Supplement to "Consistency of community detection in networks under degree-corrected stochastic block models"

Yunpeng Zhao[1], Elizaveta Levina[2], and Ji Zhu[2]

[1]Department of Statistics, George Mason University, Fairfax, VA 22030
[2]Department of Statistics, University of Michigan, Ann Arbor, MI 48109

This supplement contains the proofs of Lemma A.1 and Lemma A.2 stated in the Appendix of [3].

Lemma A.1. *Let* $\|X\|_\infty = \max_{kl} |X_{kl}|$ *and* $|\boldsymbol{e} - \boldsymbol{c}| = \sum_{i=1}^n I(e_i \neq c_i)$. *Then*

$$\mathbb{P}(\max_{\boldsymbol{e}} \|X(\boldsymbol{e})\|_\infty \geq \epsilon) \leq 2K^{n+2} \exp\left(-\frac{1}{8C}\epsilon^2 \mu_n\right), \tag{1}$$

for $\epsilon < 3C$, *where* $C = \max\{x_u x_v P_{ab}\}$.

$$\mathbb{P}(\max_{|\boldsymbol{e}-\boldsymbol{c}|\leq m} \|X(\boldsymbol{e}) - X(\boldsymbol{c})\|_\infty \geq \epsilon) \leq 2\binom{n}{m} K^{m+2} \exp\left(-\frac{3}{8}\epsilon\mu_n\right) \ , \tag{2}$$

for $\epsilon \geq 6Cm/n$.

$$\mathbb{P}(\max_{|\boldsymbol{e}-\boldsymbol{c}|\leq m} \|X(\boldsymbol{e}) - X(\boldsymbol{c})\|_\infty \geq \epsilon) \leq 2\binom{n}{m} K^{m+2} \exp\left(-\frac{n}{16mC}\epsilon^2\mu_n\right) \ , \tag{3}$$

for $\epsilon < 6Cm/n$.

Proof. This lemma is similar to Lemma 1.1 of [1], but since the proof in [1] contains some relatively minor errors, we give a full proof here for completeness. First note that in order to prove (1), it is sufficient to show

$$\mathbb{P}(|X_{kl}(\boldsymbol{e})| \geq \epsilon | \boldsymbol{c}, \boldsymbol{\theta}) \leq 2\exp\left(-\frac{1}{8C}\epsilon^2\mu_n\right) \ , \tag{4}$$

where

$$X_{kl}(\boldsymbol{e}) = \frac{1}{\mu_n}[O_{kl}(\boldsymbol{e}) - \mathbb{E}(O_{kl}(\boldsymbol{e})|\boldsymbol{c}, \boldsymbol{\theta})] \ ,$$
$$O_{kl}(\boldsymbol{e}) = \sum_{i=1}^n A_{ii} I(e_i = k, e_j = l) + 2\sum_{i<j}^n A_{ij} I(e_i = k, e_j = l) \ .$$

The proof relies on Bernstein's inequality (see e.g., [2]): If Y_i are independent, $|Y_i| \le M, EY_i = 0, S_I = \sum_{i=1}^{I} Y_i$, then

$$\mathbb{P}(|S_I| \ge w) \le 2\exp\left(-\frac{w^2/2}{\mathrm{Var}(S_I) + Bw/3}\right). \tag{5}$$

Note that conditioning on $\boldsymbol{c}, \boldsymbol{\theta}$, A_{ij} are independent and $|A_{ij}| \le 1$. Let $B = 2$, and then (5) becomes

$$\mathbb{P}(|\mu_n X_{kl}(\boldsymbol{e})| \ge w | \boldsymbol{c}, \boldsymbol{\theta}) \le 2\exp\left(-\frac{w^2/2}{\mathrm{Var}(O_{kl}|\boldsymbol{c}, \boldsymbol{\theta}) + 2w/3}\right). \tag{6}$$

In order to compare two terms in the denominator, we need to evaluate $\mathrm{Var}(O_{kl}|\boldsymbol{c}, \boldsymbol{\theta})$:

$$\begin{aligned} \mathrm{Var}(A_{ij}|\boldsymbol{c}, \boldsymbol{\theta}) &= \rho_n \theta_i \theta_j P_{c_i c_j} - (\rho_n \theta_i \theta_j P_{c_i c_j})^2 \le \rho_n C \ , \\ \mathrm{Var}(O_{kl}|\boldsymbol{c}, \boldsymbol{\theta}) &\le (n + 4(n-1)n/2)\rho_n C \le 2n^2 \rho_n C \ . \end{aligned}$$

Let $w = \epsilon \mu_n = \epsilon n^2 \rho_n$, for $\epsilon < 3C$,

$$\begin{aligned} \mathbb{P}(|X_{kl}(\boldsymbol{e})| \ge \epsilon | \boldsymbol{c}, \boldsymbol{\theta}) &\le 2\exp\left(-\frac{w^2/2}{\mathrm{Var}(O_{kl}|\boldsymbol{c}, \boldsymbol{\theta}) + 2w/3}\right) \\ &\le 2\exp\left(-\frac{\epsilon^2 n^4 \rho_n^2}{8n^2 \rho_n C}\right) = 2\exp\left(-\frac{1}{8C}\epsilon^2 \mu_n\right) \ . \end{aligned}$$

We now prove (2) and (3). If $e_{m+1} = c_{m+1}, ..., e_n = c_n$,

$$\begin{aligned} O_{kl}(\boldsymbol{e}) - O_{kl}(\boldsymbol{c}) =& \sum_{i=1}^{m} (A_{ii} I(e_i = k, e_i = l) - A_{ii} I(c_i = k, c_i = l)) \\ &+ 2\sum_{i<j}^{m} (A_{ij} I(e_i = k, e_j = l) - A_{ij} I(c_i = k, c_j = l)) \\ &+ 2\sum_{i=1}^{m} \sum_{j=m+1}^{n} (A_{ij} I(e_i = k, e_j = l) - A_{ij} I(c_i = k, c_j = l)) \ . \\ var(O_{kl}(\boldsymbol{e}) - O_{kl}(\boldsymbol{c})|\boldsymbol{c}, \boldsymbol{\theta}) \le& [m + 4(m(m-1)/2 + m(n-m))]\rho_n C \le 4mn\rho_n C \ . \end{aligned}$$

We again apply (5). For $|\boldsymbol{e} - \boldsymbol{c}| \le m$, $\epsilon \ge 6Cm/n$,

$$\begin{aligned} \mathbb{P}(|X_{kl}(\boldsymbol{e}) - X_{kl}(\boldsymbol{c})| \ge \epsilon | \boldsymbol{c}, \boldsymbol{\theta}) &\le 2\exp\left(-\frac{(\epsilon n^2 \rho_n)^2/2}{4mn\rho_n C + 2\epsilon n^2 \rho_n/3}\right) \\ &\le 2\exp\left(-\frac{3}{8}\epsilon \mu_n\right). \end{aligned}$$

For $\epsilon < 6Cm/n$,

$$\begin{aligned} \mathbb{P}(|X_{kl}(\boldsymbol{e}) - X_{kl}(\boldsymbol{c})| \ge \epsilon | \boldsymbol{c}, \boldsymbol{\theta}) &\le 2\exp\left(-\frac{(\epsilon n^2 \rho_n)^2/2}{4mn\rho_n C + 2\epsilon n^2 \rho_n/3}\right) \\ &\le 2\exp\left(-\frac{n}{16mC}\epsilon^2 \mu_n\right). \end{aligned}$$

□

Lemma A.2. *Let g, P, S be $K \times K$ matrices with nonnegative entries. Assume that*
a) P and g are symmetric;
b) P does not have two identical columns;
c) there exists at least one nonzero entry in each column of S;
d) for $1 \le k, l, a, b \le K, g_{kl} = P_{ab}$ whenever $S_{ka}S_{lb} > 0$.
Then S is a diagonal matrix or a row/column permutation of a diagonal matrix.

Proof. The proof is similar to Lemma 3.2 [1].

1) If there exists a permutation of the rows and columns of S such that its diagonals are all positive after permutation, i.e., $S_{bb} > 0$ for $b = 1, ..., K$. If S is not diagnonal, there exists $k \neq a$ such that $S_{ka} > 0$. For $b = 1, ..., K$,

$$\begin{aligned} S_{ka}S_{bb} > 0 &\Rightarrow g_{kb} = P_{ab}, \\ S_{kk}S_{bb} > 0 &\Rightarrow g_{kb} = P_{kb}, \\ &\Rightarrow P_{ab} = P_{kb}. \end{aligned}$$

This contradicts with b).

2) If there does not exist such a permutation, then we can always permute row and columns of S, such that for some $m \ge 1$, $S_{ij} = 0$ for $1 \le i, j \le m$, and $S_{bb} > 0$ for $b = m+1, ..., K$. By c), there exists $S_{k_i i} > 0$ for $i = 1, ..., m$ and some $k_i \in \{m+1, ..., K\}$. Then

$$\begin{aligned} S_{k_i i}S_{k_1 1} > 0 &\Rightarrow g_{k_i k_1} = P_{i1} = P_{1i}, \text{ for } i = 1, ..., m. \\ S_{k_i i}S_{k_1 k_1} > 0 &\Rightarrow g_{k_i k_1} = P_{ik_1} = P_{k_1 i}, \text{ for } i = 1, ..., m. \\ &\Rightarrow P_{1i} = P_{k_1 i}, \text{ for } i = 1, ..., m. \qquad (7) \\ S_{k_1 1}S_{bb} > 0 &\Rightarrow g_{k_1 b} = P_{1b}, \text{ for } b = m+1, ..., K. \\ S_{k_1 k_1}S_{bb} > 0 &\Rightarrow g_{k_1 b} = P_{k_1 b}, \text{ for } b = m+1, ..., K. \\ &\Rightarrow P_{1b} = P_{k_1 b}, \text{ for } b = m+1, ..., K. \qquad (8) \end{aligned}$$

(7) and (8) contradict with b). □

References

[1] P. J. Bickel and A. Chen. A nonparametric view of network models and Newman-Girvan and other modularities. *Proc. Natl. Acad. Sci. USA*, 106:21068–21073, 2009.

[2] G. R. Shorack and J. A. Wellner. *Empirical Processes with Applications to Statistics.* Wiley, New York, 1986.

[3] Y. Zhao, E. Levina, and J. Zhu. Consistency of community detection in networks under degree-corrected stochastic block models. *Annals of Statistics*, 2012. arxiv.org/1110.3854.