Sankhya A manuscript No.
(will be inserted by the editor)

A limit theorem for scaled eigenvectors of random dot product graphs

A. Athreya · V. Lyzinski · D.J. Marchette · C.E. Priebe · D. L. Sussman · M. Tang

the date of receipt and acceptance should be inserted later

Received: date / Accepted: date

Abstract We prove a central limit theorem for the components of the largest eigenvectors of the adjacency matrix of a finite-dimensional random dot product graph whose true latent positions are unknown. We use the spectral embedding of the adjacency matrix to construct consistent estimates for the latent positions, and we show that the appropriately scaled differences between the estimated and true latent positions converge to a mixture of Gaussian random variables. We state several corollaries, including an alternate proof of a central limit theorem for the first eigenvector of the adjacency matrix of an Erdős-Rényi random graph.

Keywords random dot product graph, central limit theorem, model-based clustering

1 Introduction and Background

Spectral analysis of the adjacency and Laplacian matrices for graphs is of both theoretical [5] and practical [18] significance. For instance, the spectrum can be used to characterize the number of connected components in a graph and various properties of random walks on graphs, and the eigenvector corresponding to the second smallest eigenvalue of the Laplacian is used in the solution to a relaxed version of the min-cut problem [7]. In our current work, we investigate the second-order properties of the eigenvectors corresponding to the largest eigenvalues of the adjacency matrix of a random graph. In particular, we show that

A. Athreya E-mail: dathrey1@jhu.edu · C.E. Priebe E-mail: cep@jhu.edu · M.Tang E-mail: mtang10@jhu.edu
Department of Applied Mathematics and Statistics, Johns Hopkins University
3400 N. Charles Street, Baltimore, Maryland 21218, U.S.A

V. Lyzinski E-mail: vlyzins1@jhu.edu
Johns Hopkins University Human Language Technology Center of Excellence
Stieff Building, 810 Wyman Park Drive, Baltimore, MD 21211

D.J. Marchette E-mail: dmarchette@gmail.com
Naval Surface Warfare Center, 17320 Dahlgren Road, Dahlgren, Virginia 22448, U.S.A

D.L. Sussman E-mail: daniellsussman@fas.harvard.edu
Department of Statistics, Harvard University
Science Center, Floor 7, One Oxford St, Cambridge, MA 02138

under the *random dot product graph* model [31], the components of the eigenvectors are asymptotically normal and centered around the true latent positions (see Section 2). We consider only undirected, loop-free graphs in which the expected number of edges grows as $\Theta(n^2)$. However, the results contained here can be extended to sparse graphs.

Random dot product graphs are a specific example of *latent position random graphs* [13], in which each vertex is associated with a latent position and, conditioned on the latent positions, the presence or absence of all edges in the graph are independent. The edge presence probability is based on *link* function, which is a symmetric function of the two latent positions.

For the random dot product graph, the latent positions are d-dimensional vectors and the link function is the inner product between the vectors of the incident vertices (see Definition 1). Random dot product graphs can be used to model certain social networks; for example, each individual in a social network might have, say, a number of different topics of interest to him or her. The d coordinates of a latent position vector represent those topics, and the value of the latent vector in a given direction represents the level of interest the individual might have in that topic [19]. Two individuals with the same interests, then, are more likely to be linked, whereas individuals with very disparate interests may be unlikely to know each other. Similar interpretations of these vectors can be made in other settings and for other latent position models.

We note briefly that, in a strong sense, latent position graphs are identical to *exchangeable random graphs* [1, 14], with the key unifying ingredient being the conditional independence of the edges. In [6], Diaconis and Janson prove several fundamental results on graph limits, constructed via subgraph counts, for exchangeable graphs. Their work has important consequences in statistical inference, for instance the method of moments for subgraph counts [3]. In a similar spirit, our current results provide asymptotic distributions for spectral statistics that have the promise to improve current statistical methodology for random graphs (see § 3).

Statistical analysis for latent position random graphs has received much recent interest: see [12] and [9] for reviews of the pertinent literature. Some fundamental results are found in [3,2,4] among many others. In the statistical analysis of latent position random graphs, a common strategy is to first estimate the latent positions based on some specified link function. For example, in random dot product graphs [31], the link function is the dot product: namely, the edge probabilities are the dot products of the latent positions. In [26], it is proved that spectral decompositions of the adjacency matrix for a random dot product graphs provide accurate estimates of the underlying latent positions. In this work, we extend the analysis in [26] to show a distributional convergence of the residuals between the estimated and true latent positions.

Our work is also influenced by the analysis of the spectra of random graphs [5]. Of special note is the classic paper of [11], in which the authors show that for an Erdős-Rényi graph with parameter p, the appropriately scaled largest eigenvalue of the adjacency matrix converges in law to a normal distribution. Other results of this type are proved for sparse graphs in both the independent edge model [17] and the d-regular random graph model [15]. More recently, general bounds for the operator norm of the difference between the adjacency matrix and its expectation have been proved in [20] and [29].

We would, of course, be remiss not to mention important recent results in random matrix theory. Specifically, [28] and [16] provide central limit theorems for the eigenvectors of a Wigner matrix. In particular, [16] gives a central limit theorem for eigenvectors near the spectral edge; assuming an additional four-moment matching condition, both [28] and [16] exhibit similar results for the bulk of the spectrum. A material difference between these

results and our present work, however, is that we consider random matrices whose entries have nonzero mean. Others have considered positive mean matrices; in particular, it was pointed out to us that in [23], the author examines the spectral radius of non-central random matrices with i.i.d entries. Our setting is different from that in [23], however, and our focus here is on the normalized eigenvectors.

The key step in our work is the application of the power method to the adjacency matrix, with the initial vector being the true latent position. Conditioned on the true latent position, this produces a vector whose components are asymptotically normally distributed. Furthermore, the difference between this vector and the true eigenvector of the adjacency matrix is asymptotically negligible, due to a large gap between the largest eigenvalue and the remaining eigenvalues.

Finally, [30] provides a proof of the asymptotic normality for maximum likelihood estimates of the parameters of a related latent position model, the logistic β-model, and gives a derivation of the associated Fisher information matrix in that case.

2 Central limit theorem for random dot product graphs

In this section, we recall the definition of a finite-dimensional random dot product graph, and we state a central limit theorem for the scaled differences between the estimated and true latent positions of the graph. We begin with the construction of our estimate for the underlying latent positions.

Definition 1 (Random Dot Product Graph (d-dimensional)) Let F be a distribution on a set $\mathcal{X} \subset \Re^d$ satisfying $x^\top y \in [0,1]$ for all $x, y \in \mathcal{X}$. We say $(X, A) \sim \mathrm{RDPG}(F)$ if the following hold. Let $X_1, \ldots, X_n \sim F$ be independent random variables and define

$$X = [X_1, \ldots, X_n]^\top \in \Re^{n \times d} \text{ and } P = XX^\top \in [0,1]^{n \times n}. \tag{1}$$

The X_i are the latent positions for the random graph. The matrix $A \in \{0,1\}^{n \times n}$ is defined to be a symmetric matrix with all zeroes on the diagonal such that for all $i < j$, conditioned on X_i, X_j the A_{ij} are independent and

$$A_{ij} \sim \mathrm{Bernoulli}(X_i^\top X_j), \tag{2}$$

namely,

$$\Pr[A|X] = \prod_{i<j} (X_i^T X_j)^{A_{ij}} (1 - X_i^T X_j)^{(1-A_{ij})} \tag{3}$$

We seek to demonstrate an asymptotically normal estimate of X, the matrix of latent positions $X_1, \ldots, X_n$. However, the model as specified above is non-identifiable: if $W \in \Re^{d \times d}$ is orthogonal, then XW generates the same distribution over adjacency matrices. As a result, we will often consider *uncentered* principal components of X.

Definition 2 Let X and P be as in Definition 1. Then P is symmetric and positive semidefinite and has rank d. Hence, P has an spectral decomposition $P = VSV^\top$ where $V \in \Re^{n \times d}$ has orthonormal columns and S is diagonal with positive decreasing entries along the diagonal. The uncentered principal components of X are then given by $VS^{1/2}$ and $VS^{1/2} = XW_n$ for some random orthogonal matrix $W_n \in \Re^{d \times d}$. We will denote the uncentered principal components of X as $\hat{X}$.

Remark 1 We denote the second moment matrix for X_i by $\Delta = \mathrm{E}(X_i X_i^\top)$. We assume, without loss of generality, that $\Delta = \mathrm{diag}(\lambda_1, \ldots, \lambda_d)$, i.e., Δ is diagonal. This is without loss of generality because if Δ is not diagonal, then we can find an orthogonal matrix which diagonalizes Δ, and we can apply this matrix to the latent positions X without changing the distribution of the random graph.

For the remainder of this work we will also assume, with slight loss of generality, that the eigenvalues of Δ are distinct and positive, so that Δ has a strictly decreasing diagonal; i.e. that $\lambda_1 > \lambda_2 > \cdots > \lambda_d > 0$. This is a mild restriction on the distribution F of the latent positions, and we impose it for technical reasons. Finally, let δ_d denote the following minimum:

$$\delta_d = \min \left\{ \min_{1 \leq i \leq d-1} |\lambda_i - \lambda_{i+1}|, \lambda_d \right\}$$

so that δ_d is the smaller of the minimum eigengap and the dth eigenvalue of Δ.

Our estimate for X, or specifically our estimate for the uncentered principal components of X, is a spectral embedding defined below, which is once again motivated by the observation that A is essentially a noisy version of P. Specifically, if we were to use the procedure below on the matrix P, we would recover the uncentered principal components on X, and this motivates our choice of the spectral embedding of A as our estimate for X.

Definition 3 (Embedding of A) Suppose that A is as in Definition 1. Let $A = U_A S_A U_A^\top$ be the (full) spectral decomposition of A. Then our estimate for the uncentered principal components of X is $\hat{X} = \hat{V}\hat{S}^{1/2}$, where $\hat{S} \in \Re^{d \times d}$ is the diagonal submatrix of S_A with the d largest eigenvalues (in magnitude) of A and $\hat{V} \in \Re^{n \times d}$ is the submatrix of U_A whose orthonormal columns are the corresponding eigenvectors.

We now state the central limit theorem for the scaled differences between the estimated and true latent positions in the finite-dimensional random dot product graph setting. Note, for two vectors $a, b \in \Re^d$, we use the notation $a \leq b$ to indicate that $a_i \leq b_i$ for each $i = 1, \ldots, d$.

Theorem 1 *Let $(X, A) \sim \mathrm{RDPG}(F)$ be a d-dimensional random dot product graph, i.e., F is a distribution for points in $\mathbb{R}^d$, and let $\hat{X}$ be our estimate for X. We assume F has a diagonal second moment matrix with distinct positive entries along the diagonal (see Remark 1). Let $\Phi(z, \Sigma)$ denote the cumulative distribution function for the multivariate normal, with mean zero and covariance matrix Σ, evaluated at z. Then there exists a sequence of orthogonal matrices W_n converging to the identity almost surely such that for each component i and any $z \in \mathbb{R}^d$,*

$$\Pr\left\{ \sqrt{n}(W_n \hat{X}_i - X_i) \leq z \right\} \to \int_{\mathcal{X}} \Phi(z, \Sigma(x)) dF(x) \tag{4}$$

where $\Sigma(x) = \Delta^{-1}\mathrm{E}(X_j X_j^\top (x^\top X_j - (x^\top X_j)^2))\Delta^{-1}$ and $\Delta = \mathrm{E}(X_1 X_1^T)$ is the second moment matrix. That is, the sequence of random variables $\sqrt{n}(W_n \hat{X}_i - X_i)$ converges in distribution to a mixture of multivariate normals. We denote this mixture by $\mathcal{N}(0, \Sigma(X_i))$.

As an immediate consequence of this theorem, we obtain the following corollary for the eigenvectors of an Erdős-Rényi random graph. It was pointed out to us that this result can also be derived from estimates in the classic work of [11]. In the Erdős-Rényi graph, the latent positions X_i have a degenerate distribution; namely, there is some $p \in [0, 1]$ such that $X_i = p^{1/2}$ for all i.

Corollary 1 *For an Erdős-Rényi graph with edge probability p, the following central limit theorem holds:*

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

We also state the following additional corollaries of Theorem 1. Corollary 2 describes the implications of conditioning on the value of the latent position; specifically, Eq. (6) demonstrates that in a stochastic blockmodel, conditioning on the block membership yields convergence of the residuals to the appropriate mixture component. Corollary 3, which follows from the Cràmer-Wold device, allows us to conclude that any fixed, finite collection of residuals is asymptotically independent. We emphasize that Corollary 3 does *not* imply that the complete n-tuple of residuals is asymptotically independent.

Corollary 2 *In the setting of Theorem 1, let $\mathcal{X} = \mathrm{supp}(F) \subset [0,1]^d$ be the support of the distribution of the X_i and suppose that $\mathcal{B} \subset [0,1]^d$ is such that $\Pr[X_i \in \mathcal{B}] > 0$. If we condition on the event $\{X_i \in \mathcal{B}\}$, we obtain*

$$\Pr\Big\{\sqrt{n}(\hat{X}_i - X_i) \leq z \mid X_i \in \mathcal{B}\Big\} \longrightarrow \frac{1}{\Pr\Big(X_i \in \mathcal{B}\Big)} \int_{\mathcal{B}} \Phi(z, \Sigma(x)) dF(x) \tag{5}$$

where $\Sigma(x)$ is as in Theorem 1. If in addition $|\mathcal{X}| = m < \infty$, then for all $x \in \mathcal{X}$

$$\Pr\Big\{\sqrt{n}(\hat{X}_i - x) \leq z \mid X_i = x\Big\} \longrightarrow \Phi(z, \Sigma(x)). \tag{6}$$

Corollary 3 *Suppose X and $\hat{X}$ are as in Theorem 1. Let $K \in \mathbb{N}$ be any fixed positive integer; let $i_1, \ldots, i_K \in \mathbb{N}$ be any fixed set of indices and let $z_1, \cdots, z_K \in \mathbb{R}^d$ be fixed. Then*

$$\lim_{n\to\infty} \Pr\Big[\bigcap_{k=1}^{K} \{\sqrt{n}(\hat{X}_{i_k} - X_{i_k}) \leq z_k\}\Big] = \prod_{k=1}^{K} \int_{\mathcal{X}} \Phi(z_k, \Sigma(x_k)) dF(x_k) \tag{7}$$

where $\Phi(\cdot, \Sigma)$ denotes the cumulative distribution function for a d-variate normal with mean zero and covariance matrix Σ. Again the covariance matrices $\Sigma(x)$ are as in Theorem 1.

2.1 Proof of Theorem 1

The proof of Theorem 1 depends on several supporting lemmas, which we present below. We begin by recalling Definition 2 of the uncentered principal components of $\tilde{X}$ of X, i.e., $\tilde{X}W_n = X$ for some orthogonal matrix W_n. That is, for a given n, there exists an orthogonal matrix W_n such that $W_n^T \tilde{X}_i = X_i$ for all $i = 1, \ldots, n$ where we recall that $X = [X_1, \ldots, X_n]^T$ and $\tilde{X} = [\tilde{X}_1, \ldots, \tilde{X}_n]^T$. Hence we shall prove Theorem 1 by showing that the right hand side of Eq. (4) holds for the scaled difference $\sqrt{n}(\hat{X}_i - \tilde{X}_i)$. To prove Theorem 1, we will need explicit control on the differences, in Frobenius norm, between $\tilde{X}$ and $\hat{X}$, as well as $\hat{V}$ and V. These bounds, given in Lemma 1 below, are proven in [26], [24], and [20].

We note that in Lemma 1 below and onwards, $\|\cdot\|_F$ denotes the Frobenius norm and the unadorned norm $\|\cdot\|$ denotes the operator norm. That is, if $B \in \mathbb{R}^{n\times m}$ is a matrix, then

$$\|B\| = \sup_{X \neq 0} \frac{\|BX\|_F}{\|X\|_F}, X \in \mathbb{R}^n$$

Also, in several of the results that follow, C represents a constant whose particular value may change from line to line.

Lemma 1 *Suppose X, A, $\tilde{X}$ and $\hat{X}$ are as defined above. As before, denote by δ_d the following minimum:*

$$\delta_d = \min\left\{\min_{1\le i\le d-1}|\lambda_i - \lambda_{i+1}|, \lambda_d\right\}$$

where $\lambda_1 > \lambda_2 > \cdots > \lambda_d$ are the ordered, distinct, positive eigenvalues of Δ. Let c be arbitrary. There exists a constant $n_0(c)$ such that if $n > n_0$, then for any η satisfying $n^{-c} < \eta < 1/2$, the following bounds hold with probability greater than $1-\eta$,

$$\|\hat{X} - \tilde{X}\|_F = \|\hat{V}\hat{S}^{1/2} - VS^{1/2}\|_F \le 6\delta_d^{-1}(2d\log(n/\eta))^{1/2}, \tag{8}$$

$$\|\hat{V} - V\|_F \le 6\delta_d^{-1}n^{-1/2}(2d\log(n/\eta))^{1/2}, \tag{9}$$

$$\text{and } \|XX^\top - A\| \le 2(n\log(n/\eta))^{1/2}. \tag{10}$$

Hence, for n sufficiently large, the events above imply that with probability greater than $1-\eta$,

$$\frac{\delta_d n}{2} \le \|S\| \le n, \quad \frac{\delta_d n}{2} \le \|\hat{S}\| \le n. \tag{11}$$

The following result provides a bound on the difference between the d largest eigenvalues of A and the corresponding eigenvalues of P.

Lemma 2 *In the setting of Lemma 1, with probability greater than $1-2\eta$*

$$\|S - \hat{S}\|_F \le C\delta_d^{-2}d\log(n/\eta)$$

Proof First, consider that we can write $\hat{S} = \hat{V}^\top A\hat{V}$ and $S = V^\top PV$. We therefore have

$$\|S - \hat{S}\|_F \le \|V^\top PV - \mathrm{diag}(V^\top AV)\|_F + \|\mathrm{diag}(V^\top AV) - \hat{V}^\top A\hat{V}\|_F, \tag{12}$$

where diag denotes the operation of setting the off-diagonal elements to zero. Since

$$\mathrm{diag}(\hat{V}^\top AV) = \mathrm{diag}(V^\top A\hat{V}),$$

for the second term in the right hand side of Eq. (12), we have

$$\begin{aligned}
\|\mathrm{diag}(V^\top AV) - \hat{V}^\top A\hat{V}\|_F &= \|\mathrm{diag}(V^\top AV - \hat{V}^\top A\hat{V})\|_F \\
&= \left\|\mathrm{diag}((V-\hat{V})^\top A(V-\hat{V}) - 2(\hat{V}-V)^\top A\hat{V})\right\|_F \\
&\le \|\mathrm{diag}\big((V-\hat{V})^T A(V-\hat{V})\big)\|_F + 2\|\mathrm{diag}\big((\hat{V}-V)^\top A\hat{V}\big)\|_F \\
&\le \|\hat{S}\|\|V-\hat{V}\|_F^2 + 2\|\mathrm{diag}\big((\hat{V}-V)^\top\hat{V}\hat{S}\big)\|_F \\
&\le \|\hat{S}\|\|V-\hat{V}\|_F^2 + 2\|\hat{S}\|\|\mathrm{diag}\big((\hat{V}-V)^\top\hat{V}\big)\|_F
\end{aligned}$$

Furthermore, we have

$$\begin{aligned}
\mathrm{diag}(\hat{V}^\top(V-\hat{V})) &= \frac{1}{2}\mathrm{diag}\Big(\hat{V}^\top V - \hat{V}^\top\hat{V} - V^\top V + V^\top\hat{V}\Big) \\
&= \frac{1}{2}\mathrm{diag}\Big((V-\hat{V})^\top(V-\hat{V})\Big),
\end{aligned} \tag{13}$$

and this implies

$$\begin{aligned}\|\mathrm{diag}(V^\top AV) - \hat{V}^\top A\hat{V}\|_F &\le \|\hat{S}\|\|V-\hat{V}\|_F^2 + \|\hat{S}\|\|\mathrm{diag}\big((\hat{V}-V)^\top\hat{V}\big)\|_F \\ &\le C\delta_d^{-2} d\log(n/\eta),\end{aligned}$$

with probability at least $1-\eta$.

Now, let us denote by $V_{\cdot k}$ the kth column of V and by $V_{k\cdot}$ the kth row of V. We chose the convention that matrix operations (such as transposition and inversion) are assumed to be done first, and the indexing of rows or columns last. Thus, for example, $V_{k\cdot}^T$ represents the kth row of V^T. Now, for the first term in the right hand side of Eq. (12), we have

$$\begin{aligned}\|\mathrm{diag}(V^\top(A-P)V)\|_F^2 &= \sum_{k=1}^{d}\Big(V_{k\cdot}^T(A-P)V_{\cdot k}\Big)^2 \\ &= \sum_{k=1}^{d}\Big(\sum_{i<j}2(A_{ij}-P_{ij})V_{ik}V_{jk} - \sum_{i=1}^{n}P_{ii}V_{ik}^2\Big)^2\end{aligned}$$

We thus have

$$\|\mathrm{diag}(V^\top(A-P)V)\|_F \le \sum_{k=1}^{d}\Big(\Big|\sum_{i<j}2(A_{ij}-P_{ij})V_{ik}V_{jk}\Big| + 1\Big) \tag{14}$$

as $\sum_i P_{ii}V_{ik}^2 \le 1$ because V is orthogonal and the entries of P_{ii} are in $[0,1]$. The right hand side in Eq.(14) is again the sum of $n(n-1)/2$ independent random variables and by Hoeffding's inequality

$$\Pr[|\sum_{i<j}2(A_{ij}-P_{ij})V_{ik}V_{jk}| \ge t] \le 2\exp\Big(\frac{-2t^2}{\sum_{i<j}(2V_{ik}V_{jk})^2}\Big) \le 2\exp(-t^2).$$

Hence, the first term in Eq. (12) satisfies

$$\Pr[\|\mathrm{diag}(V^\top(A-P)V)\|_F \le d(1+(\log(2d/\eta))^{1/2})] \ge 1-\eta.$$

Putting together the bounds for the two terms in Eq. (12) yields the result. □

In the following lemma, we prove that $V^\top\hat{V}$ is very close to the identity matrix.

Lemma 3 *In the setting of Lemma 1, with probability greater than* $1-2\eta$*,*

$$\|V^\top\hat{V} - I\|_F \le \frac{Cd\log(n/\eta)}{\delta_d^2 n}. \tag{15}$$

Proof The diagonal entries of $V^\top\hat{V}-I$ can be bounded by Eq. (13) to yield

$$\|\mathrm{diag}((V-\hat{V})\hat{V})\|_F \le \frac{1}{2}\|V-\hat{V}\|_F^2 \le \frac{Cd\log(n/\eta)}{\delta_d^2 n}$$

with probability at least $1-\eta$. To bound the off-diagonal terms, we adapt a proof from [22]. First, $V^\top(A-P)\hat{V} = V^\top\hat{V}\hat{S} - SV^\top\hat{V}$. The ijth entry of $V^\top(A-P)\hat{V}$ can be written as

$$V_{i\cdot}^\top(A-P)\hat{V}_{\cdot j} = (S_{ii}-\hat{S}_{jj})V_{i\cdot}^\top\hat{V}_{\cdot j}. \tag{16}$$

Because the eigenvalues are distinct and $\|A - P\| \leq 2(n \log(n/\eta))^{1/2}$ with probability at least $1 - \eta$, we know for $i \neq j$ that

$$|S_{ii} - \hat{S}_{jj}| \geq |S_{ii} - S_{jj}| - \|A - P\| \geq \delta_d n - 2(n \log(n/\eta))^{1/2} \geq \delta_d n/2$$

for sufficiently large n with probability at least $1 - \eta$. We also have that

$$V_{i\cdot}^\top (A - P)\hat{V}_{\cdot j} = V_{i\cdot}^\top (A - P)V_{\cdot j} + V_{i\cdot}^\top (A - P)(V_{\cdot j} - \hat{V}_{\cdot j})$$

A similar argument to the one following Eq. (14) shows that $|V_{i\cdot}^\top (A-P)V_{\cdot j}| \leq C(\log(d/\eta))^{1/2}$ and an application of the the Cauchy-Schwarz inequality and Eq. (9) yields

$$|V_{i\cdot}^\top (A - P)(V_{\cdot j} - \hat{V}_{\cdot j})| \leq C\delta_d^{-1} \log(n/\eta).$$

Dividing through by $S_{ii} - \hat{S}_{jj}$ in Eq. 16 gives

$$V_{i\cdot}^\top \hat{V}_{\cdot j} = \frac{V_{i\cdot}^\top (A - P)\hat{V}_{\cdot j}}{(S_{ii} - \hat{S}_{jj})} \leq \frac{C(\log(d/\eta))^{1/2} + C\delta_d^{-1} \log(n/\eta)}{\delta_d n/2} \leq \frac{C \log(n/\eta)}{\delta_d^2 n}$$

Eq. (14) follows from this and the bound for the diagonal terms. □

We now complete the proof of Theorem 1. Let $AVS^{-1/2} = [Y_1, \cdots, Y_n]^\top$ where $Y_i \in \mathbb{R}^d$. We first show that $\sqrt{n}(\tilde{X}_i - Y_i)$ is multivariate normal in the limit. We have

$$\begin{aligned}
\sqrt{n}(\tilde{X}_i - Y_i) &= \sqrt{n}((P\tilde{X}S^{-1})_{\cdot i}^T - (A\tilde{X}S^{-1})_{\cdot i}^T) \\
&= nS^{-1}W_n\Big(\frac{1}{\sqrt{n}} \sum_{j=1}^{n} (P_{ij} - A_{ij})X_j\Big) \\
&= nS^{-1}W_n\Big(\frac{1}{\sqrt{n}} \sum_{j \neq i} (A_{ij} - X_i^T X_j)X_j - \frac{X_i^T X_i}{\sqrt{n}} X_i\Big)
\end{aligned}$$

Conditional on $X_i = x_i$, the scaled sum

$$\frac{1}{\sqrt{n}} \sum_{j \neq i} (A_{ij} - X_i^T X_j)X_j$$

is a sum of i.i.d, mean zero random variables with covariance $\tilde{\Sigma}(x_i)$. Therefore, by the classical multivariate central limit theorem, we have

$$\Big(\frac{1}{\sqrt{n}} \sum_{j \neq i} (A_{ij} - x_i^T X_j)X_j - \frac{x_i^T x_i}{\sqrt{n}} x_i\Big) \longrightarrow \mathcal{N}(0, \tilde{\Sigma}(x_i))$$

The strong law of large numbers ensures that $nS^{-1} \to \Delta^{-1} = (\mathrm{E}(X_1 X_1^T))^{-1}$ almost surely. In addition,

$$\frac{1}{n} X^T X = (\tilde{X}W_n^T)^T (\tilde{X}W_n^T) W_n \tilde{X}^T \tilde{X} W_n^T = W_n (S/n) W_n^T \to W_n \Delta W_n^T$$

almost surely. We thus have $W_n \to I$ almost surely. Hence, by Slutsky's theorem in the multivariate setting, we have, conditional on $X_i = x_i$, that

$$\sqrt{n}(\tilde{X}_i - Y_i) \longrightarrow \mathcal{N}(0, \Delta^{-1} \tilde{\Sigma}(x_i) \Delta^{-1}) = \mathcal{N}(0, \Sigma(x_i))$$

Let $\tilde{Y}_i = \hat{S}^{-1/2}(AV)^T_{\cdot i}$ (contrast this with $Y_i = S^{-1/2}(AV)^T_{\cdot i}$. We then have

$$\begin{aligned}\sqrt{n}\|\tilde{Y}_i - Y_i\| &= \|\sqrt{n}(S^{-1/2}\hat{S}^{-1/2} - S^{-1})(A\tilde{X})^T_{\cdot i}\| \\ &\leq \|\sqrt{n}S^{-1/2}\|\|A\|\|\hat{S}^{-1/2} - S^{-1/2}\|\end{aligned}$$

We use the notation $Z = \tilde{O}_{\Pr}(f(n))$ to denote that Z is, with high probability, bounded by $f(n)$ times some multiplicative factor that does not depend on n. Since $\|\hat{S} - S\| = \tilde{O}_{\Pr}(\log n)$, i.e., $\|\hat{S} - S\|$ is bounded, with high probability, by a logarithmic function of n times some factor depending only on δ_d and d, we have

$$\|\hat{S}^{-1/2} - S^{-1/2}\| = \tilde{O}_{\Pr}(n^{-3/2}\log n).$$

We thus have

$$\sqrt{n}\|\tilde{Y}_i - Y_i\| \leq \sqrt{n}\|S^{-1/2}\|\|A\|\|\hat{S} - S\| = \tilde{O}_{\Pr}(n^{-1/2}\log n).$$

That is to say, $\sqrt{n}(\tilde{Y}_i - Y_i)$ converges to 0 in probability. Finally, we derive a bound for $\sqrt{n}(\tilde{Y}_i - \hat{X}_i)$. By Markov's inequality

$$\Pr[\sqrt{n}\|\tilde{Y}_i - \hat{X}_i\| \geq \epsilon] \leq \frac{\mathrm{E}(n\|\tilde{Y}_i - \hat{X}_i\|^2)}{\epsilon^2} = \frac{\mathrm{E}(\|AV\hat{S}^{-1/2} - A\hat{V}\hat{S}^{-1/2}\|_F^2)}{\epsilon^2}$$

where the equality holds because the exchangeability of the graph guarantees that the n rows of the matrix on the right hand side are identically distributed to $\tilde{Y}_i - \hat{X}_i$.

Let $B = A - \hat{V}\hat{S}\hat{V}^T$. We then have

$$\begin{aligned}\|A\hat{V}\hat{S}^{-1/2} - AV\hat{S}^{-1/2}\|_F &= \|(\hat{V}\hat{S}\hat{V}^T + B)(\hat{V} - V)\hat{S}^{-1/2}\|_F \\ &\leq \|\hat{V}\hat{S}\hat{V}^T(\hat{V} - V)\hat{S}^{-1/2}\|_F + \|B(\hat{V} - V)\hat{S}^{-1/2}\|_F \\ &\leq \|\hat{S}\|\|\hat{V}^TV - I\|_F\|\hat{S}^{-1/2}\| + \|B\|\|\hat{V} - V\|_F\|\hat{S}^{-1/2}\|\end{aligned}$$

By Lemma 3 and bounds for the spectral norm of S and E, we have, with probability at least $1 - 2\eta$, that

$$\|A\hat{V}\hat{S}^{-1/2} - AV\hat{S}^{-1/2}\|_F \leq Cd\delta_d^{-5/2}n^{-1/2}\log(n/\eta)$$

On the other hand, $\|A\hat{V}\hat{S}^{-1/2} - AV\hat{S}^{-1/2}\|_F$ is at most of order $\sqrt{n}$. As η is arbitrary, we thus have,

$$\frac{\mathrm{E}(\|AV\hat{S}^{-1/2} - A\hat{V}\hat{S}^{-1/2}\|_F^2)}{\epsilon^2} \leq \frac{C\log^2 n}{\epsilon^2\delta_d^5 n}$$

which converges to 0 for any fixed $\epsilon > 0$ as $n \to \infty$. Hence, $\sqrt{n}(\tilde{Y}_i - \hat{X}_i)$ also converges to 0 in probability. The above arguments can be combined to show that, conditional on $X_i = x_i$,

$$\sqrt{n}(\hat{X}_i - \tilde{X}_i) = \sqrt{n}(Y_i - \tilde{X}_i) + \sqrt{n}(\tilde{Y}_i - Y_i) + \sqrt{n}(\hat{X}_i - \tilde{Y}_i) \longrightarrow \mathcal{N}(0, \Sigma(x_i)).$$

Finally, by an application of Slutsky's theorem, conditional on $X_i = x_i$, we conclude that

$$\sqrt{n}(W_n\hat{X}_i - X_i) = W_n\sqrt{n}(\hat{X}_i - \tilde{X}_i) \to \mathcal{N}(0, \Sigma(x_i)).$$

Eq. (4) then follows by integrating over all the possible realizations of X_i and applying the dominated convergence theorem.

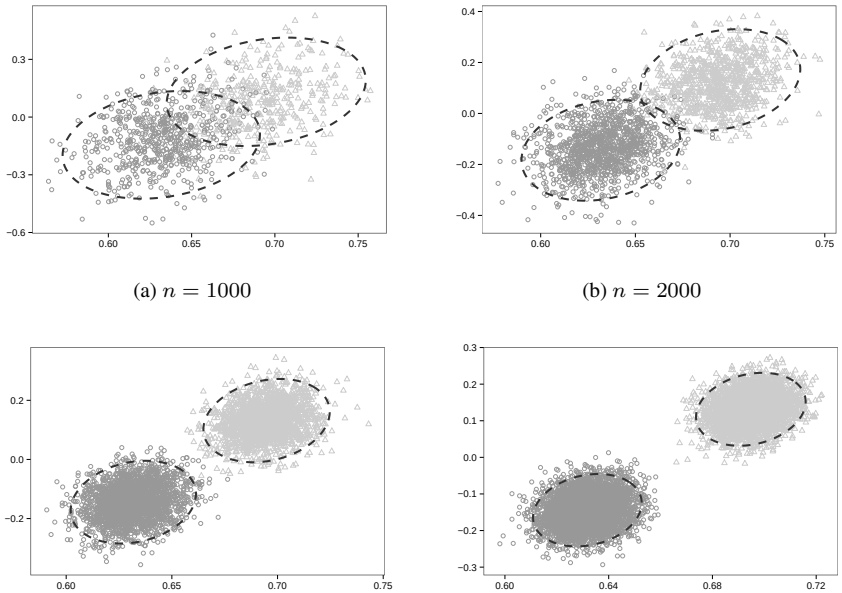
3 Simulations

To illustrate Theorem 1, we consider random graphs generated according to a stochastic block model with parameters

$$B = \begin{bmatrix} 0.42 & 0.42 \\ 0.42 & 0.5 \end{bmatrix} \quad \text{and} \quad \pi = (0.6, 0.4). \tag{17}$$

In this model, each node is either in block 1 (with probability 0.6) or block 2 (with probability 0.4). Adjacency probabilities are determined by the entries in B based on the block memberships of the incident vertices. The above stochastic blockmodel corresponds to a random dot product graph model in $\mathbb{R}^2$ where the distribution F of the latent positions is a mixture of point masses located at $x_1 \approx (0.63, -0.14)$ (with prior probability 0.6) and $x_2 \approx (0.69, 0.13)$ (with prior probability 0.4).

We sample an adjacency matrix A for graphs on n vertices from the above model for various choices of n. For each graph G, let $\hat{X} \in \Re^{n \times 2}$ denote the embedding of A and let $\hat{X}_i$ denote the ith row of $\hat{X}$. In Figure 1, we plot the n rows of $\hat{X}$ for the various choices of n. The points are denoted by symbols according to the block membership of the corresponding vertex in the stochastic blockmodel. The ellipses show the 95% level curves for the distribution of $\hat{X}_i$ for each block as specified by the limiting distribution.



(a) $n = 1000$ (b) $n = 2000$

(c) $n = 4000$ (d) $n = 8000$

Fig. 1: Plot of the estimated latent positions in a two-block stochastic blockmodel graph on n vertices. The points are denoted by symbols according to the blockmembership of the corresponding vertices. Dashed ellipses give the 95% level curves for the distributions as specified in Theorem 1.

We then estimate the covariance matrices for the residuals. The theoretical covariance matrices are given in the last column of Table 1, where Σ_1 and Σ_2 are the covariance matrices for the residual $\sqrt{n}(\hat{X}_i - X_i)$ when X_i is from the first block and second block, respectively. The empirical covariance matrices, denoted $\hat{\Sigma}_1$ and $\hat{\Sigma}_2$, are computed by evaluating the sample covariance of the rows of $\sqrt{n}\hat{X}_i$ corresponding to vertices in block 1 and 2 respectively. The estimates of the covariance matrices are given in Table 1. We see that as n increases, the sample covariances tend toward the specified limiting covariance matrix given in the last column.

n	2000	4000	8000	16000	∞
$\hat{\Sigma}_1$	$\begin{bmatrix} .58 & .54 \\ .54 & 16.56 \end{bmatrix}$	$\begin{bmatrix} .58 & .63 \\ .63 & 14.87 \end{bmatrix}$	$\begin{bmatrix} .60 & .61 \\ .61 & 14.20 \end{bmatrix}$	$\begin{bmatrix} .59 & .58 \\ .58 & 13.96 \end{bmatrix}$	$\begin{bmatrix} .59 & .55 \\ .55 & 13.07 \end{bmatrix}$
$\hat{\Sigma}_2$	$\begin{bmatrix} .58 & .75 \\ .75 & 16.28 \end{bmatrix}$	$\begin{bmatrix} .59 & .71 \\ .71 & 15.79 \end{bmatrix}$	$\begin{bmatrix} .58 & .54 \\ .54 & 14.23 \end{bmatrix}$	$\begin{bmatrix} .61 & .69 \\ .69 & 13.92 \end{bmatrix}$	$\begin{bmatrix} .60 & .59 \\ .59 & 13.26 \end{bmatrix}$

Table 1: The sample covariance matrices for $\sqrt{n}(\hat{X}_i - X_i)$ for each block in a stochastic blockmodel with two blocks. Here $n \in \{2000, 4000, 8000, 16000\}$. The last column are the theoretical covariance matrices for the limiting distribution.

We also investigate the effects of the multivariate normal distribution as specified in Theorem 1 on inference procedures. It is shown in [25, 26] that the approach of embedding a graph into some Euclidean space, followed by inference (for example, clustering or classification) in that space can be consistent. However, these consistency results are, in a sense, only first-order results. In particular, they demonstrate only that the error of the inference procedure converges to 0 as the number of vertices in the graph increases. We now illustrate how Theorem 1 may lead to a more refined error analysis.

We construct a sequence of random graphs on n vertices, where n ranges from 1000 through 4000 in increments of 250, following the stochastic blockmodel with parameters as given above in Eq. (17). For each graph G_n on n vertices, we embed G_n and cluster the embedded vertices of G_n via Gaussian mixture model and K-Means. Gaussian mixture model-based clustering was done using the MCLUST implementation of [10]. We then measure the classification error of the clustering solution. We repeat this procedure 100 times to obtain an estimate of the misclassification rate. The results are plotted in Figure 2. For comparison, we plot the Bayes optimal classification error rate under the assumption that the embedded points do indeed follow a multivariate normal mixture with covariance matrices Σ_1 and Σ_2 as given in the last column of Table 1. We also plot the misclassification rate of $(C \log n)/n$ as given in [25] where the constant C was chosen to match the misclassification rate of K-means clustering for $n = 1000$. For the number of vertices considered here, the upper bound for the constant C from [25] will give a vacuous upper bound of the order of 10^6 for the misclassification rate in this example.

4 Discussion

Our demonstration of the clustering accuracy in § 3 shows how our Theorem 1 may impact statistical inference for random graphs. First, we see that the empirical error rates are much lower than those proved in previous work on spectral methods [21, 25, 8]. Indeed, for both

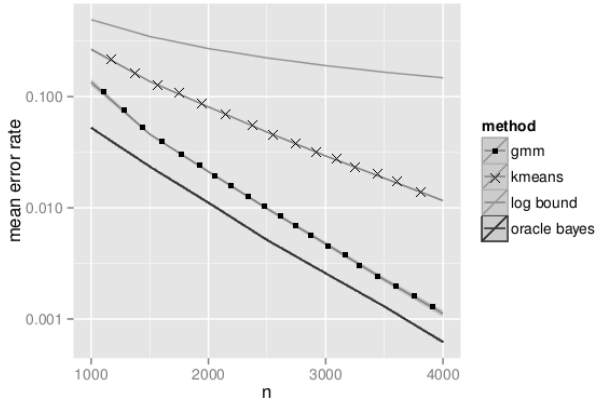


Fig. 2: Comparison of classification error for Gaussian mixture model (red curve), K-Means (green curve), and Bayes classifier (cyan curve). The classification errors for each $n \in \{1000, 1250, 1500, \ldots, 4000\}$ were obtained by averaging 100 Monte Carlo iterations and are plotted on a $\log_{10}$ scale. The plot indicates that the assumption of a mixture of multivariate normals can yield non-negligible improvement in the accuracy of the inference procedure. The log-bound curve (purple) shows an upper bound on the error rate as derived in [25].

the K-Means algorithm and Gaussian mixture model, the average clustering error decreases at an exponential rate as opposed to the $(\log n)/n$ bounds shown in previous work. Furthermore, the rate of decrease for Gaussian mixture model-based clustering closely mirrors the Bayes optimal error rate that would be achieved if the estimated latent positions were exactly distributed according to the multivariate normal distribution and the parameters of this distribution were known.

These results suggest that further investigations using our theorem could lead to more accurate bounds on the empirical error rates for adjacency spectral clustering. We believe that extending Corollary 3 to the case in which the number of residuals we consider, K, is growing with n, e.g., $K = n^{\epsilon}$ for some $\epsilon > 0$, and further work regarding distributions of spectral statistics for stochastic blockmodels will lead to foundational statistical procedures analogous to the results on estimation, hypothesis testing, and clustering in the setting of mixtures of normal distributions in Euclidean space. The relatively simple nature of our spectral procedure allows for computationally efficient statistical methodology.

Extensions of this work to a wider class of exchangeable graphs are also of interest. Though not all exchangeable random graphs can be represented as random dot product graphs, random dot product graphs can approximate any exchangeable graph in the following sense: given a sufficiently regular link function, there exists a *feature map* from the original latent position space to ℓ_2, such that the link function applied to the original latent positions is equal to the inner product applied to the feature-mapped positions in ℓ_2. In [27], the authors argue that by increasing the dimension of the estimated latent positions, it is possible to estimate these feature-mapped latent positions in a way that allows for consistent subsequent inference. Though this larger class of models is not considered here, we believe this is strong motivation to study the random dot product graph model and its eigenvalues and eigenvectors.

Acknowledgements

This work was supported in part by the National Security Science and Engineering Faculty Fellowship, Johns Hopkins University Human Language Technology Center of Excellence, and the XDATA program of the Defense Advanced Research Projects Agency. The authors thank the anonymous referees whose comments and suggestions greatly improved the manuscript.

References

1. Aldous, D.J.: Representations for partially exchangeable arrays of random variables. J. Multivariate. Anal. **11**, 581–598 (1981)
2. Bickel, P.J., Chen, A.: A nonparametric view of network models and Newman-Girvan and other modularities. Proc. Natl. Acad. Sci. USA **106**, 21,068–73 (2009)
3. Bickel, P.J., Chen, A., Levina, E.: The method of moments and degree distributions for network models. Ann. Statist. **39**, 38–59 (2011)
4. Choi, D.S., Wolfe, P.J., Airoldi, E.M.: Stochastic blockmodels with a growing number of classes. Biometrika **99**, 273–284 (2012)
5. Chung, F.R.K.: Spectral Graph Theory. American Mathematical Society (1997)
6. Diaconis, P., Janson, S.: Graph limits and exchangeable random graphs. Rend. Mat. Appl. **28**, 33–61 (2008)
7. Fiedler, M.: Algebraic connectivity of graphs. Czechoslovak Math. J. **23**, 298–305 (1973)
8. Fishkind, D.E., Sussman, D.L., Tang, M., Vogelstein, J.T., Priebe, C.E.: Consistent adjacency-spectral partitioning for the stochastic block model when the model parameters are unknown. SIAM J. Matrix Anal. Appl. **34**, 23–39 (2013)
9. Fortunato, S.: Community detection in graphs. Phys. Rep. **486**, 75–174 (2010)
10. Fraley, C., Raftery, A.E.: MCLUST: Software for model-based cluster analysis. J. Classification **16**, 297–306 (1999)
11. Füredi, Z., Komlós, J.: The eigenvalues of random symmetric matrices. Combinatorica **1**, 233–241 (1981)
12. Goldenberg, A., Zheng, A.X., Fienberg, S.E., Airoldi, E.M.: A survey of statistical network models. Foundations and Trends® in Machine Learning **2**, 129–233 (2010)
13. Hoff, P.D., Raftery, A.E., Handcock, M.S.: Latent space approaches to social network analysis. J. Am. Statist. Assoc. **97**, 1090–1098 (2002)
14. Hoover, D.N.: Relations on probability spaces and arrays of random variables. Tech. rep., Institute for Advanced Study (1979)
15. Janson, S.: The first eigenvalue of random graphs. Comb. Probab. Comput. **14**, 815–828 (2005)
16. Knowles, A., Yin, J.: Eigenvector distribution of Wigner matrices. Probab. Theory Related Fields pp. 1–40 (2011)
17. Krivelevich, M., Sudakov, B.: The largest eigenvalue of sparse random graphs. Comb. Probab. Comput. **12**, 61–72 (2003)
18. Luxburg, U.V.: A tutorial on spectral clustering. Statistics and computing **17**, 395–416 (2007)
19. Marchette, D., Priebe, C.E., Coppersmith, G.: Vertex nomination via attributed random dot product graphs. In: Proceedings of the 57th ISI World Statistics Congress (2011)
20. Oliveira, R.I.: Concentration of the adjacency matrix and of the Laplacian in random graphs with independent edges. arXiv preprint (2010)
21. Rohe, K., Chatterjee, S., Yu, B.: Spectral clustering and the high-dimensional stochastic blockmodel. Ann. Statist. **39**, 1878–1915 (2011)
22. Sarkar, P., Bickel, P.J.: Role of normalization in spectral clustering for stochastic blockmodels. Ann. Statist. **43**, 962–990 (2015)
23. Silverstein, J.W.: The spectral radii and norms of large-dimensional non-central random matrices. Comm. Statist. Stochastic Models **10**(3), 525–532 (1994)
24. Sussman, D.L.: Foundations of Adjacency Spectral Embedding. PhD Thesis, Johns Hopkins University. (2014)
25. Sussman, D.L., Tang, M., Fishkind, D.E., Priebe, C.E.: A consistent adjacency spectral embedding for stochastic blockmodel graphs. J. Am. Statist. Assoc. **107**, 1119–1128 (2012)
26. Sussman, D.L., Tang, M., Priebe, C.E.: Consistent latent position estimation and vertex classification for random dot product graphs. IEEE T. Pattern. Anal. **36**, 48–57 (2014)

27. Tang, M., Sussman, D.L., Priebe, C.E.: Universally consistent vertex classification for latent position graphs. Ann. Statist. **41**, 1406 – 1430 (2013)
28. Tao, T., Vu, V.: Random matrices: Universal properties of eigenvectors. Random Matrices: Theory and Applications **1** (2012)
29. Tropp, J.A.: Freedmans inequality for matrix martingales. Electron. Commun. Probab. **16**, 262–270 (2011)
30. Yan, T., Xu, J.: A central limit thereom in the β-model for undirected random graphs with a diverging number of vertices. Biometrika **100**, 519–524 (2013)
31. Young, S., Scheinerman, E.: Random dot product graph models for social networks. In: Algorithms and models for the web-graph, pp. 138–149. Springer (2007)