MODULARITY BASED COMMUNITY DETECTION IN HETEROGENEOUS NETWORKS

Jingfei Zhang and Yuguo Chen

University of Miami and University of Illinois at Urbana-Champaign

Supplementary Material

S1 Proof of Theorem 1

First, we state a theorem from McKay (2010) on the asymptotic number of simple graphs with forbidden edges. Consider simple graphs with m edges and degree sequence $\boldsymbol{d} = (d_1, \ldots, d_n)$. Let X be an $n \times n$ symmetric zero-one matrix that specifies the set of forbidden edges ($X_{ij} = X_{ji} = 1$ if an edge between node i and node j is forbidden, and $X_{ij} = X_{ji} = 0$ otherwise). Write the column sums of X as $\boldsymbol{x} = (x_1, \ldots, x_n)$, $\eta = \sum_i x_i/2$ and $(x)_a = x(x-1)\cdots(x-a+1)$.

Define $d_{max} = \max_i d_i$, $x_{max} = \max_i x_i$ and $\Delta = d_{max}(d_{max}+x_{max})$. Let $P_{\Sigma_{\boldsymbol{d}}}(X)$ be the probability that a simple graph with degree sequence $\boldsymbol{d}$ contains all forbidden edges in X. We have the following theorem.

Theorem S1. *(McKay, 2010) If $\Delta\eta = o(m)$, then*

$$P_{\Sigma_{\boldsymbol{d}}}(X) = \frac{\prod_{j=1}^{n}(d_j)_{x_j}}{2^{\eta}(m)_{\eta}}(1 + O(\Delta\eta/m)). \tag{S1.1}$$

From the development in Section 2.2, we have

$$E(A_{ij}^{[l]}) \quad = \quad \frac{|\Sigma_{\mathbf{d}^{[l]}|A_{ij}^{[l]}=1}|}{|\Sigma_{\mathbf{d}^{[l]}}|}, \quad l=1,2,$$

where $|\Sigma_{\mathbf{d}^{[l]}|A_{ij}^{[l]}=1}|$ is the total number of simple homogeneous networks with degree sequence $\mathbf{d}^{[l]}$ and a link between nodes i and j. Consider the matrix X with $X_{ij}=X_{ji}=1$ and 0 elsewhere. In this case, we have $\eta=1$, $x_l=1$ for $l=i,j$ and $x_l=0$ otherwise. From (S1.1) and the condition that $d_{max}^{[l]}=o(m^{1/2})$, we have

$$\begin{aligned} E(A_{ij}^{[l]}) = \frac{|\Sigma_{\mathbf{d}^{[l]}|A_{ij}^{[l]}=1}|}{|\Sigma_{\mathbf{d}^{[l]}}|} &= P_{\Sigma_{\boldsymbol{d}^{[l]}}}(D) \\ &= \frac{d_i d_j}{2m}(1+o(1)), \quad l=1,2. \end{aligned}$$

Next it remains for us to show that as $n_1, n_2 \to \infty$,

$$E(A_{ij}^{[12]}) = \frac{d_i^{[12]} d_j^{[21]}}{m^{[12]}}(1+o(1)).$$

First, it is easy to derive that

$$E(A_{ij}^{[12]}) = \frac{|\Sigma_{\mathbf{d}^{[12]},\mathbf{d}^{[21]}|A_{ij}^{[12]}=1}|}{|\Sigma_{\mathbf{d}^{[12]},\mathbf{d}^{[21]}}|},$$

where $|\Sigma_{\mathbf{d}^{[12]},\mathbf{d}^{[21]}|A_{ij}^{[12]}=1}|$ is the total number of bipartite graphs with degree sequences $\mathbf{d}^{[12]}$ for type-[1] nodes, $\mathbf{d}^{[21]}$ for type-[2] nodes and a link between the ith node of type-[1] and the jth node of type-[2]. Next, we state the following theorem from McKay (2010). It is an analog of Theorem S1 for bipartite graphs.

Consider simple bipartite graphs with m edges and degree sequence $\boldsymbol{d}=(d_1,\ldots,d_{n_1})$ for one type of nodes, referred to as type-[1] nodes, and $\boldsymbol{d'}=(d'_1,\ldots,d'_{n_2})$ for the other type of nodes, referred to type-[2] nodes. Let X be an $n_1 \times n_2$ zero-one matrix that specifies the set of forbidden edges ($X_{ij}=1$ if an edge between node i of type-[1] and node j of type-[2] is forbidden, and $X_{ij}=0$ otherwise). Write the row sums of X as $\boldsymbol{x}=(x_1,\ldots,x_{n_1})$, column sums of X as $\boldsymbol{y}=(y_1,\ldots,y_{n_2})$, $\eta=\sum_i x_i=\sum_j y_j$.

Define $d_{max} = \max_i d_i$, $d'_{max} = \max_j d'_j$, $x_{max} = \max_i x_i$, $y_{max} = \max_j y_j$ and $\Delta' = (d_{max} + d'_{max})(d_{max} + d'_{max} + x_{max} + y_{max})$. Let $P_{\Sigma_{\boldsymbol{d},\boldsymbol{d'}}}(X)$ be the probability that a bipartite graph with degree sequence $\boldsymbol{d}$ and $\boldsymbol{d'}$ contains all forbidden edges in X. We have the following theorem.

Theorem S2. *(McKay, 2010) If $\Delta'\eta = o(m)$, then*

$$P_{\Sigma_{\boldsymbol{d},\boldsymbol{d'}}}(X) = \frac{\prod_{i=1}^{n_1}(d_i)_{x_i}\prod_{j=1}^{n_2}(d'_j)_{y_j}}{(m)_\eta}(1 + O(\Delta'\eta/m)). \tag{S1.2}$$

Consider the matrix X with $X_{ij} = X_{ji} = 1$ and 0 elsewhere. In this case, we have $\eta = 1$, $x_i = 1$, $y_j = 1$, $x_l = 0$ for $l \neq i$ and $y_l = 0$ for $l \neq j$. From (S1.1), the condition that $d^{[12]}_{max} = o((m^{[12]\,1/2})$ and $d^{[21]}_{max} = o(m^{[12]\,1/2})$, we have

$$\begin{aligned} E(A^{[12]}_{ij}) = \frac{|\Sigma_{\mathbf{d}^{[12]},\mathbf{d}^{[21]}|A^{[12]}_{ij}=1}|}{|\Sigma_{\mathbf{d}^{[12]},\mathbf{d}^{[21]}}|} &= P_{\Sigma_{\boldsymbol{d}^{[1]},\boldsymbol{d}^{[2]}}}(D) \\ &= \frac{d^{[12]}_i d^{[21]}_j}{m}(1 + o(1)), \quad l = 1, 2. \end{aligned}$$

S2 Proof of Theorem 2

First we formalize the notations that will be used in the proof. Consider a heterogeneous network $\mathcal{G}(\bigcup_{i=1}^L V^{[i]}, \mathcal{E} \cup \mathcal{E}^+)$. For a community assignment label $\mathcal{E} = (\mathbf{e}^{[1]}, \ldots, \mathbf{e}^{[L]})$ with $\mathbf{e}^{[l]} = (e^{[l]}_1, \ldots, e^{[l]}_{n_l})$, $l = 1, \ldots, L$, define $K \times K$ matrices $O^{[l]}$, $l = 1, \ldots, L$, and $O^{[l_1 l_2]}$, $1 \leq l_1 \neq l_2 \leq L$, such that

$$O^{[l]}_{kh}(\mathcal{E}) = \sum_{ij} A^{[l]}_{ij} I(e^{[l]}_i = k, e^{[l]}_j = h),$$

$$O^{[l_1 l_2]}_{kh}(\mathcal{E}) = \sum_{ij} A^{[l_1 l_2]}_{ij} I(e^{[l_1]}_i = k, e^{[l_2]}_j = h).$$

Define $O^{[l]}_k = \sum_h O^{[l]}_{kh}$ and $O^{[l_1 l_2]}_k = \sum_h O^{[l_1 l_2]}_{kh}$, $l = 1, \ldots, L$, $1 \leq l_1 \neq l_2 \leq L$. Define $K \times K$ matrices $R^{[l]}(\mathcal{E})$, $V^{[l]}(\mathcal{E})$, $l = 1, \ldots, L$, such that

$$R^{[l]}_{ab}(\mathcal{E}) = \frac{1}{n}\sum_{l=1}^{n_l} I(e^{[l]}_i = a, c^{[l]}_i = b)$$

$$V_{ab}^{[l]}(\mathcal{E}) = \frac{\sum_{l=1}^{n_l} I(e_i^{[l]} = a, c_i^{[l]} = b)}{\sum_{l=1}^{n_l} I(c_i^{[l]} = b)}.$$

Write $\mathcal{O} = \{O^{[l]}, O^{[l_1 l_2]}, l = 1, \ldots, L, 1 \leq l_1 \neq l_2 \leq L\}$ and $\mathcal{R} = \{R^{[1]}, \ldots, R^{[L]}\}$.

For community assignment label $\mathcal{E}$, the contribution of the bipartite graph $G_{l_1 l_2}$ to the modularity function $Q'(\mathcal{E}, \mathcal{G})$ is

$$q_{l_1 l_2} = \frac{1}{L^2} \sum_{ij} \left(A_{ij}^{[l_1 l_2]} - \frac{d_i^{[l_1 l_2]} d_j^{[l_2 l_1]}}{m^{[l_1 l_2]}} \right) \delta(e_i^{[l_1]}, e_j^{[l_2]}),$$

where $\delta(\cdot, \cdot)$ is the Kronecker function. We have

$$\begin{aligned} q_{l_1 l_2} &= \frac{1}{L^2} \left(\sum_{ij} A_{ij}^{[l_1 l_2]} \delta(e_i^{[l_1]}, e_j^{[l_2]}) - \frac{1}{m^{[l_1 l_2]}} \sum_k \sum_{ij} d_i^{[l_1 l_2]} d_j^{[l_2 l_1]} I(e_i^{[l_1]} = k) I(e_j^{[l_2]} = k)) \right) \\ &= \frac{1}{L^2} \left(\sum_k O_k^{[l_1 l_2]} - \frac{1}{m^{[l_1 l_2]}} \sum_k O_k^{[l_1 l_2]} O_k^{[l_2 l_1]} \right). \end{aligned}$$

Following similar arguments, it is easy to show that the modularity function $Q'(\mathcal{E}, \mathcal{G})$ can be expressed as

$$\frac{1}{L^2} \left[\sum_{l=1}^{L} \sum_{k=1}^{K} \left(O_{kk}^{[l]} - \frac{{O_k^{[l]}}^2}{\sum_{kh} O_{kh}^{[l]}} \right) + \sum_{l_1 \neq l_2}^{L} \sum_{k=1}^{K} \left(O_{kk}^{[l_1 l_2]} - \frac{O_k^{[l_1 l_2]} O_k^{[l_2 l_1]}}{\sum_{kh} O_{kh}^{[l_1 l_2]}} \right) \right].$$

Here we suppress the argument $\mathcal{E}$ for brevity. Define

$$J(\mathcal{O}) = \sum_{l=1}^{L} J_1(O^{[l]}) + \sum_{l_1 \neq l_2}^{L} J_2(O^{[l_1 l_2]}, O^{[l_2 l_1]}),$$

where

$$J_1(O^{[l]}) = \sum_{k=1}^{K} \left(O_{kk}^{[l]} - \frac{{O_k^{[l]}}^2}{\sum_{kh} O_{kh}^{[l]}} \right),$$

and

$$J_2(O^{[l_1 l_2]}, O^{[l_2 l_1]}) = \sum_{k=1}^{K} \left(O_{kk}^{[l_1 l_2]} - \frac{O_k^{[l_1 l_2]} O_k^{[l_2 l_1]}}{\sum_{kh} O_{kh}^{[l_1 l_2]}} \right).$$

We show the consistency property by showing that there exists $\delta_n \to 0$ such that

$$P\left(\max_{\mathcal{E}:\ \eta(\mathcal{E}, \mathcal{C}) \geq \delta_n} J\left(\frac{\mathcal{O}(\mathcal{E})}{\mu_n} \right) \leq J\left(\frac{\mathcal{O}(\mathcal{C})}{\mu_n} \right) \right) \to 1 \quad \text{as} \quad n \to \infty,$$

where $\eta(\mathcal{E}, \mathcal{C}) = \sum_{l=1}^{L} \sum_{ab} |V_{ab}^{[l]}(\mathcal{E}) - V_{ab}^{[l]}(\mathcal{C})|$.

Define $\mu_n = n^2\rho_n$, we have

$$\begin{aligned}
&\frac{1}{\mu_n}E(O_{kh}^{[l_1l_2]}(\mathcal{E}) \mid \mathcal{C})\\
&= \frac{1}{\mu_n}E\left(\sum_{ij} A_{ij}^{[l_1l_2]} I(e_i^{[l_1]} = k, e_j^{[l_2]} = h) \mid \mathcal{C}\right)\\
&= \frac{1}{n^2}\sum_{ij}\sum_{ab} P_{ab}^{[l_1l_2]} I(e_i^{[l_1]} = k, c_i^{[l_1]} = a) I(e_j^{[l_2]} = h, e_j^{[l_2]} = b).
\end{aligned}$$

Define $H^{[l_1l_2]}(\mathcal{R}(\mathcal{E})) = \frac{1}{\mu_n}E(O^{[l_1l_2]}(\mathcal{E}) \mid \mathcal{C})$, we have

$$H^{[l_1l_2]}(\mathcal{R}(\mathcal{E})) = R^{[l_1]}(\mathcal{E})P^{[l_1l_2]}R^{[l_2]}(\mathcal{E})', \quad 1 \le l_1 \ne l_2 \le L.$$

Similarly, we can define $H^{[l]}(\mathcal{R}(\mathcal{E})) = \frac{1}{\mu_n}E(O^{[l]}(\mathcal{E}) \mid \mathcal{C})$ and write

$$H^{[l]}(\mathcal{R}(\mathcal{E})) = R^{[l]}(\mathcal{E})P^{[l]}R^{[l]}(\mathcal{E})', \quad l = 1, \ldots, L.$$

Write $\mathcal{H} = \{H^{[l]}, H^{[l_1l_2]}, l = 1, \ldots, L, 1 \le l_1 \ne l_2 \le L\}$. Since $J(.)$ is Lipschitz in all its arguments, we have

$$\left|J\left(\frac{\mathcal{O}(\mathcal{E})}{\mu_n}\right) - J\left(\mathcal{H}(\mathcal{R})\right)\right| \le M_1\left[\max_l \| \frac{O^{[l]}(\mathcal{E})}{\mu_n} - H^{[l]}(\mathcal{R}) \|_\infty + \max_{l_1\ne l_2} \| \frac{O^{[l_1l_2]}(\mathcal{E})}{\mu_n} - H^{[l_1l_2]}(\mathcal{R}) \|_\infty\right].$$

Here $||X||_\infty = \max_{kh}|X_{kh}|$. To continue with the proof, we need to use the Bernstein's inequality (Bernstein, 1924).

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

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

Since $A_{ij}^{[l]}$'s in $O^{[l]}(\mathcal{E})$ are independent Bernoulli random variables, applying the Bernstein's inequality, we have

$$P\left(|O_{kh}^{[l]}(\mathcal{E})/\mu_n - H_{kh}^{[l]}(\mathcal{R})| > \omega\right) \le 2\exp\left(-\frac{\omega^2/2}{\mathrm{var}(O_{kh}^{[l]}(\mathcal{E})) + 2\omega/3}\right).$$

Notice that $\mathrm{var}(O_{kh}^{[l]}(\mathcal{E})) \leq 2n^2 \max_{ij} \mathrm{var}(A_{ij}^{[l]})$.

Define $\tau = \max_{ij} \mathrm{var}(A_{ij}^{[l]})$. For any $\epsilon < 3\tau$, if we write $\omega = \epsilon n^2 \rho_n$, then we have

$$\begin{aligned} P\left(\left|\frac{O_{kh}^{[l]}(\mathcal{E})}{\mu_n} - H_{kh}^{[l]}(\mathcal{R})\right| > \epsilon\right) &\leq 2\exp\left(-\frac{\omega^2/2}{\mathrm{var}(O_{kh}^{[l]}(\mathcal{E})) + 2\omega/3}\right) \\ &\leq 2\exp\left(-\frac{\epsilon^2 n^4 \rho_n^2}{8n^2\rho_n\tau}\right) \\ &= 2\exp\left(-\frac{\epsilon^2\mu_n}{8\tau}\right). \end{aligned}$$

The left hand side of the inequality converges to 0 in probability uniformly over $\mathcal{E}$ as $\lambda_n \to \infty$. Following similar arguments, we can show that

$$P\left(\left|\frac{O_{kh}^{[l_1 l_2]}(\mathcal{E})}{\mu_n} - H_{kh}^{[l_1 l_2]}(\mathcal{R})\right| > \epsilon\right)$$

also converges to 0 in probability uniformly as $\lambda_n \to \infty$. Thus, there exists $\epsilon_n \to 0$, such that

$$P\left(\max_{\mathcal{E}}\left|J\left(\frac{\mathcal{O}(\mathcal{E})}{\mu_n}\right) - J\left(\mathcal{H}(\mathcal{R})\right)\right| \leq \epsilon_n\right) \to 1 \quad \text{as} \quad \lambda_n \to \infty. \tag{S2.3}$$

Next we show that $J(\mathcal{H}(\mathcal{R}))$ is uniquely maximized over $\{\mathcal{R} : R^{[l]} \geq 0,\ R^{[l]\prime}\mathbf{1} = \pi^{[l]},\ l = 1, \ldots, L\}$ at $\mathcal{S} = \mathcal{R}(\mathcal{C})$. Since $J(\mathcal{H}(\mathcal{R}))$ is the population version of $J\left(\frac{\mathcal{O}(\mathcal{E})}{\mu_n}\right)$, if $J\left(\frac{\mathcal{O}(\mathcal{E})}{\mu_n}\right)$ is maximized by the true community label $\mathcal{C}$, $J(\mathcal{H}(\mathcal{R}))$ should also be maximized by the true assignment $\mathcal{S}$.

Define

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

Using the equalities

$$\sum_k \left(H_{kk}^{[l]} - \frac{{H_k^{[l]}}^2}{\sum_{kh} H_{kh}^{[l]}}\right) + \sum_{k \neq l}\left(H_{kh}^{[l]} - \frac{H_k^{[l]} H_h^{[l]}}{\sum_{kh} H_{kh}^{[l]}}\right) = 0, \quad l = 1, \ldots, L,$$

and

$$\sum_k \left(H_{kk}^{[l_1 l_2]} - \frac{H_k^{[l_1 l_2]} H_k^{[l_2 l_1]}}{\sum_{kh} H_{kh}^{[l_1 l_2]}}\right) + \sum_{k \neq h}\left(H_{kh}^{[l_1 l_2]} - \frac{H_k^{[l_1 l_2]} H_h^{[l_2 l_1]}}{\sum_{kh} H_{kh}^{[l_1 l_2]}}\right) = 0, \quad 1 \leq l_1 \neq l_2 \leq L.$$

we have

$$
\begin{aligned}
& J(\mathcal{H}(\mathcal{R})) = \sum_{l=1}^{L} J_1(H^{[l]}(\mathcal{R})) + \sum_{l_1 \neq l_2}^{L} J_2(H^{[l_1 l_2]}(\mathcal{R}), H^{[l_2 l_1]}(\mathcal{R})) \\
= \quad & \frac{1}{2}\sum_{l=1}^{L}\sum_{kh} \triangle_{kh} \left(H_{kh}^{[l]}(\mathcal{R}) - \frac{H_k^{[l]}(\mathcal{R}) H_h^{[l]}(\mathcal{R})}{\sum_{kh} H_{kh}^{[l]}(\mathcal{R})} \right) + \frac{1}{2}\sum_{l_1 \neq l_2}^{L}\sum_{kh} \triangle_{kh} \left(H_{kh}^{[l_1 l_2]}(\mathcal{R}) - \frac{H_k^{[l_1 l_2]}(\mathcal{R}) H_h^{[l_2 l_1]}(\mathcal{R})}{\sum_{kh} H_{kh}^{[l_1 l_2]}(\mathcal{R})} \right) \\
= \quad & \frac{1}{2}\sum_{l=1}^{L}\sum_{kh} \triangle_{kh} \left(\sum_{ab} P_{ab}^{[l]} R_{ka}^{[l]}(\mathcal{E}) R_{hb}^{[l]}(\mathcal{E}) - \frac{(\sum_{as} P_{as}^{[l]} R_{ka}^{[l]}(\mathcal{E}) \pi_s^{[l]})(\sum_{bt} P_{bt}^{[l]} R_{hb}^{[l]}(\mathcal{E}) \pi_t^{[l]})}{\sum_{kh} H_{kh}^{[l]}(\mathcal{R})} \right) \\
& + \frac{1}{2}\sum_{l_1 \neq l_2}^{L}\sum_{kh} \triangle_{kh} \left(\sum_{ab} P_{ab}^{[l_1 l_2]} R_{ka}^{[l_1]}(\mathcal{E}) R_{hb}^{[l_2]}(\mathcal{E}) - \frac{(\sum_{as} P_{as}^{[l_1 l_2]} R_{ka}^{[l_1]}(\mathcal{E}) \pi_s^{[l_2]})(\sum_{bt} P_{bt}^{[l_2 l_1]} R_{hb}^{[l_2]}(\mathcal{E}) \pi_t^{[l_1]})}{\sum_{kh} H_{kh}^{[l_1 l_2]}(\mathcal{R})} \right) \\
= \quad & \frac{1}{2}\sum_{l=1}^{L}\sum_{kh}\sum_{ab} \triangle_{kh} R_{ka}^{[l]}(\mathcal{E}) R_{hb}^{[l]}(\mathcal{E}) \left(P_{ab}^{[l]} - \frac{(\sum_s P_{as}^{[l]} \pi_s^{[l]})(\sum_t P_{bt}^{[l]} \pi_t^{[l]})}{\sum_{kh} H_{kh}^{[l]}(\mathcal{R})} \right) \\
& + \frac{1}{2}\sum_{l_1 \neq l_2}^{L}\sum_{kh}\sum_{ab} \triangle_{kh} R_{ka}^{[l_1]}(\mathcal{E}) R_{hb}^{[l_2]}(\mathcal{E}) \left(P_{ab}^{[l_1 l_2]} - \frac{(\sum_s P_{as}^{[l_1 l_2]} \pi_s^{[l_2]})(\sum_t P_{bt}^{[l_2 l_1]} \pi_t^{[l_1]})}{\sum_{kh} H_{kh}^{[l_1 l_2]}(\mathcal{R})} \right) \\
\leq \quad & \frac{1}{2}\sum_{l=1}^{L}\sum_{kh}\sum_{ab} \triangle_{ab} R_{ka}^{[l]}(\mathcal{E}) R_{hb}^{[l]}(\mathcal{E}) \left(P_{ab}^{[l]} - \frac{(\sum_s P_{as}^{[l]} \pi_s^{[l]})(\sum_t P_{bt}^{[l]} \pi_t^{[l]})}{\sum_{kh} H_{kh}^{[l]}(\mathcal{R})} \right) \\
& + \frac{1}{2}\sum_{l_1 \neq l_2}^{L}\sum_{kh}\sum_{ab} \triangle_{ab} R_{ka}^{[l_1]}(\mathcal{E}) R_{hb}^{[l_2]}(\mathcal{E}) \left(P_{ab}^{[l_1 l_2]} - \frac{(\sum_s P_{as}^{[l_1 l_2]} \pi_s^{[l_2]})(\sum_t P_{bt}^{[l_1 l_2]} \pi_t^{[l_1]})}{\sum_{kh} H_{kh}^{[l_1 l_2]}(\mathcal{R})} \right)
\end{aligned}
$$

$$
\begin{aligned}
= \quad & \frac{1}{2}\sum_{l=1}^{L}\sum_{ab} \triangle_{ab} \pi_a^{[l]} \pi_b^{[l]} \left(P_{ab}^{[l]} - \frac{(\sum_s P_{as}^{[l]} \pi_s^{[l]})(\sum_t P_{bt}^{[l]} \pi_t^{[l]})}{\sum_{kh} H_{kh}^{[l]}(\mathcal{S})} \right) \\
& + \frac{1}{2}\sum_{l_1 \neq l_2}^{L}\sum_{ab} \triangle_{ab} \pi_a^{[l_1]} \pi_b^{[l_2]} \left(P_{ab}^{[l_1 l_2]} - \frac{(\sum_s P_{as}^{[l_2 l_1]} \pi_s^{[l_2]})(\sum_t P_{bt}^{[l_1 l_2]} \pi_t^{[l_1]})}{\sum_{kh} H_{kh}^{[l_1 l_2]}(\mathcal{S})} \right) \\
= \quad & \sum_{l=1}^{L} J_1(H^{[l]}(\mathcal{S})) + \sum_{l_1 \neq l_2}^{L} J_2(H^{[l_1 l_2]}(\mathcal{S}), H^{[l_2 l_1]}(\mathcal{S})) = J(\mathcal{H}(\mathcal{S})).
\end{aligned}
$$

Here we used the conditions in Theorem 2 for the inequality, and the relationship that

$$
\sum_{kh} H_{kh}^{[l]}(\mathcal{R}) = \sum_{kh}\sum_{ab} P_{ab}^{[l]} R_{ka}^{[l]}(\mathcal{E}) R_{hb}^{[l]}(\mathcal{E}) = \sum_{ab} P_{ab}^{[l]} \pi_a^{[l]} \pi_b^{[l]} = \sum_{kh} H_{kh}^{[l]}(\mathcal{S})
$$

and

$$
\sum_{kh} H_{kh}^{[l_1 l_2]}(\mathcal{R}) = \sum_{kh}\sum_{ab} P_{ab}^{[l_1 l_2]} R_{ka}^{[l_1]}(\mathcal{E}) R_{hb}^{[l_2]}(\mathcal{E}) = \sum_{ab} P_{ab}^{[l_1 l_2]} \pi_a^{[l_1]} \pi_b^{[l_2]} = \sum_{kh} H_{kh}^{[l_1 l_2]}(\mathcal{S}).
$$

We have shown that $\mathcal{S}$ is a maximizer of $J(\mathcal{H}(\mathcal{R}))$.

Next we need to show that $\mathcal{S}$ is the unique maximizer of $J(\mathcal{H}(\mathcal{R}))$. This can be shown using Lemma 3.2 in Bickel and Chen (2009). Since the inequality $J(\mathcal{H}(\mathcal{R})) \le J(\mathcal{H}(\mathcal{S}))$ holds only if $\triangle_{kh} = \triangle_{ab}$ whenever $R^{[l]}_{ka}(\mathcal{E})R^{[l]}_{hb}(\mathcal{E}) > 0$, $l = 1, \ldots, L$, and $\triangle$ does not have two identical columns, using the results in Lemma 3.2, we have $\mathcal{S}$ uniquely maximizes $J(\mathcal{H}(\mathcal{R}))$.

Now that we have shown that $J\left(\mathcal{H}(\mathcal{R})\right)$ is uniquely maximized by $\mathcal{S}$. By the continuity of $J(.)$ in the neighborhood of $\mathcal{S}$, there exists $\delta_n \to \infty$, such that

$$J\left(\mathcal{H}(\mathcal{R})\right) - J\left(\mathcal{H}(\mathcal{S})\right) \ge 2\epsilon_n \quad \text{for} \quad \eta(\mathcal{E},\mathcal{C}) \ge \delta_n.$$

Here we used the fact that

$$\begin{aligned}\eta(\mathcal{R}(\mathcal{E}),\mathcal{S}) &= \sum_{l=1}^{L}\sum_{ab}|\pi_b^{[l]}V_{ab}^{[l]}(\mathcal{E}) - \pi_b^{[l]}V_{ab}^{[l]}(\mathcal{C})| \\ &\ge (\min_{l,b}\pi_b^{[l]}) \times \sum_{l=1}^{L}\sum_{ab}|V_{ab}^{[l]}(\mathcal{E}) - V_{ab}^{[l]}(\mathcal{C})| = (\min_{l,b}\pi_b^{[l]}) \times \eta(\mathcal{E},\mathcal{C}).\end{aligned}$$

Thus, with (S2.3), we have that

$$\begin{aligned}&P\left(\max_{\mathcal{E}:\ \eta(\mathcal{E},\mathcal{C})\ge\delta_n} J\left(\frac{\mathcal{O}(\mathcal{E})}{\mu_n}\right) \le J\left(\frac{\mathcal{O}(\mathcal{C})}{\mu_n}\right)\right) \\ &\ge P\left(\left|\max_{\mathcal{E}:\ \eta(\mathcal{E},\mathcal{C})\ge\delta_n} J\left(\frac{\mathcal{O}(\mathcal{E})}{\mu_n}\right) - \max_{\mathcal{E}:\ \eta(\mathcal{E},\mathcal{C})\ge\delta_n} J(\mathcal{H}(\mathcal{R}))\right| < \epsilon_n, \left|J\left(\frac{\mathcal{O}(\mathcal{C})}{\mu_n}\right) - J(\mathcal{H}(\mathcal{S}))\right| \le \epsilon_n\right) \to 1.\end{aligned}$$

This implies that

$$P(\eta(\hat{\mathcal{C}},\mathcal{C}) \le \delta_n) \to 1,$$

where

$$\hat{\mathcal{C}} = \arg\max_{\mathcal{E}} J\left(\frac{\mathcal{O}(\mathcal{E})}{\mu_n}\right)$$

Since

$$\begin{aligned}\frac{1}{n}\sum_{l=1}^{L}\sum_{i=1}^{n_l} I(\hat{c}_i^{[l]} \ne c_i^{[l]}) = \sum_{l=1}^{L}\sum_{k}\pi_k^{[l]}(1 - V_{kk}^{[l]}(\hat{\mathcal{C}})) &\le \sum_{l}\sum_{k}(1 - V_{kk}^{[l]}(\hat{\mathcal{C}})) \\ &= \frac{1}{2}\sum_{l=1}^{L}\left(\sum_{k}(1 - V_{kk}^{[l]}(\hat{\mathcal{C}})) + \sum_{k\ne h}V_{kh}^{[l]}(\hat{\mathcal{C}})\right) \\ &= \eta(\hat{\mathcal{C}},\mathcal{C})/2,\end{aligned}$$

we have thus established the consistency property of $\hat{\mathcal{C}}$.

If we replace our proposed null model with the null model for a homogeneous network (as proposed in Zhang and Chen (2016)), it is easy to show that the new homogeneous modularity function can be written as

$$J^*(\mathcal{O}) = \sum_{k=1}^{K}\left(\sum_{l=1}^{L} O_{kk}^{[l]} + \sum_{l_1\neq l_2}^{L} O_{kk}^{[l_1 l_2]}\right) - \sum_{k=1}^{K}\left(\frac{\left(\sum_{l=1}^{L} O_{k}^{[l]} + \sum_{l_1\neq l_2}^{L} O_{k}^{[l_1 l_2]}\right)^2}{\sum_{l=1}^{L}\sum_{kh} O_{kh}^{[l]} + \sum_{l_1\neq l_2}^{L}\sum_{kh} O_{kh}^{[l_1 l_2]}}\right). \tag{S2.4}$$

In this case, we can still show that $J^*(\mathcal{O})$ is uniformly close to its population version. However, due to the complicated formulation of $J^*(\mathcal{O})$ (especially the second term in (S2.4)), we cannot find an interpretable assortative condition that guarantees the population version of $J^*(\mathcal{O})$ is maximized by the true community membership.

S3 Additional Simulation Results

In the following example, we consider heterogeneous networks with two types of nodes ($L = 2$) and two communities ($K = 2$). We consider a heterogeneous SBM structure with probability matrix $P = \begin{pmatrix} P^{[1]} & P^{[12]} \\ P^{[21]} & P^{[2]} \end{pmatrix}$, where $P^{[1]} = p_1\mathbf{1}_K\mathbf{1}_K' + r_1\boldsymbol{I}_K$, $P^{[2]} = p_2\mathbf{1}_K\mathbf{1}_K' + r_2\boldsymbol{I}_K$, $P^{[12]} = P^{[21]} = p_3\mathbf{1}_K\mathbf{1}_K' + r_3\boldsymbol{I}_K$, $\mathbf{1}_K$ is the K-vector of 1's and $\boldsymbol{I}_K$ is the K-by-K identity matrix. We can see that the strength of the community structure is regulated by r_1, r_2 and r_3. This setting is similar to the ones considered in Section 5 of the paper. In this simulation, we set the parameters $p_1 = 0.1$, $r_1 = 0.05$, $p_2 = 0.2$, $r_2 = 0.1$, $p_3 = 0.05$ and gradually change r_3 from 0.025 to 0.125. We set the number of nodes for each type to 200, and assign 100 to each community. We compare our proposed method to a homogeneous model with $K = 4$, estimated using regularized spectral clustering (Rohe et al., 2011). For each r_3 value, we simulate 50 heterogeneous networks from the model. For each heterogeneous network, we apply the proposed method and the homogeneous method. We then calculate the normalized mutual information (NMI) between

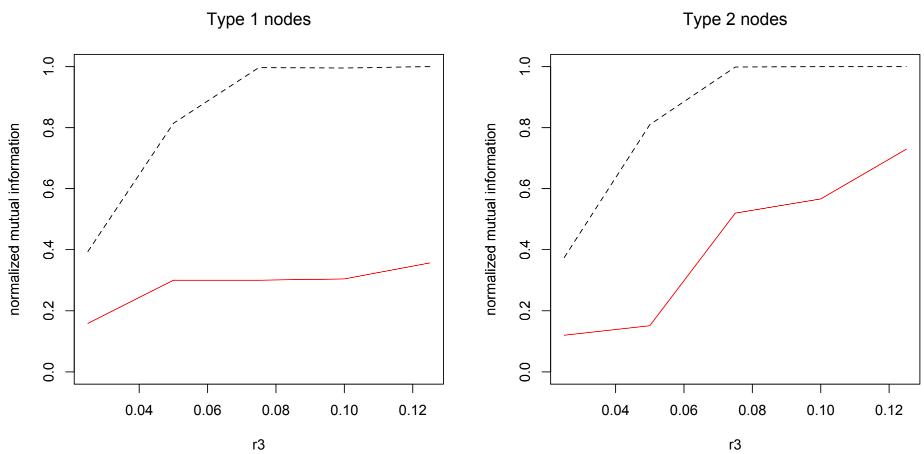

Figure 1: Average NMI between the true community membership and the community membership obtained from the proposed method (dashed line) and the homogenous model with $K = 4$ (solid line).

the obtained community detection results and the true community membership. The average NMI from the 50 simulations is summarized in Figure 1. We can see that the proposed method outperforms the homogeneous method on all values of r_3. This is because type-[1] nodes and type-[2] nodes behave very differently; compared to type-[1] nodes, type-[2] nodes are much more densely connected amongst themselves. The homogenous method does not take into consideration such information and treat all nodes equally. Note that our proposed method has good performance even when edges between type-[1] nodes and type-[2] nodes only have a weak community structure with $r_3 = 0.025$.

References

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.

McKay, B. D. (2010). Subgraphs of random graphs with specified degrees. *Proceedings of the International Congress of Mathematicians 2010* **4**, 2489-2501.

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.