# Web-based Supplementary Materials for "Logistic Regression Augmented Community Detection for Networks with Application in Identifying Autism-Related Gene Pathways"

Yunpeng Zhao[1], Qing Pan[2] and Chengan Du[1]
[1]Department of Statistics, George Mason University
[2]Department of Statistics, George Washington University

June 8, 2017

## 1 Web Appendix A: Proof of Lemma 1

Recall that $\hat{\beta}_0$ and $\hat{\beta}_1$ can be obtained by

$$(\hat{\beta}_0, \hat{\beta}_1) = \arg\max_{\beta_0,\beta_1} \sum_{i=1}^{n} \left\{ y_i(\beta_0 + x_i\beta_1) - \log(1 + e^{\beta_0 + x_i\beta_1}) \right\}.$$

Taking derivative of the log-likelihood above with respect to $\beta_0$ and $\beta_1$, we obtain

$$\sum_{i=1}^{n} \frac{e^{x_i\hat{\beta}_1+\hat{\beta}_0}}{1+e^{x_i\hat{\beta}_1+\hat{\beta}_0}} = \sum_{i=1}^{n} y_i, \tag{1.1}$$

$$\sum_{i=1}^{n} \frac{x_i e^{x_i\hat{\beta}_1+\hat{\beta}_0}}{1+e^{x_i\hat{\beta}_1+\hat{\beta}_0}} = \sum_{i=1}^{n} y_i x_i. \tag{1.2}$$

Denote $s = (1/n)\sum_{i=1}^{n} y_i$. Then $s$ is constant for $\boldsymbol{e} \in \mathcal{E}$, since $s = n_1\gamma_1 + n_2(1-\gamma_2)$. By the defintion of Riemann integral, for sufficiently large $n$,

$$(1-\epsilon)s \le \int_0^1 \frac{e^{x\hat{\beta}_1+\hat{\beta}_0}}{1+e^{x\hat{\beta}_1+\hat{\beta}_0}} dx \le (1+\epsilon)s. \tag{1.3}$$

Without loss of generality, we assume $\hat{\beta}_1 \neq 0$, since it is easy to show that $\hat{\beta}_0$ is bounded from (1.2) otherwise.

Under this assumption, the integral in (1.3) has a closed form:

$$\int_0^1 \frac{e^{x\hat{\beta}_1+\hat{\beta}_0}}{1+e^{x\hat{\beta}_1+\hat{\beta}_0}} dx = \frac{1}{\hat{\beta}_1}\{\log(1+e^{\hat{\beta}_0+\hat{\beta}_1}) - \log(1+e^{\hat{\beta}_0})\}.$$

First we consider the case that $\hat{\beta}_1 > 0$. According to (1.3),

$$\log \frac{e^{s(1-\epsilon)\hat{\beta}_1} - 1}{e^{\hat{\beta}_1} - e^{s(1-\epsilon)\hat{\beta}_1}} \leq \hat{\beta}_0 \leq \log \frac{e^{s(1+\epsilon)\hat{\beta}_1} - 1}{e^{\hat{\beta}_1} - e^{s(1+\epsilon)\hat{\beta}_1}}. \tag{1.4}$$

By (1.4), it is easy to check that

$$\lim_{\hat{\beta}_1 \to +\infty} e^{x\hat{\beta}_1 + \hat{\beta}_0} \geq \lim_{\hat{\beta}_1 \to +\infty} \frac{e^{(x+s(1-\epsilon))\hat{\beta}_1} - e^{x\hat{\beta}_1}}{e^{\hat{\beta}_1} - e^{s(1-\epsilon)\hat{\beta}_1}} = \begin{cases} +\infty & \text{if } x > 1 - s(1-\epsilon), \\ 0 & \text{if } x < 1 - s(1-\epsilon). \end{cases}$$

Therefore, for sufficiently large $n$,

$$\lim_{\hat{\beta}_1 \to +\infty} \frac{1}{n} \sum_{i=1}^{n} \frac{x_i e^{x_i\hat{\beta}_1 + \hat{\beta}_0}}{1 + e^{x_i\hat{\beta}_1 + \hat{\beta}_0}} \geq \lim_{\hat{\beta}_1 \to +\infty} (1-\epsilon) \int_0^1 \frac{x e^{x\hat{\beta}_1 + \hat{\beta}_0}}{1 + e^{x\hat{\beta}_1 + \hat{\beta}_0}} dx \geq (1-\epsilon) \int_{1-s(1-\epsilon)}^{1} x dx. \tag{1.5}$$

However,

$$\max_{e \in \mathcal{F} \cap \mathcal{E}} \frac{1}{n} \sum_{i=1}^{n} y_i x_i \leq \frac{1}{n} \sum_{i=n-\hat{n}_1\tilde{\gamma}_1+1}^{n} x_i + \frac{1}{n} \sum_{i=\hat{n}_2-\hat{n}_1(1-\tilde{\gamma}_1)+1}^{\hat{n}_2} x_i, \tag{1.6}$$

the right hand side of (1.6) converges to

$$\int_{1-\tilde{\gamma}_1 s}^{1} x dx + \int_{1-s-s(1-\tilde{\gamma}_1)}^{1-s} x dx, \tag{1.7}$$

and thus it is strictly less than (1.5). Therefore, there exists $M_1$ such that $\hat{\beta}_1 < M_1$ for sufficiently large $n$. Note that $M_1$ only depends on (1.7), and hence is independent with $n$.

Similarly, when $\hat{\beta}_1 < 0$,

$$\log \frac{e^{s(1+\epsilon)\hat{\beta}_1} - 1}{e^{\hat{\beta}_1} - e^{s(1+\epsilon)\hat{\beta}_1}} \leq \hat{\beta}_0 \leq \log \frac{e^{s(1-\epsilon)\hat{\beta}_1} - 1}{e^{\hat{\beta}_1} - e^{s(1-\epsilon)\hat{\beta}_1}}. \tag{1.8}$$

For sufficiently large $n$,

$$\lim_{\hat{\beta}_1 \to -\infty} \frac{1}{n} \sum_{i=1}^{n} \frac{x_i e^{x_i\hat{\beta}_1 + \hat{\beta}_0}}{1 + e^{x_i\hat{\beta}_1 + \hat{\beta}_0}} \leq \lim_{\hat{\beta}_1 \to -\infty} (1+\epsilon) \int_0^1 \frac{x e^{x\hat{\beta}_1 + \hat{\beta}_0}}{1 + e^{x\hat{\beta}_1 + \hat{\beta}_0}} dx \leq (1+\epsilon) \int_0^s x dx.$$

But

$$\min_{e \in \mathcal{F} \cap \mathcal{E}} \frac{1}{n} \sum_{i=1}^{n} y_i x_i \geq \frac{1}{n} \sum_{i=1}^{\tilde{\gamma}_2\hat{n}_1} x_i + \frac{1}{n} \sum_{i=\hat{n}_1+1}^{\hat{n}_1+(1-\tilde{\gamma}_2)\hat{n}_1} x_i \to \int_0^{\tilde{\gamma}_2 s} x dx + \int_s^{s+(1-\tilde{\gamma}_2)s} x dx.$$

which implies $\hat{\beta}_1 > -M_2$ for a fixed positive value of $M_2$. It implies the solution $(\hat{\beta}_0, \hat{\beta}_1)$ for (1.1) and (1.2) is bounded together with (1.4) and (1.8).

## 2 Web Appendix B: Proof of Theorem 1

With Lemma 1, the proof of Theorem 1 is closely followed the proof of Proposition 1 and Theorem 3 in (Amini et al., 2013). We give the details for completeness. We begin with notation. Recall that confusion matrix $R$ defined as $R_{ka} = (1/n) \sum_{i=1}^{n} 1(e_i = k, c_i = a)$ is constant in $\mathcal{E}$ and is given by

$$R = \begin{pmatrix} \gamma_1 \pi_1 & (1-\gamma_2)\pi_2 \\ (1-\gamma_1)\pi_1 & \gamma_2 \pi_2 \end{pmatrix}.$$

Let $\tau = \pi_2/\pi_1$ and define

$$u(x) = \frac{(1-\gamma_1)x + \gamma_2 \tau}{\gamma_1 x + (1-\gamma_2)\tau}, \quad v(x) = u(\frac{1}{x}),$$

and

$$F_1(x,y) = \log \frac{1+u(x)}{1+v(y)}, \quad F_2(x,y) = \log \frac{1+[u(x)]^{-1}}{1+[v(y)]^{-1}}.$$

Define the KullbackC-Leibler divergence of two Bernoulli distribution with success rates $p$ and $q$ respectively as

$$D(p||q) = p \log \frac{p}{q} + (1-p) \log \frac{1-p}{1-q}.$$

In addition to the conditions listed in the main text of Theorem 2, we need the following regularity condition: There exists $\delta \in (0,1)$ such that

$$\frac{\tau}{\rho_1}(1+\delta) \le \frac{D(\gamma_1||(1-\gamma_2)))}{D((1-\gamma_2)||\gamma_1)} \le (1-\delta)\rho_2 \tau.$$

Let $\mathcal{C}_\ell$ be the set of nodes in true community $\ell$, and $\mathcal{S}_k$ be the set of nodes in community $k$ according to initial labeling $e$. We set $n_\ell = |\mathcal{C}_\ell|, \hat{n}_k = |\mathcal{S}_k|$, and $\mathcal{S}_{k\ell} = \mathcal{S}_k \bigcap \mathcal{C}_\ell$.

Now we consider $i \in \mathcal{C}_1$. Then $\hat{c}_i(\boldsymbol{e}) = 1$ if

$$b_{i1} \log \frac{\hat{\theta}_{21}}{\hat{\theta}_{11}} + b_{i2} \log \frac{\hat{\theta}_{22}}{\hat{\theta}_{12}} < \log \frac{\hat{\pi}_{i1}}{1-\hat{\pi}_{i1}}.$$

Let $\hat{\pi}_{(1)}$ be the smallest value of $\hat{\pi}_{i1}$ $(i = 1, ..., n)$. Define

$$\begin{aligned} \alpha_1 &= \log \frac{\hat{\theta}_{21}}{\hat{\theta}_{11}}, \quad \alpha_2 = \log \frac{\hat{\theta}_{22}}{\hat{\theta}_{12}}, \\ \sigma_j(\boldsymbol{e}) &= \alpha_1 1\{e_j = 1\} + \alpha_2 1\{e_j = 2\}, \quad (j = 1, ..., n) \\ \hat{\tau}_{(1)} &= \frac{1-\hat{\pi}_{(1)}}{\hat{\pi}_{(1)}}. \end{aligned}$$

So that $\alpha_1 b_{i1} + \alpha_2 b_{i2} = \sum_j A_{ij}\sigma_j(\boldsymbol{e}) = \xi_i\{\sigma(\boldsymbol{e})\}$. Thus, the mis-match ratio over class 1 (with identity permutation) is,

$$\begin{aligned}
M_{n,1}(\boldsymbol{e}) &= (1/n_1)\sum_{i\in\mathcal{C}_1} 1\{\hat{c}_i(\boldsymbol{e}) \neq 1\} \\
&\leq (1/n_1)\sum_{i\in\mathcal{C}_1} 1\{\alpha_1 b_{i1} + \alpha_2 b_{i2} \geq \log\frac{\hat{\pi}_{i1}}{1-\hat{\pi}_{i1}}\} \\
&\leq (1/n_1)\sum_{i\in\mathcal{C}_1} 1\{\alpha_1 b_{i1} + \alpha_2 b_{i2} \geq -\log\hat{\tau}_{(1)}\}
\end{aligned}$$

By Bernstein inequality, we have

$$\mathrm{pr}[\xi_i(\sigma) \geq E\{\xi_i(\sigma)\} + t] \leq \exp\left\{-\frac{t^2/2}{\sum_j \mathrm{var}(A_{ij}\sigma_j) + \|\alpha\|_\infty t/3}\right\},$$

where $\|\alpha\|_\infty := \max|\alpha_1|, |\alpha_2|$ and we have used that $|\widetilde{A}_{ij}\sigma_j| \leq \|\alpha\|_\infty$ since $i \in \mathcal{C}_1$, then we have

$$\begin{aligned}
E[\xi_i(\sigma)] &= \sum_j \sigma_j E[A_{ij}] = \sum_{k=1}^2\sum_{\ell=1}^2\sum_j \sigma_j E[A_{ij}] 1\{j \in \mathcal{S}_{k\ell}\} \\
&= \sum_{k=1}^2\sum_{\ell=1}^2\sum_j \alpha_k P_{1\ell} 1\{j \in \mathcal{S}_{k\ell}\} = n[\alpha^T RP]_1 = [\Lambda\alpha]_1.
\end{aligned}$$

In which $[\Lambda\alpha]_1$ denotes the value for the first row of $\Lambda\alpha$, so is $[\alpha^T RP]_1$. By a similar argument,

$$\begin{aligned}
\sum_j \mathrm{var}(A_{ij}\sigma_j) &= \sum_j \sigma_j^2 \mathrm{var}[A_{ij}] \\
&\leq \sum_j \sigma_j^2 E[A_{ij}] \leq \|\alpha\|_\infty \sum_j |\sigma_j| E[A_{ij}] = \|\alpha\|_\infty [\Lambda|\alpha|]_1,
\end{aligned}$$

where $|\alpha| = (|\alpha_1|, |\alpha_2|)$. Combining what we've got above, we have

$$\mathrm{pr}[\xi_i(\sigma) \geq E\{\xi_i(\sigma)\} + t] \leq \exp\left[-\frac{t^2}{2\|\alpha\|_\infty \{[\Lambda|\alpha|]_1 + t/3\}}\right].$$

Take $t = z_{1,n} = -[\Lambda\alpha]_1 - \log\hat{\tau}_{(1)}$. We now show that $-[\Lambda\alpha]_1 \to \infty$ and by Lemma 1 we can conclude $z_{1,n} > 0$.

We first consider the extreme case that $\hat{\rho}_1 = \hat{\rho}_2 = \infty$. Hence we have $u(\infty) = (1-\gamma_1)/\gamma_1$, $v(\infty) = \gamma_2/(1-\gamma_2)$, $\alpha_1 = \log\{(1-\gamma_2)/\gamma_1\}$ and $\alpha_2 = \log\{\gamma_2/(1-\gamma_1)\}$. By definition of $\Lambda$,

$$\Lambda\alpha = b\pi_1 \begin{pmatrix} \rho_1 & 1 \\ 1 & \rho_2 \end{pmatrix} \begin{pmatrix} \gamma_1 & 1-\gamma_1 \\ (1-\gamma_2)\tau & \gamma_2\tau \end{pmatrix} \begin{pmatrix} \log\frac{1-\gamma_2}{\gamma_1} \\ \log\frac{\gamma_2}{1-\gamma_1} \end{pmatrix}$$

$$= b\pi_1 \begin{pmatrix} \rho_1 & \tau \\ 1 & \rho_2\tau \end{pmatrix} \begin{pmatrix} \gamma_1 & 1-\gamma_1 \\ (1-\gamma_2) & \gamma_2 \end{pmatrix} \begin{pmatrix} \log \frac{1-\gamma_2}{\gamma_1} \\ \log \frac{\gamma_2}{1-\gamma_1} \end{pmatrix}$$

$$= b\pi_1 \begin{pmatrix} \rho_1 & \tau \\ 1 & \rho_2\tau \end{pmatrix} \begin{pmatrix} -D(\gamma_1||(1-\gamma_2)) \\ D((1-\gamma_2)||\gamma_1) \end{pmatrix}.$$

So $[\Lambda\alpha]_1$ has the form

$$[\Lambda\alpha]_1 = b[\pi_1\{\tau D((1-\gamma_2)||\gamma_1) - \rho_1 D(\gamma_1||(1-\gamma_2))\}],$$

Since $\gamma_1, \gamma_2 \neq 1/2$ and

$$\frac{\tau}{\rho_1}(1+\delta) \leq \frac{D(\gamma_1||(1-\gamma_2)))}{D((1-\gamma_2)||\gamma_1)} \leq (1-\delta)\rho_2\tau,$$

it is easy to see that $[\Lambda\alpha]_1 < 0$ when $\hat{\rho}_1 = \hat{\rho}_2 = \infty$. And therefore it is also true for sufficiently large $\hat{\rho}_1$ and $\hat{\rho}_2$. Moreover, $[\Lambda\alpha]_1 \to -\infty$ when $b \to \infty$. And we can have similar result of $[\Lambda\alpha]_2 \to \infty$.

In addition, for sufficiently large value of $n$, $[\Lambda\alpha]_1 \leq 3\left\|\alpha\right\|_\infty [\Lambda|\alpha|]_1$.

Putting pieces together, we have

$$\mathrm{pr}[\xi_i(\sigma) \geq -\log\hat{\tau}_{(1)}] \leq \exp\left\{-\frac{z_{1,n}^2}{4\left\|\alpha\right\|_\infty (\Lambda|\alpha|)_1}\right\}. \tag{2.1}$$

Pick $u_n^1$ satisfying

$$u_n^1 \log u_n^1 = \frac{2C}{e\pi_1\bar{p}_1\{\log\hat{\tau}_{(1)}\}},$$

where $\bar{p}_1\{\log\hat{\tau}_{(1)}\} = \frac{1}{n_1}\sum_{i=1}^{n_1} \mathrm{pr}[\xi_i(\sigma) \geq -\hat{\tau}_{(1)}]$. We have

$$\mathrm{pr}[\sup_{e\in\mathcal{E}} M_{n,1}(\boldsymbol{e}) > \frac{1}{\pi_1}\frac{2C}{\log u_n^1}] \leq \exp\{-n(C-r_n)\},$$

by the same arguments in the supplement material of Amini et al. (2013), where $C$ is a constant and $r_n = o(1/n)$.

The right hand side of (2.1) goes to 0 as $b \to \infty$. Therefore, $\log u_n^1 \to \infty$, which implies for any $\epsilon > 0$,

$$\mathrm{pr}[\sup_{e\in\mathcal{E}} M_{n,1}(\boldsymbol{e}) > \epsilon] \to 0, \text{ as } n \to \infty. \tag{2.2}$$

By a similar argument as above, for $i \in \mathcal{C}_2$,

$$\mathrm{pr}[\sup_{e\in\mathcal{E}} M_{n,2}(\boldsymbol{e}) > \epsilon] \to 0, \text{ as } n \to \infty, \tag{2.3}$$

where $M_{n,2}(\boldsymbol{e}) = (1/n_2)\sum_{i\in\mathcal{C}_2} 1\{\hat{c}_i(\boldsymbol{e}) \neq 2\}$. The result of the theorem will automatically follows by putting (2.2) and (2.3) together, i.e., $M_n(\boldsymbol{e}) = \pi_1 M_{n,1}(\boldsymbol{e}) + \pi_2 M_{n,2}(\boldsymbol{e})$. This competes our proof to the theorem.

# References

Amini, A., Chen, A., Bickel, P., and Levina, E. (2013). Pseudo-likelihood methods for community detection in large sparse networks. Annals of Statistics **41,** 2097–2122.