Statistica Sinica: Supplement

NONLINEAR ANALYSIS OF NODAL COVARIATES IN NETWORKS: DIMENSION REDUCTION AND CLUSTERING

Zhonghan Wang[1], Shengbin Zheng[1], Junlong Zhao[1,*]

School of Statistics, Beijing Normal University

Supplementary Materials

The Supplementary Materials include proofs of all the theoretical results in the main text, additional simulations, discussions and theoretical results.

S1 Proofs

S1.1 Proof of Proposition 1

Denote $\phi(X)-\phi(X')$ as $\phi_{XX'}$ for simplicity. Denote $\tilde{G}_0 = G_0 - E(s)E(\phi_{XX'}\otimes\phi_{XX'})$. Based on the independence of X and X', it follows that $E(\phi_{XX'}\otimes\phi_{XX'}) = 2\Sigma_\phi$ and hence $\Sigma_\phi^{-1}\tilde{G}_0 = \Sigma_\phi^{-1}G_0 - 2E(s)I_{\mathcal{H}}$. Therefore, $\Sigma_\phi^{-1}\tilde{G}_0$ and $\Sigma_\phi^{-1}G_0$ share the same eigenfunctions, leading to the fact that the solution to $\max_{\boldsymbol{g}_r\in\Theta_r}\operatorname{tr}(\boldsymbol{g}_r, G_0)$ is also the solution to $\max_{\boldsymbol{g}_r\in\Theta_r}\operatorname{tr}\left(\boldsymbol{g}_r, \tilde{G}_0\right)$ and vice versa. Moreover, for $k = 1,\cdots,r$, we have

$$\begin{aligned}\langle g_k, \tilde{G}_0 g_k\rangle_{\mathcal{H}} &= E(s\langle g_k, \phi_{XX'}\rangle_{\mathcal{H}}^2) - E(s)E(\langle g_k, \phi_{XX'}\rangle_{\mathcal{H}}^2)\\ &= \operatorname{cov}(s, \langle g_k, \phi_{XX'}\rangle_{\mathcal{H}}^2).\end{aligned}$$

One can see that $\operatorname{tr}(\boldsymbol{g}_r, \tilde{G}_0) = \sum_{k=1}^{r} \operatorname{cov}(s, \langle g_k, \phi_{XX'} \rangle_{\mathcal{H}}^2)$ for any $\boldsymbol{g}_r \in \Theta_r$ with $\boldsymbol{g}_r = (g_1, \cdots, g_r)$, which completes the proof.

S1.2 Proof of Proposition 2

Let $\bar{\phi}_i = \phi(X_i) - \hat{\mu}^{\phi}$ and $\mathcal{C} = \{\bar{\phi}_i, i = 1, \cdots, n\}$. Define $[\cdot]_{\mathcal{C}}$ as the coordinate representation about the system $\mathcal{C}$. Specifically, if $u = \sum_{i=1}^{n} c_i \bar{\phi}_i$, then $[u]_{\mathcal{C}} = (c_1, \cdots, c_n)^{\top}$. For an operator $T : \mathcal{H} \to \mathcal{H}$, $[T]_{\mathcal{C}} = ([T\bar{\phi}_1]_{\mathcal{C}}, \cdots, [T\bar{\phi}_n]_{\mathcal{C}})$.

First, we find the coordinate representation of $\hat{G}$. Note that

$$
\begin{aligned}
[\hat{G}\bar{\phi}_i]_{\mathcal{C}} &= \left[(n^2 - n)^{-1} \left[\sum_{i \neq j} s_{ij} \left(\bar{\phi}_i - \bar{\phi}_j \right) \otimes \left(\bar{\phi}_i - \bar{\phi}_j \right) \right] \bar{\phi}_i \right]_{\mathcal{C}} \\
&= \left\{ (n^2 - n)^{-1} \sum_{i \neq j} s_{ij} \left([\bar{\phi}_i]_{\mathcal{C}} - [\bar{\phi}_j]_{\mathcal{C}} \right) \otimes \left([\bar{\phi}_i]_{\mathcal{C}} - [\bar{\phi}_j]_{\mathcal{C}} \right) \right\} \bar{K} [\bar{\phi}_i]_{\mathcal{C}}.
\end{aligned}
$$

It is easy to see that $[\bar{\phi}_i]_{\mathcal{C}} = e_i$ for $i = 1, \cdots, n$, where e_i is an $n \times 1$ vector with the i-th component being 1 and other components being 0. So we have

$$
[\hat{G}\bar{\phi}_i]_{\mathcal{C}} = \left[(n^2 - n)^{-1} \sum_{i \neq j} s_{ij} \left(e_i - e_j \right) \otimes \left(e_i - e_j \right) \right] \bar{K} e_i = (n^2 - n)^{-1} L_S \bar{K} e_i.
$$

Then it follows that $[\hat{G}]_{\mathcal{C}} = ([\hat{G}\bar{\phi}_1]_{\mathcal{C}}, \cdots, [\hat{G}\bar{\phi}_n]_{\mathcal{C}}) = (n^2 - n)^{-1} L_S \bar{K}$ and $[\hat{G}f]_{\mathcal{C}} = [\hat{G}]_{\mathcal{C}} [f]_{\mathcal{C}}$ for $f \in \mathcal{H}$.

Second, according to Fukumizu et al. (2009), the coordinate of the sample covariance operator $\hat{\Sigma}_{\phi}$ is $n^{-1}\bar{K}$. Therefore, by letting $[f]_{\mathcal{C}} = v$, the generalized eigendecomposition problem $\max_{f : \langle f, \hat{\Sigma}_{\phi} f \rangle_{\mathcal{H}} = 1} \langle f, \hat{G} f \rangle_{\mathcal{H}}$ can be written as

$$
\beta_1 = \arg\max_{v} \, (n^2 - n)^{-1} v^{\top} \bar{K} L_S \bar{K} v \ \text{ such that } \ \frac{1}{n} v^{\top} \bar{K}^2 v = 1.
$$

Let $u = \bar{K}vn^{-1/2}$. Then the optimization problem becomes $\nu_1 = \arg\max_u\ (n-1)^{-1}u^\top L_S u$ such that $u^\top u = 1$ and $\bar{K}\beta_1 n^{-1/2} = \nu_1$. The results for other β_k's can be similarly proved.

S1.3 Proof of Theorem 1

Write $G_{0n} = E(\hat{G})$. Then it holds that $\|\hat{G}-G_0\|_{\text{HS}} \le \|\hat{G}-G_{0n}\|_{\text{HS}}+\|G_{0n}-G_0\|_{\text{HS}} = \|\hat{G}-G_{0n}\|_{\text{HS}} + e_n$. We only need to consider the convergence rate of the first term. Write $\phi_{ij} = \phi(X_i) - \phi(X_j)$ and $H_{ij} = s_{ij}\phi_{ij}\otimes\phi_{ij} - E(s_{ij}\phi_{ij}\otimes\phi_{ij})$. Let $\{\sigma(1), \cdots, \sigma(n)\}$ be any permutation of $\{1, \cdots, n\}$. Let $g_{\sigma,t}$ for $t = 1, \cdots, m_\sigma$ be the m_σ groups of $\{\tilde{\sigma}(i) = (\sigma(2i-1), \sigma(2i)), i = 1, \dots, n/2\}$ satisfying the conditional independence property. We have

$$
\begin{aligned}
\|\hat{G}-G_{0n}\|_{\text{HS}} &= \left\| \frac{1}{n(n-1)} \sum_{i\neq j} H_{ij} \right\|_{\text{HS}} \\
&\overset{(i)}{=} \left\| \frac{1}{n!} \sum_{\textit{all } \{\sigma(1),\cdots,\sigma(n)\}} \frac{H_{\sigma(1)\sigma(2)} + \cdots + H_{\sigma(n-1)\sigma(n)}}{n/2} \right\|_{\text{HS}} \\
&\le \frac{1}{n!} \sum_{\textit{all } \{\sigma(1),\cdots,\sigma(n)\}} \left\| \frac{H_{\sigma(1)\sigma(2)} + \cdots + H_{\sigma(n-1)\sigma(n)}}{n/2} \right\|_{\text{HS}} \\
&= \frac{1}{n!} \sum_{\textit{all } \{\sigma(1),\cdots,\sigma(n)\}} \left\| \frac{2}{n} \sum_{t=1}^{m_\sigma} \sum_{(i,j)\in g_{\sigma,t}} H_{ij} \right\|_{\text{HS}} \\
&= \frac{1}{n!} \sum_{\textit{all } \{\sigma(1),\cdots,\sigma(n)\}} \left\| \frac{1}{m_\sigma} \sum_{t=1}^{m_\sigma} Y_t \right\|_{\text{HS}} \qquad \left(\text{set } Y_t = \frac{2m_\sigma}{n} \sum_{(i,j)\in g_{\sigma,t}} H_{ij} \right) \\
&\le \frac{1}{n!} \sum_{\textit{all } \{\sigma(1),\cdots,\sigma(n)\}} \frac{1}{m_\sigma} \sum_{t=1}^{m_\sigma} \|Y_t\|_{\text{HS}},
\end{aligned}
$$

where equation (i) uses the decoupling method in Serfling (1980). Next, we analyse $E\left\|Y_t\right\|_{\mathrm{HS}}^2 = 4m_\sigma^2 n^{-2}\sum_{(i,j)\in g_{\sigma,t}}\sum_{(k,l)\in g_{\sigma,t}} E\langle H_{ij}, H_{kl}\rangle_{\mathrm{HS}}$, where

$$E\langle H_{ij}, H_{kl}\rangle_{\mathrm{HS}} = E\langle s_{ij}\phi_{ij}\otimes\phi_{ij}, s_{kl}\phi_{kl}\otimes\phi_{kl}\rangle_{\mathrm{HS}} - \langle E(s_{ij}\phi_{ij}\otimes\phi_{ij}), E(s_{kl}\phi_{kl}\otimes\phi_{kl})\rangle_{\mathrm{HS}}.$$

For (i,j), $(k,l)\in g_{\sigma,t}$ and $(i,j)\neq(k,l)$, we have

$$\begin{aligned}
&E\langle s_{ij}\phi_{ij}\otimes\phi_{ij}, s_{kl}\phi_{kl}\otimes\phi_{kl}\rangle_{\mathrm{HS}}\\
&= E\left[E\left(\langle s_{ij}\phi_{ij}\otimes\phi_{ij}, s_{kl}\phi_{kl}\otimes\phi_{kl}\rangle_{\mathrm{HS}}|\{\phi_{ij}\}\right)\right]\\
&\overset{(i)}{=} E\left[\langle E(s_{ij}\phi_{ij}\otimes\phi_{ij}|\phi_{ij}), E(s_{kl}\phi_{kl}\otimes\phi_{kl}|\phi_{kl})\rangle_{\mathrm{HS}}\right]\\
&\overset{(ii)}{=} \langle E[E(s_{ij}\phi_{ij}\otimes\phi_{ij}|\phi_{ij})], E[E(s_{kl}\phi_{kl}\otimes\phi_{kl}|\phi_{kl})]\rangle_{\mathrm{HS}}\\
&= \langle E(s_{ij}\phi_{ij}\otimes\phi_{ij}), E(s_{kl}\phi_{kl}\otimes\phi_{kl})\rangle_{\mathrm{HS}},
\end{aligned}$$

where equation (i) is derived from the conditional independence property and Assumption 3 and equation (ii) is derived from the independence of ϕ_{ij} and ϕ_{kl}. So we have $E\left\|Y_t\right\|_{\mathrm{HS}}^2 = 4m_\sigma^2 n^{-2}\sum_{(i,j)\in g_{\sigma,t}} E\left\|H_{ij}\right\|_{\mathrm{HS}}^2$. Let $\tilde{\phi}_{ij} = (s_{ij})^{1/2}\phi_{ij}$. We have

$$\begin{aligned}
E\left\|H_{ij}\right\|_{\mathrm{HS}}^2 \leq\ & E\left\|s_{ij}\phi_{ij}\otimes\phi_{ij}\right\|_{\mathrm{HS}}^2 + 2E\left\|s_{ij}\phi_{ij}\otimes\phi_{ij}\right\|_{\mathrm{HS}}\left\|E(s_{ij}\phi_{ij}\otimes\phi_{ij})\right\|_{\mathrm{HS}}\\
&+ \left\|E(s_{ij}\phi_{ij}\otimes\phi_{ij})\right\|_{\mathrm{HS}}^2\\
&= E\left\|\tilde{\phi}_{ij}\right\|_{\mathcal{H}}^4 + 2E\left\|\tilde{\phi}_{ij}\right\|_{\mathcal{H}}^2\left\|E(s_{ij}\phi_{ij}\otimes\phi_{ij})\right\|_{\mathrm{HS}} + \left\|E(s_{ij}\phi_{ij}\otimes\phi_{ij})\right\|_{\mathrm{HS}}^2.
\end{aligned}$$

According to Assumption 2, $\left\|E(s_{ij}\phi_{ij}\otimes\phi_{ij})\right\|_{\mathrm{HS}} \leq \left(E\left\|\tilde{\phi}_{ij}\right\|_{\mathcal{H}}^4\right)^{1/2} < \infty$, then we have $E\left\|H_{ij}\right\|_{\mathrm{HS}}^2 = O(1)$ for any $i\neq j$, implying that $E\left\|Y_t\right\|_{\mathrm{HS}}^2 = O\left(m_\sigma^2 m_{\sigma,t} n^{-2}\right)$ with $m_{\sigma,t}$ being the number of node pairs in $g_{\sigma,t}$. According to Markov inequality,

$\|Y_t\|_{\mathrm{HS}} = O_p(m_\sigma m_{\sigma,t}^{1/2} n^{-1})$. Therefore, we have

$$m_\sigma^{-1} \sum_{t=1}^{m_\sigma} \|Y_t\|_{\mathrm{HS}} = O_p\left(n^{-1} m_\sigma \max_t m_{\sigma,t}^{1/2}\right) = O_p\left(m_\sigma n^{-1/2}\right),$$

where the last equation is derived from the fact that $\max_t m_{\sigma,t}^{1/2} = O(n^{1/2})$. Recall $m_{\mathrm{net}} = \max_{\{\sigma(1),\cdots,\sigma(n)\}} m_\sigma$. So we have

$$\frac{1}{n!} \sum_{all\ \{\sigma(1),\cdots,\sigma(n)\}} \frac{1}{m_\sigma} \sum_{t=1}^{m_\sigma} \|Y_t\|_{\mathrm{HS}} = O_p\left(\frac{m_{\mathrm{net}}}{n^{1/2}}\right),$$

implying that $\|\hat{G} - G_{0n}\|_{\mathrm{HS}} = O_p(m_{\mathrm{net}}/n^{1/2})$ and $\|\hat{G} - G_0\|_{\mathrm{HS}} = O_p(m_{\mathrm{net}}/n^{1/2} + e_n)$.

S1.4 Proof of Theorem 2

Step 1: Convergence of $\|(\hat{\Sigma}_\phi + \epsilon_n I_{\mathcal{H}})^{-1}\hat{G} - \Sigma_\phi^{-1} G_0\|_{\mathrm{HS}}$**.** Note that

$$\begin{aligned} \|(\hat{\Sigma}_\phi + \epsilon_n I_{\mathcal{H}})^{-1}\hat{G} - \Sigma_\phi^{-1} G_0\|_{\mathrm{HS}} \le{}& \|(\Sigma_\phi + \epsilon_n I_{\mathcal{H}})^{-1} G_0 - \Sigma_\phi^{-1} G_0\|_{\mathrm{HS}} \\ &+ \|[(\hat{\Sigma}_\phi + \epsilon_n I_{\mathcal{H}})^{-1} - (\Sigma_\phi + \epsilon_n I_{\mathcal{H}})^{-1}] G_0\|_{\mathrm{HS}} \\ &+ \|(\hat{\Sigma}_\phi + \epsilon_n I_{\mathcal{H}})^{-1}(\hat{G} - G_0)\|_{\mathrm{HS}}. \end{aligned} \tag{S1.1}$$

We first analyse the first term on the right hand side of (S1.1). We have

$$\begin{aligned} (\Sigma_\phi + \epsilon_n I_{\mathcal{H}})^{-1} G_0 - \Sigma_\phi^{-1} G_0 &= (\Sigma_\phi + \epsilon_n I_{\mathcal{H}})^{-1} \Sigma_\phi (\Sigma_\phi^{-1} G_0) - \Sigma_\phi^{-1} G_0 \\ &= -\epsilon_n (\Sigma_\phi + \epsilon_n I_{\mathcal{H}})^{-1} (\Sigma_\phi^{-1} G_0). \end{aligned}$$

According to Theorem 1 of Douglas (1966), there exists some bounded operator $\mathcal{Z}$ such that $G_0 = \Sigma_\phi^2 \mathcal{Z}$ by Assumption 5. It follows that $(\Sigma_\phi + \epsilon_n I_{\mathcal{H}})^{-1} G_0 - \Sigma_\phi^{-1} G_0 = -\epsilon_n (\Sigma_\phi + \epsilon_n I_{\mathcal{H}})^{-1} \Sigma_\phi \mathcal{Z}$ and

$$\|(\Sigma_\phi + \epsilon_n I_{\mathcal{H}})^{-1} G_0 - \Sigma_\phi^{-1} G_0\|_{\mathrm{HS}} \le \epsilon_n \|(\Sigma_\phi + \epsilon_n I_{\mathcal{H}})^{-1} \Sigma_\phi\|_{\mathrm{OP}} \|\mathcal{Z}\|_{\mathrm{HS}} = O(\epsilon_n),$$

where the last equation comes from the fact that $\|(\Sigma_\phi + \epsilon_n I_{\mathcal{H}})^{-1}\Sigma_\phi\|_{\mathrm{OP}} < \|(\Sigma_\phi + \epsilon_n I_{\mathcal{H}})^{-1}(\Sigma_\phi + \epsilon_n I_{\mathcal{H}})\|_{\mathrm{OP}} = 1$.

For the second term on the right hand side of (S1.1), since

$$\begin{aligned}
&[(\hat{\Sigma}_\phi + \epsilon_n I_{\mathcal{H}})^{-1} - (\Sigma_\phi + \epsilon_n I_{\mathcal{H}})^{-1}]G_0 \\
&= (\hat{\Sigma}_\phi + \epsilon_n I_{\mathcal{H}})^{-1}[(\Sigma_\phi + \epsilon_n I_{\mathcal{H}}) - (\hat{\Sigma}_\phi + \epsilon_n I_{\mathcal{H}})](\Sigma_\phi + \epsilon_n I_{\mathcal{H}})^{-1}G_0 \\
&= (\hat{\Sigma}_\phi + \epsilon_n I_{\mathcal{H}})^{-1}[\Sigma_\phi - \hat{\Sigma}_\phi](\Sigma_\phi + \epsilon_n I_{\mathcal{H}})^{-1}G_0,
\end{aligned}$$

we have

$$\begin{aligned}
&\|[(\hat{\Sigma}_\phi + \epsilon_n I_{\mathcal{H}})^{-1} - (\Sigma_\phi + \epsilon_n I_{\mathcal{H}})^{-1}]G_0\|_{\mathrm{HS}} \\
&\le \|(\hat{\Sigma}_\phi + \epsilon_n I_{\mathcal{H}})^{-1}\|_{\mathrm{OP}}\|\Sigma_\phi - \hat{\Sigma}_\phi\|_{\mathrm{OP}}\|(\Sigma_\phi + \epsilon_n I_{\mathcal{H}})^{-1}\Sigma_\phi\|_{\mathrm{OP}}\|\Sigma_\phi^{-1}G_0\|_{\mathrm{HS}}.
\end{aligned}$$

According to Lemma 4 in Virta et al. (2022) and Assumption 4, we have $\|\Sigma_\phi - \hat{\Sigma}_\phi\|_{\mathrm{OP}} = O_p(n^{-1/2})$. Then by Assumption 6, it follows that $\|[(\hat{\Sigma}_\phi + \epsilon_n I_{\mathcal{H}})^{-1} - (\Sigma_\phi + \epsilon_n I_{\mathcal{H}})^{-1}]G_0\|_{\mathrm{HS}} = O_p([\epsilon_n n^{1/2}]^{-1})$.

For the last term on the right hand side of (S1.1), we have

$$\|(\hat{\Sigma}_\phi + \epsilon_n I_{\mathcal{H}})^{-1}(\hat{G} - G_0)\|_{\mathrm{HS}} \le \|(\hat{\Sigma}_\phi + \epsilon_n I_{\mathcal{H}})^{-1}\|_{\mathrm{OP}}\|\hat{G} - G_0\|_{\mathrm{HS}} = O_p\left(\frac{m_{\mathrm{net}}}{\epsilon_n n^{1/2}} + \frac{e_n}{\epsilon_n}\right).$$

Combining all results above, we have

$$\|(\hat{\Sigma}_\phi + \epsilon_n I_{\mathcal{H}})^{-1}\hat{G} - \Sigma_\phi^{-1}G_0\|_{\mathrm{HS}} = O_p\left(\epsilon_n + \frac{m_{\mathrm{net}}}{\epsilon_n n^{1/2}} + \frac{e_n}{\epsilon_n}\right).$$

Step 2: Convergence of $\|\hat{\psi}_k - c_0\psi_k\|_{\mathcal{H}}$.

Recall that ψ_k's and $\hat{\psi}_k$'s are the unit eigenfunctions of $\Sigma_\phi^{-1}G_0$ and $(\hat{\Sigma}_\phi + \epsilon_n I_{\mathcal{H}})^{-1}\hat{G}$ respectively. By Lemma 8 in the supplementary material of Ying and

Yu (2022), we have

$$\|\hat{\psi}_k - c_0\psi_k\|_{\mathcal{H}} = O_p\left(\epsilon_n + \frac{m_{\text{net}}}{\epsilon_n n^{1/2}} + \frac{e_n}{\epsilon_n}\right),$$

where $c_0 \in \{-1, 1\}$ such that $c_0\langle\hat{\psi}_k, \psi_k\rangle_{\mathcal{H}} > 0$.

Step 3: Convergence of $\|\hat{f}_{\epsilon_n,k} - c_0 f_k\|_{\mathcal{H}}$. Note that $\hat{f}_{\epsilon_n,k} - c_0 f_k = \hat{\psi}_k/\|(\hat{\Sigma}_\phi + \epsilon_n I_{\mathcal{H}})^{1/2}\hat{\psi}_k\|_{\mathcal{H}} - c_0\psi_k/\|\Sigma_\phi^{1/2}\psi_k\|_{\mathcal{H}}$. It follows that

$$\begin{aligned}\|\hat{f}_{\epsilon_n,k} - c_0 f_k\|_{\mathcal{H}} &= \frac{1}{\|\Sigma_\phi^{1/2}\psi_k\|_{\mathcal{H}}}\left\| c_0\psi_k - \hat{\psi}_k\frac{\|\Sigma_\phi^{1/2}\psi_k\|_{\mathcal{H}}}{\|(\hat{\Sigma}_\phi + \epsilon_n I_{\mathcal{H}})^{1/2}\hat{\psi}_k\|_{\mathcal{H}}}\right\|_{\mathcal{H}} \\ &\leq \frac{1}{\|\Sigma_\phi^{1/2}\psi_k\|_{\mathcal{H}}}\left[\|\hat{\psi}_k - c_0\psi_k\|_{\mathcal{H}} + \left\|\hat{\psi}_k\frac{\|(\hat{\Sigma}_\phi + \epsilon_n I_{\mathcal{H}})^{1/2}\hat{\psi}_k\|_{\mathcal{H}} - \|\Sigma_\phi^{1/2}\psi_k\|_{\mathcal{H}}}{\|(\hat{\Sigma}_\phi + \epsilon_n I_{\mathcal{H}})^{1/2}\hat{\psi}_k\|_{\mathcal{H}}}\right\|_{\mathcal{H}}\right].\end{aligned}$$

According to Assumption 5, we have $\text{span}(\Sigma_\phi^{-1}G_0) \subseteq \text{span}(\Sigma_\phi)$. Since $\psi_k \in \text{span}(\Sigma_\phi^{-1}G_0)$, it follows that $\psi_k \in \text{span}(\Sigma_\phi)$, and there exists some constant $c_1 > 0$ such that $\|\Sigma_\phi^{1/2}\psi_k\|_{\mathcal{H}} > c_1$ for $k = 1, \cdots, r$.

Next, we study the convergence of $\|(\hat{\Sigma}_\phi + \epsilon_n I_{\mathcal{H}})^{1/2}\hat{\psi}_k\|_{\mathcal{H}} - \|\Sigma_\phi^{1/2}\psi_k\|_{\mathcal{H}}$. Note that

$$\begin{aligned}\|(\hat{\Sigma}_\phi + \epsilon_n I_{\mathcal{H}})^{1/2}\hat{\psi}_k\|_{\mathcal{H}} - \|\Sigma_\phi^{1/2}\psi_k\|_{\mathcal{H}} \leq &\|(\hat{\Sigma}_\phi + \epsilon_n I_{\mathcal{H}})^{1/2}(c_0\hat{\psi}_k - \psi_k)\|_{\mathcal{H}} \\ &+ \|[(\hat{\Sigma}_\phi + \epsilon_n I_{\mathcal{H}})^{1/2} - \Sigma_\phi^{1/2}]\psi_k\|_{\mathcal{H}},\end{aligned}$$

where the first and second terms on the right hand side are $O_p(\|c_0\hat{\psi}_k - \psi_k\|_{\mathcal{H}})$ and $O_p\left(\|(\hat{\Sigma}_\phi + \epsilon_n I_{\mathcal{H}})^{1/2} - \Sigma_\phi^{1/2}\|_{\text{OP}}\right)$ respectively, and the latter is

$$O_p\left(\|(\hat{\Sigma}_\phi + \epsilon_n I_{\mathcal{H}})^{1/2} - (\Sigma_\phi + \epsilon_n I_{\mathcal{H}})^{1/2}\|_{\text{OP}} + \|(\Sigma_\phi + \epsilon_n I_{\mathcal{H}})^{1/2} - \Sigma_\phi^{1/2}\|_{\text{OP}}\right).$$

From the equation $A^{1/2} - B^{1/2} = A^{1/2}(B^{3/2} - A^{3/2})B^{-3/2} + (A^2 - B^2)B^{-3/2}$ and

Lemma 8 in Fukumizu et al. (2007), we have

$$\begin{aligned}
&\|(\hat{\Sigma}_\phi + \epsilon_n I_{\mathcal{H}})^{1/2} - (\Sigma_\phi + \epsilon_n I_{\mathcal{H}})^{1/2}\|_{\mathrm{OP}} \\
&= O_p\Big(\|(\Sigma_\phi + \epsilon_n I_{\mathcal{H}})^{1/2}\|_{\mathrm{OP}}\|\hat{\Sigma}_\phi - \Sigma_\phi\|_{\mathrm{OP}}\|\|(\hat{\Sigma}_\phi + \epsilon_n I_{\mathcal{H}})^{-3/2}\|_{\mathrm{OP}} + \\
&\quad (\|\hat{\Sigma}_\phi\|_{\mathrm{OP}} + \|\Sigma_\phi\|_{\mathrm{OP}}\|)\|\hat{\Sigma}_\phi - \Sigma_\phi\|_{\mathrm{OP}}\|\|(\hat{\Sigma}_\phi + \epsilon_n I_{\mathcal{H}})^{-3/2}\|_{\mathrm{OP}}\Big) \\
&= O_p\left(\|\hat{\Sigma}_\phi - \Sigma_\phi\|_{\mathrm{OP}}\|(\hat{\Sigma}_\phi + \epsilon_n I_{\mathcal{H}})^{-3/2}\|_{\mathrm{OP}}\right)
\end{aligned}$$

and hence is $O_p\left(n^{-1/2}\epsilon_n^{-3/2}\right)$. Similarly, we have $\|(\Sigma_\phi + \epsilon_n I_{\mathcal{H}})^{1/2} - \Sigma_\phi^{1/2}\|_{\mathrm{OP}} = O_p(\epsilon_n^{1/2})$. Therefore, we have

$$\|(\hat{\Sigma}_\phi + \epsilon_n I_{\mathcal{H}})^{1/2}\hat{\psi}_k\|_{\mathcal{H}} - \|\Sigma_\phi^{1/2}\psi_k\|_{\mathcal{H}} = O_p\left(\epsilon_n^{1/2} + \frac{m_{\mathrm{net}}}{\epsilon_n n^{1/2}} + \frac{e_n}{\epsilon_n} + \frac{1}{n^{1/2}\epsilon_n^{3/2}}\right).$$

For $\|(\hat{\Sigma}_\phi + \epsilon_n I_{\mathcal{H}})^{1/2}\hat{\psi}_k\|_{\mathcal{H}}$, since $\|(\hat{\Sigma}_\phi + \epsilon_n I_{\mathcal{H}})^{1/2}\hat{\psi}_k\|_{\mathcal{H}} = \|\Sigma_\phi^{1/2}\psi_k\|_{\mathcal{H}} + [\|(\hat{\Sigma}_\phi + \epsilon_n I_{\mathcal{H}})^{1/2}\hat{\psi}_k\|_{\mathcal{H}} - \|\Sigma_\phi^{1/2}\psi_k\|_{\mathcal{H}}]$, it is easy to see that there exists some constant $c_2 > 0$ such that $\|(\hat{\Sigma}_\phi + \epsilon_n I_{\mathcal{H}})^{1/2}\hat{\psi}_k\|_{\mathcal{H}} > c_2$ with probability approaching to 1.

Combining all the bounds above, we have

$$\|\hat{f}_{\epsilon_n,k} - c_0 f_k\|_{\mathcal{H}} = O_p\left(\epsilon_n^{1/2} + \frac{m_{\mathrm{net}}}{\epsilon_n n^{1/2}} + \frac{e_n}{\epsilon_n} + \frac{1}{n^{1/2}\epsilon_n^{3/2}}\right),$$

where $c_0 \in \{-1, 1\}$ such that $c_0\langle\hat{f}_{\epsilon_n,k}, f_k\rangle_{\mathcal{H}} > 0$.

S1.5 Proof of Theorem 3

The difference between the proof of Theorem 3 and those of Theorems 1 and 2 is in proving the rate of $\|Y_t\|_{\mathrm{HS}}$ for $t = 1, \cdots, m_\sigma$. According to Lemma 1 in the supplementary material of Zhao et al. (2022), for any permutation $\{\sigma(1), \cdots, \sigma(n)\}$, the index pairs $[\tilde{\sigma}(i) = \{\sigma(2i-1), \sigma(2i)\}, i = 1, \ldots, n/2]$ can always be split into $\tilde{m}_\sigma$

groups $g_{\sigma,1},\cdots,g_{\sigma,\tilde{m}_\sigma}$ such that the conditional independence property holds for the first $\tilde{m}_\sigma-1$ groups, where $\tilde{m}_\sigma \le \tilde{m}_{\text{net}} = \log(n/4)/\log\{4d_{\max}/(4d_{\max}-1)\}+1$ and $d_{\max} = \max_{i=1,\ldots,n}\text{Card}(\{j : N_j \cap N_i \neq \varnothing\})$. So we need to analyse $\|Y_t\|_{\text{HS}}$ for $t=1,\cdots,\tilde{m}_\sigma-1$ and $t=\tilde{m}_\sigma$ separately.

For $t=1,\cdots,\tilde{m}_\sigma-1$, similar to the proof of Theorem 1, we have $\|Y_t\|_{\text{HS}} = O_p(\tilde{m}_\sigma n^{-1/2})$.

For $t=\tilde{m}_\sigma$, note that $Y_{\tilde{m}_\sigma} = |g_{\sigma,\tilde{m}_\sigma}|^{-1}\sum_{(i,j)\in g_{\sigma,\tilde{m}_\sigma}} 2\tilde{m}_\sigma|g_{\sigma,\tilde{m}_\sigma}|n^{-1}H_{ij}$, where $H_{ij} = s_{ij}\phi_{ij}\otimes\phi_{ij} - E(s_{ij}\phi_{ij}\otimes\phi_{ij})$. Denoting $2\tilde{m}_\sigma|g_{\tilde{\sigma},\tilde{m}_\sigma}|n^{-1}H_{ij}$ by $\bar{H}_{ij}$, by Jensen's inequality, we have

$$E\|Y_{\tilde{m}_\sigma}\|_{\text{HS}}^2 \le \frac{1}{|g_{\sigma,\tilde{m}_\sigma}|}\sum_{(i,j)\in g_{\sigma,\tilde{m}_\sigma}} E\|\bar{H}_{ij}\|_{\text{HS}}^2.$$

For any $(i,j)\in g_{\sigma,\tilde{m}_\sigma}$, we have

$$E\|\bar{H}_{ij}\|_{\text{HS}}^2 = \frac{4\tilde{m}_\sigma^2|g_{\sigma,\tilde{m}_\sigma}|^2}{n^2}E\|H_{ij}\|_{\text{HS}}^2.$$

As shown in the proof of Theorem 1, we have $E\|H_{ij}\|_{\text{HS}}^2 = O(1)$, implying that $E\|\bar{H}_{ij}\|_{\text{HS}}^2 = O(4\tilde{m}_\sigma^2|g_{\tilde{\sigma},\tilde{m}_\sigma}|^2 n^{-2})$. According to Lemma 1 in the supplementary material of Zhao et al. (2022), we have $|g_{\tilde{\sigma},\tilde{m}_\sigma}| \le 4d_{\max}$. By Assumption 7, it follows that $E\|\bar{H}_{ij}\|_{\text{HS}}^2 = O(\tilde{m}_\sigma^2 n^{-1})$ and hence $\|Y_{\tilde{m}_\sigma}\|_{\text{HS}} = O_p(\tilde{m}_\sigma n^{-1/2})$. Therefore, $\frac{1}{m_\sigma}\sum_{t=1}^{m_\sigma}\|Y_t\|_{\text{HS}} = O_p(\tilde{m}_\sigma n^{-1/2})$. The rest of the proof is exactly the same as those of Theorems 1 and 2.

S1.6 Proof of Lemma 1

According to the triangle inequality, we have

$$\|\hat{\boldsymbol{f}}_{r,\epsilon_n}(\phi(X_i)) - \boldsymbol{f}_r(\mu^{\phi}_{C_i})\| \leq \|\hat{\boldsymbol{f}}_{r,\epsilon_n}(\phi(X_i)) - \boldsymbol{f}_r(\phi(X_i))\| + \|\boldsymbol{f}_r(\phi(X_i)) - \boldsymbol{f}_r(\mu^{\phi}_{C_i})\|.$$

For $\|\boldsymbol{f}_r(\phi(X_i)) - \boldsymbol{f}_r(\mu^{\phi}_{C_i})\|$, by the Cauchy-Schwarz inequality, we have

$$\|\boldsymbol{f}_r(\phi(X_i)) - \boldsymbol{f}_r(\mu^{\phi}_{C_i})\| \leq \left(r \max_k \|f_k\|^2_{\mathcal{H}} \|\phi(X_i) - \mu^{\phi}_{C_i}\|^2_{\mathcal{H}}\right)^{1/2}.$$

Since $\max_{i=1,\cdots,n} \mathcal{K}(X_i, X_i) = O_p(\delta^2_{\phi})$ and $\max_{t=1,\cdots,N_c} \|\mu_t\|_{\mathcal{H}} = O(\delta_{\mu})$, we have

$$\left(r \max_k \|f_k\|^2_{\mathcal{H}} \|\phi(X_i) - \mu^{\phi}_{C_i}\|^2_{\mathcal{H}}\right)^{1/2} = O_p\left(r^{1/2}(\delta_{\phi} + \delta_{\mu}) \max_k \|f_k\|_{\mathcal{H}}\right).$$

Since $f_k = \psi_k / \|\Sigma^{1/2}_{\phi}\psi_k\|_{\mathcal{H}}$ where $\|\psi_k\|_{\mathcal{H}} = 1$ and $\|\Sigma^{1/2}_{\phi}\psi_k\|_{\mathcal{H}} > c_1$ with some constant $c_1 > 0$, we have $\|\boldsymbol{f}_r(\phi(X_i)) - \boldsymbol{f}_r(\mu^{\phi}_{C_i})\| = O_p(r^{1/2}(\delta_{\phi} + \delta_{\mu}))$.

For $\|\hat{\boldsymbol{f}}_{r,\epsilon_n}(\phi(X_i)) - \boldsymbol{f}_r(\phi(X_i))\|$, it is easy to see that it is equal to

$$\left(\sum_{k=1}^{r} \langle \hat{f}_{\epsilon_n,k} - f_k, \phi(X_i) \otimes \phi(X_i)(\hat{f}_{\epsilon_n,k} - f_k)\rangle_{\mathcal{H}}\right)^{1/2}.$$

Without loss of generality, we assume that $\hat{f}_{\epsilon_n,k}$ converges to f_k instead of $-f_k$. Then for any k, $\langle \hat{f}_{\epsilon_n,k} - f_k, \phi(X_i) \otimes \phi(X_i)(\hat{f}_{\epsilon_n,k} - f_k)\rangle_{\mathcal{H}} \leq \lambda_1(\phi(X_i) \otimes \phi(X_i))\|\hat{f}_{\epsilon_n,k} - f_k\|^2_{\mathcal{H}} \leq \|\phi(X_i) \otimes \phi(X_i)\|_{\mathrm{HS}} \|\hat{f}_{\epsilon_n,k} - f_k\|^2_{\mathcal{H}}$, where $\lambda_1(\phi(X_i) \otimes \phi(X_i))$ denotes the largest eigenvalue of $\phi(X_i) \otimes \phi(X_i)$. According to Theorem 2, we have

$$\|\hat{f}_{\epsilon_n,k} - f_k\|_{\mathcal{H}} = O_p\left(\epsilon_n^{1/2} + \frac{m_{\mathrm{net}}}{\epsilon_n n^{1/2}} + \frac{e_n}{\epsilon_n} + \frac{1}{n^{1/2}\epsilon_n^{3/2}}\right).$$

Since $\|\phi(X_i) \otimes \phi(X_i)\|_{\mathrm{HS}} = \mathcal{K}(X_i, X_i) = O_p(\delta^2_{\phi})$, it follows that for any k,

$$\langle \hat{f}_{\epsilon_n,k} - f_k, \phi(X_i) \otimes \phi(X_i)(\hat{f}_{\epsilon_n,k} - f_k)\rangle_{\mathcal{H}} = O_p\left(\delta^2_{\phi}\left[\epsilon_n^{1/2} + \frac{m_{\mathrm{net}}}{\epsilon_n n^{1/2}} + \frac{e_n}{\epsilon_n} + \frac{1}{n^{1/2}\epsilon_n^{3/2}}\right]^2\right),$$

implying that

$$
\begin{aligned}
&\|\hat{\boldsymbol{f}}_{r,\epsilon_n}(\phi(X_i)) - \boldsymbol{f}_r(\mu^{\phi}_{C_i})\| \\
&= O_p\left(r^{1/2}(\delta_\phi + \delta_\mu) + r^{1/2}\delta_\phi\left[\epsilon_n^{1/2} + \frac{m_{\text{net}}}{\epsilon_n n^{1/2}} + \frac{e_n}{\epsilon_n} + \frac{1}{n^{1/2}\epsilon_n^{3/2}}\right]\right) \\
&= O_p\left(r^{1/2}\delta_\mu + r^{1/2}\delta_\phi\left[1 + \frac{m_{\text{net}}}{\epsilon_n n^{1/2}} + \frac{e_n}{\epsilon_n} + \frac{1}{n^{1/2}\epsilon_n^{3/2}}\right]\right).
\end{aligned}
$$

S1.7 Proof of Theorem 4

We give the proof in a similar framework as the proof of Theorem S4 in Hu and Wang (2024). Denote the convergence rate of $\|\hat{\boldsymbol{f}}_{r,\epsilon_n}(\phi(X_i)) - \boldsymbol{f}_r(\mu^{\phi}_{C_i})\|$ obtained in Lemma 1 by R_{ϵ_n}. Then there exists $D_n = O(1)$ such that the probability that $\|\hat{\boldsymbol{f}}_{r,\epsilon_n}(\phi(X_i)) - \boldsymbol{f}_r(\mu^{\phi}_{C_i})\|^2 \le D_n R^2_{\epsilon_n}$ tends to 1 as n tends to infinity. For simplicity, let $\hat{r}_i = \hat{\boldsymbol{f}}_{r,\epsilon_n}(\phi(X_i))$ and $r_i = \boldsymbol{f}_r(\mu^{\phi}_{C_i})$ for $i = 1, \cdots, n$. Then we have $\|\hat{r}_i - r_i\| \le D_n^{1/2} R_{\epsilon_n}$ with probability approaching to 1. Denote the cluster centers of r_i's by m_k for $k = 1, \cdots, N_c$. It is easy to see that $m_k = r_i$ when $C_i = k$ for $k = 1, \cdots, N_c$. Let $\hat{C} = (\hat{C}_1, \cdots, \hat{C}_n)$ denote the estimated labels by K-means. Let $\hat{m} = (\hat{m}_1, \cdots, \hat{m}_{N_c})$ denote the estimated centers by K-means. Define the within-group distance as

$$
L(\hat{C}, \hat{m}) = \sum_{i=1}^{n} \|\hat{r}_i - \hat{m}_{\hat{C}_i}\|^2.
$$

For the true labels C_i's and true centers m_k's, according to the procedure of K-means, we have $L(\hat{C}, \hat{m}) \le L(C, m)$. For $k = 1, \cdots, N_c$, let $\rho(k) = \arg\min_{1\le j\le N_c} \|m_k - \hat{m}_j\|$, that is, $\rho(k)$ is the index of the estimated center which is closest to m_k.

We first control the distance between m_k and $\hat{m}_{\rho(k)}$. For any $k \in \{1, \cdots, N_c\}$, we have

$$\begin{aligned}
L(\hat{C}, \hat{m}) &\geq \sum_{C_i=k} \|\hat{r}_i - \hat{m}_{\hat{C}_i}\|^2 \\
&\geq \sum_{C_i=k} (-\|\hat{r}_i - m_k\|^2 + \frac{1}{2}\|m_k - \hat{m}_{\hat{C}_i}\|^2) \\
&= \sum_{C_i=k} (-\|\hat{r}_i - m_{C_i}\|^2 + \frac{1}{2}\|m_k - \hat{m}_{\hat{C}_i}\|^2) \\
&\geq -\sum_{i=1}^{n} \|\hat{r}_i - m_{C_i}\|^2 + \frac{1}{2} n_k \|m_k - \hat{m}_{\rho(k)}\|^2 \\
&= -L(C, m) + \frac{1}{2} n_k \|m_k - \hat{m}_{\rho(k)}\|^2.
\end{aligned}$$

Combining with $L(\hat{C}, \hat{m}) \leq L(C, m)$, we have $\|m_k - \hat{m}_{\rho(k)}\|^2 \leq 4L(C, m)/n_k \leq 4L(C, m)/c_b n$ where the last inequality comes from Assumption 8. According to Assumption 9, each m_k is paired with a unique $\hat{m}_{\rho(k)}$ such that $\|m_k - \hat{m}_{\rho(k)}\| \leq 2(L(C, m)/c_b n)^{1/2}$. Since $L(C, m) = \sum_{i=1}^{n} \|\hat{r}_i - m_{C_i}\|^2 = \sum_{i=1}^{n} \|\hat{r}_i - r_i\|^2$, according to the results in Lemma 1, the probability that $L(C, m) \leq nD_n R_{\epsilon_n}^2$ tends to 1 when $n \to \infty$. Therefore, we find that $\|m_k - \hat{m}_{\rho(k)}\| \leq 2(D_n R_{\epsilon_n}^2 / c_b)^{1/2}$ with probability approaching to 1.

Suppose that there exists a node i such that $\hat{C}_i \neq \rho(C_i)$. Next, we will show that the corresponding $\hat{C}$ is not the optimal solution to K-means. Consider a different solution $(\hat{C}', \hat{m}')$ with $\hat{m}' = \hat{m}$ and $\hat{C}'$ differs with $\hat{C}$ only at the i-th component, that $\hat{C}'_i = \rho(C_i)$. Then we have $L(\hat{C}', \hat{m}') - L(\hat{C}, \hat{m}) = \|\hat{r}_i - \hat{m}_{\hat{C}'_i}\|^2 - \|\hat{r}_i - \hat{m}_{\hat{C}_i}\|^2$. Note that with probability approaching to 1, $\|\hat{r}_i - \hat{m}_{\hat{C}'_i}\| \leq \|\hat{r}_i - r_i\| +$

$\|r_i - \hat{m}_{\rho(C_i)}\| = \|\hat{r}_i - r_i\| + \|m_{C_i} - \hat{m}_{\rho(C_i)}\| \leq D_n^{1/2} R_{\epsilon_n} + 2(D_n R_{\epsilon_n}^2/c_b)^{1/2}$. Assume that $\hat{C}_i = \rho(t)$. Then it is easy to see that $t \neq C_i$. According to Assumption 9, we have the following result with probability approaching to 1

$$\begin{aligned}
\|\hat{r}_i - \hat{m}_{\hat{C}_i}\| &\geq \|m_t - m_{C_i}\| - \|m_t - \hat{m}_{\rho(t)}\| - \|r_i - \hat{r}_i\| \quad \text{(Note that } m_{C_i} = r_i) \\
&\geq (4 + 4/{c_b}^{1/2}) D_n^{1/2} R_{\epsilon_n} - 2(D_n R_{\epsilon_n}^2/c_b)^{1/2} - D_n^{1/2} R_{\epsilon_n} \\
&= 3 D_n^{1/2} R_{\epsilon_n} + 2(D_n R_{\epsilon_n}^2/c_b)^{1/2} \\
&\geq D_n^{1/2} R_{\epsilon_n} + 2(D_n R_{\epsilon_n}^2/c_b)^{1/2} \\
&\geq \|\hat{r}_i - \hat{m}_{\hat{C}_i'}\|,
\end{aligned}$$

which contradicts the optimality of $\hat{C}$. Combining all results above, we prove that each node can be exactly recovered.

S1.8 Proof of Proposition 3

We give a more general version of Proposition 3 in Proposition B and the proof of Proposition 3 is included in the proof of Proposition B.

S2 Additional simulations

S2.1 Results of the adjusted Rand index

In this section, we report the results of the adjusted Rand Index (ARI) under the same settings as those in Sections 5-6 in the main text. Since the results and conclusions of ARI are consistent with those of the community detection accuracy (ACC) reported in the main text, we only present the ARI values here without further discussions. For detailed interpretations, readers can refer to Sections 5-6 in the main text.

- **Community detection performance.** The settings are referred to Section 5.1 in the main text. We present the average ARIs along with the standard deviations in Figures 1-2.
- **Performance under varying n, p and N_c.** The settings are referred to Section 5.2 in the main text. The results are presented in Figure 3.
- **Robustness to the number of projection directions.** The settings are referred to Section 5.3 in the main text. The boxplots of ARI are presented in Figure 4.
- **Robustness to kernel functions.** The settings are referred to Section 5.4 in the main text. The boxplots of ARI are presented in Figure 5.
- **Real data analysis.** The details are referred to Section 6 in the main text. The results are presented in Figures 6-7.

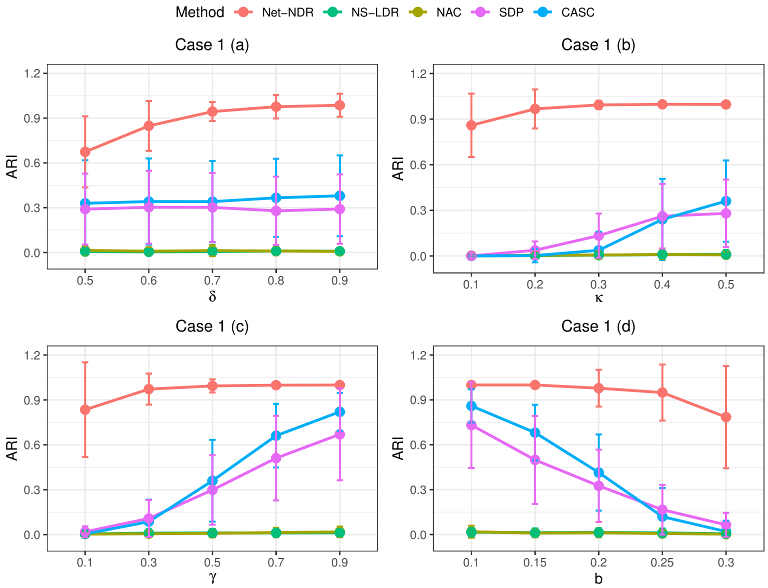

Figure 1: The adjusted Rand index (ARI) of different methods for Case 1. Each dot represents the mean and the vertical bar represents the mean ± standard deviation.

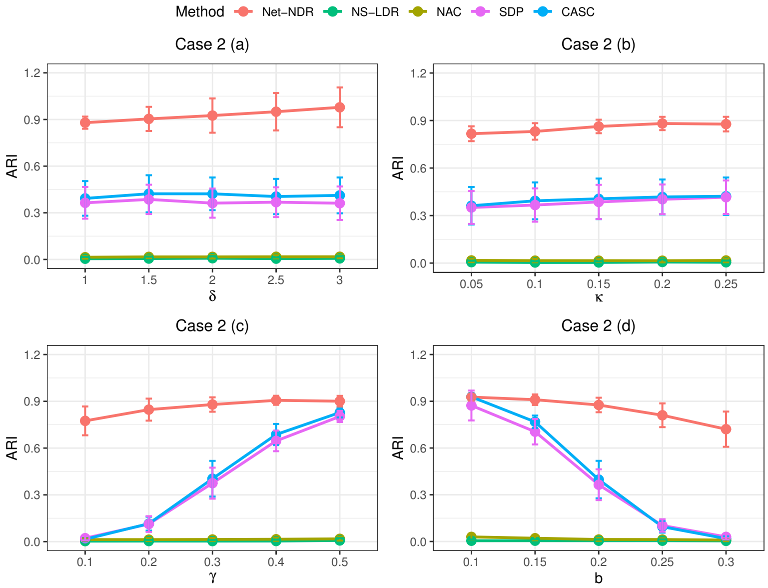

Figure 2: The adjusted Rand index (ARI) of different methods for Case 2. Each dot represents the mean and the vertical bar represents the mean ± standard deviation.

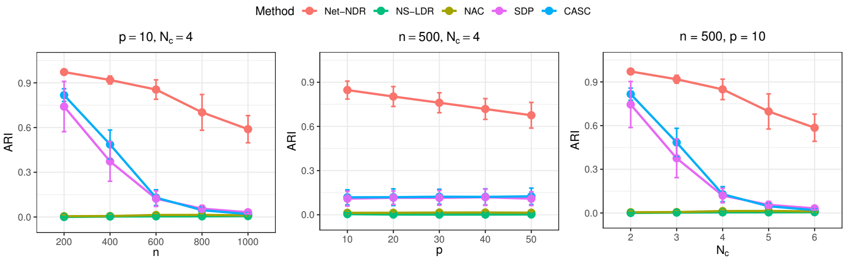

Figure 3: The adjusted Rand index (ARI) for Case 2 with varying sample size n (left), dimension p (middle), and the number of communities N_c (right). Each dot represents the mean and the vertical bar represents the mean ± standard deviation.

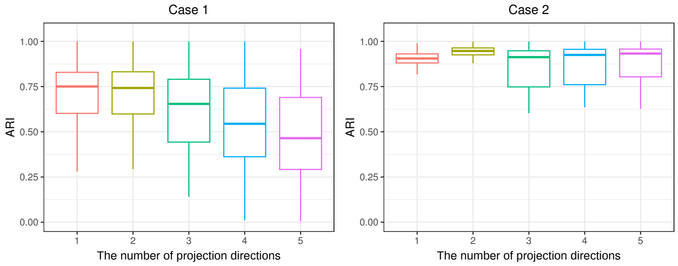

Figure 4: The boxplots of the adjusted Rand index (ARI) of Net-NDR with different numbers of projection directions in Case 1 and Case 2.

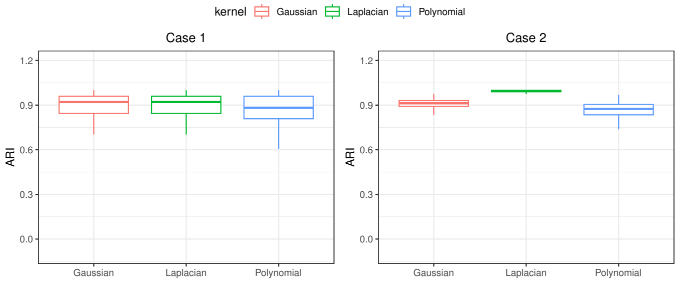

Figure 5: The boxplots of adjusted Rand index (ARI) of Net-NDR with different kernel functions in Case 1 and Case 2.

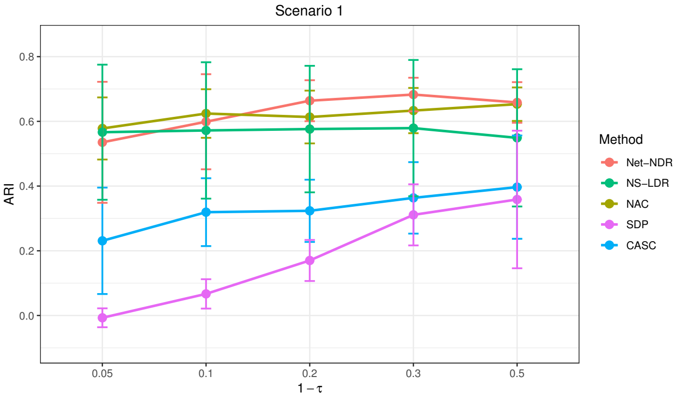

Figure 6: The adjusted Rand index (ARI) of different methods on the pulsar dataset in Scenario 1. Each dot represents the mean, and vertical bars represent the mean ± standard deviation.

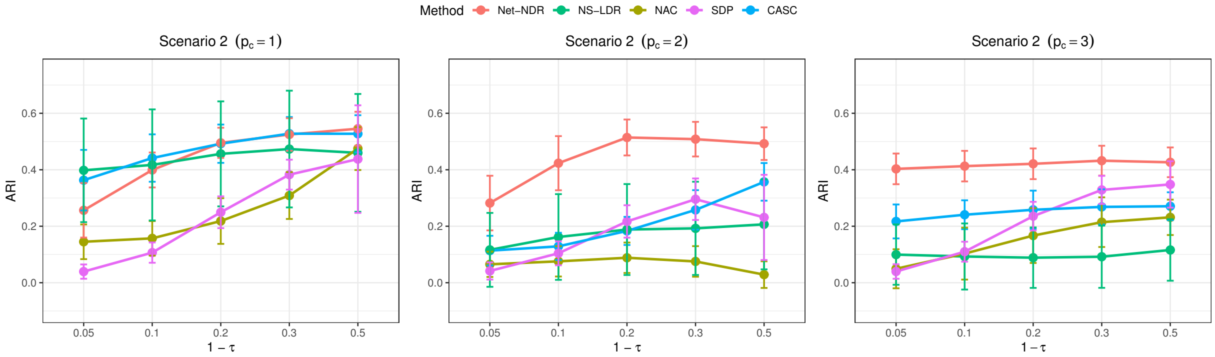

Figure 7: The adjusted Rand index (ARI) of different methods on the pulsar dataset in Scenario 2 for $p_c = 1, 2, 3$ (from left to right). Each dot represents the mean, and vertical bars represent the mean ± standard deviation.

S2.2 Computational cost and memory usage

First, we show the superiority of `eigs_sym` on the computational time and memory usage when performing eigen decomposition on the matrix K_L defined in Section 2.3 of the main text, by comparing it with the classical `eigen` function in R, which calculates all eigenvalues. Taking the default setting in Case 2 of Section 5.1 as an example, we set the range of the sample size as $n \in \{200, 500, 800, 1000, 3000\}$. Specifically, we calculate the first $p/2$ eigenvalues for the function `eigs_sym`, where $p/2$ is chosen according to the rank selection rule of Lam and Yao (2012). Each experiment is repeated 200 times and the average results are reported in Table 1. The results clearly demonstrate the computational advantages of `eigs_sym`.

Table 1: Computational time and memory usage of functions `eigen` and `eigs_sym` versus varying sample size n.

	Time used (s)					Memory usage (kb)				
n	200	500	800	1000	3000	200	500	800	1000	3000
`eigen`()	0.06	0.07	2.58	4.99	105.87	321.70	2004.72	5127.12	8008.72	72024.72
`eigs_sym`()	0.01	0.03	0.08	0.13	1.70	10.50	24.90	39.30	48.90	144.90

Table 2: Computational time (s) of different methods under different sample sizes in Case 2.

	NS-LDR	Net-NDR	NAC	SDP	CASC
$n = 200$	0.05	0.06	0.02	0.11	0.12
$n = 500$	0.68	0.70	0.15	1.25	1.48
$n = 800$	2.66	2.73	0.55	4.55	5.76
$n = 1000$	4.77	5.32	1.07	8.72	11.70
$n = 3000$	153.68	159.98	30.23	246.16	359.28

Second, we compare the average computational time of five methods introduced in Section 5.1, taking Case 2 in Section 5.1 with $(\delta, \kappa, \gamma, b) = (1, 0.25, 0.5, 0.2)$ as an example. We vary n from 200 to 3000, and each experiment is repeated 300 times. The results are shown in Table 2, demonstrating that the computational speed of Net-NDR is competitive.

S2.3 Demonstration of ϵ_n's effect

In this subsection, we illustrate the effect of ϵ_n in a single simulation repetition under the default setting of Case 2 in Section 5.1. Specifically, we calculate the Calinski-Harabasz (CH) index and adjusted Rand index (ARI) values corresponding to different values of ϵ_n. Results are presented in Figure 8. It shows that the ϵ_n that maximizes the CH index corresponds to the value that maximizes the ARI, thus achieving satisfactory community detection outcome.

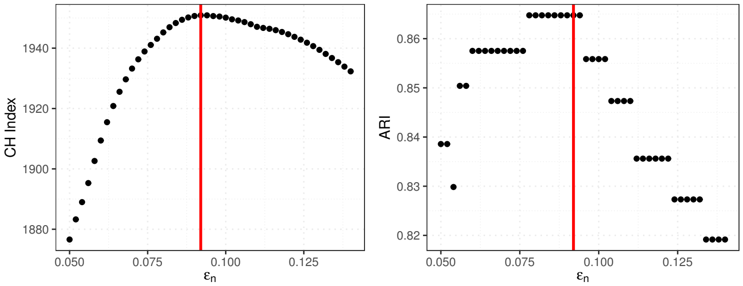

Figure 8: The effect of ϵ_n on the Calinski-Harabasz (CH) index and the adjusted Rand index (ARI) under Case 2. The red vertical line marks the value of ϵ_n that maximizes the CH index.

S2.4 Consistency of rank selection

In this section, we further evaluate the consistency of rank selection; the selection criterion is presented in Remark 4 in the main text. The data are generated as follows.

(Covariates) Consider the following two cases:

Case S1. Set the covariates $X_i = (X_{i,1}, X_{i,2})^\top$ for $i = 1, \cdots, 300$, where $X_{i,j} \overset{i.i.d.}{\sim} U(0,1)$. The feature map is considered as $\phi(X_i) = (\sin(2\pi X_{i,1}), \cos(2\pi X_{i,2}))^\top$;

Case S2. $X_1, \cdots, X_{800} \overset{i.i.d.}{\sim} N_p(\mathbf{0}, \Sigma)$ with $\Sigma = (0.5^{|i-j|})$ and $p = 20$. The feature map is taken as $\phi(X_i) = X_i$.

(Network data) We generate the network from the following model

$$P(\omega_{ij} = 1 | C_i, C_j, \phi_{ij}) = P(\omega_{ij} = 1 | C_i, C_j) \frac{\exp(\gamma - \beta_1 \|F_0^\top \phi_{ij}\|)}{1 + \exp(\gamma - \beta_1 \|F_0^\top \phi_{ij}\|)},$$

where $\phi_{ij} = \phi(X_i) - \phi(X_j)$, γ controls the network sparsity level and $\beta_1 > 0$ controls the influence of the distance between projected covariates on the connection probability. Recall that $\hat{\Sigma}_\phi$ is the sample estimator of the covariance matrix $\Sigma_\phi = \mathrm{cov}(\phi(X_i))$. The coefficient matrix $F_0 \in \mathbb{R}^{p \times r}$ consists of the first r eigenvectors of $\hat{\Sigma}_\phi$ such that $F_0^\top \hat{\Sigma}_\phi F_0 = I_r$. The latent community labels C_i's are generated independently with $P(C_i = 1) = P(C_i = 2) = 1/2$.

Table 3: The confusion matrices of estimated ranks under different values of (γ, β_1) in Case S1.

	$(\gamma,\beta_1)=(1,1)$		$(\gamma,\beta_1)=(1,2)$		$(\gamma,\beta_1)=(3,1)$		$(\gamma,\beta_1)=(3,2)$	
	$r=1$	$r=2$	$r=1$	$r=2$	$r=1$	$r=2$	$r=1$	$r=2$
$\hat{r}=1$	192	2	187	0	192	0	194	1
$\hat{r}=2$	2	192	6	196	3	194	3	194
$\hat{r}=3$	3	2	6	0	4	3	2	3
$\hat{r}=4$	1	4	1	4	1	3	1	2

For Case S1, we set the true rank $r \in \{1, 2\}$. The parameter M appeared in our rule of rank selection in Remark 4 in the main text is set to 5. A Gaussian kernel function is adopted. We compute the confusion matrices with four parameter configurations of (γ, β_1) across 200 repetitions and report the results in Table 3.

Table 4: The accuracy of estimated ranks for $r \in \{1, 5, 10, 15\}$ under different values of (γ, β_1) in Case S2.

	$(\gamma, \beta_1) = (10, 4)$				$(\gamma, \beta_1) = (12, 5)$			
	$r = 1$	$r = 5$	$r = 10$	$r = 15$	$r = 1$	$r = 5$	$r = 10$	$r = 15$
Accuracy	1	1	1	0.98	1	1	1	0.99

Table 5: The confusion matrices of estimated ranks for $r \in \{11, 12, 13, 14, 15\}$ under different values of (γ, β_1) in Case S2.

	$(\gamma, \beta_1) = (10, 4)$					$(\gamma, \beta_1) = (12, 5)$				
	$r = 11$	$r = 12$	$r = 13$	$r = 14$	$r = 15$	$r = 11$	$r = 12$	$r = 13$	$r = 14$	$r = 15$
$\hat{r} = 11$	200	0	0	0	0	200	0	0	0	0
$\hat{r} = 12$	0	200	0	0	0	0	200	0	0	0
$\hat{r} = 13$	0	0	200	2	0	0	0	195	0	0
$\hat{r} = 14$	0	0	0	191	4	0	0	2	196	3
$\hat{r} = 15$	0	0	0	7	196	0	0	3	4	197

For Case S2, we expand the range of rank to $\{1, 5, 10, 11, \cdots, 14, 15\}$. The linear kernel function is used and $M = 15$. To intuitively present the results from 200 independent repetitions with two parameter configurations of (γ, β_1), we first calculate the rank estimation accuracy when $r \in \{1, 5, 10, 15\}$, and then compute the confusion matrices for $r \in \{11, \cdots, 15\}$. These results are reported in Tables

4 and 5, respectively. It shows that our method can select the true rank with high probability.

S3 Additional discussions and theoretical results

S3.1 Comparisons with Yan and Sarkar (2021)

Our work differs from Yan and Sarkar (2021) in the following aspects.

- **Different assumptions on covariates:** Recall that N_c denotes the number of communities. Let $\{\mu_k, k = 1, \cdots, N_c\}$ be the community-specific means of the covariates. Yan and Sarkar (2021) assumed the following mixture model for p-dimensional nodal covariates X_i's:

  $$X_i = \mu_{C_i} + \xi_i,\ i = 1, \cdots, n, \tag{S3.2}$$

  where ξ_i's are *i.i.d.* zero-mean sub-Gaussian random errors with the covariance matrix $\sigma^2_{C_i} I_p$. In contrast, it is assumed in our work that

  $$\phi(X_i) = \mu^{\phi}_{C_i} + \varsigma_i,\ i = 1, \cdots, n \tag{S3.3}$$

  with a given feature map $\phi(\cdot)$, where $\{\mu^{\phi}_k, k = 1, \cdots, N_c\}$ are the community-specific mean functions and ς_i's are independent zero-mean random elements in a reproducing kernel Hilbert space (RKHS).

  From (S3.2), we can see that the effectiveness of Yan and Sarkar (2021) depends heavily on the separability of the means μ_k's from different communities. So when μ_k's are close to each other, their method will not provide satisfactory results (e.g. Case 1 in Subsection 5.1 of the main text where μ_1 and μ_2 are both zero vectors). In contrast, our method requires that the community-specific means after transformation (i.e. μ_k^ϕ's) are separable. This is analogous to using a kernel in support vector machine in the setting where the original data are not linearly separable but can become separable after a feature mapping. Recall the concentric ring data introduced in Section 1 of the main text – a classic example of linear inseparability. Here we present the scatter plot of projected data from the method SDP in Yan and Sarkar (2021). Figure 9 shows that SDP cannot clearly distinguish two communities. In contrast, according to Figure 1 in the main text, our nonlinear dimension reduction method Net-NDR can linearly separate two groups nicely.

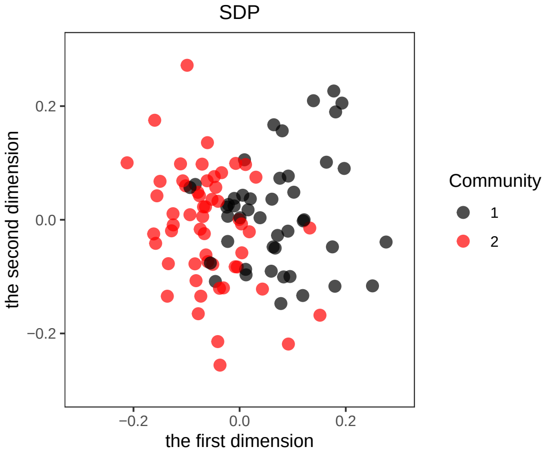

Figure 9: The Scatter plot of projected data via SDP in Yan and Sarkar (2021).

- **Difference in other aspects:** In terms of research objectives, we develop a dimension reduction framework for nodal covariates, which can be applied to multiple downstream tasks. In contrast, Yan and Sarkar (2021) solely focused on the community detection task. In terms of assumptions, Yan and Sarkar (2021) assumed that the network follows the stochastic block-model, while we do not require such assumption. Last, our theoretical framework accounts for dependence structure among w_{ij}'s ($i < j$), while Yan and Sarkar (2021) required w_{ij}'s ($i < j$) to be independent.

S3.2 Discussions on applications of our nonlinear dimension reduction method

Our dimension reduction method Net-NDR can be applied to many models that incorporate nodal covariates, including Ma et al. (2020), Huang et al. (2024), Stein and Leng (2025) and many others. For example, the model proposed in Stein and Leng (2025) assumes

$$P(w_{ij} = 1 | Z_{ij}) = \frac{\exp(\beta_i + \beta_j + \mu + Z_{ij}^\top \gamma)}{1 + \exp(\beta_i + \beta_j + \mu + Z_{ij}^\top \gamma)}, \tag{S3.4}$$

where β_i's are heterogeneity parameters, μ is the global density parameter, and $Z_{ij} = g(X_i, X_j)$ with a symmetric function g. By setting $Z_{ij} = \|\boldsymbol{f}_r(\phi(X)) - \boldsymbol{f}_r(\phi(X'))\|^2$, the covariates after nonlinear dimension reduction are integrated into this model. Due to the model-free property, our method can identify the infinite-dimensional subspace spanned by $\boldsymbol{f}_r$ and obtain the low-dimensional nonlinear embedding $\boldsymbol{f}_r(\phi(X_i))$'s of the original covariates; please refer to Proposition A.

Besides community detection discussed in the main text, our dimension reduction method can be combined with multiple downstream tasks. First, the low-dimensional nonlinear embedding of the original covariates can be integrated as covariates into existing models to support tasks such as link prediction. For example, the link prediction model in Graham (2017) incorporates $Z_{ij} = g(X_i, X_j)$ with a symmetric function g. The model's prediction performance depends on how Z_{ij} is constructed. Setting $Z_{ij} = \|\boldsymbol{f}_r(\phi(X_i)) - \boldsymbol{f}_r(\phi(X_j))\|$ improves upon the Euclidean distance $\|X_i - X_j\|$ used in Graham (2017) by better capturing nonlinear relationships among nodal covariates. Second, the embedding is applicable to a wide range of unsupervised tasks. For example, the embeddings $\boldsymbol{f}_r(\phi(X_i))$'s can be utilized for anomaly detection by applying relevant algorithms to them (Liu et al., 2022).

S3.3 A discussion on subspace recovery

The following proposition interprets $\boldsymbol{f}_r$ defined in (2.1) of the main text from the perspective of subspace recovery.

Proposition A. *Suppose the following conditions hold:*

(1) $E(s|X, X') = h(\boldsymbol{f}_{r,0}(\phi(X)) - \boldsymbol{f}_{r,0}(\phi(X')))$ *where* $\boldsymbol{f}_{r,0} = (f_{0,1}, \cdots, f_{0,r}) \in \Theta_r$ *and* h *is unspecified;*

(2) the r *largest eigenvalues of* $\Sigma_\phi^{-1} G_0$ *are distinct;*

(3) for any $v \in \Theta_1$ *satisfying* $\langle v, \Sigma_\phi f_{0,m}\rangle_{\mathcal{H}} = 0$ *for* $m = 1, \cdots, r$, *we have*

$$\operatorname{cov}\left\{s, \langle f_{0,m}, \phi(X) - \phi(X')\rangle_{\mathcal{H}}^2\right\} > \operatorname{cov}\left\{s, \langle v, \phi(X) - \phi(X')\rangle_{\mathcal{H}}^2\right\}.$$

Then it follows that $\mathrm{span}(\boldsymbol{f}_r) = \mathrm{span}(\boldsymbol{f}_{r,0})$.

The condition (1) means that the expectation of s depends on X and X' only through the projections $\boldsymbol{f}_{r,0}(\phi(X))$ and $\boldsymbol{f}_{r,0}(\phi(X'))$. When the conditional distribution of s given (X, X') depends solely on $E\left(s|X, X'\right)$, for example when w is a Bernoulli random variable and $s = 1 - w$, the condition (1) implies that s is independent of X and X' given $\boldsymbol{f}_{r,0}(\phi(X)) - \boldsymbol{f}_{r,0}(\phi(X'))$. The condition (3) requires that the covariance between the distance after projection and s peaks in the true directions. This is a general requirement on s, imposing no strict restrictions on its functional form. For example, if $\mathrm{cov}\left\{w, \langle f_{0,m}, \phi(X) - \phi(X')\rangle_{\mathcal{H}}^2\right\} < \mathrm{cov}\left\{w, \langle v, \phi(X) - \phi(X')\rangle_{\mathcal{H}}^2\right\}$ and $s = \alpha_0 - \alpha_1 w$ $(\alpha_0 \in \mathbb{R}, \alpha_1 > 0)$, it is easy to see condition (3) holds for all values of (α_0, α_1).

Since Proposition A does not restrict the specific form of the function h, Net-NDR recovers the true directions in a model-free style. While Net-NDR shares similarities with sufficient dimension reduction (SDR) techniques (Li, 1991; Cook and Li, 2002; Ying and Yu, 2022; Zhang et al., 2024), it differs from them in two major respects. First, s_{ij}'s $(i < j)$ are dependent due to the network structure, while the responses in SDR are generally assumed to be independent. Second, the variable s corresponds to a pair of covariates (X, X'), while the response variable in SDR corresponds to covariates of only one individual.

The proof of Proposition A is given as follows.

Proof. Denote $\phi(X) - \phi(X')$ as $\phi_{XX'}$ for simplicity. Write $M_0 = E(s\phi_{XX'} \otimes$

$\phi_{XX'})$. Consider the operator $M_1 = M_0 - E(s)E(\phi_{XX'} \otimes \phi_{XX'})$. For any $u \in \mathcal{H}$ s.t. $\langle u, \Sigma_\phi u\rangle_{\mathcal{H}} = 1$, we have $E(s)E(\langle u, \phi_{XX'}\rangle^2_{\mathcal{H}}) = 2E(s)$. For any $v \in \Theta_1$, we have $\langle v, M_1 v\rangle_{\mathcal{H}} = E(s\langle v, \phi_{XX'}\rangle^2_{\mathcal{H}}) - 2E(s) = \mathrm{cov}(s, \langle v, \phi_{XX'}\rangle^2_{\mathcal{H}})$. Similarly, since $\boldsymbol{f}_{r,0} \in \Theta_r$, it holds that, for $m = 1, \cdots, r$, $\langle f_{0,m}, M_1 f_{0,m}\rangle_{\mathcal{H}} = E(s\langle f_{0,m}, \phi_{XX'}\rangle^2_{\mathcal{H}}) - 2E(s) = \mathrm{cov}(s, \langle f_{0,m}, \phi_{XX'}\rangle^2_{\mathcal{H}})$. By the assumption that $\mathrm{cov}\left\{s, \langle f_{0,m}, \phi_{XX'}\rangle^2_{\mathcal{H}}\right\} > \mathrm{cov}\left\{s, \langle v, \phi_{XX'}\rangle^2_{\mathcal{H}}\right\}$, we have $E(s\langle f_{0,m}, \phi_{XX'}\rangle^2_{\mathcal{H}}) > E(s\langle v, \phi_{XX'}\rangle^2_{\mathcal{H}})$. According to the distinctiveness of the r largest eigenvalues of $\Sigma_\phi^{-1}G_0$ and the definition of $\boldsymbol{f}_r$, it is easy to see that $\mathrm{span}(\boldsymbol{f}_r) = \mathrm{span}(\boldsymbol{f}_{r,0})$.

$\square$

S3.4 Discussions on s

The invariance of subspace to s. According to Proposition A, under a general model $E\left(s|X, X'\right) = h\left(\boldsymbol{f}_{r,0}(\phi(X)) - \boldsymbol{f}_{r,0}(\phi(X'))\right)$ with an unspecified function h, our optimization problem can be seen as finding the dimension reduction subspace $\mathrm{span}(\boldsymbol{f}_{r,0})$ which is invariant to the choice of s under some conditions, sharing similar spirits with sufficient dimension reduction.

Implications of the linear function s. Recall the definition of L_S and ν_1 in Subsection 2.3 in the main text. We begin by discussing the eigenvectors of L_S under an unweighted network. We take $r = 1$ as an example for illustration and the discussions can be easily extended to the case where $r > 1$. Denote $n^{1/2}\nu_1$ by $\nu_1^\dagger$. By Proposition 2 in the main text, it holds that $\nu_1^\dagger = \arg\max_{z:\|z\|^2=n} z^\top L_S z$.

With the notation $z = (z_1, \cdots, z_n)^\top$, we have

$$z^\top L_S z = [2n(n-1)]^{-1} \sum_{i \neq j} s_{ij}(z_i - z_j)^2. \tag{S3.5}$$

To probe deeper into (S3.5), we consider a linear function $s_{ij} = \alpha_0 - \alpha_1 w_{ij}$ with $0 \leq \alpha_0 < \alpha_1$, which is also used by Zhao et al. (2022). Then we have

$$z^\top L_S z = [2n(n-1)]^{-1} \left[\sum_{i,j:w_{ij}=0} \alpha_0 (z_i - z_j)^2 + \sum_{i,j:w_{ij}=1} (\alpha_0 - \alpha_1)(z_i - z_j)^2 \right]. \tag{S3.6}$$

From (S3.6), it can be seen that $\nu_1^\dagger$ maximizes the sum of the distances between projections of unconnected node pairs (i.e. $w_{ij} = 0$) and minimizes the sum of the distances between projections of connected pairs (i.e. $w_{ij} = 1$) simultaneously, with the weights being α_0 and $\alpha_0 - \alpha_1$ respectively. When $\alpha_0 = 0$, $\nu_1^\dagger$ only minimizes the sum of the distances between projections of connected pairs, disregarding information from unconnected ones. To balance the information from both connected and unconnected pairs, we recommend setting $\alpha_0 = 1$ and $\alpha_1 = 2$.

S3.5 A general version of Proposition 3

Recall that the optimization problem (2.1) in the main text requires the directions $f_1, \cdots, f_r$ to satisfy the constraint (i): $\langle f_k, \Sigma_\phi f_l \rangle_\mathcal{H} = \mathbb{I}(k = l)$. In this section, apart from constraint (i), we also consider the constraint (ii): $\langle f_k, f_l \rangle_\mathcal{H} = \mathbb{I}(k = l)$. Under constraints (i) and (ii), the relationships between our method Net-NDR and kernel discriminant analysis (KDA) as well as kernel principal component analysis

(KPCA) are established respectively under the setting outlined in Section 4.2 in the main text.

Proposition B. *Assume that* $W = (w_{ij})_{n\times n}$ *is generated from the degree-corrected stochastic block-model and that* $\phi(X_i)$*'s follow the model (2.8) outlined in the main text. Denote* $\bar{\eta} = \lim_n \{n(n-1)\}^{-1}\sum_{i\neq j}\sum_{t=1}^{N_c}\pi_t^2\eta_{t,ij}$, $\bar{\nu} = \lim_n\{n(n-1)\}^{-1}\sum_{i\neq j}\nu_{ij}$ *and* $\bar{M} = \bar{\eta}+\bar{\nu}\sum_{t_1\neq t_2}\pi_{t_1}\pi_{t_2}$*. Consider the generalized eigen-decomposition problem*

$$f_{A,k} = \arg\max_{\substack{f:f\perp\mathcal{L}_{A,k-1}\\ \langle f,Af\rangle_{\mathcal{H}}=1}} \langle f, G_0 f\rangle_{\mathcal{H}}$$

where $\mathcal{L}_{A,k-1} = \mathrm{span}(Af_{A,1},\cdots,Af_{A,k-1})$ *for* $k = 2,\cdots,r$ *and* $\mathcal{L}_{A,0} = \{\mathbf{0}\}$ *and* A *is an operator that will be specified as* Σ_ϕ *or* $I_{\mathcal{H}}$ *later. Assume that (i)* $\bar{\eta}$ *and* $\bar{\nu}$ *both exist and are bounded; (ii) the* r *largest eigenvalues of* $A^{-1/2}G_0A^{-1/2}$ *are distinct. Then the following conclusions hold.*

(1) When $A = I_{\mathcal{H}}$*, that is,* $\langle f_k, f_l\rangle_{\mathcal{H}} = \mathbb{I}(k=l)$*, Net-NDR is equivalent to KPCA at the population level in the sense that for any given* $r \le n$*,* $f_{I_{\mathcal{H}},k}$ *is exactly the eigenfunction associated with the* k*-th largest eigenvalue of* Σ_ϕ *for* $k = 1,\ldots,r$ *if and only if* $\bar{M} \neq 0$ *and* $\bar{\nu} = \bar{M}$*.*

(2) When $A = \Sigma_\phi$*, that is,* $\langle f_k, \Sigma_\phi f_l\rangle_{\mathcal{H}} = \mathbb{I}(k=l)$*, if* $\bar{\nu} > \bar{M}$*, Net-NDR is equivalent to KDA at the population level in the sense that for any given* $r \le n$*,* $f_{\Sigma_\phi,k}$ *is proportional to the* k*-th direction of KDA for* $k = 1,\ldots,r$*.*

Proposition B tells us that when the projection directions are orthonormal, Net-NDR only utilizes the covariance information of mapped covariates. When $A = \Sigma_\phi$, the projection directions are estimated as if the unknown community membership is exploited.

The proof of Proposition B is given as follows.

Proof. In this proof, we solve the generalized eigendecomposition problem

$$f_{A,k} = \arg \max_{\substack{\langle f, Af \rangle_{\mathcal{H}}=1 \\ f \perp \mathcal{L}_{A,k-1}}} \langle f, G_0 f \rangle_{\mathcal{H}}$$

with $\mathcal{L}_{A,k-1} = \operatorname{span}(Af_{A,1}, \cdots, Af_{A,k-1})$ $(k = 2, \cdots, r)$ and $\mathcal{L}_{A,0} = \{0\}$ in a slightly different way. We first find the eigenfunctions φ_i^A's corresponding to the r largest eigenvalues of $A^{-1/2} G_0 A^{-1/2}$ and then set $f_{A,i} = A^{-1/2} \varphi_i^A$ for $i = 1, \cdots, r$. The following proof is based on this process.

Write $\phi_{ij} = \phi(X_i) - \phi(X_j)$, $\varsigma_{ij} = \varsigma_i - \varsigma_j$ and $\mu_{t_1 t_2} = \mu_{t_1}^{\phi} - \mu_{t_2}^{\phi}$. Then

$$E(\phi_{ij} \otimes \phi_{ij} \mid C_i = t_1, C_j = t_2) = \begin{cases} \mu_{t_1 t_2} \otimes \mu_{t_1 t_2} + 2\Sigma_{\varsigma}, & t_1 \neq t_2 \\ 2\Sigma_{\varsigma}, & t_1 = t_2 \end{cases}.$$

Let $\xi = E\left(\phi(X_i) \mid C_i\right)$ be a discrete random element taking values in $\left\{\mu_1^{\phi}, \cdots, \mu_{N_c}^{\phi}\right\}$ with $P\left(\xi = \mu_i\right) = \pi_i$. Suppose that ξ_1 and ξ_2 are independent copies of ξ. Then it is easy to see that

$$\operatorname{cov}\left(\xi_1 - \xi_2\right) = 2 \operatorname{cov}(\xi) = \sum_{t_1 \neq t_2} \pi_{t_1} \pi_{t_2} \mu_{t_1 t_2} \otimes \mu_{t_1 t_2}.$$

Under the setting in Section 4.2, s_{ij} is independent of ϕ_{ij} given (C_i, C_j), then it follows that for any $i \neq j$

$$
\begin{aligned}
&E\left(s_{ij}\phi_{ij} \otimes \phi_{ij}\right) \\
&= E\left[E\left(s_{ij}\phi_{ij} \otimes \phi_{ij} \mid C_i, C_j\right)\right] \\
&= \sum_{t_1,t_2=1}^{N_c} P\left(C_i = t_1, C_j = t_2\right) E\left(s_{ij} \mid C_i = t_1, C_j = t_2\right) E\left(\phi_{ij} \otimes \phi_{ij} \mid C_i = t_1, C_j = t_2\right) \\
&= \sum_{t=1}^{N_c} 2\pi_t^2 \eta_{t,ij} \Sigma_\varsigma + \sum_{t_1 \neq t_2} \pi_{t_1}\pi_{t_2}\nu_{ij}\left(\mu_{t_1t_2} \otimes \mu_{t_1t_2} + 2\Sigma_\varsigma\right) \\
&= 2\left(\sum_{t=1}^{N_c} \pi_t^2 \eta_{t,ij} + \nu_{ij} \sum_{t_1 \neq t_2} \pi_{t_1}\pi_{t_2}\right)\Sigma_\varsigma + \nu_{ij} \sum_{t_1 \neq t_2} \pi_{t_1}\pi_{t_2}\mu_{t_1t_2} \otimes \mu_{t_1t_2} \\
&= 2M_{ij}\Sigma_\varsigma + 2\nu_{ij}\operatorname{cov}(\xi),
\end{aligned}
$$

where we write $M_{ij} = \sum_{t=1}^{N_c} \pi_t^2 \eta_{t,ij} + \nu_{ij} \sum_{t_1 \neq t_2} \pi_{t_1}\pi_{t_2}$. Since

$$
\operatorname{cov}\left(\phi(X_i) \mid C_i = t\right) \equiv \Sigma_\varsigma
$$

for $t = 1, \cdots, N_c$, we have $E\left\{\operatorname{cov}\left(\phi(X_i) \mid C_i\right)\right\} = \Sigma_\varsigma$. Thus, by the formula $\operatorname{cov}\left(\phi(X_i)\right) = \operatorname{cov}\left[E\left(\phi(X_i) \mid C_i\right)\right] + E\left[\operatorname{cov}\left(\phi(X_i) \mid C_i\right)\right]$, it holds that $\operatorname{cov}\left(\phi(X_i)\right) = \Sigma_\varsigma + \operatorname{cov}(\xi)$. Then we have

$$
\begin{aligned}
&E\left(s_{ij}\phi_{ij} \otimes \phi_{ij}\right) = 2M_{ij}\operatorname{cov}\left(\phi(X_i)\right) + 2\left(\nu_{ij} - M_{ij}\right)\operatorname{cov}(\xi), \quad \text{and} \\
&G_0 = \lim_n E(\hat{G}) = \lim_n \frac{1}{n(n-1)} \sum_{i \neq j} E\left(s_{ij}\phi_{ij} \otimes \phi_{ij}\right) \\
&\ = 2\bar{M}\operatorname{cov}\left(\phi(X_i)\right) + 2(\bar{\nu} - \bar{M})\operatorname{cov}(\xi). \qquad \text{(S3.7)}
\end{aligned}
$$

(1) If $A = I_{\mathcal{H}}$, from (S3.7) we see that Net-NDR is equivalent to KPCA if and only if $\bar{M} \neq 0$ and $\bar{\nu} - \bar{M} = 0$.

(2) Recall the assumption that the r largest eigenvalues of $G_{0A} = A^{-1/2}G_0A^{-1/2}$ are distinct. Denote the spectral decomposition of G_{0A} by

$$G_{0A} = \sum_{i=1}^{\infty} q_i^A \varphi_i^A \otimes \varphi_i^A,$$

where $q_1^A \geq q_2^A \geq \cdots$ are the eigenvalues, and φ_i^A's are the associated eigenfunctions. We have $f_{A,i} = A^{-1/2}\varphi_i^A$ $(i \leq r)$. Moreover, combining (S3.7) with the fact that $A = \operatorname{cov}(\phi(X))$, we have,

$$G_{0A} = A^{-1/2}G_0A^{-1/2} = 2\bar{M}I_{\mathcal{H}} + 2\left(\bar{\nu} - \bar{M}\right) A^{-1/2}\operatorname{cov}(\xi)A^{-1/2}.$$

Let $G_{\xi,A} = A^{-1/2}\operatorname{cov}(\xi)A^{-1/2}$. Since the r largest eigenvalues of G_{0A} are distinct, the r largest eigenvalues of $G_{\xi,A}$ are also distinct. Denote the i-th largest eigenvalue of $G_{\xi,A}$ by $\lambda_{\xi,i}$. By the condition $\bar{\nu} > \bar{M}$, φ_i^A is the eigenfunction of $G_{\xi,A}$ associated with $\lambda_{\xi,i}$. Then by Courant-Fischer min-max theorem on an infinite dimensional Hilbert space (Courant, 1920), we have

$$\lambda_{\xi,i} = \sup_{V\subset\mathcal{H},\dim(V)=i}\ \min_{v\in V:\|v\|_{\mathcal{H}}=1} \langle v, G_{\xi,A}v\rangle_{\mathcal{H}} = \langle \varphi_i^A, G_{\xi,A}\varphi_i^A\rangle_{\mathcal{H}},$$

where $\dim(V)$ denotes the dimension of V. Letting V_i be the optimal V with $\dim(V) = i$, then φ_i^A is the unique optimal v in V_i, that is,

$$\begin{aligned}\left(V_i, \varphi_i^A\right) &= \arg \sup_{V\subset\mathcal{H},\dim(V)=i}\ \min_{v\in V:\|v\|_{\mathcal{H}}=1} \langle v, G_{\xi,A}v\rangle_{\mathcal{H}} \\ &= \arg \sup_{V\subset\mathcal{H},\dim(V)=i}\ \min_{v\in V:\|v\|_{\mathcal{H}}=1} \langle v, A^{-1/2}\operatorname{cov}(\xi)A^{-1/2}v\rangle_{\mathcal{H}} \\ &= \arg \sup_{V\subset\mathcal{H},\dim(V)=i}\ \min_{v\in V:v\neq 0} \frac{\langle v, A^{-1/2}\operatorname{cov}(\xi)A^{-1/2}v\rangle_{\mathcal{H}}}{\|v\|_{\mathcal{H}}},\end{aligned}$$

where for the optimization problem in the last step, the optimal v equals $c\varphi_i^A$ for any constant $c \neq 0$ and we take $c = 1$ here. Letting $\tilde{v} = A^{-1/2}v$ and recalling the fact that $A = \text{cov}(\phi(X)) = \text{cov}(\xi) + \Sigma_\varsigma$, it follows that

$$
\begin{aligned}
\left(A^{-1/2}V_i, A^{-1/2}\varphi_i^A\right) &= \arg \sup_{V \subset \mathcal{H}, \dim(V)=i} \min_{\tilde{v} \in V: \tilde{v} \neq 0} \frac{\langle \tilde{v}, \text{cov}(\xi)\tilde{v}\rangle_{\mathcal{H}}}{\langle \tilde{v}, A\tilde{v}\rangle_{\mathcal{H}}} \\
&= \arg \sup_{V \subset \mathcal{H}, \dim(V)=i} \min_{\tilde{v} \in V: \tilde{v} \neq 0} \frac{\langle \tilde{v}, \text{cov}(\xi)\tilde{v}\rangle_{\mathcal{H}}}{\langle \tilde{v}, \Sigma_\varsigma\tilde{v}\rangle_{\mathcal{H}} + \langle \tilde{v}, \text{cov}(\xi)\tilde{v}\rangle_{\mathcal{H}}} \\
&= \arg \sup_{V \subset \mathcal{H}, \dim(V)=i} \min_{\tilde{v} \in V: \tilde{v} \neq 0} \frac{\langle \tilde{v}, \text{cov}(\xi)\tilde{v}\rangle_{\mathcal{H}}/\langle \tilde{v}, \Sigma_\varsigma\tilde{v}\rangle_{\mathcal{H}}}{1 + \langle \tilde{v}, \text{cov}(\xi)\tilde{v}\rangle_{\mathcal{H}}/\langle \tilde{v}, \Sigma_\varsigma\tilde{v}\rangle_{\mathcal{H}}} \\
&= \arg \sup_{V \subset \mathcal{H}, \dim(V)=i} \min_{\tilde{v} \in V: \tilde{v} \neq 0} \langle \tilde{v}, \text{cov}(\xi)\tilde{v}\rangle_{\mathcal{H}}/\langle \tilde{v}, \Sigma_\varsigma\tilde{v}\rangle_{\mathcal{H}}.
\end{aligned}
$$

Let $(\lambda_{\text{kda },i}, v_{\text{kda },i})$ be the pair of the i-th eigenvalue and direction in kernel linear discriminant analysis and let $G_{\xi,\Sigma_\varsigma} = \Sigma_\varsigma^{-1/2}\text{cov}(\xi)\Sigma_\varsigma^{-1/2}$. Then $\lambda_{\text{kda },i}$ is the i-th eigenvalue of G_{ξ,Σ_ς}, and $v_{\text{kda },i} = \Sigma_\varsigma^{-1/2}\omega_i$, where ω_i is the i-th eigenfunction of G_{ξ,Σ_ς}. By Courant-Fischer min-max theorem again, it follows that

$$
\begin{aligned}
\lambda_{\text{kda },i} &= \sup_{V \subset \mathcal{H}, \dim(V)=i} \min_{\tilde{v} \in V: \tilde{v} \neq 0} \langle \tilde{v}, \text{cov}(\xi)\tilde{v}\rangle_{\mathcal{H}}/\langle \tilde{v}, \Sigma_\varsigma\tilde{v}\rangle_{\mathcal{H}} \qquad \text{(S3.8)} \\
&= \langle v_{\text{kda },i}, \text{cov}(\xi)v_{\text{kda },i}\rangle_{\mathcal{H}}/\langle v_{\text{kda },i}, \Sigma_\varsigma v_{\text{kda },i}\rangle_{\mathcal{H}}. \qquad \text{(S3.9)}
\end{aligned}
$$

On the other hand, due to the fact that the r largest eigenvalues of G_{0A} are distinct, it is easy to see that $\lambda_{\text{kda}, i}$'s are all distinct, and there is a unique solution for the optimization problem (S3.9). Therefore, $v_{\text{kda },i} = cA^{-1/2}\varphi_i^A = f_{A,i}$ $(i \leq r)$ with $c = \pm 1$.

□

Bibliography

Cook, R. D. and B. Li (2002). Dimension reduction for conditional mean in regression. *The Annals of Statistics 30*(2), 455–474.

Courant, R. (1920). Über die eigenwerte bei den differentialgleichungen der mathematischen physik. *Mathematische Zeitschrift 7*(1), 1–57.

Douglas, R. G. (1966). On majorization, factorization, and range inclusion of operators on hilbert space. *Proceedings of the American Mathematical Society 17*(2), 413–415.

Fukumizu, K., F. R. Bach, and A. Gretton (2007). Statistical consistency of kernel canonical correlation analysis. *Journal of Machine Learning Research 8*(2).

Fukumizu, K., F. R. Bach, and M. I. Jordan (2009). Kernel dimension reduction in regression. *The Annals of Statistics 37*(4), 1871–1905.

Graham, B. S. (2017). An econometric model of network formation with degree heterogeneity. *Econometrica 85*(4), 1033–1063.

Hu, Y. and W. Wang (2024). Network-adjusted covariates for community detection. *Biometrika 111*(4), 1221–1240.

Huang, S., J. Sun, and Y. Feng (2024). PCABM: Pairwise covariates-adjusted block model for community detection. *Journal of the American Statistical Association 119*(547), 2092–2104.

Lam, C. and Q. Yao (2012). Factor modeling for high-dimensional time series: inference for the number of factors. *The Annals of Statistics 40*(2), 694–726.

Li, K.-C. (1991). Sliced inverse regression for dimension reduction. *Journal of the American Statistical Association 86*(414), 316–327.

Liu, K., Y. Dou, Y. Zhao, X. Ding, X. Hu, R. Zhang, K. Ding, C. Chen, H. Peng, K. Shu, et al. (2022). Bond: benchmarking unsupervised outlier node detection on static attributed graphs. *Advances in Neural Information Processing Systems 35*, 27021–27035.

Ma, Z., Z. Ma, and H. Yuan (2020). Universal latent space model fitting for large networks with edge covariates. *Journal of Machine Learning Research 21*(4), 1–67.

Serfling, R. J. (1980). Approximation theorems of mathematical statistics. John Wiley & Sons.

Stein, S. and C. Leng (2025). A sparse beta regression model for network analysis. *Journal of the American Statistical Association 120*(550), 1281–1293.

Virta, J., K.-Y. Lee, and L. Li (2022). Sliced inverse regression in metric spaces. *Statistica Sinica 32*, 2315–2337.

Yan, B. and P. Sarkar (2021). Covariate regularized community detection in sparse graphs. *Journal of the American Statistical Association 116*(534), 734–745.

Ying, C. and Z. Yu (2022). Fréchet sufficient dimension reduction for random objects. *Biometrika 109*(4), 975–992.

Zhang, Q., B. Li, and L. Xue (2024). Nonlinear sufficient dimension reduction for distribution-on-distribution regression. *Journal of Multivariate Analysis 202*, 105302.

Zhao, J., X. Liu, H. Wang, and C. Leng (2022). Dimension reduction for covariates in network data. *Biometrika 109*(1), 85–102.