') < 2\exp(-c \log p), \end{aligned} \tag{18}$$

i.e. between λ_l and λ_u, $p_{r,b,\lambda}(t')$ goes from a value close to 1, to a value close to 0. Now, continuity of $p_{r,b,\lambda}(t')$ implies that for any $k \in (\gamma, 1/4]$, there exists a λ s.t. $p_{r,b,\lambda}(t')\,(1 - p_{r,b,\lambda}(t')) \geq k$, which implies $\mathcal{M}_{r,b,\lambda} \geq k$. □

3.1. Proof of Claim 1

Proof. Let D be any set of b samples, $D = \{x^{(1)} \ldots, x^{(b)}\}$. Any solution, $\widetilde{\theta}_{\setminus r}$, of (7) (with the samples D) must satisfy

$$\nabla \mathcal{L}(\widetilde{\theta}_{\setminus r}; D) + \lambda z = 0 \tag{19}$$

for some $z \in \partial \|\widetilde{\theta}_{\setminus r}\|_1$.

Suppose we have $\lambda > \|\nabla \mathcal{L}(0; D)\|_\infty$ and we pick $z_i = -[\nabla \mathcal{L}(0; D)]_i / \lambda$. Then, $z \in \partial \|\widetilde{\theta}_{\setminus r}\|_1$ for $\widetilde{\theta}_{\setminus r} = 0$ and $(0, z)$ satisfies (19). Thus, 0 is an optimum for (7). Also, since we have shown the existence of a subgradient z s.t. $\|z\|_\infty < 1$, by Lemma 1 in (Ravikumar et al., 2010) we know that 0 is the only solution. If we pick $\lambda_s = \max_{D \in \{-1,1\}^{pb}} \|\nabla \mathcal{L}(0; D)\|_\infty$, then for any $\lambda > \lambda_s$, 0 is the unique optimum for any choice of D. This implies that $p_{r,b,\lambda}(t) = 0 \,\forall t \in V \setminus r$ and $\mathcal{M}_{r,b,\lambda} = 0$. Finally, note that

$$\|\nabla \mathcal{L}(0; D)\|_\infty = \max_{t \in V \setminus r} \left| \frac{1}{n} \sum_{i=1}^{b} x_r^{(i)} x_t^{(i)} \right| \leq 1 \Rightarrow \lambda_s \leq 1 \tag{20}$$

□

4. Proof of Proposition 4

Proof. Consider any $t \in V \setminus r$. We have

$$\text{Either } \ \lambda_u < \lambda_{\min}(t) \ \text{ or } \ \lambda_u > \lambda_{\max}(t). \tag{21}$$

This can be seen as at λ_u, we have $\mathcal{M}_{r,b,\lambda_u} > \gamma = 2\exp(-c\log p)\left(1 - 2\exp(-c\log p)\right)$. This implies that

$$\text{Either } \; p_{r,b,\lambda_u}(t) > 1 - 2\exp(-c\log p) \;\text{ or }\; p_{r,b,\lambda_u}(t) < 2\exp(-c\log p). \tag{22}$$

Based on Assumption 1(a), this implies equation (21).

Now, consider this for any two irrelevant variables $t_1, t_2 \notin \mathcal{N}^*(r)$. We cannot have $\lambda_u < \lambda_{\min}(t_1)$ and $\lambda_u > \lambda_{\max}(t_2)$ (or vice-versa), as this would violate Assumption 1(b). Thus, we must have

$$\text{Either } \; \lambda_u < \min_{t \notin \mathcal{N}^*(r)} \lambda_{\min}(t) \;\text{ or }\; \lambda_u > \max_{t \notin \mathcal{N}^*(r)} \lambda_{\max}(t). \tag{23}$$

We shall show that the former possibility cannot happen. To see this, assume $\lambda_u < \min_{t \notin \mathcal{N}^*(r)} \lambda_{\min}(t)$. Then, using Assumption 1(c), this means that $\lambda_u < \lambda_{\max}(\tilde{t})$, for any $\tilde{t} \in V \setminus r$. But, from (21), this must imply that $\lambda_u < \lambda_{\min}\left(\tilde{t}\right)$, for any $\tilde{t} \in V \setminus r$. However, this is a contradiction, since $\lambda_u > \lambda_l = \min_{t \in V \setminus r} \lambda_{\min}(t)$, where the equality comes through the same argument used to show (16).

Thus, $\lambda_u > \max_{t \notin \mathcal{N}^*(r)} \lambda_{\max}(t)$. This implies that $p_{r,b,\lambda_u}(t) < 2\exp(-c\log p)$ for any $t \notin \mathcal{N}^*(r)$ *i.e.*

$$\text{For any } \; t \notin \mathcal{N}^*(r), \; \mathbb{P}\left(t \notin \widehat{\mathcal{N}}_{b,\lambda_u}(r; D)\right) \geq 1 - 2\exp(-c\log p). \tag{24}$$

Using union bound on the irrelevant variables, we get that $\mathbb{P}\left(\widehat{\mathcal{N}}_{b,\lambda_u}(r; D) \subseteq \mathcal{N}^*(r)\right) \geq 1 - 2\exp\left(-(c-1)\log p\right)$. □

5. Proof of Proposition 3

Proof. Following the same argument as in Proposition 4 above, we can infer that for any $t \notin \mathcal{N}^*(r)$, $p_{r,b,\lambda_u}(t) < 2\exp(-c\log p)$.

Using Corollary 1, we know that there exists a λ_0 s.t.

$$\begin{aligned} &p_{r,b,\lambda_0}(t) \geq 1 - 2\exp(-c_1 c_4 \log p) > 1 - 2\exp(-c\log p) \quad \text{ if } t \in \mathcal{N}^*_{sub}(r) \\ &p_{r,b,\lambda_0}(t) \leq 2\exp(-c_1 c_4 \log p) < 2\exp(-c\log p) \quad \text{ if } t \notin \mathcal{N}^*_{sub}(r). \end{aligned} \tag{25}$$

Based on Assumption 1, this means for any $t \in \mathcal{N}^*_{sub}(r)$ we have $\lambda_0 < \lambda_{\min}(t)$, and for any $t \notin \mathcal{N}^*_{sub}(r)$ we have $\lambda_0 > \lambda_{\max}(t)$.

Observe that $\lambda_0 > \lambda_l$. This is because for any $t' \notin \mathcal{N}^*_{sub}(r)$, $\lambda_0 > \lambda_{\max}(t')$ which implies $\lambda_0 > \lambda_{\min}(t')$, whereas $\lambda_l = \min_{t'' \in V \setminus r} \lambda_{\min}(t'')$, using arguments used to show (16).

Now, we shall show that we cannot have $\lambda_0 < \lambda_u$. Suppose $\lambda_0 < \lambda_u$. From (25), we have that $\mathcal{M}_{r,b,\lambda_0} < \gamma$, where γ is as defined in Assumption 1. So, we get $\lambda_0 \in (\lambda_l, \lambda_u)$ s.t. $\mathcal{M}_{r,b,\lambda_0} < \gamma$. This is a contradiction since $\lambda_u = \inf\left\{\lambda > \lambda_l : \mathcal{M}_{r,b,\lambda} < \gamma\right\}$. Therefore, we must have $\lambda_u \leq \lambda_0$.

So, for any $t \in \mathcal{N}^*_{sub}(r)$, $\lambda_u < \lambda_{\min}(t)$, which means that $p_{r,b,\lambda_u}(t) > 1 - 2\exp(-c\log p)$. Now, taking a union bound over the exclusion of all irrelevant variables and the inclusion of all variables in $\mathcal{N}^*_{sub}(r)$ proves the proposition. □

6. Proof of Theorem 2

Since this is a simple corollary, we shall only provide an outline of the proof here. The conditions specified in the theorem ensure that Proposition 3 is true for any node $r \in V$ with degree, $d(r) \leq d$, and that, Proposition 4 is true for any other node. In addition, owing to the choice of n and N, Proposition 2 guarantees that $\widehat{\mathcal{M}}_{r,b,\lambda}$ would be reliable estimate for $\mathcal{M}_{r,b,\lambda}$ upto a tolerance of ϵ w.h.p. Thus, running Algorithm 2, with the parameters specified, for all nodes would yield the $\mathcal{N}^*_{sub}(r)$ neighbourhoods of nodes with degree at most d, and yield subsets of the true neighbourhoods for the rest. E_d is defined to be the set of edges (u, v) such that atleast one of its endpoints is a node with degree at most d (say u), and the other belongs to the $\mathcal{N}^*_{sub}$ neighbourhood of the first (*i.e.* $v \in \mathcal{N}^*_{sub}(u)$). Then, if we consider the union of all neighbourhoods obtained from Algorithm 2, clearly, the set E_d gets recovered with high probability.

7. Proof of Corollary 2

This is again a simple consequence of Theorem 2. Under the conditions specified here, the set E_d, defined in Theorem 2, becomes the set of true edges E^*. Thus, we are guaranteed exact graph recovery in this setting.

References

Liu, Han, Roeder, Kathryn, and Wasserman, Larry A. Stability approach to regularization selection (stars) for high dimensional graphical models. In *NIPS*, pp. 1432–1440, 2010.

McDiarmid, C. On the method of bounded differences. In *Surveys in Combinatorics*, number 141 in London Mathematical Society Lecture Note Series, pp. 148–188. Cambridge University Press, August 1989.

Ravikumar, P., Wainwright, M. J., and Lafferty, J. High-dimensional ising model selection using ℓ_1-regularized logistic regression. *Annals of Statistics*, 38(3):1287–1319, 2010.

Serfling, Robert J. *Approximation Theorems of Mathematical Statistics*. Wiley-Interscience, 1981.