LIMIT THEOREMS FOR EIGENVECTORS OF THE NORMALIZED LAPLACIAN FOR RANDOM GRAPHS

BY MINH TANG AND CAREY E. PRIEBE

Johns Hopkins University

We prove a central limit theorem for the components of the eigenvectors corresponding to the d largest eigenvalues of the normalized Laplacian matrix of a finite dimensional random dot product graph. As a corollary, we show that for stochastic blockmodel graphs, the rows of the spectral embedding of the normalized Laplacian converge to multivariate normals and, furthermore, the mean and the covariance matrix of each row are functions of the associated vertex's block membership. Together with prior results for the eigenvectors of the adjacency matrix, we then compare, via the Chernoff information between multivariate normal distributions, how the choice of embedding method impacts subsequent inference. We demonstrate that neither embedding method dominates with respect to the inference task of recovering the latent block assignments.

CONTENTS

1. Introduction. Statistical inference on graphs is a burgeoning field of research in machine learning and statistics, with numerous applications to social network, neuroscience, etc. Many statistical inference procedures for graphs involve a preprocessing step of finding a representation of the vertices as points in some low-dimensional Euclidean space. This representation is usually given by the truncated eigendecomposition of the adjacency matrix or related matrices such as the combinatorial Laplacian or the normalized Laplacian. For example, given a point cloud lying in some purported low-dimensional manifold in a high-dimensional ambient space, many manifold learning or nonlinear dimension reduction algorithms such as Laplacian eigenmaps [5] and diffusion maps [14] use the eigenvectors of the normalized Laplacian constructed from a neighborhood graph of the points as a low-dimensional Euclidean representation of the point cloud before performing inference such as clustering or classification. Spectral clustering algorithms such as the normalized cuts algorithm [36] proceed by embedding a graph into a low-dimensional Euclidean space followed by running K-means on the embedding to obtain a partitioning of the vertices. Some network comparison procedures embed the graphs and then compute a kernel-based distance measure between the resulting point clouds [3, 42].

The choice of the matrix used in the embedding step and its effect on subsequent inference is, however, rarely addressed in the literature. In a recent pioneering work, the authors of [35] addressed this issue by analyzing, in the context of spectral embedding of the adjacency matrix or the normalized Laplacian matrix for stochastic blockmodel graphs where the subsequent inference task is the recovery of the block assignments, a metric given by the average distance between the vertices of a block and its cluster centroid for the spectral embedding. The metric is then used as a surrogate measure for the performance of the subsequent inference task, that is, the metric is a surrogate measure for the error rate in recovering the vertices to block assignments using the spectral embedding. The stochastic blockmodel [19] is a popular generative model for random graphs with latent community structure and many results are known regarding consistent recovery of the block assignments; see, for example, [6, 12, 23, 27, 28, 30, 34, 37, 40] and the references therein.

It was shown in [35] that for two-block stochastic blockmodels, for a large regime of parameters the normalized Laplacian spectral embedding reduces the within-block variance (occasionally by a factor of four) while preserving the between-block variance, as compared to that of the adjacency spectral embedding. This suggests that for a large region of the parameters space for two-block stochastic blockmodels, the spectral embedding of the Laplacian is to be preferred over that of the adjacency matrix for subsequent inference. However, we observed that the metric in [35] is intrinsically tied to the use of K-means as the clustering procedure, that is, a smaller value of the metric for the Laplacian spectral embedding as compared to that for the adjacency spectral embedding only implies that clustering

the Laplacian spectral embedding using K-means is possibly better than clustering the adjacency spectral embedding using K-means.

Motivated by the above observation, one main goal of this paper is to propose a metric that is *independent* of any specific clustering procedure, that is, a metric that characterizes the minimum error achievable by *any* clustering procedure that uses only the spectral embedding, for the recovery of block assignments in stochastic blockmodel graphs. We achieve this by establishing distributional limit results for the eigenvectors corresponding to the few largest eigenvalues of the adjacency or Laplacian matrix and then characterizing, through the notion of statistical information, the distributional differences between the blocks for either embedding method. Roughly speaking, smaller statistical information implies less information to discriminate between the blocks of the stochastic blockmodel.

More specifically, the limit result in [4] states that, for stochastic blockmodel graphs, conditional on the block assignments the entries of the scaled eigenvectors corresponding to the few largest eigenvalues of the adjacency matrix converge to a multivariate normal (see Theorem 2.2) as the number of vertices increases. Furthermore, the associated covariance matrix is not necessarily spherical, and hence K-means clustering for the adjacency spectral embedding does not always yield minimum error for recovering the block assignment. Analogous limit results (see Theorem 3.2) for the eigenvectors of the normalized Laplacian matrix then facilitate comparison between the two embedding methods via the classical notion of Chernoff information [10]. The Chernoff information is a supremum of the Chernoff α-divergences for $\alpha \in (0, 1)$ and characterizes the error rate of the Bayes' decision rule in hypothesis testing; the Chernoff α-divergence is an example of a f-divergence [1, 15] and it satisfies the information processing lemma and is invariant with respect to invertible transformations [24].

Our paper is thus structured as follows. We recall in Section 2 the definition of random dot product graphs, stochastic blockmodel graphs, and spectral embedding of the adjacency and Laplacian matrices. We then state in Section 2.1 several limit results for the eigenvectors of the adjacency spectral embedding. These results are generalizations of results from [4, 41]. The main technical contribution of this paper, namely analogous limit results for the eigenvectors of the Laplacian spectral embedding, are then given in Section 3. All of the limit results derived in this paper are for relatively dense graphs; more precisely, those for which the average degree grows at rate at least $\log^k(n)$ for some $k \geq 4$ as n, the number of vertices in the graph increases. We then discuss the implications of these limit results in Section 4; in particular Section 4.3 characterizes, via the notion of Chernoff statistical information, the *large-sample optimal* error rate of spectral clustering procedures. We demonstrate that neither embedding method dominates for the inference task of recovering block assignments in stochastic blockmodels. We conclude the paper with some brief remarks on potential extensions of the results presented herein. Proofs of stated results are given in the Appendix.

2. Background and setting. We first recall the notion of a random dot product graph [31].

DEFINITION 1. Let F be a distribution on a set $\mathcal{X} \subset \mathbb{R}^d$ satisfying $x^\top y \in [0, 1]$ for all $x, y \in \mathcal{X}$. We say $(\mathbf{X}, \mathbf{A}) \sim \mathrm{RDPG}(F)$ with sparsity factor $\rho_n \le 1$ if the following hold. Let $X_1, \ldots, X_n \sim F$ be independent and identically distributed random variables and define

$$(2.1) \qquad \mathbf{X} = [X_1 \mid \cdots \mid X_n]^\top \in \mathbb{R}^{n \times d} \quad \text{and} \quad \mathbf{P} = \rho_n \mathbf{X} \mathbf{X}^\top \in [0, 1]^{n \times n}.$$

The X_i are the *latent* positions for the random graph, that is, we do not observe $\mathbf{X}$, rather we observe only the matrix $\mathbf{A}$. The matrix $\mathbf{A} \in \{0, 1\}^{n \times n}$ is defined to be symmetric with all zeros on the diagonal such that for all $i < j$, conditioned on X_i, X_j the A_{ij} are independent and

$$(2.2) \qquad A_{ij} \sim \mathrm{Bernoulli}\big(\rho_n X_i^\top X_j\big),$$

namely,

$$(2.3) \qquad \mathbb{P}[\mathbf{A} \mid \mathbf{X}] = \prod_{i<j} \big(\rho_n X_i^\top X_j\big)^{A_{ij}} \big(1 - \rho_n X_i^\top X_j\big)^{(1-A_{ij})}.$$

REMARK. We note that nonidentifiability is an intrinsic property of random dot product graphs. More specifically, if $(\mathbf{X}, \mathbf{A}) \sim \mathrm{RDPG}(F)$ where F is a distribution on $\mathbb{R}^d$, then for any orthogonal transformation U, $(\mathbf{Y}, \mathbf{B}) \sim \mathrm{RDPG}(F \circ U)$ is identically distributed to $(\mathbf{X}, \mathbf{A})$; we write $F \circ U$ to denote the distribution of $Y = UX$ whenever $X \sim F$. Furthermore, there also exists a distribution F' on $\mathbb{R}^{d'}$ with $d' > d$ such that $(\mathbf{Y}, \mathbf{B}) \sim \mathrm{RDPG}(F')$ is identically distributed to $(\mathbf{X}, \mathbf{A})$. Nonidentifiability due to orthogonal transformations cannot be avoided given the observed $\mathbf{A}$. We avoid the other source of nonidentifiability by assuming throughout this paper that if $(\mathbf{X}, \mathbf{A}) \sim \mathrm{RDPG}(F)$ then F is nondegenerate, that is, $\mathbb{E}[XX^\top]$ is of full rank.

As an example of random dot product graphs, we could take $\mathcal{X}$ to be the unit simplex in $\mathbb{R}^d$ and let F be a mixture of Dirichlet distributions or logistic-normal distribution. Random dot product graphs are a specific example of latent position graphs or inhomogeneous random graphs [7, 18], in which each vertex is associated with a latent position X_i and, conditioned on the latent positions, the presence or absence of the edges in the graph are independent Bernoulli random variables where the probability of an edge between any two vertices with latent positions X_i and X_j is given by $\kappa(X_i, X_j)$ for some symmetric function κ. A random dot product graph on n vertices is also, when viewed as an induced subgraph of an infinite graph, an exchangeable random graph [16]. Random dot product graphs are related to stochastic block model graphs [19] and degree-corrected stochastic block model graphs [21]; for example, a stochastic blockmodel graph on K blocks

with a positive semidefinite block probability matrix $\mathbf{B}$ corresponds to a random dot product graph where F is a mixture of K point masses.

For a given matrix $\mathbf{M}$ with nonnegative entries, denote by $\mathcal{L}(\mathbf{M})$ the *normalized* Laplacian of $\mathbf{M}$ defined as

$$(2.4) \qquad \mathcal{L}(\mathbf{M}) = \big(\mathrm{diag}(\mathbf{M1})\big)^{-1/2}\mathbf{M}\big(\mathrm{diag}(\mathbf{M1})\big)^{-1/2},$$

where, given $z = (z_1, \ldots, z_n) \in \mathbb{R}^n$, $\mathrm{diag}(z)$ is the $n \times n$ diagonal matrix whose diagonal entries are the z_i's. Our definition of the normalized Laplacian is slightly different from that often found in the literature, for example, in [13, 36] the normalized Laplacian is $\mathbf{I} - \mathcal{L}(\mathbf{M})$. For the purpose of this paper, namely the notion of the Laplacian spectral embedding via the eigenvalues and eigenvectors of the normalized Laplacian, these two definitions of the normalized Laplacian are equivalent. We shall henceforth refer to $\mathcal{L}(\mathbf{M})$ as the Laplacian of $\mathbf{M}$, in contrast to the *combinatorial* Laplacian $\mathrm{diag}(\mathbf{M1}) - \mathbf{M}$ of $\mathbf{M}$. See [29] for a survey of the combinatorial Laplacian and its connection to graph theory.

DEFINITION 2 (Adjacency and Laplacian spectral embedding). Let $\mathbf{A}$ be a $n \times n$ adjacency matrix. Suppose the eigendecomposition of $\mathbf{A}$ is given by $\mathbf{A} = \sum_{i=1}^{n} \lambda_i \boldsymbol{u}_i \boldsymbol{u}_i^\top$ where $|\lambda_1| \geq |\lambda_2| \geq \ldots$ are the eigenvalues and $\boldsymbol{u}_1, \boldsymbol{u}_2, \ldots, \boldsymbol{u}_n$ are the corresponding orthonormal eigenvectors. Given a positive integer $d \leq n$, denote by $\mathbf{S_A} = \mathrm{diag}(|\lambda_1|, \ldots, |\lambda_d|)$ the diagonal matrix whose diagonal entries are the $|\lambda_1|, \ldots, |\lambda_d|$, and denote by $\mathbf{U_A}$ the $n \times d$ matrix whose columns are the corresponding eigenvectors $\boldsymbol{u}_1, \ldots, \boldsymbol{u}_d$. The *adjacency spectral embedding* (ASE) of $\mathbf{A}$ into $\mathbb{R}^d$ is then the $n \times d$ matrix $\hat{\mathbf{X}} = \mathbf{U_A}\mathbf{S_A}^{1/2}$. Similarly, let $\mathcal{L}(\mathbf{A})$ denote the normalized Laplacian of $\mathbf{A}$ and suppose the eigendecomposition of $\mathcal{L}(\mathbf{A})$ is given by $\mathcal{L}(\mathbf{A}) = \sum_{i=1}^{n} \tilde{\lambda}_i \tilde{\boldsymbol{u}}_i \tilde{\boldsymbol{u}}_i^\top$ where $|\tilde{\lambda}_1| \geq |\tilde{\lambda}_2| \geq \cdots \geq |\tilde{\lambda}_n| \geq 0$ are the eigenvalues and $\tilde{\boldsymbol{u}}_1, \tilde{\boldsymbol{u}}_2, \ldots, \tilde{\boldsymbol{u}}_n$ are the corresponding orthonormal eigenvectors. Then given a positive integer $d \leq n$, denote by $\tilde{\mathbf{S}}_{\mathbf{A}} = \mathrm{diag}(|\tilde{\lambda}_1|, \ldots, |\tilde{\lambda}_d|)$ the diagonal matrix whose diagonal entries are the $|\tilde{\lambda}_1|, \ldots, |\tilde{\lambda}_d|$ and denote by $\tilde{\mathbf{U}}_{\mathbf{A}}$ the $n \times d$ matrix whose columns are the eigenvectors $\tilde{\boldsymbol{u}}_1, \ldots, \tilde{\boldsymbol{u}}_d$. The *Laplacian spectral embedding* of $\mathbf{A}$ into $\mathbb{R}^d$ is then the $n \times d$ matrix $\check{\mathbf{X}} = \tilde{\mathbf{U}}_{\mathbf{A}}\tilde{\mathbf{S}}_{\mathbf{A}}^{1/2}$.

REMARK. Let $(\mathbf{X}, \mathbf{A}) \sim \mathrm{RDPG}(F)$ with sparsity factor ρ_n and suppose that the $d \times d$ matrix $\mathbb{E}[XX^\top]$ is of full-rank where $X \sim F$. The $n \times d$ matrix $\hat{\mathbf{X}}$, the adjacency spectral embedding $\hat{\mathbf{X}}$ of $\mathbf{A}$ into $\mathbb{R}^d$, can then be viewed as a consistent estimate of $\rho_n^{1/2}\mathbf{X}$. See [39] for a comprehensive overview of the consistency results and their implications for subsequent inference. On the other hand, as $\mathcal{L}(c\mathbf{M}) = \mathcal{L}(\mathbf{M})$ for any constant $c > 0$, the $n \times d$ matrix $\check{\mathbf{X}}$—the normalized Laplacian embedding of $\mathbf{A}$ into $\mathbb{R}^d$—can be viewed as a consistent estimate of $(\rho_n \mathrm{diag}(\mathbf{XX}^\top\mathbf{1}))^{-1/2}\rho_n^{1/2}\mathbf{X}$ which does not depend on the sparsity factor ρ_n. This is in contrast to the adjacency spectral embedding. For previous consistency re-

sults of $\check{\mathbf{X}}$ as an estimator for $\widetilde{\mathbf{X}}$ in various random graphs models, the reader is referred to [33, 34, 43] among others. However, to the best of our knowledge, Theorem 3.2—namely the distributional convergence of $\check{\mathbf{X}}$ to a mixture of multivariate normals in the context of random dot product graphs and stochastic blockmodel graphs—had not been established prior to this paper. Finally, we remark that $\hat{\mathbf{X}}$ and $\check{\mathbf{X}}$ are estimating quantities that, while closely related—$\mathbf{X}$ and $(\text{diag}(\mathbf{X}\mathbf{X}^\top \mathbf{1}))^{-1/2}\mathbf{X}$ are one-to-one transformations of each other—are in essence distinct "parametrizations" of random dot product graphs. It is therefore not entirely straightforward to facilitate a direct comparison of the "efficiency" of $\hat{\mathbf{X}}$ and $\check{\mathbf{X}}$ as estimators. This thus motivates our consideration of the f-divergences between the multivariate normals since the family of f-divergences satisfy the information processing lemma and are invariant with respect to invertible transformations.

REMARK. For simplicity, we shall assume henceforth that either $\rho_n = 1$ for all n, or that $\rho_n \to 0$ with $n\rho_n = \omega(\log^4 n)$. We note that for our purpose, namely the distributional limit results in Section 2.1 and Section 3, the assumption that $\rho_n = 1$ for all n is equivalent to the assumption that there exists a constant $c > 0$ such that $\rho_n \to c$. The assumption that $n\rho_n = \omega(\log^4 n)$ is so that we can apply the concentration inequalities from [25] to show concentration, in spectral norm, of $\mathbf{A}$ and $\mathcal{L}(\mathbf{A})$ around $\rho_n\mathbf{X}\mathbf{X}^\top$ and $\mathcal{L}(\mathbf{X}\mathbf{X}^\top)$, respectively.

2.1. *Limit results for the adjacency spectral embedding.* We now recall several limit results for $\hat{\mathbf{X}} - \mathbf{X}$. These results are restatements of earlier results from [4] and [41]. Theorem 2.2 as stated below is a slight generalization of Theorem 1 in [4]; the result in [4] assumed a more restrictive distinct eigenvalues assumption for the matrix $\mathbb{E}[XX^\top]$ where $X \sim F$. We shall assume throughout this paper that d, the rank of $\mathbb{E}[XX^\top]$ where $X \sim F$, is fixed and known a priori.

REMARK. For ease of exposition, many of the bounds in this paper are said to hold "with high probability." We say that a random variable $\xi \in \mathbb{R}$ is $O_{\mathbb{P}}(f(n))$ if, for any positive constant $c > 0$ there exists a n_0 and a constant $C > 0$ (both of which possibly depend on c) such that for all $n \geq n_0$, $|\xi| \leq Cf(n)$ with probability at least $1 - n^{-c}$; in addition, we say that a random variable $\xi \in \mathbb{R}$ is $o_{\mathbb{P}}(f(n))$ if for any positive constant $c > 0$ and any $\varepsilon > 0$ there exists a n_0 such that for all $n \geq n_0$, $|\xi| \leq \varepsilon f(n)$ with probability at least $1 - n^{-c}$. Similarly, when ξ is a random vector in $\mathbb{R}^d$ or a random matrix in $\mathbb{R}^{d_1 \times d_2}$, $\xi = O_{\mathbb{P}}(f(n))$ or $\xi = o_{\mathbb{P}}(f(n))$ if $\|\xi\| = O_{\mathbb{P}}(f(n))$ or $\|\xi\| = o_{\mathbb{P}}(f(n))$, respectively. Here, $\|x\|$ denotes the Euclidean norm of x when x is a vector and the spectral norm of x when x is a matrix. We write $\xi = \zeta + O_{\mathbb{P}}(f(n))$ or $\xi = \zeta + o_{\mathbb{P}}(f(n))$ if $\xi - \zeta = O_{\mathbb{P}}(f(n))$ or $\xi - \zeta = o_{\mathbb{P}}(f(n))$, respectively.

THEOREM 2.1. *Let* $(\mathbf{X}_n, \mathbf{A}_n) \sim \mathrm{RDPG}(F)$ *with sparsity factor* ρ_n. *Then there exists a* $d \times d$ *orthogonal matrix* $\mathbf{W}_n$ *and a* $n \times d$ *matrix* $\mathbf{R}_n$ *such that*

$$(2.5) \qquad \hat{\mathbf{X}}_n \mathbf{W}_n - \rho_n^{1/2} \mathbf{X}_n = \rho_n^{-1/2} (\mathbf{A}_n - \mathbf{P}_n) \mathbf{X}_n (\mathbf{X}_n^\top \mathbf{X}_n)^{-1} + \mathbf{R}_n.$$

Furthermore, $\|\mathbf{R}_n\| = O_\mathbb{P}((n\rho_n)^{-1/2})$. *Let* $\mu_F = \mathbb{E}[X_1] \in \mathbb{R}^d$ *and* $\Delta = \mathbb{E}[X_1 X_1^\top]$ *be a* $d \times d$ *matrix. If* $\rho_n = 1$ *for all* n, *then there exists a sequence of orthogonal matrices* $\mathbf{W}_n$ *such that*

$$(2.6) \qquad \|\hat{\mathbf{X}}_n \mathbf{W}_n - \mathbf{X}_n\|_F^2 \xrightarrow{\text{a.s.}} \operatorname{tr} \Delta^{-1} \big(\mathbb{E}[X_1 X_1^\top (X_1^\top \mu_F - X_1^\top \Delta X_1)]\big) \Delta^{-1}.$$

If, however, $\rho_n \to 0$ *and* $n\rho_n = \omega(\log^4 n)$, *then*

$$(2.7) \qquad \|\hat{\mathbf{X}}_n \mathbf{W}_n - \rho_n^{1/2} \mathbf{X}_n\|_F^2 \xrightarrow{\text{a.s.}} \operatorname{tr} \Delta^{-1} \big(\mathbb{E}[X_1 X_1^\top (X_1^\top \mu_F)]\big) \Delta^{-1}.$$

THEOREM 2.2. *Assume the setting and notation of Theorem* 2.1. *Denote by* $\hat{X}_i$ *the* i*th row of* $\hat{\mathbf{X}}_n$. *Let* $\Phi(z, \Sigma)$ *denote the cumulative distribution function for the multivariate normal, with mean zero and covariance matrix* Σ, *evaluated at* z. *Also denote by* $\Sigma(x)$ *the* $d \times d$ *matrix*

$$(2.8) \qquad \Sigma(x) = \Delta^{-1} \mathbb{E}[X_1 X_1^\top (x^\top X_1 - x^\top X_1 X_1^\top x)] \Delta^{-1}.$$

If $\rho_n = 1$ *for all* n, *then there exists a sequence of orthogonal matrices* $\mathbf{W}_n$ *such that for each fixed index* i *and any* $z \in \mathbb{R}^d$,

$$(2.9) \qquad \mathbb{P}\{\sqrt{n}(\mathbf{W}_n \hat{X}_i - X_i) \le z\} \longrightarrow \int \Phi(z, \Sigma(x)) \, dF(x).$$

That is, the sequence $\sqrt{n}(\mathbf{W}_n \hat{X}_i - X_i)$ *converges in distribution to a mixture of multivariate normals. We denote this mixture by* $\mathcal{N}(0, \tilde{\Sigma}(X_i))$. *If, however,* $\rho_n \to 0$ *and* $n\rho_n = \omega(\log^4 n)$ *then there exists a sequence of orthogonal matrices* $\mathbf{W}_n$ *such that*

$$(2.10) \qquad \mathbb{P}\{\sqrt{n}(\mathbf{W}_n \hat{X}_i - \rho_n^{1/2} X_i) \le z\} \longrightarrow \int \Phi(z, \Sigma_{o(1)}(x)) \, dF(x),$$

where $\Sigma_{o(1)}(x)$ *is the* $d \times d$ *matrix defined by*

$$(2.11) \qquad \Sigma_{o(1)}(x) = \Delta^{-1} \mathbb{E}[X_1 X_1^\top x^\top X_1] \Delta^{-1}.$$

An important corollary of Theorem 2.2 is the following result for when F is a mixture of K point masses, that is, $(\mathbf{X}, \mathbf{A}) \sim \mathrm{RDPG}(F)$ is a K-block stochastic blockmodel graph. Then for any fixed index i, the event that X_i is assigned to block $k \in \{1, 2, \ldots, K\}$ has nonzero probability, and hence one can conditioned on the block assignment of X_i to show that the conditional distribution of $\sqrt{n}(\mathbf{W}_n \hat{X}_i - X_i)$ converges to a multivariate normal. This is in contrast to the unconditional distribution being a mixture of multivariate normals as in equation (2.9) and equation (2.10).

COROLLARY 2.3. *Assume the setting and notation of Theorem* 2.1 *and let*

$$F = \sum_{k=1}^{K} \pi_k \delta_{v_k}, \qquad \pi_1, \ldots, \pi_K > 0, \sum_k \pi_k = 1$$

be a mixture of K point masses in $\mathbb{R}^d$ where δ_{v_k} is the Dirac delta measure at v_k. Then if $\rho_n \equiv 1$, there exists a sequence of orthogonal matrices $\mathbf{W}_n$ such that for any fixed index i,

$$(2.12) \qquad \mathbb{P}\{\sqrt{n}(\mathbf{W}_n \hat{X}_i - X_i) \le z \mid X_i = v_k\} \longrightarrow \Phi(z, \Sigma_k),$$

where $\Sigma_k = \Sigma(v_k)$ is as defined in equation (2.8). *If $\rho_n \to 0$ and $n\rho_n = \omega(\log^4(n))$ as $n \to \infty$, then the sequence of orthogonal matrices $\mathbf{W}_n$ satisfies*

$$(2.13) \qquad \mathbb{P}\{\sqrt{n}(\mathbf{W}_n \hat{X}_i - \rho_n^{1/2} X_i) \le z \mid X_i = v_k\} \longrightarrow \Phi(z, \Sigma_{o(1),k}),$$

where $\Sigma_{o(1),k} = \Sigma_{o(1)}(v_k)$ is as defined in equation (2.11).

3. Limit results for Laplacian spectral embedding. We now present the main technical results of this paper, namely analogues of the limit results in Section 2.1 for the Laplacian spectral embedding.

THEOREM 3.1. *Let $(\mathbf{A}_n, \mathbf{X}_n) \sim \text{RDPG}(F)$ for $n \ge 1$ be a sequence of random dot product graphs with sparsity factors $(\rho_n)_{n\ge1}$. Denote by $\mathbf{D}_n$ and $\mathbf{T}_n$ the $n \times n$ diagonal matrices $\text{diag}(\mathbf{A}_n \mathbf{1})$ and $\text{diag}(\rho_n \mathbf{X}_n \mathbf{X}_n^\top \mathbf{1})$, respectively, that is, the diagonal entries of $\mathbf{D}_n$ are the vertex degrees of $\mathbf{A}_n$ and the diagonal entries of $\mathbf{T}_n$ are the* expected *vertex degrees. Let $\tilde{\mathbf{X}}_n = \rho_n^{1/2} \mathbf{T}_n^{-1/2} \mathbf{X}_n = \text{diag}(\mathbf{X}_n \mathbf{X}_n^\top \mathbf{1})^{-1/2} \mathbf{X}_n$. Then for any n, there exists a $d \times d$ orthogonal matrix $\mathbf{W}_n$ and a $n \times d$ matrix $\mathbf{R}_n$ such that $\zeta_n := (\check{\mathbf{X}}_n \mathbf{W}_n - \tilde{\mathbf{X}}_n)$ satisfies*

$$(3.1) \qquad \zeta_n = \mathbf{T}_n^{-1/2}(\mathbf{A}_n - \mathbf{P}_n)\mathbf{T}_n^{-1/2}\tilde{\mathbf{X}}_n(\tilde{\mathbf{X}}_n^\top \tilde{\mathbf{X}}_n)^{-1} + \frac{1}{2}(\mathbf{I} - \mathbf{D}_n \mathbf{T}_n^{-1})\tilde{\mathbf{X}}_n + \mathbf{R}_n.$$

Furthermore, $\|\mathbf{R}_n\|_F = O_\mathbb{P}((n\rho_n)^{-1})$, that is, $\|\mathbf{R}_n\|/\|\zeta_n\| \xrightarrow{\text{a.s.}} 0$ as $n \to \infty$. Define the following quantities:

$$\mu = \mathbb{E}[X_1] \in \mathbb{R}^d; \qquad \tilde{\mu} = \mathbb{E}\left[\frac{X_1}{X_1^\top \mu}\right] \in \mathbb{R}^d;$$

$$(3.2)$$

$$\tilde{\Delta} = \mathbb{E}\left[\frac{X_1 X_1^\top}{X_1^\top \mu}\right] \in \mathbb{R}^{d\times d}; \quad and$$

$$(3.3) \qquad g(X_1, X_2) = \left(\frac{\tilde{\Delta}^{-1} X_1}{X_1^\top \mu} - \frac{X_2}{2X_2^\top \mu}\right)\left(\frac{\tilde{\Delta}^{-1} X_1}{X_1^\top \mu} - \frac{X_2}{2X_2^\top \mu}\right)^\top \in \mathbb{R}^{d\times d}.$$

If $\rho_n \equiv 1$, then the sequence of orthogonal matrices $(\mathbf{W}_n)_{n\ge1}$ satisfies

$$(3.4) \qquad n\|\check{\mathbf{X}}_n \mathbf{W}_n - \tilde{\mathbf{X}}_n\|_F^2 \xrightarrow{\text{a.s.}} \text{tr}\, \mathbb{E}\left[g(X_1, X_2) \frac{X_1^\top X_2 - X_1^\top X_2 X_2^\top X_1}{X_2^\top \mu}\right],$$

where the expectation in equation (3.4) *is taken with respect to* X_1 *and* X_2 *being i.i.d. drawn according to* F. *Equivalently, with* $\Delta = \mathbb{E}[X_1 X_1^\top]$,

$$n \| \check{\mathbf{X}}_n \mathbf{W}_n - \widetilde{\mathbf{X}}_n \|_F^2 \xrightarrow{\text{a.s.}} \operatorname{tr} \mathbb{E}\left[\frac{\widetilde{\Delta}^{-2} X_1 X_1^\top (X_1^\top \widetilde{\mu} - X_1^\top \widetilde{\Delta} X_1)}{(X_1^\top \mu)^2} - \frac{3 X_1 X_1^\top}{4(X_1^\top \mu)^2} \right]$$

$$+ \operatorname{tr} \mathbb{E}\left[\frac{\widetilde{\Delta}^{-1} X_1 X_1^\top X_2 X_2^\top (X_1^\top X_2)}{X_1^\top \mu (X_2^\top \mu)^2} - \frac{X_1 X_1^\top (X_1^\top \Delta X_1)}{4(X_1^\top \mu)^3} \right].$$

If $\rho_n \to 0$ *and* $n\rho_n = \omega(\log^4 n)$, *then the sequence* $(\mathbf{W}_n)_{n \geq 1}$ *satisfies*

$$(3.5) \qquad n\rho_n \| \check{\mathbf{X}} \mathbf{W}_n - \widetilde{\mathbf{X}}_n \|_F^2 \xrightarrow{\text{a.s.}} \operatorname{tr} \mathbb{E}\left[\frac{\widetilde{\Delta}^{-2} X_1 X_1^\top (X_1^\top \widetilde{\mu})}{(X_1^\top \mu)^2} - \frac{3 X_1 X_1^\top}{4(X_1^\top \mu)^2} \right].$$

As a companion of Theorem 3.1, we have the following result on the asymptotic normality of the rows of $\check{\mathbf{X}}_n \mathbf{W}_n - \widetilde{\mathbf{X}}_n$.

THEOREM 3.2. *Assume the setting and notation of Theorem* 3.1. *Denote by* $\check{X}_i$ *and* $\widetilde{X}_i$ *the* i*th row of* $\check{\mathbf{X}}_n$ *and* $\widetilde{\mathbf{X}}_n$, *respectively. We note that* $\widetilde{X}_i = \frac{X_i}{\sqrt{\sum_j X_i^\top X_j}}$.
Also denote by $\widetilde{\Sigma}(x)$ *the* $d \times d$ *matrix*

$$(3.6) \qquad \mathbb{E}\left[\left(\frac{\widetilde{\Delta}^{-1} X_1}{X_1^\top \mu} - \frac{x}{2x^\top \mu} \right) \left(\frac{X_1^\top \widetilde{\Delta}^{-1}}{X_1^\top \mu} - \frac{x^\top}{2x^\top \mu} \right) \frac{(x^\top X_1 - x^\top X_1 X_1^\top x)}{x^\top \mu} \right].$$

If $\rho_n \equiv 1$, *then there exists a sequence of* $d \times d$ *orthogonal matrices* $\mathbf{W}_n$ *such that for each fixed index* i *and any* $z \in \mathbb{R}^d$,

$$(3.7) \qquad \mathbb{P}\left\{ n\left(\mathbf{W}_n \check{X}_i - \frac{X_i}{\sqrt{\sum_j X_i^\top X_j}} \right) \leq z \right\} \longrightarrow \int \Phi(z, \widetilde{\Sigma}(x)) \, dF(x).$$

That is, the sequence $n(\mathbf{W}_n \check{X}_i - \widetilde{X}_i)$ *converges in distribution to a mixture of multivariate normals. We denote this mixture by* $\mathcal{N}(0, \widetilde{\Sigma}(X_i))$. *If* $\rho_n \to 0$ *and* $n\rho_n = \omega(\log^4 n)$, *then there exists a sequence of* $d \times d$ *orthogonal matrices* $\mathbf{W}_n$ *such that*

$$(3.8) \qquad \mathbb{P}\left\{ n\rho_n^{1/2}\left(\mathbf{W}_n \check{X}_i - \frac{X_i}{\sqrt{\sum_j X_i^\top X_j}} \right) \leq z \right\} \longrightarrow \int \Phi(z, \widetilde{\Sigma}_{o(1)}(x)) \, dF(x),$$

where $\widetilde{\Sigma}_{o(1)}(x)$ *is the* $d \times d$ *matrix defined by*

$$(3.9) \qquad \widetilde{\Sigma}_{o(1)}(x) = \mathbb{E}\left[\left(\frac{\widetilde{\Delta}^{-1} X_1}{X_1^\top \mu} - \frac{x}{2x^\top \mu} \right) \left(\frac{X_1^\top \widetilde{\Delta}^{-1}}{X_1^\top \mu} - \frac{x^\top}{2x^\top \mu} \right) \frac{x^\top X_1}{x^\top \mu} \right].$$

The proofs of Theorem 3.1 and Theorem 3.2 are given in Section B. We end this section by stating the conditional distribution of $n\rho_n(\mathbf{W}_n \check{X}_i - \widetilde{X}_i)$ when $(\mathbf{X}, \mathbf{A}) \sim$ RDPG(F) is a K-block stochastic blockmodel graph.

COROLLARY 3.3. *Assume the setting and notation of Theorem* 3.1 *and let*

$$F = \sum_{k=1}^{K} \pi_k \delta_{v_k}, \qquad \pi_1, \ldots, \pi_K > 0, \sum_k \pi_k = 1$$

be a mixture of K point masses in $\mathbb{R}^d$. *Then if* $\rho_n \equiv 1$, *there exists a sequence of* $d \times d$ *orthogonal matrices* $\mathbf{W}_n$ *such that for any fixed index i,*

$$(3.10) \qquad \mathbb{P}\left\{ n\left(\mathbf{W}_n \check{X}_i - \frac{v_k}{\sqrt{\sum_l n_l v_k^\top v_l}} \right) \le z \,\middle|\, X_i = v_k \right\} \longrightarrow \mathbf{\Phi}(z, \widetilde{\Sigma}_k),$$

where $\widetilde{\Sigma}_k = \widetilde{\Sigma}(v_k)$ *is as defined in equation* (3.6) *and* n_k *for* $k \in \{1, 2, \ldots, K\}$ *denote the number of vertices in* $\mathbf{A}_n$ *that are assigned to block k. If instead* $\rho_n \to 0$ *and* $n\rho_n = \omega(\log^4(n))$ *as* $n \to \infty$, *then the sequence of orthogonal matrices* $\mathbf{W}_n$ *satisfies*

$$(3.11) \qquad \mathbb{P}\left\{ n\rho_n^{1/2}\left(\mathbf{W}_n \check{X}_i - \frac{v_k}{\sqrt{\sum_l n_l v_k^\top v_l}} \right) \le z \,\middle|\, X_i = v_k \right\} \longrightarrow \mathbf{\Phi}(z, \widetilde{\Sigma}_{o(1),k}),$$

where $\widetilde{\Sigma}_{o(1),k} = \widetilde{\Sigma}_{o(1)}(v_k)$ *is as defined in equation* (3.9).

REMARK. As a special case of Corollary 3.3, we have that if $\mathbf{A}$ is an Erdős–Rényi graph on n vertices with edge probability p^2—which corresponds to a random dot product graph where the latent positions are identically p—then for each fixed index i, the normalized Laplacian embedding satisfies

$$n\left(\check{X}_i - \frac{1}{\sqrt{n}} \right) \xrightarrow{\text{d}} \mathcal{N}\left(0, \frac{1 - p^2}{4p^2} \right).$$

Recall that $\check{X}_i$ is proportional to $1/\sqrt{d_i}$ where d_i is the degree of the ith vertex. On the other hand, the adjacency spectral embedding satisfies

$$\sqrt{n}(\hat{X}_i - p) \xrightarrow{\text{d}} \mathcal{N}(0, 1 - p^2).$$

As another example, let $\mathbf{A}$ be a stochastic blockmodel graph with block probabilities matrix $\mathbf{B} = \left[\begin{smallmatrix} p^2 & pq \\ pq & q^2 \end{smallmatrix} \right]$ and block assignment probabilities $(\pi, 1 - \pi)$. Since $\mathbf{B}$ has rank 1, this model corresponds to a random dot product graph where the latent positions are either p with probability π or q with probability $1 - \pi$—then for each fixed index i, the normalized Laplacian embedding satisfies

$$(3.12) \qquad \begin{aligned} & n\left(\check{X}_i - \frac{p}{\sqrt{n_1 p^2 + n_2 pq}} \right) \\ & \xrightarrow{\text{d}} \mathcal{N}\left(0, \frac{\pi p(1 - p^2) + (1 - \pi)q(1 - pq)}{4(\pi p + (1 - \pi)q)^3} \right) \qquad \text{if } X_i = p, \end{aligned}$$

(3.13)
$$n\left(\check{X}_i - \frac{q}{\sqrt{n_1 pq + n_2 q^2}}\right)$$
$$\xrightarrow{d} \mathcal{N}\left(0, \frac{\pi p(1 - pq) + (1 - \pi)q(1 - q^2)}{4(\pi p + (1 - \pi)q)^3}\right) \quad \text{if } X_i = q,$$

where n_1 and $n_2 = n - n_1$ are the number of vertices of **A** with latent positions p and q. The adjacency spectral embedding meanwhile satisfies

(3.14)
$$\sqrt{n}(\hat{X}_i - p)$$
$$\xrightarrow{d} \mathcal{N}\left(0, \frac{\pi p^4(1 - p^2) + (1 - \pi)pq^3(1 - pq)}{(\pi p^2 + (1 - \pi)q^2)^2}\right) \quad \text{if } X_i = p,$$

(3.15)
$$\sqrt{n}(\hat{X}_i - q)$$
$$\xrightarrow{d} \mathcal{N}\left(0, \frac{\pi p^3 q(1 - pq) + (1 - \pi)q^4(1 - q^2)}{(\pi p^2 + (1 - \pi)q^2)^2}\right) \quad \text{if } X_i = q.$$

REMARK. We note that the quantity n_k appears in equation (3.12) and equation (3.13). Replacing n_k by $n\pi_k$ in equation (3.12) and equation (3.13) is, however, not straightforward. More precisely, letting $\zeta = \frac{np}{\sqrt{n_1 p^2 + n_2 pq}} - \frac{np}{\sqrt{n\pi p^2 + n(1-\pi)pq}}$ we have

$$\zeta = \frac{np(\sqrt{n\pi p^2 + n(1 - \pi)pq} - \sqrt{n_1 p^2 + n_2 pq})}{\sqrt{n_1 p^2 + n_2 pq}\sqrt{n\pi p^2 + n(1 - \pi)pq}}$$
$$= \left(np(n\pi p^2 + n(1 - \pi)pq - n_1 p^2 - n_2 pq)\right)$$
$$\Big/ \left(\left(\sqrt{n\pi p^2 + n(1 - \pi)pq} + \sqrt{n_1 p^2 + n_2 pq}\right)\right.$$
$$\left.\times \sqrt{n_1 p^2 + n_2 pq}\sqrt{n\pi p^2 + n(1 - \pi)pq}\right)$$
$$= \left(np(n\pi - n_1)(p^2 - pq)\right)$$
$$\Big/ \left(\left(\sqrt{n\pi p^2 + n(1 - \pi)pq} + \sqrt{n_1 p^2 + n_2 pq}\right)\right.$$
$$\left.\times \sqrt{n_1 p^2 + n_2 pq}\sqrt{n\pi p^2 + n(1 - \pi)pq}\right).$$

By the strong law of large numbers and Slutsky's theorem, we have

$$\frac{n^{3/2}}{(\sqrt{n\pi p^2 + n(1-\pi)pq} + \sqrt{n_1 p^2 + n_2 pq})\sqrt{n_1 p^2 + n_2 pq}\sqrt{n\pi p^2 + n(1-\pi)pq}}$$

$$\xrightarrow{\text{a.s.}} \frac{1}{2(p^2 + pq)^{3/2}}.$$

We note that, as the n_k are assumed to be random variables, that is, we are not conditioning on the block sizes, by the central limit theorem we have

$$\frac{1}{\sqrt{n}}(n\pi - n_1) \xrightarrow{\text{d}} \mathcal{N}(0, \pi(1-\pi)).$$

Therefore, by Slutsky's theorem, we have

$$\zeta = \frac{np}{\sqrt{n_1 p^2 + n_2 pq}} - \frac{np}{\sqrt{n\pi p^2 + n(1-\pi)pq}} \xrightarrow{\text{d}} \mathcal{N}\left(0, \frac{\pi(1-\pi)p(p-q)^2}{4(p+q)^3}\right).$$

To replace n_k by $n\pi_k$ in equation (3.12) and equation (3.13), we thus need to include the random term ζ. While we surmise that this can be done, and more generally, equation (3.10) and equation (3.11) can be modified to handle the randomness in n_k, we shall not do so in this paper.

3.1. *Proofs sketch for Theorem* 3.1 *and Theorem* 3.2. We present in this subsection a sketch of the main ideas in the proofs of Theorem 3.1 and Theorem 3.2; the detailed proofs are given in Section B of the Appendix. We start with the motivation behind equation (3.1). Given $\widetilde{\mathbf{X}}_n$, the entries of the right-hand side of equation (3.1), except for the term $\mathbf{R}_n$, can be expressed explicitly in terms of linear combinations of the entries $a_{ij} - p_{ij}$ of $\mathbf{A}_n - \mathbf{P}_n$. This is in contrast with the left-hand side of equation (3.1) which depends on the quantities $\widetilde{\mathbf{U}}_{\mathbf{A}}$ and $\widetilde{\mathbf{S}}_{\mathbf{A}}$ (recall Definition 2); since the quantities $\widetilde{\mathbf{U}}_{\mathbf{A}}$ and $\widetilde{\mathbf{S}}_{\mathbf{A}}$ cannot be expressed explicitly in terms of the entries of $\mathbf{A}_n$ and $\mathbf{P}_n$, we conclude that the right-hand side of equation (3.1) is simpler to analyze. From equation (3.1), the squared Frobenius norm $n\rho_n \|\check{\mathbf{X}}_n \mathbf{W}_n - \widetilde{\mathbf{X}}_n\|_F^2$ is

$$n\rho_n \left\| \mathbf{T}_n^{-1/2}(\mathbf{A}_n - \mathbf{P}_n)\mathbf{T}_n^{-1/2}\widetilde{\mathbf{X}}_n(\widetilde{\mathbf{X}}_n^\top \widetilde{\mathbf{X}}_n)^{-1} + \frac{1}{2}(\mathbf{I} - \mathbf{D}_n \mathbf{T}_n^{-1})\widetilde{\mathbf{X}}_n \right\|_F^2 + O_{\mathbb{P}}((n\rho_n)^{-1/2}).$$

Then conditional on $\mathbf{P}_n$, the above expression is, up to the term of order $O_{\mathbb{P}}((n\rho_n)^{-1/2})$, a function of the *independent* random variables $\{a_{ij} - p_{ij}\}_{i<j}$. We can then apply concentration inequalities such as those in [8] to show that the squared Frobenius norm $n\rho_n \|\check{\mathbf{X}}_n \mathbf{W}_n - \widetilde{\mathbf{X}}_n\|_F^2$ is, conditional on $\mathbf{P}_n$, concentrated around its expectation. Here, the expectation is taken with respect to the random entries of $\mathbf{A}_n$. Equation (3.4) and equation (3.5) then follows by direct evaluation of this expectation, for the case when $\rho_n \equiv 1$ and for when $\rho_n \to 0$, respectively.

Once equation (3.1) is established, we can derive Theorem 3.2 as follows. Let ξ_i denotes the ith row of $n\rho_n^{1/2}(\mathbf{W}_n\check{\mathbf{X}}_n - \widetilde{\mathbf{X}}_n)$ and let r_i denotes the ith row of $\mathbf{R}_n$. Equation (3.1) then implies

$$
\begin{aligned}
\xi_i &= (\widetilde{\mathbf{X}}_n^\top \widetilde{\mathbf{X}}_n)^{-1} \frac{n\rho_n^{1/2}}{\sqrt{t_i}} \left(\sum_j \frac{a_{ij} - p_{ij}}{\sqrt{t_j}} \widetilde{X}_j \right) + \frac{n\rho_n^{1/2}(t_i - d_i)}{2t_i} \widetilde{X}_i + n\rho_n^{1/2} r_i \\
&= (\widetilde{\mathbf{X}}_n^\top \widetilde{\mathbf{X}}_n)^{-1} \frac{\sqrt{n\rho_n}}{\sqrt{t_i}} \left(\sum_j \frac{\sqrt{n\rho_n}(a_{ij} - p_{ij})X_j}{t_j} \right) \\
&\quad - \frac{n\rho_n X_i}{2t_i^{3/2}} \sum_j (a_{ij} - p_{ij}) + n\rho_n^{1/2} r_i \\
&= \frac{\sqrt{n\rho_n}}{\sqrt{t_i}} \sum_j \frac{(a_{ij} - p_{ij})}{\sqrt{n\rho_n}} \left(\frac{(\widetilde{\mathbf{X}}_n^\top \widetilde{\mathbf{X}}_n)^{-1} X_j}{t_j/(n\rho_n)} - \frac{X_i}{2t_i/(n\rho_n)} \right) + n\rho_n^{1/2} r_i.
\end{aligned}
$$

We then show that $n\rho_n^{1/2} r_i \xrightarrow{\mathrm{d}} 0$. Indeed, there are n rows in $\mathbf{R}_n$ and $\|\mathbf{R}_n\|_F = O((n\rho_n)^{-1})$; hence, on average, for each index i, $\|r_i\|^2 = O_\mathbb{P}(n^{-3}\rho_n^{-2})$. Furthermore, $t_i/(n\rho_n) = \sum_j X_i^\top X_j / n \xrightarrow{\text{a.s.}} X_i^\top \mu$ as $n \to \infty$. Finally, $\widetilde{\mathbf{X}}_n^\top \widetilde{\mathbf{X}}_n = \sum_i (X_i X_i^\top / (\sum_j X_i^\top X_j))$ which, as we show in Section B, converges to $\widetilde{\Delta} = \mathbb{E}[\frac{X_1 X_1^\top}{X_1^\top \mu}]$ as $n \to \infty$. We therefore have, after additional manipulations, that

$$
\begin{aligned}
\xi_i &= \frac{\sqrt{n\rho_n}}{\sqrt{t_i}} \sum_j \frac{(a_{ij} - p_{ij})}{\sqrt{n\rho_n}} \left(\frac{\widetilde{\Delta}^{-1} X_j}{X_j^\top \mu} - \frac{X_i}{2X_i^\top \mu} \right) + o_\mathbb{P}(1) \\
&= \frac{\sqrt{n\rho_n}}{\sqrt{t_i}} \sum_j \frac{(a_{ij} - \rho_n X_i^\top X_j)}{\sqrt{n\rho_n}} \left(\frac{\widetilde{\Delta}^{-1} X_j}{X_j^\top \mu} - \frac{X_i}{2X_i^\top \mu} \right) + o_\mathbb{P}(1).
\end{aligned}
$$

Then conditioning on $X_i = x$, the above expression for ξ_i is roughly a sum of independent and identically distributed mean 0 random variables. The multivariate central limit theorem can then be applied to the above expression for ξ_i, thereby yielding Theorem 3.2.

We now sketch the derivation of equation (3.1). For simplicity, we ignore the subscript n in the matrices $\mathbf{A}_n$, $\mathbf{X}_n$, $\mathbf{P}_n$ and related matrices. First, consider the following expression:

$$
\begin{aligned}
\widetilde{\mathbf{U}}_\mathbf{A} \widetilde{\mathbf{S}}_\mathbf{A}^{1/2} - \widetilde{\mathbf{U}}_\mathbf{P} \widetilde{\mathbf{S}}_\mathbf{P}^{1/2} \widetilde{\mathbf{U}}_\mathbf{P}^\top \widetilde{\mathbf{U}}_\mathbf{A} &= \mathcal{L}(\mathbf{A}) \widetilde{\mathbf{U}}_\mathbf{A} \widetilde{\mathbf{S}}_\mathbf{A}^{-1/2} - \mathcal{L}(\mathbf{P}) \widetilde{\mathbf{U}}_\mathbf{P} \widetilde{\mathbf{S}}_\mathbf{P}^{-1/2} \widetilde{\mathbf{U}}_\mathbf{P}^\top \widetilde{\mathbf{U}}_\mathbf{A} \\
&= \mathcal{L}(\mathbf{A}) \widetilde{\mathbf{U}}_\mathbf{A} \widetilde{\mathbf{U}}_\mathbf{A}^\top \widetilde{\mathbf{U}}_\mathbf{A} \widetilde{\mathbf{S}}_\mathbf{A}^{-1/2} - \mathcal{L}(\mathbf{P}) \widetilde{\mathbf{U}}_\mathbf{P} \widetilde{\mathbf{S}}_\mathbf{P}^{-1/2} \widetilde{\mathbf{U}}_\mathbf{P}^\top \widetilde{\mathbf{U}}_\mathbf{A}.
\end{aligned}
$$

Now $\mathcal{L}(\mathbf{A})$ is "concentrated" around $\mathcal{L}(\mathbf{P})$, that is, $\|\mathcal{L}(\mathbf{A}) - \mathcal{L}(\mathbf{P})\| = O_\mathbb{P}((n\rho_n)^{-1/2})$ (see Theorem 2 in [25]). Since $\|\mathcal{L}(\mathbf{P})\| = \Theta(1)$ and the nonzero eigenvalues of $\mathcal{L}(\mathbf{P})$ are all of order $\Theta(1)$, this implies, by the Davis–Kahan

theorem, that the eigenspace spanned by the d largest eigenvalues of $\mathcal{L}(\mathbf{A})$ is "close" to that spanned by the d largest eigenvalues of $\mathcal{L}(\mathbf{P})$. More precisely, $\widetilde{\mathbf{U}}_{\mathbf{A}}\widetilde{\mathbf{U}}_{\mathbf{A}}^{\top} = \widetilde{\mathbf{U}}_{\mathbf{P}}\widetilde{\mathbf{U}}_{\mathbf{P}}^{\top} + O_{\mathbb{P}}((n\rho_n)^{-1/2})$ and

$$\widetilde{\mathbf{U}}_{\mathbf{A}}\widetilde{\mathbf{S}}_{\mathbf{A}}^{1/2} - \widetilde{\mathbf{U}}_{\mathbf{P}}\widetilde{\mathbf{S}}_{\mathbf{P}}^{1/2}\widetilde{\mathbf{U}}_{\mathbf{P}}^{\top}\widetilde{\mathbf{U}}_{\mathbf{A}} = \mathcal{L}(\mathbf{A})\widetilde{\mathbf{U}}_{\mathbf{P}}\widetilde{\mathbf{U}}_{\mathbf{P}}^{\top}\widetilde{\mathbf{U}}_{\mathbf{A}}\widetilde{\mathbf{S}}_{\mathbf{A}}^{-1/2} - \mathcal{L}(\mathbf{P})\widetilde{\mathbf{U}}_{\mathbf{P}}\widetilde{\mathbf{S}}_{\mathbf{P}}^{-1/2}\widetilde{\mathbf{U}}_{\mathbf{P}}^{\top}\widetilde{\mathbf{U}}_{\mathbf{A}}$$
$$+ O_{\mathbb{P}}((n\rho_n)^{-1}).$$

We then consider the terms $\widetilde{\mathbf{S}}_{\mathbf{P}}^{-1/2}\widetilde{\mathbf{U}}_{\mathbf{P}}^{\top}\widetilde{\mathbf{U}}_{\mathbf{A}}$ and $\widetilde{\mathbf{U}}_{\mathbf{P}}^{\top}\widetilde{\mathbf{U}}_{\mathbf{A}}\widetilde{\mathbf{S}}_{\mathbf{A}}^{-1/2}$. Since $\widetilde{\mathbf{U}}_{\mathbf{P}}$ and $\widetilde{\mathbf{U}}_{\mathbf{A}}$ both have orthonormal columns, $\widetilde{\mathbf{U}}_{\mathbf{A}}\widetilde{\mathbf{U}}_{\mathbf{A}}^{\top} = \widetilde{\mathbf{U}}_{\mathbf{P}}\widetilde{\mathbf{U}}_{\mathbf{P}}^{\top} + O_{\mathbb{P}}((n\rho_n)^{-1/2})$ implies that there exists an orthogonal matrix $\mathbf{W}^{*}$ such that $\widetilde{\mathbf{U}}_{\mathbf{P}}^{\top}\widetilde{\mathbf{U}}_{\mathbf{A}} = \mathbf{W}^{*} + O_{\mathbb{P}}((n\rho_n)^{-1})$ (see Proposition B.2). Furthermore, $\mathbf{W}^{*}$ satisfies an important property, namely that $\mathbf{W}^{*}\widetilde{\mathbf{S}}_{\mathbf{A}}^{-1/2} - \widetilde{\mathbf{S}}_{\mathbf{P}}^{-1/2}\mathbf{W}^{*} = O_{\mathbb{P}}((n\rho_n)^{-1})$. (see Lemma B.3). We can thus juxtapose $\widetilde{\mathbf{U}}_{\mathbf{P}}^{\top}\widetilde{\mathbf{U}}_{\mathbf{A}}$ and $\widetilde{\mathbf{S}}_{\mathbf{A}}^{-1/2}$ in the above expression and replace $\widetilde{\mathbf{U}}_{\mathbf{P}}^{\top}\widetilde{\mathbf{U}}_{\mathbf{A}}$ by the orthogonal matrix $\mathbf{W}^{*}$, thereby yielding

$$\widetilde{\mathbf{U}}_{\mathbf{A}}\widetilde{\mathbf{S}}_{\mathbf{A}}^{1/2} - \widetilde{\mathbf{U}}_{\mathbf{P}}\widetilde{\mathbf{S}}_{\mathbf{P}}^{1/2}\mathbf{W}^{*} = (\mathcal{L}(\mathbf{A}) - \mathcal{L}(\mathbf{P}))\widetilde{\mathbf{U}}_{\mathbf{P}}\widetilde{\mathbf{S}}_{\mathbf{P}}^{-1/2}\mathbf{W}^{*} + O_{\mathbb{P}}((n\rho_n)^{-1}).$$

As $\widetilde{\mathbf{X}}\widetilde{\mathbf{X}}^{\top} = \mathcal{L}(\mathbf{P}) = \widetilde{\mathbf{U}}_{\mathbf{P}}\widetilde{\mathbf{S}}_{\mathbf{P}}^{1/2}\widetilde{\mathbf{U}}_{\mathbf{P}}^{\top}$, we have $\widetilde{\mathbf{X}} = \widetilde{\mathbf{U}}_{\mathbf{P}}^{1/2}\widetilde{\mathbf{S}}_{\mathbf{P}}\widetilde{\mathbf{W}}$ for some orthogonal matrix $\widetilde{\mathbf{W}}$. Therefore,

$$\widetilde{\mathbf{U}}_{\mathbf{A}}\widetilde{\mathbf{S}}_{\mathbf{A}}^{1/2} - \widetilde{\mathbf{X}}\widetilde{\mathbf{W}}^{\top}\mathbf{W}^{*}$$
$$= (\mathcal{L}(\mathbf{A}) - \mathcal{L}(\mathbf{P}))\widetilde{\mathbf{U}}_{\mathbf{P}}\widetilde{\mathbf{S}}_{\mathbf{P}}^{-1/2}\mathbf{W}^{*} + O_{\mathbb{P}}((n\rho_n)^{-1})$$
$$= (\mathcal{L}(\mathbf{A}) - \mathcal{L}(\mathbf{P}))\widetilde{\mathbf{U}}_{\mathbf{P}}\widetilde{\mathbf{S}}_{\mathbf{P}}^{1/2}\widetilde{\mathbf{W}}\widetilde{\mathbf{W}}^{\top}\widetilde{\mathbf{S}}_{\mathbf{P}}^{-1}\widetilde{\mathbf{W}}\widetilde{\mathbf{W}}^{\top}\mathbf{W}^{*} + O_{\mathbb{P}}((n\rho_n)^{-1})$$
$$= (\mathcal{L}(\mathbf{A}) - \mathcal{L}(\mathbf{P}))\widetilde{\mathbf{X}}(\widetilde{\mathbf{X}}^{\top}\widetilde{\mathbf{X}})^{-1}\widetilde{\mathbf{W}}^{\top}\mathbf{W}^{*} + O_{\mathbb{P}}((n\rho_n)^{-1}).$$

Equivalently,

$$(3.16) \qquad \widetilde{\mathbf{U}}_{\mathbf{A}}\widetilde{\mathbf{S}}_{\mathbf{A}}^{1/2}(\mathbf{W}^{*})^{\top}\widetilde{\mathbf{W}} - \widetilde{\mathbf{X}} = (\mathcal{L}(\mathbf{A}) - \mathcal{L}(\mathbf{P}))\widetilde{\mathbf{X}}(\widetilde{\mathbf{X}}^{\top}\widetilde{\mathbf{X}})^{-1} + O_{\mathbb{P}}((n\rho_n)^{-1}).$$

The right-hand side of equation (3.16) can be written explicitly in terms of the entries of $\mathbf{A}$. However, since $\mathcal{L}(\mathbf{A}) = \mathbf{D}^{-1/2}\mathbf{A}\mathbf{D}^{-1/2}$ and $\mathbf{D} = \mathrm{diag}(\mathbf{A}\mathbf{1})$, the entries of the right-hand side of equation (3.16) are not linear/affine combinations of the entries of $\mathbf{A}$. Nevertheless, by a Taylor-series expansion of the entries of $\mathbf{D}^{-1/2}$, we have $\mathbf{D}^{-1/2} = \mathbf{T}^{-1/2} + \frac{1}{2}\mathbf{T}^{-3/2}(\mathbf{T} - \mathbf{D}) + O_{\mathbb{P}}((n\rho_n)^{-3/2})$. Substituting this into equation (3.16) followed by further simplifications yield equation (3.1).

4. Subsequent inference.

In this section, we demonstrate how the results of Section 2.1 and Section 3 provide insights into subsequent inference. We first consider graphs generated according to a stochastic blockmodel with parameters

$$(4.1) \qquad \mathbf{B} = \begin{bmatrix} 0.42 & 0.42 \\ 0.42 & 0.5 \end{bmatrix}; \quad \text{and} \quad \pi = (0.6, 0.4).$$
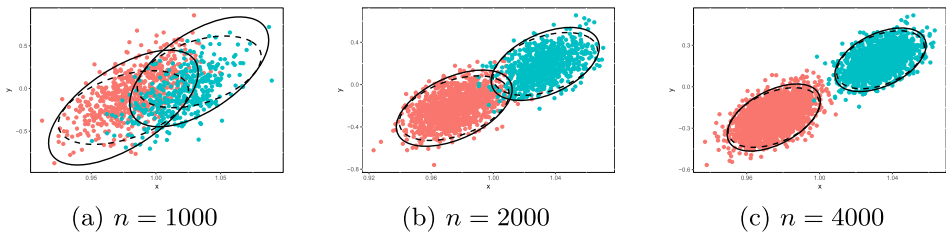
(a) $n = 1000$ (b) $n = 2000$ (c) $n = 4000$

FIG. 1. *Plot of the estimated latent positions in a two-block stochastic blockmodel graph on n vertices. The points are colored according to the block membership of the corresponding vertices. Dashed ellipses give the* 95% *level curves for the* empirical *distributions, that is, the* 95% *level curves using the sample covariance matrices as estimated from the rows of* $\check{\mathbf{X}}$. *Solid ellipses give the* 95% theoretical *level curves for the distributions as specified by Theorem* 3.2. *For each choice of n, the points are scaled by* $\sqrt{n}$ *so that they are on a that is scale invariant with respect to n.*

We sample an adjacency matrix **A** for graphs on n vertices from the above model for various choices of n. For each adjacency matrix **A**, we compute the normalized Laplacian embedding of **A**. Figure 1 presents examples of the scatter plots for these embeddings for $n = 1000$, 2000 and 4000. The points in the scatter plots are colored according to the block membership of the corresponding vertices in the blockmodel. For each block, we also plot the ellipses showing the empirical (dashed lines) and theoretical (solid lines) 95% level curves for the distribution of $\check{X}_i$. The theoretical level curves are as specified in Theorem 3.2.

We next investigate the implication of the multivariate normal distribution from Theorem 3.2 on subsequent inference. Spectral clustering refers to a large class of techniques used in partitioning data points into clusters that proceed by first performing a truncated eigendecomposition of a similarity matrix between the data points to obtain a low-dimensional Euclidean representation of these data points followed by clustering of the data points in this low-dimensional representation; see [26] for a comprehensive introduction. The *normalized cuts* algorithm of [36] is a popular and widely-used instance of spectral clustering where the similarity matrix is a normalized Laplacian matrix and clustering is done using the K-means algorithm.

It was shown in [34] that the normalized cuts algorithm, that is, the normalized Laplacian embedding followed by K-means, is consistent for estimating the block memberships of stochastic blockmodels graphs. The result of Corollary 3.3, however, suggests that K-means clustering is suboptimal unless the covariance matrices of the estimated latent positions for the blocks are spherical. We illustrate this by generating sequences of stochastic blockmodel graphs on n vertices with parameters as given in equation (4.1) where $n \in \{1000, 1250, 1500, \ldots, 4000\}$. For each graph, we embed its normalized Laplacian matrix into $\mathbb{R}^2$ and cluster the embedded vertices via either K-means or the MCLUST Gaussian mixture model-based clustering algorithm [17]. We then measure the error rate of the clustering
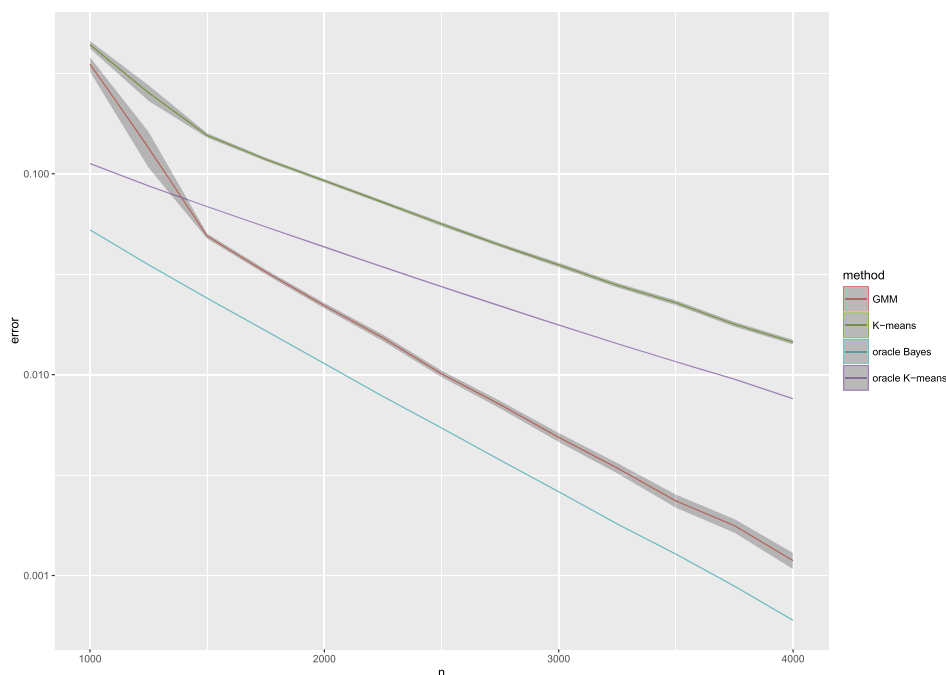
FIG. 2. *Comparison of clustering error rates for Gaussian mixture model (GMM) cluster-ing, K-means clustering, linear classifier and Bayes-optimal classifier. The error rate for each* $n \in \{1000, 1250, 1500, \ldots, 4000\}$ *was obtained by averaging* 100 *Monte Carlo iterations and are plotted on a* $\log_{10}$ *scale. The plot indicates that the assumption of a mixture of multivariate normals can yield significant improvement in the accuracy of the spectral clustering procedure.*

solution. The error rates, averaged over 100 replicates of the experiment, are presented on log-scale in Figure 2. We see that the Gaussian mixture model-based clustering does yield significant improvement over K-means clustering. For further comparison, we plot the Bayes-optimal error rate and that of a linear classifier which assign an embedded point to the closest theoretical centroid. The error rate of the linear classifier is computed under the assumption that the rows of the Laplacian spectral embedding are indeed multivariate normal with known covariance matrices and centered around the centroid of the respective blocks; this error rate serves as a lower-bound for that of K-means clustering.

4.1. *Comparison of ASE and LSE via within-class covariances.* We now discuss a comparison of the use of adjacency spectral embedding and Laplacian spectral embedding for subsequent inference. We consider as our subsequent inference task the problem of recovering the block assignments in stochastic blockmodel graphs. Our first metric of comparison is the notion of within-block variance for each block of the stochastic blockmodel, following the work of Sarkar and Bickely

[35]. We partially extend the results of [35] for two-block stochastic blockmodels to K-block stochastic blockmodels with positive semidefinite block probability matrices. However, while the collection of within-block variances is a meaningful surrogate for the performance of our subsequent inference task, we argue that it is not the "right" metric as it captures only the trace of the block-conditional covariance matrices and not the form of the block-conditional covariance matrices. That is to say, the use of the within-block variances as a surrogate measure is similar to the oracle K-means lower bound in Figure 2. A more appropriate surrogate is the collection of pairwise Chernoff information between the block-conditional multivariate normals, which behave similarly to the oracle Bayes' lower bound in Figure 2. The discussion of Chernoff information is postponed to the next subsection.

DEFINITION 3 (Within-block variances). Let $(\mathbf{X}, \mathbf{A}) \sim \text{RDPG}(F)$ with sparsity factor ρ_n where $F = \sum_k \pi_k \delta_{v_k}$ is a mixture of K point masses at $v_1, v_2, \ldots,$ $v_K \in \mathbb{R}^d$ and δ_{v_k} denotes the Dirac delta function. Given $\mathbf{A}$, let C_k for $k \in \{1, 2, \ldots, K\}$ denote the set of vertices of $\mathbf{A}$ assigned to block k. Recall the definitions of $\mathbf{U_A}$ and $\widetilde{\mathbf{U}}_\mathbf{A}$ in Definition 2, that is, $\mathbf{U_A}$ and $\widetilde{\mathbf{U}}_\mathbf{A}$ are the $n \times d$ matrices containing the d largest eigenvectors of the adjacency matrix and the Laplacian matrix, respectively. For any index i, let $\mathbf{U_A}(i, :)$ and $\widetilde{\mathbf{U}}_\mathbf{A}(i, :)$ denote the ith row of $\mathbf{U_A}$ and $\widetilde{\mathbf{U}}_\mathbf{A}(i, :)$, respectively. Then for any $k, l \in \{1, 2, \ldots, K\}$, the ASE variance between block k and block l is defined as

$$(4.2) \quad \hat{d}_{kl}^2 = \hat{d}_{kl}(\mathbf{A}) = \frac{1}{|C_k|} \sum_{i \in C_k} \|\mathbf{U_A}(i, :) - \hat{\mu}_l\|^2; \qquad \hat{\mu}_l = \frac{1}{|C_l|} \sum_{j \in C_l} \mathbf{U_A}(j, :).$$

Similarly, the LSE variance between block k and block l is

$$(4.3) \quad \widetilde{d}_{kl}^2 = \widetilde{d}_{kl}(\mathbf{A}) = \frac{1}{|C_k|} \sum_{i \in C_k} \|\widetilde{\mathbf{U}}_\mathbf{A}(i, :) - \widetilde{\mu}_l\|^2; \qquad \widetilde{\mu}_l = \frac{1}{|C_l|} \sum_{j \in C_k} \widetilde{\mathbf{U}}_\mathbf{A}(j, :).$$

When $k = l$, $\hat{d}_{kk}^2$ and $\widetilde{d}_{kk}^2$ are referred to as the ASE within-block variance for block k and the LSE within-block variance for block k, respectively.

We note that the $\hat{d}_{kl}^2$ and $\widetilde{d}_{kl}^2$ are defined in terms of $\mathbf{U_A}$ and $\widetilde{\mathbf{U}}_\mathbf{A}$ and not in terms of $\hat{\mathbf{X}} = \mathbf{U_A}\mathbf{S}_\mathbf{A}^{1/2}$ and $\check{\mathbf{X}} = \widetilde{\mathbf{U}}_\mathbf{A}\widetilde{\mathbf{S}}_\mathbf{A}^{1/2}$. This is because $\|\mathbf{S}_\mathbf{A}^{1/2}\| \gg \|\widetilde{\mathbf{S}}_\mathbf{A}^{1/2}\|$.

We then have the following large-sample limit results for $\hat{d}_{kk}^2$ and $\widetilde{d}_{kk}^2$. Their proofs are similar to those of Theorem 2.1 and Theorem 3.1 and, therefore, will be omitted. The limit results for $\hat{d}_{kl}^2$ and $\widetilde{d}_{kl}^2$ when $k \neq l$ are much simpler and will not be presented here. Nevertheless, we will verify later in this section that Theorem 4.1 and Theorem 4.2 are indeed generalizations of Theorem 3.1 and Theorem 3.2 from [35]. We emphasize that neither Theorem 4.1 nor Theorem 4.2 assume distinct eigenvalues of the matrix $\mathbf{XX}^\top$ or $\mathcal{L}(\mathbf{XX}^\top)$; distinct eigenvalues is a necessary assumption used in the proofs of Theorem 3.1 and Theorem 3.2 in [35] (see Section 8 of the cited paper).

THEOREM 4.1. *Assume the setting and notation of Theorem* 2.1 *and suppose furthermore that $F = \sum_k \pi_k \delta_{v_k}$ is a mixture of K distinct point masses at $v_1, v_2, \ldots, v_K \in \mathbb{R}^d$. Let $\mathbf{U}_{\mathbf{P}_n}$ denote the $n \times d$ matrix whose columns are the orthonormal eigenvectors corresponding to the nonzero eigenvalues of the matrix $\mathbf{P}_n = \rho_n \mathbf{X}_n \mathbf{X}_n^\top$. For any $k \in \{1, 2, \ldots, K\}$, let $\mathbf{S}_k$ be the $n \times n$ diagonal matrix with diagonal entries $(s_k(1), s_k(2), \ldots, s_k(n))$ such that $s_k(i) = 1$ if $X_i = v_k$ and $s_k(i) = 0$ otherwise. We then have, for any $k \in \{1, 2, \ldots, K\}$*

$$
(4.4) \qquad
\begin{aligned}
n^2 \hat{d}_{kk}^2 &= \frac{n^2}{|C_k|} \big\| \mathbf{S}_k (\mathbf{U}_{\mathbf{A}_n} \mathbf{W}_n - \mathbf{U}_{\mathbf{P}_n}) \big\|_F^2 + o_{\mathbb{P}}(1) \\
&= \frac{n^2}{|C_k|} \big\| \mathbf{S}_k (\mathbf{A}_n - \mathbf{P}_n) \mathbf{X}_n (\mathbf{X}_n^\top \mathbf{X}_n)^{-3/2} \big\|_F^2 + o_{\mathbb{P}}(1).
\end{aligned}
$$

Therefore, if $\rho_n \equiv 1$, then for any $k \in \{1, 2, \ldots, K\}$

$$
(4.5) \qquad n^2 \hat{d}_{kk}^2 \xrightarrow{\text{a.s.}} \operatorname{tr} \Delta^{-3} \mathbb{E}\big[X_1 X_1^\top \big(v_k^\top X_1 - v_k^\top X_1 X_1^\top v_k \big) \big]
$$

as $n \to \infty$. If, however, $\rho_n \to 0$ and $n\rho_n = \omega(\log^4(n))$, then

$$
(4.6) \qquad n^2 \hat{d}_{kk}^2 \xrightarrow{\text{a.s.}} \operatorname{tr} \Delta^{-3} \mathbb{E}\big[X_1 X_1^\top v_k^\top X_1 \big]
$$

as $n \to \infty$.

For the $\widetilde{d}_{kk}^2$, we have the following result.

THEOREM 4.2. *Assume the setting and notation of Theorem* 3.1 *and suppose furthermore that $F = \sum_k \pi_k \delta_{v_k}$ is a mixture of K distinct point masses at $v_1, v_2, \ldots, v_K \in \mathbb{R}^d$. Let $\widetilde{\mathbf{U}}_{\mathbf{P}_n}$ denote the $n \times d$ matrix whose columns are the orthonormal eigenvectors corresponding to the nonzero eigenvalues of the matrix $\mathcal{L}(\mathbf{P}_n) = \mathcal{L}(\rho_n \mathbf{X}_n \mathbf{X}_n^\top) = \mathcal{L}(\mathbf{X}_n \mathbf{X}_n^\top)$. For any $k \in \{1, 2, \ldots, K\}$, let $\mathbf{S}_k$ be the $n \times n$ diagonal matrix with diagonal entries $(s_k(1), s_k(2), \ldots, s_k(n))$ such that $s_k(i) = 1$ if $X_i = v_k$ and $s_k(i) = 0$ otherwise. We then have, for any $k \in \{1, 2, \ldots, K\}$,*

$$
(4.7) \qquad
\begin{aligned}
n^2 \widetilde{d}_{kk}^2 &= \frac{n^2}{|C_k|} \big\| \mathbf{S}_k (\widetilde{\mathbf{U}}_{\mathbf{A}_n} \mathbf{W}_n - \widetilde{\mathbf{U}}_{\mathbf{P}_n}) \big\|_F^2 + o_{\mathbb{P}}(1) \\
&= \frac{n^2}{|C_k|} \big\| \mathbf{S}_k \mathbf{M}_1 (\widetilde{\mathbf{X}}_n^\top \widetilde{\mathbf{X}}_n)^{-3/2} + \tfrac{1}{2} \mathbf{S}_k \mathbf{M}_2 (\widetilde{\mathbf{X}}_n^\top \widetilde{\mathbf{X}}_n)^{-1/2} \big\|_F^2 + o_{\mathbb{P}}(1),
\end{aligned}
$$

where $\mathbf{M}_1$ and $\mathbf{M}_2$ are defined as

$$
(4.8) \qquad \mathbf{M}_1 = \mathbf{T}_n^{-1/2} (\mathbf{A}_n - \mathbf{P}_n) \mathbf{T}_n^{-1/2} \widetilde{\mathbf{X}}_n,
$$

$$
(4.9) \qquad \mathbf{M}_2 = \mathbf{T}_n^{-1/2} (\mathbf{T}_n - \mathbf{D}_n) \mathbf{T}_n^{-1/2} \widetilde{\mathbf{X}}_n.
$$

Therefore, if $\rho_n \equiv 1$, then for any $k \in \{1, 2, \ldots, K\}$

$$
\begin{aligned}
(4.10) \quad n^2 \widetilde{d}_{kk}^2 \xrightarrow{\text{a.s.}} \ & \operatorname{tr} \widetilde{\Delta}^{-3} \mathbb{E}\bigg[\bigg(\frac{X_1}{X_1^\top \mu} - \frac{\widetilde{\Delta} v_k}{2 v_k^\top \mu}\bigg)\bigg(\frac{X_1^\top}{X_1^\top \mu} - \frac{v_k^\top \widetilde{\Delta}}{2 v_k^\top \mu}\bigg) \\
& \times \frac{(v_k^\top X_1 - v_k^\top X_1 X_1^\top v_k)}{v_k^\top \mu}\bigg]
\end{aligned}
$$

as $n \to \infty$. If, however, $\rho_n \to 0$ and $n\rho_n = \omega(\log^4(n))$, then

$$
(4.11) \quad n^2 \widetilde{d}_{kk}^2 \xrightarrow{\text{a.s.}} \operatorname{tr} \widetilde{\Delta}^{-3} \mathbb{E}\bigg[\bigg(\frac{X_1}{X_1^\top \mu} - \frac{\widetilde{\Delta} v_k}{2 v_k^\top \mu}\bigg)\bigg(\frac{X_1^\top}{X_1^\top \mu} - \frac{v_k^\top \widetilde{\Delta}}{2 v_k^\top \mu}\bigg)\frac{v_k^\top X_1}{v_k^\top \mu}\bigg]
$$

as $n \to \infty$.

We now verify that Theorem 4.1 and Theorem 4.2 are indeed generalizations of Theorem 3.1 and Theorem 3.2 from [35]. Suppose that $K = d$, that is, that **B** is invertible. Then denoting by $\boldsymbol{v}$ the $d \times d$ matrix $\boldsymbol{v} = [v_1 \mid v_2 \mid \cdots \mid v_d]$, we have that $\boldsymbol{v}$ is also invertible and that $\mathbf{B} = \boldsymbol{v}^\top \boldsymbol{v}$ and $\Delta = \boldsymbol{v} \operatorname{diag}(\boldsymbol{\pi}) \boldsymbol{v}^\top$. Let $z_k = (v_k^\top v_1 (1 - v_k^\top v_1), \ldots, v_k^\top v_d (1 - v_k^\top v_d))$. Then

$$
\mathbb{E}[X_1 X_1^\top (v_k^\top X_1 - v_k^\top X_1 X_1^\top v_k)] = \boldsymbol{v}(\operatorname{diag}(\boldsymbol{\pi}) \operatorname{diag}(z_k))\boldsymbol{v}^\top.
$$

Then equation (4.5) in Theorem 4.1 simplifies to

$$
\begin{aligned}
(4.12) \quad n^2 \hat{d}_{kk}^2 \xrightarrow{\text{a.s.}} \ & \operatorname{tr} \Delta^{-3} \mathbb{E}[X_1 X_1^\top (v_k^\top X_1 - v_k^\top X_1 X_1^\top v_k)] \\
& = \operatorname{tr}(\boldsymbol{v} \operatorname{diag}(\boldsymbol{\pi}) \boldsymbol{v}^\top)^{-3} \boldsymbol{v}(\operatorname{diag}(\boldsymbol{\pi}) \operatorname{diag}(z_k))\boldsymbol{v}^\top \\
& = \operatorname{tr}((\boldsymbol{v}^\top)^{-1} \operatorname{diag}(\boldsymbol{\pi})^{-1} \boldsymbol{v}^{-1})^3 \boldsymbol{v}(\operatorname{diag}(\boldsymbol{\pi}) \operatorname{diag}(z_k))\boldsymbol{v}^\top \\
& = \operatorname{tr}(\operatorname{diag}(\boldsymbol{\pi})^{-1} \boldsymbol{v}^{-1} (\boldsymbol{v}^\top)^{-1})^2 \operatorname{diag}(z_k) \\
& = \operatorname{tr}(\operatorname{diag}(\boldsymbol{\pi})^{-1} \mathbf{B}^{-1})^2 \operatorname{diag}(z_k) \\
& = \operatorname{tr}(\operatorname{diag}(\boldsymbol{\pi})^{-1/2} \mathbf{B}^{-1} \operatorname{diag}(\boldsymbol{\pi})^{-1/2})^2 \operatorname{diag}(z_k) \\
& = \sum_l \sum_{l'} \frac{v_k^\top v_l (1 - v_k^\top v_l)(\mathbf{B}_{ll'}^{-1})^2}{\pi_l \pi_{l'}} \\
& = \sum_l \sum_{l'} \frac{\mathbf{B}_{kl}(1 - \mathbf{B}_{kl})(\mathbf{B}_{ll'}^{-1})^2}{\pi_l \pi_{l'}},
\end{aligned}
$$

where $\mathbf{B}_{ll'}^{-1}$ is the ll'th entry of $\mathbf{B}^{-1}$. We emphasize that the above expression for $\hat{d}_{kk}^2$ can be written purely in terms of the entries of **B** and $\boldsymbol{\pi}$ without the need to find the $v_1, v_2, \ldots, v_d$ explicitly.

We compare equation (4.12) with Theorem 3.1 in [35]. Let $\mathbf{A}$ be sampled from a stochastic blockmodel with parameters $\mathbf{B} = \begin{bmatrix} \alpha_n & \beta_n \\ \beta_n & \gamma_n \end{bmatrix}$ and $\boldsymbol{\pi} = (\pi_1, \pi_2)$ with $\alpha_n \beta_n \neq \gamma_n^2$. In [35], it is assumed that the number of vertices assigned to block 1 and block 2 are $n\pi_1$ and $n\pi_2$, respectively. For ease of exposition and without loss of generality, suppose that the row indices of $\mathbf{A}$ are such that the first $n\pi_1$ rows correspond to vertices assigned to block 1 and the last $n\pi_2 = n - n\pi_1$ rows correspond to vertices assigned to block 2. Let $\boldsymbol{v}_1$ and $\boldsymbol{v}_2$ denote the eigenvectors corresponding to the largest and second largest eigenvector of $\mathbf{P} = \mathbf{Z}\mathbf{B}\mathbf{Z}^\top$ where $\mathbf{Z}$ is a $n \times 2$ matrix whose ith row is $(1, 0)$ for $i = 1, 2, \ldots, n\pi_1$ and is $(0, 1)$ for $i = n\pi_1 + 1, n\pi_1 + 2, \ldots, n$. We then have that $\boldsymbol{v}_1 = (x_1, x_1, \ldots, x_1, y_1, y_1, \ldots, y_1)$ for some x_1, y_1, that is, the first $n\pi_1$ elements of $\boldsymbol{v}_1$ are x_1 and the remaining $n\pi_2$ elements are x_2. Similarly, we have $\boldsymbol{v}_2 = (x_2, x_2, \ldots, x_2, y_2, y_2, \ldots, y_2)$ for some x_2, y_2. Then equation (3.1) in [35] states that [the notation $a_n \sim b_n$ in [35] means $a_n / b_n = 1 + o_{\mathbb{P}}(1)$]

$$(4.13) \qquad \hat{d}_{11}^2 \sim \left[\left(\frac{x_1^2}{\lambda_1^2} + \frac{x_2^2}{\lambda_2^2} \right) n\pi_1 \alpha_n (1 - \alpha_n) + \left(\frac{y_1^2}{\lambda_1^2} + \frac{y_2^2}{\lambda_2^2} \right) n\pi_2 \gamma_n (1 - \gamma_n) \right],$$

where λ_1 and λ_2 are the largest and second largest eigenvalues of $\mathbf{P}$. We can rewrite equation (4.13) as

$$(4.14) \qquad \hat{d}_{11}^2 \sim \mathrm{tr}(\mathbf{P}^\dagger)^2 \mathrm{diag}((\alpha_n(1 - \alpha_n), \ldots \gamma_n(1 - \gamma_n), \ldots)),$$

where $\mathbf{P}^\dagger$ is the Moore–Penrose pseudo-inverse of $\mathbf{P}$ and the first $n\pi_1$ entries of the diagonal matrix $\mathrm{diag}(\alpha_n(1 - \alpha_n), \ldots \gamma_n(1 - \gamma_n), \ldots)$ are $\alpha_n(1 - \alpha_n)$ while the remaining $n\pi_2$ diagonal entries are $\gamma_n(1 - \gamma_n)$. As $\mathbf{Z}$ is of full-column rank, we have $\mathbf{Z}^\dagger = (\mathbf{Z}^\top \mathbf{Z})^{-1} \mathbf{Z}^\top = \mathrm{diag}((1/(n\pi_1), 1/(n\pi_2)))\mathbf{Z}^\top$. Furthermore, $\mathbf{B}$ is invertible, and hence

$$\mathbf{P}^\dagger = (\mathbf{Z}\mathbf{B}\mathbf{Z}^\top)^\dagger = (\mathbf{Z}^\top)^\dagger \mathbf{B}^{-1} \mathbf{Z}^\dagger = n^{-2} \mathbf{Z}\, \mathrm{diag}(\boldsymbol{\pi})^{-1} \mathbf{B}^{-1}\, \mathrm{diag}(\boldsymbol{\pi})^{-1} \mathbf{Z}^\top.$$

Therefore,

$$(\mathbf{P}^\dagger)^2 = n^{-3} \mathbf{Z}\, \mathrm{diag}(\boldsymbol{\pi})^{-1} \mathbf{B}^{-1}\, \mathrm{diag}(\boldsymbol{\pi})^{-1} \mathbf{B}^{-1}\, \mathrm{diag}(\boldsymbol{\pi})^{-1} \mathbf{Z}^\top$$

and hence

$$\begin{aligned}
n^2 \hat{d}_{11}^2 &\sim n^2 \, \mathrm{tr}(\mathbf{P}^\dagger)^2 \, \mathrm{diag}((\alpha_n(1 - \alpha_n), \ldots, \gamma_n(1 - \gamma_n), \ldots)) \\
&\sim n^{-1} \, \mathrm{tr}\, \mathbf{Z}(\mathrm{diag}(\boldsymbol{\pi})^{-1} \mathbf{B}^{-1})^2 \, \mathrm{diag}(\boldsymbol{\pi})^{-1} \mathbf{Z}^\top \\
&\qquad \times \mathrm{diag}((\alpha_n(1 - \alpha_n), \ldots \gamma_n(1 - \gamma_n), \ldots)) \\
&\sim \mathrm{tr}(\mathrm{diag}(\boldsymbol{\pi})^{-1} \mathbf{B}^{-1})^2 \, \mathrm{diag}((\alpha_n(1 - \alpha_n), \gamma_n(1 - \gamma_n)))
\end{aligned}$$

which is a special case of equation (4.12). Theorem 4.1 is thus an extension of Theorem 3.1 in [35] to general K-block stochastic blockmodels, provided that the block probability matrix is positive semidefinite.

We now consider $n^2 \widetilde{d}_{kk}^2$. When $\mathbf{B}$ is invertible, equation (4.10) in Theorem 4.2 can be simplified as follows. Let $\boldsymbol{\mu} = (\mu_1, \mu_2, \ldots \mu_d)$ where $\mu_k = v_k^\top \boldsymbol{\mu}$. Then $\widetilde{\Delta} = \boldsymbol{v}(\mathrm{diag}(\boldsymbol{\pi})\,\mathrm{diag}(\boldsymbol{\mu})^{-1})\boldsymbol{v}^\top$. The right-hand side of equation (4.10) can be decomposed as $\zeta_1 - \zeta_2 + \zeta_3$ with ζ_1 given by

$$
\begin{aligned}
\zeta_1 &= \mathrm{tr}\, \widetilde{\Delta}^{-3} \mathbb{E}\left[\frac{X_1 X_1^\top}{(X_1^\top \mu)^2} \frac{v_k^\top X_1 - v_k^\top X_1 X_1^\top v_k}{\mu_k} \right] \\
&= \frac{1}{\mu_k}\,\mathrm{tr}\, \widetilde{\Delta}^{-3} \boldsymbol{v}^\top \big(\mathrm{diag}(\boldsymbol{\pi})\,\mathrm{diag}(\boldsymbol{\mu})^{-2}\,\mathrm{diag}(z_k)\big)\boldsymbol{v}^\top \\
&= \frac{1}{v_k^\top \mu}\,\mathrm{tr}\big(\mathrm{diag}(\boldsymbol{\pi})^{-1}\,\mathrm{diag}(\boldsymbol{\mu})\boldsymbol{v}^{-1}(\boldsymbol{v}^\top)^{-1}\big)^2 \mathrm{diag}(\boldsymbol{\mu})^{-1}\,\mathrm{diag}(z_k) \\
&= \frac{1}{\mu_k}\,\mathrm{tr}\big(\mathrm{diag}(\boldsymbol{\pi})^{-1}\,\mathrm{diag}(\boldsymbol{\mu})\mathbf{B}^{-1}\big)^2 \mathrm{diag}(\boldsymbol{\mu})^{-1}\,\mathrm{diag}(z_k) \\
&= \frac{1}{\mu_k}\,\mathrm{tr}\big(\mathrm{diag}(\boldsymbol{\pi})^{-1/2}\,\mathrm{diag}(\boldsymbol{\mu})^{1/2}\mathbf{B}^{-1}\,\mathrm{diag}(\boldsymbol{\mu})^{1/2}\,\mathrm{diag}(\boldsymbol{\pi})^{-1/2}\big)^2 \\
&\quad \times \mathrm{diag}(\boldsymbol{\mu})^{-1}\,\mathrm{diag}(z_k) \\
&= \sum_l \sum_{l'} \frac{(\mathbf{B}_{ll'}^{-1})^2 \mu_l \mu_{l'}}{\pi_l \pi_{l'}} \frac{v_k^\top v_l (1 - v_k^\top v_l)}{\mu_l \mu_k} = \sum_l \sum_{l'} \frac{\mathbf{B}_{kl}(1 - \mathbf{B}_{kl})(\mathbf{B}_{ll'}^{-1})^2 \mu_{l'}}{\pi_l \pi_{l'} \mu_k},
\end{aligned}
$$

ζ_2 given by (here $\boldsymbol{e}_k$ is the vector whose ith element is 1 if $i = k$ and 0 otherwise)

$$
\begin{aligned}
\zeta_2 &= \mathrm{tr}\, \widetilde{\Delta}^{-2} \mathbb{E}\left[\frac{X_1 v_k^\top}{X_1^\top \mu} \frac{v_k^\top X_1 - v_k^\top X_1 X_1^\top v_k}{\mu_k^2} \right] \\
&= \frac{1}{\mu_k^2}\,\mathrm{tr}\, \widetilde{\Delta}^{-2} \boldsymbol{v}^\top \big(\mathrm{diag}(\boldsymbol{\pi})\,\mathrm{diag}(\boldsymbol{\mu})^{-1}\,\mathrm{diag}(z_k)\big)\mathbf{1}v_k^\top \\
&= \frac{1}{\mu_k^2}\,\mathrm{tr}\,\mathrm{diag}(\boldsymbol{\pi})^{-1}\,\mathrm{diag}(\boldsymbol{\mu})\boldsymbol{v}^{-1}(\boldsymbol{v}^\top)^{-1}\,\mathrm{diag}(z_k)\mathbf{1}e_k^\top \\
&= \frac{1}{\mu_k^2}\,\mathrm{tr}\,\mathrm{diag}(\boldsymbol{\pi})^{-1}\,\mathrm{diag}(\boldsymbol{\mu})\mathbf{B}^{-1}\,\mathrm{diag}(z_k)\mathbf{1}e_k^\top \\
&= \frac{1}{\pi_k \mu_k} \sum_l v_k^\top v_l (1 - v_k^\top v_l)\mathbf{B}_{kl}^{-1} = \frac{1}{\pi_k \mu_k} \sum_l \mathbf{B}_{kl}(1 - \mathbf{B}_{kl})\mathbf{B}_{kl}^{-1},
\end{aligned}
$$

and ζ_3 given by

$$
\begin{aligned}
\zeta_3 &= \frac{1}{4\mu_k^3}\,\mathrm{tr}\, \widetilde{\Delta}^{-1} v_k v_k^\top \mathbb{E}\big[v_k^\top X_1 - v_k^\top X_1 X_1^\top v_k\big] \\
&= \frac{\mathbb{E}\big[v_k^\top X_1 - v_k^\top X_1 X_1^\top v_k\big]}{4\mu_k^3}\,\mathrm{tr}\, v_k^\top \widetilde{\Delta}^{-1} v_k
\end{aligned}
$$

$$= \frac{\mathbb{E}[\nu_k^\top X_1 - \nu_k^\top X_1 X_1^\top \nu_k]}{4\mu_k^3} \operatorname{tr} \nu_k^\top (\boldsymbol{\nu}^\top)^{-1} (\operatorname{diag}(\boldsymbol{\pi})^{-1} \operatorname{diag}(\boldsymbol{\mu})) \boldsymbol{\nu}^{-1} \nu_k$$

$$= \frac{\mathbb{E}[\nu_k^\top X_1 - \nu_k^\top X_1 X_1^\top \nu_k]}{4\mu_k^3} \frac{\mu_k}{\pi_k} = \frac{\sum_l \pi_l \mathbf{B}_{kl}(1 - \mathbf{B}_{kl})}{4\pi_k \mu_k^2}.$$

As $\mu_k = \sum_l \pi_l \nu_k^\top \nu_l = \sum_l \pi_l \mathbf{B}_{kl}$, ζ_1, ζ_2 and ζ_3 can also be written purely in terms of the entries of $\mathbf{B}$ and $\boldsymbol{\pi}$.

For the two-block stochastic blockmodel, equation (3.3) in [35] states that

$$(4.15) \qquad n^2 \widetilde{d}_{11}^2 \sim \frac{\alpha_n(1 - \alpha_n)}{\mu_1^2} \left(\frac{1}{4} + \frac{\pi_2 \gamma_n}{\mu_1 \widetilde{\lambda}_2^2} \right) + \frac{\gamma_n(1 - \gamma_n)}{\mu_1^2} \left(\frac{\pi_2}{4\pi_1} + \frac{\pi_1 \alpha_n}{\mu_2 \widetilde{\lambda}_2^2} \right),$$

where $\widetilde{\lambda}_2 = \pi_1 \pi_2 (\alpha_n \beta_n - \gamma_n^2)/(\mu_1 \mu_2)$ is the second largest eigenvalue of $\mathcal{L}(\mathbf{P})$ (cf. Lemma 6.1 in [35]). Verifying that $\zeta_1 - \zeta_2 + \zeta_3$ does indeed yield equation (4.15) for the two-block stochastic blockmodel is a straightforward computation. We omit the details. Theorem 4.2 is thus an extension of Theorem 3.2 in [35] for general K-blocks stochastic blockmodels whenever the matrix of block probabilities is positive semidefinite.

4.2. *Chernoff information.* Let F_0 and F_1 be two absolutely continuous multivariate distributions in $\Omega = \mathbb{R}^d$ with density functions f_0 and f_1, respectively. Suppose that $Y_1, Y_2, \ldots, Y_m$ are independent and identically distributed random variables, with Y_i distributed either F_0 or F_1. We are interested in testing the simple null hypothesis $\mathbb{H}_0: F = F_0$ against the simple alternative hypothesis $\mathbb{H}_1: F = F_1$. A test T can be viewed as a sequence of mappings $T_m: \Omega^m \mapsto \{0, 1\}$ such that given $Y_1 = y_1, Y_2 = y_2, \ldots, Y_m = y_m$, the test rejects $\mathbb{H}_0$ in favor of $\mathbb{H}_1$ if $T_m(y_1, y_2, \ldots, y_m) = 1$; similarly, the test favors $\mathbb{H}_0$ if $T_m(y_1, y_2, \ldots, y_m) = 0$.

The Neyman–Pearson lemma states that, given $Y_1 = y_1, Y_2 = y_2, \ldots, Y_m = y_m$ and a threshold $\eta_m \in \mathbb{R}$, the likelihood ratio test which rejects $\mathbb{H}_0$ in favor of $\mathbb{H}_1$ whenever

$$\left(\sum_{i=1}^m \log f_0(y_i) - \sum_{i=1}^m \log f_1(y_i) \right) \leq \eta_m$$

is the most powerful test at significance level $\alpha_m = \alpha(\eta_m)$, that is, the likelihood ratio test minimizes the type-II error β_m subject to the constraint that the type-I error is at most α_m.

Assuming that $\pi \in (0, 1)$ is a prior probability that $\mathbb{H}_0$ is true. Then, for a given $\alpha_m^* \in (0, 1)$, let $\beta_m^* = \beta_m^*(\alpha_m^*)$ be the type-II error associated with the likelihood ratio test when the type-I error is at most α_m^*. The quantity $\inf_{\alpha_m^* \in (0,1)} \pi \alpha_m^* + (1 - \pi)\beta_m^*$ is then the Bayes' risk in deciding between $\mathbb{H}_0$ and $\mathbb{H}_1$ given the m independent random variables $Y_1, Y_2, \ldots, Y_m$. A classical result of Chernoff [10, 11]

states that the Bayes' risk is intrinsically linked to a quantity known as the *Chernoff information*. More specifically, let $C(F_0, F_1)$ be the quantity

(4.16)
$$C(F_0, F_1) = -\log\left[\inf_{t\in(0,1)} \int_{\mathbb{R}^d} f_0^t(\boldsymbol{x}) f_1^{1-t}(\boldsymbol{x})\,d\boldsymbol{x}\right]$$
$$= \sup_{t\in(0,1)}\left[-\log \int_{\mathbb{R}^d} f_0^t(\boldsymbol{x}) f_1^{1-t}(\boldsymbol{x})\,d\boldsymbol{x}\right].$$

Then we have

(4.17)
$$\lim_{m\to\infty} \frac{1}{m} \inf_{\alpha_m^*\in(0,1)} \log(\pi\alpha_m^* + (1-\pi)\beta_m^*) = -C(F_0, F_1).$$

Thus $C(F_0, F_1)$, the Chernoff information between F_0 and F_1, is the *exponential rate* at which the Bayes' error $\inf_{\alpha_m^*\in(0,1)} \pi\alpha_m^* + (1-\pi)\beta_m^*$ decreases as $m\to\infty$; we note that the Chernoff information is independent of π. We also define, for a given $t\in(0,1)$ the Chernoff divergence $C_t(F_0, F_1)$ between F_0 and F_1 by

$$C_t(F_0, F_1) = -\log \int_{\mathbb{R}^d} f_0^t(\boldsymbol{x}) f_1^{1-t}(\boldsymbol{x})\,d\boldsymbol{x}.$$

The Chernoff divergence is an example of a f-divergence as defined in [1, 15]. When $t = 1/2$, $C_t(F_0, F_1)$ is the Bhattacharyya distance between F_0 and F_1. As we mentioned previously, any f-divergence satisfies the information processing lemma and is invariant with respect to invertible transformations [24]. Thus any f-divergence such as the Kullback–Liebler divergence can also be used to compare the two embedding methods. We chose the Chernoff information mainly because of its explicit relationship with the Bayes' risk.

The result of equation (4.17) can be extended to $K + 1 \geq 2$ hypotheses. Let $F_0, F_1, \ldots, F_K$ be distributions on $\mathbb{R}^d$ and suppose that $Y_1, Y_2, \ldots, Y_m$ are independent and identically distributed random variables with Y_i distributed $F \in \{F_0, F_1, \ldots, F_K\}$. We are thus interested in determining the distribution of the Y_i among the $K + 1$ hypothesis $\mathbb{H}_0\colon F = F_0, \ldots, \mathbb{H}_K\colon F = F_K$. Suppose also that hypothesis $\mathbb{H}_k$ has a priori probability π_k. Then for any decision rule δ, the risk of δ is $r(\delta) = \sum_k \pi_k \sum_{l\neq k} \alpha_{lk}(\delta)$ where $\alpha_{lk}(\delta)$ is the probability of accepting hypothesis $\mathbb{H}_l$ when hypothesis $\mathbb{H}_k$ is true. Then we have [22]

(4.18)
$$\inf_{\delta} \lim_{m\to\infty} \frac{r(\delta)}{m} = -\min_{k\neq l} C(F_k, F_l),$$

where the infimum is over all decision rules δ. That is to say, for any δ, $r(\delta)$ decreases to 0 as $m\to\infty$ at a rate no faster than $\exp(-m \min_{k\neq l} C(F_k, F_l))$. It was also shown in [22] that the *Maximum A Posterior* decision rule achieves this rate.

For this paper, we are interested in computing the Chernoff information $C(F_0, F_1)$ when F_0 and F_1 are multivariate normals. Suppose $F_0 = \mathcal{N}(\mu_0, \Sigma_0)$ and $F_1 = \mathcal{N}(\mu_1, \Sigma_1)$; then, denoting by $\Sigma_t = t\Sigma_0 + (1-t)\Sigma_1$, we have

$$C(F_0, F_1) = \sup_{t\in(0,1)} \left(\frac{t(1-t)}{2}(\mu_1 - \mu_2)^\top \Sigma_t^{-1}(\mu_1 - \mu_2) + \frac{1}{2}\log \frac{|\Sigma_t|}{|\Sigma_0|^t |\Sigma_1|^{1-t}}\right).$$

4.3. *Comparison of ASE and LSE via Chernoff information.* We now employ the limit results of Corollary 2.3 and Corollary 3.3 to compare the performance of the Laplacian spectral embedding and the adjacency spectral embedding for subsequent inference. Our subsequent inference task is once again the problem of recovering the block assignments in stochastic blockmodel graphs; furthermore, we are interested in estimating the *large-sample optimal* error rate possible for recovering the underlying block assignments after the spectral embedding step is carried out. The discussion in Section 4.2 indicates that an appropriate measure for the large-sample optimal error rate for spectral clustering using adjacency or Laplacian spectral embedding is in terms of the minimum of the pairwise Chernoff information between the multivariate normal distributions as specified in Corollary 2.3 or Corollary 3.3. More specifically, let $\mathbf{B} \in [0,1]^{K \times K}$ and $\boldsymbol{\pi} \in \mathbb{R}^K$ be the matrix of block probabilities and the vector of block assignment probabilities for a K-block stochastic blockmodel. We shall assume that $\mathbf{B}$ is positive semidefinite. Then given an n vertex instantiation of the SBM graph with parameters $(\boldsymbol{\pi}, \mathbf{B})$, for sufficiently large n, the large-sample optimal error rate for recovering the block assignments when adjacency spectral embedding is used as the initial embedding step can be characterized by the quantity $\rho_{\mathrm{A}} = \rho_{\mathrm{A}}(n)$ defined by

$$
\begin{aligned}
(4.19) \quad \rho_{\mathrm{A}} = \min_{k \neq l} \sup_{t \in (0,1)} \frac{1}{2} \log \frac{|\Sigma_{kl}(t)|}{|\Sigma_k|^t |\Sigma_l|^{1-t}} \\
+ \frac{nt(1-t)}{2} (v_k - v_l)^\top \Sigma_{kl}^{-1}(t)(v_k - v_l),
\end{aligned}
$$

where $\Sigma_{kl}(t) = t\Sigma_k + (1-t)\Sigma_l$; $\Sigma_k = \Sigma(v_k)$ and $\Sigma_l = \Sigma(v_l)$ are as defined in equation (2.8). We recall equation (4.18), in particular the fact that as ρ_{A} increases, the large-sample optimal error rate decreases. Similarly, the large-sample optimal error rate when Laplacian spectral embedding is used as the pre-processing step can be characterized by the quantity $\rho_{\mathrm{L}} = \rho_{\mathrm{L}}(n)$ defined by

$$
\begin{aligned}
(4.20) \quad \rho_{\mathrm{L}} = \min_{k \neq l} \sup_{t \in (0,1)} \frac{1}{2} \log \frac{|\widetilde{\Sigma}_{kl}(t)|}{|\widetilde{\Sigma}_k|^t |\widetilde{\Sigma}_l|^{1-t}} \\
+ \frac{nt(1-t)}{2} (\widetilde{v}_k - \widetilde{v}_l)^\top \widetilde{\Sigma}_{kl}^{-1}(t)(\widetilde{v}_k - \widetilde{v}_l),
\end{aligned}
$$

where $\widetilde{\Sigma}_{kl}(t) = t\widetilde{\Sigma}_k + (1-t)\widetilde{\Sigma}_l$ with $\widetilde{\Sigma}_k = \widetilde{\Sigma}(v_k)$ and $\widetilde{\Sigma}_l = \widetilde{\Sigma}(v_l)$ as defined in equation (3.6), and $\widetilde{v}_k = v_k / (\sum_{k'} \pi_{k'} v_k^\top v_{k'})^{1/2}$. We emphasize that we have made the simplifying assumption that $n_k = n\pi_k$ in our expression for $\widetilde{v}_k$ in equation (4.20). This is for ease of comparison between ρ_{A} and ρ_{L} in our subsequent discussion.

We thus propose to use the ratio $\rho_{\mathrm{A}}/\rho_{\mathrm{L}}$ as a measure of the relative large-sample performance of the adjacency spectral embedding as compared to the Laplacian spectral embedding for subsequent inference, at least in the context of stochastic

blockmodel graphs. That is to say, for given parameters $\boldsymbol{\pi}$ and $\mathbf{B}$, if $\rho_\mathrm{A}/\rho_\mathrm{L} > 1$ then adjacency spectral embedding is to be preferred over Laplacian spectral embedding when n, the number of vertices in the graph, is sufficiently large; similarly, if $\rho_\mathrm{A}/\rho_\mathrm{L} < 1$ then Laplacian spectral embedding is to be preferred over adjacency spectral embedding.

REMARK. We note that if the block-conditional covariance matrices Σ_k are all nonsingular, then for sufficiently large n, the term $\log \frac{|\Sigma_{kl}(t)|}{|\Sigma_k|^t|\Sigma_l|^{1-t}}$ in the definition of ρ_A is negligible; similarly, the term $\log \frac{|\widetilde{\Sigma}_{kl}(t)|}{|\widetilde{\Sigma}_k|^t|\widetilde{\Sigma}_l|^{1-t}}$ in the definition of ρ_L is also negligible. However, on occasion, some of the block-conditional covariance matrices Σ_k are singular. As an example, we consider a completely associative two-block stochastic blockmodel with $\mathbf{B} = \begin{bmatrix} p^2 & 0 \\ 0 & q^2 \end{bmatrix}$ and $\boldsymbol{\pi} = (\pi_1, \pi_2)$. Then the block-conditional covariance matrices are

$$\Sigma_1 = (1 - p^2) \begin{bmatrix} \pi_1^{-1} & 0 \\ 0 & 0 \end{bmatrix}; \qquad \Sigma_2 = (1 - q^2) \begin{bmatrix} 0 & 0 \\ 0 & \pi_2^{-1} \end{bmatrix},$$

$$\widetilde{\Sigma}_1 = \frac{(1 - p^2)}{4p^2} \begin{bmatrix} \pi_1^{-2} & 0 \\ 0 & 0 \end{bmatrix}; \qquad \widetilde{\Sigma}_2 = \frac{(1 - q^2)}{4q^2} \begin{bmatrix} 0 & 0 \\ 0 & \pi_2^{-2} \end{bmatrix},$$

and $\rho_\mathrm{A} = \rho_\mathrm{L} = \infty$. Therefore, ASE and LSE are equivalent with respect to the subsequent inference task. In contrast, [35] showed that the within-block variances for ASE are four times larger than that of the within-block variances of LSE, while the between-block variances for ASE and LSE are the same. We conclude that the within-block variances measure fails to capture the fact that the block-conditional covariance matrices Σ_1 and Σ_2 are singular but in different subspaces, and similarly $\widetilde{\Sigma}_1$ and $\widetilde{\Sigma}_2$ are also singular but in different subspaces, and thus if we had used the within-block variances measure as a surrogate, we would have been misled into believing that LSE is preferable to ASE for this particular subsequent inference task. Indeed, had we ignored the terms $\log \frac{|\Sigma_{kl}(t)|}{|\Sigma_k|^t|\Sigma_l|^{1-t}}$ and $\log \frac{|\widetilde{\Sigma}_{kl}(t)|}{|\widetilde{\Sigma}_k|^t|\widetilde{\Sigma}_l|^{1-t}}$ in the definitions of ρ_A and ρ_L, we would have come to the similar conclusion that $\rho_\mathrm{L} = \max\{\frac{2\pi_1 p^2}{1-p^2}, \frac{2\pi_2 q^2}{1-q^2}\} = 4\max\{\frac{\pi_1 p^2}{1-p^2}, \frac{\pi_2 q^2}{1-q^2}\} = 4\rho_\mathrm{A}$ for sufficiently large n.

As an illustration of the ratio $\rho_\mathrm{A}/\rho_\mathrm{L}$, we first consider the collection of 2-block stochastic blockmodels where $\mathbf{B} = \begin{bmatrix} p^2 & pq \\ pq & q^2 \end{bmatrix}$ for $p, q \in (0, 1)$ and $\boldsymbol{\pi} = (\pi_1, \pi_2)$ with $\pi_1 + \pi_2 = 1$. We note that these $\mathbf{B}$ also have rank 1, and thus the Chernoff information can be computed explicitly. Then for sufficiently large n, ρ_A is approximately

$$\rho_\mathrm{A} \approx \sup_{t \in (0,1)} \frac{nt(1 - t)}{2} (p - q)^2 (t\sigma_1^2 + (1 - t)\sigma_2^2)^{-1},$$

where σ_1 and σ_2 are as specified in equation (3.14) and equation (3.15), respectively. Simple calculations yield

$$\rho_A \approx \left(n(p-q)^2(\pi_1 p^2 + \pi_2 q^2)^2\right)$$
$$/\left(2\left(\sqrt{\pi_1 p^4(1-p^2) + \pi_2 pq^3(1-pq)}\right.\right.$$
$$\left.\left. + \sqrt{\pi_1 p^3 q(1-pq) + \pi_2 q^4(1-q^2)}\right)^2\right)$$

for sufficiently large n. Similarly, denoting by $\widetilde{\sigma}_1^2$ and $\widetilde{\sigma}_2^2$ the variances specified in equation (3.12) and equation (3.13), we have

$$\rho_L \approx \sup_{t \in (0,1)} \frac{nt(1-t)}{2}\left(\frac{p}{\sqrt{\pi_1 p^2 + \pi_2 pq}} - \frac{q}{\sqrt{\pi_1 pq + \pi_2 q^2}}\right)^2 (t\widetilde{\sigma}_1^2 + (1-t)\widetilde{\sigma}_2^2)^{-1}$$

$$\approx \frac{2n(\sqrt{p}-\sqrt{q})^2(\pi_1 p + \pi_2 q)^2}{(\sqrt{\pi_1 p(1-p^2) + \pi_2 q(1-pq)} + \sqrt{\pi_1 p(1-pq) + \pi_2 q(1-q^2)})^2}$$

$$\approx \left(2n(p-q)^2(\pi_1 p + \pi_2 q)^2\right)$$
$$/\left((\sqrt{p}+\sqrt{q})^2(\sqrt{\pi_1 p(1-p^2) + \pi_2 q(1-pq)}\right.$$
$$\left. + \sqrt{\pi_1 p(1-pq) + \pi_2 q(1-q^2)})^2\right)$$

for sufficiently large n. Fixing $\boldsymbol{\pi} = (0.6, 0.4)$, we computed the ratio ρ_A/ρ_L for a range of p and q values, with $p \in [0.2, 0.8]$ and $q = p + r$ where $r \in [-0.15, 0.15]$. The results are plotted in Figure 3. The y-axis of Figure 3 denotes the values of p and the x axis are the values of r. We see from the above figure that in general, neither of the methods, namely adjacency spectral clustering or normalized Laplacian spectral embedding followed by clustering via Gaussian mixture models, dominates over the whole (p, r) parameter space. However, in general one can easily show that LSE is preferable over ASE whenever the block probability matrix $\mathbf{B}$ becomes sufficiently sparse. For example, if we let $p = c\widetilde{p}$ and $q = c\widetilde{q}$ in the setting of Figure 3 and let $c \to 0$, then for sufficiently small c, we have

$$\rho_A \approx \frac{nc^2(\widetilde{p}-\widetilde{q})^2(\pi_1\widetilde{p}^2 + \pi_2\widetilde{q}^2)^2}{2(\sqrt{\pi_1\widetilde{p}^4 + \pi_2\widetilde{p}\widetilde{q}^3} + \sqrt{\pi_1\widetilde{p}^3\widetilde{q} + \pi_2\widetilde{q}^4})^2} \approx \frac{nc^2(\widetilde{p}-\widetilde{q})^2(\pi_1\widetilde{p}^2 + \pi_2\widetilde{q}^2)^2}{2(\sqrt{\widetilde{p}} + \sqrt{\widetilde{q}})^2(\pi_1\widetilde{p}^3 + \pi_2\widetilde{q}^3)},$$

$$\rho_L \approx \frac{nc^2(\widetilde{p}-\widetilde{q})^2(\pi_1\widetilde{p} + \pi_2\widetilde{q})}{2(\sqrt{\widetilde{p}} + \sqrt{\widetilde{q}})^2}$$

and thus, by the Cauchy–Schwarz inequality,

$$\frac{\rho_A}{\rho_L} \approx \frac{(\pi_1\widetilde{p}^2 + \pi_2\widetilde{q}^2)^2}{(\pi_1\widetilde{p}^3 + \pi_2\widetilde{q}^3)(\pi_1\widetilde{p} + \pi_2\widetilde{q})} \leq 1.$$
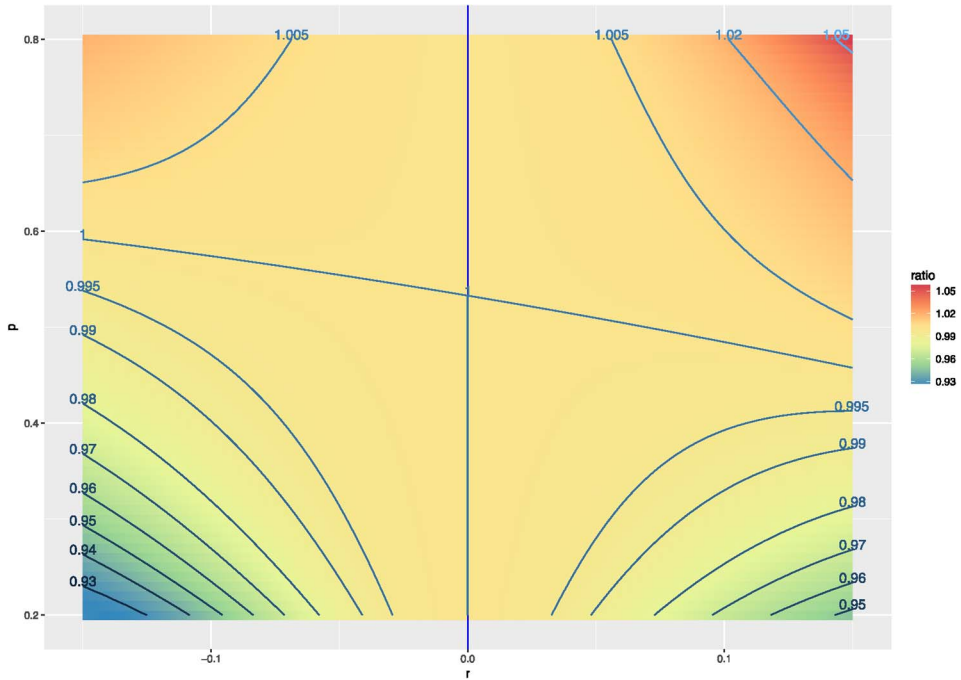
FIG. 3. *The ratio $\rho_{\mathrm{A}}/\rho_{\mathrm{L}}$ displayed for various values of $p \in [0.2, 0.8]$ and $r = q - p \in [-0.15, 0.15]$. The labeled lines are the contour lines for $\rho_{\mathrm{A}}/\rho_{\mathrm{L}}$.*

Determination of similarly intuitive conditions for which ASE dominates over LSE is considerably more subtle and is the topic of current research. But in general, we observe that ASE dominates over LSE whenever the entries of **B** are relatively large.

We also generate instances of a stochastic blockmodel graph on 200 vertices with parameters $p = 0.75$ and $q = 0.6$. For each graph, we measure the error rate of the spectral embedding followed by the Gaussian mixture-model based clustering procedure in recovering the block assignments. The error rate for the GMM ∘ ASE procedure, averaged over 1000 Monte Carlo replicates, is 0.079 with a standard error of 6.6×10^{-4}; meanwhile the error rate for the GMM ∘ LSE procedure, also averaged over 1000 Monte Carlo replicates, is 0.083 with a standard error of 7.2×10^{-6}. The difference in the mean error rate is statistically significant at $\alpha = 0.001$. Conversely, when $p = 0.2$ and $q = 0.3$ and the graphs are on 400 vertices, the mean error rate, over 1000 Monte Carlo replicates, for the GMM ∘ ASE procedure is 0.161 while the mean error rate for the GMM ∘ LSE procedure is 0.151 and this difference is also statistically significant at $\alpha = 0.001$.
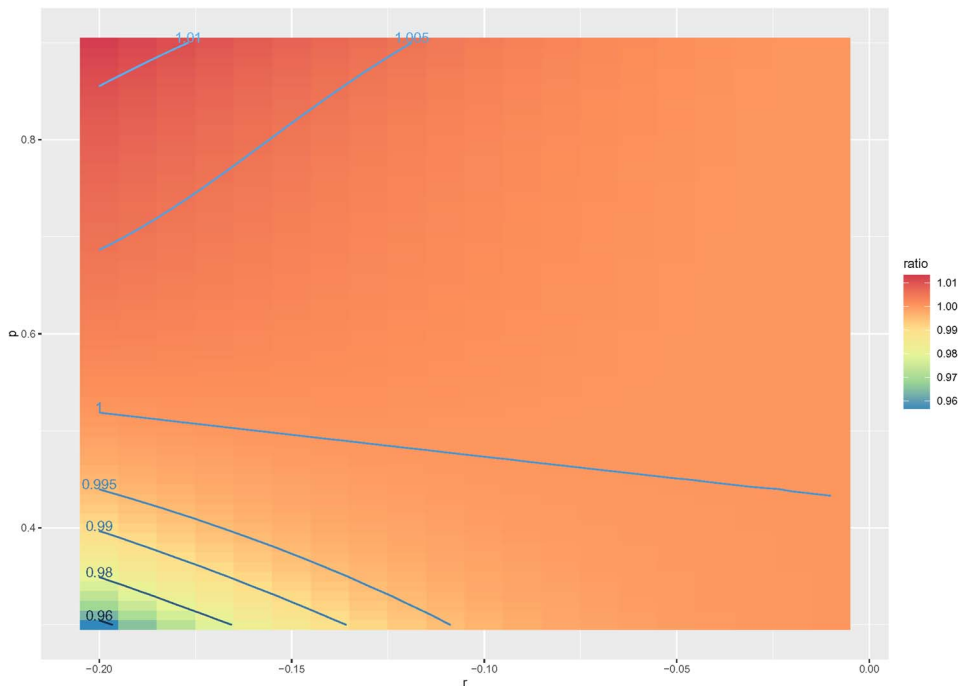
FIG. 4. *The ratio ρ_A/ρ_L displayed for various values of $p \in [0.2, 0.8]$ and $r = q - p \in$ $[-0.2, -0.01]$ for the 3-block stochastic blockmodel of equation* (4.21). *The labeled lines are the contour lines for ρ_A/ρ_L.*

We next consider the collection of stochastic blockmodels with parameters $\boldsymbol{\pi}$ and $\mathbf{B}$ where

$$(4.21) \qquad \mathbf{B} = \begin{bmatrix} p & q & q \\ q & p & q \\ q & q & p \end{bmatrix}, \qquad p, q \in (0, 1) \quad \text{and} \quad \boldsymbol{\pi} = (0.8, 0.1, 0.1).$$

First, we compute the ratio ρ_A/ρ_L for $p \in [0.3, 0.9]$ and $r = q - p$ with $r \in [-0.2, -0.01]$. The results are plotted in Figure 4, with the y-axis of Figure 4 being the values of p and the x-axis being the values of r. Once again we see that for the purpose of subsequent inference, neither embedding methods dominates over the whole parameter space and that LSE is still preferable to ASE for smaller values of p and q and that ASE is preferable to LSE for larger values of p and q. We also generate instances of a stochastic blockmodel graph on 800 vertices with $p = 0.9$ and $q = 0.72$ and estimate the error rate of the GMM ∘ ASE and the GMM ∘ LSE procedures in recovering the block assignments. The GMM ∘ ASE and GMM ∘ LSE error rates, averaged over 1000 Monte Carlo replicates, are 0.29 and 0.38, respectively. For these choice of parameters, $\rho_A/\rho_L \approx 1.01$. We also generate instances of a stochastic blockmodel graph on 1600 vertices with

$p = 0.34$ and $q = 0.15$. The ratio ρ_A/ρ_L in this case is ≈ 0.98; the GMM $\circ$ ASE and GMM $\circ$ LSE error rates, averaged over 1000 Monte Carlo replicates, are 0.18 and 0.06, respectively.

5. Summary and conclusions. We have shown in this paper several limit results for the eigenvectors corresponding to the largest eigenvalues of the normalized Laplacian matrix of random graphs. In particular, we show that for stochastic blockmodel graphs, conditioned on the block assignments, each row of the Laplacian spectral embedding converges to a multivariate normal distribution. We then discuss the relationship between spectral embeddings of the adjacency and normalized Laplacian matrices and subsequent inference. When the subsequent inference task is the problem of clustering the vertices of a graph, we show that the Chernoff information between the multivariate normals approximation of the embedding is a suitable measure for the large-sample optimal error rate, that is, it characterizes the minimum error rate achievable by *any* clustering procedure that operates only on the spectral embedding. As a result, we are able to theoretically compare the use of spectral embedding of the adjacency matrix versus that of the normalized Laplacian for subsequent inference, thereby refining and extending the pioneering work of [35].

The limit results in Section 2.1 and Section 3 of this paper assume that $n\rho_n$ grows at the rate of $\omega(\log^4(n))$ as $n \to \infty$. This technical condition is from Theorem 6 of [25], which provides us with concentration inequalities for $\|\mathbf{A} - \mathbf{P}\|$ and $\|\mathcal{L}(\mathbf{A}) - \mathcal{L}(\mathbf{P})\|$. Similar results to Theorem 6 of [25] but under weaker conditions exist; for example, Theorem 5.2 of [23] and Theorem 4 of [20] showed concentration of the adjacency matrix and the Laplacian matrix, around their respective expectation, under the weaker condition that $n\rho_n = \omega(\log n)$. These results can then be applied to this paper to show that the limit results in Section 2.1 and Section 3 hold for $n\rho_n = \omega(\log(n))$. Finally, the limit results of this paper do not hold for $n\rho_n = o(\log n)$ since, as it is widely known, neither the adjacency matrix nor the normalized Laplacian matrix concentrate around their expectation in this case.

We now mention several potential extensions of this work. The normalized Laplacian considered in this paper is just one example of possible normalization. In particular, given $\tau > 0$ one can define the τ-regularized normalized Laplacian $\mathcal{L}_\tau$ via $\mathcal{L}_\tau(\mathbf{A}) = (\mathbf{D} + \tau\mathbf{I})^{-1/2}\mathbf{A}(\mathbf{D} + \tau\mathbf{I})^{-1/2}$ or $\mathcal{L}_\tau(\mathbf{A}) = (\mathbf{D} + \tau\mathbf{I})^{-1/2}(\mathbf{A} + \tau\mathbf{1}\mathbf{1}^\top)(\mathbf{D} + \tau\mathbf{I})^{-1/2}$[2, 9, 33]. It had been shown that regularization is particularly useful for spectral clustering in sparse graphs. It will thus be of interest to derive limit results for the eigenvectors of $\mathcal{L}_\tau(\mathbf{A})$ analogous to those in this paper; such results can potentially allow one to choose the regularization parameter τ.

The limit results in this paper are for the spectral embedding of $(\mathbf{X}, \mathbf{A}) \sim$ RPDG(F) into $\mathbb{R}^d$ when d, the rank of the matrix $\mathbb{E}[XX^\top]$ where $X \sim F$, is fixed and known. Similar results can be derived when the spectral embedding of $\mathbf{A}$ is into $\mathbb{R}^{d'}$ where $d' < d$. Limit results for spectral embedding of the adjacency matrix or Laplacian matrix into $\mathbb{R}^{d'}$ when $d' > d$ is, to the best of our knowledge, an open

problem. A related inquiry is limit results for spectral embedding into $\mathbb{R}^{d'}$ when $d' < d$ but d varies with n and is not fixed, such as when the graph arises from a latent position model where the link function, viewed as an integral operator, has infinite rank. Since new results on stochastic blockmodels indicate that they can be regarded as a universal approximation to latent positions model graphs or graphons of exchangeable random graphs [44, 45], limit results for the adjacency and Laplacian spectral embedding will be useful in further understanding of this approximation property.

Finally, the Chernoff information used in this paper is a measure of the effect of spectral embedding on subsequent inference for a single graph. Recently, there have been interests in the two-sample inference for graphs, for example, network comparisons or two-sample hypothesis testing for graphs [3, 41, 42]. As an example, given two distributions F and G, the problem of testing whether $F = G$ given two random dot product graphs $\mathbf{A} \sim \mathrm{RDPG}(F)$ and $\mathbf{B} \sim \mathrm{RDPG}(G)$ was considered in [42]; the proposed test statistic is a kernel-based distance measure between the spectral embedding $\hat{\mathbf{X}}$ of $\mathbf{A}$ and $\hat{\mathbf{Y}}$ of $\mathbf{B}$. Determining a measure that characterizes the effect of spectral embedding for two-sample graphs inference problems, akin to how the Chernoff information characterizes the effect of spectral embedding for single graph inference, is of significant interest.

APPENDIX A: PROOF OF THEOREM 2.1 AND THEOREM 2.2

We first present a sketch of the proof of Theorem 2.1, noting that the main arguments are given in [41]. We also note that similar, albeit more involved, arguments are used in the proof of Theorem 3.1. Since the proof of Theorem 3.1 will be presented in much greater detail in Section B, to avoid repetitions, we chose to omit the details in the current proof. Nevertheless, we emphasize that the statements of the results in [41] are slightly different from how they are stated in the current paper; these differences stem mainly from how sparseness in the graphs is incorporated. More specifically, [41] considered a sequence of random dot product graphs where for each n, the matrix of latent positions $\mathbf{X}_n$ are fixed but unknown (see Definition 1 in [41]) and furthermore, there need not exist any relationship between $\mathbf{X}_n$ and $\mathbf{X}_{n'}$ for $n \neq n'$. Sparseness of the graphs is thus implicit (see, e.g., the condition on the minimum vertex's degree in Assumption 1 in [41]). The current paper, however, assumes that the rows of $\mathbf{X}_n$ are independently sampled according to a distribution F. As such, sparseness needs to be made explicit through the sparsity factor ρ_n.

REMARK. For ease of exposition, henceforth we shall on many occasions remove the subscript n from the matrices $\mathbf{X}_n, \hat{\mathbf{X}}_n, \mathbf{A}_n, \mathbf{P}_n$ and other related matrices such as $\mathbf{U}_{\mathbf{A}_n}, \mathbf{U}_{\mathbf{P}_n}$, etc. The subsequent statements are thus to be interpreted as holding for sufficient large n. Since we are concerned with limit results, this should lead to minimal confusion.

We first note that equation (2.5) follows from Theorem A.5 in [41]. More specifically, if $(\mathbf{X}, \mathbf{A}) \sim \mathrm{RDPG}(F)$ with sparsity factor ρ_n, then Theorem A.5 in [41] yields

$$\big\|\hat{\mathbf{X}} - \rho_n^{1/2}\mathbf{X}\mathbf{W}\big\|_F = \big\|(\mathbf{A} - \mathbf{P})\mathbf{U_P}\mathbf{S_P}^{-1/2}\big\|_F + O_{\mathbb{P}}\big((n\rho_n)^{-1/2}\big).$$

Since $\mathbf{P} = \rho_n \mathbf{X}\mathbf{X}^\top$ we have $\mathbf{U_P}\mathbf{S_P}^{1/2}\mathbf{W} = \rho_n^{1/2}\mathbf{X}$ for some orthogonal matrix $\mathbf{W}$. Therefore,

$$\begin{aligned}
\big\|(\mathbf{A} - \mathbf{P})\mathbf{U_P}\mathbf{S_P}^{-1/2}\big\|_F &= \big\|(\mathbf{A} - \mathbf{P})\mathbf{U_P}\mathbf{S_P}^{1/2}\mathbf{W}\mathbf{W}^\top\mathbf{S_P}^{-1}\mathbf{W}\big\|_F \\
&= \big\|(\mathbf{A} - \mathbf{P})\rho_n^{1/2}\mathbf{X}(\rho_n\mathbf{X}^\top\mathbf{X})^{-1}\big\|_F \\
&= \rho_n^{-1/2}\big\|(\mathbf{A} - \mathbf{P})\mathbf{X}(\mathbf{X}^\top\mathbf{X})^{-1}\big\|_F.
\end{aligned}$$

Equation (2.5) is thus established. We now show equation (2.6) and equation (2.7). We shall use the convention that, unless stated otherwise, expectation of a random variable dependent on $\mathbf{A}$ is taken with respect to $\mathbf{A}$ conditional on $\mathbf{P}$. Let $\zeta = \rho_n\|(\mathbf{A} - \mathbf{P})\mathbf{U_P}\mathbf{S_P}^{-1/2}\|_F^2$. Then, conditional on $\mathbf{P}$, ζ is a linear function of the *independent* random variables $\{a_{ij} - p_{ij}\}_{i<j}$. Lemma A.5 in [41] shows that ζ is tightly concentrated around its expectation $\mathbb{E}[\zeta]$. We then have

$$\begin{aligned}
\mathbb{E}[\zeta] &= \mathbb{E}\big[\big\|(\mathbf{A} - \mathbf{P})\mathbf{U_P}\mathbf{S_P}^{-1/2}\big\|_F^2\big] \\
&= \rho_n^{-1}\mathbb{E}\big[\big\|(\mathbf{A} - \mathbf{P})\mathbf{X}(\mathbf{X}^\top\mathbf{X})^{-1}\big\|_F^2\big] \\
&= \operatorname{tr} n(\mathbf{X}^\top\mathbf{X})^{-1}\big(n^{-2}\rho_n^{-1}\mathbf{X}^\top\mathbb{E}[(\mathbf{A} - \mathbf{P})^2]\mathbf{X}\big)n(\mathbf{X}^\top\mathbf{X})^{-1}.
\end{aligned}$$

Now, the ijth entry of $(\mathbf{A} - \mathbf{P})^2$ is of the form $\sum_k (a_{ik} - p_{ik})(a_{kj} - p_{kj})$. As the upper diagonal entries of $\mathbf{A}$ are independent conditional on $\mathbf{P}$, we have

$$\mathbb{E}\Bigg[\sum_k (a_{ik} - p_{ik})(a_{kj} - p_{kj})\Bigg] = \begin{cases} 0 & \text{if } i \neq j, \\ \displaystyle\sum_{k \neq i} p_{kj}(1 - p_{kj}) & \text{if } i = j. \end{cases}$$

By the strong law of large numbers, $n^{-1}\mathbf{X}^T\mathbf{X}_n = n^{-1}\sum_i X_i X_i^\top$ converges to $\Delta = \mathbb{E}[X_1 X_1^\top]$ almost surely as $n \to \infty$. Hence $n(\mathbf{X}^\top\mathbf{X})^{-1}$ converges to Δ^{-1} almost surely. In addition,

$$\begin{aligned}
n^{-2}\rho_n^{-1}\mathbf{X}^T\mathbb{E}[(\mathbf{A} - \mathbf{P})^2]\mathbf{X} &= n^{-2}\rho_n^{-1}\sum_{i=1}^n\sum_{j \neq i} X_i X_i^\top p_{ik}(1 - p_{ik}) \\
&= n^{-2}\rho_n^{-1}\sum_{i=1}^n\sum_{k \neq i} X_i X_i^\top\big(\rho_n X_i^\top X_k - \rho_n^2(X_i^\top X_k)^2\big) \\
&= n^{-2}\sum_{i=1}^n\sum_{k \neq i} X_i X_i^\top\big(X_i^\top X_k - \rho_n X_i^\top X_k X_k^\top X_i\big).
\end{aligned}$$

If $\rho_n = 1$ for all n, the above term converges to $\mathbb{E}[X_1 X_1^\top (X_1^\top \mu_F - X_1^\top \Delta_F X_1)]$ almost surely. When $\rho_n \to 0$, the above term converges to $\mathbb{E}[X_1 X_1^\top X_1^\top \mu_F]$ almost surely. Equation (2.6) and equation (2.7) is thus established.

We now sketch the proof of Theorem 2.2. We emphasize that Theorem 2.2 is a generalization of the corresponding result in [4, 39], the generalization being that Theorem 2.2 does not assume distinct eigenvalues of the matrix $\mathbb{E}[XX^\top]$ where $X \sim F$; distinct eigenvalues is a necessary assumption for the proof given in [4, 39].

Let a_{ij} and p_{ij} denote the ijth entry of $\mathbf{A}$ and $\mathbf{P}$. From equation (2.5), by exchangeability of the collection $\{\mathbf{W}_n \hat{X}_j - \rho_n^{1/2} X_j\}_{j=1}^n$, for any fixed index i we have

$$\sqrt{n}(\mathbf{W}_n \hat{X}_i - \rho_n^{1/2} X_i) = \sqrt{n} \rho_n^{-1/2} (\mathbf{X}^\top \mathbf{X})^{-1} \sum_{j \neq i} (a_{ij} - p_{ij}) X_j + o_\mathbb{P}(1)$$

$$= \rho_n^{-1/2} (n^{-1} \mathbf{X}^\top \mathbf{X})^{-1} \sum_{j \neq i} \frac{(a_{ij} - p_{ij})}{\sqrt{n}} X_j + o_\mathbb{P}(1)$$

$$= (n^{-1} \mathbf{X}^\top \mathbf{X})^{-1} \sum_{j \neq i} \frac{(a_{ij} - \rho_n X_i^\top X_j)}{\sqrt{n\rho_n}} X_j + o_\mathbb{P}(1).$$

Now conditional on X_i, the quantity $\sum_{j \neq i} \frac{(a_{ij} - \rho_n X_i^\top X_j)}{\sqrt{n\rho_n}} X_j$ is a sum of independent and identically distributed mean 0 random variables. Thus by the multivariate central limit theorem, conditioning on $X_i = x$ yields

$$\sum_{j \neq i} \frac{(a_{ij} - \rho_n x^\top X_j)}{\sqrt{n\rho_n}} X_j \xrightarrow{\mathrm{d}} \mathcal{N}(0, \mathbb{E}[X_1 X_1^\top (x^\top X_1 - \rho_n x^\top X_1 X_1^\top x)]).$$

Furthermore, since $n^{-1} \mathbf{X}^\top \mathbf{X} = n^{-1} \sum X_i X_i^\top \xrightarrow{\text{a.s.}} \Delta$ as $n \to \infty$, we have by Slutsky's theorem that

$$\sqrt{n}(\mathbf{W}_n \hat{X}_i - \rho_n^{1/2} X_i) \xrightarrow{\mathrm{d}} \mathcal{N}(0, \Delta^{-1} \mathbb{E}[X_1 X_1^\top (x^\top X_1 - \rho_n x^\top X_1 X_1^\top x)] \Delta^{-1}),$$

thereby establishing Theorem 2.2.

APPENDIX B: PROOF OF THEOREM 3.1 AND THEOREM 3.2

For ease of exposition, we present in Section B.1 a proof of Theorem 3.2, assuming equation (3.1) in Theorem 3.1 holds. We next derive, in Section B.2, equation (3.1) in Theorem 3.1. We then show, in Section B.4 that the Frobenius norms in equation (3.4) and equation (3.5) are tightly concentrated around their expectations. We complete the proof of Theorem 3.1 by computing these expectations explicitly when $\rho_n \equiv 1$ and when $\rho_n \to 0$.

B.1. Proof of Theorem 3.2. Recall that we suppress the dependency on n in the subscript of the matrices $\mathbf{A}_n$, $\mathbf{X}_n$, $\mathbf{P}_n$ and other related matrices. In addition, recall that $\widetilde{\mathbf{X}} = \rho_n^{1/2}\mathbf{T}^{-1/2}\mathbf{X} = \mathrm{diag}(\mathbf{XX}^\top\mathbf{1})^{-1/2}\mathbf{X}$. Equation (3.1) from Theorem 3.1 then implies

$$\check{\mathbf{X}}\mathbf{W} - \widetilde{\mathbf{X}} = \mathbf{T}^{-1/2}(\mathbf{A} - \mathbf{P})\mathbf{T}^{-1/2}\widetilde{\mathbf{X}}(\widetilde{\mathbf{X}}^\top\widetilde{\mathbf{X}})^{-1} + \frac{1}{2}\mathbf{T}^{-1}(\mathbf{T} - \mathbf{D})\widetilde{\mathbf{X}} + \mathbf{R}$$

for some orthogonal matrix $\mathbf{W}$ and $n \times d$ matrix $\mathbf{R}$ with $\|\mathbf{R}\|_F = O_\mathbb{P}((n\rho_n)^{-1})$. For a fixed index i, let ζ_i denotes the ith row of $n\rho_n^{1/2}(\check{\mathbf{X}}\mathbf{W} - \widetilde{\mathbf{X}})$. Also let r_i denote the ith row of $\mathbf{R}$. Now exchangeability of the $\{X_j\}_{j=1}^n$ implies exchangeability of the $\{\check{X}_i\}_{j=1}^n$ and exchangeability of the $\{\widetilde{X}_i\}_{j=1}^n$. This also implies exchangeability of the $\{\zeta_j\}_{j=1}^n$ and thus exchangeability of the $\{r_j\}_{j=1}^n$. Now, for any fixed index i, by exchangeability of the $\{r_j\}_{j=1}^n$, we have

$$n^2\rho_n\mathbb{E}[\|r_i\|^2] = n^2\rho_n\frac{1}{n}\mathbb{E}\left[\sum_j\|r_j\|^2\right] = n\rho_n\mathbb{E}[\|\mathbf{R}\|_F^2]$$

Now, with probability at least $1 - n^{-3}$, $\|\mathbf{R}\|_F \leq C_0(n\rho_n)^{-1})$ for some constant C_0. In addition, $\|\mathbf{R}\|_F \leq n$ almost surely. Thus $\mathbb{E}[\|\mathbf{R}\|_F^2] \leq C_0^2(n\rho_n)^{-2}(1 - n^{-3}) + n \times n^{-3} = O((n\rho_n)^{-2})$. Therefore, $n^2\rho_n\mathbb{E}[\|r_i\|^2] = O((n\rho_n)^{-1})$. Since $n\rho_n = \omega(\log^4(n))$, we therefore have $n^2\rho_n\mathbb{E}[\|r_i\|^2] \to 0$ as $n \to \infty$, that is, $n\rho_n^{1/2}r_i \xrightarrow{d} 0$ as $n \to \infty$.

Let a_{ij} and p_{ij} denote the ijth entry of $\mathbf{A}$ and $\mathbf{P}$, respectively. The above reasoning implies that for a fixed index i, ζ_i is of the form

$$\zeta_i = (\widetilde{\mathbf{X}}^\top\widetilde{\mathbf{X}})^{-1}\frac{n\rho_n^{1/2}}{\sqrt{t_i}}\left(\sum_j\frac{a_{ij} - p_{ij}}{\sqrt{t_j}}\widetilde{X}_j\right) + \frac{n\rho_n^{1/2}(t_i - d_i)}{2t_i}\widetilde{X}_i + o_\mathbb{P}(1)$$

$$= (\widetilde{\mathbf{X}}^\top\widetilde{\mathbf{X}})^{-1}\frac{\sqrt{n\rho_n}}{\sqrt{t_i}}\left(\sum_{j\neq i}\frac{\sqrt{n\rho_n}(a_{ij} - p_{ij})X_j}{t_j}\right)$$

$$- \frac{(n\rho_n)^{3/2}X_i}{2t_i^{3/2}}\sum_{j\neq i}\frac{(a_{ij} - p_{ij})}{\sqrt{n\rho_n}} + o_\mathbb{P}(1).$$

We first note that $\widetilde{\mathbf{X}}^\top\widetilde{\mathbf{X}} = \mathbf{X}^\top\mathrm{diag}(\mathbf{XX1})^{-1}\mathbf{X}$ converges almost surely to $\widetilde{\Delta}$ as $n \to \infty$. This can be seen as follows. Denoting $\mu = \mathbb{E}[X_1]$, we have

$$\widetilde{\mathbf{X}}^\top\widetilde{\mathbf{X}} = \sum_{i=1}^n\frac{X_iX_i^\top}{\sum_j X_i^\top X_j}$$

$$= \left(\sum_{i=1}^n\frac{X_iX_i^\top}{nX_i^\top\mu}\right) + \sum_{i=1}^n X_iX_i^\top\left(\frac{1}{\sum_j X_i^\top X_j} - \frac{1}{nX_i^\top\mu}\right)$$

$$= \left(\sum_{i=1}^n\frac{X_iX_i^\top}{nX_i^\top\mu}\right) + \sum_{i=1}^n\frac{X_iX_i^\top}{nX_i^\top\mu}\left(\frac{nX_i^\top\mu - \sum_j X_i^\top X_j}{\sum_j X_i^\top X_j}\right).$$

Now, for any index i, let $c_i = |(nX_i^\top \mu - \sum_j X_i^\top X_j)/(\sum_j X_i^\top X_j)|$. Then by Hoeffding's inequality, $c_i = O_\mathbb{P}(n^{-1/2})$. As $X_i X_i^\top$ is positive semidefinite for each index i, we thus have

$$-c_i \frac{X_i X_i^\top}{nX_i^\top \mu} \preceq \frac{X_i X_i^\top}{nX_i^\top \mu}\left(\frac{nX_i^\top \mu - \sum_j X_i^\top X_j}{\sum_j X_i^\top X_j}\right) \preceq c_i \frac{X_i X_i^\top}{nX_i^\top \mu},$$

where $\preceq$ denotes the positive semidefinite ordering of matrices. Hence

$$-\left(\sup_{j\in[n]} c_j\right)\sum_i \frac{X_i X_i^\top}{nX_i^\top \mu} \preceq \sum_i \frac{X_i X_i^\top}{nX_i^\top \mu}\left(\frac{nX_i^\top \mu - \sum_j X_i^\top X_j}{\sum_j X_i^\top X_j}\right)$$

$$\preceq \left(\sup_{j\in[n]} c_j\right)\sum_i \frac{X_i X_i^\top}{nX_i^\top \mu}.$$

We then have by a union bound that $\sup_{i\in[n]} c_i = O_\mathbb{P}(\sqrt{n^{-1}\log n})$, and hence $\sup_{i\in[n]} c_i \xrightarrow{\text{a.s.}} 0$ as $n \to \infty$. In addition, by the strong law of large numbers

(B.1) $$\sum_i \frac{X_i X_i^\top}{nX_i^\top \mu} \xrightarrow{\text{a.s.}} \mathbb{E}\left[\frac{X_1 X_1^\top}{X_1^\top \mu}\right]$$

as $n \to \infty$. Thus

$$\sum_i \frac{X_i X_i^\top}{nX_i^\top \mu}\left(\frac{nX_i^\top \mu - \sum_j X_i^\top X_j}{\sum_j X_i^\top X_j}\right) \xrightarrow{\text{a.s.}} 0$$

as $n \to \infty$. We thus conclude that

(B.2) $$\widetilde{\mathbf{X}}^\top \widetilde{\mathbf{X}} = \left(\sum_{i=1}^n \frac{X_i X_i^\top}{nX_i^\top \mu}\right) + \sum_{i=1}^n \frac{X_i X_i^\top}{nX_i^\top \mu}\frac{nX_i^\top \mu - \sum_j X_i^\top X_j}{\sum_j X_i^\top X_j} \xrightarrow{\text{a.s.}} \mathbb{E}\left[\frac{X_1 X_1^\top}{X_1^\top \mu}\right]$$

as $n \to \infty$.

Therefore, $(\widetilde{\mathbf{X}}^\top \widetilde{\mathbf{X}})^{-1}$ converges almost surely to $\widetilde{\Delta}^{-1}$ as $n \to \infty$. In addition, $t_i/(n\rho_n) \to X_i^\top \mu$ as $n \to \infty$, and hence $\sqrt{n\rho_n/t_i} \to (X_i^\top \mu)^{-1/2}$ as $n \to \infty$. We next consider the term

$$\sum_{j\neq i} \frac{\sqrt{n\rho_n}(a_{ij} - p_{ij})X_j}{t_j} = \sum_{j\neq i} \frac{(a_{ij} - p_{ij})X_j}{\sqrt{n\rho_n}X_j^\top \mu} + \sum_{j\neq i} \frac{(a_{ij} - p_{ij})X_j}{\sqrt{n\rho_n}X_j^\top \mu}\frac{n\rho_n X_j^\top \mu - t_j}{t_j}.$$

The second sum on the right-hand side of the above display is, conditioned on $\mathbf{P}$, a sum of mean 0 random variables. Hoeffding's inequality implies that the event

$$\left\|\sum_{j\neq i} \frac{(a_{ij} - p_{ij})X_j}{\sqrt{n\rho_n}X_j^\top \mu}\frac{n\rho_n X_j^\top \mu - t_j}{t_j}\right\| \geq s$$

occurs with probability at most

$$2\exp\left(\frac{-Cn\rho_n s^2}{\sum_{j\neq i}\|X_j\|^2(n\rho_n X_j^\top\mu-t_j)^2(X_j^\top\mu)^{-2}t_j^{-2}}\right)$$

for some constant $C>0$. Therefore,

$$\sum_{j\neq i}\frac{(a_{ij}-p_{ij})X_j}{\sqrt{n\rho_n}X_j^\top\mu}\frac{n\rho_n X_j^\top\mu-t_j}{t_j}\xrightarrow{\text{a.s.}}0$$

as $n\to\infty$. We thus have

$$\zeta_i=(\widetilde{\mathbf{X}}^\top\widetilde{\mathbf{X}})^{-1}\frac{\sqrt{n\rho_n}}{\sqrt{t_i}}\left(\sum_{j\neq i}\frac{(a_{ij}-p_{ij})X_j}{\sqrt{n\rho_n}X_j^\top\mu}\right)$$

(B.3)
$$-\frac{n\rho_n\sqrt{n\rho_n}X_i}{2t_i\sqrt{t_i}}\sum_{j\neq i}\frac{(a_{ij}-p_{ij})}{\sqrt{n\rho_n}}+o_{\mathbb{P}}(1).$$

We now show that

$$\frac{n\rho_n\sqrt{n\rho_n}X_i}{2t_i\sqrt{t_i}}\sum_{j\neq i}\frac{(a_{ij}-p_{ij})}{\sqrt{n\rho_n}}$$

(B.4)
$$=\frac{\sqrt{n\rho_n}(\widetilde{\mathbf{X}}^\top\widetilde{\mathbf{X}})^{-1}\widetilde{\Delta}X_i}{2\sqrt{t_i}X_i^\top\mu}\sum_{j\neq i}\frac{(a_{ij}-p_{ij})}{\sqrt{n\rho_n}}+o_{\mathbb{P}}(1).$$

This can be done as follows. We first consider the term

$$\frac{n\rho_n\sqrt{n\rho_n}X_i}{2t_i\sqrt{t_i}}\sum_{j\neq i}\frac{(a_{ij}-p_{ij})}{\sqrt{n\rho_n}}$$

$$=\frac{\sqrt{n\rho_n}}{2\sqrt{t_i}}\left(\sum_{j\neq i}\frac{(a_{ij}-p_{ij})}{\sqrt{n\rho_n}}\right)\left(\frac{X_i}{X_i^\top\mu}+\frac{n\rho_n X_i}{t_i}-\frac{X_i}{X_i^\top\mu}\right).$$

Once again, conditional on $\mathbf{P}$,

$$\frac{\sqrt{n\rho_n}}{2\sqrt{t_i}}\left(\sum_{j\neq i}\frac{(a_{ij}-p_{ij})}{\sqrt{n\rho_n}}\right)\left(\frac{n\rho_n X_i}{t_i}-\frac{X_i}{X_i^\top\mu}\right)$$

is a sum of mean 0 random variable. Hence, by Hoeffding's inequality, we also have that

$$\frac{\sqrt{n}\rho_n}{2\sqrt{t_i}}\left(\sum_{j\neq i}\frac{(a_{ij}-p_{ij})}{\sqrt{n\rho_n}}\right)\left(\frac{n\rho_n X_i}{t_i}-\frac{X_i}{X_i^\top\mu}\right)\xrightarrow{\text{a.s.}}0$$

as $n\to\infty$. We thus have

(B.5) $\quad\dfrac{n\rho_n\sqrt{n\rho_n}X_i}{2t_i\sqrt{t_i}}\displaystyle\sum_{j\neq i}\dfrac{(a_{ij}-p_{ij})}{\sqrt{n\rho_n}}=\dfrac{\sqrt{n\rho_n}}{2\sqrt{t_i}}\sum_{j\neq i}\dfrac{(a_{ij}-p_{ij})X_i}{\sqrt{n\rho_n}X_i^\top\mu}+o_{\mathbb{P}}(1).$

We next write

$$\frac{\sqrt{n\rho_n}}{2\sqrt{t_i}}\sum_{j\neq i}\frac{(a_{ij}-p_{ij})X_i}{\sqrt{n\rho_n}X_i^\top\mu}=\frac{\sqrt{n\rho_n}(\widetilde{\mathbf{X}}^\top\widetilde{\mathbf{X}})^{-1}}{2\sqrt{t_i}}\sum_{j\neq i}\frac{(a_{ij}-p_{ij})(\widetilde{\Delta}+\widetilde{\mathbf{X}}^\top\widetilde{\mathbf{X}}-\widetilde{\Delta})X_i}{\sqrt{n\rho_n}X_i^\top\mu}.$$

We again evoke Hoeffding's inequality conditionally on $\mathbf{P}$ to conclude that

$$(B.6) \qquad \frac{\sqrt{n\rho_n}(\widetilde{\mathbf{X}}^\top\widetilde{\mathbf{X}})^{-1}}{2\sqrt{t_i}}\sum_{j\neq i}\frac{(a_{ij}-p_{ij})(\widetilde{\mathbf{X}}^\top\widetilde{\mathbf{X}}-\widetilde{\Delta})X_i}{\sqrt{n\rho_n}X_i^\top\mu}\xrightarrow{\text{a.s.}}0$$

as $n\to\infty$. Equation (B.4) then follows from equation (B.5) and equation (B.6).

Combining equation (B.3) and equation (B.4), we arrive at

$$\zeta_i=(\widetilde{\mathbf{X}}^\top\widetilde{\mathbf{X}})^{-1}\frac{\sqrt{n\rho_n}}{\sqrt{t_i}}\left(\sum_{j\neq i}\frac{(a_{ij}-p_{ij})}{\sqrt{n\rho_n}}\left(\frac{X_j}{X_j^\top\mu}-\frac{\widetilde{\Delta}X_i}{2X_i^\top\mu}\right)\right)+o_{\mathbb{P}}(1)$$

$$=(\widetilde{\mathbf{X}}^\top\widetilde{\mathbf{X}})^{-1}\frac{\sqrt{n\rho_n}}{\sqrt{t_i}}\left(\sum_{j\neq i}\frac{(a_{ij}-\rho_n X_i^\top X_j)}{\sqrt{n\rho_n}}\left(\frac{X_j}{X_j^\top\mu}-\frac{\widetilde{\Delta}X_i}{2X_i^\top\mu}\right)\right)+o_{\mathbb{P}}(1).$$

Now, for each fixed index i, conditioning on $X_i=x$, the quantity

$$(B.7) \qquad \frac{1}{\sqrt{n\rho_n}}\sum_{j\neq i}(a_{ij}-\rho_n X_i^\top X_j)\left(\frac{X_j}{X_j^\top\mu}-\frac{\widetilde{\Delta}X_i}{2X_i^\top\mu}\right)$$

is a sum of independent and identically distributed mean 0 random variables. Therefore, by the multivariate central limit theorem, we have that conditional on $X_i=x$, the term in equation (B.7) converges in distribution to

$$\mathcal{N}\left(0,\mathbb{E}\left[\left(\frac{X_j}{X_j^\top\mu}-\frac{\widetilde{\Delta}x}{2x^\top\mu}\right)(x^\top X_j-\rho_n x^\top X_j X_j^\top x)\left(\frac{X_j}{X_j^\top\mu}-\frac{\widetilde{\Delta}x}{2x^\top\mu}\right)^\top\right]\right).$$

Finally, recall that $(\widetilde{\mathbf{X}}^\top\widetilde{\mathbf{X}})^{-1}$ and $\sqrt{n\rho_n/t_i}$ converge almost surely to $\widetilde{\Delta}^{-1}$ and $(X_i^\top\mu)^{-1/2}$ as $n\to\infty$. Therefore, by Slutsky's theorem, conditional on $X_i=x$, $\zeta_i=n\rho_n^{1/2}(\mathbf{W}\breve{X}_i-\widetilde{X}_i)$ converges in distribution to

$$\mathcal{N}\left(0,\mathbb{E}\left[\left(\frac{\widetilde{\Delta}^{-1}X_j}{X_j^\top\mu}-\frac{x}{2x^\top\mu}\right)\left(\frac{x^\top X_j-\rho_n x^\top X_j X_j^\top x}{x^\top\mu}\right)\left(\frac{\widetilde{\Delta}^{-1}X_j}{X_j^\top\mu}-\frac{x}{2x^\top\mu}\right)^\top\right]\right)$$

as desired.

B.2. Proof of equation (3.1). We start with a concentration inequality for the spectral norm of $\mathbf{A}-\mathbf{P}$ and $\mathcal{L}(\mathbf{A})-\mathcal{L}(\mathbf{P})$ in the case when $\mathbf{A}$ is an edge-independent inhomogeneous random graph.

LEMMA B.1 ([25, 32]).   *Let* $\mathbf{A} \sim$ Bernoulli($\mathbf{P}$), *that is,* $\mathbf{A}$ *is a symmetric matrix whose upper triangular entries are independent Bernoulli random variables with* $\mathbb{P}[a_{ij} = 1] = p_{ij}$. *Let* $\Delta = \max_i \sum_{j \neq i} p_{ij}$ *and* $\delta = \min_i \sum_{j \neq i} p_{ij}$ *denote the maximum and minimum row sums of* $\mathbf{P}$. *Suppose* δ *satisfies* $\delta \gg \log^4(n)$. *Then*

$$\|\mathbf{A} - \mathbf{P}\| = O_{\mathbb{P}}(\sqrt{\Delta}),$$

$$\|\mathcal{L}(\mathbf{A}) - \mathcal{L}(\mathbf{P})\| = O_{\mathbb{P}}(\delta^{-1/2}).$$

When $\mathbf{P} = \rho_n \mathbf{X}\mathbf{X}^\top$ then δ and Δ are both of order $\Theta(n\rho_n)$. Furthermore, the nonzero eigenvalues of $\mathbf{P}$ are all of order $\Theta(n\rho_n)$ while the nonzero eigenvalues of $\mathcal{L}(\mathbf{P})$ are all of order $\Theta(1)$. In light of Lemma B.1, for our subsequent derivation, we shall assume that $\rho_n = \omega(\log^k(n))$ for some positive integer $k \geq 4$.

Lemma B.1 implies the following proposition.

PROPOSITION B.2.   *Let* $(\mathbf{A}, \mathbf{X}) \sim$ RDPG(F) *with sparsity factor* ρ_n. *Let* $\mathbf{W}_1 \mathbf{\Sigma} \mathbf{W}_2^\top$ *be the singular value decomposition of* $\widetilde{\mathbf{U}}_{\mathbf{P}}^\top \widetilde{\mathbf{U}}_{\mathbf{A}}$. *Then*

$$\|\widetilde{\mathbf{U}}_{\mathbf{P}}^\top \widetilde{\mathbf{U}}_{\mathbf{A}} - \mathbf{W}_1 \mathbf{W}_2^\top\|_F = O_{\mathbb{P}}((n\rho_n)^{-1}).$$

PROOF.   Let $\sigma_1, \sigma_2, \ldots, \sigma_d$ denote the singular values of $\widetilde{\mathbf{U}}_{\mathbf{P}}^\top \widetilde{\mathbf{U}}_{\mathbf{A}}$ (the diagonal entries of $\mathbf{\Sigma}$). Then $\sigma_i = \cos(\theta_i)$ where the θ_i are the principal angles between the subspaces spanned by $\widetilde{\mathbf{U}}_{\mathbf{A}}$ and $\mathbf{U}_{\mathbf{P}}$. Furthermore, by the Davis–Kahan $\sin(\Theta)$ theorem (see, e.g., Theorem 3.6 in [38]),

$$\|\widetilde{\mathbf{U}}_{\mathbf{A}}\widetilde{\mathbf{U}}_{\mathbf{A}}^\top - \widetilde{\mathbf{U}}_{\mathbf{P}}\widetilde{\mathbf{U}}_{\mathbf{P}}^\top\| = \max_i |\sin(\theta_i)| \leq \frac{\|\mathcal{L}(\mathbf{A}) - \mathcal{L}(\mathbf{P})\|}{\lambda_d(\mathcal{L}(\mathbf{P}))} = O_{\mathbb{P}}((n\rho_n)^{-1/2}).$$

Here, $\lambda_d(\mathcal{L}(\mathbf{P}))$ denotes the d largest eigenvalue of $\mathcal{L}(\mathbf{P})$. We thus have

$$\|\widetilde{\mathbf{U}}_{\mathbf{P}}^\top \widetilde{\mathbf{U}}_{\mathbf{A}} - \mathbf{W}_1 \mathbf{W}_2^\top\|_F = \|\mathbf{\Sigma} - \mathbf{I}\|_F = \left(\sum_{i=1}^{d}(1 - \sigma_i)^2\right)^{1/2}$$

$$\leq \sum_{i=1}^{d}(1 - \sigma_i^2) = \sum_{i=1}^{d}\sin^2(\theta_i).$$

Therefore, $\|\widetilde{\mathbf{U}}_{\mathbf{P}}^\top \widetilde{\mathbf{U}}_{\mathbf{A}} - \mathbf{W}_1 \mathbf{W}_2^\top\|_F = O_{\mathbb{P}}((n\rho_n)^{-1})$ as desired. $\square$

From now on, we shall denote by $\mathbf{W}^*$ the orthogonal matrix $\mathbf{W}_1 \mathbf{W}_2^\top$ as defined in the above proposition. Next, we state the following lemma.

LEMMA B.3.   *Let* $(\mathbf{A}, \mathbf{X}) \sim$ RDPG(F) *with sparsity factor* ρ_n. *Then*

(B.8) $n\rho_n \|\widetilde{\mathbf{U}}_{\mathbf{P}}^\top \widetilde{\mathbf{U}}_{\mathbf{A}} \widetilde{\mathbf{S}}_{\mathbf{A}} - \widetilde{\mathbf{S}}_{\mathbf{P}} \widetilde{\mathbf{U}}_{\mathbf{P}}^\top \widetilde{\mathbf{U}}_{\mathbf{A}}\| = O_{\mathbb{P}}(1),$

(B.9) $n\rho_n \|\widetilde{\mathbf{U}}_{\mathbf{P}}^\top \widetilde{\mathbf{U}}_{\mathbf{A}} \widetilde{\mathbf{S}}_{\mathbf{A}}^{1/2} - \widetilde{\mathbf{S}}_{\mathbf{P}}^{1/2} \widetilde{\mathbf{U}}_{\mathbf{P}}^\top \widetilde{\mathbf{U}}_{\mathbf{A}}\| = O_{\mathbb{P}}(1),$

(B.10) $n\rho_n \|\widetilde{\mathbf{U}}_{\mathbf{P}}^\top \widetilde{\mathbf{U}}_{\mathbf{A}} \widetilde{\mathbf{S}}_{\mathbf{A}}^{-1/2} - \widetilde{\mathbf{S}}_{\mathbf{P}}^{-1/2} \widetilde{\mathbf{U}}_{\mathbf{P}}^\top \widetilde{\mathbf{U}}_{\mathbf{A}}\| = O_{\mathbb{P}}(1).$

In proving Lemma B.3, we need the following technical result. Lemma B.3 and Lemma B.4 are the key technical lemmas of this paper. Roughly speaking, Lemma B.3 along with Proposition B.2 allows us to interchange the order of the orthogonal transformation $\mathbf{W}^*$ with the diagonal scaling matrices $\mathbf{S_A}$ or $\mathbf{S_A}$; Lemma B.4 simplifies various expressions involving $\mathbf{A}, \mathbf{D}, \mathcal{L}(\mathbf{A})$ and $\widetilde{\mathbf{U}}_\mathbf{A}$.

LEMMA B.4. *Let* $(\mathbf{A}, \mathbf{X}) \sim \mathrm{RDPG}(F)$ *with sparsity factor* ρ_n. *Then the following holds simultaneously*:

$$\text{(B.11)} \qquad \mathbf{D}^{-1/2} - \mathbf{T}^{-1/2} = \frac{1}{2}\mathbf{T}^{-3/2}(\mathbf{T} - \mathbf{D}) + O_{\mathbb{P}}((n\rho_n)^{-3/2}),$$

$$\mathcal{L}(\mathbf{A}) = \mathbf{T}^{-1/2}(\mathbf{A} - \mathbf{P})\mathbf{T}^{-1/2}$$
$$\text{(B.12)}$$
$$+ \mathbf{D}^{-1/2}\mathbf{P}\mathbf{D}^{-1/2} + O_{\mathbb{P}}((n\rho_n)^{-1}),$$

$$\mathbf{D}^{-1/2}\mathbf{P}\mathbf{D}^{-1/2} - \mathcal{L}(\mathbf{P}) = \frac{1}{2}\mathbf{T}^{-3/2}(\mathbf{T} - \mathbf{D})\mathbf{P}\mathbf{D}^{-1/2}$$
$$\text{(B.13)}$$
$$+ \frac{1}{2}\mathbf{T}^{-1/2}\mathbf{P}\mathbf{T}^{-3/2}(\mathbf{T} - \mathbf{D}) + O_{\mathbb{P}}((n\rho_n)^{-1}),$$

$$\text{(B.14)} \qquad \widetilde{\mathbf{U}}_\mathbf{A} - \widetilde{\mathbf{U}}_\mathbf{P}\widetilde{\mathbf{U}}_\mathbf{P}^\top\widetilde{\mathbf{U}}_\mathbf{A} = O_{\mathbb{P}}((n\rho_n)^{-1/2}),$$

$$\text{(B.15)} \quad \mathbf{T}^{-1/2}\mathbf{P}\mathbf{T}^{-3/2}(\mathbf{T} - \mathbf{D})\widetilde{\mathbf{U}}_\mathbf{P} = O_{\mathbb{P}}((n\rho_n)^{-1}),$$

$$\text{(B.16)} \quad \widetilde{\mathbf{U}}_\mathbf{P}^\top\mathbf{T}^{-3/2}(\mathbf{T} - \mathbf{D})\mathbf{P}\mathbf{D}^{-1/2} = O_{\mathbb{P}}((n\rho_n)^{-1}),$$

$$\text{(B.17)} \qquad \widetilde{\mathbf{U}}_\mathbf{P}^\top(\mathcal{L}(\mathbf{A}) - \mathcal{L}(\mathbf{P}))\widetilde{\mathbf{U}}_\mathbf{P} = O_{\mathbb{P}}((n\rho_n)^{-1}).$$

We continue with the proof of equation (3.1). Let $\boldsymbol{\Pi} = \widetilde{\mathbf{U}}_\mathbf{P}\widetilde{\mathbf{U}}_\mathbf{P}^\top$ and $\boldsymbol{\Pi}^\perp = \mathbf{I} - \boldsymbol{\Pi}$. Proposition B.2 and Lemma B.3 then yield

$$\widetilde{\mathbf{U}}_\mathbf{A}\widetilde{\mathbf{S}}_\mathbf{A}^{1/2} - \widetilde{\mathbf{U}}_\mathbf{P}\widetilde{\mathbf{S}}_\mathbf{P}^{1/2}\mathbf{W}^* = \widetilde{\mathbf{U}}_\mathbf{A}\widetilde{\mathbf{S}}_\mathbf{A}^{1/2} - \widetilde{\mathbf{U}}_\mathbf{P}\widetilde{\mathbf{S}}_\mathbf{P}^{1/2}\widetilde{\mathbf{U}}_\mathbf{P}^\top\widetilde{\mathbf{U}}_\mathbf{A} + O_{\mathbb{P}}((n\rho_n)^{-1})$$
$$= \widetilde{\mathbf{U}}_\mathbf{A}\widetilde{\mathbf{S}}_\mathbf{A}^{1/2} - \widetilde{\mathbf{U}}_\mathbf{P}\widetilde{\mathbf{U}}_\mathbf{P}^\top\widetilde{\mathbf{U}}_\mathbf{A}\widetilde{\mathbf{S}}_\mathbf{A}^{1/2} + O_{\mathbb{P}}((n\rho_n)^{-1})$$
$$= \boldsymbol{\Pi}^\perp\mathcal{L}(\mathbf{A})\widetilde{\mathbf{U}}_\mathbf{A}\widetilde{\mathbf{S}}_\mathbf{A}^{-1/2} + O_{\mathbb{P}}((n\rho_n)^{-1}).$$

Since $\mathcal{L}(\mathbf{P}) = \widetilde{\mathbf{U}}_\mathbf{P}\widetilde{\mathbf{S}}_\mathbf{P}\widetilde{\mathbf{U}}_\mathbf{P}^\top$, $\boldsymbol{\Pi}^\perp\mathcal{L}(\mathbf{P}) = \mathbf{0}$, and hence

$$\text{(B.18)} \quad \widetilde{\mathbf{U}}_\mathbf{A}\widetilde{\mathbf{S}}_\mathbf{A}^{1/2} - \widetilde{\mathbf{U}}_\mathbf{P}\widetilde{\mathbf{S}}_\mathbf{P}^{1/2}\mathbf{W}^* = \boldsymbol{\Pi}^\perp(\mathcal{L}(\mathbf{A}) - \mathcal{L}(\mathbf{P}))\widetilde{\mathbf{U}}_\mathbf{A}\widetilde{\mathbf{S}}_\mathbf{A}^{-1/2} + O_{\mathbb{P}}((n\rho_n)^{-1}).$$

In addition,

$$(\mathcal{L}(\mathbf{A}) - \mathcal{L}(\mathbf{P}))\widetilde{\mathbf{U}}_\mathbf{A}\widetilde{\mathbf{S}}_\mathbf{A}^{-1/2} = (\mathcal{L}(\mathbf{A}) - \mathcal{L}(\mathbf{P}))\boldsymbol{\Pi}^\perp\widetilde{\mathbf{U}}_\mathbf{A}\widetilde{\mathbf{S}}_\mathbf{A}^{-1/2}$$
$$+ (\mathcal{L}(\mathbf{A}) - \mathcal{L}(\mathbf{P}))\boldsymbol{\Pi}\widetilde{\mathbf{U}}_\mathbf{A}\widetilde{\mathbf{S}}_\mathbf{A}^{-1/2}$$
$$= O_{\mathbb{P}}((n\rho_n)^{-1}) + (\mathcal{L}(\mathbf{A}) - \mathcal{L}(\mathbf{P}))\boldsymbol{\Pi}\widetilde{\mathbf{U}}_\mathbf{A}\widetilde{\mathbf{S}}_\mathbf{A}^{-1/2},$$

where we bound $(\mathcal{L}(\mathbf{A}) - \mathcal{L}(\mathbf{P}))\mathbf{\Pi}^{\perp}\widetilde{\mathbf{U}}_{\mathbf{A}}\widetilde{\mathbf{S}}_{\mathbf{A}}^{-1/2}$ using Lemma B.1, equation (B.14), and the sub-multiplicativity of the spectral norm. Equation (B.18) then implies

$$
\begin{aligned}
(B.19) \quad \widetilde{\mathbf{U}}_{\mathbf{A}}\widetilde{\mathbf{S}}_{\mathbf{A}}^{1/2} - \widetilde{\mathbf{U}}_{\mathbf{P}}\widetilde{\mathbf{S}}_{\mathbf{P}}^{1/2}\mathbf{W}^* &= \mathbf{\Pi}^{\perp}(\mathcal{L}(\mathbf{A}) - \mathcal{L}(\mathbf{P}))\widetilde{\mathbf{U}}_{\mathbf{A}}\widetilde{\mathbf{S}}_{\mathbf{A}}^{-1/2} + O_{\mathbb{P}}((n\rho_n)^{-1}) \\
&= \mathbf{\Pi}^{\perp}(\mathcal{L}(\mathbf{A}) - \mathcal{L}(\mathbf{P}))\mathbf{\Pi}\widetilde{\mathbf{U}}_{\mathbf{A}}\widetilde{\mathbf{S}}_{\mathbf{A}}^{-1/2} + O_{\mathbb{P}}((n\rho_n)^{-1}).
\end{aligned}
$$

By equation (B.17) and sub-multiplicativity of the Frobenius norm, we also have

$$
\begin{aligned}
\mathbf{\Pi}^{\perp}(\mathcal{L}(\mathbf{A}) - \mathcal{L}(\mathbf{P}))\mathbf{\Pi}\widetilde{\mathbf{U}}_{\mathbf{A}}\widetilde{\mathbf{S}}_{\mathbf{A}}^{-1/2} &= (\mathcal{L}(\mathbf{A}) - \mathcal{L}(\mathbf{P}))\mathbf{\Pi}\widetilde{\mathbf{U}}_{\mathbf{A}}\widetilde{\mathbf{S}}_{\mathbf{A}}^{-1/2} \\
&\quad - \mathbf{\Pi}(\mathcal{L}(\mathbf{A}) - \mathcal{L}(\mathbf{P}))\mathbf{\Pi}\widetilde{\mathbf{U}}_{\mathbf{A}}\widetilde{\mathbf{S}}_{\mathbf{A}}^{-1/2} \\
&= (\mathcal{L}(\mathbf{A}) - \mathcal{L}(\mathbf{P}))\mathbf{\Pi}\widetilde{\mathbf{U}}_{\mathbf{A}}\widetilde{\mathbf{S}}_{\mathbf{A}}^{-1/2} + O_{\mathbb{P}}((n\rho_n)^{-1}).
\end{aligned}
$$

Equation (B.19) then becomes

$$
\begin{aligned}
(B.20) \quad \widetilde{\mathbf{U}}_{\mathbf{A}}\widetilde{\mathbf{S}}_{\mathbf{A}}^{1/2} - \widetilde{\mathbf{U}}_{\mathbf{P}}\widetilde{\mathbf{S}}_{\mathbf{P}}^{1/2}\mathbf{W}^* &= \mathbf{\Pi}^{\perp}(\mathcal{L}(\mathbf{A}) - \mathcal{L}(\mathbf{P}))\mathbf{\Pi}\widetilde{\mathbf{U}}_{\mathbf{A}}\widetilde{\mathbf{S}}_{\mathbf{A}}^{-1/2} + O_{\mathbb{P}}((n\rho_n)^{-1}) \\
&= (\mathcal{L}(\mathbf{A}) - \mathcal{L}(\mathbf{P}))\mathbf{\Pi}\widetilde{\mathbf{U}}_{\mathbf{A}}\widetilde{\mathbf{S}}_{\mathbf{A}}^{-1/2} + O_{\mathbb{P}}((n\rho_n)^{-1}) \\
&= (\mathcal{L}(\mathbf{A}) - \mathcal{L}(\mathbf{P}))\widetilde{\mathbf{U}}_{\mathbf{P}}\widetilde{\mathbf{U}}_{\mathbf{P}}^{\top}\widetilde{\mathbf{U}}_{\mathbf{A}}\widetilde{\mathbf{S}}_{\mathbf{A}}^{-1/2} + O_{\mathbb{P}}((n\rho_n)^{-1}) \\
&= (\mathcal{L}(\mathbf{A}) - \mathcal{L}(\mathbf{P}))\widetilde{\mathbf{U}}_{\mathbf{P}}\mathbf{W}^*\widetilde{\mathbf{S}}_{\mathbf{A}}^{-1/2} + O_{\mathbb{P}}((n\rho_n)^{-1}) \\
&= (\mathcal{L}(\mathbf{A}) - \mathcal{L}(\mathbf{P}))\widetilde{\mathbf{U}}_{\mathbf{P}}\widetilde{\mathbf{S}}_{\mathbf{P}}^{-1/2}\mathbf{W}^* + O_{\mathbb{P}}((n\rho_n)^{-1}),
\end{aligned}
$$

where we once again use Proposition B.2 and Lemma B.3. Recall from equation (B.12) the decomposition

$$
\mathcal{L}(\mathbf{A}) = \mathbf{T}^{-1/2}(\mathbf{A} - \mathbf{P})\mathbf{T}^{-1/2} + \mathbf{D}^{-1/2}\mathbf{P}\mathbf{D}^{-1/2} + O_{\mathbb{P}}((n\rho_n)^{-1}).
$$

Therefore, from equation (B.20), we have

$$
\begin{aligned}
(B.21) \quad \widetilde{\mathbf{U}}_{\mathbf{A}}\widetilde{\mathbf{S}}_{\mathbf{A}}^{1/2} &- \widetilde{\mathbf{U}}_{\mathbf{P}}\widetilde{\mathbf{S}}_{\mathbf{P}}^{1/2}\widetilde{\mathbf{W}}^* \\
&= O_{\mathbb{P}}((n\rho_n)^{-1}) + \mathbf{T}^{-1/2}(\mathbf{A} - \mathbf{P})\mathbf{T}^{-1/2}\widetilde{\mathbf{U}}_{\mathbf{P}}\widetilde{\mathbf{S}}_{\mathbf{P}}^{-1/2}\mathbf{W}^* \\
&\quad + (\mathbf{D}^{-1/2}\mathbf{P}\mathbf{D}^{-1/2} - \mathbf{T}^{-1/2}\mathbf{P}\mathbf{T}^{-1/2})\widetilde{\mathbf{U}}_{\mathbf{P}}\widetilde{\mathbf{S}}_{\mathbf{P}}^{-1/2}\mathbf{W}^*.
\end{aligned}
$$

We next recall from equation (B.13) the decomposition

$$
\begin{aligned}
\mathbf{D}^{-1/2}\mathbf{P}\mathbf{D}^{-1/2} - \mathbf{T}^{-1/2}\mathbf{P}\mathbf{T}^{-1/2} &= \frac{1}{2}\mathbf{T}^{-3/2}(\mathbf{T} - \mathbf{D})\mathbf{P}\mathbf{D}^{-1/2} \\
&\quad + \frac{1}{2}\mathbf{T}^{-1/2}\mathbf{P}\mathbf{T}^{-3/2}(\mathbf{T} - \mathbf{D}) + O_{\mathbb{P}}((n\rho_n)^{-3/2}).
\end{aligned}
$$

In addition, we recall from equation (B.15) that

$$
\mathbf{T}^{-1/2}\mathbf{P}\mathbf{T}^{-3/2}(\mathbf{T} - \mathbf{D})\widetilde{\mathbf{U}}_{\mathbf{P}} = O_{\mathbb{P}}((n\rho_n)^{-1}).
$$

Equation (B.21) therefore reduces to

$$(B.22) \quad \widetilde{\mathbf{U}}_{\mathbf{A}}\widetilde{\mathbf{S}}_{\mathbf{A}}^{1/2} - \widetilde{\mathbf{U}}_{\mathbf{P}}\widetilde{\mathbf{S}}_{\mathbf{P}}^{1/2}\mathbf{W}^* = O_{\mathbb{P}}\big((n\rho_n)^{-1}\big) + \mathbf{T}^{-1/2}(\mathbf{A} - \mathbf{P})\mathbf{T}^{-1/2}\widetilde{\mathbf{U}}_{\mathbf{P}}\widetilde{\mathbf{S}}_{\mathbf{P}}^{-1/2}\mathbf{W}^* \\ + \frac{1}{2}\mathbf{T}^{-3/2}(\mathbf{T} - \mathbf{D})\mathbf{P}\mathbf{D}^{-1/2}\widetilde{\mathbf{U}}_{\mathbf{P}}\widetilde{\mathbf{S}}_{\mathbf{P}}^{-1/2}\mathbf{W}^*.$$

Equation (B.11) now implies

$$\mathbf{T}^{-3/2}(\mathbf{T} - \mathbf{D})\mathbf{P}\mathbf{D}^{-1/2} = \mathbf{T}^{-3/2}(\mathbf{T} - \mathbf{D})\mathbf{P}(\mathbf{D}^{-1/2} - \mathbf{T}^{-1/2} + \mathbf{T}^{-1/2}) \\ = \mathbf{T}^{-3/2}(\mathbf{T} - \mathbf{D})\mathbf{P}\mathbf{T}^{-1/2} + O_{\mathbb{P}}\big((n\rho_n)^{-1}\big),$$

and thus equation (B.22) further simplifies to

$$(B.23) \quad \widetilde{\mathbf{U}}_{\mathbf{A}}\widetilde{\mathbf{S}}_{\mathbf{A}}^{1/2} - \widetilde{\mathbf{U}}_{\mathbf{P}}\widetilde{\mathbf{S}}_{\mathbf{P}}^{1/2}\mathbf{W}^* = O_{\mathbb{P}}\big((n\rho_n)^{-1}\big) + \mathbf{T}^{-1/2}(\mathbf{A} - \mathbf{P})\mathbf{T}^{-1/2}\widetilde{\mathbf{U}}_{\mathbf{P}}\widetilde{\mathbf{S}}_{\mathbf{P}}^{-1/2}\mathbf{W}^* \\ + \frac{1}{2}\mathbf{T}^{-3/2}(\mathbf{T} - \mathbf{D})\mathbf{P}\mathbf{T}^{-1/2}\widetilde{\mathbf{U}}_{\mathbf{P}}\widetilde{\mathbf{S}}_{\mathbf{P}}^{-1/2}\mathbf{W}^*.$$

Since $\mathbf{T}$ and $\mathbf{D}$ are diagonal matrices, we have

$$\mathbf{T}^{-3/2}(\mathbf{T} - \mathbf{D})\mathbf{P}\mathbf{T}^{-1/2}\widetilde{\mathbf{U}}_{\mathbf{P}}\widetilde{\mathbf{S}}_{\mathbf{P}}^{-1/2}\mathbf{W}^* = \mathbf{T}^{-1}(\mathbf{T} - \mathbf{D})\mathbf{T}^{-1/2}\mathbf{P}\mathbf{T}^{-1/2}\widetilde{\mathbf{U}}_{\mathbf{P}}\widetilde{\mathbf{S}}_{\mathbf{P}}^{-1/2}\mathbf{W}^* \\ = \mathbf{T}^{-1}(\mathbf{T} - \mathbf{D})\mathcal{L}(\mathbf{P})\widetilde{\mathbf{U}}_{\mathbf{P}}\widetilde{\mathbf{S}}_{\mathbf{P}}^{-1/2}\mathbf{W}^* \\ = \mathbf{T}^{-1}(\mathbf{T} - \mathbf{D})\widetilde{\mathbf{U}}_{\mathbf{P}}\widetilde{\mathbf{S}}_{\mathbf{P}}\widetilde{\mathbf{S}}_{\mathbf{P}}^{-1/2}\mathbf{W}^* \\ = \mathbf{T}^{-1}(\mathbf{T} - \mathbf{D})\widetilde{\mathbf{U}}_{\mathbf{P}}\widetilde{\mathbf{S}}_{\mathbf{P}}^{1/2}\mathbf{W}^*.$$

We therefore arrive at

$$(B.24) \quad \widetilde{\mathbf{U}}_{\mathbf{A}}\widetilde{\mathbf{S}}_{\mathbf{A}}^{1/2} - \widetilde{\mathbf{U}}_{\mathbf{P}}\widetilde{\mathbf{S}}_{\mathbf{P}}^{1/2}\mathbf{W}^* = O_{\mathbb{P}}\big((n\rho_n)^{-1}\big) + \mathbf{T}^{-1/2}(\mathbf{A} - \mathbf{P})\mathbf{T}^{-1/2}\widetilde{\mathbf{U}}_{\mathbf{P}}\widetilde{\mathbf{S}}_{\mathbf{P}}^{-1/2}\mathbf{W}^* \\ + \frac{1}{2}\mathbf{T}^{-1}(\mathbf{T} - \mathbf{D})\widetilde{\mathbf{U}}_{\mathbf{P}}\widetilde{\mathbf{S}}_{\mathbf{P}}^{1/2}\mathbf{W}^*.$$

To conclude the proof of equation (3.1), we recall that $\widetilde{\mathbf{X}}\widetilde{\mathbf{X}}^{\top} = \mathcal{L}(\mathbf{P}) = \widetilde{\mathbf{U}}_{\mathbf{P}}\widetilde{\mathbf{S}}_{\mathbf{P}}\widetilde{\mathbf{U}}_{\mathbf{P}}^{\top}$; hence $\widetilde{\mathbf{X}} = \widetilde{\mathbf{U}}_{\mathbf{P}}\widetilde{\mathbf{S}}_{\mathbf{P}}^{1/2}\widetilde{\mathbf{W}}$ for some orthogonal matrix $\widetilde{\mathbf{W}}$. Therefore,

$$\widetilde{\mathbf{U}}_{\mathbf{P}}\widetilde{\mathbf{S}}_{\mathbf{P}}^{1/2}\mathbf{W}^* = \widetilde{\mathbf{U}}_{\mathbf{P}}\widetilde{\mathbf{S}}_{\mathbf{P}}^{1/2}\widetilde{\mathbf{W}}\widetilde{\mathbf{W}}^{\top}\mathbf{W}^* = \widetilde{\mathbf{X}}\widetilde{\mathbf{W}}^{\top}\mathbf{W}^*,$$

$$\widetilde{\mathbf{U}}_{\mathbf{P}}\widetilde{\mathbf{S}}_{\mathbf{P}}^{-1/2}\mathbf{W}^* = \widetilde{\mathbf{U}}_{\mathbf{P}}\widetilde{\mathbf{S}}_{\mathbf{P}}^{1/2}\widetilde{\mathbf{W}}\widetilde{\mathbf{W}}^{\top}\widetilde{\mathbf{S}}_{\mathbf{P}}^{-1}\widetilde{\mathbf{W}}\widetilde{\mathbf{W}}^{\top}\mathbf{W}^* = \widetilde{\mathbf{X}}(\widetilde{\mathbf{X}}^{\top}\widetilde{\mathbf{X}})^{-1}\widetilde{\mathbf{W}}^{\top}\mathbf{W}^*.$$

Substituting the above equations into equation (B.24) yield

$$\widetilde{\mathbf{U}}_{\mathbf{A}}\widetilde{\mathbf{S}}_{\mathbf{A}}^{1/2} - \widetilde{\mathbf{X}}\widetilde{\mathbf{W}}^{\top}\mathbf{W}^* = O_{\mathbb{P}}\big((n\rho_n)^{-1}\big) + \mathbf{T}^{-1/2}(\mathbf{A} - \mathbf{P})\mathbf{T}^{-1/2}\widetilde{\mathbf{X}}(\widetilde{\mathbf{X}}^{\top}\widetilde{\mathbf{X}})^{-1}\widetilde{\mathbf{W}}^{\top}\mathbf{W}^* \\ + \frac{1}{2}\mathbf{T}^{-1}(\mathbf{T} - \mathbf{D})\widetilde{\mathbf{X}}\widetilde{\mathbf{W}}^{\top}\mathbf{W}^*.$$

Equivalently,

$$\widetilde{\mathbf{U}}_{\mathbf{A}}\widetilde{\mathbf{S}}_{\mathbf{A}}^{1/2}(\mathbf{W}^*)^\top \widetilde{\mathbf{W}} - \widetilde{\mathbf{X}} = O_{\mathbb{P}}((n\rho_n)^{-1}) + \mathbf{T}^{-1/2}(\mathbf{A} - \mathbf{P})\mathbf{T}^{-1/2}\widetilde{\mathbf{X}}(\widetilde{\mathbf{X}}^\top \widetilde{\mathbf{X}})^{-1}$$

$$+ \frac{1}{2}\mathbf{T}^{-1}(\mathbf{T} - \mathbf{D})\widetilde{\mathbf{X}}.$$

Equation (3.1) is thereby established.

B.3. Proof of Lemma B.3 and Lemma B.4. We first present the proof of Lemma B.4. We recall the notation $\mathbf{D} = \mathrm{diag}(\mathbf{A1})$ and $\mathbf{T} = \mathrm{diag}(\mathbf{P1})$. Denote by d_i and t_i the ith diagonal elements of $\mathbf{D}$ and $\mathbf{T}$. The ith diagonal element of $\mathbf{D}^{-1/2} - \mathbf{T}^{-1/2}$ can be written as

$$\frac{1}{\sqrt{d_i}} - \frac{1}{\sqrt{t_i}} = \frac{t_i - d_i}{(\sqrt{d_i} + \sqrt{t_i})\sqrt{d_i}\sqrt{t_i}}$$

$$= \frac{t_i - d_i}{2t_i^{3/2}} + (t_i - d_i)\left(\frac{1}{d_i\sqrt{t_i} + t_i\sqrt{d_i}} - \frac{1}{2t_i^{3/2}}\right)$$

$$= \frac{t_i - d_i}{2t_i^{3/2}} + (t_i - d_i)\frac{t_i(\sqrt{t_i} - \sqrt{d_i}) + (t_i - d_i)\sqrt{t_i}}{2t_i^{3/2}(d_i\sqrt{t_i} + t_i\sqrt{d_i})}.$$

We have, by Chernoff's bound, that $|t_i - d_i| = O_{\mathbb{P}}(\sqrt{n\rho_n})$ for any given index i, and hence $|\sqrt{t_i} - \sqrt{d_i}| = O_{\mathbb{P}}(1)$. Therefore,

$$(t_i - d_i)\frac{t_i(\sqrt{t_i} - \sqrt{d_i}) + (t_i - d_i)\sqrt{t_i}}{2t_i^{3/2}(d_i\sqrt{t_i} + t_i\sqrt{d_i})} = O_{\mathbb{P}}(\sqrt{n\rho_n})\frac{O_{\mathbb{P}}(n\rho_n)}{\Omega_{\mathbb{P}}(n^3\rho_n^3)} = O_{\mathbb{P}}((n\rho_n)^{-3/2}).$$

Upon taking an union bound over all indices $i = 1, 2, \ldots, n$, we have

(B.25) $$\mathbf{D}^{-1/2} - \mathbf{T}^{-1/2} = \frac{1}{2}\mathbf{T}^{-3/2}(\mathbf{T} - \mathbf{D}) + O_{\mathbb{P}}((n\rho_n)^{-3/2}\log n).$$

Equation (B.11) is thereby established. Equation (B.13) follows directly from equation (B.11) and the definition of $\mathcal{L}(\mathbf{P}) = \mathbf{T}^{-1/2}\mathbf{P}\mathbf{T}^{-1/2}$. We next show equation (B.12). Consider the following decomposition of $\mathcal{L}(\mathbf{A})$:

$$\mathcal{L}(\mathbf{A}) = \mathbf{D}^{-1/2}(\mathbf{A} - \mathbf{P})\mathbf{D}^{-1/2} + \mathbf{D}^{-1/2}\mathbf{P}\mathbf{D}^{-1/2}$$

$$= \mathbf{T}^{-1/2}(\mathbf{A} - \mathbf{P})\mathbf{T}^{-1/2} + \mathbf{T}^{-1/2}(\mathbf{A} - \mathbf{P})(\mathbf{D}^{-1/2} - \mathbf{T}^{-1/2})$$

$$+ (\mathbf{D}^{-1/2} - \mathbf{T}^{-1/2})(\mathbf{A} - \mathbf{P})\mathbf{D}^{-1/2} + \mathbf{D}^{-1/2}\mathbf{P}\mathbf{D}^{-1/2}.$$

By Lemma B.1, we have

(B.26) $$\|(\mathbf{A} - \mathbf{P})\mathbf{T}^{-1/2}\| \leq \|\mathbf{A} - \mathbf{P}\| \times \|\mathbf{T}^{-1/2}\| = O_{\mathbb{P}}(1).$$

Similarly, Lemma B.1 and Chernoff bound yield

(B.27) $$\|(\mathbf{A} - \mathbf{P})\mathbf{D}^{-1/2}\| \leq \|(\mathbf{A} - \mathbf{P})\| \times \|\mathbf{D}^{-1/2}\| = O_{\mathbb{P}}(1).$$

Combining equation (B.25) and equation (B.27), we have

$$\|(\mathbf{D}^{-1/2} - \mathbf{T}^{-1/2})(\mathbf{A} - \mathbf{P})\mathbf{D}^{-1/2}\|$$
$$\leq (\|\mathbf{T}^{-3/2}(\mathbf{D} - \mathbf{T})\|/2 + O_{\mathbb{P}}((n\rho_n)^{-3/2})) \times O_{\mathbb{P}}(1)$$
$$= O_{\mathbb{P}}((n\rho_n)^{-1}).$$

Similarly, equation (B.25) and equation (B.26) implies

$$\|\mathbf{T}^{-1/2}(\mathbf{A} - \mathbf{P})(\mathbf{D}^{-1/2} - \mathbf{T}^{-1/2})\| = O_{\mathbb{P}}((n\rho_n)^{-1}).$$

We thus have

(B.28) $$\mathcal{L}(\mathbf{A}) = \mathbf{T}^{-1/2}(\mathbf{A} - \mathbf{P})\mathbf{T}^{-1/2} + \mathbf{D}^{-1/2}\mathbf{P}\mathbf{D}^{-1/2} + O_{\mathbb{P}}((n\rho_n)^{-1}).$$

Equation (B.12) is thereby established.

We next derive equation (B.15) through equation (B.17). From equation (B.28), we have

(B.29)
$$\widetilde{\mathbf{U}}_{\mathbf{P}}^{\top}(\mathcal{L}(\mathbf{A}) - \mathcal{L}(\mathbf{P}))\widetilde{\mathbf{U}}_{\mathbf{P}} = \widetilde{\mathbf{U}}_{\mathbf{P}}^{\top}\mathbf{T}^{-1/2}(\mathbf{A} - \mathbf{P})\mathbf{T}^{-1/2}\widetilde{\mathbf{U}}_{\mathbf{P}}$$
$$+ \widetilde{\mathbf{U}}_{\mathbf{P}}^{\top}(\mathbf{D}^{-1/2}\mathbf{P}\mathbf{D}^{-1/2} - \mathbf{T}^{-1/2}\mathbf{P}\mathbf{T}^{-1/2})\widetilde{\mathbf{U}}_{\mathbf{P}}$$
$$+ O_{\mathbb{P}}((n\rho_n)^{-1}).$$

We first bound the spectral norm of $\widetilde{\mathbf{U}}_{\mathbf{P}}^{\top}\mathbf{T}^{-1/2}(\mathbf{A} - \mathbf{P})\mathbf{T}^{-1/2}\widetilde{\mathbf{U}}_{\mathbf{P}}$. Let $\widetilde{\boldsymbol{u}}_i$ be the ith column of $\widetilde{\mathbf{U}}_{\mathbf{P}}\mathbf{T}^{-1/2}$; the ijth entry of $\widetilde{\mathbf{U}}_{\mathbf{P}}^{\top}\mathbf{T}^{-1/2}(\mathbf{A} - \mathbf{P})\mathbf{T}^{-1/2}\widetilde{\mathbf{U}}_{\mathbf{P}}$ is then of the form

$$\widetilde{\boldsymbol{u}}_i^{\top}(\mathbf{A} - \mathbf{P})\widetilde{\boldsymbol{u}}_j = \sum_{k<l} 2\widetilde{\boldsymbol{u}}_{ik}(a_{kl} - p_{kl})\widetilde{\boldsymbol{u}}_{jl} + \sum_k \widetilde{\boldsymbol{u}}_{ik} p_{kk} \widetilde{\boldsymbol{u}}_{jk},$$

where $\widetilde{\boldsymbol{u}}_{ik}$ is the kth element of the vector $\widetilde{\boldsymbol{u}}_i$. We note that

$$\left|\sum_k \widetilde{\boldsymbol{u}}_{ik} p_{kk} \widetilde{\boldsymbol{u}}_{jk}\right| \leq \rho_n \|\widetilde{\boldsymbol{u}}_i\| \times \|\widetilde{\boldsymbol{u}}_j\| \leq \rho_n \delta_n^{-1} = O_{\mathbb{P}}(\rho_n(n\rho_n)^{-1}).$$

In addition, $\sum_{k<l} 2\widetilde{\boldsymbol{u}}_{ik}(a_{kl} - p_{kl})\widetilde{\boldsymbol{u}}_{jl}$ is, conditioned on $\mathbf{P}$, a sum of mean 0 random variables. Hoeffding's inequality then implies

$$\mathbb{P}\left[\left|\sum_{k<l} 2\widetilde{\boldsymbol{u}}_{ik}(a_{kl} - p_{kl})\widetilde{\boldsymbol{u}}_{lj}\right| \geq t\right] \leq \exp\left(-\frac{t^2}{2(\sum_{k<l} \widetilde{\boldsymbol{u}}_{ik}^2 \widetilde{\boldsymbol{u}}_{jl}^2)}\right)$$
$$\leq \exp\left(-\frac{t^2}{2\sum_k \sum_l \widetilde{\boldsymbol{u}}_{ik}^2 \widetilde{\boldsymbol{u}}_{jl}^2}\right)$$
$$\leq \exp\left(-\frac{t^2}{2\delta^{-2}}\right).$$

Hence $\widetilde{\boldsymbol{u}}_i^\top (\mathbf{A} - \mathbf{P})\widetilde{\boldsymbol{u}}_j = O_{\mathbb{P}}(\delta^{-1})$. As $\widetilde{\mathbf{U}}_{\mathbf{P}}^\top \mathbf{T}^{-1/2}(\mathbf{A} - \mathbf{P})\mathbf{T}^{-1/2}\widetilde{\mathbf{U}}_{\mathbf{P}}$ is a $d \times d$ matrix, a union bound then implies

$$(B.30) \qquad \widetilde{\mathbf{U}}_{\mathbf{P}}^\top \mathbf{T}^{-1/2}(\mathbf{A} - \mathbf{P})\mathbf{T}^{-1/2}\widetilde{\mathbf{U}}_{\mathbf{P}} = O_{\mathbb{P}}(\delta^{-1}) = O_{\mathbb{P}}((n\rho_n)^{-1}).$$

We next bound the spectral norm of $\widetilde{\mathbf{U}}_{\mathbf{P}}^\top (\mathbf{D}^{-1/2}\mathbf{P}\mathbf{D}^{-1/2} - \mathbf{T}^{-1/2}\mathbf{P}\mathbf{T}^{-1/2})\widetilde{\mathbf{U}}_{\mathbf{P}}$. Let ζ_{ij} denote the ijth entry of $\widetilde{\mathbf{U}}_{\mathbf{P}}^\top (\mathbf{D}^{-1/2}\mathbf{P}\mathbf{D}^{-1/2} - \mathbf{T}^{-1/2}\mathbf{P}\mathbf{T}^{-1/2})\widetilde{\mathbf{U}}_{\mathbf{P}}$. From equation (B.25), we have

$$\zeta_{ij} = \widetilde{\boldsymbol{u}}_i^\top \big((\mathbf{D}^{-1/2} - \mathbf{T}^{-1/2})\mathbf{P}\mathbf{D}^{-1/2} + \mathbf{T}^{-1/2}\mathbf{P}(\mathbf{D}^{-1/2} - \mathbf{T}^{-1/2})\big)\widetilde{\boldsymbol{u}}_j$$

$$= \frac{1}{2}\widetilde{\boldsymbol{u}}_i(\mathbf{T}^{-3/2}(\mathbf{T} - \mathbf{D})\mathbf{P}\mathbf{D}^{-1/2} + \mathbf{T}^{-1/2}\mathbf{P}\mathbf{T}^{-3/2}(\mathbf{T} - \mathbf{D}))\widetilde{\boldsymbol{u}}_j + O_{\mathbb{P}}((n\rho_n)^{-3/2}).$$

Now let $\zeta_{ij}^{(1)}$ and $\zeta_{ij}^{(2)}$ denote the quantities

$$\zeta_{ij}^{(1)} = \frac{1}{2}\widetilde{\boldsymbol{u}}_i^\top \mathbf{T}^{-3/2}(\mathbf{T} - \mathbf{D})\mathbf{P}\mathbf{D}^{-1/2}\widetilde{\boldsymbol{u}}_j,$$

$$\zeta_{ij}^{(2)} = \frac{1}{2}\widetilde{\boldsymbol{u}}_i^\top \mathbf{T}^{-1/2}\mathbf{P}\mathbf{T}^{-3/2}(\mathbf{T} - \mathbf{D})\widetilde{\boldsymbol{u}}_j.$$

Because $\mathbf{P} = \rho_n \mathbf{X}\mathbf{X}^\top$, we have

$$\zeta_{ij}^{(1)} \le \frac{1}{2}\|\rho_n^{1/2}\widetilde{\boldsymbol{u}}_i^\top \mathbf{T}^{-3/2}(\mathbf{T} - \mathbf{D})\mathbf{X}\| \times \|\rho_n^{1/2}\mathbf{X}^\top \mathbf{D}^{-1/2}\widetilde{\boldsymbol{u}}_j\|,$$

$$\zeta_{ij}^{(2)} \le \frac{1}{2}\|\rho_n^{1/2}\widetilde{\boldsymbol{u}}_i^\top \mathbf{T}^{-1/2}\mathbf{X}\| \times \|\rho_n^{1/2}\mathbf{X}^\top \mathbf{T}^{-3/2}(\mathbf{T} - \mathbf{D})\widetilde{\boldsymbol{u}}_j\|.$$

For $k \in \{1, 2, \ldots, d\}$, let $\boldsymbol{x}_k$ denote the kth column of $\mathbf{X}$. Furthermore, for $l \in \{1, 2, \ldots, n\}$, let x_{kl} denote the lth entry of $\boldsymbol{x}_k$—equivalently the kth entry of X_l (recall that $\mathbf{X} = [X_1 \mid \cdots \mid X_n]^\top$). Also let $\widetilde{u}_{kl}$ denotes the lth entry of $\widetilde{\boldsymbol{u}}_k$. Then $\rho_n^{1/2}\widetilde{\boldsymbol{u}}_i^\top \mathbf{T}^{-3/2}(\mathbf{T} - \mathbf{D})\mathbf{X}$ is a vector in $\mathbb{R}^d$ whose kth element is of the form

$$\rho_n^{1/2}\widetilde{\boldsymbol{u}}_i^\top \mathbf{T}^{-3/2}(\mathbf{D} - \mathbf{T})\boldsymbol{x}_k = \rho_n^{1/2} \sum_l \frac{\widetilde{u}_{il}}{t_l^{3/2}}(d_l - t_l)x_{kl}$$

$$= \rho_n^{1/2} \sum_l \sum_m \frac{\widetilde{u}_{il}}{t_l^{3/2}}(a_{lm} - p_{lm})x_{kl}$$

$$= 2\rho_n^{1/2} \sum_{l<m} \frac{\widetilde{u}_{il}}{t_l^{3/2}}(a_{lm} - p_{lm})x_{kl} + \rho_n^{1/2} \sum_l \frac{\widetilde{u}_{il}}{t_l^{3/2}}p_{ll}x_{kl}.$$

Conditioned on $\mathbf{P}$, the above is a sum of mean 0 random variables and a term of order $O((n\rho_n)^{-3/2})$. Hoeffding's inequality then yields

$$\mathbb{P}\left[\left|2\rho_n^{1/2}\sum_{l<m}\frac{\widetilde{u}_{il}}{t_l^{3/2}}(a_{lm}-p_{lm})x_{kl}\right|\geq s\right] \leq 2\exp\left(-\frac{s^2}{2\rho_n\sum_{l<m}t_l^{-3}\widetilde{u}_{il}^2 x_{kl}^2}\right)$$

$$\leq 2\exp\left(-\frac{s^2}{2n\rho_n\sum_l t_l^{-3}\widetilde{u}_{il}^2 x_{kl}^2}\right)$$

$$\leq 2\exp\left(-\frac{s^2}{2n\rho_n\delta_n^{-3}\sum_l \widetilde{u}_{il}^2}\right)$$

$$\leq 2\exp\left(-\frac{s^2\delta_n^3}{2n\rho_n\|\widetilde{\boldsymbol{u}}_i\|^2}\right)$$

$$\leq 2\exp\left(-\frac{s^2\delta^3}{2n\rho_n}\right),$$

where we used the fact that $x_{kl}^2 \leq 1$ for all indices k and l [as $(\mathbf{A},\mathbf{X})\sim\mathrm{RDPG}(F)$]. We thus have

(B.31) $$\rho_n^{1/2}\widetilde{\boldsymbol{u}}_i^\top\mathbf{T}^{-3/2}(\mathbf{D}-\mathbf{T})\boldsymbol{x}_k = O_{\mathbb{P}}((n\rho_n)^{-1})$$

A union bound over the d entries of $\rho_n^{1/2}\widetilde{\boldsymbol{u}}_i^\top\mathbf{T}^{-3/2}(\mathbf{T}-\mathbf{D})\mathbf{X}$ along with the bound $\|\rho_n^{1/2}\mathbf{T}^{-1/2}\mathbf{X}\| = O_{\mathbb{P}}(1)$ yield that $\zeta_{ij}^{(1)} = O_{\mathbb{P}}((n\rho_n)^{-1})$. An identical argument also yield that $\zeta_{ij}^{(2)} = O_{\mathbb{P}}((n\rho_n)^{-1})$. Therefore, $\zeta_{ij} = O((n\rho_n)^{-1})$. A union bound over the indices $i, j \in \{1, 2, \ldots, d\}$ also implies

(B.32) $$\mathbf{T}^{-1/2}\mathbf{P}\mathbf{T}^{-3/2}(\mathbf{T}-\mathbf{D})\widetilde{\mathbf{U}}_{\mathbf{P}} = O_{\mathbb{P}}((n\rho_n)^{-1}),$$

(B.33) $$\widetilde{\mathbf{U}}_{\mathbf{P}}^\top\mathbf{T}^{-3/2}(\mathbf{T}-\mathbf{D})\mathbf{P}\mathbf{D}^{-1/2} = O_{\mathbb{P}}((n\rho_n)^{-1}),$$

(B.34) $$\widetilde{\mathbf{U}}_{\mathbf{P}}^\top(\mathbf{D}^{-1/2}\mathbf{P}\mathbf{D}^{-1/2}-\mathbf{T}^{-1/2}\mathbf{P}\mathbf{T}^{-1/2})\widetilde{\mathbf{U}}_{\mathbf{P}} = O_{\mathbb{P}}((n\rho_n)^{-1}).$$

We thus derive equation (B.15) and equation (B.16). Equation (B.17) follows from equation (B.29), equation (B.30) and equation (B.34). Lemma B.4 is thereby established.

Lemma B.3 now follows directly from Lemma B.4. Indeed, by equation (B.14) and equation (B.17), we have

$$\widetilde{\mathbf{U}}_{\mathbf{P}}^\top\widetilde{\mathbf{U}}_{\mathbf{A}}\widetilde{\mathbf{S}}_{\mathbf{A}} - \widetilde{\mathbf{S}}_{\mathbf{P}}\widetilde{\mathbf{U}}_{\mathbf{P}}^\top\widetilde{\mathbf{U}}_{\mathbf{A}}$$

$$= \widetilde{\mathbf{U}}_{\mathbf{P}}^\top\mathcal{L}(\mathbf{A})\widetilde{\mathbf{U}}_{\mathbf{A}} - \widetilde{\mathbf{U}}_{\mathbf{P}}^\top\mathcal{L}(\mathbf{P})\widetilde{\mathbf{U}}_{\mathbf{A}}$$

(B.35) $$= \widetilde{\mathbf{U}}_{\mathbf{P}}^\top(\mathcal{L}(\mathbf{A})-\mathcal{L}(\mathbf{P}))(\widetilde{\mathbf{U}}_{\mathbf{A}} - \widetilde{\mathbf{U}}_{\mathbf{P}}\widetilde{\mathbf{U}}_{\mathbf{P}}^\top\widetilde{\mathbf{U}}_{\mathbf{A}} + \widetilde{\mathbf{U}}_{\mathbf{P}}\widetilde{\mathbf{U}}_{\mathbf{P}}^\top\widetilde{\mathbf{U}}_{\mathbf{A}})$$

$$= O_{\mathbb{P}}((n\rho_n)^{-1}) + \widetilde{\mathbf{U}}_{\mathbf{P}}(\mathcal{L}(\mathbf{A})-\mathcal{L}(\mathbf{P}))\widetilde{\mathbf{U}}_{\mathbf{P}}\widetilde{\mathbf{U}}_{\mathbf{P}}^\top\widetilde{\mathbf{U}}_{\mathbf{A}}$$

$$= O_{\mathbb{P}}((n\rho_n)^{-1}).$$

Equation (B.8) is thereby established. We now establish equation (B.9), noting that the same argument applies also to equation (B.10). For $i, j \in \{1, 2, \ldots, d\}$, let r_{ij} denote the ijth entry of $\widetilde{\mathbf{U}}_{\mathbf{P}}^{\top}\widetilde{\mathbf{U}}_{\mathbf{A}}$. Also, for $i \in \{1, 2, \ldots, d\}$, let $\widetilde{\lambda}_i(\mathbf{A})$ and $\widetilde{\lambda}_j(\mathbf{P})$ denote the ith eigenvalue of $\mathcal{L}(\mathbf{A})$ and $\mathcal{L}(\mathbf{P})$, respectively. Then the ijth entry of $\widetilde{\mathbf{U}}_{\mathbf{P}}^{\top}\widetilde{\mathbf{U}}_{\mathbf{A}}\widetilde{\mathbf{S}}_{\mathbf{A}}^{1/2} - \widetilde{\mathbf{S}}_{\mathbf{P}}^{1/2}\widetilde{\mathbf{U}}_{\mathbf{P}}^{\top}\widetilde{\mathbf{U}}_{\mathbf{A}}$ is of the form

$$r_{ij}(\widetilde{\lambda}_j^{1/2}(\mathbf{A}) - \widetilde{\lambda}_i^{1/2}(\mathbf{P})) = \frac{r_{ij}(\widetilde{\lambda}_j(\mathbf{A}) - \widetilde{\lambda}_i(\mathbf{P}))}{\widetilde{\lambda}_j^{1/2}(\mathbf{A}) + \widetilde{\lambda}_i^{1/2}(\mathbf{P})}.$$

Since $\widetilde{\lambda}_i(\mathbf{A}) = \Theta_{\mathbb{P}}(1)$ and $\widetilde{\lambda}_j(\mathbf{P}) = \Theta_{\mathbb{P}}(1)$, the previous expression and equation (B.35) yield

$$r_{ij}(\widetilde{\lambda}_j^{1/2}(\mathbf{A}) - \widetilde{\lambda}_i^{1/2}(\mathbf{P})) = O_{\mathbb{P}}((n\rho_n)^{-1}).$$

A union bound over i, j then implies equation (B.9).

B.4. Proof of equation (3.4) **and equation** (3.5). Recall equation (3.1), that is, with $\zeta = (\check{\mathbf{X}}\mathbf{W} - \widetilde{\mathbf{X}})$, we have

$$\|\zeta\|_F = \left\|\mathbf{T}^{-1/2}(\mathbf{A} - \mathbf{P})\mathbf{T}^{-1/2}\widetilde{\mathbf{X}}(\widetilde{\mathbf{X}}^{\top}\widetilde{\mathbf{X}})^{-1} + \frac{1}{2}\mathbf{T}^{-1}(\mathbf{T} - \mathbf{D})\widetilde{\mathbf{X}}\right\|_F + O_{\mathbb{P}}((n\rho_n)^{-1}).$$

The above implies

$$\|\zeta\|_F^2 = \left\|\mathbf{T}^{-1/2}(\mathbf{A} - \mathbf{P})\mathbf{T}^{-1/2}\widetilde{\mathbf{X}}(\widetilde{\mathbf{X}}^{\top}\widetilde{\mathbf{X}})^{-1}\right\|_F^2 + \frac{1}{4}\left\|\mathbf{T}^{-1}(\mathbf{T} - \mathbf{D})\widetilde{\mathbf{X}}\right\|_F^2$$
$$+ \operatorname{tr}\widetilde{\mathbf{X}}^{\top}\mathbf{T}^{-1}(\mathbf{T} - \mathbf{D})\mathbf{T}^{-1/2}(\mathbf{A} - \mathbf{P})\mathbf{T}^{-1/2}\widetilde{\mathbf{X}}(\widetilde{\mathbf{X}}^{\top}\widetilde{\mathbf{X}})^{-1} + O_{\mathbb{P}}((n\rho_n)^{-3/2}).$$

We show equation (3.4) and equation (3.5) by analyzing each term in the right-hand side of the above display. In particular, we shall show that these terms are concentrated around their expected values; evaluation of these expected values, in the limit as $n \to \infty$, yield equation (3.4) and equation (3.5).

We first consider the term $Z = \|\mathbf{T}^{-1/2}(\mathbf{A} - \mathbf{P})\mathbf{T}^{-1/2}\widetilde{\mathbf{X}}(\widetilde{\mathbf{X}}^{\top}\widetilde{\mathbf{X}})^{-1}\|_F^2$. We note that conditional on $\mathbf{P}$, Z is a function of the $n(n-1)/2$ independent random variables $\{a_{ij}\}_{i<j}$. It is therefore expected that Z will be concentrated around its expectation $\mathbb{E}[Z]$ where the expectation is taken with respect to $\mathbf{A}$, conditional on $\mathbf{P}$. We verify this below.

Let $\mathbf{A}' = (a_{ij}')$ be an independent copy of $\mathbf{A}$, that is, the upper triangular entries of $\mathbf{A}'$ are independent Bernoulli random variables with mean parameters $\{p_{ij}\}_{i<j}$. Let $\mathbf{A}^{(ij)}$ be the matrix obtained by replacing the (i, j) and (j, i) entries of $\mathbf{A}$ by a_{ij}' and let $Z^{(ij)} = \|\mathbf{T}^{-1/2}(\mathbf{A}^{(ij)} - \mathbf{P})\mathbf{T}^{-1/2}\widetilde{\mathbf{X}}(\widetilde{\mathbf{X}}^{\top}\widetilde{\mathbf{X}})^{-1}\|_F^2$. We show concentration of Z around $\mathbb{E}[Z]$ using the following concentration inequality from [8], Theorem 5 and Theorem 6.

THEOREM B.5. *Assume that there exists positive constants a and b such that*

$$\sum_{i<j}(Z - Z^{(ij)})^2 \le aZ + b.$$

Then for all $t > 0$,

(B.36) $$\mathbb{P}[Z - \mathbb{E}[Z] \ge t] \le \exp\left(\frac{-t^2}{4a\mathbb{E}[Z] + 4b + 2at}\right),$$

(B.37) $$\mathbb{P}[Z - \mathbb{E}[Z] \le -t] \le \exp\left(\frac{-t^2}{4a\mathbb{E}[Z]}\right).$$

We now bound $\sum_{i<j}(Z - Z^{(ij)})^2$. For notational convenience, we denote the ith row of $\widetilde{\mathbf{X}}(\widetilde{\mathbf{X}}^\top\widetilde{\mathbf{X}})^{-1}$ by ζ_i and the ith row of $\mathbf{T}^{-1/2}(\mathbf{A} - \mathbf{P})\mathbf{T}^{-1/2}\widetilde{\mathbf{X}}(\widetilde{\mathbf{X}}^\top\widetilde{\mathbf{X}})^{-1}$ by ξ_i. We shall also denote the inner product between vectors in Euclidean space by $\langle\cdot,\cdot\rangle$. For each i, $\xi_i = \sum_{j=1}^{n}\frac{a_{ij}-p_{ij}}{\sqrt{t_i t_j}}\zeta_j$, and hence

$$Z = \sum_{k=1}^{n}\xi_k^2 = \sum_{k=1}^{n}\sum_{\ell=1}^{n}\sum_{\ell'=1}^{n}\frac{(a_{k\ell} - p_{k\ell})(a_{k\ell'} - p_{k\ell'})}{t_k\sqrt{t_\ell t_{\ell'}}}\langle\zeta_\ell, \zeta_{\ell'}\rangle.$$

Now $\mathbf{A}$ and $\mathbf{A}^{(ij)}$ differs possibly only in the (i, j) and (j, i) entries; furthermore, the $\{t_i\}$ do not depend on the entries of $\mathbf{A}$ and $\mathbf{A}^{(ij)}$. We thus have, upon considering the cases where $k = i$ and $\ell = j$, $k = j$ and $\ell = i$, $k = i$ and $\ell' = j$, and $k = j$ and $\ell' = i$, that

$$Z - Z^{(ij)} = \sum_{\ell'=1}^{n}\frac{(a_{ij} - a'_{ij})(a_{i\ell'} - p_{i\ell'})}{t_i\sqrt{t_j t_{\ell'}}}\langle\zeta_j, \zeta_{\ell'}\rangle$$

$$+ \sum_{\ell'=1}^{n}\frac{(a_{ji} - a'_{ji})(a_{j\ell'} - p_{j\ell'})}{t_j\sqrt{t_i t_{\ell'}}}\langle\zeta_i, \zeta_{\ell'}\rangle$$

$$+ \sum_{\ell=1}^{n}\frac{(a_{i\ell} - p_{i\ell})(a_{ij} - a'_{ij})}{t_i\sqrt{t_j t_\ell}}\langle\zeta_j, \zeta_\ell\rangle$$

$$+ \sum_{\ell=1}^{n}\frac{(a_{j\ell} - p_{j\ell})(a_{ji} - a'_{ji})}{t_j\sqrt{t_i t_\ell}}\langle\zeta_i, \zeta_\ell\rangle.$$

Since $a_{ij} = a_{ji}$ and $a'_{ij} = a'_{ji}$, the above simplifies to

$$Z - Z^{(ij)} = 2(a_{ij} - a'_{ij})\sum_{\ell=1}^{n}\left(\frac{a_{i\ell} - p_{i\ell}}{t_i\sqrt{t_j t_\ell}}\langle\zeta_j, \zeta_\ell\rangle + \frac{a_{j\ell} - p_{j\ell}}{t_j\sqrt{t_i t_\ell}}\langle\zeta_i, \zeta_\ell\rangle\right).$$

We then have, since a_{ij} and a'_{ij} are binary variables, that is, $|a_{ij} - a'_{ij}| \le 1$, that

$$(Z - Z^{(ij)})^2 \le 4\left(\sum_{\ell=1}^{n}\frac{a_{i\ell} - p_{i\ell}}{t_i\sqrt{t_j t_\ell}}\langle\zeta_j, \zeta_\ell\rangle\right)^2 + 4\left(\sum_{\ell=1}^{n}\frac{a_{j\ell} - p_{j\ell}}{t_j\sqrt{t_i t_\ell}}\langle\zeta_i, \zeta_\ell\rangle\right)^2.$$

Now $(t_j t_l)^{-1/2}\langle \zeta_j, \zeta_l \rangle$ is the (l, j)th entry of $\mathbf{T}^{-1/2}\widetilde{\mathbf{X}}(\widetilde{\mathbf{X}}^{\top}\widetilde{\mathbf{X}})^{-1}(\widetilde{\mathbf{X}}^{\top}\widetilde{\mathbf{X}})^{-1}\widetilde{\mathbf{X}}\mathbf{T}^{-1/2}$.
Thus $\sum_{\ell=1}^n \frac{a_{i\ell} - p_{i\ell}}{t_i\sqrt{t_j t_\ell}}\langle \zeta_j, \zeta_\ell \rangle$ is the (i, j)th entry of $\mathbf{T}^{-1}(\mathbf{A} - \mathbf{P})\mathbf{T}^{-1/2}\widetilde{\mathbf{X}}(\widetilde{\mathbf{X}}^{\top}\widetilde{\mathbf{X}})^{-2} \times$
$\widetilde{\mathbf{X}}\mathbf{T}^{-1/2}$. We therefore have

$$
\sum_{i<j}(Z - Z^{(ij)})^2
$$

$$
\leq 4\sum_{i<j}\left(\sum_{\ell=1}^n \frac{a_{i\ell} - p_{i\ell}}{t_i\sqrt{t_j t_\ell}}\langle \zeta_j, \zeta_\ell \rangle\right)^2 + 4\sum_{i<j}\left(\sum_{\ell=1}^n \frac{a_{j\ell} - p_{j\ell}}{t_j\sqrt{t_i t_\ell}}\langle \zeta_i, \zeta_\ell \rangle\right)^2
$$

$$
\leq 8\sum_{i=1}^n\sum_{j=1}^n\left(\sum_{\ell=1}^n \frac{a_{i\ell} - p_{i\ell}}{t_i\sqrt{t_j t_\ell}}\langle \zeta_j, \zeta_\ell \rangle\right)^2
$$

$$
\leq 8\|\mathbf{T}^{-1}(\mathbf{A} - \mathbf{P})\mathbf{T}^{-1/2}\widetilde{\mathbf{X}}(\widetilde{\mathbf{X}}^{\top}\widetilde{\mathbf{X}})^{-2}\widetilde{\mathbf{X}}\mathbf{T}^{-1/2}\|_F^2
$$

$$
\leq 8\|\mathbf{T}^{-1/2}(\mathbf{A} - \mathbf{P})\mathbf{T}^{-1/2}\widetilde{\mathbf{X}}(\widetilde{\mathbf{X}}^{\top}\widetilde{\mathbf{X}})^{-1}\|_F^2\|(\widetilde{\mathbf{X}}^{\top}\widetilde{\mathbf{X}})^{-1}\widetilde{\mathbf{X}}\mathbf{T}^{-1/2}\|^2\|\mathbf{T}^{-1/2}\|^2
$$

$$
\leq 8Z\|(\widetilde{\mathbf{X}}^{\top}\widetilde{\mathbf{X}})^{-1}\widetilde{\mathbf{X}}\mathbf{T}^{-1/2}\|^2\|\mathbf{T}^{-1/2}\|^2
$$

$$
\leq 8Z\|\rho_n^{1/2}(\mathbf{X}^{\top}\mathbf{T}^{-1}\mathbf{X})^{-1}\mathbf{X}\mathbf{T}^{-1}\|^2\|\mathbf{T}^{-1/2}\|^2
$$

$$
\leq 8\rho_n Z\|(\mathbf{X}^{\top}\mathbf{T}^{-1}\mathbf{X})^{-1}\|^2\|\mathbf{X}\|^2\|\mathbf{T}^{-1}\|^2\|\mathbf{T}^{-1/2}\|^2
$$

$$
\leq 8\rho_n Z\|(\mathbf{X}^{\top}\mathbf{T}^{-1}\mathbf{X})^{-1}\|^2 n\delta^{-3}
$$

$$
\leq C\rho_n Z\|(\mathbf{X}^{\top}\mathbf{T}^{-1}\mathbf{X})^{-1}\|^2 n(n\rho_n)^{-3}
$$

$$
\leq C(n\rho_n)^{-2}Z
$$

for some constant C; note that C denote a generic constant, not depending on Z, in the above display and could change from line to line. In the above derivation, we have used the fact that $C_0\sqrt{n} \leq \|\mathbf{X}\| \leq \sqrt{n}$ for some constant $C_0 > 0$ and $\|\mathbf{T}\| \geq \delta \geq C_1 n\rho_n$ for some constant $C_1 > 0$.

We then have, by Theorem B.5, that for all $t > 0$,

$$
(\text{B.38}) \qquad \mathbb{P}\big[Z - \mathbb{E}[Z] > t\big] \leq \exp\left(\frac{-Ct^2}{(n\rho_n)^{-2}\mathbb{E}[Z] + 2(n\rho_n)^{-2}t}\right),
$$

$$
(\text{B.39}) \qquad \mathbb{P}\big[Z - \mathbb{E}[Z] > -t\big] \leq \exp\left(\frac{-Ct^2}{(n\rho_n)^{-2}\mathbb{E}[Z]}\right).
$$

In addition, it is straightforward to see that $\mathbb{E}[Z] \leq C_3(n\rho_n)^{-1}$, for some constant $C_3 > 0$; here, the expectation is taken with respect to $\mathbf{A}$ conditional on $\mathbf{P}$. We therefore have that there exists a constant $C > 0$ such that $t = C(n\rho_n)^{-3/2}\log^{1/2} n$

yield

(B.40)
$$Z = \mathbb{E}[Z] + O_{\mathbb{P}}\big((n\rho_n)^{-3/2}\log^{1/2}\big)$$

$$= \mathbb{E}\|\mathbf{T}^{-1/2}(\mathbf{A}-\mathbf{P})\mathbf{T}^{-1/2}\widetilde{\mathbf{X}}(\widetilde{\mathbf{X}}^\top\widetilde{\mathbf{X}})^{-1}\|_F^2 + O_{\mathbb{P}}\big((n\rho_n)^{-3/2}\log^{1/2}\big).$$

We now evaluate $\mathbb{E}[Z]$. We have

$$\mathbb{E}[Z] = \mathbb{E}\big[\operatorname{tr}(\widetilde{\mathbf{X}}^\top\widetilde{\mathbf{X}})^{-1}\widetilde{\mathbf{X}}^\top\mathbf{T}^{-1/2}(\mathbf{A}-\mathbf{P})\mathbf{T}^{-1}(\mathbf{A}-\mathbf{P})\mathbf{T}^{-1/2}\widetilde{\mathbf{X}}(\widetilde{\mathbf{X}}^\top\widetilde{\mathbf{X}})^{-1}\big]$$

$$= \operatorname{tr}(\widetilde{\mathbf{X}}^\top\widetilde{\mathbf{X}})^{-1}\widetilde{\mathbf{X}}^\top\mathbb{E}\big[\mathbf{T}^{-1/2}(\mathbf{A}-\mathbf{P})\mathbf{T}^{-1}(\mathbf{A}-\mathbf{P})\mathbf{T}^{-1/2}\big]\widetilde{\mathbf{X}}(\widetilde{\mathbf{X}}^\top\widetilde{\mathbf{X}})^{-1}.$$

We note that $\mathbf{T}^{-1/2}(\mathbf{A}-\mathbf{P})\mathbf{T}^{-1}(\mathbf{A}-\mathbf{P})\mathbf{T}^{-1/2}$ is a $n \times n$ matrix whose ijth entry ξ_{ij} is of the form

$$\xi_{ij} := \sum_k t_i^{-1/2} t_k^{-1} t_j^{-1/2}(a_{ik}-p_{ik})(a_{kj}-p_{kj})$$

and hence

$$\mathbb{E}[\xi_{ij}] = \begin{cases} 0 & \text{if } i \neq j, \\ \sum_k t_i^{-1} t_k^{-1} p_{ik}(1-p_{ik}) & \text{if } i = j. \end{cases}$$

We shall denote by $\widetilde{\mathbf{M}}$ the diagonal matrix $(\mathbb{E}[\xi_{ij}])$ as given above. Then

$$n\rho_n\mathbb{E}[Z] = n\rho_n\operatorname{tr}(\widetilde{\mathbf{X}}^\top\widetilde{\mathbf{X}})^{-1}\widetilde{\mathbf{X}}^\top\widetilde{\mathbf{M}}\widetilde{\mathbf{X}}(\widetilde{\mathbf{X}}^\top\widetilde{\mathbf{X}})^{-1}.$$

We first recall from equation (B.2) that $\widetilde{\mathbf{X}}^\top\widetilde{\mathbf{X}} \overset{\text{a.s.}}{\to} \widetilde{\Delta}$ and $(\widetilde{\mathbf{X}}^\top\widetilde{\mathbf{X}})^{-1} \overset{\text{a.s.}}{\to} \widetilde{\Delta}^{-1}$ as $n \to \infty$. We next consider $n\rho_n\widetilde{\mathbf{X}}^\top\widetilde{\mathbf{M}}\widetilde{\mathbf{X}}$. Let $\widetilde{m}_i$ denote the ith diagonal element of $\widetilde{\mathbf{M}}$. We have

$$n\rho_n\widetilde{\mathbf{X}}^\top\widetilde{\mathbf{M}}\widetilde{\mathbf{X}} = n\rho_n\sum_i \frac{\rho_n X_i X_i^\top \widetilde{m}_i}{t_i}$$

$$= n\rho_n\sum_i \frac{\rho_n X_i X_i^\top \widetilde{m}_i}{\rho_n\sum_j X_i^\top X_j}$$

$$= \sum_i \frac{X_i X_i^\top n\rho_n\widetilde{m}_i}{n X_i^\top \mu} + \sum_i \frac{X_i X_i^\top n\rho_n\widetilde{m}_i}{n X_i^\top \mu}\left(\frac{n X_i^\top \mu - \sum_j X_i^\top X_j}{\sum_j X_i^\top X_j}\right).$$

Similar to our derivation of equation (B.2), we have

$$-\Big(\sup_{j\in[n]} n\rho_n\widetilde{m}_j c_j\Big)\sum_i \frac{X_i X_i^\top}{n X_i^\top \mu} \preceq \sum_i \frac{X_i X_i^\top n\rho_n\widetilde{m}_i}{n X_i^\top \mu}\left(\frac{n X_i^\top \mu - \sum_j X_i^\top X_j}{\sum_j X_i^\top X_J}\right)$$

$$\preceq \Big(\sup_{j\in[n]} n\rho_n\widetilde{m}_j c_j\Big)\sum_i \frac{X_i X_i^\top}{n X_i^\top \mu}.$$

In addition, for each index i,

$$n\rho_n \widetilde{m}_i = n\rho_n \sum_k t_i^{-1} t_k^{-1} p_{ik}(1 - p_{ik}) = O_{\mathbb{P}}(1)$$

and hence $\sup_{i \in [n]} n\rho_n \widetilde{m}_i c_i \xrightarrow{\text{a.s.}} 0$ as $n \to \infty$. Therefore,

$$(B.41) \qquad \sum_i \frac{X_i X_i^\top n\rho_n \widetilde{m}_i}{n X_i^\top \mu} \left(\frac{n X_i^\top \mu - \sum_j X_i^\top X_j}{\sum_j X_i^\top X_J} \right) \xrightarrow{\text{a.s}} \mathbf{0}$$

as $n \to \infty$. We thus only need to consider

$$\sum_i \frac{X_i X_i^\top n\rho_n \widetilde{m}_i}{n X_i^\top \mu} = \sum_i \sum_k \frac{\rho_n X_i X_i^\top p_{ik}(1 - p_{ik})}{(X_i^\top \mu) t_i t_k}$$

$$= \sum_i \sum_k \frac{\rho_n X_i X_i^\top \rho_n X_i^\top X_k (1 - \rho_n X_i^\top X_k)}{(X_i^\top \mu) \sum_j \rho_n X_i^\top X_j \sum_l \rho_n X_k^\top X_l}$$

$$= \sum_i \sum_k \frac{X_i X_i^\top (X_i^\top X_k - \rho_n X_i^\top X_k X_k^\top X_i)}{(X_i^\top \mu) \sum_j X_i^\top X_j \sum_l X_k^\top X_l}$$

$$= \sum_i \sum_k \frac{X_i X_i^\top (X_i^\top X_k - \rho_n X_i^\top X_k X_k^\top X_i)}{n^2 (X_i^\top \mu)^2 (X_k^\top \mu)}$$

$$+ \sum_i \sum_k \frac{X_i X_i^\top (X_i^\top X_k - \rho_n X_i^\top X_k X_k^\top X_i)}{n^2 (X_i^\top \mu)^2 (X_k^\top \mu)}$$

$$\times \left(\frac{n^2 (X_i^\top \mu)(X_k^\top \mu) - \sum_j X_i^\top X_j \sum_l X_k^\top X_l}{\sum_j X_i^\top X_j \sum_l X_k^\top X_l} \right).$$

An analogous argument to that used in deriving equation (B.41) yields

$$\sum_i \sum_k \frac{X_i X_i^\top (X_i^\top X_k - \rho_n X_i^\top X_k X_k^\top X_i)}{n^2 (X_i^\top \mu)^2 (X_k^\top \mu)}$$

$$\times \left(\frac{n^2 (X_i^\top \mu)(X_k^\top \mu) - \sum_j X_i^\top X_j \sum_l X_k^\top X_l}{\sum_j X_i^\top X_j \sum_l X_k^\top X_l} \right) \xrightarrow{\text{a.s.}} \mathbf{0}$$

as $n \to \infty$. It thus remains to evaluate

$$\sum_i \sum_k \frac{X_i X_i^\top (X_i^\top X_k - \rho_n X_i^\top X_k X_k^\top X_i)}{n^2 (X_i^\top \mu)^2 (X_k^\top \mu)}.$$

The strong law of large numbers implies

$$\sum_i \sum_k \frac{X_i X_i^\top X_i^\top X_k}{n^2 (X_i^\top \mu)^2 (X_k^\top \mu)} \xrightarrow{\text{a.s.}} \mathbb{E}\left[\frac{X_1 X_1^\top X_1^\top \tilde{\mu}}{(X_1^\top \mu)^2}\right],$$

$$\rho_n \sum_i \sum_k \frac{X_i X_i^\top X_i^\top X_k X_k^\top X_i}{n^2 (X_i^\top \mu)^2 (X_k^\top \mu)} \rightarrow \rho_n \mathbb{E}\left[\frac{X_1 X_1^\top X_1^\top \tilde{\Delta} X_1}{(X_1^\top \mu)^2}\right].$$

We invoke Slutsky's theorem and conclude that

(B.42)
$$\begin{aligned}
n\rho_n Z &= n\rho_n \|\mathbf{T}^{-1/2}(\mathbf{A} - \mathbf{P})\mathbf{T}^{-1/2}\tilde{\mathbf{X}}(\tilde{\mathbf{X}}^\top \tilde{\mathbf{X}})^{-1}\|_F^2 \\
&= n\rho_n \operatorname{tr}(\tilde{\mathbf{X}}^\top \tilde{\mathbf{X}})^{-1} \tilde{\mathbf{X}}^\top \tilde{\mathbf{M}}\tilde{\mathbf{X}}(\tilde{\mathbf{X}}^\top \tilde{\mathbf{X}})^{-1} + O_{\mathbb{P}}((n\rho_n)^{-1/2}\log^{1/2} n) \\
&\rightarrow \operatorname{tr}\tilde{\Delta}^{-1}\mathbb{E}\left[\frac{X_1 X_1^\top (X_1^\top \tilde{\mu} - \rho_n X_1^\top \tilde{\Delta} X_1)}{(X_1^\top \mu)^2}\right]\tilde{\Delta}^{-1}.
\end{aligned}$$

We next bound $Z := \|(\mathbf{T} - \mathbf{D})\mathbf{T}^{-1}\tilde{\mathbf{X}}\|_F^2$. Z is again a function of the $n(n-1)/2$ independent random variables $\{a_{ij}\}_{i<j}$. Let $Z^{(ij)} = \|(\mathbf{T} - \mathbf{D}^{(ij)})\mathbf{T}^{-1}\tilde{\mathbf{X}}\|$ where $\mathbf{D}^{(ij)}$ is the diagonal matrix whose diagonal entries are the degrees of $\mathbf{A}^{(ij)}$; we recall that $\mathbf{A}^{(ij)}$ is obtained by replacing the (i, j) and (j, i) entries of $\mathbf{A}$ with an independent copy a'_{ij} of a_{ij}. We now bound $\sum_{i<j}(Z - Z^{(ij)})^2$. Let $\tilde{X}_i$ denote the ith row of $\tilde{\mathbf{X}}$. Then

$$Z = \sum_k \frac{(t_k - d_k)^2}{t_k^2}\|\tilde{X}_k\|^2,$$

and hence (with $d_k^{(ij)}$ denoting the degree of vertex k in $\mathbf{A}^{(ij)}$)

$$\begin{aligned}
Z - Z^{(ij)} &= \sum_k \left((t_k - d_k)^2 - (t_k - d_k^{(ij)})^2\right)\frac{\|\tilde{X}_k\|^2}{t_k^2} \\
&= \sum_k (d_k^{(ij)} - d_k)(2t_k - d_k - d_k^{(ij)})\frac{\|\tilde{X}_k\|^2}{t_k^2} \\
&= (a'_{ij} - a_{ij})\Big((2t_i - 2d_i + a_{ij} - a'_{ij})\frac{\|\tilde{X}_i\|^2}{t_i^2} \\
&\qquad + (2t_j - 2d_j + a_{ij} - a'_{ij})\frac{\|\tilde{X}_j\|^2}{t_j^2}\Big).
\end{aligned}$$

Using the fact that $(b + c)^2 \leq 2b^2 + 2c^2$ and that $a_{ij} = a_{ji}$, $a'_{ij} = a'_{ji}$ we have

$$
(Z - Z^{(ij)})^2 \leq 2(a'_{ij} - a_{ij})^2 (2t_i - 2d_i + a_{ij} - a'_{ij})^2 \frac{\|\widetilde{X}_i\|^4}{t_i^4}
$$

$$
+ 2(a'_{ji} - a_{ji})^2 (2t_j - 2d_j + a_{ji} - a'_{ji})^2 \frac{\|\widetilde{X}_j\|^4}{t_j^4},
$$

from which we derive

$$
\sum_{i<j} (Z - Z^{(ij)})^2 \leq \sum_{i=1}^{n} \sum_{j=1}^{n} (a'_{ij} - a_{ij})^2 (16(t_i - d_i)^2 + 4) \frac{\|\widetilde{X}_i\|^4}{t_i^4}
$$

$$
\leq \sum_{i=1}^{n} \sum_{j=1}^{n} (16(t_i - d_i)^2 + 4) \frac{\|\widetilde{X}_i\|^4}{t_i^4}
$$

$$
\leq \sum_{i=1}^{n} \sum_{j=1}^{n} (16(t_i - d_i)^2 + 4) \frac{\|\widetilde{X}_i\|^2}{t_i^2} \frac{\rho_n t_i^{-1} \|X_i\|^2}{t_i^2}
$$

$$
\leq C \sum_{i=1}^{n} \sum_{j=1}^{n} (16(t_i - d_i)^2 + 4) \frac{\|\widetilde{X}_i\|^2}{t_i^2} n^{-3} \rho_n^{-2}
$$

$$
\leq C_1 (n\rho_n)^{-2} Z + C_2 (n\rho_n)^{-4} \leq C_3 (n\rho_n)^{-2} Z
$$

for some constants $C_1, C_2, C_3 > 0$. Once again, we apply Theorem B.5 to conclude

$$
(B.43) \qquad \mathbb{P}[Z - \mathbb{E}[Z] > t] \leq \exp\left(\frac{-Ct^2}{(n\rho_n)^{-2}\mathbb{E}[Z] + 2(n\rho_n)^{-2}t} \right),
$$

$$
(B.44) \qquad \mathbb{P}[Z - \mathbb{E}[Z] > -t] \leq \exp\left(\frac{-Ct^2}{(n\rho_n)^{-2}\mathbb{E}[Z]} \right).
$$

In addition, $\mathbb{E}[Z] = \mathbb{E}[\sum_k (t_k - d_k)^2 t_k^{-2} \|\widetilde{X}_k\|^2] \leq C(n\rho_n)^{-1}$ for some constant $C > 0$; here, the expectation is taken with respect to $\mathbf{D}$ conditional on $\mathbf{P}$. We thus conclude

$$
(B.45) \qquad
\begin{aligned}
Z &= \|(\mathbf{T} - \mathbf{D})\mathbf{T}^{-1}\widetilde{\mathbf{X}}\|_F^2 \\
&= \mathbb{E}[\|\mathbf{T} - \mathbf{D})\mathbf{T}^{-1}\widetilde{\mathbf{X}}\|_F^2 + O_{\mathbb{P}}((n\rho_n)^{-3/2}\log^{1/2}(n)).
\end{aligned}
$$

We now evaluate $n\rho_n \mathbb{E}[\|\frac{1}{2}\mathbf{T}^{-1}(\mathbf{T} - \mathbf{D})\widetilde{\mathbf{X}}\|_F^2]$. We only sketch the argument, noting that the details follow in a similar manner to that used in deriving equation (B.42). We have that

$$
n\rho_n \mathbb{E}\left[\left\|\frac{1}{2}\mathbf{T}^{-1}(\mathbf{T} - \mathbf{D})\widetilde{\mathbf{X}}\right\|_F^2\right] = n\rho_n \frac{1}{4} \operatorname{tr} \widetilde{\mathbf{X}}^\top \mathbf{T}^{-1} \mathbb{E}[(\mathbf{T} - \mathbf{D})^2] \mathbf{T}^{-1} \widetilde{\mathbf{X}}.
$$

Now $\mathbf{T}^{-1}\mathbb{E}[(\mathbf{T} - \mathbf{D})^2]\mathbf{T}^{-1}$ is a diagonal matrix whose ith diagonal entry is of the form $t_i^{-2} \sum_j p_{ij}(1 - p_{ij})$. Hence,

$$
\begin{aligned}
n\rho_n \widetilde{\mathbf{X}}^\top \mathbf{T}^{-1} &\mathbb{E}[(\mathbf{T} - \mathbf{D})^2]\mathbf{T}^{-1}\widetilde{\mathbf{X}} \\
&= n\rho_n^2 \sum_i t_i^{-3} X_i X_i^\top \sum_j p_{ij}(1 - p_{ij}) \\
&= \sum_i \frac{n\rho_n^2 X_i X_i^\top}{(n\rho_n X_i^\top \mu)^{-3}} \sum_j \rho_n X_i^\top X_j (1 - \rho_n X_i^\top X_j) + o_{\mathbb{P}}(1) \\
&= \sum_i n^{-1} \frac{X_i X_i^\top}{(X_i^\top \mu)^3} \sum_j n^{-1} X_i^\top X_j (1 - \rho_n X_i^\top X_j) + o_{\mathbb{P}}(1) \\
&= \sum_i n^{-1} \frac{X_i X_i^\top}{(X_i^\top \mu)^3} \sum_j n^{-1} X_i^\top X_j (1 - \rho_n X_j^\top X_i) + o_{\mathbb{P}}(1).
\end{aligned}
$$

We therefore have

$$
(\text{B.46}) \quad n\rho_n \mathbb{E}\left[\left\|\frac{1}{2}\mathbf{T}^{-1}(\mathbf{T} - \mathbf{D})\widetilde{\mathbf{X}}\right\|_F^2\right] \xrightarrow{\text{a.s.}} \frac{1}{4}\operatorname{tr}\left(\mathbb{E}\left[\frac{X_i X_i^\top}{(X_i^\top \mu)^2}\left(1 - \frac{\rho_n X_i^\top \Delta X_i}{(X_i^\top \mu)^3}\right)\right]\right)
$$

as $n \to \infty$.

Finally, we consider $Z := n\rho_n \operatorname{tr} \widetilde{\mathbf{X}}^\top \mathbf{T}^{-1}(\mathbf{T} - \mathbf{D})\mathbf{T}^{-1/2}(\mathbf{A} - \mathbf{P})\mathbf{T}^{-1/2}\widetilde{\mathbf{X}}(\widetilde{\mathbf{X}}^\top \widetilde{\mathbf{X}})^{-1}$. A similar, albeit slightly more tedious argument to that used in deriving equation (B.40) and equation (B.45) yields

$$
\begin{aligned}
Z &= \operatorname{tr} \widetilde{\mathbf{X}}^\top \mathbf{T}^{-1}(\mathbf{T} - \mathbf{D})\mathbf{T}^{-1/2}(\mathbf{A} - \mathbf{P})\mathbf{T}^{-1/2}\widetilde{\mathbf{X}}(\widetilde{\mathbf{X}}^\top \widetilde{\mathbf{X}})^{-1} \\
&= \operatorname{tr} \mathbb{E}\left[\widetilde{\mathbf{X}}^\top \mathbf{T}^{-1}(\mathbf{T} - \mathbf{D})\mathbf{T}^{-1/2}(\mathbf{A} - \mathbf{P})\mathbf{T}^{-1/2}\widetilde{\mathbf{X}}(\widetilde{\mathbf{X}}^\top \widetilde{\mathbf{X}})^{-1}\right] \\
&\quad + O_{\mathbb{P}}((n\rho_n)^{-3/2} \log^{1/2} n).
\end{aligned}
$$

We now evaluate $\mathbb{E}[Z]$. We have

$$
\mathbb{E}[Z] = \operatorname{tr} \widetilde{\mathbf{X}}^\top \mathbf{T}^{-3/2}\mathbb{E}[(\mathbf{T} - \mathbf{D})(\mathbf{A} - \mathbf{P})]\mathbf{T}^{-1/2}\widetilde{\mathbf{X}}(\widetilde{\mathbf{X}}^\top \widetilde{\mathbf{X}})^{-1}.
$$

Now the ijth entry of $\mathbb{E}[(\mathbf{A} - \mathbf{P})(\mathbf{T} - \mathbf{D})]$ is of the form

$$
\begin{aligned}
\mathbb{E}[(a_{ij} - p_{ij})(t_j - d_j)] &= \mathbb{E}\left[(a_{ij} - p_{ij})\sum_k (p_{jk} - a_{jk})\right] \\
&= \sum_k \mathbb{E}[(a_{ij} - p_{ij})(p_{kj} - a_{kj})] = -p_{ij}(1 - p_{ij}),
\end{aligned}
$$

and hence, with $\circ$ denoting the Hadamard product of matrices,

$$
\begin{aligned}
n\rho_n \mathbb{E}[Z] &= -n\rho_n \operatorname{tr} \widetilde{\mathbf{X}}^\top \mathbf{T}^{-3/2}(\mathbf{P} - \mathbf{P} \circ \mathbf{P})\mathbf{T}^{-1/2}\widetilde{\mathbf{X}}(\widetilde{\mathbf{X}}^\top \widetilde{\mathbf{X}})^{-1} \\
&= -n\rho_n \operatorname{tr} \widetilde{\mathbf{X}}^\top \mathbf{T}^{-1}(\mathbf{T}^{-1/2}\mathbf{P}\mathbf{T}^{-1/2} \\
&\qquad - \mathbf{T}^{-1/2}(\mathbf{P} \circ \mathbf{P})\mathbf{T}^{-1/2})\widetilde{\mathbf{X}}(\widetilde{\mathbf{X}}^\top \widetilde{\mathbf{X}})^{-1} \\
&= -n\rho_n \operatorname{tr} \widetilde{\mathbf{X}}^\top \mathbf{T}^{-1}(\widetilde{\mathbf{X}}\widetilde{\mathbf{X}}^\top - \mathbf{T}^{-1/2}(\mathbf{P} \circ \mathbf{P})\mathbf{T}^{-1/2})\widetilde{\mathbf{X}}(\widetilde{\mathbf{X}}^\top \widetilde{\mathbf{X}})^{-1} \\
&= -n\rho_n \operatorname{tr} \widetilde{\mathbf{X}}^\top \mathbf{T}^{-1}\widetilde{\mathbf{X}} + n\rho_n \operatorname{tr} \widetilde{\mathbf{X}}^\top \mathbf{T}^{-3/2}(\mathbf{P} \circ \mathbf{P})\mathbf{T}^{-1/2}\widetilde{\mathbf{X}}(\widetilde{\mathbf{X}}^\top \widetilde{\mathbf{X}})^{-1}.
\end{aligned}
\tag{B.47}
$$

We first consider the term $n\rho_n \operatorname{tr} \widetilde{\mathbf{X}}^\top \mathbf{T}^{-1}\widetilde{\mathbf{X}}$. We have

$$
n\rho_n \operatorname{tr} \widetilde{\mathbf{X}}^\top \mathbf{T}^{-1}\widetilde{\mathbf{X}} = -n\rho_n \sum_i \frac{\rho_n X_i X_i^\top}{t_i^2} = -\frac{1}{n}\sum_i \frac{X_i X_i^\top}{(X_i^\top \mu)^2} + o_\mathbb{P}(1),
$$

and hence

$$
-n\rho_n \operatorname{tr} \widetilde{\mathbf{X}}^\top \mathbf{T}^{-1}\widetilde{\mathbf{X}} \xrightarrow{\text{a.s.}} -\operatorname{tr} \mathbb{E}\left[\frac{X_1 X_1^\top}{(X_1^\top \mu)^2}\right].
\tag{B.48}
$$

Finally, we consider the term $n\rho_n \operatorname{tr} \widetilde{\mathbf{X}}^\top \mathbf{T}^{-3/2}(\mathbf{P} \circ \mathbf{P})\mathbf{T}^{-1/2}\widetilde{\mathbf{X}}(\widetilde{\mathbf{X}}^\top \widetilde{\mathbf{X}})^{-1}$. We recall that $(\widetilde{\mathbf{X}}^\top \widetilde{\mathbf{X}})^{-1} \xrightarrow{\text{a.s.}} \widetilde{\Delta}^{-1}$ as $n \to \infty$. In addition,

$$
\begin{aligned}
n\rho_n \widetilde{\mathbf{X}}^\top \mathbf{T}^{-3/2}(\mathbf{P} \circ \mathbf{P})\mathbf{T}^{-1/2}\widetilde{\mathbf{X}} &= n\rho_n^2 \mathbf{X}^\top \mathbf{T}^{-2}(\mathbf{P} \circ \mathbf{P})\mathbf{T}^{-1}\mathbf{X} \\
&= n\rho_n^2 \sum_i \sum_j \frac{p_{ij}^2}{t_i^2 t_j} X_i X_j^\top \\
&= n\rho_n^2 \sum_i \sum_j \frac{p_{ij}^2}{(n\rho_n)^3 (X_i^\top \mu)^2 X_j^\top \mu} X_i X_j^\top + o_\mathbb{P}(1) \\
&= n\rho_n^2 \sum_i \sum_j \frac{\rho_n^2 (X_i^\top X_j)^2}{(n\rho_n)^3 (X_i^\top \mu)^2 X_j^\top \mu} X_i X_j^\top + o_\mathbb{P}(1) \\
&= \rho_n \sum_i \frac{1}{n} \sum_j \frac{1}{n} \frac{X_i^\top X_j X_j^\top X_i}{(X_i^\top \mu)^2 X_j^\top \mu} X_i X_j^\top + o_\mathbb{P}(1).
\end{aligned}
$$

We thus conclude

$$
\begin{aligned}
&n\rho_n \widetilde{\mathbf{X}}^\top \mathbf{T}^{-3/2}(\mathbf{P} \circ \mathbf{P})\mathbf{T}^{-1/2}\widetilde{\mathbf{X}}(\widetilde{\mathbf{X}}^\top \widetilde{\mathbf{X}})^{-1} \\
&\qquad \xrightarrow{\text{a.s.}} \rho_n \operatorname{tr} \mathbb{E}\left[\frac{X_1^\top X_2 X_2^\top X_1}{(X_1^\top \mu)^2 X_2^\top \mu} X_1 X_2^\top\right]\widetilde{\Delta}^{-1},
\end{aligned}
\tag{B.49}
$$

where the expectation is taken with respect to X_1, X_2 being i.i.d. drawn from F. Combining equation (B.48) and equation (B.49) yield

$$n\rho_n \operatorname{tr} \mathbb{E}\big[\widetilde{\mathbf{X}}^\top \mathbf{T}^{-1}(\mathbf{T} - \mathbf{D})\mathbf{T}^{-1/2}(\mathbf{A} - \mathbf{P})\mathbf{T}^{-1/2}\widetilde{\mathbf{X}}(\widetilde{\mathbf{X}}^\top \widetilde{\mathbf{X}})^{-1}\big]$$

(B.50)

$$\xrightarrow{\text{a.s.}} \rho_n \operatorname{tr} \mathbb{E}\left[\frac{X_1^\top X_2 X_2^\top X_1}{(X_1^\top \mu)^2 X_2^\top \mu} X_1 X_2^\top\right]\widetilde{\Delta}^{-1} - \operatorname{tr} \mathbb{E}\left[\frac{X_1 X_1^\top}{(X_1^\top \mu)^2}\right].$$

Equation (3.4) and equation (3.5) then follows directly from equation (B.42), equation (B.46) and equation (B.50).

REFERENCES

[1] ALI, S. M. and SILVEY, S. D. (1966). A general class of coefficients of divergence of one distribution from another. *J. Roy. Statist. Soc. Ser. B* **28** 131–142. MR0196777

[2] AMINI, A., CHEN, A., BICKEL, P. and LEVINA, E. (2013). Pseudo-likelihood methods for community detection in large sparse networks. *Ann. Statist.* **41** 2097–2122. MR3127859

[3] ASTA, D. and SHALIZI, C. (2014). Geometric network comparison. arXiv preprint at arXiv:1411.1350.

[4] ATHREYA, A., LYZINSKI, V., MARCHETTE, D. J., PRIEBE, C. E., SUSSMAN, D. L. and TANG, M. (2016). A limit theorem for scaled eigenvectors of random dot product graphs. *Sankhya A* **78** 1–18. MR3494576

[5] BELKIN, M. and NIYOGI, P. (2003). Laplacian eigenmaps for dimensionality reduction and data representation. *Neural Comput.* **15** 1373–1396.

[6] BICKEL, P. J. and CHEN, A. (2009). A nonparametric view of network models and Newman–Girvan and other modularities. *Proc. Natl. Acad. Sci. USA* **106** 21068–21073.

[7] BOLLOBÁS, B., JANSON, S. and RIORDAN, O. (2007). The phase transition in inhomogeneous random graphs. *Random Structures Algorithms* **31** 3–122. MR2337396

[8] BOUCHERON, S., LUGOSI, G. and MASSART, P. (2003). Concentration inequalities using the entropy method. *Ann. Probab.* **31** 1583–1614. MR1989444

[9] CHAUDHURI, K., CHUNG, F. and TSIATAS, A. (2012). Spectral partitioning of graphs with general degrees and the extended planted partition model. In *Proceedings of the 25th Conference on Learning Theory*.

[10] CHERNOFF, H. (1952). A measure of asymptotic efficiency for tests of a hypothesis based on the sum of observations. *Ann. Math. Stat.* **23** 493–507.

[11] CHERNOFF, H. (1956). Large sample theory: Parametric case. *Ann. Math. Stat.* **27** 1–22.

[12] CHOI, D. S., WOLFE, P. J. and AIROLDI, E. M. (2012). Stochastic blockmodels with a growing number of classes. *Biometrika* **99** 273–284.

[13] CHUNG, F. R. K. (1997). *Spectral Graph Theory* **92**. Amer. Math. Soc.

[14] COIFMAN, R. and LAFON, S. (2006). Diffusion maps. *Appl. Comput. Harmon. Anal.* **21** 5–30.

[15] CSIZÁR, I. (1967). Information-type measures of difference of probability distributions and indirect observations. *Studia Sci. Math. Hungar.* **2** 229–318.

[16] DIACONIS, P. and JANSON, S. (2008). Graph limits and exchangeable random graphs. *Rend. Mat. Appl. (7)* **28** 33–61.

[17] FRALEY, C. and RAFTERY, A. E. (1999). MCLUST: Software for model-based cluster analysis. *J. Classification* **16** 297–306.

[18] HOFF, P. D., RAFTERY, A. E. and HANDCOCK, M. S. (2002). Latent space approaches to social network analysis. *J. Amer. Statist. Assoc.* **97** 1090–1098. MR1951262

[19] HOLLAND, P. W., LASKEY, K. B. and LEINHARDT, S. (1983). Stochastic blockmodels: First steps. *Soc. Netw.* **5** 109–137.

[20] JOSEPH, A. and YU, B. (2016). Impact of regularization on spectral clustering. *Ann. Statist.* **44** 1765–1791. MR3519940

[21] KARRER, B. and NEWMAN, M. E. J. (2011). Stochastic blockmodels and community structure in networks. *Phys. Rev. E (3)* **83** 016107.

[22] LEANG, C. C. and JOHNSON, D. H. (1997). On the asymptotics of M-hypothesis Bayesian detection. *IEEE Trans. Inform. Theory* **43** 280–282.

[23] LEI, J. and RINALDO, A. (2015). Consistency of spectral clustering in stochastic block models. *Ann. Statist.* **43** 215–237. MR3285605

[24] LIESE, F. and VAJDA, I. (2006). On divergences and informations in statistics and information theory. *IEEE Trans. Inform. Theory* **52** 4394–4412. MR2300826

[25] LU, L. and PENG, X. (2013). Spectra of edge-independent random graphs. *Electron. J. Combin.* **20**.

[26] LUXBURG, U. V. (2007). A tutorial on spectral clustering. *Stat. Comput.* **17** 395–416. MR2409803

[27] LYZINSKI, V., SUSSMAN, D. L., TANG, M., ATHREYA, A. and PRIEBE, C. E. (2014). Perfect clustering for stochastic blockmodel graphs via adjacency spectral embedding. *Electron. J. Stat.* **8** 2905–2922. MR3299126

[28] MCSHERRY, F. (2001). Spectral partitioning of random graphs. In *Proceedings of the 42nd IEEE Symposium on Foundations of Computer Science* 529–537.

[29] MERRIS, R. (1994). Laplacian matrices of graphs: A survey. *Linear Algebra Appl.* **197** 143–176.

[30] MOSSEL, E., NEEMAN, J. and SLY, A. (2015). Reconstruction and estimation in the planted partition model. *Probab. Theory Related Fields* **162** 431–461. MR3383334

[31] NICKEL, C. L. M. (2006). Random dot product graphs: A model for social networks. Ph.D. thesis, Johns Hopkins Univ., Baltimore, MD.

[32] OLIVEIRA, R. I. (2009). Concentration of the adjacency matrix and of the Laplacian in random graphs with independent edges. arXiv preprint at arXiv:0911.0600.

[33] QIN, T. and ROHE, K. (2013). Regularized spectral clustering under the degree-corrected stochastic blockmodel. In *NIPS*.

[34] ROHE, K., CHATTERJEE, S. and YU, B. (2011). Spectral clustering and the high-dimensional stochastic blockmodel. *Ann. Statist.* **39** 1878–1915. MR2893856

[35] SARKAR, P. and BICKEL, P. J. (2015). Role of normalization for spectral clustering in stochastic blockmodels. *Ann. Statist.* **43** 962–990. MR3346694

[36] SHI, J. and MALIK, J. (2000). Normalized cuts and image segmentation. *IEEE Trans. Pattern Anal. Mach. Intell.* **22** 888–905.

[37] SNIJDERS, T. A. B. and NOWICKI, K. (1997). Estimation and prediction for stochastic blockmodels for graphs with latent block structure. *J. Classification* **14** 75–100.

[38] STEWART, G. W. and SUN, J. (1990). *Matrix Pertubation Theory*. Academic Press, New York.

[39] SUSSMAN, D. L. (2014). Foundations of adjacency spectral embedding. PhD Thesis, Johns Hopkins Univ., Baltimore, MD.

[40] SUSSMAN, D. L., TANG, M., FISHKIND, D. E. and PRIEBE, C. E. (2012). A consistent adjacency spectral embedding for stochastic blockmodel graphs. *J. Amer. Statist. Assoc.* **107** 1119–1128. MR3010899

[41] TANG, M., ATHREYA, A., SUSSMAN, D. L., LYZINSKI, V., PARK, Y. and PRIEBE, C. E. (2017). A semiparametric two-sample hypothesis testing problem for random dot product graphs. *J. Comput. Graph. Statist.* **26** 344–354.

[42] TANG, M., ATHREYA, A., SUSSMAN, D. L., LYZINSKI, V. and PRIEBE, C. E. (2017). A nonparametric two-sample hypothesis testing problem for random graphs. *Bernoulli* **23** 1599–1630. MR3624872

[43] VON LUXBURG, U., BELKIN, M. and BOUSQUET, O. (2008). Consistency of spectral clustering. *Ann. Statist.* **36** 555–586. MR2396807

[44] WOLFE, P. J. and OLHEDE, S. C. (2013). Nonparametric graphon estimation. arXiv preprint at arXiv:1309.5936.

[45] YANG, J. J., HAN, Q. and AIROLDI, E. M. (2014). Nonparametric estimation and testing of exchangeable graph models. In *Proceedings of the Seventeenth International Conference on Artificial Intelligence and Statistics* 1060–1067.

DEPARTMENT OF APPLIED MATHEMATICS
    AND STATISTICS
JOHNS HOPKINS UNIVERSITY
3400 N. CHARLES ST
BALTIMORE, MARYLAND 21218
USA
E-MAIL: minh@jhu.edu
           cep@jhu.edu