A HYPOTHESIS TESTING FRAMEWORK FOR MODULARITY BASED NETWORK COMMUNITY DETECTION

Jingfei Zhang and Yuguo Chen

University of Miami and University of Illinois at Urbana-Champaign

Supplementary Material

S1 Proof of Theorem 1

We first state a theorem from McKay (1985) on the asymptotic number of simple graphs with restricted edges. Consider a simple graph $G(V,E)$ with m edges and degree sequence $\boldsymbol{d}=(d_1,\dots,d_n)$. Let D be an $n\times n$ symmetric zero-one matrix with a zero diagonal that specifies the set of edges that are not allowed, i.e., $D_{ij}=D_{ji}=1$ if an edge between node i and node j is forbidden, and $D_{ij}=D_{ji}=0$ otherwise. Let $\lambda=\sum_{i=1}^n d_i(d_i-1)/(4m)$, $\gamma=\sum_{i<j,D_{ij}=1}\frac{d_id_j}{2m}$, $d_{max}=\max_i d_i$, and $\tilde{d}=2+d_{max}(1.5d_{max}+x_{max}+1)$, where x_{max} is the maximum column sum of D. Then we have the following theorem.

Theorem. *(McKay, 1985) Suppose $d_{max}\geq 1$ and $\tilde{d}\leq\epsilon_1 m$, where $\epsilon_1<2/3$. Then the number of simple graphs with degree sequence $\boldsymbol{d}$ and none of the forbidden edges specified in D is uniformly*

$$\frac{(2m)!}{(m)!2^m\prod_{i=1}^n d_i!}\exp\{-\lambda-\lambda^2-\gamma+O(\tilde{d}^2/m)\} \tag{S1.1}$$

as $n\to\infty$.

This conclusion is uniform over all possibilities for $\boldsymbol{d}$ satisfying the constraints as $n\to\infty$. We use this theorem to approximate $|\Sigma_{\boldsymbol{d}}|$ and $|\Sigma_{\boldsymbol{d}|A_{ij}=0}|$, which will then lead to an approximation of P_{ij}. For the set $\Sigma_{\boldsymbol{d}}$, the matrix D has 0 for all entries. Therefore $x_{max}=0$ and $\gamma=0$. Since the condition of Theorem 1 requires $d_{max}=o(m^{1/4})$, we immediately have $d_{max}\geq 1$ and $\tilde{d}\leq\epsilon_1 m$ for $\epsilon_1<2/3$, i.e., the conditions for McKay's theorem are satisfied. Also

$O(\tilde{d}^2/m)$ becomes $o(1)$. Applying McKay's theorem, we have $|\Sigma_{\boldsymbol{d}}|$ is uniformly

$$\frac{(2m)!}{(m)!2^m \prod_{i=1}^n d_i!} \exp\{-\lambda - \lambda^2 + o(1)\} \tag{S1.2}$$

as $n \to \infty$.

For the set $|\Sigma_{\boldsymbol{d}|A_{ij}=0}|$, the matrix D has $D_{ij} = D_{ji} = 1$ and 0 elsewhere. In this case, $x_{max} = 1$ and $\gamma = d_i d_j / 2m$. Based on the same argument, we have that the conditions for McKay's theorem are satisfied and $O(\tilde{d}^2/m)$ becomes $o(1)$. Therefore, Applying McKay's theorem, we have $|\Sigma_{\boldsymbol{d}|A_{ij}=0}|$ is uniformly

$$\frac{(2m)!}{(m)!2^m \prod_{i=1}^n d_i!} \exp\{-\lambda - \lambda^2 - \frac{d_i d_j}{2m} + o(1)\} \tag{S1.3}$$

as $n \to \infty$.

From (3.4), we know $P_{ij} = 1 - |\Sigma_{\boldsymbol{d}|A_{ij}=0}|/|\Sigma_{\boldsymbol{d}}|$. Plugging in the approximations in (S1.2) and (S1.3), we have P_{ij} is uniformly $1 - e^{-\frac{d_i d_j}{2m} + o(1)}$ as $n \to \infty$.

S2 Proof of Theorem 2

We follow the structure in Zhao et al. (2012) for the proof. However, the consistency results under the logit link is not a trivial extension of the results under a log link. First we formalize the notations that will be used in the proof. For any label $\boldsymbol{e} = (e_1, \ldots, e_n)$, define the $K \times K$ matrix $O(\boldsymbol{e})$ by

$$O_{kl}(\boldsymbol{e}) = \sum_{ij} A_{ij} I\{e_i = k, e_j = l\},$$

and define

$$O_k(\boldsymbol{e}) = \sum_l O_{kl}(\boldsymbol{e}).$$

For $k \neq l$, O_{kl} is the number of edges between block k and block l (we shall often suppress the argument $\boldsymbol{e}$ for brevity).

Define the arrays $\hat{S}_{K\times K\times M}$, $V_{K\times K\times M}$ and $\hat{\Pi}_{K\times M}$ as

$$\begin{aligned}
\hat{S}_{kau}(\boldsymbol{e}) &= \frac{1}{n}\sum_{i=1}^n I(e_i = k, c_i = a, \theta_i = x_u), \\
V_{kau}(\boldsymbol{e}) &= \frac{\sum_{i=1}^n I(e_i = k, c_i = a, \theta_i = x_u)}{\sum_{i=1}^n I(c_i = a, \theta_i = x_u)}, \\
\hat{\Pi}_{au} &= \frac{1}{n}\sum_{i=1}^n I(c_i = a, \theta_i = x_u).
\end{aligned}$$

Roughly speaking, $\hat{S}$ can be thought of as the empirical joint distribution of $\boldsymbol{e}, \boldsymbol{c}$ and $\boldsymbol{\theta}$, V can be thought of as the conditional distribution of $\boldsymbol{e}$ given $\boldsymbol{c}$ and $\boldsymbol{\theta}$, and $\hat{\Pi}$ is the empirical joint distribution of $\boldsymbol{c}$ and $\boldsymbol{\theta}$.

Before we proceed to the proof, we first state a lemma.

Lemma 1. *Let $P_{K\times K}$ and $S_{K\times K\times N}$ be two matrices, and $S_{K\times K\times N}$ has nonnegative entries. Define the $K \times K$ matrix S^u as $S^u_{ij} = S_{iju}$. Assume that*

a) P is symmetric;

b) P does not have two identical columns;

c) there exist at least one nonzero entry in each column of $\sum_{u=1}^N S^u$;

d) for $1 \le a, b, k, l \le K$ and $1 \le u, v \le N$, $P_{kl} = P_{ab}$ whenever $S_{kau}S_{lbv} > 0$.

Then S^u, $u = 1, \ldots, N$, are all diagonal matrices or permutations of a diagonal matrix by the same permutation matrix.

See Section S3 for the proof of the lemma. Define $\mu_n = n^2\rho_n$, we will show Theorem 2 in four steps:

Step 1: Show that the modularity function in (4.9) can be written in the form of $F(\frac{O(\boldsymbol{e})}{\mu_n})$, for some function $F(\cdot)$.

Step 2: Show that $F(\frac{O(\boldsymbol{e})}{\mu_n})$ is uniformly close to its population version.

Step 3: Show the weak consistency by showing that there exist $\delta_n \to 0$, such that

$$P\left(\max_{\boldsymbol{e}:\ ||V(\boldsymbol{e})-V(\boldsymbol{c})||_1\ge\delta_n} F\left(\frac{O(\boldsymbol{e})}{\mu_n}\right) < F\left(\frac{O(\boldsymbol{c})}{\mu_n}\right)\right) \to 1, \quad \text{as} \quad \lambda_n \to \infty.$$

Here $||S||_1 = \sum_{kau} |S_{kau}|$.

Step 4: Show that

$$P\left(\max_{\boldsymbol{e}:\ \boldsymbol{e}\neq\boldsymbol{c}} F\left(\frac{O(\boldsymbol{e})}{\mu_n}\right) < F\left(\frac{O(\boldsymbol{c})}{\mu_n}\right)\right) \to 1, \quad \text{when} \quad \frac{\lambda_n}{\log n} \to \infty,$$

which implies the strong consistency.

Details of Step 1: The modularity in (4.9) can also be written as

$$Q(\boldsymbol{e}) = \sum_k \left(\frac{O_{kk}}{2m} - \frac{O_k^2}{(2m)^2}\right),$$

since it is true that

$$\frac{O_k^2}{2m} = \frac{(\sum_i d_i I(e_i = k))^2}{2m} = \sum_{ij} \frac{d_i d_j I(e_i = k, e_j = k)}{2m}.$$

Moreover, define

$$F(O)=\sum_k\left(\frac{O_{kk}}{\sum_{kh}O_{kh}}-\left(\frac{\sum_l O_{kl}}{\sum_{kh}O_{kh}}\right)^2\right).$$

We have

$$F\left(\frac{O(\boldsymbol{e})}{\mu_n}\right)=\sum_k\left(\frac{O_{kk}}{2m}-\frac{O_k^2}{(2m)^2}\right). \tag{S2.4}$$

Details of Step 2: Define $H_{kl}(R)=\sum_{abuv}\text{logit}^{-1}(x_u+x_v+Z_{ab})R_{kau}R_{lbv}$, we have

$$\begin{aligned}
&\frac{1}{\mu_n}E(O_{kl}|\boldsymbol{c},\boldsymbol{\theta})\\
=\ &\frac{1}{\mu_n}\sum_{ij}\sum_{abuv}E(A_{ij}I(e_i=k,c_i=a,\theta_i=x_u)I(e_j=l,c_j=b,\theta_j=x_v)|\boldsymbol{c},\boldsymbol{\theta})\\
=\ &\frac{1}{\mu_n}\sum_{ij}\sum_{abuv}P_{ij}^{(n)}I(e_i=k,c_i=a,\theta_i=x_u)I(e_j=l,c_j=b,\theta_j=x_v)\\
=\ &\frac{1}{\mu_n}\sum_{ij}\sum_{abuv}\rho_n\times\text{logit}^{-1}(x_u+x_v+Z_{ab})I(e_i=k,c_i=a,\theta_i=x_u)I(e_j=l,c_j=b,\theta_j=x_v)\\
=\ &\frac{1}{\mu_n}\sum_{abuv}\rho_n\times\text{logit}^{-1}(x_u+x_v+Z_{ab})\times n\times\hat{S}_{kau}(\boldsymbol{e})\times n\times\hat{S}_{lbv}(\boldsymbol{e})\\
=\ &\sum_{abuv}\text{logit}^{-1}(x_u+x_v+Z_{ab})\hat{S}_{kau}(\boldsymbol{e})\hat{S}_{lbv}(\boldsymbol{e})\\
=\ &H_{kl}(\hat{S}(\boldsymbol{e})).
\end{aligned}$$

Here we used the fact that $P_{ij}^{(n)}=\rho_n\times\text{logit}^{-1}(x_u+x_v+Z_{ab})$. Define

$$\hat{T}_{kl}(\boldsymbol{e})=\frac{1}{\mu_n}E(O_{kl}(\boldsymbol{e})|\boldsymbol{c},\boldsymbol{\theta}).$$

We have

$$\begin{aligned}
\hat{T}_{kl}(\boldsymbol{e})&=\sum_{abuv}\text{logit}^{-1}(x_u+x_v+Z_{ab})\hat{S}_{kau}(\boldsymbol{e})\hat{S}_{lbv}(\boldsymbol{e})\\
&=\sum_{abuv}\text{logit}^{-1}(x_u+x_v+Z_{ab})V_{kau}(\boldsymbol{e})\hat{\Pi}_{au}V_{lbv}(\boldsymbol{e})\hat{\Pi}_{bv}.
\end{aligned}$$

Replacing $\hat{\Pi}$ by Π, we define

$$T_{kl}(\boldsymbol{e})=\sum_{abuv}\text{logit}^{-1}(x_u+x_v+Z_{ab})V_{kau}(\boldsymbol{e})\Pi_{au}V_{lbv}(\boldsymbol{e})\Pi_{bv}. \tag{S2.5}$$

To show $F\left(\frac{O(\boldsymbol{e})}{\mu_n}\right)$ is uniformly close to its population version, we show that there exist $\epsilon_n\to 0$, such that

$$P\left(\max_{\boldsymbol{e}}\left|F\left(\frac{O(\boldsymbol{e})}{\mu_n}\right)-F(T(\boldsymbol{e}))\right|<\epsilon_n\right)\to 1\quad\text{as}\quad\lambda_n\to\infty. \tag{S2.6}$$

Since

$$\left|F\left(\frac{O(\boldsymbol{e})}{\mu_n}\right)-F(T(\boldsymbol{e}))\right|\le\left|F\left(\frac{O(\boldsymbol{e})}{\mu_n}\right)-F(\hat{T}(\boldsymbol{e}))\right|+\left|F(\hat{T}(\boldsymbol{e}))-F(T(\boldsymbol{e}))\right|,$$

we can bound the two terms on the right hand side. Since $F(\cdot)$ is Lipschitz in all its arguments, we have

$$\left|F(\hat{T}(\boldsymbol{e})) - F(T(\boldsymbol{e}))\right| \leq M_1 ||\hat{T}(\boldsymbol{e}) - T(\boldsymbol{e})||_\infty. \quad \text{(S2.7)}$$

Here $||X||_\infty = \max_{kl} |X_{kl}|$. Further,

$$\begin{aligned} |\hat{T}_{kl}(\boldsymbol{e}) - T_{kl}(\boldsymbol{e})| &= \left|\sum_{abuv} \text{logit}^{-1}(x_u + x_v + Z_{ab}) V_{kau}(\boldsymbol{e}) V_{lbv}(\boldsymbol{e})(\hat{\Pi}_{au}\hat{\Pi}_{bv} - \Pi_{au}\Pi_{bv})\right| \\ &\leq \sum_{abuv} \text{logit}^{-1}(x_u + x_v + Z_{ab}) |\hat{\Pi}_{au}\hat{\Pi}_{bv} - \Pi_{au}\Pi_{bv}|. \quad \text{(S2.8)} \end{aligned}$$

Since $\hat{\Pi}_{au} \to_p \Pi_{au}$ for all a, u, we have the left hand side of (S2.7) converges to 0 in probability uniformly over all $\boldsymbol{e}$ as $\lambda_n \to \infty$. Next, we have

$$\left|F\left(\frac{O(\boldsymbol{e})}{\mu_n}\right) - F(\hat{T}(\boldsymbol{e}))\right| \leq M_1 ||\frac{O(\boldsymbol{e})}{\mu_n} - \hat{T}(\boldsymbol{e})||_\infty. \quad \text{(S2.9)}$$

To continue the proof, we need to use Bernstein's inequality (Bernstein, 1924).

Bernstein's inequality: *Let $X_1, \ldots, X_n$ be independent variables. Suppose that $|X_i| \leq M$ for all i. Then, for all positive t,*

$$P\left(\left|\sum_{i=1}^n X_i - \sum_{i=1}^n E(X_i)\right| > t\right) \leq 2\exp\left(-\frac{t^2/2}{\sum var(X_i) + Mt/3}\right).$$

Since the A_{ij} in $O_{kl}(\boldsymbol{e})$ are independent and $|A_{ij}| < 2$, according to Bernstein's inequality, we have

$$P(|O_{kl}(\boldsymbol{e}) - \mu_n \hat{T}_{kl}(\boldsymbol{e})| > \omega | \boldsymbol{c}, \boldsymbol{\theta}) \leq 2\exp\left(-\frac{\omega^2/2}{\text{var}(O_{kl}|\boldsymbol{c}, \boldsymbol{\theta}) + 2\omega/3}\right). \quad \text{(S2.10)}$$

Notice that $\text{var}(O_{kl}|\boldsymbol{c}, \boldsymbol{\theta}) \leq 2n^2 \max_{ij} \text{var}(A_{ij}|\boldsymbol{c}, \boldsymbol{\theta}) \leq 2n^2\rho_n \max_{uvab}(\text{logit}^{-1}(x_u + x_v + Z_{ab}))$. Define $\tau = \max_{uvab}(\text{logit}^{-1}(x_u + x_v + Z_{ab}))$. Let $\omega = \epsilon n^2 \rho_n$. For $\epsilon < 3\tau$, we have

$$\begin{aligned} P\left(\left|\frac{O_{kl}(\boldsymbol{e})}{\mu_n} - \hat{T}_{kl}(\boldsymbol{e})\right| > \epsilon | \boldsymbol{c}, \boldsymbol{\theta}\right) &\leq 2\exp\left(-\frac{\omega^2/2}{\text{var}(O_{kl}|\boldsymbol{c}, \boldsymbol{\theta}) + 2\omega/3}\right) \\ &\leq 2\exp\left(-\frac{\epsilon^2 n^4 \rho_n^2}{8n^2\rho_n\tau}\right) \\ &= 2\exp\left(-\frac{\epsilon^2\mu_n}{8\tau}\right). \end{aligned}$$

We have the left hand side of (S2.9) converges to 0 in probability uniformly if $\lambda_n \to \infty$. Hence, we have shown that (S2.6) holds.

Details of Step 3: We show that there exists $\delta_n \to 0$, such that

$$P\left(\max_{\boldsymbol{e}:\ ||V(\boldsymbol{e}) - V(\boldsymbol{c})||_1 \geq \delta_n} F\left(\frac{O(\boldsymbol{e})}{\mu_n}\right) < F\left(\frac{O(\boldsymbol{c})}{\mu_n}\right)\right) \to 1, \quad \text{as} \quad \lambda_n \to \infty.$$

We first show that $F(H(S))$ is uniquely maximized over $\{S : S \geq 0,\ \sum_k S_{kau} = \Pi_{au}\}$ by $S = \mathbb{D}$, where $\mathbb{D}_{kau} = \Pi_{au} E_{ka}$. Here $S_{kau} = V_{kau}\Pi_{au}$ is the limit of $\hat{S}$ and E is any row permutation of the $K \times K$ identity matrix. The matrix E is for the case with permutation equivalence class. For simplicity, in the following proof, we assume E is the identity matrix.

If $Q(\mathbf{e})$ is maximized by the true label $\mathbf{c}$, then $F(H(S))$ should be maximized by the true assignment $S = \mathbb{D}$. Since $\sum_k \hat{S}_{kau}(\boldsymbol{e}) \to \Pi_{au}$, the limit S must satisfy $\sum_k S_{kau} = \Pi_{au}$. Therefore, we only need to consider the maximizer of $F(H(S))$ under the constraint $\{S : S \geq 0,\ \sum_k S_{kau} = \Pi_{au}\}$.

Define $H_0(S) = \sum_{kl} H_{kl}(S)$ and $H_k(S) = \sum_l H_{kl}(S)$. For simplicity of the notations, we leave out the dependence of H on S. Then we have

$$F(H) = \sum_k \left(\frac{H_{kk}}{H_0} - \frac{H_k^2}{H_0^2} \right).$$

Define

$$\triangle_{kl} = \begin{cases} 1 & \text{for} \quad k = l, \\ -1 & \text{for} \quad k \neq l. \end{cases}$$

Using the equality that

$$\sum_k \left(\frac{H_{kk}}{H_0} - \frac{H_k^2}{H_0^2} \right) + \sum_{k \neq l} \left(\frac{H_{kl}}{H_0} - \frac{H_k H_l}{H_0^2} \right) = 0,$$

we have

$$
\begin{aligned}
F(H(S)) &= \frac{1}{2H_0}\sum_{kl}\triangle_{kl}\sum_{abuv}\text{logit}^{-1}(x_u+x_v+Z_{ab})S_{kau}S_{lbv}\\
&-\frac{1}{2H_0^2}\sum_{kl}\triangle_{kl}\left[\sum_{abuv}\text{logit}^{-1}(x_u+x_v+Z_{ab})\Pi_{bv}S_{kau}\right]\left[\sum_{abuv}\text{logit}^{-1}(x_u+x_v+Z_{ab})\Pi_{au}S_{lbv}\right]\\
&= \frac{1}{2H_0}\sum_{kl}\sum_{ab}\triangle_{kl}\sum_{uv}\text{logit}^{-1}(x_u+x_v+Z_{ab})S_{kau}S_{lbv}\\
&-\frac{1}{2H_0^2}\sum_{kl}\sum_{a'b}\triangle_{kl}\sum_{ab'uu'vv'}\text{logit}^{-1}(x_u+x_v+Z_{ab})\text{logit}^{-1}(x_{u'}+x_{v'}+Z_{a'b'})\Pi_{bv}\Pi_{a'u'}S_{kau}S_{lb'v'}\\
&= \frac{1}{2H_0}\sum_{kl}\sum_{ab'}\sum_{uv'}\triangle_{kl}S_{kau}S_{lb'v'}\Bigg[\text{logit}^{-1}(x_u+x_{v'}+Z_{ab'})\\
&-\frac{1}{H_0}\sum_{a'bu'v}\text{logit}^{-1}(x_u+x_v+Z_{ab})\text{logit}^{-1}(x_{u'}+x_{v'}+Z_{a'b'})\Pi_{bv}\Pi_{a'u'}\Bigg]\\
&\leq \frac{1}{2H_0}\sum_{kl}\sum_{ab'}\sum_{uv'}\triangle_{ab'}S_{kau}S_{lb'v'}\Bigg[\text{logit}^{-1}(x_u+x_{v'}+Z_{ab'})\\
&-\frac{1}{H_0}\sum_{a'bu'v}\text{logit}^{-1}(x_u+x_v+Z_{ab})\text{logit}^{-1}(x_{u'}+x_{v'}+Z_{a'b'})\Pi_{bv}\Pi_{a'u'}\Bigg]\\
&= \frac{1}{2H_0}\sum_{ab'}\triangle_{ab'}\sum_{uv}\text{logit}^{-1}(x_u+x_v+Z_{ab'})\Pi_{au}\Pi_{b'v}\\
&-\frac{1}{2H_0^2}\sum_{ab'}\triangle_{ab'}\sum_{a'uv}\sum_{bu'v'}\text{logit}^{-1}(x_u+x_v+Z_{ab})\text{logit}^{-1}(x_{u'}+x_{v'}+Z_{a'b'})\Pi_{bv}\Pi_{b'v'}\Pi_{au}\Pi_{a'u'}\\
&= F(H(\mathbb{D})).
\end{aligned}
$$

The inequality is true because of conditions (4.10) and (4.11). Now we need to show that $\mathbb{D}$ is the unique maximizer of $F(H(S))$. The inequality $F(H(S))\leq F(H(\mathbb{D}))$ holds only if $\triangle_{kl}=\triangle_{ab'}$ when $S_{kau}S_{lb'v'}>0$. Since $\triangle$ is symmetric, does not have two identical columns and $\sum_u S_{kau}$ has at least one non-zero entry in each column, following the result in Lemma 1, we have S^u are all diagonal matrices or permutations of the diagonal matrix by the same permutation matrix. With the constraint $\{S: S\geq 0,\ \sum_k S_{kau}=\Pi_{au}\}$, we have $F(H(S))=F(H(\mathbb{D}))$ holds only when $S=\mathbb{D}$.

From (S2.5), the definition of $H_{kl}(S)$ and $\mathbb{D}$, it is straightforward that $H_{kl}(S)=T_{kl}(\boldsymbol{e})$ and $H_{kl}(\mathbb{D})=T_{kl}(\boldsymbol{c})$. Since $F(H(S))$ is uniquely maximized by $\mathbb{D}$, by the continuity of $F(\cdot)$ in the neighborhood of $\mathbb{D}$, there exists $\delta_n\to 0$ such that

$$
F(T(\mathbf{c}))-F(T(\mathbf{e}))>2\epsilon_n \quad \text{for} \quad ||V(\mathbf{e})-V(\mathbf{c})||_1\geq\delta_n. \tag{S2.11}
$$

Here we used the fact that

$$
\begin{aligned}
||S-\mathbb{D}||_1=\sum_{kau}|V_{kau}(\boldsymbol{e})\Pi_{au}-V_{kau}(\boldsymbol{c})\Pi_{au}| \quad &\geq \quad (\min_{au}\Pi_{au})\times\sum_{kau}|V_{kau}(\boldsymbol{e})-V_{kau}(\boldsymbol{c})|\\
&= \quad (\min_{au}\Pi_{au})\times||V(\mathbf{e})-V(\mathbf{c})||_1.
\end{aligned}
$$

Thus, with (S2.6), we have

$$\begin{aligned}
&P\left(\max_{\boldsymbol{e}:\ ||V(\boldsymbol{e})-V(\boldsymbol{c})||_1\geq\delta_n} F\left(\frac{O(\boldsymbol{e})}{\mu_n}\right) < F\left(\frac{O(\boldsymbol{c})}{\mu_n}\right)\right)\\
&\geq P\left(\left|\max_{\boldsymbol{e}:\ ||V(\boldsymbol{e})-V(\boldsymbol{c})||_1\geq\delta_n} F\left(\frac{O(\boldsymbol{e})}{\mu_n}\right) - \max_{\boldsymbol{e}:\ ||V(\boldsymbol{e})-V(\boldsymbol{c})||_1\geq\delta_n} F(T(\boldsymbol{e}))\right| < \epsilon_n,\right.\\
&\qquad\left.\left|F\left(\frac{O(\boldsymbol{c})}{\mu_n}\right) - F(T(\boldsymbol{c}))\right| < \epsilon_n\right) \to 1.
\end{aligned}$$

This implies that

$$P(||V(\hat{\boldsymbol{c}}) - V(\boldsymbol{c})||_1 \leq \delta_n) \to 1,$$

where $\hat{\boldsymbol{c}} = \arg\max_{\boldsymbol{e}} F\left(\frac{O(\boldsymbol{e})}{\mu_n}\right)$ is the estimator. Since

$$\begin{aligned}
\frac{1}{n}|\boldsymbol{e}-\boldsymbol{c}| = \frac{1}{n}\sum_{i=1}^n I(c_i \neq e_i) &= \sum_{au}\hat{\Pi}_{au}(1 - V_{aau}(\boldsymbol{e}))\\
&\leq \sum_{au}(1 - V_{aau}(\boldsymbol{e}))\\
&= \frac{1}{2}\left(\sum_{au}(1 - V_{aau}(\boldsymbol{e})) + \sum_{au}\sum_{k\neq a} V_{kau}(\boldsymbol{e})\right)\\
&= \frac{1}{2}||V(\boldsymbol{e}) - V(\boldsymbol{c})||_1,
\end{aligned}$$

we have established the weak consistency of the estimator.

Details of Step 4: In order to show

$$P\left(\max_{\boldsymbol{e}:\ \boldsymbol{e}\neq\boldsymbol{c}} F\left(\frac{O(\boldsymbol{e})}{\mu_n}\right) < F\left(\frac{O(\boldsymbol{c})}{\mu_n}\right)\right) \to 1, \quad \text{as} \quad \frac{\lambda_n}{\log n} \to \infty,$$

we only need to show that there exists $\delta_n \to 0$, such that

$$P\left(\max_{\boldsymbol{e}:\ 0<||V(\boldsymbol{e})-V(\boldsymbol{c})||_1\leq\delta_n} F\left(\frac{O(\boldsymbol{e})}{\mu_n}\right) < F\left(\frac{O(\boldsymbol{c})}{\mu_n}\right)\right) \to 1, \quad \text{as} \quad \frac{\lambda_n}{\log n} \to \infty, \tag{S2.12}$$

since the results in Step 3 implies that there exists $\delta_n \to 0$, such that

$$P\left(\max_{\boldsymbol{e}:\ ||V(\boldsymbol{e})-V(\boldsymbol{c})||_1\geq\delta_n} F\left(\frac{O(\boldsymbol{e})}{\mu_n}\right) < F\left(\frac{O(\boldsymbol{c})}{\mu_n}\right)\right) \to 1, \quad \text{as} \quad \frac{\lambda_n}{\log n} \to \infty.$$

Following equation (A.11) in Zhao et al. (2012), we have

$$F\left(\frac{O(\boldsymbol{e})}{\mu_n}\right) - F\left(\frac{O(\boldsymbol{c})}{\mu_n}\right) = F(\hat{T}(\boldsymbol{e})) - F(\hat{T}(\boldsymbol{c})) + \boldsymbol{\Delta}(\boldsymbol{e},\boldsymbol{c}). \tag{S2.13}$$

By the continuity of the derivatives of F in $||V(\boldsymbol{e}) - V(\mathbf{c})||_1 \leq \delta_n$, we have

$$F(T(\boldsymbol{e})) - F(T(\boldsymbol{c})) \leq -C||V(\boldsymbol{e}) - V(\boldsymbol{c})||_1 + o(||V(\boldsymbol{e}) - V(\boldsymbol{c})||_1).$$

Since the derivative of F is continuous with respect to $V(\boldsymbol{e})$ in $||V(\boldsymbol{e}) - V(\mathbf{c})||_1 \leq \delta_n$, there exists a δ^* such that

$$F(\hat{T}(\boldsymbol{e})) - F(\hat{T}(\boldsymbol{c})) \leq -\frac{C}{2}||V(\boldsymbol{e}) - V(\boldsymbol{c})||_1 + o(||V(\boldsymbol{e}) - V(\boldsymbol{c})||_1), \tag{S2.14}$$

for $||\hat{\Pi} - \Pi||_\infty \leq \delta^*$. Based on (S2.13) and (S2.14), we can see that in order to show (S2.12), it is sufficient to show

$$P(\max_{\{\boldsymbol{e}\neq\boldsymbol{c}\}} |\boldsymbol{\Delta}(\boldsymbol{e}, \boldsymbol{c})| \leq C||V(\boldsymbol{e}) - V(\boldsymbol{c})||_1/4) \to 1. \tag{S2.15}$$

The conclusion (S2.15) is true following the results in Bickel et al. (2015). Therefore we have established the strong consistency.

S3 Proof of Lemma 1

Define $\mathcal{S} = \sum_{u=1}^N S^u$. We have

$$\begin{aligned} \mathcal{S}_{ka}\mathcal{S}_{lb} > 0 \implies \left(\sum_{u=1}^N S^u_{ka}\right)\left(\sum_{u=1}^N S^u_{lb}\right) > 0 \quad &\implies \quad \exists\, i, j \quad \text{such that } S_{kai}S_{lbj} > 0 \\ &\implies \quad P_{ka} = P_{lb}. \end{aligned}$$

Following Lemma 3.2 in Bickel and Chen (2009), we have $\mathcal{S}$ is diagonal or a permutation of the diagonal matrix, since there exists at least one non-zero entry in each column of $\mathcal{S}$. Since all entries in S are non-negative, we have S^u, $u = 1, \ldots, N$ are all diagonal matrices or permutations of the diagonal matrix by the same permutation matrix.

S4 Modularity Maximization

We discuss modularity maximization techniques for finding the partition that maximizes the modularity function, i.e., finding

$$\hat{\boldsymbol{e}} = \arg \max_{\substack{\boldsymbol{e}=(e_1,\ldots,e_n) \\ e_i \in \{1,\ldots,n\}}} Q(\boldsymbol{e}, G), \tag{S4.16}$$

where $Q(\boldsymbol{e}, G)$ is defined in (3.3). Modularity based community detection techniques are among the most popular approaches in detecting communities in networks (Fortunato, 2010). Existing approaches for maximizing the modulation function come from various fields, including computer science, physics, sociology and statistics. Some are fast techniques that can be applied to large graphs, but may not find good approximations to the optimum of the modularity function (Clauset et al., 2004; Newman, 2004). Some are accurate methods that find good approximations to the optimum but are limited to graphs of moderate sizes (Massen and Doye, 2005; Guimera et al, 2004; Agrawal and Kempe, 2008). Some algorithms achieve a good balance between accuracy and complexity (Newman, 2006a; Wang et al., 2008). See Chapter VI of Fortunato (2010) for a review.

To simplify the notation, we define an assignment matrix $B_{n\times K}$, which is a 0-1 matrix with $B_{ij}=1$ if the i-th node is in the j-th community and $B_{ij}=0$ otherwise. Each row of B sums to unity, and the columns $\boldsymbol{b}_1,\ldots,\boldsymbol{b}_K$ of B are mutually orthogonal. Maximizing the modularity in (S4.16) can therefore be expressed as

$$\max_B\{\text{Trace}(B^TMB)\} \quad \text{such that} \quad \text{Trace}(B^TB)=n, \tag{S4.17}$$

where $M=A-E_{p,\Sigma_{\boldsymbol{d}}}(A)$ is the modularity matrix. Newman (2006b) pointed out the intimate relationship between community structures and the eigen-spectrum of the Newman-Girvan modularity matrix. Here we extend that relationship to the community structure and the eigen-spectrum of the modularity matrix M.

Denote the eigenvalues of M as $\lambda_1,\ldots,\lambda_n$ and the corresponding normalized pairwise orthogonal eigenvectors as $\boldsymbol{v}_1,\ldots,\boldsymbol{v}_n$. Without loss of generality, assume $\lambda_1\geq\cdots\geq\lambda_n$. Denote $\kappa=\sum_{i=1}^n I(\lambda_i>0)$ as the number of positive eigenvalues. We have

$$Q=\text{Trace}(B^TMB)=\sum_{i=1}^n\sum_{k=1}^K\lambda_i(\boldsymbol{v}_i^T\boldsymbol{b}_k)^2. \tag{S4.18}$$

Maximizing the modularity is equivalent to choosing $K-1$ orthogonal columns $\boldsymbol{b}_1,\ldots,\boldsymbol{b}_{K-1}$ such that the summation in (S4.18) is maximized. Since $\boldsymbol{v}_1,\ldots,\boldsymbol{v}_n$ form an orthonormal basis of an n-dimensional vector space, we have

$$\boldsymbol{b}_k=\sum_{i=1}^n\alpha_{ki}\boldsymbol{v}_i, \quad \text{for} \quad k=1,\ldots,K,$$

where $\alpha_{ki}=\boldsymbol{v}_i^T\boldsymbol{b}_k$. Therefore, we have

$$Q=\sum_{i=1}^n\sum_{k=1}^K\lambda_i(\boldsymbol{v}_i^T\boldsymbol{b}_k)^2=\sum_{i=1}^n\lambda_i\left(\sum_{k=1}^K\alpha_{ki}^2\right). \tag{S4.19}$$

This shows that the major contribution to the modularity value comes from the projection of the columns $\boldsymbol{b}_1,\ldots,\boldsymbol{b}_K$ onto the subspace spanned by the leading eigenvectors. For a partition that achieves large Q, vectors $\boldsymbol{b}_1,\ldots,\boldsymbol{b}_K$ necessarily have large projections onto the leading eigenvectors with positive eigenvalues. Newman (2006b) showed that if we do not have the binary constraint on the entries in B, then Q will be maximized when $\boldsymbol{b}_k$ is proportional to $\boldsymbol{v}_k$, $k=1,\ldots,K-1$, and the number of orthogonal columns in B is the same as the number of positive eigenvalues, i.e., $K=\kappa+1$. However, the entries in B are binary. With this constraint, the number of positive eigenvalues κ becomes an upper bound for $K-1$.

When the graph is large, we can drop the terms in (S4.18) that are proportional to the $n-K+1$ smallest eigenvalues λ_i (Newman, 2006b; Ng et al., 2001; Wang et al., 2008). The $K-1$ largest positive eigenvalues $\lambda_1, \ldots, \lambda_{K-1}$ can be used to form a diagonal matrix $\Lambda = \text{diag}(\lambda_1, \ldots, \lambda_{K-1})$, and their corresponding eigenvectors $\boldsymbol{v}_1, \ldots, \boldsymbol{v}_{K-1}$ can be used to form a matrix $V = (\boldsymbol{v}_1, \ldots, \boldsymbol{v}_{K-1})$. Then we have

$$\begin{aligned}
\max_B Q = \max_B \sum_{i=1}^{n}\sum_{k=1}^{K} \lambda_i (\boldsymbol{v}_i^T \boldsymbol{b}_k)^2 &\approx \max_B \sum_{i=1}^{K-1}\sum_{k=1}^{K} \lambda_i (\boldsymbol{v}_i^T \boldsymbol{b}_k)^2 \\
&= \max_B \Big(\text{Trace}(B^T V \Lambda V^T B)\Big) \\
&= \max_B \Big((\text{Trace}(\Lambda^{\frac{1}{2}} V^T B)^T (\Lambda^{\frac{1}{2}} V^T B))\Big) \\
&= \max_B \sum_{k=1}^{K} ||\boldsymbol{w}_k||^2,
\end{aligned}$$

where $\boldsymbol{w}_k = \sum_i \boldsymbol{y}_i I(B_{ik=1})$ and $\boldsymbol{y}_i$ is the i-th row of matrix $V\Lambda^{\frac{1}{2}}$. The modularity maximization is now a problem of grouping vectors $\boldsymbol{y}_i$ into K groups such that the magnitude of the vector $\boldsymbol{w}_k$ is maximized. One simple approach for this problem is to apply the k-means clustering on the normalized vectors $\boldsymbol{y}_1, \ldots, \boldsymbol{y}_n$ (Ng et al., 2001).

Here is a summary of our approach for detecting up to $K \leq \kappa + 1$ communities in the graph $G(V, E)$.

1. Find the modularity matrix M, its eigenvalues $\lambda_1, \ldots, \lambda_n$ and their corresponding normalized orthogonal eigenvectors $\boldsymbol{v}_1, \ldots, \boldsymbol{v}_n$.

2. For each value k, $2 \leq k \leq K$:

- **a.** Construct the corresponding diagonal matrix Λ and the eigenvector matrix V. Calculate $Y = V\Lambda^{\frac{1}{2}}$.
- **b.** Perform k-means clustering on the normalized rows $\boldsymbol{y}_i$ of matrix Y.
- **c.** With the membership output from the k-means clustering, calculate the modularity function Q_k.

3. Output the k that has the largest Q_k and its corresponding community labels.

In step 2(b), more sophisticated clustering methods can be used. This type of problems are referred to as the *vector partitioning* problems, i.e., grouping vectors $\boldsymbol{y}_i$ into K groups such that the magnitude of the vector $\boldsymbol{w}_k$ is maximized (Alpert et al., 1999).

References

Agrawal, G., and Kempe, D. (2008). Modularity-Maximizing Graph Communities Via Mathematical Programming. *The European Physics Journal B* **66**, 409-418.

Alpert, C., Kahng, A., and Yao, S. (1999). Spectral Partitioning with Multiple Eigenvectors. *Discrete Applied Mathematics* **90**, 3-26.

Bernstein, S. N. (1924). On a Modification of Chebyshev's Inequality and of the Error Formula of Laplace. *Annals Science Institute SAV. Ukraine* **4**, 38-49.

Bickel, P., and Chen, A. (2009). A Non-parametric View of Network Models and Newman-Girvan and Other Modularities. *Proceedings of the National Academy of Sciences* **106**, 21068-21073.

Bickel, P., Chen, A., Zhao, Y., Levina, E., and Zhu, J. (2015). Correction to the Proof of Consistency of Community Detection. *The Annals of Statistics* **43**, 462-466.

Clauset, A., Newman, M. E. J., and Moore, C. (2004). Finding Community Structure in Very Large Networks. *Physical Review E* **70**, 066111.

Fortunato, S. (2010). Community Detection in Graphs. *Physics Reports* **428**, 75-174.

Guimera, R., Sales-Pardo, M., and Amaral, L. A. N. (2004). Modularity from Fluctuations in Random Graphs and Complex Networks. *Physical Review E* **70**, 025101.

Massen, C., and Doye, J. (2005). Identifying Communities within Energy Landscapes. *Physical Review E* **71**, 046101.

McKay, B. D. (1985). Asymptotics for Symmetric 0-1 Matrices with Prescribed Row Sums. *Ars Combinatoria* **19A**, 15-26.

Newman, M. E. J. (2004). Fast Algorithm for Detecting Community Structure in Networks. *Physical Review E* **69**, 066133.

Newman, M. E. J. (2006a). Modularity and Community Structure in Networks. *Proceedings of the National Academy of Sciences*, 8577-8582.

Newman, M. E. J. (2006b). Finding Community Structure in Networks Using the Eigenvectors of Matrices. *Physical Review E* **74**, 035104.

Ng, A. Y., Jordan, M., and Weiss, Y. (2001). On Spectral Clustering: Analysis and an Algorithm. *Advances in Neural Information Processing Systems* **14**, 849-856.

Wang, G., Shen, Y., and Ouyang, M. (2008). A Vector Partitioning Approach to Detecting Community Structure in Complex Networks. *Computers and Mathematics with Applications* **55**, 2746-2752.

Zhao, Y., Levina, E., and Zhu, J. (2012). Consistency of Community Detection in Networks Under Degree-Corrected Stochastic Block Models. *The Annals of Statistics* **40**, 2266-2292.