ESTIMATING STOCHASTIC BLOCK MODELS IN THE PRESENCE OF COVARIATES

YUICHI KITAMURA AND LOUISE LAAGE

Abstract. In the standard stochastic block model for networks, the probability of a connection between two nodes, often referred to as the edge probability, depends on the unobserved communities each of these nodes belongs to. We consider a flexible framework in which each edge probability, together with the probability of community assignment, are also impacted by observed covariates. We propose a computationally tractable two-step procedure to estimate the conditional edge probabilities as well as the community assignment probabilities. The first step relies on a spectral clustering algorithm applied to a localized adjacency matrix of the network. In the second step, k-nearest neighbor regression estimates are computed on the extracted communities. We study the statistical properties of these estimators by providing non-asymptotic bounds.

1. Introduction

The Stochastic Block Model (SBM) is a powerful yet convenient framework for network data analysis, by postulating that each node in a network belongs to a community, in the presence of a finite number of communities. The community assignments are unobserved to the researcher. A SBM is highly effective in applications where community membership can be regarded as a discretized version of unobserved heterogeneity. The standard SBM has been applied widely in diverse areas, where an algorithm based on spectral clustering is often used. Statistical properties of these methodologies have been investigated intensively in the recent literature; see, Lei and Rinaldo (2015), Joseph and Yu (2016) and Rohe, Qin, and Yu (2016), just to name a few.

As an alternative approach, one can employ a model with more specific structures for the edge probability functions, while incorporating (possibly continuous-valued) unobserved heterogeneity through node-specific fixed effects. An advantage of such an approach is its ease of incorporating observed covariates into the model. If one is willing to accept a specific form of the edge probability function, e.g. scalar valued fixed effects representing unobserved heterogeneity, and the estimation

Date: This Version: Sep 11, 2024.
Keywords: Network Data.
JEL Classification Number: C14.

algorithm is computationally feasible, then such a method is practical and intuitive to use, and by having covariates in the model, it can shed valuable insights on, for example, the magnitude of homophily (or heterophily) effects in terms of observed characteristics in a given network.

The main goal of our paper is to develop a procedure that incorporates both unobserved heterogeneity (through unobserved community assignments) and observed heterogeneity (through covariates) building on the SBM framework. More specifically, our version of SBM lets each edge probability, together with the probability of community assignment, be impacted by observed covariates in a flexible manner. That is, we postulate that the edge probabilities and the community assignment probabilities are nonparametric function of covariates. Unobserved community assignment enters the model in a fully unrestricted way, since it simply indexes (or used as a label for) the edge probability and the community assignment probability. We note that letting the community assignment probability depend on covariates nonparametrically is an important feature of our model. It means that the observed covariates (or observed heterogeneity) and the unobserved heterogeneity can be correlated in an unspecified way, a feature often considered to be highly desirable in econometrics. Of course, this generality comes at the cost of additional technical complications, a part of the many theoretical challenges presented by our model, as is discussed shortly.

We propose a computationally tractable two-step procedure to estimate the conditional edge probabilities as well as the community assignment probabilities. The first step relies on a spectral clustering algorithm applied to a localized adjacency matrix of the network. We build on the k-nearest neighbor (k-nn) algorithm in our localization procedure, followed by the Singular Value Decomposition (SDV) to compute left/right singular vectors of the localized adjacency matrix, or, more precisely, a localized and normalized version of the Laplacian. Note that we need to employ SVD, as opposed to the eigen-value decomposition often used for the standard spectral clustering, even though we consider undirected network. This is because our localized and normalized network Laplacian depends on the covariate values at each of the two nodes, resulting its asymmetry. We then apply K-means clustering to extract communities, as in the standard spectral clustering algorithm, though it is implemented at each covariate value. As in the literature of the standard SBM without covariates, we provide non-asymptotic bounds for misclassificaiton error. Since we allow for correlation of unknown form between the community assignment and the covariates, it naturally induces independent but non-identically distributed (i.n.i.d) covariates when conditioned on community assignments, even though we assume random sampling for community assignments and observed covariates. We thus obtain some non-asymptotic results for the k-nn estimator under an i.n.i.d sampling; those can be of independent theoretical interest.

In the second step, once again we employ the k-nn regression algorithm, but this time in order to estimate the edge probability matrix and the community assignment probability vector. We apply the k-nn estimator to the extracted communities obtained from the first step. We study the statistical properties of these estimators while taking account of the effects of classification error in the first step, by obtaining non-asymptotic bounds for them.

1.1. **Notation.**

For a matrix M, $\|M\|$ denotes its spectral norm, $\|M\|_F$ its Frobenius norm, $\|M\|_{\max}$ its max norm. $\mathcal{B}(x,r) \subset \mathbb{R}^d$ is the closed ball of center x and radius r, λ is the Lebesgue measure on $\mathbb{R}^d$, $V_d = \int_{\mathcal{B}(0,1)} \mathrm{d}\lambda$ and I_k is the identity matrix of size k.

2. Model and Estimator

2.1. **Stochastic Block Model with Covariates.**

We consider a network of N nodes such that each node i has a d-vector of covariates $x(i)$ and belongs to a community $g(i) \in [G]$ where $G \in \mathbb{N}$. For each node i, the researcher observes $x(i)$ but does not observe the community $g(i)$. Let $\mathbb{M}_{N,G} \subset \mathbb{R}^{N\times G}$ the set of membership matrices, i.e., of matrices such that each row has exactly one nonzero coefficient set to 1. Let $\theta_i \in \mathbb{R}^G$ be a vector such that $\theta_{ig(i)} = 1$ and all other coefficients are 0. Then $\Theta := (\theta_1, ..., \theta_N)^\top \in \mathbb{M}_{N,G}$. The researcher also observes the adjacency matrix $A = (A_{ij})_{1\le i,j\le N}$. Define $\mathbf{x} = \{x(i), i \in [N]\}$ and $\mathbf{g} = \{g(i), i \in [N]\}$. In our stochastic block model with covariates, the distribution of the adjacency matrix conditional on $\mathbf{x}$ and $\mathbf{g}$ is given by a matrix-valued function $B : (x, x') \mapsto B(x, x') \in \mathbb{R}^{G\times G}$: conditional on $\mathbf{x}$ and $\mathbf{g}$, the entries A_{ij} are i.i.d Bernouilli random variables with

$$\Pr(A_{ij} = 1|\mathbf{x}, \mathbf{g}) = \Pr(A_{ij} = 1|x(i), x(j), g(i), g(j)) = B_{g(i)g(j)}(x(i), x(j)).$$

We note that the function B may vary with N and thus allows for sparsity. The distribution of $g(i)$ conditional on $x(i)$ is given by the functions $x \mapsto \pi_g(x) = \Pr(g(i) = g|x(i) = x)$, for $g \in [G]$. We assume that $(x(i), g(i))_{i\in[N]}$ are i.i.d random variables, and let $\mathcal{S}_X = \operatorname{supp}(x(i))$. However we state some of our results as nonasymptotic bounds conditional on $\mathbf{g}$, in which case it is understood that we only assume that $(x(i))_{i\in[N]}$ are independent draws and the marginal distribution of each depend on $g(i)$ only.

2.2. Construction of the estimator.

Our primary parameters of interest are $B(x, x')$ and $\pi(x) = (\pi_1(x), ..., \pi_G(x))$ for some $(x, x') \in \mathcal{S}_X \times \mathcal{S}_X$. Our estimators of these quantities rely on a spectral clustering algorithm applied to a truncated adjacency matrix, following the intuition of k-nearest neighbor regression. Before describing the algorithm and our estimators, we introduce the following notations.

Let $k \in \mathbb{N}$. We define the k-nearest neighbor (k-NN) radius of x as $r_k(x) := \inf\{r > 0 : |\mathcal{B}(x, r) \cap \mathbf{x}| = k\}$. The k-neighborhood of x is $\eta_N(x) := \{i \in [N] : ||x(i) - x|| \leq r_k(x)\}$. Let $\eta_N(x, x') := \{(x(i), x(j)) : x(i) \in \eta_N(x) \text{ and } x(j) \in \eta_N(x')\}$. For ease of notation, we sometimes write $\eta(x)$ and $\eta(x, x')$. For these units, we define the adjacency matrix $A^\eta \in \mathbb{R}^{k \times k}$: it is a submatrix of the network adjacency matrix A where rows index units in $\eta(x)$ and columns index units in $\eta(x')$. Similarly, $\Theta^\eta(x) \in \mathbb{M}_{k,G}$ is the membership matrix for the individuals in $\eta(x)$: it is a submatrix of Θ where rows index units in $\eta(x)$ and columns index communities. Note that we drop the dependence of A^η in x and x' for ease of notation. Let $O^\eta \in \mathbb{R}^{k \times k}$ and $Q^\eta \in \mathbb{R}^{k \times k}$ be the diagonal matrices such that $O^\eta_{ii} := \sum_{j \in \eta_N(x')} A^\eta_{ij}$ and $Q^\eta_{jj} := \sum_{i \in \eta_N(x)} A^\eta_{ij}$. For $\tau > 0$, we also define $O^\eta_\tau := O^\eta + \tau I_k$ and $Q^\eta_\tau := Q^\eta + \tau I_{k'}$. Finally, by analogy with the regularized graph Laplacian in the symmetric case, let

$$L^\eta_\tau := (O^\eta_\tau)^{-\frac{1}{2}} A^\eta (Q^\eta_\tau)^{-\frac{1}{2}}.$$

Our estimation procedure first applies a spectral clustering algorithm to L^η_τ to estimate the communities of units in $\eta(x)$ and $\eta(x')$. This spectral clustering algorithm proceeds as follows.

Spectral Clustering Algorithm:

Input: L^η_τ, G, approximation parameter for K-means.

Output: $\widehat{\Theta}^\eta(x) \in \mathbb{M}_{k,G}$ and $\widehat{\Theta}^\eta(x') \in \mathbb{M}_{k,G}$ estimators of the membership matrices.

Steps:

(1) Obtain the singular value decomposition of L^η_τ. Let $U \in \mathbb{R}^{k \times G}$ and $V \in \mathbb{R}^{k \times G}$ be the matrices of the top G left and right singular vectors, respectively.
(2) Apply K-means clustering on the rows of U, output $\widehat{\Theta}^\eta(x)$.
(3) Apply K-means clustering on the rows of V, output $\widehat{\Theta}^\eta(x')$.

Once these communities are estimated, we estimate $B_{gh}(x, x')$ and $\pi_g(x)$ by running k-NN regressions on the appropriate communities. For $i \in \eta(x)$, let $\hat{g}(i)$ be the estimated community, i.e., such that $\left(\widehat{\Theta}^\eta(x)\right)_{i,\hat{g}(i)} = 1$ and $\left(\widehat{\Theta}^\eta(x)\right)_{i,g} = 0$ for any $g \neq \hat{g}(i)$. Estimated communities of units in $\eta(x')$ are similarly defined. Let $\mathcal{G}_{hN}(x) = \{g(i) = h, i \in \eta_N(x)\}$ and $n_h(x) = |\mathcal{G}_{hN}(x)|$. Our estimator for $n_h(x)$ is $\hat{n}_h(x) := \#\{i \in \eta(x) \,:\, \hat{g}(i) = h\}$. We can finally introduce our estimators of the connection probabilities and community probabilities. The estimator of $\pi_h(x)$ is

$$\widehat{\pi}_h(x) = \frac{\hat{n}_h(x)}{k}. \tag{2.1}$$

The estimator of $B_{gh}(x, x')$ is

$$\widehat{B}_{gh}(x, x') = \frac{1}{\hat{n}_g(x)\hat{n}_h(x')} \sum_{\substack{i \in \eta(x)\,:\,\hat{g}(i)=g \\ j \in \eta(x')\,:\,\hat{g}(j)=h}} A_{ij}. \tag{2.2}$$

2.3. **Assumptions.**

We will maintain the following assumptions where S is a subset of $\mathcal{S}_X$, the support of x.

Assumption 2.1. *There exist constants $c > 0$ and $T > 0$ such that*

$$\lambda(S \cap \mathcal{B}(x, t)) \geq c\lambda(\mathcal{B}(x, t)), \;\; \forall t \in (0, T], \, \forall x \in S. \tag{2.3}$$

We consider the case where the covariates are continuous. Define the density of $x(i)$ given $g(i) = g$ as $f(.|g)$, for $g \in [G]$. Define also $\underline{f}(x) := \min_{g \in [G]} \; f(x|g)$.

Assumption 2.2. *There exist constants U_X and $b_X > 0$ such that*

$$U_X \geq \underline{f}(x) \geq b_X, \;\; \forall\, x \in S. \tag{2.4}$$

Similarly, define $\overline{f}(x) := \max_{g \in [G]} \; f(x|g)$.

Assumption 2.3. *There exist constants $\overline{U}_X$ and $\overline{b}_X > 0$ such that*

$$\overline{U}_X \geq \overline{f}(x) \geq \overline{b}_X, \;\; \forall\, x \in S. \tag{2.5}$$

Some smoothness assumptions will be imposed to study our estimators.

Assumption 2.4. $(x, x') \mapsto B_{gh}(x, x')$ *is Lipschitz continuous for all $(g, h) \in [G]^2$. We denote l_B the smallest Lipschitz constant.*

Assumption 2.5. $x \mapsto \pi_g(x)$ *is Lipschitz continuous for all $g \in [G]$. We denote l_π the smallest Lipschitz constant.*

Note that l_B may depend on N and thus captures sparsity.

3. Clustering

We fix a pair $(x,x') \in S^2$. The main result of this section is a high probability finite sample bound on a misclustering measure applied to $\widehat{\Theta}^\eta(x)$ and $\widehat{\Theta}^\eta(x')$. This result is obtained in 3 steps. We first obtain a finite sample bound on the difference between the graph Laplacian and its population counterpart. We then apply a version of Davis-Kahan theorem to bound the difference between $\hat{U}$, $\hat{V}$, and their population counterparts. Finally, we use theoretical properties of the K-means algorithm to bound the misclustering error.

3.1. Convergence of the graph Laplacian.

To introduce the population counterpart of L_τ^η, we define the following objects. Let $P(\mathbf{x},\mathbf{g}) := \mathbb{E}[A|\mathbf{x},\mathbf{g}]$. Thus, $P_{ij}(\mathbf{x},\mathbf{g}) = \mathbb{E}[A_{ij}|x(i),x(j),g(i),g(j)]$. Let also $P_{ij}(x,x',\mathbf{g}) = \mathbb{E}[A_{ij}|x(i)=x,x(j)=x',g(i),g(j)]$ and $P(x,x',\mathbf{g}) = (P_{ij}(x,x',\mathbf{g}))_{1\le i,j\le N}$. As in Section 2.2, we also define the localized matrices $P^\eta(\mathbf{x},\mathbf{g})$ and $P^\eta(x,x',\mathbf{g})$ as submatrices of the matrices $P(\mathbf{x},\mathbf{g})$ and $P(x,x',\mathbf{g})$ where rows index units in $\eta(x)$ and columns index units in $\eta(x')$. Let $\mathcal{O}^\eta(\mathbf{x},\mathbf{g}) \in \mathbb{R}^{k\times k}$, $\mathcal{O}^\eta(x,x',\mathbf{g}) \in \mathbb{R}^{k\times k}$, $\mathcal{Q}^\eta(\mathbf{x},\mathbf{g}) \in \mathbb{R}^{k\times k}$ and $\mathcal{Q}^\eta(x,x',\mathbf{g}) \in \mathbb{R}^{k\times k}$ be the diagonal matrices such that $\mathcal{O}_{ii}^\eta(\mathbf{x},\mathbf{g}) := \sum_{j\in\eta_N(x')} \mathbb{E}[A_{ij}^\eta|\mathbf{x},\mathbf{g}]$, $\mathcal{O}_{ii}^\eta(x,x',\mathbf{g}) := \sum_{j\in\eta_N(x')} P_{ij}(x,x',\mathbf{g})$, $\mathcal{Q}_{jj}^\eta(\mathbf{x},\mathbf{g}) := \sum_{i\in\eta_N(x)} \mathbb{E}[A_{ij}^\eta|\mathbf{x},\mathbf{g}]$ and finally $\mathcal{Q}_{jj}^\eta(x,x',\mathbf{g}) := \sum_{i\in\eta_N(x)} P_{ij}(x,x',\mathbf{g})$. Define also $\mathcal{O}_\tau^\eta(\mathbf{x},\mathbf{g}) := \mathcal{O}^\eta(\mathbf{x},\mathbf{g}) + \tau I_k$, $\mathcal{O}_\tau^\eta(x,x',\mathbf{g}) := \mathcal{O}^\eta(x,x',\mathbf{g}) + \tau I_k$, $\mathcal{Q}_\tau^\eta(\mathbf{x},\mathbf{g}) := \mathcal{Q}^\eta(\mathbf{x},\mathbf{g}) + \tau I_k$ and $\mathcal{Q}_\tau^\eta(x,x',\mathbf{g}) := \mathcal{Q}^\eta(x,x',\mathbf{g}) + \tau I_k$. Finally, let

$$\mathcal{L}_\tau^\eta(\mathbf{x},\mathbf{g}) := (\mathcal{O}_\tau^\eta(\mathbf{x},\mathbf{g}))^{-\frac{1}{2}} P^\eta(\mathbf{x},\mathbf{g}) (\mathcal{Q}_\tau^\eta(\mathbf{x},\mathbf{g}))^{-\frac{1}{2}},$$
$$\mathcal{L}_\tau^\eta(x,x',\mathbf{g}) := (\mathcal{O}_\tau^\eta(x,x',\mathbf{g}))^{-\frac{1}{2}} P^\eta(x,x',\mathbf{g}) (\mathcal{Q}_\tau^\eta(x,x',\mathbf{g}))^{-\frac{1}{2}}.$$

To obtain a finite sample bound, we will rely on concentration inequalities for symmetric matrices. However, $P^\eta(x,x',\mathbf{g})$ and A^η are not symmetric in general. Following Rohe, Qin, and Yu (2016), we first focus on their Hermitian dilations, where the Hermitian dilation of a matrix M, denoted $\widetilde{M}$, is given by

$$\widetilde{M} = \begin{pmatrix} 0 & M \\ M^\top & 0 \end{pmatrix}.$$

Define the minimum degree

$$d_{\min}(x,x',\mathbf{x}) := \min\left(\min_{i\in\eta_N(x)} \mathcal{O}_{ii}(x,x',\mathbf{g}), \min_{j\in\eta_N(x')} \mathcal{Q}_{jj}(x,x',\mathbf{g}), \min_{i\in\eta_N(x)} \mathcal{O}_{ii}(\mathbf{x},\mathbf{g}), \min_{j\in\eta_N(x')} \mathcal{Q}_{jj}(\mathbf{x},\mathbf{g}) \right).$$

In most of the computations, we drop the dependence in $(x,x',\mathbf{x})$. Define

$$R_k := \left(\frac{2k}{N b_X c V_d} \right)^{1/d}. \tag{3.1}$$

Lemma 3.1. *Let Assumptions 2.1, 2.2, 2.3 and 2.4 hold. For all N, $\delta \in (0,1)$, $\tau > 0$ and $1 \le k \le N$ such that*

(1) $\sup_{x \in S} r_k(x) \le R_k$,

(2) $3\ln(8k/\delta) \le d_{\min} + \tau$,

with probability at least $1-\delta$ conditional on $(\mathbf{g}, \mathbf{x})$, it holds that

$$\|\widetilde{L^\eta_\tau} - \widetilde{\mathcal{L}^\eta_\tau}(x, x', \mathbf{g})\| \le 4\sqrt{\frac{3\ln(8k/\delta)}{d_{\min}+\tau}} + \frac{2kl_B R_k}{d_{\min}+\tau}\left(\frac{2kl_B R_k}{d_{\min}+\tau} + 3\right). \tag{3.2}$$

Note that this finite sample bound holds conditionally on $(\mathbf{g}, \mathbf{x})$. It cannot be integrated directly because Conditions (1) and (2) are restrictions on $\mathbf{x}$. We show in Lemma A 1.5 that under certain assumptions, $\sup_{x\in S} r_k(x) \le R_k$ holds with high probability and we derive in Lemma 3.3 a high probability bound for $d_{\min}$. Lemma 3.4 obtains a high probability bound on $\|\widetilde{L^\eta_\tau} - \widetilde{\mathcal{L}^\eta_\tau}(x, x', \mathbf{g})\|$ which will hold conditional on $\mathbf{g}$ only.

Proof.

Let

$$D^\eta_\tau := \begin{pmatrix} O^\eta_\tau & 0 \\ 0 & Q^\eta_\tau \end{pmatrix},\ \mathcal{D}^\eta_\tau(\mathbf{x},\mathbf{g}) := \begin{pmatrix} \mathcal{O}^\eta_\tau(\mathbf{x},\mathbf{g}) & 0 \\ 0 & \mathcal{Q}^\eta_\tau(\mathbf{x},\mathbf{g}) \end{pmatrix} \text{ and } \mathcal{D}^\eta_\tau(x,x',\mathbf{g}) := \begin{pmatrix} \mathcal{O}^\eta_\tau(x,x',\mathbf{g}) & 0 \\ 0 & \mathcal{Q}^\eta_\tau(x,x',\mathbf{g}) \end{pmatrix}.$$

Note that

$$\begin{aligned}
\widetilde{L^\eta_\tau} &= (D^\eta_\tau)^{-\frac12}\widetilde{A^\eta}(D^\eta_\tau)^{-\frac12},\\
\widetilde{\mathcal{L}^\eta_\tau}(\mathbf{x},\mathbf{g}) &= \mathcal{D}^\eta_\tau(\mathbf{x},\mathbf{g})^{-\frac12}\widetilde{P^\eta}(\mathbf{x},\mathbf{g})\mathcal{D}^\eta_\tau(\mathbf{x},\mathbf{g})^{-\frac12}\\
\widetilde{\mathcal{L}^\eta_\tau}(x,x',\mathbf{g}) &= \mathcal{D}^\eta_\tau(x,x',\mathbf{g})^{-\frac12}\widetilde{P^\eta}(x,x',\mathbf{g})\mathcal{D}^\eta_\tau(x,x',\mathbf{g})^{-\frac12}.
\end{aligned}$$

We decompose

$$\begin{aligned}
\widetilde{L^\eta_\tau} - \widetilde{\mathcal{L}^\eta_\tau}(x,x',\mathbf{g}) &= (D^\eta_\tau)^{-\frac12}\widetilde{A^\eta}(D^\eta_\tau)^{-\frac12} - (\mathcal{D}^\eta_\tau(x,x',\mathbf{g}))^{-\frac12}\widetilde{P^\eta}(x,x',\mathbf{g})(\mathcal{D}^\eta_\tau(x,x',\mathbf{g}))^{-\frac12}\\
&= \underbrace{[(D^\eta_\tau)^{-\frac12} - (\mathcal{D}^\eta_\tau(\mathbf{x},\mathbf{g}))^{-\frac12}]\widetilde{A^\eta}(D^\eta_\tau)^{-\frac12}}_{=B_1} + \underbrace{(\mathcal{D}^\eta_\tau(\mathbf{x},\mathbf{g}))^{-\frac12}\widetilde{A^\eta}[(D^\eta_\tau)^{-\frac12} - (\mathcal{D}^\eta_\tau(\mathbf{x},\mathbf{g}))^{-\frac12}]}_{=B_2}\\
&\quad + \underbrace{(\mathcal{D}^\eta_\tau(\mathbf{x},\mathbf{g}))^{-\frac12}\widetilde{A^\eta}(\mathcal{D}^\eta_\tau(\mathbf{x},\mathbf{g}))^{-\frac12} - (\mathcal{D}^\eta_\tau(\mathbf{x},\mathbf{g}))^{-\frac12}\widetilde{P^\eta}(\mathbf{x},\mathbf{g})(\mathcal{D}^\eta_\tau(\mathbf{x},\mathbf{g}))^{-\frac12}}_{=B_3}\\
&\quad + \underbrace{(\mathcal{D}^\eta_\tau(\mathbf{x},\mathbf{g}))^{-\frac12}\widetilde{P^\eta}(\mathbf{x},\mathbf{g})(\mathcal{D}^\eta_\tau(\mathbf{x},\mathbf{g}))^{-\frac12} - (\mathcal{D}^\eta_\tau(x,x',\mathbf{g}))^{-\frac12}\widetilde{P^\eta}(x,x',\mathbf{g})(\mathcal{D}^\eta_\tau(x,x',\mathbf{g}))^{-\frac12}}_{=B_4}.
\end{aligned}$$

Note that by Lemma A 1.6, $\|\widetilde{L^\eta_\tau}\| \leq 1$. Thus

$$\begin{aligned} \|B_1\| &= \|[I - (\mathcal{D}^\eta_\tau(\mathbf{x},\mathbf{g}))^{-\frac{1}{2}}(D^\eta_\tau)^{\frac{1}{2}}](D^\eta_\tau)^{-\frac{1}{2}}\widetilde{A^\eta}(D^\eta_\tau)^{-\frac{1}{2}}\| \\ &= \|I - (\mathcal{D}^\eta_\tau(\mathbf{x},\mathbf{g}))^{-\frac{1}{2}}(D^\eta_\tau)^{\frac{1}{2}}\widetilde{L^\eta_\tau}\| \\ &\leq \|I - (\mathcal{D}^\eta_\tau(\mathbf{x},\mathbf{g}))^{-\frac{1}{2}}(D^\eta_\tau)^{\frac{1}{2}}\|. \end{aligned}$$

Likewise

$$\begin{aligned} \|B_2\| &= \|(\mathcal{D}^\eta_\tau(\mathbf{x},\mathbf{g}))^{-\frac{1}{2}}(D^\eta_\tau)^{\frac{1}{2}}(D^\eta_\tau)^{-\frac{1}{2}}\widetilde{A^\eta}(D^\eta_\tau)^{-\frac{1}{2}}[I - (D^\eta_\tau)^{\frac{1}{2}}(\mathcal{D}^\eta_\tau(\mathbf{x},\mathbf{g}))^{-\frac{1}{2}}]\| \\ &\leq \|(\mathcal{D}^\eta_\tau(\mathbf{x},\mathbf{g}))^{-\frac{1}{2}}(D^\eta_\tau)^{\frac{1}{2}}\|\|I - (D^\eta_\tau)^{\frac{1}{2}}(\mathcal{D}^\eta_\tau(\mathbf{x},\mathbf{g}))^{-\frac{1}{2}}\|. \end{aligned}$$

We follow Rohe, Qin, and Yu (2016) and use the two-sided Chernoff concentration inequality of Chung and Lu (2006) (see Theorem 2.4) to obtain for $i \in \eta_N(x) \cup \eta_N(x')$,

$$\Pr\left(|(\mathcal{D}^\eta(\mathbf{x},\mathbf{g}))_{ii} - (D^\eta)_{ii}| \geq a_0|\mathbf{g},\mathbf{x}\right) \leq \exp\left(-\frac{a_0^2}{2(\mathcal{D}^\eta(\mathbf{x},\mathbf{g}))_{ii}}\right) + \exp\left(-\frac{a_0^2}{2(\mathcal{D}^\eta(\mathbf{x},\mathbf{g}))_{ii} + \frac{2}{3}a_0}\right)$$

where the absence of τ as a subscript indicates $\tau = 0$. Take $a_0 = a_1(\mathcal{D}^\eta_\tau(\mathbf{x},\mathbf{g}))_{ii}$ where $0 < a_1 \leq 1$, then

$$\begin{aligned} \Pr\left[|(\mathcal{D}^\eta_\tau(\mathbf{x},\mathbf{g}))_{ii} - (D^\eta_\tau)_{ii}| \geq a_1(\mathcal{D}^\eta_\tau(\mathbf{x},\mathbf{g}))_{ii}|\mathbf{g},\mathbf{x}\right] &\leq 2\exp\left(-\frac{a_1^2(\mathcal{D}^\eta_\tau(\mathbf{x},\mathbf{g}))_{ii}}{3}\right) \\ &= 2\exp\left(-\frac{a_1^2\left[(\mathcal{D}^\eta(\mathbf{x},\mathbf{g}))_{ii} + \tau\right]}{3}\right) \\ &\leq 2\exp\left(-\frac{a_1^2\left[d_{\min} + \tau\right]}{3}\right). \end{aligned}$$

Note that for any $x \geq 0$, $|\sqrt{x} - 1| \leq |x - 1|$. Thus we have

$$\begin{aligned} \Pr\left[\|I - (\mathcal{D}^\eta_\tau(\mathbf{x},\mathbf{g}))^{-\frac{1}{2}}(D^\eta_\tau)^{\frac{1}{2}}\| \geq a_1|\mathbf{g},\mathbf{x}\right] &= \Pr\left[\max_{i\in\eta_N(x)\cup\eta_N(x')}\left|1 - \sqrt{\frac{(D^\eta_\tau)_{ii}}{(\mathcal{D}^\eta_\tau(\mathbf{x},\mathbf{g}))_{ii}}}\right| \geq a_1|\mathbf{g},\mathbf{x}\right] \\ &\leq \Pr\left[\max_{i\in\eta_N(x)\cup\eta_N(x')}\left|1 - \frac{(D^\eta_\tau)_{ii}}{(\mathcal{D}^\eta_\tau(\mathbf{x},\mathbf{g}))_{ii}}\right| \geq a_1|\mathbf{g},\mathbf{x}\right] \\ &\leq \sum_{i\in\eta_N(x)\cup\eta_N(x')} \Pr\left[|(\mathcal{D}^\eta_\tau(\mathbf{x},\mathbf{g}))_{ii} - (D^\eta_\tau)_{ii}| \geq a_1(\mathcal{D}^\eta_\tau(\mathbf{x},\mathbf{g}))_{ii}|\mathbf{g},\mathbf{x}\right] \\ &\leq 4k\exp\left(-\frac{a_1^2\left[d_{\min} + \tau\right]}{3}\right) \end{aligned}$$

On the event $\{\|I - (\mathcal{D}^\eta_\tau(\mathbf{x},\mathbf{g}))^{-\frac{1}{2}}(D^\eta_\tau)^{\frac{1}{2}}\| \leq a_1\}$,

$$\|(\mathcal{D}^\eta_\tau(\mathbf{x},\mathbf{g}))^{-\frac{1}{2}}(D^\eta_\tau)^{\frac{1}{2}}\| \leq \|I\| + \|I - (D^\eta_\tau)^{\frac{1}{2}}(\mathcal{D}^\eta_\tau(\mathbf{x},\mathbf{g}))^{-\frac{1}{2}}\| \leq 1 + a_1,$$

which implies that $\|B_1\| + \|B_2\| \leq a_1^2 + 2a_1 \leq 3a_1$. Thus,

$$\Pr\left(\|B_1\| + \|B_2\| \leq 3a_1 | \mathbf{g}, \mathbf{x}\right) \geq 1 - 4k \exp\left(-\frac{a_1^2 \left[d_{\min} + \tau\right]}{3}\right). \tag{3.3}$$

The expression for B_3 simplifies to

$$B_3 = (\mathcal{D}_\tau^\eta(\mathbf{x}, \mathbf{g}))^{-\frac{1}{2}} \left[\widetilde{A^\eta} - \widetilde{P^\eta}(\mathbf{x}, \mathbf{g})\right] (\mathcal{D}_\tau^\eta(\mathbf{x}, \mathbf{g}))^{-\frac{1}{2}}.$$

We follow Rohe, Qin, and Yu (2016) and decompose $B_3 = \sum_{i \in \eta_N(x)} \sum_{j \in \eta_N(x')} Y_{i,j}$ with

$$Y_{i,j} = \frac{A_{ij} - P_{ij}(\mathbf{x}, \mathbf{g})}{\left([\mathcal{O}_{ii}^\eta(\mathbf{x}, \mathbf{g}) + \tau][\mathcal{Q}_{ii}^\eta(\mathbf{x}, \mathbf{g}) + \tau]\right)^{1/2}} E^{i,k+j}$$

where $E^{i,j}$ is the square matrix of size $2k$ with ones at position (i, j) and at position (j, i) and zeros everywhere else. We apply a concentration inequality for symmetric matrices, see Theorem 5.4.1 of Vershynin (2018). $\|E^{i,j}\| = 1$ thus a bound on $\|Y_{i,j}\|$ is

$$\begin{aligned}
\|Y_{i,j}\| &\leq \left([\mathcal{O}_{ii}^\eta(\mathbf{x}, \mathbf{g}) + \tau][\mathcal{Q}_{ii}^\eta(\mathbf{x}, \mathbf{g}) + \tau]\right)^{-1/2} \\
&\leq \left([d_{\min} + \tau]\,[d_{\min} + \tau]\right)^{-1/2} = \frac{1}{d_{\min} + \tau}
\end{aligned}$$

and a bound on $\| \sum_{i \in \eta_N(x)} \sum_{j \in \eta_N(x')} \mathbb{E}[(Y_{i,j})^2 | \mathbf{g}, \mathbf{x}]\|$ is

$$\begin{aligned}
\| \sum_{i \in \eta_N(x)} \sum_{j \in \eta_N(x')} \mathbb{E}[(Y_{i,j})^2 | \mathbf{g}, \mathbf{x}]\| &= \| \sum_{i \in \eta_N(x)} \sum_{j \in \eta_N(x')} \left[\frac{P_{ij}(\mathbf{x}, \mathbf{g}) - P_{ij}(\mathbf{x}, \mathbf{g})^2}{[\mathcal{O}_{ii}^\eta(\mathbf{x}, \mathbf{g}) + \tau][\mathcal{Q}_{jj}^\eta(\mathbf{x}, \mathbf{g}) + \tau]} (E^{i,i} + E^{k+j,k+j})\right] \| \\
&= \| \sum_{i \in \eta_N(x)} \left[\sum_{j \in \eta_N(x')} \frac{P_{ij}(\mathbf{x}, \mathbf{g}) - P_{ij}(\mathbf{x}, \mathbf{g})^2}{[\mathcal{O}_{ii}^\eta(\mathbf{x}, \mathbf{g}) + \tau][\mathcal{Q}_{jj}^\eta(\mathbf{x}, \mathbf{g}) + \tau]}\right] E^{i,i} \\
&\qquad + \sum_{j \in \eta_N(x')} \left[\sum_{i \in \eta_N(x)} \frac{P_{ij}(\mathbf{x}, \mathbf{g}) - P_{ij}(\mathbf{x}, \mathbf{g})^2}{[\mathcal{O}_{ii}^\eta(\mathbf{x}, \mathbf{g}) + \tau][\mathcal{Q}_{jj}^\eta(\mathbf{x}, \mathbf{g}) + \tau]}\right] E^{k+j,k+j} \| \\
&= \max\left(\max_{i \in \eta_N(x)} \left(\sum_{j \in \eta_N(x')} \frac{P_{ij}(\mathbf{x}, \mathbf{g}) - P_{ij}(\mathbf{x}, \mathbf{g})^2}{[\mathcal{O}_{ii}^\eta(\mathbf{x}, \mathbf{g}) + \tau][\mathcal{Q}_{jj}^\eta(\mathbf{x}, \mathbf{g}) + \tau]}\right), \right. \\
&\qquad \left. \max_{j \in \eta_N(x')} \left(\sum_{i \in \eta_N(x)} \frac{P_{ij}(\mathbf{x}, \mathbf{g}) - P_{ij}(\mathbf{x}, \mathbf{g})^2}{[\mathcal{O}_{ii}^\eta(\mathbf{x}, \mathbf{g}) + \tau][\mathcal{Q}_{jj}^\eta(\mathbf{x}, \mathbf{g}) + \tau]}\right)\right) \\
&\leq \max\left(\frac{1}{d_{\min} + \tau} \max_{i \in \eta_N(x)} \left(\sum_{j \in \eta_N(x')} \frac{P_{ij}(\mathbf{x}, \mathbf{g})}{\mathcal{O}_{ii}^\eta(\mathbf{x}, \mathbf{g}) + \tau}\right), \frac{1}{d_{\min} + \tau} \max_{j \in \eta_N(x')} \left(\sum_{i \in \eta_N(x)} \frac{P_{ij}(\mathbf{x}, \mathbf{g})}{\mathcal{Q}_{jj}^\eta(\mathbf{x}, \mathbf{g}) + \tau}\right)\right) \\
&\leq \frac{1}{d_{\min} + \tau},
\end{aligned}$$

where the third equality holds by definition of the spectral norm. We apply Theorem 5.4.1 of Vershynin (2018) and obtain

$$\Pr\left(\|B_3\| \le a_1|\mathbf{g},\mathbf{x}\right) \ge 1 - 4k\exp\left(-\frac{a_1^2[d_{\min}+\tau]}{2+2a_1/3}\right). \tag{3.4}$$

The remaining term is B_4. We decompose

$$\begin{aligned} B_4 \;=\; & \underbrace{[(\mathcal{D}^\eta_\tau(\mathbf{x},\mathbf{g}))^{-\frac12} - (\mathcal{D}^\eta_\tau(x,x',\mathbf{g}))^{-\frac12}]\widetilde{P^\eta}(\mathbf{x},\mathbf{g})(\mathcal{D}^\eta_\tau(\mathbf{x},\mathbf{g}))^{-\frac12}}_{=B_{41}} \\ & + \underbrace{(\mathcal{D}^\eta_\tau(x,x',\mathbf{g}))^{-\frac12}\widetilde{P^\eta}(\mathbf{x},\mathbf{g})[(\mathcal{D}^\eta_\tau(\mathbf{x},\mathbf{g}))^{-\frac12} - (\mathcal{D}^\eta_\tau(x,x',\mathbf{g}))^{-\frac12}]}_{=B_{42}} \\ & + \underbrace{(\mathcal{D}^\eta_\tau(x,x',\mathbf{g}))^{-\frac12}\widetilde{P^\eta}(\mathbf{x},\mathbf{g})(\mathcal{D}^\eta_\tau(x,x',\mathbf{g}))^{-\frac12} - (\mathcal{D}^\eta_\tau(x,x',\mathbf{g}))^{-\frac12}\widetilde{P^\eta}(x,x',\mathbf{g})(\mathcal{D}^\eta_\tau(x,x',\mathbf{g}))^{-\frac12}}_{=B_{43}}. \end{aligned}$$

By Lemma A 1.7, $\|\widetilde{\mathcal{L}^\eta_\tau}(\mathbf{x},\mathbf{g})\| \le 1$. Thus,

$$\begin{aligned} \|B_{41}\| &\le \|I - (\mathcal{D}^\eta_\tau(x,x',\mathbf{g}))^{-\frac12}(\mathcal{D}^\eta_\tau(\mathbf{x},\mathbf{g}))^{\frac12}\|\|\widetilde{\mathcal{L}^\eta_\tau}(\mathbf{x},\mathbf{g})\| \\ &\le \|I - (\mathcal{D}^\eta_\tau(x,x',\mathbf{g}))^{-\frac12}(\mathcal{D}^\eta_\tau(\mathbf{x},\mathbf{g}))^{\frac12}\|, \end{aligned}$$

and

$$\begin{aligned} \|B_{42}\| &\le \|(\mathcal{D}^\eta_\tau(x,x',\mathbf{g}))^{-\frac12}(\mathcal{D}^\eta_\tau(\mathbf{x},\mathbf{g}))^{\frac12}\|\|\widetilde{\mathcal{L}^\eta_\tau}(\mathbf{x},\mathbf{g})\|\|I - (\mathcal{D}^\eta_\tau(\mathbf{x},\mathbf{g}))^{\frac12}(\mathcal{D}^\eta_\tau(x,x',\mathbf{g}))^{-\frac12}\| \\ &\le \|(\mathcal{D}^\eta_\tau(x,x',\mathbf{g}))^{-\frac12}(\mathcal{D}^\eta_\tau(\mathbf{x},\mathbf{g}))^{\frac12}\|\|I - (\mathcal{D}^\eta_\tau(\mathbf{x},\mathbf{g}))^{\frac12}(\mathcal{D}^\eta_\tau(x,x',\mathbf{g}))^{-\frac12}\|. \end{aligned}$$

Note that

$$\|I - (\mathcal{D}^\eta_\tau(x,x',\mathbf{g}))^{-\frac12}(\mathcal{D}^\eta_\tau(\mathbf{x},\mathbf{g}))^{\frac12}\| = \max_{i\in\eta_N(x)\cup\eta_N(x')}\left|1 - \sqrt{\frac{(\mathcal{D}^\eta_\tau(\mathbf{x},\mathbf{g}))_{ii}}{(\mathcal{D}^\eta_\tau(x,x',\mathbf{g}))_{ii}}}\right| \le \max_{i\in\eta_N(x)\cup\eta_N(x')}\left|1 - \frac{(\mathcal{D}^\eta_\tau(\mathbf{x},\mathbf{g}))_{ii}}{(\mathcal{D}^\eta_\tau(x,x',\mathbf{g}))_{ii}}\right|$$

If $i \in \eta_N(x)$,

$$(\mathcal{D}^\eta_\tau(\mathbf{x},\mathbf{g}))_{ii} = (\mathcal{O}^\eta_\tau(\mathbf{x},\mathbf{g}))_{ii} = \sum_{j\in\eta_N(x')}\mathbb{E}[A^\eta_{ij}|\mathbf{x},\mathbf{g}] + \tau = \sum_{j\in\eta_N(x')} B_{g(i),g(j)}(x(i),x(j)) + \tau$$

thus

$$\begin{aligned} \left|(\mathcal{D}^\eta_\tau(\mathbf{x},\mathbf{g}))_{ii} - (\mathcal{D}^\eta_\tau(x,x',\mathbf{g}))_{ii}\right| &\le \sum_{j\in\eta_N(x')}|B_{g(i),g(j)}(x(i),x(j)) - B_{g(i),g(j)}(x,x')| \\ &\le \sum_{j\in\eta_N(x')} l_B\|(x(i),x(j)) - (x,x')\| \\ &\le \sum_{j\in\eta_N(x')} l_B\left[||(x(i),x(j)) - (x(i),x')|| + ||(x(i),x') - (x,x')||\right] \end{aligned}$$

$$= \sum_{j \in \eta_N(x')} l_B \left[||x(j) - x'|| + ||x(i) - x|| \right] \leq k l_B (r_k(x) + r_k(x'))$$
$$\leq 2 k l_B R_k,$$

where the second inequality holds by Assumption 2.4 and the last inequality by Condition (1). We also have for $i \in \eta_N(x)$,

$$(\mathcal{D}^\eta_\tau(x, x', \mathbf{g}))_{ii} = \mathcal{O}_{ii}(x, x', \mathbf{g}) + \tau \geq d_{\min} + \tau \tag{3.5}$$

which implies

$$\left| 1 - \frac{(\mathcal{D}^\eta_\tau(\mathbf{x}, \mathbf{g}))_{ii}}{(\mathcal{D}^\eta_\tau(x, x', \mathbf{g}))_{ii}} \right| \leq \frac{2 k l_B R_k}{d_{\min} + \tau}.$$

The same bound holds for $i \in \eta_N(x')$. Thus

$$\|B_{41}\| \leq \frac{2 k l_B R_k}{d_{\min} + \tau}. \tag{3.6}$$

We also obtain

$$\|B_{42}\| \leq \frac{2 k l_B R_k}{d_{\min} + \tau} \left(\frac{2 k l_B R_k}{d_{\min} + \tau} + 1 \right). \tag{3.7}$$

The spectral norm of B_{43} can be bounded as follows

$$\|B_{43}\| \leq \|\mathcal{D}^\eta_\tau(x, x', \mathbf{g})\|^{-1} \, \|\widetilde{P^\eta}(\mathbf{x}, \mathbf{g}) - \widetilde{P^\eta}(x, x', \mathbf{g})\|.$$

By Equation (3.5), we have $\|\mathcal{D}^\eta_\tau(x, x', \mathbf{g})\| \geq d_{\min} + \tau$. As for the second term in the inequality above,

$$\|\widetilde{P^\eta}(\mathbf{x}, \mathbf{g}) - \widetilde{P^\eta}(x, x', \mathbf{g})\| \leq k \max_{\substack{i \in \eta_N(x) \\ j \in \eta_N(x')}} \left(\widetilde{P^\eta_{ij}}(\mathbf{x}, \mathbf{g}) - \widetilde{P^\eta_{ij}}(x, x', \mathbf{g}) \right)$$
$$= k \max_{\substack{i \in \eta_N(x) \\ j \in \eta_N(x')}} \left(B_{g(i), g(j)}(x(i), x(j)) - B_{g(i), g(j)}(x, x') \right)$$
$$\leq 2 k l_B R_k,$$

where the first inequality comes from the fact that for any matrix A of size $k \times k$, $\|A\| \leq \|A\|_F \leq k \max_{i,j} |A_{ij}|$ and the following equality from Assumption 2.4. Thus

$$\|B_{43}\| \leq \frac{2 k l_B R_k}{d_{\min} + \tau}. \tag{3.8}$$

Adding (3.6), (3.7) and (3.8), we obtain

$$\|B_4\| \leq \frac{2 k l_B R_k}{d_{\min} + \tau} \left(\frac{2 k l_B R_k}{d_{\min} + \tau} + 3 \right). \tag{3.9}$$

Putting (3.3), (3.4) and (3.9) together, and using the fact that for L events V_1,...,V_L, $\Pr(\cap_{l=1}^{L} V_l) \geq \sum_{l=1}^{L} \Pr(V_l) - L + 1$, we obtain

$$\begin{aligned}\Pr\left(\|\widetilde{L_\tau^\eta} - \widetilde{\mathcal{L}_\tau^\eta}(x,x',\mathbf{g})\| \leq 4a_1 + \frac{2kl_BR_k}{d_{\min}+\tau}\left(\frac{2kl_BR_k}{d_{\min}+\tau}+3\right) \mid \mathbf{g},\mathbf{x}\right) \\ \geq \Pr\left(\|B_1\| + \|B_2\| \leq 3a_1|\mathbf{g}, V(\mathbf{g})\right) + \Pr\left(\|B_3\| \leq a_1|\mathbf{g},\mathbf{x}\right) \\ + \Pr\left(\|B_4\| \leq \frac{2kl_BR_k}{d_{\min}+\tau}\left(\frac{2kl_BR_k}{d_{\min}+\tau}+3\right)|\mathbf{g},\mathbf{x}\right) - 2 \\ \geq 1 - 8k\exp\left(-\frac{a_1^2\left[d_{\min}+\tau\right]}{3}\right).\end{aligned}$$

Taking

$$a_1 = \sqrt{\frac{3\ln(8k/\delta)}{d_{\min}+\tau}},$$

then $a_1 \leq 1$ by Condition (2), and

$$\Pr\left(\|\widetilde{L_\tau^\eta} - \widetilde{\mathcal{L}_\tau^\eta}(x,x',\mathbf{g})\| \leq 4\sqrt{\frac{3\ln(8k/\delta)}{d_{\min}+\tau}} + \frac{2kl_BR_k}{d_{\min}+\tau}\left(\frac{2kl_BR_k}{d_{\min}+\tau}+3\right) \mid \mathbf{g},\mathbf{x}\right) \geq 1-\delta.$$

□

3.2. Lower bound on $d_{\min}(x, x', \mathbf{x})$.

The finite sample bound obtained in Lemma 3.1 is valid under restrictions on $d_{\min}(x,x',\mathbf{x})$, a localized minimum expected degree. In this section, we derive a nonasymptotic probability bound on $d_{\min}(x,x',\mathbf{x})$ and combine it with Lemmas 3.1 and A 1.5 to obtain a finite sample bound on $\|\widetilde{L_\tau^\eta} - \widetilde{\mathcal{L}_\tau^\eta}(x,x',\mathbf{g})\|$ conditional on $\mathbf{g}$ only and with explicit dependence in k. For $h \in [G]$, let

$$\begin{aligned} n_h(x) &:= \#\{g(i) = h, i \in \eta_N(x)\} \\ N_h &:= \#\{g(i) = h, i \in [N]\} \\ \underline{n_h} &:= \min_{x\in S} n_h(x). \end{aligned}$$

Define

$$\Delta = \min_{g\in[G]} \max_{h\in[G]} \inf_{(x,x')\in S^2} B_{gh}(x,x')$$

Note that as for B and l_B, Δ may vary with N and thus captures sparsity. We assume that $\Delta > 0$. A natural lower bound on $d_{\min}(x,x',\mathbf{x})$ therefore involves $\min_{h\in[G]} \underline{n_h}$. We show in Lemma A 1.5 that under certain assumptions, $\inf_{x\in S} r_k(x) \geq \underline{R}_k$ holds with high probability, where

$$\underline{R}_k := \left(\frac{k - 12d\ln(12N/\delta)}{4N\overline{U}_X V_d}\right)^{1/d}. \tag{3.10}$$

The following lemma uses this result and establishes a lower bound on $\min_{h\in[G]} \underline{n_h}$.

Lemma 3.2. *Let Assumptions 2.1, 2.2 and 2.3 hold. Then for all N, $\delta \in (0,1)$, $1 \le k \le N$ such that*

(1) $k \ge 12d\ln(24GN/\delta)$,

(2) $k \le 8T^d V_d \overline{U}_X N$,

(3) and for all $h \in [G]$, $\frac{c}{16}\frac{N_h}{N}\frac{b_X}{\overline{U}_X}\ k \ge 24d\ln(24GN_h/\delta) + 1$,

with probability at least $1-\delta$ *conditional on* $\mathbf{g}$*, it holds that*

$$\min_{h\in[G]} \underline{n_h} \ge \min_{h\in[G]} \left\lfloor \frac{c}{16}\frac{N_h}{N}\frac{b_X}{\overline{U}_X}\ k \right\rfloor.$$

Proof. Fix $\delta \in (0,1)$ and k satisfying the conditions of the lemma. Note that

$$\underline{R}_k \ge \left(\frac{k}{8N\overline{U}_X V_d}\right)^{1/d}.$$

We apply Lemma 4 of Portier (2021) to the subpopulation $\{i : g(i) = h\}$. We define $r_l^h(x)$ to be the l-NN radius of $x \in S$ for this subpopulation, that is,

$$r_l^h(x) := \inf\{r > 0 : |\mathcal{B}(x,r) \cap \{x(i) : g(i) = h\}| = l\}.$$

By Assumptions 2.2 and 2.3, we have

$$0 < b_X \le f(x|h) \le \overline{U}_X,\ \forall x \in S, \tag{3.11}$$

and by Assumption 2.1, Condition (7) of Portier (2021) also holds for $f(x|h)$. For any l and $\delta' > 0$ such that $24d\ln(12N_h/\delta') \le l \le T^d N_h b_X c V_d/2$, let

$$\overline{\tau}_l^h := \left(\frac{2l}{N_h b_X c V_d}\right)^{1/d}.$$

By Lemma 4 of Portier (2021), with probability at least $1-\delta'$ conditional on $\mathbf{g}$, it holds that

$$\sup_{x\in S} r_l^h(x) \le \overline{\tau}_l^{S,h}.$$

We take $l_h = \left\lfloor \frac{c}{16}\frac{N_h}{N}\frac{b_X}{\overline{U}_X}\ k \right\rfloor$ and $\delta' = \frac{\delta}{2G}$. Then

$$l_h \le \frac{c}{16}\frac{N_h}{N}\frac{b_X}{\overline{U}_X}\ k \le T^d N_h b_X c V_d/2,$$

by Condition (2) and

$$l_h \ge 24d\ln(12N_h/\delta'),$$

by Condition (3). Moreover,

$$\overline{r}^h_{l_h} \leq \left(\frac{k}{8N\overline{U}_X V_d}\right)^{1/d} \leq \underline{R}_k.$$

This implies

$$\{\underline{n_h} \geq l_h\} = \{n_h(x) \geq l_h \text{ for every } x \in S\} = \{r^h_{l_h}(x) \leq r_k(x) \text{ for every } x \in S\ \}$$
$$\supseteq \{\sup_{x\in S} r^h_{l_h}(x) \leq \overline{r}^h_{l_h}\} \cap \{\inf_{x\in S} r_k(x) \geq \underline{R}_k\}.$$

Define $L = \min_{h\in[G]} \left\lfloor \frac{c}{16}\frac{N_h}{N}\frac{b_X}{\overline{U}_X}\ k\right\rfloor = \min_{h\in[G]} l_h$. We apply the equation above to all $h \in [G]$ and obtain

$$\begin{aligned}
&\Pr\left(\min_{h\in[G]} \underline{n_h} \geq L \,|\, \mathbf{g}\right) \geq \sum_{h\in[G]} \Pr\left(\underline{n_h} \geq L \,|\, \mathbf{g}\right) - G + 1\\
&\geq \sum_{h\in[G]} \Pr\left(\underline{n_h} \geq l_h \,|\, \mathbf{g}\right) - G + 1\\
&\geq \sum_{h\in[G]} \Pr\left(\{\sup_{x\in S} r^h_{l_h}(x) \leq \overline{r}^h_{l_h}\} \cap \{\inf_{x\in S} r_k(x) \geq \underline{R}_k\} \,|\, \mathbf{g}\right) - G + 1\\
&\geq \sum_{h\in[G]} \left[\Pr\left(\sup_{x\in S} r^h_{l_h}(x) \leq \overline{r}^h_{l_h} \,|\, \mathbf{g}\right) + \Pr\left(\inf_{x\in S} r_k(x) \geq \underline{R}_k \,|\, \mathbf{g}\right) - 1\right] - G + 1\\
&\geq G(1-2\delta') - G + 1 = 1 - \delta.
\end{aligned}$$

where in the last inequality, we used $k \geq 12d\ln(12N/\delta')$, guaranteed by Condition (1), together with (A 1.8) of Lemma A 1.5. □

Lemma 3.3. *Let Assumptions 2.1, 2.2 and 2.3 hold. Then for all N, $\delta \in (0,1)$, $1 \leq k \leq N$ such that*

(1) $k \geq 12d\ln(24GN/\delta)$,
(2) $k \leq 8T^d V_d \overline{U}_X N$,
(3) and for all $h \in [G]$, $\frac{c}{16}\frac{N_h}{N}\frac{b_X}{\overline{U}_X}\ k \geq 24d\ln(24GN_h/\delta) + 1$,

with probability at least $1-\delta$ conditional on $\mathbf{g}$, it holds that

$$d_{\min}(x, x', \mathbf{x}) \geq \Delta \min_{h\in[G]} \left\lfloor \frac{c}{16}\frac{N_h}{N}\frac{b_X}{\overline{U}_X}\ k\right\rfloor.$$

Proof. Recall that

$$d_{\min}(x,x',\mathbf{x}) = \min\left(\min_{i\in\eta_N(x)} \mathcal{O}_{ii}(x,x',\mathbf{g}),\ \min_{j\in\eta_N(x')} \mathcal{Q}_{jj}(x,x',\mathbf{g}),\ \min_{i\in\eta_N(x)} \mathcal{O}_{ii}(\mathbf{x},\mathbf{g}),\ \min_{j\in\eta_N(x')} \mathcal{Q}_{jj}(\mathbf{x},\mathbf{g})\right).$$

Note that

$$\begin{aligned}
\mathcal{O}_{ii}(x,x',\mathbf{g}) &= \sum_{j\in\eta_N(x')} P_{ij}(x,x',\mathbf{g}) = \sum_{h\in[G]} n_h(x')B_{g(i)h}(x,x') \\
&\geq \sum_{h\in[G]} n_h(x') \inf_{(x,x')\in S^2} B_{g(i)h}(x,x') \\
&\geq \min_{h\in[G]} n_h(x') \max_{h'\in[G]} \inf_{(x,x')\in S^2} B_{g(i)h'}(x,x') \\
&\geq \min_{h\in[G]} \min_{x\in S} n_h(x) \min_{g\in[G]} \max_{h'\in[G]} \inf_{(x,x')\in S^2} B_{gh'}(x,x') \\
&\geq \Delta \min_{h\in[G]} \underline{n_h}.
\end{aligned}$$

The same inequality holds for $\mathcal{Q}_{jj}(x,x',\mathbf{g})$. For the remaining terms in the definitions of $d_{\min}(x,x',\mathbf{x})$, note that

$$\begin{aligned}
\mathcal{O}_{ii}(\mathbf{x},\mathbf{g}) &= \sum_{j\in\eta_N(x')} P_{ij}(\mathbf{x},\mathbf{g}) \\
&= \sum_{j\in\eta_N(x')} \mathbb{E}[A^{\eta}_{ij}|x(i),x(j),g(i),g(j)] \\
&= \sum_{h\in[G]} \sum_{j\in\eta_N(x'):g(j)=h} B_{g(i)h}(x(i),x(j)) \\
&\geq \sum_{h\in[G]} n_h(x') \inf_{(x,x')\in S^2} B_{g(i)h}(x,x') \\
&\geq \min_{h\in[G]} n_h(x') \max_{h\in[G]} \inf_{(x,x')\in S^2} B_{g(i)h}(x,x') \\
&\geq \min_{h\in[G]} \min_{x\in S} n_h(x) \min_{g\in[G]} \max_{h\in[G]} \inf_{(x,x')\in S^2} B_{gh}(x,x') \\
&\geq \Delta \min_{h\in[G]} \underline{n_h}
\end{aligned}$$

This also holds for $\mathcal{Q}_{jj}(\mathbf{x},\mathbf{g})$. Thus $d_{\min}(x,x',\mathbf{x}) \geq \Delta \min_{h\in[G]} \underline{n_h}$ and the result holds by Lemma 3.2. □

Combining Lemma 3.1 together with Lemma 3.3, we obtain a new bound on $\|\widetilde{L^{\eta}} - \widetilde{\mathcal{L}^{\eta}}(x,x',\mathbf{g})\|$.

Lemma 3.4. *Let Assumptions 2.1, 2.2, 2.3 and 2.4 hold. Then for all N, $\delta \in (0,1)$, $1 \leq k \leq N$ such that*

(1) $\Delta \min_{h\in[G]} \left\lfloor \frac{c}{16} \frac{N_h}{N} \frac{b_X}{\overline{U}_X}\ k \right\rfloor + \tau \geq 3\ln(24k/\delta)$,
(2) $k \geq \max(12d\ln(72GN/\delta), 24d\ln(36N/\delta))$,
(3) $k \leq \min(8T^d V_d \overline{U}_X N, (1/2)T^d V_d b_X cN)$,

(4) and for all $h \in [G]$, $\frac{c}{16}\frac{N_h}{N}\frac{b_X}{\overline{U}_X}\ k \geq 24d\ln(72GN_h/\delta)+1$,

with probability at least $1-\delta$ *conditional on* $\mathbf{g}$*, it holds that*

$$\|\widetilde{L^\eta_\tau} - \widetilde{\mathcal{L}^\eta_\tau}(x,x',\mathbf{g})\| \leq 4\sqrt{\frac{3\ln(24k/\delta)}{\Delta\min_{h\in[G]}\left\lfloor \frac{c}{16}\frac{N_h}{N}\frac{b_X}{\overline{U}_X}\ k\right\rfloor + \tau}}$$

$$+\frac{2kl_BR_k}{\Delta\min_{h\in[G]}\left\lfloor \frac{c}{16}\frac{N_h}{N}\frac{b_X}{\overline{U}_X}\ k\right\rfloor + \tau}\left(\frac{2kl_BR_k}{\Delta\min_{h\in[G]}\left\lfloor \frac{c}{16}\frac{N_h}{N}\frac{b_X}{\overline{U}_X}\ k\right\rfloor + \tau}+3\right). \tag{3.12}$$

Proof. In what follows we use some simplified notation. Let C be the function such that $C(d_{\min})$ is the right hand side of (3.2). We also define the following two events, $V(\mathbf{g},\mathbf{x}) := \{\sup_{x\in S} r_k(x) \leq R_k\}$ and $W(\mathbf{g},\mathbf{x}) := \left\{\min_{h\in[G]} \underline{n_h} \geq \min_{h\in[G]}\left\lfloor \frac{c}{16}\frac{N_h}{N}\frac{b_X}{\overline{U}_X}\ k\right\rfloor\right\} \subset \left\{d_{\min} \geq \Delta\min_{h\in[G]}\left\lfloor \frac{c}{16}\frac{N_h}{N}\frac{b_X}{\overline{U}_X}\ k\right\rfloor\right\}$ where the inclusion follows by the proof of Lemma 3.3. Then,

$$\begin{aligned}
&\Pr\left[\|\widetilde{L^\eta_\tau} - \widetilde{\mathcal{L}^\eta_\tau}(x,x',\mathbf{g})\| \leq C\left(\Delta\min_{h\in[G]}\left\lfloor \frac{c}{16}\frac{N_h}{N}\frac{b_X}{\overline{U}_X}\ k\right\rfloor\right)\ \Big|\, \mathbf{g}\right]\\
&\geq \Pr\left[\left\{\|\widetilde{L^\eta_\tau} - \widetilde{\mathcal{L}^\eta_\tau}(x,x',\mathbf{g})\| \leq C\left(\Delta\min_{h\in[G]}\left\lfloor \frac{c}{16}\frac{N_h}{N}\frac{b_X}{\overline{U}_X}\ k\right\rfloor\right)\right\}\cap V(\mathbf{g},\mathbf{x})\cap W(\mathbf{g},\mathbf{x})\,\Big|\, \mathbf{g}\right]\\
&\geq \Pr\left[\left\{\|\widetilde{L^\eta_\tau} - \widetilde{\mathcal{L}^\eta_\tau}(x,x',\mathbf{g})\| \leq C\left(d_{\min}\right)\right\}\cap V(\mathbf{g},\mathbf{x})\cap W(\mathbf{g},\mathbf{x})\,\Big|\, \mathbf{g}\right]\\
&= \mathbb{E}\left[\Pr\left[\left\{\|\widetilde{L^\eta_\tau} - \widetilde{\mathcal{L}^\eta_\tau}(x,x',\mathbf{g})\| \leq C\left(d_{\min}\right)\right\}\ \Big|\, \mathbf{g},\mathbf{x},V(\mathbf{g},\mathbf{x}),W(\mathbf{g},\mathbf{x})\right]\ \Big|\, \mathbf{g},V(\mathbf{g},\mathbf{x}),W(\mathbf{g},\mathbf{x})\right]\\
&\quad\times\Pr\left[V(\mathbf{g},\mathbf{x})\cap W(\mathbf{g},\mathbf{x})\,\Big|\,\mathbf{g}\right].
\end{aligned}$$

By Lemma 3.1 and $\Delta\min_{h\in[G]}\left\lfloor \frac{c}{16}\frac{N_h}{N}\frac{b_X}{\overline{U}_X}\ k\right\rfloor + \tau \geq 3\ln(24k/\delta)$,

$$\Pr\left[\left\{\|\widetilde{L^\eta_\tau} - \widetilde{\mathcal{L}^\eta_\tau}(x,x',\mathbf{g})\| \leq C\left(d_{\min}\right)\right\}\ \Big|\, \mathbf{g},\mathbf{x},V(\mathbf{g},\mathbf{x}),W(\mathbf{g},\mathbf{x})\right] \geq 1-\delta/3.$$

Moreover, by Lemma A 1.5 and $24d\ln(36N/\delta) \leq k \leq T^dNb_XcV_d/2$,

$$\Pr\left[V(\mathbf{g},\mathbf{x})\,\Big|\,\mathbf{g}\right] \geq 1-\delta/3.$$

Since

(1) $k \geq 12d\ln(72GN/\delta)$,
(2) $k \leq 8T^dV_d\overline{U}_XN$,
(3) and for all $h\in[G]$, $\frac{c}{16}\frac{N_h}{N}\frac{b_X}{\overline{U}_X}\ k \geq 24d\ln(72GN_h/\delta)+1$,

we also have by Lemma 3.3,

$$\Pr\left[W(\mathbf{g},\mathbf{x})\,\Big|\,\mathbf{g}\right] \geq 1-\delta/3.$$

Thus

$$\Pr\left[\|\widetilde{L^{\eta}_{\tau}} - \widetilde{\mathcal{L}^{\eta}_{\tau}}(x,x',\mathbf{g})\| \le C\left(\Delta \min_{h\in[G]}\left\lfloor \frac{c}{16}\frac{N_h}{N}\frac{b_X}{\overline{U}_X}\; k\right\rfloor\right)\,\middle|\,\mathbf{g}\right]$$
$$\ge (1-\delta/3)\left(\Pr\left[V(\mathbf{g},\mathbf{x})\,\middle|\,\mathbf{g}\right] + \Pr\left[W(\mathbf{g},\mathbf{x})\,\middle|\,\mathbf{g}\right] - 1\right) \ge (1-\delta/3)(1-2\delta/3) \ge 1-\delta.$$

□

Remark 3.1. The variance term is the first term in (3.12). Consider the case where no regularization is made ($\tau = 0$) and sparsity is captured by a parameter ρ_N, i.e., such that $B = \rho_N B_0$ where B_0 does not vary with N, see e.g. Bickel, Chen, and Levina (2011)). Approximating $N_h/N \approx C$ for all $h \in [G]$, the variance term in (3.12) simplifies to

$$C\sqrt{\frac{\ln(24k/\delta)}{\rho_N k}}, \tag{3.13}$$

where C is a constant subject to change in value. The normalization in the definition of the Laplacian is not unlike that of the k-NN regression estimator. However in comparison to standard k-NN regression, e.g., Jiang (2019), the variance term has an extra $\sqrt{\ln k}$ term in the numerator, which comes from the growing dimension of $\widetilde{L^{\eta}_{\tau}}$ and appears when applying various matrix concentration inequalities. On the other hand, (3.13) does not have a $\ln N$ term in the numerator because it is a pointwise bound at $(x, x') \in S^2$.

Remark 3.2. Let $\tau = 0$, sparsity be captured by a parameter ρ_N and $N_h/N \approx C$ for all $h \in [G]$. Then the bias term in (3.12) simplifies to $\approx R_k(R_k + C) \approx R_k$ if $R_k \to 0$ as $N \to \infty$, where C is a constant subject to change in value. This is similar to Jiang (2019), confirming the intuition that the effective dimension is that of the covariates and not twice as much. Note that the bias, a novel term in this type of calculations, is not impacted by sparsity.

Remark 3.3. Let $\tau = 0$, sparsity be captured by a parameter ρ_N and $N_h/N \approx C$ for all $h \in [G]$. Then as long as $N\rho_N \to \infty$, taking $k \approx (N^2/\rho_N^d)^{\frac{1}{d+2}}$ gives $\|\widetilde{L^{\eta}} - \widetilde{\mathcal{L}^{\eta}}(x,x',\mathbf{g})\| \approx (\rho_N N)^{\frac{-1}{d+2}}$, up to $\ln N$ and $\ln \rho_N$ factors: the rate obtained is as Jiang (2019), see Remark 1, where the sample size is replaced with $N\rho_N$ due to sparsity.

3.3. **Clustering.**

In this section, for ease of readability, we do not display dependence of many of the defined objects in x, x', τ, k, etc. Using the Laplacian L^{η}_{τ} we compute its top G left/right singular vectors, to obtain the SVD of L_τ^η

$$L_\tau^\eta = U\widehat{\Lambda}V^\top$$

where $U, V \in \mathbb{R}^{k\times G}$. We then apply K-means clustering to U and V, solving

$$(\widehat{\Theta}^\eta(x), Z_U) = \underset{\Theta\in\mathbb{M}_{k,G}, Z\in\mathbb{R}^{G\times G}}{\operatorname{argmin}} \|\Theta Z - U\|_F \tag{3.14}$$

$$(\widehat{\Theta}^\eta(x'), Z_V) = \underset{\Theta\in\mathbb{M}_{k,G}, Z\in\mathbb{R}^{G\times G}}{\operatorname{argmin}} \|\Theta Z - V\|_F \tag{3.15}$$

It is possible to use a faster clustering algorithm by using an approximate solution instead, incorporating the approximation error explicitly in the following analysis as in Lei and Rinaldo (2015).

Note that the i-th element of the diagonal matrix $\mathcal{O}_\tau^\eta(x, x', \mathbf{g})$ only depends on $g(i)$. Define the diagonal matrix $\overline{\mathcal{O}} \in \mathbb{R}^{G\times G}$ collecting these coefficients, i.e., such that $\overline{\mathcal{O}}_{gg} = \sum_{j\in\eta_N(x')} B_{gg(j)}(x, x') + \tau$. Similarly, define the diagonal matrix $\overline{\mathcal{Q}} \in \mathbb{R}^{G\times G}$ such that $\overline{\mathcal{Q}}_{gg} = \sum_{i\in\eta_N(x)} B_{g(i)g}(x, x') + \tau$.

Define $\mathcal{N}(x) = \operatorname{diag}\left(\sqrt{n_1(x)}, ..., \sqrt{n_G(x)}\right)$ and let

$$\mathcal{N}(x)\overline{\mathcal{O}}^{-1/2}B(x, x')\overline{\mathcal{Q}}^{-1/2}\mathcal{N}(x') = z_U\Lambda_\tau z_V^\top \tag{3.16}$$

be the SVD of $B(x, x')$ after normalization by $\mathcal{N}(x)\overline{\mathcal{O}}^{-1/2}$ and $\mathcal{N}(x')\overline{\mathcal{Q}}^{-1/2}$. Note that

$$\mathcal{L}_\tau^\eta(x, x', \mathbf{g}) = (\mathcal{O}_\tau^\eta(x, x', \mathbf{g}))^{-\frac{1}{2}}\Theta^\eta(x)B(x, x')\Theta^\eta(x')^\top(\mathcal{Q}_\tau^\eta(x, x', \mathbf{g}))^{-\frac{1}{2}},$$

with $(\mathcal{O}_\tau^\eta(x, x', \mathbf{g}))^{-\frac{1}{2}}\Theta^\eta(x) = \Theta^\eta(x)\overline{\mathcal{O}}^{-1/2}$ and $(\mathcal{Q}_\tau^\eta(x, x', \mathbf{g}))^{-\frac{1}{2}}\Theta^\eta(x') = \Theta^\eta(x')\overline{\mathcal{Q}}^{-1/2}$. Thus

$$\mathcal{L}_\tau^\eta(x, x', \mathbf{g}) = \Theta^\eta(x)\mathcal{N}(x)^{-1}z_U\Lambda_\tau z_V^\top\mathcal{N}(x')^{-1}\Theta^\eta(x')^\top. \tag{3.17}$$

From this expression, by slight modification of the argument in the proof of Lemma 2.1 in Lei and Rinaldo (2015), considering SVD instead of eigenvalue decomposition, and taking account of the normalization of the Laplacian, we see that the SVD of $\mathcal{L}^\eta(x, x', \mathbf{g})$ is given by

$$\mathcal{L}^\eta(x, x', \mathbf{g}) = \mathcal{U}\Lambda_\tau\mathcal{V}^\top$$

with $\mathcal{U} = \Theta^\eta(x)\mathcal{Z}_U$ and $\mathcal{V} = \Theta^\eta(x')\mathcal{Z}_V$ where $\mathcal{Z}_U = \mathcal{N}(x)^{-1}z_U$ and $\mathcal{Z}_V = \mathcal{N}(x)^{-1}z_V$.

Note

$$\mathcal{Z}_U\mathcal{Z}_U^\top = \mathcal{N}(x)^{-1}z_Uz_U^\top\mathcal{N}(x)^{-1} = \operatorname{diag}\left(\frac{1}{n_1(x)}, ..., \frac{1}{n_G(x)}\right) \tag{3.18}$$

As in Rohe, Qin, and Yu (2016), define

$$\tilde{\mathcal{U}} := \frac{1}{\sqrt{2}}\begin{pmatrix}\mathcal{U}\\ \mathcal{V}\end{pmatrix},$$

$$\tilde{U} := \frac{1}{\sqrt{2}}\binom{U}{V},$$

then $\tilde{\mathcal{U}}$ and $\tilde{U}$ are the eigenvectors corresponding to top G eigenvalues of $\widetilde{\mathcal{L}^\eta}$ and $\widetilde{L^\eta}$, respectively. We will use $\lambda_1(x,x') \geq \lambda_2(x,x') \geq ... \geq \lambda_G(x,x')$ and $\hat{\lambda}_1(x,x') \geq \hat{\lambda}_2(x,x') \geq ... \geq \hat{\lambda}_G(x,x')$ to denote them although we drop (x,x') in most of our computations. We now have

Lemma 3.5.

$$\|\tilde{U}\tilde{U}^\top - \tilde{\mathcal{U}}\tilde{\mathcal{U}}^\top\|_F \leq \frac{2\sqrt{2G}}{\lambda_G}\|\widetilde{L^\eta_\tau} - \widetilde{\mathcal{L}^\eta_\tau}(x,x',\mathbf{g})\|.$$

Moreover, for some $G \times G$ orthogonal matrices Q_U and Q_V,

$$\|U - \mathcal{U}Q_U\|_F \leq \frac{4\sqrt{2G}}{\lambda_G}\|\widetilde{L^\eta_\tau} - \widetilde{\mathcal{L}^\eta_\tau}(x,x',\mathbf{g})\|.$$

and

$$\|V - \mathcal{V}Q_V\|_F \leq \frac{4\sqrt{2G}}{\lambda_G}\|\widetilde{L^\eta_\tau} - \widetilde{\mathcal{L}^\eta_\tau}(x,x',\mathbf{g})\|.$$

Proof. We employ the proof strategy developed by Lei and Rinaldo (2015) and Rohe, Qin, and Yu (2016) with suitable modification. Note that these eigenvalues are equal to the top G singular values of $\mathcal{L}^\eta$ and L^η, respectively. By Davis-Kahan Theorem (see Theorem VII.3.1 in Bhatia (2013)) and noting that the eigengap between $\hat{\lambda}_j, j \leq G$ and $\lambda_j, j \geq G+1$ is $\hat{\lambda}_G$ we have

$$\|(I - \tilde{\mathcal{U}}\tilde{\mathcal{U}}^\top)\tilde{U}\tilde{U}^\top\| \leq \frac{1}{\hat{\lambda}_G}\|\widetilde{L^\eta_\tau} - \widetilde{\mathcal{L}^\eta_\tau}(x,x',\mathbf{g})\|.$$

But

$$\begin{aligned}\hat{\lambda}_G &\geq \lambda_G - |\hat{\lambda}_G - \lambda_G| \\ &\geq \lambda_G - \|\widetilde{L^\eta_\tau} - \widetilde{\mathcal{L}^\eta_\tau}(x,x',\mathbf{g})\|\end{aligned}$$

where the second inequality follows from Weyl's Theorem. It follows that

$$\|(I - \tilde{\mathcal{U}}\tilde{\mathcal{U}}^\top)\tilde{U}\tilde{U}^\top\| \leq \frac{1}{\lambda_G - \|\widetilde{L^\eta_\tau} - \widetilde{\mathcal{L}^\eta_\tau}(x,x',\mathbf{g})\|}\|\widetilde{L^\eta_\tau} - \widetilde{\mathcal{L}^\eta_\tau}(x,x',\mathbf{g})\|.$$

If $\|\widetilde{L^\eta_\tau} - \widetilde{\mathcal{L}^\eta_\tau}(x,x',\mathbf{g})\| \leq \frac{\lambda_G}{2}$ then we have

$$\|(I - \tilde{\mathcal{U}}\tilde{\mathcal{U}}^\top)\tilde{U}\tilde{U}^\top\| \leq \frac{2}{\lambda_G}\|\widetilde{L^\eta_\tau} - \widetilde{\mathcal{L}^\eta_\tau}(x,x',\mathbf{g})\|.$$

If, on the other hand, $\|\widetilde{L^\eta_\tau} - \widetilde{\mathcal{L}^\eta_\tau}(x,x',\mathbf{g})\| > \frac{\lambda_G}{2}$ then we directly have

$$\|(I - \tilde{\mathcal{U}}\tilde{\mathcal{U}}^\top)\tilde{U}\tilde{U}^\top\| \leq 1 \leq \frac{2}{\lambda_G}\|\widetilde{L^\eta_\tau} - \widetilde{\mathcal{L}^\eta_\tau}(x,x',\mathbf{g})\|$$

again. Now, noting $\operatorname{rank}(\hat{U})$ is at most G it holds that

$$\begin{aligned}\|(I-\tilde{\mathcal{U}}\tilde{\mathcal{U}}^\top)\tilde{U}\tilde{U}^\top\| &\geq \frac{1}{\sqrt{G}}\|(I-\tilde{\mathcal{U}}\tilde{\mathcal{U}}^\top)\tilde{U}\tilde{U}^\top\|_F \\ &\geq \frac{1}{\sqrt{2G}}\|\tilde{U}\tilde{U}^\top - \tilde{\mathcal{U}}\tilde{\mathcal{U}}^\top\|_F.\end{aligned}$$

by Proposition 2.1 in Vu and Lei (2013). In sum, we have

$$\|\tilde{U}\tilde{U}^\top - \tilde{\mathcal{U}}\tilde{\mathcal{U}}^\top\|_F \leq \frac{2\sqrt{2G}}{\lambda_G}\|\widetilde{L^\eta_\tau} - \widetilde{\mathcal{L}^\eta_\tau}(x,x',\mathbf{g})\|$$

as desired. For the second assertion, as in Rohe, Qin, and Yu (2016) we note that for some $G\times G$ orthogonal matrix Q_U

$$\begin{aligned}\|\tilde{U}\tilde{U}^\top - \tilde{\mathcal{U}}\tilde{\mathcal{U}}^\top\|_F &\geq \frac{1}{2}\|UU^\top - \mathcal{U}\mathcal{U}^\top\|_F \\ &\geq \frac{1}{\sqrt{2}}\|\sin\Theta(\operatorname{col}(U),\operatorname{col}(\mathcal{U}))\|_F \\ &\geq \frac{1}{2}\|U-\mathcal{U}Q_U\|_F\end{aligned}$$

where the second inequality follows from Proposition 2.1 in Vu and Lei (2013) and the third follows from Proposition 2.2 in Vu and Lei (2013). We now have

$$\|U-\mathcal{U}Q_U\|_F \leq \frac{4\sqrt{2G}}{\lambda_G}\|\widetilde{L^\eta_\tau} - \widetilde{\mathcal{L}^\eta_\tau}(x,x',\mathbf{g})\|.$$

A similar argument shows the last inequality. □

We now bound misclassification rates. Let

$$\bar{U} = \widehat{\Theta}^\eta(x)Z_U, \quad \bar{V} = \widehat{\Theta}^\eta(x')Z_V,$$

then

$$\begin{aligned}\|\bar{U}-\mathcal{U}Q_U\|_F^2 &\leq 2\|\bar{U}-U\|_F^2 + 2\|U-\mathcal{U}Q_U\|_F^2 \\ &\leq 2\|\mathcal{U}Q_U - U\|_F^2 + 2\|U-\mathcal{U}Q_U\|_F^2 \\ &= 4\|U-\mathcal{U}Q_U\|_F^2\end{aligned}$$

and likewise

$$\|\bar{V}-\mathcal{V}Q_V\|_F^2 \leq 4\|V-\mathcal{V}Q_V\|_F^2.$$

Define

$$\mathcal{Z}_U' = \mathcal{Z}_U Q_U, \quad \mathcal{Z}_V' = \mathcal{Z}_V Q_V,$$

Note that for $g \in [G]$, by (3.18)

$$\begin{aligned}\min_{\ell \neq g} \|\mathcal{Z}'_{U_{\ell\cdot}} - \mathcal{Z}'_{U_{g\cdot}}\| &= \min_{\ell \neq g} \|\mathcal{Z}_{U_{\ell\cdot}} - \mathcal{Z}_{U_{g\cdot}}\| \\ &= \min_{\ell \neq g} \sqrt{\frac{1}{n_g(x)} + \frac{1}{n_\ell(x)}} \\ &= \sqrt{\frac{1}{n_g(x)} + \frac{1}{\max\{n_\ell(x) : \ell \neq g\}}}.\end{aligned} \tag{3.19}$$

Moreover, let

$$\mathcal{S}_{gN}(x) := \left\{ i \in \mathcal{G}_{gN}(x) : \|\bar{U}_{i\cdot} - (\mathcal{U}Q_U)_{i\cdot}\| \geq \frac{1}{2}\sqrt{\frac{1}{n_g(x)} + \frac{1}{\max\{n_\ell(x) : \ell \neq g\}}} \right\}.$$

Then

$$\begin{aligned}\|\bar{U} - \mathcal{U}Q_U\|_F^2 &= \sum_{g=1}^{G} \frac{1}{4}\left[\frac{1}{n_g(x)} + \frac{1}{\max\{n_\ell(x) : \ell \neq g\}}\right]\left(\frac{4}{\frac{1}{n_g(x)} + \frac{1}{\max\{n_\ell(x):\ell\neq g\}}} \sum_{i \in \mathcal{G}_{gN}(x)} \|\bar{U}_{i\cdot} - (\mathcal{U}Q_U)_{i\cdot}\|^2\right) \\ &\geq \sum_{g=1}^{G} \frac{1}{4}\left[\frac{1}{n_g(x)} + \frac{1}{\max\{n_\ell(x) : \ell \neq g\}}\right] \\ &\qquad \left(\sum_{i \in \mathcal{G}_{gN}(x)} \mathbf{1}\left\{\|\bar{U}_{i\cdot} - (\mathcal{U}Q_U)_{i\cdot}\| \geq \frac{1}{2}\sqrt{\frac{1}{n_g(x)} + \frac{1}{\max\{n_\ell(x) : \ell \neq g\}}}\right\}\right) \\ &= \sum_{g=1}^{G} |\mathcal{S}_{gN}(x)| \frac{1}{4}\left[\frac{1}{n_g(x)} + \frac{1}{\max\{n_\ell(x) : \ell \neq g\}}\right].\end{aligned}$$

Therefore we have

$$\begin{aligned}\sum_{g=1}^{G} \frac{|\mathcal{S}_{gN}(x)|}{n_g(x)} &\leq \sum_{g=1}^{G} |\mathcal{S}_{gN}(x)| \left(\frac{1}{n_g(x)} + \frac{1}{\max\{n_\ell(x) : \ell \neq g\}}\right) \\ &\leq 4\|\bar{U} - \mathcal{U}Q_U\|_F^2 \\ &\leq 16\|U - \mathcal{U}Q_U\|_F^2.\end{aligned}$$

By Lemma 3.5 it follows that

$$\sum_{g=1}^{G} \frac{|\mathcal{S}_{gN}(x)|}{n_g(x)} \leq \frac{512G}{\lambda_G(x,x')^2}\|\widetilde{L^\eta_\tau} - \widetilde{\mathcal{L}^\eta_\tau}(x,x',\mathbf{g})\|^2. \tag{3.20}$$

The same holds with $\frac{|\mathcal{S}_{hN}(x')|}{n_h(x')}$. Moreover, one can show that following the proof of Lemma 5.3 of Lei and Rinaldo (2015), the definition of $S_{gN}(x)$, in view of (3.19), guarantees that for each $g \in [G]$, under the condition

$$\frac{16\sqrt{2G}}{\lambda_G(x,x')}\|\widetilde{L^{\eta}_{\tau}} - \widetilde{\mathcal{L}^{\eta}_{\tau}}(x,x',\mathbf{g})\| < 1, \tag{3.21}$$

the estimated membership matrix $\widehat{\Theta}^{\eta}(x)$ assigns the correct membership for every $i \in \mathcal{G}_{gN}(x)\backslash\mathcal{S}_{gN}(x)$. That is, for $\mathcal{G}_N(x) = \cup_{g\in[G]}(\mathcal{G}_{gN}(x) \setminus \mathcal{S}_{gN}(x))$, there exists a permutation matrix $J(x)$ such that $(\widehat{\Theta}\eta(x))_{\mathcal{G}_N(x)\cdot}J(x) = (\Theta^{\eta}(x))_{\mathcal{G}_N(x)\cdot}$ and the same holds for x'.

By (3.16) and (3.17), the singular values of $\mathcal{L}^{\eta}(x,x',\mathbf{g})$ are that of $\mathcal{N}(x)\overline{\mathcal{O}}^{-1/2}B(x,x')\overline{\mathcal{Q}}^{-1/2}\mathcal{N}(x')$. Denote $\sigma_1(M) \geq ... \geq \sigma_r(M)$ the singular values of a matrix M of rank r. We use Theorem 3.3.16 of Horn and Johnson (1990), see also Wang and Xi (1997), and write

$$\begin{aligned}
\lambda_G(x,x') &= \sigma_G\left(\mathcal{N}(x)\overline{\mathcal{O}}^{-1/2}B(x,x')\overline{\mathcal{Q}}^{-1/2}\mathcal{N}(x')\right)\\
&\geq \frac{\sigma_G\left(\mathcal{N}(x)\overline{\mathcal{O}}^{-1/2}B(x,x')\right)}{\sigma_1\left(\left[\overline{\mathcal{Q}}^{-1/2}\mathcal{N}(x')\right]^{-1}\right)} = \frac{\sigma_G\left(\mathcal{N}(x)\overline{\mathcal{O}}^{-1/2}B(x,x')\right)}{\sigma_1\left(\mathcal{N}(x')^{-1}\overline{\mathcal{Q}}^{1/2}\right)}\\
&\geq \frac{\sigma_G\left(B(x,x')\right)}{\sigma_1\left(\mathcal{N}(x')^{-1}\overline{\mathcal{Q}}^{1/2}\right)\sigma_1\left(\left[\mathcal{N}(x)\overline{\mathcal{O}}^{-1/2}\right]^{-1}\right)} = \frac{\sigma_G\left(B(x,x')\right)}{\sigma_1\left(\mathcal{N}(x')^{-1}\overline{\mathcal{Q}}^{1/2}\right)\sigma_1\left(\mathcal{N}(x)^{-1}\overline{\mathcal{O}}^{1/2}\right)}
\end{aligned}$$

where we recall that the matrices $\mathcal{N}(x)$, $\overline{\mathcal{O}}$ and $\overline{\mathcal{Q}}$ are diagonal. Their singular values are equal to their diagonal coefficients, and

$$\left(\mathcal{N}(x)^{-1}\overline{\mathcal{O}}^{1/2}\right)^2_{gg} = \left(\sum_{j\in\eta_N(x')} B_{gg(j)}(x,x') + \tau\right)/n_g(x) \leq \left(\|B(x,x')\|_{\max}k + \tau\right)/\min_h \underline{n_h}.$$

This implies that

$$\lambda_G(x,x') \geq \frac{\sigma_G\left(B(x,x')\right)\min_h \underline{n_h}}{\|B(x,x')\|_{\max}k + \tau}. \tag{3.22}$$

We plug in (3.22) and the bounds obtained on the relevant quantities in (3.21). This condition becomes

$$\begin{aligned}
&4\sqrt{\frac{3\ln(8k/\tilde{\delta})}{\Delta \min_{h\in[G]}\left\lfloor \frac{c}{16}\frac{N_h}{N}\frac{b_X}{\overline{U}_X}\ k\right\rfloor + \tau}} + \frac{2kl_BR_k}{\Delta \min_{h\in[G]}\left\lfloor \frac{c}{16}\frac{N_h}{N}\frac{b_X}{\overline{U}_X}\ k\right\rfloor + \tau}\left(\frac{2kl_BR_k}{\Delta \min_{h\in[G]}\left\lfloor \frac{c}{16}\frac{N_h}{N}\frac{b_X}{\overline{U}_X}\ k\right\rfloor + \tau} + 3\right)\\
&\quad < \ 16\sqrt{2G}\sigma_G\left(B(x,x')\right)\frac{\min_{h\in[G]}\left\lfloor \frac{c}{16}\frac{N_h}{N}\frac{b_X}{\overline{U}_X}\ k\right\rfloor}{\|B(x,x')\|_{\max}k + \tau}.
\end{aligned} \tag{3.23}$$

We combine (3.20) and (3.23) to obtain a finite sample probability bound on $\sum_{g=1}^{G}|\mathcal{S}_{gN}(x)|/n_g(x)$ in the following Lemma.

Lemma 3.6. *Let Assumptions 2.1, 2.2, 2.3 and 2.4 hold. Then for all N, $\delta \in (0,1)$, $1 \leq k \leq N$ such that*

(1) $\Delta \min_{h\in[G]} \left\lfloor \frac{c}{16} \frac{N_h}{N} \frac{b_X}{\overline{U}_X}\ k \right\rfloor + \tau \geq 3\ln(24k/\delta)$,

(2) $k \geq \max(12d\ln(72GN/\delta), 24d\ln(36N/\delta))$,

(3) $k \leq \min(8T^dV_d\overline{U}_XN, (1/2)T^dV_db_XcN)$,

(4) and for all $h \in [G]$, $\frac{c}{16}\frac{N_h}{N}\frac{b_X}{\overline{U}_X}\ k \geq 24d\ln(72GN_h/\delta) + 1$,

(5) (3.23) *holds for* $\tilde{\delta} = \delta/3$

with probability at least $1-\delta$ *conditional on* $\mathbf{g}$, *it holds that*

$$
\begin{aligned}
\sum_{g=1}^{G} \frac{|\mathcal{S}_{gN}(x)|}{n_g(x)} \leq & \frac{512G\left[\|B(x,x')\|_{\max}k+\tau\right]^2}{\sigma_G\left(B(x,x')\right)^2 \min_{h\in[G]}\left\lfloor \frac{c}{16}\frac{N_h}{N}\frac{b_X}{\overline{U}_X}\ k\right\rfloor^2} \left[4\sqrt{\frac{3\ln(24k/\delta)}{\Delta\min_{h\in[G]}\left\lfloor \frac{c}{16}\frac{N_h}{N}\frac{b_X}{\overline{U}_X}\ k\right\rfloor+\tau}}\right. \\
&\left. + \frac{2kl_BR_k}{\Delta\min_{h\in[G]}\left\lfloor \frac{c}{16}\frac{N_h}{N}\frac{b_X}{\overline{U}_X}\ k\right\rfloor+\tau}\left(\frac{2kl_BR_k}{\Delta\min_{h\in[G]}\left\lfloor \frac{c}{16}\frac{N_h}{N}\frac{b_X}{\overline{U}_X}\ k\right\rfloor+\tau}+3\right)\right]^2, \qquad (3.24)
\end{aligned}
$$

and that there exist permutation matrices $J(x)$ *and* $J(x')$ *such that* $(\widehat{\Theta}^{\eta}(x))_{\mathcal{G}_N(x)\cdot}J(x) = (\Theta^{\eta}(x))_{\mathcal{G}_N(x)\cdot}$ *and the same holds for* x'.

Proof. We use the events $V(\mathbf{g},\mathbf{x})$ and $W(\mathbf{g},\mathbf{x})$ as well as the function C defined in the proof of Lemma 3.4. Let $\tilde{C}$ be the function such that $\tilde{C}(d_{\min})$ is the right hand side of (3.24).

$$
\begin{aligned}
&\Pr\left[\sum_{g=1}^{G}\frac{|\mathcal{S}_{gN}(x)|}{n_g(x)} \leq \tilde{C}\left(\Delta\min_{h\in[G]}\left\lfloor\frac{c}{16}\frac{N_h}{N}\frac{b_X}{\overline{U}_X}\ k\right\rfloor\right)\ \Big|\, \mathbf{g}\right] \\
&\geq \Pr\left[\left\{\|\widetilde{L_{\tau}^{\eta}} - \widetilde{\mathcal{L}_{\tau}^{\eta}}(x,x',\mathbf{g})\| \leq C\left(\Delta\min_{h\in[G]}\left\lfloor\frac{c}{16}\frac{N_h}{N}\frac{b_X}{\overline{U}_X}\ k\right\rfloor\right)\right\}\cap V(\mathbf{g},\mathbf{x})\cap W(\mathbf{g},\mathbf{x})\ \Big|\, \mathbf{g}\right]
\end{aligned}
$$

by (3.20) and (3.22). By the proof of Lemma 3.4, the probability on the right hand side is larger than $1-\delta$. ∎

Remark 3.4. Let $\tau = 0$, sparsity be captured by a parameter ρ_N and $N_h/N \approx C$ for all $h \in [G]$. Remark 3.3 explains that $\|\widetilde{L^{\eta}} - \widetilde{\mathcal{L}^{\eta}}(x,x',\mathbf{g})\| \approx (\rho_N N)^{\frac{-1}{d+2}}$, up to $\ln N$ and $\ln\rho_N$ factors. According to Lemma 3.6, this implies that

$$
\sum_{g=1}^{G}\frac{|\mathcal{S}_{gN}(x)|}{n_g(x)} \approx (\rho_N N)^{\frac{-2}{d+2}},
$$

whereas Lei and Rinaldo (2015) obtains a rate of $(\rho_N N)^{-1}$, see Corollary 3.2.

4. Estimation of B and π

As in the previous section, we fix a pair $(x, x') \in S^2$. Recall that $\mathcal{G}_{hN}(x) = \{g(i) = h, i \in \eta_N(x)\}$ and $n_h(x) = |\mathcal{G}_{hN}(x)|$. We introduce the following new notation,

- $\mathcal{G}_{hN}(x) = \{g(i) = h, i \in \eta_N(x)\}$, then $n_h(x) = |\mathcal{G}_{hN}(x)|$
- $\widehat{\mathcal{G}}_{hN}(x) = \{\hat{g}(i) = h, i \in \eta_N(x)\}$, then $\hat{n}_h(x) = |\widehat{\mathcal{G}}_{hN}(x)|$
- $\mathcal{G}_{hN} = \{g(i) = h, i \in [N]\}$, then $N_h = |\mathcal{G}_{hN}|$
- $\mathbf{1}_{h,x}(i) = \mathbf{1}\{i \in \mathcal{G}_{hN}(x)\}$
- $\widehat{\mathbf{1}}_{h,x}(i) = \mathbf{1}\{i \in \widehat{\mathcal{G}}_{hN}(x)\}$
- $\mathbf{1}_x(i) = \mathbf{1}\{i \in \eta_N(x)\}$

Note that the estimators of $\pi_h(x)$ and $B_{gh}(x, x')$ can be written

$$\widehat{\pi}_h(x) = \frac{\hat{n}_h(x)}{k} = \frac{\sum_{i \in \eta_N(x)} \widehat{\mathbf{1}}_{h,x}(i)}{k},$$

$$\widehat{B}_{gh}(x, x') = \frac{1}{\hat{n}_g(x)\hat{n}_h(x')} \sum_{\substack{i \in \widehat{\mathcal{G}}_{gN}(x) \\ j \in \widehat{\mathcal{G}}_{hN}(x')}} A_{ij} = \frac{1}{\hat{n}_g(x)\hat{n}_h(x')} \sum_{i,j} A_{ij} \widehat{\mathbf{1}}_{g,x}(i) \widehat{\mathbf{1}}_{h,x'}(j).$$

Define the oracle estimators for $\pi_g(x)$ and $B_{gh}(x, x')$ as

$$\pi_g^{or}(x) = \frac{n_g(x)}{k} = \frac{\sum_i \mathbf{1}_{g,x}(i)}{k},$$

$$B_{gh}^{or}(x, x') = \frac{1}{n_g(x) n_h(x')} \sum_{i,j} A_{ij} \mathbf{1}_{g,x}(i) \mathbf{1}_{h,x'}(j).$$

In this section, we derive probability bounds on $\widehat{\pi}_h(x)$ and $\widehat{B}_{gh}(x, x')$. Note that according to Lemma 3.6, with probability at least $1 - \delta$ conditional on $\mathbf{g}$, the memberships of $i \in (\mathcal{G}_{gN}(x) \setminus \mathcal{S}_{gN}(x))$ are all correctly estimated up to a permutation. Thus all the following probability bounds hold up to a permutation. We explore this identification issue this raises in Remark 4.1.

4.1. Result conditional on g.

We can decompose

$$|\widehat{B}_{gh}(x, x') - B_{gh}^{or}(x, x')| \leq \underbrace{\left| \frac{1}{\hat{n}_g(x)\hat{n}_h(x')} - \frac{1}{n_g(x) n_h(x')} \right| \sum_{i,j} \widehat{\mathbf{1}}_{g,x}(i) \widehat{\mathbf{1}}_{h,x'}(j)}_{T_1}$$
$$+ \underbrace{\frac{1}{n_g(x) n_h(x')} \sum_{i,j} \left| \widehat{\mathbf{1}}_{g,x}(i) \widehat{\mathbf{1}}_{h,x'}(j) - \mathbf{1}_{g,x}(i) \mathbf{1}_{h,x'}(j) \right|}_{T_2},$$

where by a slight abuse of notation the summations are over $i \in \eta_N(x)$ and $j \in \eta_N(x')$. We look at T_1 and T_2 separately.

$$T_1 = \left|1 - \frac{\hat{n}_g(x)\hat{n}_h(x')}{n_g(x)n_h(x')}\right| = \frac{|n_g(x)n_h(x') - \hat{n}_g(x)\hat{n}_h(x')|}{n_g(x)n_h(x')},$$
$$T_2 \le \underbrace{\frac{1}{n_g(x)n_h(x')}\sum_{i,j}\left|\widehat{\mathbf{1}}_{g,x}(i) - \mathbf{1}_{g,x}(i)\right|\widehat{\mathbf{1}}_{h,x'}(j)}_{T_{21}} + \underbrace{\frac{1}{n_g(x)n_h(x')}\sum_{i,j}\mathbf{1}_{g,x}(i)\left|\widehat{\mathbf{1}}_{h,x'}(j) - \mathbf{1}_{h,x'}(j)\right|}_{T_{22}}.$$

When (3.21) holds,

$$\begin{aligned} T_{21} &= \frac{\hat{n}_h(x')}{n_h(x')}\frac{1}{\hat{n}_h(x')}\sum_j \widehat{\mathbf{1}}_{h,x'}(j)\frac{1}{n_g(x)}\sum_i\left|\widehat{\mathbf{1}}_{g,x}(i) - \mathbf{1}_{g,x}(i)\right| \\ &\le \frac{\hat{n}_h(x')}{n_h(x')}\frac{1}{n_g(x)}\sum_{l\in[G]}|\mathcal{S}_{lN}(x)|, \end{aligned}$$

where the second inequality holds by the argument used in (4.7). Similarly, when (3.21) holds we obtain

$$T_{22} \le \frac{1}{n_h(x')}\sum_{l\in[G]}|\mathcal{S}_{lN}(x')|.$$

Note that when (3.21) holds,

$$\begin{aligned} T_1 &\le \frac{|n_g(x) - \hat{n}_g(x)|\, n_h(x') + \hat{n}_g(x)\,|\hat{n}_h(x') - n_h(x')|}{n_g(x)n_h(x')} \\ &\le \frac{1}{n_g(x)}\sum_{l\in[G]}|\mathcal{S}_{lN}(x)| + \frac{\hat{n}_g(x)}{n_g(x)}\frac{1}{n_h(x')}\sum_{l\in[G]}|\mathcal{S}_{lN}(x')| \\ &\le \left[\frac{k}{\min_{h\in[G]}\underline{n_h}} + \frac{k^2}{\left(\min_{h\in[G]}\underline{n_h}\right)^2}\right]\sum_{l\in[G]}\frac{|\mathcal{S}_{lN}(x)|}{n_l(x)}, \end{aligned}$$

where we used the same argument as in (4.7) in the second inequality. We use similar arguments to bound T_{21} and T_{22} and obtain

$$|\widehat{B}_{gh}(x,x') - B^{or}_{gh}(x,x')| \le 2\left[\frac{k}{\min_{h\in[G]}\underline{n_h}} + \frac{k^2}{\left(\min_{h\in[G]}\underline{n_h}\right)^2}\right]\sum_{l\in[G]}\frac{|\mathcal{S}_{lN}(x)|}{n_l(x)}. \tag{4.1}$$

We now bound $|B^{or}_{gh}(x,x') - B_{gh}(x,x')|$. The oracle estimator is not a standard k-nearest neighbor estimator with, say, $k = n_g(x)n_h(x')$, because $n_g(x)$ and $n_h(x')$ are not chosen by the statistician but random. Moreover, the neighborhood for x is chosen separately from that of x'. It is also not a Nadaraya-Watson estimator with uniform kernel as $r^g_{n_g(x)}(x)$ and $r^h_{n_h(x')}(x')$ are random. Write

$$A_{ij} = B_{g(i)g(j)}(x(i), x(j)) + \zeta_{ij},$$

then we can decompose

$$|B^{or}_{gh}(x,x') - B_{gh}(x,x')| \leq \underbrace{\frac{1}{n_g(x)n_h(x')} \sum_{i,j} \big| \left[B_{gh}(x(i),x(j)) - B_{gh}(x,x')\right] \big| \mathbf{1}_{g,x}(i)\mathbf{1}_{h,x'}(j)}_{T_3}$$

$$+ \underbrace{\frac{1}{n_g(x)n_h(x')} \big| \sum_{i,j} \zeta_{ij}\mathbf{1}_{g,x}(i)\mathbf{1}_{h,x'}(j)\big|}_{T_4}. \tag{4.2}$$

In this decomposition, T_3 is a bias term and T_4 is a variance term. As in the proof of Lemma 3.1, the bias term can be bounded using Assumption 2.4,

$$T_3 \leq l_B[r_k(x) + r_k(x')] \tag{4.3}$$

by $r^g_{n_g(x)}(x) \leq r_k(x)$ and $r^g_{n_h(x')}(x') \leq r_k(x')$. Note that the relevant regressor dimension is d. As for the variance term T_4, note that we can rewrite

$$T_4 = \frac{1}{n_g(x)n_h(x')} \big| \sum_{\substack{i\in\mathcal{G}_{gN}(x)\\ j\in\mathcal{G}_{hN}(x')}} \zeta_{ij}\big|,$$

where conditional on $(\mathbf{x},\mathbf{g})$, $\{\zeta_{ij},\ i \in \mathcal{G}_{gN}(x),\ j \in \mathcal{G}_{hN}(x')\}$ are independent mean-zero bounded random variables. We apply a Hoeffding inequality for bounded random variables, see Theorem 2.2.6 in Vershynin (2018), and obtain for any $a_0 \geq 0$,

$$\Pr\left(\big|\sum_{i,j} \zeta_{ij}\mathbf{1}_{g,x}(i)\mathbf{1}_{h,x'}(j)\big| \geq a_0 \middle| \mathbf{x},\mathbf{g}\right) \leq 2\exp\left(-a_0^2/[2n_g(x)n_h(x')]\right), \tag{4.4}$$

where we used that $\zeta_{ij} \in [-1,1]$. Taking $a_0 = \sqrt{2n_g(x)n_h(x')\ln(2/\tilde{\delta})}$ implies

$$\Pr\left(T_4 \leq \frac{\sqrt{2\ln(2/\tilde{\delta})}}{\min_{h\in[G]} \underline{n_h}} \middle| \mathbf{x},\mathbf{g}\right) \geq 1-\tilde{\delta}. \tag{4.5}$$

We combine (4.1) with (4.3) and (4.5) to obtain the following Lemma.

Lemma 4.1. *Let Assumptions 2.1, 2.2, 2.3 and 2.4 hold. Then for all N, $\delta \in (0,1)$, $1 \leq k \leq N$ such that*

(1) $\Delta \min_{h\in[G]} \left\lfloor \frac{c}{16}\frac{N_h}{N}\frac{b_X}{\overline{U}_X}\ k\right\rfloor + \tau \geq 3\ln(48k/\delta)$,
(2) $k \geq \max(12d\ln(144GN/\delta), 24d\ln(72N/\delta))$,
(3) $k \leq \min(8T^dV_d\overline{U}_XN, (1/2)T^dV_db_XcN)$,
(4) and for all $h \in [G]$, $\frac{c}{16}\frac{N_h}{N}\frac{b_X}{\overline{U}_X}\ k \geq 24d\ln(144GN_h/\delta)+1$,

(5) (3.23) *holds for* $\tilde{\delta} = \delta/6$,

with probability at least $1-\delta$ *conditional on* $\mathbf{g}$*, it holds that*

$$
\begin{aligned}
&|\widehat{B}_{gh}(x,x') - B_{gh}(x,x')| \\
&\le \frac{1024G\left[\|B(x,x')\|_{\max}k+\tau\right]^2}{\sigma_G\left(B(x,x')\right)^2 \min_{h\in[G]}\left\lfloor \frac{c}{16}\frac{N_h}{N}\frac{b_X}{\overline{U}_X}\ k\right\rfloor^2}\left[\frac{k}{\min_{h\in[G]}\left\lfloor \frac{c}{16}\frac{N_h}{N}\frac{b_X}{\overline{U}_X}\ k\right\rfloor} + \frac{k^2}{\left(\min_{h\in[G]}\left\lfloor \frac{c}{16}\frac{N_h}{N}\frac{b_X}{\overline{U}_X}\ k\right\rfloor\right)^2}\right]\times \\
&\left[4\sqrt{\frac{3\ln(48k/\delta)}{\Delta\min_{h\in[G]}\left\lfloor \frac{c}{16}\frac{N_h}{N}\frac{b_X}{\overline{U}_X}\ k\right\rfloor+\tau}} + \frac{2kl_BR_k}{\Delta\min_{h\in[G]}\left\lfloor \frac{c}{16}\frac{N_h}{N}\frac{b_X}{\overline{U}_X}\ k\right\rfloor+\tau}\left(\frac{2kl_BR_k}{\Delta\min_{h\in[G]}\left\lfloor \frac{c}{16}\frac{N_h}{N}\frac{b_X}{\overline{U}_X}\ k\right\rfloor+\tau}+3\right)\right]^2
\end{aligned}
$$

(4.6)

$$
+\,2l_BR_k + \frac{\sqrt{2\ln(4/\delta)}}{\min_{h\in[G]}\left\lfloor \frac{c}{16}\frac{N_h}{N}\frac{b_X}{\overline{U}_X}\ k\right\rfloor}.
$$

Proof. We use once more the definitions introduced in the proof of Lemma 3.4.

$$
\begin{aligned}
&\Pr\Bigg[\left\{\|\widetilde{L^\eta_\tau} - \widetilde{\mathcal{L}^\eta_\tau}(x,x',\mathbf{g})\| \le C\left(\Delta\min_{h\in[G]}\left\lfloor \frac{c}{16}\frac{N_h}{N}\frac{b_X}{\overline{U}_X}\ k\right\rfloor\right)\right\} \\
&\qquad\cap\left\{T_4 \le \frac{\sqrt{2\ln(4/\delta)}}{\min_{h\in[G]}\left\lfloor \frac{c}{16}\frac{N_h}{N}\frac{b_X}{\overline{U}_X}\ k\right\rfloor}\right\}\cap V(\mathbf{g},\mathbf{x})\cap W(\mathbf{g},\mathbf{x})\,\Big|\,\mathbf{g}\Bigg] \\
&\ge \Pr\left[\left\{\|\widetilde{L^\eta_\tau} - \widetilde{\mathcal{L}^\eta_\tau}(x,x',\mathbf{g})\| \le C\left(d_{\min}\right)\right\}\cap\left\{T_4 \le \frac{\sqrt{2\ln(4/\delta)}}{\min_{h\in[G]}\underline{n_h}}\right\}\cap V(\mathbf{g},\mathbf{x})\cap W(\mathbf{g},\mathbf{x})\,\Big|\,\mathbf{g}\right] \\
&\ge \left[\mathbb{E}\left[\Pr\left[\left\{\|\widetilde{L^\eta_\tau} - \widetilde{\mathcal{L}^\eta_\tau}(x,x',\mathbf{g})\| \le C\left(d_{\min}\right)\right\}\,\Big|\,\mathbf{g},\mathbf{x},V(\mathbf{g},\mathbf{x}),W(\mathbf{g},\mathbf{x})\right]\,\Big|\,\mathbf{g},V(\mathbf{g},\mathbf{x}),W(\mathbf{g},\mathbf{x})\right]\right. \\
&\qquad\times\Pr\left[V(\mathbf{g},\mathbf{x})\cap W(\mathbf{g},\mathbf{x})\,|\,\mathbf{g}\right] + \mathbb{E}\left[\Pr\left[\left\{T_4 \le \frac{\sqrt{2\ln(4/\delta)}}{\min_{h\in[G]}\underline{n_h}}\right\}\,\Big|\,\mathbf{g},\mathbf{x}\right]\,\Big|\,\mathbf{g}\right]\;-\;1 \\
&\ge 1-\delta/2+1-\delta/2-1 = 1-\delta
\end{aligned}
$$

where the last inequality holds by (4.5) and the proof of Lemma 3.4. The result holds by

$$
\left\{\|\widetilde{L^\eta_\tau} - \widetilde{\mathcal{L}^\eta_\tau}(x,x',\mathbf{g})\| \le C\left(\Delta\min_{h\in[G]}\left\lfloor \frac{c}{16}\frac{N_h}{N}\frac{b_X}{\overline{U}_X}\ k\right\rfloor\right)\right\}\cap V(\mathbf{g},\mathbf{x})\cap W(\mathbf{g},\mathbf{x})
$$

$$\Rightarrow \begin{cases} \sum_{l\in[G]} \frac{|\mathcal{S}_{lN}(x)|}{n_l(x)} \leq \epsilon_N(\delta/6), \\ |\widehat{B}_{gh}(x,x') - B^{or}_{gh}(x,x')| \leq 2\left[\frac{k}{\min_{h\in[G]}\left\lfloor \frac{c}{16}\frac{N_h}{N}\frac{b_X}{\bar{U}_X}\; k\right\rfloor} + \frac{k^2}{\left(\min_{h\in[G]}\left\lfloor \frac{c}{16}\frac{N_h}{N}\frac{b_X}{\bar{U}_X}\; k\right\rfloor\right)^2}\right]\epsilon_N(\delta/6), \\ T_3 \leq 2 l_B R_k, \end{cases}$$

where the first line holds as in Lemma 3.6 under condition (5), the second holds by (4.1), and the third by (4.3). □

4.2. **Results unconditional on g.**

4.2.1. *Result on the community assignment probabilities.*

The following results hold up to a permutation but for ease of readability we let $J(x) = J(x') = I$. Note that when (3.21) holds,

$$\begin{aligned} |\widehat{\pi}_g(x) - \pi^{or}_g(x)| &\leq \frac{1}{k}\sum_i \left|\widehat{\mathbf{1}}_{g,x}(i) - \mathbf{1}_{g,x}(i)\right| \\ &\leq \frac{1}{k}\sum_{\substack{h\in[G]\\ h\neq g}}\sum_{i\in\mathcal{G}_{hN}(x)} \widehat{\mathbf{1}}_{g,x}(i) + \frac{1}{k}\sum_{i\in\mathcal{G}_{gN}(x)}\left|\widehat{\mathbf{1}}_{g,x}(i) - 1\right| \qquad (4.7) \\ &\leq \frac{1}{k}\sum_{h\in[G]} |\mathcal{S}_{hN}(x)| \leq \sum_{h\in[G]} \frac{|\mathcal{S}_{hN}(x)|}{n_h(x)} \qquad (4.8) \end{aligned}$$

The behavior of $\pi^{or}_g(x)$ is given by Theorem 1 of Jiang (2019). Write

$$\mathbf{1}_{g,x}(i) = \pi_g(x) + \xi_i$$

with $\mathbb{E}(\xi_i|x(i) = x) = 0$. Assumptions 1-3 of Jiang (2019) hold by Assumptions 2.1, 2.2 and taking the sub-gaussian parameter to be 1 since $\xi_i \in [-1, 1]$ almost surely, see Exercise 2.4 in **?**. We note that the imposed assumption of independence between ξ and x is not needed for his Theorem 1[1]. Assume also that

$$2^8 d\ln(4/\tilde{\delta})\ln N \leq k \leq cV_d b_X T^d N/2, \qquad (4.9)$$

then under Assumption 2.5, Theorem 1 of Jiang (2019) implies that the following holds with probability at least $1 - \tilde{\delta}$,

$$|\pi^{or}_g(x) - \pi_g(x)| \leq l_\pi R_k + 2\sqrt{\frac{d\ln N + \ln(2/\tilde{\delta})}{k}}. \qquad (4.10)$$

[1]The application of Hoeffding's inequality in the Proof of Theorem 1 (see p4004) can be done conditional on $\mathbf{x}$ as long as $(x(i), \xi_i)$ for $i = 1...n$ is i.i.d.

Let $\pi_g := \Pr(g(i) = g) = \mathbb{E}(\pi_g(x(i)))$ and $\underline{\pi} := \min_{g\in[G]} \pi_g$. We combine Equations (4.8) and (4.10) to obtain a probability bound unconditional on $\mathbf{g}$ in the following Lemma, where (3.23) is replaced with

$$4\sqrt{\frac{3\ln(24k/\tilde{\delta})}{\Delta\left\lfloor \frac{\pi c b_X}{32\overline{U}_X}\ k\right\rfloor + \tau}} + \frac{2kl_BR_k}{\Delta\left\lfloor \frac{\pi c b_X}{32\overline{U}_X}\ k\right\rfloor + \tau}\left(\frac{2kl_BR_k}{\Delta\left\lfloor \frac{\pi c b_X}{32\overline{U}_X}\ k\right\rfloor + \tau} + 3\right) < \frac{16\sqrt{2G}\sigma_G\left(B(x,x')\right)\left\lfloor \frac{\pi c b_X}{32\overline{U}_X}\ k\right\rfloor}{\|B(x,x')\|_{\max}k + \tau}. \tag{4.11}$$

Lemma 4.2. *Let Assumptions 2.1, 2.2, 2.3, 2.4 and 2.5 hold. Assume moreover that* $\underline{\pi} > 0$. *Then for all* N, $\delta \in (0,1)$, $1 \le k \le N$ *such that*

(1) $\Delta\left\lfloor \frac{\pi c b_X}{32\overline{U}_X}\ k\right\rfloor + \tau \ge 3\ln(72k/\delta)$,
(2) $k \ge \max(12d\ln(216GN/\delta), 24d\ln(108N/\delta))$,
(3) $k \le \min(8T^dV_d\overline{U}_XN, (1/2)T^dV_db_XcN)$,
(4) $\frac{\pi c b_X}{32\overline{U}_X}\ k \ge 24d\ln(216GN/\delta) + 1$,
(5) (4.11) *holds for* $\tilde{\delta} = \delta/3$,
(6) $2^8d\ln(24/\delta)\ln N \le k \le cV_db_XT^dN/2$,
(7) $N \ge 8\ln(3G/\delta)/\underline{\pi}^2$,

with probability at least $1-\delta$, *it holds that*

$$|\widehat{\pi}_g(x) - \pi_g(x)| \le \frac{512G\left[\|B(x,x')\|_{\max}k + \tau\right]^2}{\sigma_G\left(B(x,x')\right)^2\left\lfloor \frac{\pi c b_X}{32\overline{U}_X}\ k\right\rfloor^2}\left[4\sqrt{\frac{3\ln(72k/\delta)}{\Delta\left\lfloor \frac{\pi c b_X}{32\overline{U}_X}\ k\right\rfloor + \tau}} + \frac{2kl_BR_k}{\Delta\left\lfloor \frac{\pi c b_X}{32\overline{U}_X}\ k\right\rfloor + \tau}\right.$$
$$\left.\times\left(\frac{2kl_BR_k}{\Delta\left\lfloor \frac{\pi c b_X}{32\overline{U}_X}\ k\right\rfloor + \tau} + 3\right)\right]^2 + l_\pi R_k + 2\sqrt{\frac{d\ln N + \ln(6/\delta)}{k}}. \tag{4.12}$$

Proof. For ease of readability, we denote with $\epsilon_N(\delta)$ the bound on the right hand side of (3.24). We apply a Hoeffding inequality for bounded random variables, see Theorem 2.2.6 in Vershynin (2018), and obtain,

$$\Pr\left(N_g - \pi_gN \ge -\pi_gN/2\right) = \Pr\left(\sum_{i\in[N]}\mathbf{1}\{g(i) = g\} - \pi_g \ge -\pi_gN/2\right)$$
$$\le \exp\left(-\pi_g^2N/8\right) \le \exp\left(-\underline{\pi}^2N/8\right).$$

Define the event $T(\mathbf{g}) = \{\forall\, g \in [G],\ N \ge N_g \ge \underline{\pi}N/2\}$. Then

$$\Pr\left(T(\mathbf{g})\right) \ge 1 - G\exp\left(-\underline{\pi}^2N/8\right) \ge 1 - \delta/3,$$

by Condition (7). Note that on $T(\mathbf{g})$, Conditions (1) - (5) of Lemma 3.6 hold. Thus,

$$\Pr\left(\sum_{h\in[G]}\frac{|\mathcal{S}_{hN}(x)|}{n_h(x)}\le\epsilon_N(\delta/3)\middle|\mathbf{g},T(\mathbf{g})\right)\ge 1-\delta/3.$$

Moreover on $T(\mathbf{g})$, note that the upper bound on $\epsilon_N(\delta/3)$ is bounded above by

$$\frac{512G\left[\|B(x,x')\|_{\max}k+\tau\right]^2}{\sigma_G\left(B(x,x')\right)^2\left\lfloor\frac{\underline{\pi}cb_X}{32\overline{U}_X}\ k\right\rfloor^2}\left[4\sqrt{\frac{3\ln(72k/\delta)}{\Delta\left\lfloor\frac{\underline{\pi}cb_X}{32\overline{U}_X}\ k\right\rfloor+\tau}}+\frac{2kl_BR_k}{\Delta\left\lfloor\frac{\underline{\pi}cb_X}{32\overline{U}_X}\ k\right\rfloor+\tau}\left(\frac{2kl_BR_k}{\Delta\left\lfloor\frac{\underline{\pi}cb_X}{32\overline{U}_X}\ k\right\rfloor+\tau}+3\right)\right]^2.$$

The result follows by Theorem 1 of Jiang (2019). □

4.2.2. *Result on the connection probabilities.*

We integrate Lemma 4.1 with respect to $\mathbf{g}$ on $T(\mathbf{g})$ and obtain the following Lemma.

Lemma 4.3. *Let Assumptions 2.1, 2.2, 2.3 and 2.4 hold. Assume moreover that* $\underline{\pi}>0$. *Then for all* N, $\delta\in(0,1)$, $1\le k\le N$ *such that*

(1) $\Delta\left\lfloor\frac{\underline{\pi}cb_X}{32\overline{U}_X}\ k\right\rfloor+\tau\ge 3\ln(96k/\delta)$,
(2) $k\ge\max(12d\ln(288GN/\delta),24d\ln(144N/\delta))$,
(3) $k\le\min(8T^dV_d\overline{U}_XN,(1/2)T^dV_db_XcN)$,
(4) $\frac{\underline{\pi}cb_X}{32\overline{U}_X}\ k\ge 24d\ln(288GN/\delta)+1$,
(5) (4.11) *holds for* $\tilde{\delta}=\delta/12$,
(6) $N\ge 8\ln(2G/\delta)/\underline{\pi}^2$,

with probability at least $1-\delta$ *conditional on* $\mathbf{g}$, *it holds that*

$$\begin{aligned}
&|\widehat{B}_{gh}(x,x')-B_{gh}(x,x')|\\
&\le\frac{1024G\left[\|B(x,x')\|_{\max}k+\tau\right]^2}{\sigma_G\left(B(x,x')\right)^2\min_{h\in[G]}\left\lfloor\frac{\underline{\pi}cb_X}{32\overline{U}_X}\ k\right\rfloor^2}\left[\frac{k}{\left\lfloor\frac{\underline{\pi}cb_X}{32\overline{U}_X}\ k\right\rfloor}+\frac{k^2}{\left\lfloor\frac{\underline{\pi}cb_X}{32\overline{U}_X}\ k\right\rfloor^2}\right]\times \quad (4.13)\\
&\left[4\sqrt{\frac{3\ln(48k/\delta)}{\Delta\left\lfloor\frac{\underline{\pi}cb_X}{32\overline{U}_X}\ k\right\rfloor+\tau}}+\frac{2kl_BR_k}{\Delta\left\lfloor\frac{\underline{\pi}cb_X}{32\overline{U}_X}\ k\right\rfloor+\tau}\left(\frac{2kl_BR_k}{\Delta\left\lfloor\frac{\underline{\pi}cb_X}{32\overline{U}_X}\ k\right\rfloor+\tau}+3\right)\right]^2 \quad (4.14)\\
&+2l_BR_k+\frac{\sqrt{2\ln(4/\delta)}}{\left\lfloor\frac{\underline{\pi}cb_X}{32\overline{U}_X}\ k\right\rfloor}.
\end{aligned}$$

Proof. Under Condition (6), $\Pr(T(\mathbf{g})) \geq 1 - \delta/2$ and under the remaining conditions, by Lemma 4.1,

$$\Pr\left(\text{ (4.13) holds} \mid \mathbf{g}, T(\mathbf{g})\right) \geq 1 - \delta/2.$$

□

Remark 4.1. As in the existing literature on the SBM (or other models with mixture structure in general) the preceding results hold up to relabeling of communities. This usually causes no issues since community labels have little practical implications in those models. While this observation partially applies to our procedure, the issue of community labeling still poses a novel challenge in terms of the interpretability of our estimators.

Take a simple, special case with $G = 2$. The results obtained above enable us to identify and estimate each of the four elements of the edge probability matrix $B(x, x')$, $(x, x') \in S^2$, as well as the two vectors of community assignment probabilities, up to relabeling of the rows and the columns. We are clearly free to choose the labels for either the rows or the columns, so let us say we fix the two labels for the rows (those corresponding to the nodes with covariate value being x): this essentially amounts to normalization. Given this normalization, however, one may wish to identify the community labels for the columns (i.e. the order of the two columns of B), at least for two reasons. First, even when one is interested in the edge probabilities and the community assignment probabilities at just one point (x, x') in S^2, interpreting and using these probabilities might demand identification of the column (row) order, relative to a given choice of the row (column) order. For example, a diagonal element of the edge probability matrix represents connections within an unobservable community (though they generally differ in terms of the observed heterogeneity, as far as $x \neq x'$); Such within-community connections are often associated with homophily/heterophily or (dis)assortativity. Of course, such issues potentially affect off-diagonal elements of the edge probability matrix with general $G \geq 2$. Second, if, for example, one is interested in the partial effect of moving x' to x'' in $B_{gh}(x, \cdot)$ or $\pi_h(\cdot)$, then it is obvious that we need to have the correspondence between the community labels remain consistent between (x, x') and (x, x'').

This issue does not arise in the standard SBM without covariates as in Lei and Rinaldo (2015), as their edge probability matrix are defined for the same population. This holds true even in the analysis of asymmetric networks in Rohe, Qin, and Yu (2016). In our analysis, if $x \neq x'$, then the edge probability matrix $B(x, x')$ is concerned with edges between two separate populations, and this feature gives rise to difficulties in guaranteeing proper matching between the row-clusters and the column clusters in the absence of further information/restrictions. Of course, by setting $x = x'$ in

$B(x,x')$, matching the community labels between the two sides is trivial, and then to the extent that continuity of $B_{g,h}(\cdot,\cdot),(g,h)\in[G]\times[G]$ and the connectedness of the support of covariate X permit, the labels can be matched as we move x' away from x. Such approach may fail to be reasonable or practical in actual applications, however.

If we are willing to impose additional restrictions, it is possible to address this issue directly. For example, once again, we are free to choose community labeling such that $\pi_1(x) > \pi_2(x) > \cdots > \pi_G(x)$, assuming no ties. Suppose we also choose community labels for the columns, so that $\pi_1(x') > \pi_2(x') > \cdots > \pi_G(x')$. Under the assumption that the ranking of the magnitudes of community assignment probabilities is invariant between those at x and those at x', then trivially the community labels on both sides can be matched in a consistent manner.

One may also wish to introduce qualitative restrictions that are motivated by concepts developed in the literature of network analysis, in order to achieve successful community label matching. Examples of such restrictions include homophily/heterophily, or, assortativity/disasortativity, *conditional on the covariates*. Suppose the edge probability matrix $B(x,x')$ is conditionally weakly assortative at $(x,x')\in S^2$, in the sense that[2]

$$B_{gg}(x,x') > \max(B_{g,h}(x,x'),B_{h,g}(x,x')) \text{ for every } (g,h)\in[G]\times[G] \text{ with } g\neq h. \tag{4.15}$$

This can be justified under homophily in terms of unobserved community membership, while the impact of covariates (x,x') on edge probabilities remain fully unspecified (and can be correlated with unobserved heterogeneity in an arbitrary manner). The restriction (4.15) suffices to achieve identification of matched community labels. Let $\mathcal{E}_G$ denote the set of $G\times G$ permutation matrices. Without loss of generality, assign G labels on the rows of the edge probability matrix; let $B(x,x')$ be the resulting matrix to be (uniquely) recovered. Without a restriction such as (4.15), we can only identify the set $\{B(x,x')E_G, E_G\in\mathcal{E}_G\}$. Let $B^\circ(x,x')$ an arbitrary element of the set. To exploit the

[2]Technically, even a weaker condition such as

$$B_{gg}(x,x') > B_{g,h}(x,x') \text{ for every } (g,h)\in[G]\times[G] \text{ with } g\neq h.$$

which ensures that each diagonal element dominates the rest of its row elements — or its column elements, by symmetry — suffices. The strong assortative version of (4.15) can be obtained by strengthening the inequality by

$$B_{gg}(x,x') > B_{h,f}(x,x') \text{ for every } (f,g,h)\in[G]\times[G]\times[G] \text{ with } h\neq f.$$

asortativity restriction (4.15), we can simply solve

$$\overline{E}_G = \underset{E_G \in \mathcal{E}_G}{\operatorname{argmax}} \operatorname{tr}[B^\circ(x,x')E_G],$$

and then we recover $B(x,x')$ by $B^\circ(x,x')\overline{E}_G$. This offers a practical algorithm, as the maximization problem $\max_{E_G \in \mathcal{E}_G} \operatorname{tr}[AE_G]$ is well defined for any G by G matrix A as far as the maximum entry of each row has no ties. Likewise, if one is willing to impose a heterophily, we can flip the inequality sign in (4.15) to restrict $B(x,x')$ to be disasortative, then solve

$$\underline{E}_G = \underset{E_G \in \mathcal{E}_G}{\operatorname{argmin}} \operatorname{tr}[B^\circ(x,x')E_G],$$

to recover the desired edge probability matrix.

5. Conclusion

This paper demonstrates that it is possible to incorporate both observed and unobserved heterogeneity in network data analysis in a flexible way, at least when we have discrete values of heterogeneity represented by community assignments. It offers a highly versatile, yet computationally tractable procedure, which is expected to complement the existing methodology for analyzing networks with covariates under more specific structures for the edge probabilities and the community assignment probabilities. Our results build upon recent developments in spectral clustering in SBMs and k-nn algorithms, and we contribute to the literature by addressing novel theoretical challenges presented by our multi-step procedure. Though our estimators can be computed in a straightforward manner, an extensive simulation exercise is called for in order to assess the efficacy of the new procedure in a practical setting.

Supplement: Some Useful results

A 1.1. **Nearest Neighbor Radius.**

To obtain uniform upper and lower bounds on the radiuses, we use inequalities for relative deviations see Anthony and Shawe-Taylor (1993) and Section 1.4.2 of Lugosi (2002). Before stating the two inequalities we will use, we introduce some definitions. Let x denote a $\mathbf{R}^d$ valued covariate vector. Let $P_{\mathbf{x}N}$ denote the empirical measure based on $\mathbf{x} = (x_1, ..., x_N)$, that is, $P_{\mathbf{x}N}(A) := \#\{x(i) \in A, i \in [N]\}/N$ for $A \in \mathcal{C}$. Likewise define $P_{\mathbf{y}N}(A) := \#\{y(i) \in A, i \in [N]\}/N$ and $P_{\mathbf{xy}N}(A) := \#\{x(i) \in A, y(i) \in A, i \in [N]\}/2N$. We consider $\mathbf{x} = (x_1, ..., x_N)$ and $\mathbf{y} = (y_1, ..., y_N)$ defined on the sample space S^N, independent of each other and both distributed according to $P(.)^n$ for a probability measure P. Let $\mathcal{C}$ be a collection of subsets of S. For a sample $\mathbf{x} = (x_1, ..., x_N)$ viewed as a collection of draws $\{x_1, ..., x_N\}$, we define as in Giné and Nickl (2021) (Section 3.6.1) the trace of $\mathcal{C}$ on $\mathbf{x}$ as all the subsamples of $\mathbf{x}$ obtained by intersection of $\mathbf{x}$ with sets $A \in \mathcal{C}$. Define $\Delta_{\mathcal{C}}(\mathbf{x})$ as the cardinal of the trace of the collection $\mathcal{C}$ and

$$m_{\mathcal{C}}(N) = \sup_{\mathbf{x} \in S^N} \Delta_{\mathcal{C}}(\mathbf{x}).$$

$m_{\mathcal{C}}(N)$ is the shattering coefficient of the collection $\mathcal{C}$. Following Anthony and Shawe-Taylor (1993), we define a complete set of distinct representatives (CSDR) of $\mathcal{C}$ for $\mathbf{x}$ as a collection $\mathcal{A} = \{A^1, ..., A^{\Delta_{\mathcal{C}}(\mathbf{x})}\}$ if for any $1 \leq i \neq j \leq \Delta_{\mathcal{C}}(\mathbf{x})$ then $A^i \cap \{x_1, ..., x_N\} \neq A^j \cap \{x_1, ..., x_N\}$. For all $A \in \mathcal{C}$, there exists $1 \leq i \leq \Delta_{\mathcal{C}}(\mathbf{x})$ such that $A \cap \{x_1, ..., x_N\} = A^i \cap \{x_1, ..., x_N\}$.

We are in particular interested in the following inequalities, see e.g. Theorem 1.11 in Lugosi (2002),

$$\forall \eta > 0, \; \Pr\left(\sup_{a \in \mathcal{C}} \frac{P(A) - P_{\mathbf{x}N}(A)}{\sqrt{P(A)}} > \eta\right) \leq 4 m_{\mathcal{C}}(2N) \exp\left(-\eta^2 N/4\right) \tag{A 1.1}$$

$$\forall \eta > 0, \; \Pr\left(\sup_{a \in \mathcal{C}} \frac{P_{\mathbf{x}N}(A) - P(A)}{\sqrt{P_{\mathbf{x}N}(A)}} > \eta\right) \leq 4 m_{\mathcal{C}}(2N) \exp\left(-\eta^2 N/4\right) \tag{A 1.2}$$

where $\mathcal{C}$ is any collection of Borelian sets. We first give a proof of (A 1.2). As we could not find one in the literature, we restate one which we will later modify to accommodate non-identically (but independently) distributed random variables.

Lemma A 1.1. *For $\mathcal{C}$ any collection of Borelian sets,* (A 1.2) *holds.*

Proof. The proof adapts the steps of Anthony and Shawe-Taylor (1993). We define the sets

$$Q := \left\{(x_1, ..., x_N) \in S^N : \exists A \in \mathcal{C} \text{ such that } \frac{P_{\mathbf{x}N}(A) - P(A)}{\sqrt{P_{\mathbf{x}N}(A)}} > \eta\right\}$$

$$R := \left\{(x_1, ..., x_N, y_1, ..., y_N) \in S^{2N} : \exists A \in \mathcal{C} \text{ such that } P_{\mathbf{x}N}(A) - P_{\mathbf{y}N}(A) > \eta\sqrt{P_{\mathbf{xy}N}(A)}\right\}.$$

We first look at the case $N > 2/\eta$. First note that for each $\mathbf{x} \in Q$ there exists a set $A_{\mathbf{x}} \in \mathcal{C}$, indexed by $\mathbf{x}$, such that $P_{\mathbf{x}N}(A_{\mathbf{x}}) - P(A_{\mathbf{x}}) > \eta\sqrt{P_{\mathbf{x}N}(A_{\mathbf{x}})}$. Define

$$F_{\mathbf{xy}}(A_{\mathbf{x}}) := \frac{P_{\mathbf{x}N}(A_{\mathbf{x}}) - P_{\mathbf{y}N}(A_{\mathbf{x}})}{\sqrt{P_{\mathbf{xy}N}(A_{\mathbf{x}})}}.$$

Take $P_{\mathbf{y}N}(A_{\mathbf{x}})$ such that $P_{\mathbf{y}N}(A_{\mathbf{x}}) < P(A_{\mathbf{x}})$. Then

$$F_{\mathbf{xy}}(A_{\mathbf{x}}) > \frac{P(A_{\mathbf{x}}) + \eta\sqrt{P_{\mathbf{x}N}(A_{\mathbf{x}})} - P_{\mathbf{y}N}(A_{\mathbf{x}})}{\sqrt{[P_{\mathbf{x}N}(A_{\mathbf{x}})) + P_{\mathbf{y}N}(A_{\mathbf{x}})]/2}}$$

$$> \frac{\eta\sqrt{P_{\mathbf{x}N}(A)}}{\sqrt{[P_{\mathbf{x}N}(A_{\mathbf{x}}) + P(A_{\mathbf{x}})]/2}} > \eta,$$

where the second inequality holds by monotonicity in $P_{\mathbf{x}N}(A_{\mathbf{x}})$ and the third by $P_{\mathbf{x}N}(A_{\mathbf{x}}) > P(A_{\mathbf{x}})$.

$$\begin{aligned}
\Pr\{(\mathbf{x},\mathbf{y}) \in R|\mathbf{g}\} &\geq \Pr\{F_{\mathbf{xy}}(A_{\mathbf{x}}) > \eta \,|P_{\mathbf{y}N}(A_{\mathbf{x}}) < P(A_{\mathbf{x}}), \mathbf{x} \in Q\} \Pr\{P_{\mathbf{y}N}(A_{\mathbf{x}}) < P(A_{\mathbf{x}}), \mathbf{x} \in Q\} \\
&\geq \left(\inf_{\mathbf{x}\in Q} \Pr\{P_{\mathbf{y}N}(A_{\mathbf{x}}) < P(A_{\mathbf{x}})\}\right) \Pr\{\mathbf{x} \in Q\} = \left(\inf_{\mathbf{x}\in Q} \Pr\{P_{\mathbf{y}N}(A^c_{\mathbf{x}}) > P(A^c_{\mathbf{x}})\}\right) \Pr\{\mathbf{x} \in Q\} \\
&\geq \frac{1}{4} \Pr\{\mathbf{x} \in Q\}.
\end{aligned}$$

where $A^c_{\mathbf{x}}$ is the complement set of $A_{\mathbf{x}}$. The last inequality uses Theorem 1 of Greenberg and Mohri (2014) thus we need to check that for all $\mathbf{x} \in Q$, $P(A^c_{\mathbf{x}}) > 1/N$, that is, $P(A_{\mathbf{x}}) < 1 - 1/N$. Note that $P(A_{\mathbf{x}}) < P_{\mathbf{x}N}(A_{\mathbf{x}}) - \eta\sqrt{P_{\mathbf{x}N}(A_{\mathbf{x}})} =: g(P_{\mathbf{x}N}(A_{\mathbf{x}}))$. The function g decreases on $[0, \eta^2/4]$ and is negative on this interval, and increases on $[\eta^2/4, +\infty)$. Thus $P(A_{\mathbf{x}}) < g(1) = 1 - \eta$, which imposes $\eta < 1$, and $P(A_{\mathbf{x}}) < 1 - 1/N$ follows by $N > 2/\eta$. The inequality (A 1.2) is obtained by

$$\Pr\{(\mathbf{x},\mathbf{y}) \in R|\mathbf{g}\} \leq m_{\mathcal{C}}(2N) \exp\left(-\eta^2 N/4\right),$$

which holds by the proof of Theorem 2.1 in Anthony and Shawe-Taylor (1993).

If $N < 2/\eta$, the upper bound in (A 1.2) is $4m_{\mathcal{C}}(2N)\exp\left(-\eta^2 N/4\right) \geq 4m_{\mathcal{C}}(2N)\exp(-\eta/2)$. For $\mathbf{x} \in Q$, $P(A_{\mathbf{x}}) < g(P_{\mathbf{x}N}(A_{\mathbf{x}}))$ guarantees that $\eta < 1$ which implies that $4m_{\mathcal{C}}(2N)\exp\left(-\eta^2 N/4\right) \geq 4\exp(-1/2) \geq 2$. Thus (A 1.2) holds naturally. Note that if $Q = \emptyset$, (A 1.2) holds naturally as well. □

Before elaborating on our results, we state two preliminary lemmas.

Lemma A 1.2. *Suppose* $\{z_i\}_{i=1}^N$ *are independently distributed, with* $z_i \sim$*Bernoulli*(q_i), $0 < q_i < 1$ *for every* $i \in [N]$. *Let* $S := \sum_{i=1}^N z_i$, *then* $\Pr\{S > N \min_{i\in[N]} q_i\} > \frac{1}{4}$ *if* $\min_{i\in[N]} q_i > \frac{1}{N}$.

Proof. It is easy to see that, if $\underline{z}_i \sim_{iid}$Bernoulli$(\min_{i\in[N]} q_i)$, and we let $\underline{S} := \sum_{i=1}^N \underline{z}_i$, then $\Pr\{S > N \min_{i\in[N]} q_i\} \geq \Pr\{\underline{S} > N \min_{i\in[N]} q_i\}$. Indeed, note that by definition z_i first-order stochastically dominates $\underline{z}_i$, or $z_i \geq_1 \underline{z}_i$ for each i. Then by Theorem 1.A.3(b) in Shaked and Shanthikumar (2007), we have $S \geq_1 \underline{S}$ and the statement indeed follows. Since $\underline{S} \sim$Binom$(N, \min_{i\in[N]} q_i)$, by Theorem 1 of Greenberg and Mohri (2014) $\Pr\{\underline{S} > N \min_{i\in[N]} q_i\} > \frac{1}{4}$ if $\min_{i\in[N]} q_i > \frac{1}{N}$ and the result follows. □

Corollary A 1.1. *Suppose* $\{z_i\}_{i=1}^N$ *are independently distributed, with* $z_i \sim$*Bernoulli*(q_i), $0 < q_i < 1$ *for every* $i \in [N]$. *Let* $S := \sum_{i=1}^N z_i$, *then* $\Pr\{S < N \max_{i\in[N]} q_i\} > \frac{1}{4}$ *if* $\max_{i\in[N]} q_i < 1 - \frac{1}{N}$.

Proof. Apply Lemma A 1.2 to $\{\tilde{z}_i\}_{i=1}^N = \{1 - z_i\}_{i=1}^N$. □

We now extend (A 1.1) and (A 1.2) to accommodate non-identically (but independently) distributed random variables (i.ni.d), induced by conditioning. We first introduce additional notation as we now consider $\mathbf{x} = (x_1, ..., x_N)$ and $\mathbf{y} = (y_1, ..., y_N)$ defined on the sample space S^N, independent of each other and both distributed according to $\prod_{i=1}^N f(.|g_i)$ conditional on $g(i) = g_i, i \in [N]$. We use $\Pr\{\cdot|\mathbf{g}\}$ to denote the probability of event $\cdot$ conditional on $(g(1), ..., g(N)) = \mathbf{g}$. Define

$$\underline{P}(A) := \int \underline{f}(x)dx, \tag{A 1.3}$$

$$\overline{P}(A) := \int_A \overline{f}(x)dx. \tag{A 1.4}$$

where $\underline{f}$ and $\overline{f}$ are as defined in Section 2.3. Define $\mathcal{B} = \{\mathcal{B}(x,\tau) \,|\, x \in \mathbb{R}^d,\, \tau > 0\}$ and

$$\underline{Q} := \left\{(x_1, ..., x_N) \in S^N : \exists A \in \mathcal{C} \text{ such that } \frac{\underline{P}(A) - P_{\mathbf{x}N}(A)}{\sqrt{\underline{P}(A)}} > \eta\right\},$$

$$\overline{Q} := \left\{(x_1, ..., x_N) \in S^N : \exists A \in \mathcal{C} \text{ such that } \frac{P_{\mathbf{x}N}(A) - \overline{P}(A)}{\sqrt{P_{\mathbf{x}N}(A)}} > \eta\right\},$$

$$R := \left\{(x_1, ..., x_N, y_1, ..., y_N) \in S^{2N} : \exists A \in \mathcal{C} \text{ such that } P_{\mathbf{x}N}(A) - P_{\mathbf{y}N}(A) > \eta\sqrt{P_{\mathbf{xy}N}(A)}\right\}.$$

Lemma A 1.3 and Lemma A 1.4 extend (A 1.1) and (A 1.2) to i.ni.d data.

Lemma A 1.3. *For any collection $\mathcal{C}$ of Borel sets,*

$$\Pr\{\mathbf{x} \in \underline{Q}|\mathbf{g}\} \leq 4m_{\mathcal{C}}(2N)\exp\left(-\eta^2 N/4\right).$$

Proof. The proof relies on two claims.

First claim: $\Pr\{\mathbf{x} \in \underline{Q}|\mathbf{g}\} \leq 4\Pr\{(\mathbf{x},\mathbf{y}) \in R|\mathbf{g}\}$ if $N > 2/\eta^2$.

Proof of the first claim: We follow the proof of Theorem 2.1 of Anthony and Shawe-Taylor (1993), which deals with IID sequences, while accomodating heterogeneity induced by conditioning on $\mathbf{g}$. First note that for each $\mathbf{x} \in \underline{Q}$ there exists a set $A_{\mathbf{x}} \in \mathcal{C}$, indexed by $\mathbf{x}$, such that $\underline{P}(A_{\mathbf{x}}) - P_{\mathbf{x}N}(A_{\mathbf{x}}) > \eta\sqrt{\underline{P}(A_{\mathbf{x}})}$. It then follows that $\inf_{\mathbf{x}\in\underline{Q}} \underline{P}(A_{\mathbf{x}}) \geq \eta^2$ for $\mathbf{x} \in \underline{Q}$. Define

$$F_{\mathbf{xy}}(A_{\mathbf{x}}) := \frac{P_{\mathbf{y}N}(A_{\mathbf{x}}) - P_{\mathbf{x}N}(A_{\mathbf{x}})}{\sqrt{P_{\mathbf{xy}N}(A_{\mathbf{x}})}}.$$

Note that for every $\mathbf{x} \in \underline{Q}$

$$\begin{aligned}\min_{g\in[G]} \int_{A_{\mathbf{x}}} f(x|g)dx &\geq \inf_{\xi\in\underline{Q}} \underline{P}(A_\xi) \\ &\geq \eta^2\end{aligned}$$

$$> \frac{2}{N}.$$

Therefore by Lemma A 1.2 we have $\Pr\{P_{\mathbf{y}N}(A_{\mathbf{x}}) > \underline{P}(A_{\mathbf{x}})|\mathbf{g}\} > \frac{1}{4}$ for every $x \in \underline{Q}$. Then noting that $\frac{d}{dy}\left((y-x)/\sqrt{\frac{x+y}{2}}\right)$ is nonnegative,

$$\begin{aligned}
\Pr\{(\mathbf{x},\mathbf{y}) \in R|\mathbf{g}\} &\geq \Pr\{F_{\mathbf{xy}}(A_{\mathbf{x}}) > \eta \;|P_{\mathbf{y}N}(A_{\mathbf{x}}) > \underline{P}(A_{\mathbf{x}}), \mathbf{x} \in \underline{Q}, \mathbf{g}\} \Pr\{P_{\mathbf{y}N}(A_{\mathbf{x}}) > \underline{P}(A_{\mathbf{x}}), \mathbf{x} \in \underline{Q}|\mathbf{g}\} \\
&\geq \Pr\left\{\frac{\underline{P}(A_{\mathbf{x}}) - P_{\mathbf{x}N}(A_{\mathbf{x}})}{(\sqrt{(P_{\mathbf{x}N}(A_{\mathbf{x}}) + \underline{P}(A_{\mathbf{x}}))/2}} > \eta|\mathbf{x} \in \underline{Q}, \mathbf{g}\right\}\left(\inf_{\mathbf{x}\in\underline{Q}} \Pr\{P_{\mathbf{y}N}(A_{\mathbf{x}}) > \underline{P}(A_{\mathbf{x}})|\mathbf{g}\}\right) \Pr\{\mathbf{x} \in \underline{Q}|\mathbf{g}\} \\
&\geq \frac{1}{4} \Pr\left\{\frac{\eta\sqrt{\underline{P}(A_{\mathbf{x}})}}{\sqrt{(P_{\mathbf{x}N}(A_{\mathbf{x}}) + \underline{P}(A_{\mathbf{x}}))/2}} > \eta|\mathbf{x} \in \underline{Q}, \mathbf{g}\right\} \Pr\{\mathbf{x} \in \underline{Q}|\mathbf{g}\} \\
&= \frac{1}{4} \;\Pr\{\mathbf{x} \in \underline{Q}|\mathbf{g}\}.
\end{aligned}$$

Second claim: if $N > 2/\eta^2$,

$$\Pr\{(\mathbf{x},\mathbf{y}) \in R|\mathbf{g}\} \leq m_{\mathcal{C}}(2N) \exp\left(-\eta^2 N/4\right).$$

Proof of the second claim: Define Λ as in Anthony and Shawe-Taylor (1993), i.e., the group generated by all transpositions of the form $(i, N+i)$ for $1 \leq i \leq N$. Consider $\tau \in \Lambda$ and define for $z \in S^{2N}$, $\tau\mathbf{z} := (\mathbf{z}_{\tau(1)}, ..., \mathbf{z}_{\tau(2N)})$. If $\mathbf{x}$ and $\mathbf{y}$ are two samples independent of each other and both distributed according to $\prod_{i=1}^N f(.|g_i)$, then $\mathbf{xy}$ and $\tau\mathbf{xy}$ have the same distribution. Thus

$$\begin{aligned}
\Pr(R|\mathbf{g}) &= \Pr(\exists A_{\tau\mathbf{xy}} \in \mathcal{C} \text{ such that } F_{\tau\mathbf{xy}}(A_{\tau\mathbf{xy}}) > \eta \;|\mathbf{g}) \\
&= \frac{1}{|\Lambda|}\, \mathbb{E}\left(\sum_{\tau\in\Lambda} \mathbf{1}\left[\exists A_{\tau\mathbf{xy}} \in \mathcal{C} \text{ such that } F_{\tau\mathbf{xy}}(A_{\tau\mathbf{xy}}) > \eta\right] |\mathbf{g}\right)
\end{aligned}$$

We use the notation $(A^1_{\mathbf{xy}}, ..., A^{\Delta_{\mathcal{C}}(\mathbf{xy})}_{\mathbf{xy}})$ for a CSDR of $\mathbf{xy}$. Note that any CSDR of $\tau\mathbf{xy}$ is a CSDR of $\mathbf{xy}$ and vice versa. Then for all $A \in \mathcal{C}$, there exists $1 \leq t \leq \Delta_{\mathcal{C}}(\mathbf{xy})$ such that $F_{\tau\mathbf{xy}}(A) = F_{\tau\mathbf{xy}}(A^t_{\mathbf{xy}})$. Thus

$$\Pr\{(\mathbf{x},\mathbf{y}) \in R|\mathbf{g}\} \leq \frac{1}{|\Lambda|}\, \mathbb{E}\left(\sum_{\tau\in\Lambda} \sum_{t=1}^{\Delta_{\mathcal{C}}(\mathbf{xy})} \mathbf{1}\left[F_{\tau\mathbf{xy}}(A^t_{\mathbf{xy}}) > \eta\right] |\mathbf{g}\right)$$

Define $\Theta^t(\mathbf{xy})$ as in Anthony and Shawe-Taylor (1993), that is, the number of permutations $\tau \in \Lambda$ such that $F_{\tau\mathbf{xy}}(A^t_{\mathbf{xy}}) > \eta$. The inequality above can be rewritten

$$\Pr\{(\mathbf{x},\mathbf{y}) \in R|\mathbf{g}\} \leq \frac{1}{|\Lambda|}\, \mathbb{E}\left(\sum_{t=1}^{\Delta_{\mathcal{C}}(\mathbf{xy})} \sum_{\tau\in\Lambda} \mathbf{1}\left[F_{\tau\mathbf{xy}}(A^t_{\mathbf{xy}}) > \eta\right] |\mathbf{g}\right) = \frac{1}{|\Lambda|}\, \mathbb{E}\left(\sum_{t=1}^{\Delta_{\mathcal{C}}(\mathbf{xy})} \Theta^t(\mathbf{xy})\,|\mathbf{g}\right).$$

As in Anthony and Shawe-Taylor (1993),

$$\frac{\Theta^t(\mathbf{xy})}{|\Lambda|} \leq \exp(-\eta^2 N/4),$$

thus

$$\Pr\{(\mathbf{x},\mathbf{y}) \in R|\mathbf{g}\} \leq \Delta_{\mathcal{C}}(\mathbf{xy})\exp(-\eta^2 N/4) \leq m_{\mathcal{C}}(2N)\exp\left(-\eta^2 N/4\right).$$

By the first and second claim, if $N > 2/\eta^2$, $\Pr\{\mathbf{x} \in \underline{Q}|\mathbf{g}\} \leq 4m_{\mathcal{C}}(2N)\exp\left(-\eta^2 N/4\right)$. Note that if $N \leq 2/\eta^2$, $4\exp(-\eta^2 N/4) \geq 2$ thus the previous inequality also holds. □

The next lemma adapts Lemma A 1.4 to i.ni.d data.

Lemma A 1.4. *For any collection $\mathcal{C}$ of Borel sets,*

$$\Pr\{\mathbf{x} \in \overline{Q}|\mathbf{g}\} \leq 4m_{\mathcal{C}}(2N)\exp\left(-\eta^2 N/4\right).$$

Proof. We look first at the case $N > 2/\eta$ and prove that $\Pr\{\mathbf{x} \in \overline{Q}|\mathbf{g}\} \leq 4\Pr\{(\mathbf{x},\mathbf{y}) \in R|\mathbf{g}\}$. The rest of the proof follows by the second claim in the proof of Lemma A 1.3.

First note that for each $\mathbf{x} \in \overline{Q}$ there exists a set $A_{\mathbf{x}} \in \mathcal{C}$, indexed by $\mathbf{x}$, such that $P_{\mathbf{x}N}(A_{\mathbf{x}}) - \overline{P}(A_{\mathbf{x}}) > \eta\sqrt{P_{\mathbf{x}N}(A_{\mathbf{x}})}$. As in the proof of Lemma A 1.4, this implies that $P_{\mathbf{x}N}(A_{\mathbf{x}}) > \overline{P}(A_{\mathbf{x}})$ and $\overline{P}(A_{\mathbf{x}}) \leq 1-\eta$. Define

$$F_{\mathbf{xy}}(A_{\mathbf{x}}) := \frac{P_{\mathbf{x}N}(A_{\mathbf{x}}) - P_{\mathbf{y}N}(A_{\mathbf{x}})}{\sqrt{P_{\mathbf{xy}N}(A_{\mathbf{x}})}}.$$

If $P_{\mathbf{y}N}(A_{\mathbf{x}}) < \overline{P}(A_{\mathbf{x}})$,

$$\begin{aligned} F_{\mathbf{xy}}(A_{\mathbf{x}}) &> \frac{\overline{P}(A_{\mathbf{x}}) + \eta\sqrt{P_{\mathbf{x}N}(A_{\mathbf{x}})} - P_{\mathbf{y}N}(A_{\mathbf{x}})}{\sqrt{[P_{\mathbf{x}N}(A_{\mathbf{x}})) + P_{\mathbf{y}N}(A_{\mathbf{x}})]/2}} \\ &> \frac{\eta\sqrt{P_{\mathbf{x}N}(A)}}{\sqrt{[P_{\mathbf{x}N}(A_{\mathbf{x}}) + \overline{P}(A_{\mathbf{x}})]/2}} > \eta. \end{aligned}$$

Note that for every $\mathbf{x} \in \overline{Q}$, $\overline{P}(A_{\mathbf{x}}) > \max_{g\in[G]} P(A_{\mathbf{x}}|g)$. Thus $\Pr\{P_{\mathbf{y}N}(A_{\mathbf{x}}) < \overline{P}(A_{\mathbf{x}})|\mathbf{g}\} \geq \Pr\{P_{\mathbf{y}N}(A_{\mathbf{x}}) < \max_{g\in[G]} P(A_{\mathbf{x}}|g)|\mathbf{g}\} > 1/4$ by Corollary A 1.1 as long as $\max_{g\in[G]} P(A_{\mathbf{x}}|g) \leq 1 - 1/N$. This holds by

$$\begin{aligned} \max_{g\in[G]} P(A_{\mathbf{x}}|g) &\leq \overline{P}(A_{\mathbf{x}}) \\ &\leq 1-\eta \\ &< 1-\frac{2}{N}. \end{aligned}$$

Therefore

$$\Pr\{(\mathbf{x},\mathbf{y}) \in R|\mathbf{g}\} \geq \Pr\{F_{\mathbf{xy}}(A_{\mathbf{x}}) > \eta \,|P_{\mathbf{y}N}(A_{\mathbf{x}}) < \overline{P}(A_{\mathbf{x}}), \mathbf{x} \in \overline{Q}, \mathbf{g}\}\Pr\{P_{\mathbf{y}N}(A_{\mathbf{x}}) < \overline{P}(A_{\mathbf{x}}), \mathbf{x} \in \overline{Q}|\mathbf{g}\}$$

$$\geq \left(\inf_{\mathbf{x}\in\overline{Q}} \Pr\{P_{\mathbf{y}N}(A_{\mathbf{x}}) < \overline{P}(A_{\mathbf{x}})|\mathbf{g}\}\right) \Pr\{\mathbf{x}\in\overline{Q}|\mathbf{g}\}$$

$$\geq \frac{1}{4} \Pr\{\mathbf{x}\in\overline{Q}|\mathbf{g}\}.$$

The case $N < 2/\eta$ is handled as in the proof of Lemma A 1.4. □

We now apply these results to derive bounds on radiuses. Lemma A 1.3 implies the following version of Theorem 4 in Portier (2021).

Corollary A 1.2. *Let $(x(i))_{i\leq N}$ be a sequence of independent and nonidentically distributed random vectors valued in $\mathbb{R}^d$ and $\underline{P}$ and $\overline{P}$ defined as above. For any $\delta > 0$,*

With probability at least $1-\delta$ conditional on $\mathbf{g}$,

$$\forall B\in\mathcal{B}, \frac{1}{N}\sum_{i=1}^{N}\mathbf{1}\left(x(i)\in B\right) \geq \underline{P}(B)\left(1-\sqrt{\frac{12d\ln(12N/\delta)}{N\underline{P}(B)}}\right), \tag{A 1.5}$$

With probability at least $1-\delta$ conditional on $\mathbf{g}$,

$$\forall B\in\mathcal{B}, \frac{1}{N}\sum_{i=1}^{N}\mathbf{1}\left(x(i)\in B\right) \leq \frac{12d\ln(12N/\delta)}{N} + 4\overline{P}(B). \tag{A 1.6}$$

Proof. The proof of (A 1.5) applies Lemma A 1.3 to the collection $\mathcal{B}$ and follows the lines of the proof of Theorem 4 in Portier (2021).

To obtain (A 1.6), note that Lemma A 1.4 applied to the collection $\mathcal{B}$ implies that with probability at least $1-\delta$,

$$\forall B\in\mathcal{B}, \frac{P_{\mathbf{x}N}(B)-\overline{P}(B)}{\sqrt{P_{\mathbf{x}N}(B)}} \leq \sqrt{\frac{4[\ln\left(4m_{\mathcal{C}}(2N)/\delta\right)]}{N}} \leq \sqrt{\frac{12d\ln(12N/\delta)}{N}}$$

where the second inequality is obtained through the same arguments as in the proof of Theorem 4 in Portier (2021). Define $\beta_n := \sqrt{\frac{12d\ln(12N/\delta)}{N}}$ and the function $g : x\mapsto x^2-\beta_n x-\overline{P}(B)$. The function g has two roots, thus

$$[P_{\mathbf{x}N}(B)-\overline{P}(B)]/\sqrt{P_{\mathbf{x}N}(B)} \leq \beta_n$$

$$\Rightarrow g(\sqrt{P_{\mathbf{x}N}(B)}) \leq 0$$

$$\Rightarrow \sqrt{P_{\mathbf{x}N}(B)} \leq \frac{1}{2}\left(\beta_n+\sqrt{\beta_n^2+4\overline{P}(B)}\right) \leq \sqrt{\beta_n^2+4\overline{P}(B)}.$$

□

We finally obtain the following upper and lower bounds on the radius.

Lemma A 1.5. *Let Assumptions 2.1, 2.2 and 2.3 hold. Then,* *(1) for all N, $\delta \in (0,1)$ and $1 \leq k \leq N$ such that $24d\ln(12N/\delta) \leq k \leq T^d N b_X c V_d/2$,*

$$\Pr\left(\sup_{x\in S} r_k(x) \leq R_k|\mathbf{g}\right) \geq 1-\delta, \tag{A 1.7}$$

(2) for all N, $\delta \in (0,1)$ and $1 \leq k \leq N$ such that $k \geq 12d\ln(12N/\delta)$,

$$\Pr\left(\inf_{x\in S} r_k(x) \geq \underline{R}_k|\mathbf{g}\right) \geq 1-\delta. \tag{A 1.8}$$

Proof. (A 1.7) is derived replacing P with $\underline{P}$ in the proof of Lemma 4 in Portier (2021) and using (A 1.5).

To prove (A 1.8), note that

$$\begin{aligned}\overline{P}(\mathcal{B}(x,\underline{R}_k)) &= \int_{\mathcal{B}(x,\underline{R}_k)\cap S}\overline{f}(x)dx\\ &\leq \overline{U}_X\lambda(\mathcal{B}(x,\underline{R}_k)\cap S)\\ &\leq \frac{k-12d\ln(12N/\delta)}{4N}\end{aligned}$$

where λ is the Lebesgue measure. By (A 1.6), with probability at least $1-\delta$,

$$\forall x \in S, \frac{1}{N}\sum_{i=1}^{N}\mathbf{1}\left(x(i)\in\mathcal{B}(x,\underline{R}_k)\right) \leq k/N$$

which implies (A 1.8). □

A 1.2. **Results on Laplacian.**

Lemma A 1.6. $\|\widetilde{L^\eta_\tau}\| \leq 1$.

Proof. By Theorem 7.3.3 of Horn and Johnson (2012), the eigenvalues $\widetilde{L^\eta_\tau}$ are $\sigma_1(L^\eta_\tau) \geq ... \geq \sigma_G(L^\eta_\tau) \geq 0 \geq -\sigma_G(L^\eta_\tau) \geq ... \geq -\sigma_1(L^\eta_\tau)$. Thus the claim holds if $\sigma_1(L^\eta_\tau) \leq 1$. We equivalently show that $I-\widetilde{L^\eta_\tau}$ is a symmetric positive semidefinite matrix. Note that

$$I-\widetilde{L^\eta_\tau} = \begin{pmatrix} I & -L^\eta_\tau \\ -\left(L^\eta_\tau\right)^\top & I\end{pmatrix}.$$

Take $c \in \mathbb{R}^{2k}$, write $c = \begin{pmatrix} y \\ z\end{pmatrix}$ where $y, z \in \mathbb{R}^k$. Then

$$\begin{aligned}c^\top\left[I-\widetilde{L^\eta_\tau}\right]c &= y^\top y + z^\top z - 2y^\top L^\eta_\tau z\\ &= \sum_{i\in\eta_N(x)} y_i^2 + \sum_{j\in\eta_N(x')} z_j^2 - 2\sum_{\substack{i\in\eta_N(x)\\ j\in\eta_N(x')}}\frac{y_i z_j A_{ij}}{\left[\left(O^\eta_\tau\right)_{ii}\left(Q^\eta_\tau\right)_{jj}\right]^{1/2}}\end{aligned}$$

$$\geq \sum_{\substack{i\in\eta_N(x)\\ j\in\eta_N(x')}} \left(\frac{y_i A_{ij}}{(O^{\eta}_{\tau})^{1/2}_{ii}} - \frac{z_j A_{ij}}{(Q^{\eta}_{\tau})^{1/2}_{jj}} \right)^2 \geq 0,$$

where the first inequality comes from $\tau \geq 0$ thus guaranteeing $I - \widetilde{L^{\eta}_{\tau}}$ is positive semidefinite. □

Lemma A 1.7. $\|\widetilde{\mathcal{L}^{\eta}_{\tau}}(\mathbf{x},\mathbf{g})\| \leq 1$.

Proof. We proceed as for L^{η}_{τ}: we can equivalently show that $I - \widetilde{\mathcal{L}^{\eta}_{\tau}}(\mathbf{x},\mathbf{g})$ is a symmetric positive semidefinite matrix. Take $c \in \mathbb{R}^{2k}$, write $c = \begin{pmatrix} y \\ z \end{pmatrix}$ where $y, z \in \mathbb{R}^k$. Then note that

$$\sum_{\substack{i\in\eta_N(x)\\ j\in\eta_N(x')}} \left(\frac{y_i\sqrt{P^{\eta}_{ij}(\mathbf{x},\mathbf{g})}}{\left[(\mathcal{O}^{\eta}_{\tau}(\mathbf{x},\mathbf{g}))_{ii}\right]^{1/2}} - \frac{z_j\sqrt{P^{\eta}_{ij}(\mathbf{x},\mathbf{g})}}{\left[(\mathcal{Q}^{\eta}_{\tau}(\mathbf{x},\mathbf{g}))_{jj}\right]^{1/2}} \right)^2$$

$$= \sum_{i\in\eta_N(x)} \sum_{j\in\eta_N(x')} \frac{y_i^2 P^{\eta}_{ij}(\mathbf{x},\mathbf{g})}{(\mathcal{O}^{\eta}_{\tau}(\mathbf{x},\mathbf{g}))_{ii}} + \sum_{j\in\eta_N(x')} \sum_{i\in\eta_N(x)} \frac{z_j^2 P^{\eta}_{ij}(\mathbf{x},\mathbf{g})}{(\mathcal{Q}^{\eta}_{\tau}(\mathbf{x},\mathbf{g}))_{jj}} - 2 \sum_{\substack{i\in\eta_N(x)\\ j\in\eta_N(x')}} \frac{y_i z_j P^{\eta}_{ij}(\mathbf{x},\mathbf{g})}{\left[(\mathcal{O}^{\eta}_{\tau}(\mathbf{x},\mathbf{g}))_{ii}\,(\mathcal{Q}^{\eta}_{\tau}(\mathbf{x},\mathbf{g}))_{jj}\right]^{1/2}}$$

$$\leq \sum_{i\in\eta_N(x)} y_i^2 + \sum_{j\in\eta_N(x')} z_j^2 - 2 \sum_{\substack{i\in\eta_N(x)\\ j\in\eta_N(x')}} \frac{y_i z_j P^{\eta}_{ij}(\mathbf{x},\mathbf{g})}{\left[(\mathcal{O}^{\eta}_{\tau}(\mathbf{x},\mathbf{g}))_{ii}\,(\mathcal{Q}^{\eta}_{\tau}(\mathbf{x},\mathbf{g}))_{jj}\right]^{1/2}},$$

by $(\mathcal{O}^{\eta}_{\tau}(\mathbf{x},\mathbf{g}))_{ii} = \sum_{j\in\eta_N(x')} P^{\eta}_{ij}(\mathbf{x},\mathbf{g}) + \tau$, $(\mathcal{Q}^{\eta}_{\tau}(\mathbf{x},\mathbf{g}))_{jj} = \sum_{i\in\eta_N(x)} P^{\eta}_{ij}(\mathbf{x},\mathbf{g}) + \tau$ and $\tau > 0$. Thus we obtain

$$\begin{aligned} c^{\top}\left[I - \widetilde{\mathcal{L}^{\eta}_{\tau}}(\mathbf{x},\mathbf{g})\right] c &= y^{\top}y + z^{\top}z - 2y^{\top}\mathcal{L}^{\eta}_{\tau}(\mathbf{x},\mathbf{g})z \\ &= \sum_{i\in\eta_N(x)} y_i^2 + \sum_{j\in\eta_N(x')} z_j^2 - 2 \sum_{\substack{i\in\eta_N(x)\\ j\in\eta_N(x')}} \frac{y_i z_j P^{\eta}_{ij}(\mathbf{x},\mathbf{g})}{\left[(\mathcal{O}^{\eta}_{\tau}(\mathbf{x},\mathbf{g}))_{ii}\,(\mathcal{Q}^{\eta}_{\tau}(\mathbf{x},\mathbf{g}))_{jj}\right]^{1/2}} \\ &\geq \sum_{\substack{i\in\eta_N(x)\\ j\in\eta_N(x')}} \left(\frac{y_i\sqrt{P^{\eta}_{ij}(\mathbf{x},\mathbf{g})}}{\left[(\mathcal{O}^{\eta}_{\tau}(\mathbf{x},\mathbf{g}))_{ii}\right]^{1/2}} - \frac{z_j\sqrt{P^{\eta}_{ij}(\mathbf{x},\mathbf{g})}}{\left[(\mathcal{Q}^{\eta}_{\tau}(\mathbf{x},\mathbf{g}))_{jj}\right]^{1/2}} \right)^2 \geq 0. \end{aligned}$$

□

References

Anthony, M., and J. Shawe-Taylor (1993): "A result of Vapnik with applications," *Discrete Applied Mathematics*, 47(3), 207–217.

Bhatia, R. (2013): *Matrix analysis*, vol. 169. Springer Science & Business Media.

Bickel, P. J., A. Chen, and E. Levina (2011): "The method of moments and degree distributions for network models," *The Annals of Statistics*, 39(5), 2280–2301.

Chung, F. R., and L. Lu (2006): *Complex graphs and networks*, no. 107. American Mathematical Soc.

Giné, E., and R. Nickl (2021): *Mathematical foundations of infinite-dimensional statistical models*. Cambridge university press.

Greenberg, S., and M. Mohri (2014): "Tight lower bound on the probability of a binomial exceeding its expectation," *Statistics & Probability Letters*, 86, 91–98.

Horn, R. A., and C. R. Johnson (1990): *Matrix Analysis*. Cambridge University Press.

——— (2012): *Matrix analysis*. Cambridge university press.

Jiang, H. (2019): "Non-asymptotic uniform rates of consistency for k-nn regression," in *Proceedings of the AAAI Conference on Artificial Intelligence*, vol. 33, pp. 3999–4006.

Joseph, A., and B. Yu (2016): "Impact of regularization on spectral clustering," *The Annals of Statistics*, 44(4), 1765–1791.

Lei, J., and A. Rinaldo (2015): "Consistency of spectral clustering in stochastic block models," *The Annals of Statistics*, pp. 215–237.

Lugosi, G. (2002): "Pattern classification and learning theory," in *Principles of nonparametric learning*. Springer, pp. 1–56.

Portier, F. (2021): "Nearest neighbor process: weak convergence and non-asymptotic bound," *arXiv preprint arXiv:2110.15083*.

Rohe, K., T. Qin, and B. Yu (2016): "Co-clustering directed graphs to discover asymmetries and directional communities," *Proceedings of the National Academy of Sciences*, 113(45), 12679–12684.

Shaked, M., and J. G. Shanthikumar (2007): *Stochastic orders*. Springer.

Vershynin, R. (2018): *High-dimensional probability: An introduction with applications in data science*, vol. 47. Cambridge university press.

Vu, V. Q., and J. Lei (2013): "Minimax sparse principal subspace estimation in high dimensions," .

Wang, B.-Y., and B.-Y. Xi (1997): "Some inequalities for singular values of matrix products," *Linear algebra and its applications*, 264, 109–115.

Cowles Foundation for Research in Economics, Yale University, New Haven, CT 06520.

Email address: `yuichi.kitamura@yale.edu`

Department of Economics, Georgetown University, Washington, D.C., 20007.

Email address: `louise.laage@georgetown.edu`