A Central Limit Theorem for an Omnibus Embedding of Random Dot Product Graphs

Keith Levin and Avanti Athreya and Minh Tang and Vince Lyzinski and Carey E. Priebe
Department of Applied Mathematics and Statistics
Johns Hopkins University
July 19, 2017

Abstract

Performing statistical inference on collections of graphs is of import to many disciplines. Graph embedding, in which the vertices of a graph are mapped to vectors in a low-dimensional Euclidean space, has gained traction as a basic tool for graph analysis. In this paper, we describe an omnibus embedding in which multiple graphs on the same vertex set are jointly embedded into a single space with a distinct representation for each graph. We prove a central limit theorem for this omnibus embedding, and we show that this simultaneous embedding into a common space allows comparison of graphs without the need to perform pairwise alignments of graph embeddings. Experimental results demonstrate that the omnibus embedding improves upon existing methods, allowing better power in multiple-graph hypothesis testing and yielding better estimation in a latent position model.

Keywords: graph inference, graph embedding, multiple-graph hypothesis testing

*The authors gratefully acknowledge the support of DARPA grants FA8750-12-2-0303, N66001-14-1-4028 and N66001-15-C-4041
1 Introduction

Statistical inference across multiple graphs is of vital interdisciplinary interest in domains as varied as machine learning, neuroscience, and epidemiology, to name but a few. Inference on random graphs frequently depends on appropriate low-dimensional Euclidean representations of the vertices of these graphs, known as graph embeddings, typically given by spectral decompositions of adjacency or Laplacian matrices (see, for example, Belkin and Niyogi, 2003; Sussman et al., 2012a). However, for inference tasks involving multiple graphs—for instance, determining whether two graphs on the same vertex set are similar—it is unclear how to apply such an embedding to all graphs simultaneously in such a way that individual graph structures are preserved. In this paper, we describe an omnibus embedding, in which the adjacency matrices of multiple graphs on the same vertex set are jointly embedded into a single space with a distinct representation for each graph. This simultaneous embedding into a shared space allows comparison of graphs without the need to perform pairwise alignments of graph embeddings. Further, a distinct representation of each graph renders the omnibus embedding especially useful for subsequent comparative graph inference.

The main theoretical results of this paper are a consistency theorem for the omnibus embedding, akin to Lyzinski et al. (2014), and a central limit theorem, akin to Athreya et al. (2016), for the distribution of any finite collection of rows of this omnibus embedding. We emphasize that distributional results for spectral decompositions of random graphs are comparatively few. The classic results of Füredi and Komlós (1981) describe the eigenvalues of the Erdős-Rényi random graph and the work of Tao and Vu (2012) concerns distributions of eigenvectors of more general random matrices under moment restrictions, but Athreya et al. (2016) and Tang and Priebe (2016) are among the only references for central limit theorems for spectral decompositions of adjacency and Laplacian matrices for a class of independent-edge random graphs far broader than the Erdős-Rényi model.

In our consistency result, we show that the omnibus embedding provides consistent estimates of certain unobserved vectors, called latent positions, that are associated to vertices of the graphs. At present, the best available spectral estimates of such latent positions involve averaging across graphs followed by an embedding, resulting in a single set of estimated latent positions, rather than one set for each graph. We find, in simulations, that our omnibus-derived estimates perform competitively with these spectral estimates of the latent positions, while still retaining graph-specific information. In addition, we show that the omnibus embedding allows for a test statistic that improves on the state-of-the-art two-sample test procedure presented in Tang et al. (2016a) for determining whether two random dot product graphs (Young and Scheinerman, 2007) have the same latent positions.

Multiple-graph inference is a nascent field, and the existing literature on even two-sample graph inference is scarce. We point to Tang et al. (2016a) and Tang et al. (2017), which provide theoretical and empirical results for testing whether two random dot product graphs are statistically similar or not, as among very few principled methodologies currently available for testing network similarity. The first of these papers on two-sample graph testing, Tang et al. (2016a), introduces a test statistic generated by performing a Euclidean, lower-dimensional embedding of the adjacency matrix (see Sussman et al. 2012a) of each of the two networks,
followed by a Procrustes alignment \cite{Gower1975, Dryden1998} of the two embeddings. The second of the papers on two-sample graph testing, Tang et al.\cite{Tang2017}, uses the embeddings of each graph to estimate associated density functions. We find, however, that Procrustes alignment both complicates our test statistic and, empirically, weakens the power of the test (see Section \ref{sec:power}). Furthermore, it is unclear how to effectively adapt pairwise Procrustes alignments to tests involving more than two graphs. The omnibus embedding allows us to eliminate these inferentially suboptimal subspace alignments altogether, and provides a multiple-graph representation that is well-suited to latent position estimation and comparative graph inference methods, from two-sample hypothesis testing to joint graph clustering (see, for example, an earlier connectomic analysis in Chen et al.\cite{Chen2016}).

\section{Background and notation}

The present work focuses on \textit{random dot product graphs} (RDPG \cite{Young2007}, independent-edge random graphs in which each vertex $i$ has an associated unobserved vector $X_i \in \mathbb{R}^d$, called the \textit{latent position} or \textit{latent vector}. The probability $p_{ij}$ of an edge joining vertices $i$ and $j$ is simply the dot product of their associated latent positions, independent of all other edges in the graph. For an RDPG with $n$ vertices, the $n \times d$ matrix of latent positions $X$ is formed by taking vector $X_i$ associated to vertex $i$ to be the $i$-th row of $X$. Then $P = [p_{ij}]$, the matrix of probabilities of edges between vertices, is easily expressed as $P = XX^T$. Given such a model, a natural inference task is that of estimating the latent position matrix $X$.

Because of the assumption that the matrix $P$ is of comparatively low rank, random dot product graphs can be analyzed with a number of tools from classical linear algebra, such as singular-value decompositions of their adjacency matrices. Nevertheless, this tractability does not compromise the utility of the model. Random dot product graphs are flexible enough to approximate a wide class of independent-edge random graphs \cite{Tang2013}, including the stochastic block model \cite{Holland1983, Karrer2011}. Under mild assumptions, the adjacency matrix $A$ of a random dot product graph is a rough approximation of the matrix $P = [p_{ij}]$ of edge probabilities in the sense that the spectral norm of $A - P$ can be controlled; see for example Oliveira\cite{Oliveira2009} and Lu and Peng\cite{Lu2013}. In Sussman et al.\cite{Sussman2012a, Sussman2012b, Lyzinski2014}, it is established that, under eigengap assumptions on $P$, a partial spectral decomposition of the adjacency matrix $A$, known as the \textit{adjacency spectral embedding} (ASE), allows for consistent estimation of the true, unobserved latent positions $X$. That is, the rows of the truncated eigendecomposition of $A$ are consistent estimates $\{\hat{X}_i\}$ of the latent positions $\{X_i\}$, which are often the parameters we wish to estimate. In Lyzinski et al.\cite{Lyzinski2017}, it is shown that embedding the adjacency matrix and then performing a novel angle-based clustering of the rows is key to decomposing large, hierarchical networks into structurally similar subcommunities. In Athreya et al.\cite{Athreya2016}, it is shown that the suitably-scaled eigenvectors of the adjacency matrix converge in distribution to a Gaussian mixture. In this paper, we prove a similar result for an omnibus matrix generated from multiple independent graphs.
Since the ASE provides a consistent estimate for the true latent positions in a random dot product graph, a Procrustes distance between the adjacency spectral embedding of two graphs on the same vertex set serves as a test statistic for determining whether two random dot product graphs have the same latent positions (Tang et al., 2016a). Specifically, let \( A^{(1)} \) and \( A^{(2)} \) be the adjacency matrices of two random dot product graphs on the same vertex set (with known vertex correspondence), and let \( \hat{X} \) and \( \hat{Y} \) be their respective adjacency spectral embeddings. If the two graphs have the same generating \( P \) matrices, it is reasonable to surmise that
\[
\min_{W \in O_{d \times d}} \| \hat{X} - \hat{Y}W \|_F
\]
will be relatively small. Observe that the minimum over orthogonal matrices \( W \)—that is, the Procrustes distance between the embeddings \( \hat{X} \) and \( \hat{Y} \)—is taken because of the non-identifiability arising from the fact that if \( W \) is a unitary \( d \times d \) matrix, then the matrix \( XW \) generates the same matrix of probabilities \( P \) as does \( X \). In Tang et al. (2016a), the authors show that a scaled version of the Procrustes distance in (1) provides a valid and consistent test for the equality of latent positions for a pair of random dot product graphs. Unfortunately, the fact that a Procrustes fit must be performed both complicates the test statistic and compromises its power. We demonstrate here (see Section 4) that it is possible instead to perform a \( d \)-dimensional singular value decomposition of the \( 2n \times 2n \) omnibus matrix of the form
\[
\begin{bmatrix}
A^{(1)} & A^{(1)} + A^{(2)} \\
A^{(1) + A^{(2)}} & A^{(2)}
\end{bmatrix}
\]
and consider only the Frobenius norm of the difference between the matrices defined by, respectively, the first \( n \) and the second \( n \) rows of this decomposition. This matrix difference, without any further Procrustes fit, also serves as a test statistic for the equality of latent positions, and we find that it yields an improvement in power.

### 2.1 Notation and Definitions

We begin by establishing notation. For a positive integer \( n \), we let \( [n] = \{1, 2, \ldots, n\} \), and denote the identity matrix and zero matrix by, respectively, \( I \) and \( 0 \). For an \( n \times n \) matrix \( H \), we let \( \lambda_i(H) \) denote the \( i \)-th eigenvalue of \( H \). We let \( \sigma_i(H) \) denote the \( i \)-th singular value of \( H \). We let \( J \) denote the square matrix of all ones. We use \( \otimes \) to denote the Kronecker product. For a vector \( v \), we let \( \| v \| \) denote the Euclidean norm of \( v \). For a matrix \( H \in \mathbb{R}^{n_1 \times n_2} \), we denote by \( H_j \) the column vector formed by the \( j \)-th column of \( H \), and let \( H_i \) denote the row vector formed by the \( i \)-th row of \( H \). For ease of notation, we let \( H_i \in \mathbb{R}^{n_2} \) denote the column vector formed by transposing the \( i \)-th row of \( H \). That is, \( H_i = (H_i)^T \). We let \( \| H \| \) denote the spectral norm of \( H \) and \( \| H \|_F \) denote the Frobenius norm of \( H \). We let \( \| H \|_{2 \to \infty} \) denote the maximum of the Euclidean norms of the rows of \( H \), i.e. \( \| H \|_{2 \to \infty} = \max_i \| H_i \| \). We let \( \Pi_n \) denote the set of all \( n \)-by-\( n \) permutation matrices. Where there is no danger of confusion, we will often refer to a graph \( G \) and its adjacency matrix \( A \) interchangeably. Throughout, we will use \( C > 0 \) to denote a constant, not depending on \( n \), which may vary from one line.
to another. For an event \( E \), we denote its complement by \( E^c \). Given a sequence of events \( \{ E_n \} \), we say that \( E_n \) occurs with high probability, and write \( E_n \) w.h.p., if \( \Pr[E^c_n] \leq Cn^{-2} \) for \( n \) sufficiently large. We note that \( E_n \) w.h.p. implies, by the Borel-Cantelli Lemma, that with probability 1 there exists an \( n_0 \) such that \( E_n \) holds for all \( n \geq n_0 \).

Our focus here is on \( d \)-dimensional random dot product graphs, for which the edge connection probabilities arise as inner products between vectors, called latent positions, that are associated to the vertices. Therefore, we define an an inner product distribution as a probability distribution over a suitable subset of \( \mathbb{R}^d \), as follows:

**Definition 1.** \((d\text{-dimensional Inner Product Distribution})\) Let \( F \) be a probability distribution on \( \mathbb{R}^d \). We say that \( F \) is a \( d \)-dimensional inner product distribution on \( \mathbb{R}^d \) if for all \( x, y \in \text{supp} \ F \), we have \( x^T y \in [0, 1] \).

Next, we define a random dot product graph as an independent-edge random graph for which the edge probabilities are given by the dot products of the latent positions associated to the vertices. We restrict our attention here to graphs that are undirected and in which no vertex has an edge to itself.

**Definition 2.** \((Random Dot Product Graph)\) Let \( F \) be a \( d \)-dimensional inner product distribution with \( X_1, X_2, \ldots, X_n \) i.i.d. \( \sim F \), collected in the rows of the matrix \( X = [X_1, X_2, \ldots, X_n]^T \in \mathbb{R}^{n \times d} \). Suppose \( A \) is a random adjacency matrix given by

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

We then write \((A, X) \sim \text{RDPG}(F, n)\) and say that \( A \) is the adjacency matrix of a random dot product graph with latent positions given by the rows of \( X \).

We see that, given \( X \), the probability \( p_{ij} \) of observing an edge between vertex \( i \) and vertex \( j \) is simply \( X_i^T X_j \), the dot product of the associated latent positions \( X_i \) and \( X_j \). We define the matrix \( P = [p_{ij}] \) of such probabilities by \( P = XX^T \). We will also write \( A \sim \text{Bernoulli}(P) \) to denote that the existence of an edge between any two vertices \( 1 \leq i < j \leq n \) is a Bernoulli random variable with probability \( p_{ij} \) in which edges are independent. That is, if \( P = XX^T \), then \( A \sim \text{Bernoulli}(P) \) implies that conditioned on \( X \), \( A \) is distributed as in (2).

**Remark 1.** Given a graph distributed as an RDPG, the natural task is to recover the latent positions \( X \) that gave rise to the observed graph. However, the RDPG model has an inherent nonidentifiability in this respect: let \( X \in \mathbb{R}^{n \times d} \) be a matrix of latent positions and let \( W \in \mathbb{R}^{d \times d} \) be a unitary matrix. Since \( XX^T = (XW)(XW)^T \), it is clear that the latent positions \( X \) and \( XW \) give rise to the same distribution over graphs in Equation (2). This implies that the true latent positions can be recovered only up to some orthogonal rotation \( W \), which is evident in the statement of our main result, Theorem [1].

The focus of this paper is on multi-graph inference. In particular, we consider a collection of \( m \) random dot product graphs, all with the same latent positions. This motivates the following definition:
Definition 3. (Joint Random Dot Product Graph) Let $F$ be a $d$-dimensional inner product distribution on $\mathbb{R}^d$. We say that random graphs $A^{(1)}, A^{(2)}, \ldots, A^{(m)}$ are distributed as a joint random dot product graph (JRDPG) and write $(A^{(1)}, A^{(2)}, \ldots, A^{(m)}, X) \sim$ JRDPG($F, n, m$) if $X = [X_1, X_2, \ldots, X_n]^T \in \mathbb{R}^{n \times d}$ has its (transposed) rows distributed i.i.d. as $X_i \sim F$, and we have marginal distributions $(A^{(k)}, X) \sim$ RDPG($F, n$) for each $k = 1, 2, \ldots, m$. That is, the $A^{(k)}$ are conditionally independent given $X$, with edges independently distributed as $A_{i,j}^{(k)} \sim \text{Bernoulli}((XX^T)_{ij})$ for all $1 \leq i < j \leq n$ and all $k \in [m]$.

Throughout, we let $\delta > 0$ denote the eigengap of

$$\Delta = \mathbb{E}X_1X_1^T \in \mathbb{R}^{d \times d},$$

the second moment matrix of $X_1 \sim F$. That is, $\delta = \lambda_d(\Delta) > 0 = \lambda_{d+1}(\Delta)$. We note that $\Delta$ can be chosen diagonal without loss of generality after a suitable change of basis, as in [Athreya et al. (2016)]. We assume further that $\Delta$ is such that its diagonal entries are in nonincreasing order, so that $\Delta_{1,1} \geq \Delta_{2,2} \geq \cdots \geq \Delta_{d,d} = \delta$. We assume that the matrix $\Delta$ is constant in $n$, so that $d$ and $\delta$ are constants, while the number of graphs $m$ is allowed to grow with $n$. We leave for future work the exploration of the case where the model parameters are allowed to vary with the number of vertices $n$.

Since we rely on spectral decompositions, we begin with a straightforward one: the spectral decomposition of the positive semidefinite matrix $P = XX^T$.

Definition 4. (Spectral Decomposition of $P$) Since $P$ is symmetric and positive semidefinite, let $P = U_P S_P U_P^T$ denote its spectral decomposition, with $U_P \in \mathbb{R}^{n \times d}$ having orthonormal columns and $S_p \in \mathbb{R}^{d \times d}$ diagonal with nonincreasing entries $(S_p)_{1,1} \geq (S_p)_{2,2} \geq \cdots \geq (S_p)_{d,d} > 0$.

We note that while $P = XX^T$ is not observed, existing spectral norm bounds (e.g., Oliveira [2009]; Lu and Peng [2013]) establish that for a given $P$ and independent-edge graph $A \sim \text{Bernoulli}(P)$, the spectral norm of $A - P$ is comparatively small. As a result, we regard $A$ as a noisy version of $P$, and we begin our inference procedures with a spectral decomposition of $A$.

Definition 5. (Adjacency Spectral Embedding: Sussman et al. [2012b]) Let $A \in \mathbb{R}^{n \times n}$ be the adjacency matrix of an undirected $d$-dimensional random dot product graph. The $d$-dimensional adjacency spectral embedding (ASE) of $A$ is a spectral decomposition of $A$ based on its top $d$ eigenvalues, obtained by $\text{ASE}(A, d) = U_A S_A^{1/2}$, where $S_A \in \mathbb{R}^{d \times d}$ is a diagonal matrix whose entries are the top eigenvalues of $A$ (in nonincreasing order) and $U_A \in \mathbb{R}^{n \times d}$ is the matrix whose columns are the orthonormal eigenvectors corresponding to the eigenvalues in $S_A$.

Remark 2. We observe that without any additional assumptions, the top $d$ eigenvalues of $A$ are not guaranteed to be nonnegative. However, under our eigengap assumptions on $\Delta$, the i.i.d.-ness of the latent positions ensures that for large $n$, the eigenvalues of $A$ will be nonnegative with high probability (see Observation 2 in the Supplementary Material).
Given a set of $m$ adjacency matrices distributed as

$$(A^{(1)}, A^{(2)}, \ldots, A^{(m)}, X) \sim \text{JRDGP}(F, n, m)$$

for distribution $F$ on $\mathbb{R}^d$, a natural inference task is to recover the $n$ latent positions $X_1, X_2, \ldots, X_n \in \mathbb{R}^d$ shared by the vertices of the $m$ graphs. To estimate the underlying latent positions from these $m$ graphs, Tang et al. (2016b) provides justification for the estimate $\bar{X} = \text{ASE}(\bar{A}, d)$, where $\bar{A}$ is the sample mean of the adjacency matrices $A^{(1)}, A^{(2)}, \ldots, A^{(m)}$. However, $\bar{X}$ is ill-suited to any task that requires comparing latent positions across the $m$ graphs, since the $\bar{X}$ estimate collapses the $m$ graphs into a single set of $n$ latent positions. This motivates the omnibus embedding, which still yields a single spectral decomposition, but with a separate $d$-dimensional representation for each of the $m$ graphs. This makes the omnibus embedding useful for simultaneous inference across all $m$ observed graphs.

**Definition 6. (Omnibus embedding)** Let $A^{(1)}, A^{(2)}, \ldots, A^{(m)} \in \mathbb{R}^{n \times n}$ be (possibly weighted) adjacency matrices of a collection of $m$ undirected graphs. We define the $mn$-by-$mn$ omnibus matrix of $A^{(1)}, A^{(2)}, \ldots, A^{(m)}$ by

$$M = \begin{bmatrix}
\frac{1}{2} (A^{(1)} + A^{(1)}) & \frac{1}{2} (A^{(1)} + A^{(2)}) & \cdots & \frac{1}{2} (A^{(1)} + A^{(m)}) \\
\vdots & \vdots & \ddots & \vdots \\
\frac{1}{2} (A^{(m)} + A^{(1)}) & \frac{1}{2} (A^{(m)} + A^{(2)}) & \cdots & A^{(m)}
\end{bmatrix},$$

and the $d$-dimensional omnibus embedding of $A^{(1)}, A^{(2)}, \ldots, A^{(m)}$ is the adjacency spectral embedding of $M$:

$$\text{OMNI}(A^{(1)}, A^{(2)}, \ldots, A^{(m)}, d) = \text{ASE}(M, d).$$

If $(A^{(1)}, A^{(2)}, \ldots, A^{(m)}, X) \sim \text{JRDGP}(F, n, m)$, then the omnibus embedding provides a natural approach to estimating $X$ without collapsing the $m$ graphs into a single representation as with $\bar{X} = \text{ASE}(\bar{A}, d)$. Under the JRDGP, the omnibus matrix has expected value

$$\mathbb{E}M = \bar{P} = J_m \otimes P = U_P S_P U_P^T$$

for $U_P \in \mathbb{R}^{mn \times d}$ having $d$ orthonormal columns and $S_P \in \mathbb{R}^{d \times d}$ diagonal. Since $M$ is a reasonable estimate for $\bar{P} = \mathbb{E}M$ (see, for example, Oliveira, 2009), the matrix $\bar{Z} = \text{OMNI}(A^{(1)}, A^{(2)}, \ldots, A^{(m)}, d)$ is a natural estimate of the $mn$ latent positions collected in the matrix $Z = [X^T X^T \ldots X^T]^T \in \mathbb{R}^{mn \times d}$. Here again, as in Remark 1, $\bar{Z}$ only recovers the true latent positions $Z$ up to an orthogonal rotation. The matrix

$$Z^* = \begin{bmatrix}
X^* \\
X^* \\
\vdots \\
X^*
\end{bmatrix} = U_P S_P^{1/2} \in \mathbb{R}^{mn \times d},$$

provides a reasonable canonical choice of latent positions, so that $Z = Z^* W$ for some suitably-chosen orthogonal matrix $W \in \mathbb{R}^{d \times d}$, and our main theorem shows that we can recover $Z$ (up to orthogonal rotation) by recovering $Z^*$.  

7
3 Main Results

In this section, we give theoretical results on the consistency and asymptotic distribution of the estimated latent positions based on the omnibus matrix $\mathbf{M}$. In the next section, we demonstrate from simulations that the omnibus embedding can be successfully leveraged for subsequent inference, specifically two-sample testing.

Lemma 1 shows that the omnibus embedding provides uniformly consistent estimates of the true latent positions, up to an orthogonal transformation, roughly analogous to Lemma 5 in Lyzinski et al. (2014). Lemma 1 shows consistency of the omnibus embedding under the $2 \rightarrow \infty$ norm, implying that all $mn$ of the estimated latent positions are near their corresponding true positions. We recall that the orthogonal transformation $\tilde{\mathbf{W}}$ in the statement of the lemma is necessary since, as discussed in Remark 1, $\mathbf{P} = \mathbf{X}\mathbf{X}^T = (\mathbf{XW})(\mathbf{XW})^T$ for any orthogonal $\mathbf{W} \in \mathbb{R}^{d \times d}$.

Lemma 1. With $\tilde{\mathbf{P}}$, $\mathbf{M}$, $\mathbf{U}_M$, and $\mathbf{U}_{\tilde{\mathbf{P}}}$ defined as above, there exists an orthogonal matrix $\tilde{\mathbf{W}} \in \mathbb{R}^{d \times d}$ such that with high probability,

$$\|\mathbf{U}_M \mathbf{S}_M^{1/2} - \mathbf{U}_{\tilde{\mathbf{P}}} \mathbf{S}_{\tilde{\mathbf{P}}}^{1/2} \tilde{\mathbf{W}}\|_{2 \rightarrow \infty} \leq \frac{Cm^{1/2} \log mn}{\sqrt{n}}.$$ (6)

Proof. This result is proved in the supplemental material.

As noted earlier, our central limit theorem for the omnibus embedding is analogous to a similar result proved in Athreya et al. (2016), but with the crucial difference that we no longer require that the second moment matrix have distinct eigenvalues. As in Athreya et al. (2016), our proof here depends on writing the difference between a row of the omnibus embedding and its corresponding latent position as a pair of summands: the first, to which a classical Central Limit Theorem can be applied, and the second, essentially a combination of residual terms, which converges to zero. The weakening of the assumption of distinct eigenvalues necessitates significant changes in how to bound the residual terms. In fact, Athreya et al. (2016) adapts a result of Bickel and Sarkar (2015)—the latter of which depends on the assumption of distinct eigenvalues—to control these terms. Here, we resort to somewhat different methodology: we prove instead that analogous bounds to those in Lyzinski et al. (2017); Tang and Priebe (2016) hold for the estimated latent positions based on the omnibus matrix $\mathbf{M}$, and this enables us to establish that here, too, the rows of the omnibus embedding are also approximately normally distributed. Further, en route to this limiting result, we compute the explicit variance of the omnibus matrix, and show that as $m$, the number of graphs embedded, increases, this contributes to a reduction in the variance of the estimated latent positions.

Theorem 1. Let $(\mathbf{A}^{(1)}, \mathbf{A}^{(2)}, \ldots, \mathbf{A}^{(m)}, \mathbf{X}) \sim \text{JRPD}(\mathbf{F}, n, m)$ for some $d$-dimensional inner product distribution $\mathbf{F}$ and let $\mathbf{M}$ denote the omnibus matrix as in (4). Let $\mathbf{Z} = \mathbf{Z}'\mathbf{W}$ with $\mathbf{Z}'$ as defined in Equation (5), with estimate $\hat{\mathbf{Z}} = \text{OMNI}(\mathbf{A}^{(1)}, \mathbf{A}^{(2)}, \ldots, \mathbf{A}^{(m)}, d)$. Let $h = m(s - 1) + i$ for $i \in [n], s \in [m]$, so that $\hat{\mathbf{Z}}_h$ denotes the estimated latent position of the $i$-th vertex in the $s$-th graph $\mathbf{A}^{(s)}$. That is, $\hat{\mathbf{Z}}_h$ is the column vector formed by transposing the
h-th row of the matrix $\hat{Z} = U_M S_M^{1/2} = \text{OMNI}(A^{(1)}, A^{(2)}, \ldots, A^{(m)}, d)$. Let $\Phi(x, \Sigma)$ denote the cdf of a (multivariate) Gaussian with mean zero and covariance matrix $\Sigma$, evaluated at $x \in \mathbb{R}^d$. There exists a sequence of orthogonal $d$-by-$d$ matrices $(\tilde{W}_n)_{n=1}^\infty$ such that for all $x \in \mathbb{R}^d$,

$$\lim_{n \to \infty} \Pr\left[ n^{1/2} \left( \hat{Z} \tilde{W}_n - Z \right)_h \leq x \right] = \int_{\text{supp}_F} \Phi(x, \Sigma(y)) \, dF(y),$$

where $\Sigma(y) = (m + 3)\Delta^{-1}\tilde{\Sigma}(y)\Delta^{-1}/(4m)$, $\Delta$ is as defined in (3) and

$$\tilde{\Sigma}(y) = \mathbb{E}\left[(y^T X_1 - (y^T X_1)^2)X_1 X_1^T\right].$$

Proof. This result is proved in the supplemental material. \qed

4 Experimental results

In this section, we present experiments on synthetic data exploring the efficacy of the omnibus embedding described above. We consider both estimation of latent positions and two-sample graph testing.

4.1 Recovery of Latent Positions

Perhaps the most ubiquitous estimation problem for RDPG data is that of estimating the latent positions (i.e., the rows of the matrix $X$); consequently, we begin by exploring how well the omnibus embedding recovers the latent positions of a given random dot product graph. If one wishes merely to estimate the latent positions $X$ of a set of $m$ graphs $(A^{(1)}, A^{(2)}, \ldots, A^{(m)}, X) \sim \text{JRDPG}(F, n, m)$, the estimate $\bar{X} = \text{ASE}(\sum_{i=1}^m A^{(i)}/m, d)$, the embedding of the sample mean of the adjacency matrices performs well asymptotically (Tang et al., 2016b). Indeed, all else equal, the embedding $\bar{X}$ is preferable to the omnibus embedding if only because it requires an eigendecomposition of an $n$-by-$n$ matrix rather than the much larger $mn$-by-$mn$ omnibus matrix.

Of course, the omnibus embedding can still be used to to estimate the latent positions, potentially at the cost of increased variance. Figure 1 compares the mean-squared error of various techniques for estimating the latent positions for a random dot product graph. The figure plots the (empirical) mean squared error in recovering the latent positions of a 3-dimensional JRDPG as a function of the number of vertices $n$. Each point in the plot is the empirical mean of 50 independent trials. In each trial, the vertex latent positions are drawn i.i.d. from a Dirichlet with parameter $[1, 1, 1]^T \in \mathbb{R}^3$. Having generated a random set of latent positions, we generate two graphs, $A^{(1)}, A^{(2)} \in \mathbb{R}^{n \times n}$ independently, based on this set of latent positions. Thus, we have $(A^{(1)}, A^{(2)}, X) \sim \text{JRDPG}(F, n, 2)$, where $F = \text{Dir}([1, 1, 1]^T)$ is a Dirichlet with parameter $[1, 1, 1]^T \in \mathbb{R}^3$, and $n$ varies. The lines correspond to

1. **ASE1 (red)**: we embed only one of the two observed graphs, and use only the ASE of that graph to estimate the latent positions in $X$. That is, we consider $\text{ASE}(A^{(1)})$
Figure 1: Mean squared error (MSE) in recovery of latent positions (up to rotation) in a 2-graph joint RDPG model as a function of the number of vertices. The figure shows the performance of ASE applied to a single graph (red), ASE embedding of the mean graph (gold), the Procrustes-based pairwise embedding (blue), the omnibus embedding (green) and the mean omnibus embedding (purple). Each point is the mean of 50 trials, with error bars indicating two times the standard error. We see that the mean omnibus embedding (OMNIbar) achieves performance competitive with that of the optimal embedding ASE(Å, d), while the Procrustes alignment estimation is notably inferior to the other two-graph techniques for graphs of size between 80 and 200 vertices (and we note that the gap appears to persist at larger graph sizes, though it shrinks).

2. Abar (gold): we embed the average of the two graphs, $\hat{A} = (A^{(1)} + A^{(2)})/2$ as $\hat{X} = ASE(\hat{A}, 3)$. As discussed in, for example, Tang et al. (2016b), this is the lowest-variance estimate of the latent positions $X$.

3. OMNI (green): We apply the omnibus embedding to obtain $\hat{Z} = ASE(M, 3)$, where $M$ is as in Equation (4). We then use only the first $n$ rows of $\hat{Z} \in \mathbb{R}^{2n \times d}$ as our estimate of $X$. Thus, this embedding takes advantage of the information available in both graphs $A^{(1)}$ and $A^{(2)}$, but does not use both graphs equally, since the first rows of $\hat{Z}$ are based primarily on the information contained in $A^{(1)}$.

4. OMNIbar (purple): We again apply the omnibus embedding to obtain estimated latent positions $\hat{Z} = ASE(M, 3)$, but this time we use all available information by averaging the first $n$ rows and the second $n$ rows of $\hat{Z}$.
5. **PROCbar (blue)**: We separately embed the graphs $A^{(1)}$ and $A^{(2)}$, obtaining two separate estimates of the latent positions in $\mathbb{R}^3$. We then align these two sets of estimated latent positions via Procrustes alignment, and average the aligned embeddings to obtain our final estimate of the latent positions.

First, let us note that ASE applied to a single graph (red) lags all other methods. This is expected, since all other methods assessed in Figure 1 use information from both observed graphs $A^{(1)}$ and $A^{(2)}$ rather than only $A^{(1)}$. We see that all other methods perform essentially equally well on graphs of 50 vertices or fewer. Given the dearth of signal in these smaller graphs, we do not expect any method to recover the latent positions accurately.

Crucially, however, we see that the OMNIbar estimate (purple) performs nearly identically to the Abar estimate (gold), the natural choice among spectral methods for the estimation latent positions (for more on the efficiency of Abar, see Tang et al., 2016b). The Procrustes estimate (in blue) provides a two-graph analogue of ASE (red): it combines two ASE estimates via Procrustes alignment, but does not enforce an *a priori* alignment of the estimated latent positions in the manner of the omnibus embedding does (we discuss this enforced alignment in 5 as well.) As predicted by the results in Lyzinski et al. (2014) and Tang et al. (2016a), the Procrustes estimate is competitive with the Abar (gold) estimate for suitably large graphs. The OMNI estimate (in green) serves, in a sense, as an in-between method, in that it uses information available from both graphs, but in contrast to Procrustes (blue), OMNIbar (purple) and Abar (gold), it does not make complete use of the information available in the second graph. For this reason, it is noteworthy that the OMNI estimate outperforms the Procrustes estimate for graphs of 80-100 vertices. That is, for certain graph sizes, the omnibus estimate appears to more optimally leverage the information in both graphs than the Procrustes estimate does, despite the fact that the information in the second graph has comparatively little influence on the OMNI embedding.

### 4.2 Two-graph Hypothesis Testing

We now turn to the matter of using the omnibus embedding for testing the semiparametric hypothesis that two observed graphs are drawn from the same underlying latent positions. Suppose we have a set of points $X_1, X_2, \ldots, X_n, Y_1, Y_2, \ldots, Y_n \in \mathbb{R}^d$. Let the graph $G_1$ with adjacency matrix $A^{(1)}$ have edges distributed independently as $A_{ij}^{(1)} \sim \text{Bernoulli}(X_i^T X_j)$. Similarly, let $G_2$ have adjacency matrix $A^{(2)}$ with edges distributed independently as $A_{ij}^{(2)} \sim \text{Bernoulli}(Y_i^T Y_j)$. As discussed previously, while $\bar{A} = (A^{(1)} + A^{(2)})/2$ may be optimal for estimation of latent positions, it is not clear how to use the embedding $\text{ASE}(\bar{A}, d)$ to test the following hypothesis:

$$H_0 : X_i = Y_i \ \forall i \in [n]. \quad (7)$$

On the other hand, the omnibus embedding provides a natural test of the null hypothesis (7) by comparing the first $n$ and last $n$ embeddings of the omnibus matrix

$$M = \begin{bmatrix} \frac{A^{(1)}}{(A^{(1)} + A^{(2)})/2} & \frac{(A^{(1)} + A^{(2)})/2}{A^{(2)}} \end{bmatrix}.$$
Intuitively, when $H_0$ holds, the distributional result in Theorem 1 holds, and the $i$-th and $(n+i)$-th rows of OMNI($A^{(1)}, A^{(2), d}$) are equidistributed (though they are not independent). On the other hand, when $H_0$ fails to hold, there exists at least one $i \in [n]$ for which the $i$-th and $(n+i)$-th rows of $M$ are not identically distributed, and thus the corresponding embeddings are also distributionally distinct. This suggests a test that compares the first $n$ rows of OMNI($A^{(1)}, A^{(2), d}$) against the last $n$ rows (see below for details). Here, we empirically explore the power this test against its Procrustes-based alternative from Tang et al. (2016a).

Our setup is as follows. We draw $X_1, X_2, \ldots, X_n \in \mathbb{R}^3$ i.i.d. according to a Dirichlet distribution $F$ with parameter $\alpha = [1, 1, 1]^T$. Assembling these $n$ points into a matrix $X = [X_1 X_2 \ldots X_n]^T \in \mathbb{R}^{n \times 3}$, we can generate a graph $G_1$ with adjacency matrix $A^{(1)}$ with entries $A_{ij}^{(1)} \sim \text{Bernoulli}((XX^T)_{ij})$. Thus, $(A^{(1)}, X) \sim \text{RDPG}(F, n)$. We generate a second graph $G_2$ by first drawing random points $Z_1, Z_2, \ldots, Z_n \overset{i.i.d.}{\sim} F$. Selecting a set of indices $I \subset [n]$ of size $k < n$ uniformly at random from among all such $\binom{n}{k}$ sets, we let $G_2$ have latent positions

$$Y_i = \begin{cases} Z_i & \text{if } i \in I \\ X_i & \text{otherwise.} \end{cases}$$

Assembling these points into a matrix $Y = [Y_1, Y_2, \ldots, Y_n]^T \in \mathbb{R}^{n \times 3}$, we generate graph $G_2$ with adjacency matrix $A^{(2)}$ with edges generated independently according to $A_{ij}^{(2)} \sim \text{Bernoulli}(YY^T)_{ij}$. The task is then to test the hypothesis

$$H_0 : X = Y.$$ (8)

To test this hypothesis, we consider two different tests, one based on a Procrustes alignment of the adjacency spectral embeddings of $G_1$ and $G_2$ (Tang et al. 2016a) and the other based on the omnibus embedding. Both approaches are based on estimates of the latent positions of the two graphs. In both cases we use a test statistic of the form $T = \sum_{i=1}^n \|\hat{X}_i - \hat{Y}_i\|^2_F$, and accept or reject based on a Monte Carlo estimate of the critical value of $T$ under the null hypothesis, in which $X_i = Y_i$ for all $i \in [n]$. In each trial, we use 500 Monte Carlo iterates to estimate the distribution of $T$.

We note that in the experiments presented here, we assume that the latent positions $X_1, X_2, \ldots, X_n$ of graph $G_1$ are known for sampling purposes, so that the matrix $P = \mathbb{E}A^{(1)}$ is known exactly, rather than estimated from the observed adjacency matrix $A^{(1)}$. This allows us to sample from the true null distribution. As proved in Lyzinski et al. (2014), the estimated latent positions $\hat{X}_1 = \text{ASE}(A^{(1)})$ and $\hat{X}_2 = \text{ASE}(A^{(2)})$ recover the true latent positions $X_1$ and $X_2$ (up to rotation) to arbitrary accuracy in $(2, \infty)$-norm for suitably large $n$ (Lyzinski et al. 2014). Without using this known matrix $P$, we would require that our matrices have tens of thousands of vertices before the variance associated with estimating the latent positions would no longer overwhelm the signal present in the few altered latent positions.

Three major factors influence the complexity of testing the null hypothesis in Equation (8): the number of vertices $n$, the number of changed latent positions $k = |I|$, and the
Figure 2: Power of the ASE-based (blue) and omnibus-based (green) tests to detect when the two graphs being testing differ in (a) one, (b) five, and (c) ten of their latent positions. Each point is the proportion of 1000 trials for which the given technique correctly rejected the null hypothesis, and error bars denote two standard errors of this empirical mean in either direction.

The distances $\|X_i - Y_i\|_F$ between the latent positions. The three plots in Figure 2 illustrate the first two of these three factors. These three plots show the power of two different approaches to testing the null hypothesis (8) for different sized graphs and for different values of $k$, the number of altered latent positions. In all three conditions, both methods improve as the number of vertices increases, as expected, especially since we do not require estimation of the underlying expected matrix $P$ for Monte Carlo estimation of the null distribution of the test statistic. We see that when only one vertex is changed, neither method has power much above 0.25. However, in the case of $k = 5$ and $k = 10$, is it clear that the omnibus-based test achieves higher power than the Procrustes-based test, especially in the range of 30 to 250 vertices.

Figure 3 shows the effect of the difference between the latent position matrices under null and alternative. We consider a 3-dimensional RDGP on $n$ vertices, in which one latent position, $i \in [n]$, is fixed to be equal to $x_i = (0.8, 0.1, 0.1)^T$ and the remaining latent positions are drawn i.i.d. from a Dirichlet with parameter $\vec{\alpha} = (1, 1, 1)^T$. We collect these latent positions in the rows of the matrix $X \in \mathbb{R}^{n \times 3}$. To produce the latent positions $Y \in \mathbb{R}^{n \times 3}$ of the second graph, we use the same latent positions in $X$, but we alter the $i$-th position to be $Y_i = (1 - \lambda)x_i + \lambda(0.1, 0.1, 0.8)^T$ for $\lambda \in [0, 1]$ a “drift” parameter, controlling how much the latent position changes between the two graphs. Intuitively, correctly rejecting $H_0: X = Y$ is easier for larger values of $\lambda$; the greater the gap between latent position matrices under null and alternative, the more easily our test procedure should discriminate between them. Figure 3 shows how the size of the drift parameter influences the power. We see that for $n = 30$ vertices (top left), neither the omnibus nor Procrustes test has power appreciably better than approximately 0.05, largely in agreement with the what we observed in Figure 2. Similarly, when $n = 200$ vertices (bottom right), both methods perform approximately equally (though omnibus does appear to consistently outperform Procrustes testing). The case of $n = 50$ and $n = 100$ vertices (upper right and bottom left, respectively), though, offers a fascinating instance in which the omnibus test consistently outperforms the Procrustes test. Particularly interesting to note is the $n = 50$ case (top right), in which we
Figure 3: Power of the ASE-based (blue) and omnibus-based (green) tests to detect when the two graphs being testing differ in their latent positions. Subplots show power as a function of the drift parameter $\lambda$ for (a) $n = 30$, (b) $n = 50$, (c) $n = 100$ and (d) $n = 200$ vertices. Each point is the proportion of 500 trials for which the given technique correctly rejected the null hypothesis, and error bars denote two standard errors of this empirical mean.

see that performance of the Procrustes test is more or less flat as a function of drift parameter $\lambda$, while the omnibus embedding clearly improves as $\lambda$ increases, with performance climbing well above that of Procrustes for $\lambda > 0.8$.

5 Discussion and Conclusion

We have presented a method for the simultaneous “omnibus” embedding of multiple graphs on the same vertex set, and a central limit theorem for the embeddings of individual vertices. Our distributional results for the asymptotic normality of the rows of the omnibus embedding quantify the impact of multiple graphs on the variance of the rows, specifically in relation to the variance given in [Athreya et al., 2016]. This result shows that as the number of graphs, $m$, grows, a significant reduction in the variance is achievable. Experimental data suggest that the omnibus embedding is competitive with state-of-the-art, multiple-graph spectral
estimation of latent positions. We surmise that the variance of the rows in the omnibus embedding is close to optimal for latent position estimators derived from the adjacency spectral embedding. That is, the variance of the omnibus embedding is asymptotically equal to the variance obtained by first averaging the $m$ graphs to get $\bar{A}$ (which corresponds, in essence, to the maximum likelihood estimate for $P$), and then performing an adjacency spectral embedding of $\bar{A}$. Let $\hat{Z}_i$ correspond to the $i$-th row of $\hat{Z} = \text{OMNI}(A^{(1)}, A^{(2)}, \ldots, A^{(m)}, d)$, and let $\bar{X}_i$ denote the $i$-th row of the adjacency spectral embedding of $\bar{A}$. Let $\bar{\hat{Z}}_i$ denote the average value of the $m$ rows of $\hat{Z}$ corresponding to the $i$-th vertex, i.e., the average of the $m$ vectors corresponding to the $i$-th vertex in the omnibus embedding. We conjecture that averaging the rows of the omnibus embedding accounts for all of the reduction in variance when one compares a single row of the omnibus embedding and a single row of $\bar{X}$. Figure 1 provides weak evidence in favor of this conjecture, since it illustrates that the MSE of both the omnibus- and Procrustes-based estimates of the latent position estimates are very close to that of the estimate based on the mean adjacency matrix.

**Conjecture 1.** (Decomposition of Variance) With notation as above, for large $n$,

$$\text{Var}(\sqrt{n}\bar{X}_i) \approx \text{Var}(\sqrt{n}\bar{\hat{Z}}_i) < \text{Var}(\sqrt{n}\hat{Z}_i).$$

We have also demonstrated that the omnibus embedding can be profitably deployed for two-sample semiparametric hypothesis testing of graph-valued data. Our omnibus embedding provides a natural mechanism for the simultaneous embedding of multiple graphs into a single vector space. This eliminates the need for multiple Procrustes alignments, which were required in previously-explored approaches to multiple-graph testing (Tang et al., 2016a). In the two-graph hypothesis testing framework of Tang et al. (2016a), each graph is embedded separately. Under the assumption of equality of latent positions (i.e., under $H_0$ in Equation (7)), we note that embedding the first graph estimates the true latent positions $X$ up to a unitary transformation in $\mathbb{R}^{d \times d}$. Call this estimate $\hat{X}_1$. Similarly, $\hat{X}_2$, the estimates based on the second graph, estimates $X$ only up to some potentially different unitary rotation, i.e., $\hat{X}_2 \approx XW^*$ for some unitary $W^*$. Procrustes alignment is thus required to discover the rotation aligning $\hat{X}_1$ with $\hat{X}_2$. In Tang et al. (2016a), it was shown that this Procrustes alignment, given by

$$\min_{W \in O_d} \|\hat{X}_1 - \hat{X}_2 W\|_F, \quad (9)$$

converges under the null hypothesis. The effect of this Procrustes alignment on subsequent inference is ill-understood. At the very least, it has the potential to introduce variance, and our simulations in Section 4 suggest that it negatively impacts performance in both estimation and testing settings. Furthermore, when the matrix $P = XX^T$ does not have distinct eigenvalues (i.e., is not uniquely diagonalizable), this Procrustes step is unavoidable, since the difference $\|\hat{X}_1 - \hat{X}_2\|_F$ need not converge at all.

In contrast, our omnibus embedding builds an alignment of the graphs into its very structure. To see this, consider, for simplicity, the $m = 2$ case. Let $X \in \mathbb{R}^{n \times d}$ be the matrix whose rows are the latent positions of both graphs $G_1$ and $G_2$, and let $M \in \mathbb{R}^{2n \times 2n}$ be their
omnibus matrix. Then
\[ \mathbf{EM} = \tilde{\mathbf{P}} = \begin{bmatrix} \mathbf{P} & \mathbf{P} \\ \mathbf{P} & \mathbf{P} \end{bmatrix} = \begin{bmatrix} \mathbf{X} \\ \mathbf{X} \end{bmatrix} \begin{bmatrix} \mathbf{X} \\ \mathbf{X} \end{bmatrix}^T. \]

Suppose now that we wish to factorize \( \tilde{\mathbf{P}} \) as
\[ \tilde{\mathbf{P}} = \begin{bmatrix} \mathbf{X} \\ \mathbf{XW}^* \end{bmatrix} \begin{bmatrix} \mathbf{X} \\ \mathbf{XW}^* \end{bmatrix}^T = \begin{bmatrix} \mathbf{P} & \mathbf{X(W^*)^TX}^T \\ \mathbf{XW^XT} & \mathbf{P} \end{bmatrix}. \]

That is, we want to consider graphs \( G_1 \) and \( G_2 \) as being generated from the same latent positions, but in one case, say, under a different rotation. This possibility necessitates the Procrustes alignment in the case of separately-embedded graphs. In the case of the omnibus matrix, the structure of the \( \tilde{\mathbf{P}} \) matrix implies that \( \mathbf{W}^* = \mathbf{I}_d \). Thus, in contrast to the Procrustes alignment, the omnibus matrix incorporates an alignment \textit{a priori}. Simulations show that the omnibus embedding outperforms the Procrustes-based test for equality of latent positions, especially in the case of moderately-sized graphs.

To further illustrate the utility of this omnibus embedding, consider the case of testing whether three different random dot product graphs have the same generating latent positions. The omnibus embedding gives us a \textit{single} canonical representation of all three graphs: Let \( \hat{\mathbf{X}}_1^O, \hat{\mathbf{X}}_2^O, \) and \( \hat{\mathbf{X}}_3^O \) be the estimates for the three latent position matrices generated from the omnibus embedding. To test whether any two of these random graphs have the same generating latent positions, we merely have to compare the Frobenius norms of their differences, as opposed to computing three separate Procrustes alignments. In the latter case, in effect, we do not have a canonical choice of coordinates in which to compare our graphs simultaneously.

Finally, we remark that open problems abound, including an analysis of the omnibus embedding when the \( m \) graphs are correlated, or when some are corrupted by occlusion or noise; a closer examination of the impact of the Procrustes alignment on power; the development of an analogue to a Tukey test for determining which graphs differ when we test equality of multiple graphs; and the comparative efficiency of the omnibus embedding relative to other spectral estimates.

References

A. Athreya, V. Lyzinski, D. J. Marchette, C. E. Priebe, D. L. Sussman, and M. Tang. A limit theorem for scaled eigenvectors of random dot product graphs. \textit{Sankhya A}, 78:1–18, 2016.

M. Belkin and P. Niyogi. Laplacian eigenmaps for dimensionality reduction and data representation. \textit{Neural Computation}, 15:1373–1396, 2003.

R. Bhatia. \textit{Matrix Analysis}. 1997, 1997.

P. Bickel and P. Sarkar. Role of normalization for spectral clustering in stochastic blockmodels. \textit{Annals of Statistics}, 43:962–990, 2015.
L. Chen, J.T. Vogelstein, V. L. Lyzinski, and C. E. Priebe. A joint graph inference case study: the c. elegans chemical and electrical connectomes. *Worm*, 5(2), 2016.

C. Davis and W. M. Kahan. The rotation of eigenvectors by a perturbation. *SIAM J. Numerical Analysis*, 7(1), March 1970.

I. L. Dryden and K. V. Mardia. *Statistical Shape Analysis*. Wiley, 1998.

Z. Füredi and J. Komlós. The eigenvalues of random symmetric matrices. *Combinatorica*, 1(3):233–241, 1981.

J. C. Gower. Generalized procrustes analysis. *Psychometrika*, 40:33–51, 1975.

P. W. Holland, K. B. Laskey, and S. Leinhardt. Stochastic blockmodels: first steps. *Social Networks*, 5:109–137, 1983.

R. Horn and C. Johnson. *Matrix Analysis*. Cambridge University Press, 1985.

B. Karrer and M. E. J. Newman. Stochastic blockmodels and community structure in networks. *Physical Review E*, 83:016107, 2011.

L. Lu and X. Peng. Spectra of edge-independent random graphs. *Electronic Journal of Combinatorics*, 20, 2013.

V. Lyzinski, D. L. Sussman, M. Tang, A. Athreya, and C. E. Priebe. Perfect clustering for stochastic blockmodel graphs via adjacency spectral embedding. *Electronic Journal of Statistics*, 8:2905–2922, 2014.

V. Lyzinski, M. Tang, A. Athreya, Y. Park, and C. E. Priebe. Community detection and classification in hierarchical stochastic blockmodels. *IEEE Transactions in Network Science and Engineering*, 2017.

R. I. Oliveira. Concentration of the adjacency matrix and of the Laplacian in random graphs with independent edges. [http://arxiv.org/abs/0911.0600](http://arxiv.org/abs/0911.0600), 2009.

D. L. Sussman, M. Tang, D. E. Fishkind, and C. E. Priebe. A consistent adjacency spectral embedding for stochastic blockmodel graphs. *J. Amer. Statist. Assoc.*, 107(499):1119–1128, 2012a.

D. L. Sussman, M. Tang, D. E. Fishkind, and C. E. Priebe. A consistent adjacency spectral embedding for stochastic blockmodel graphs. *Journal of the American Statistical Association*, 107:1119–1128, 2012b.

M. Tang and C. E. Priebe. Limit theorems for eigenvectors of the normalized laplacian for random graphs. Arxiv preprint at [https://arxiv.org/abs/1607.08601](https://arxiv.org/abs/1607.08601), 2016.

M. Tang, D. L. Sussman, and C. E. Priebe. Universally consistent vertex classification for latent position graphs. *Ann. Statist.*, 41:1406 – 1430, 2013.
M. Tang, A. Athreya, D. L. Sussman, V. Lyzinski, Y. Park, and C. E. Priebe. A semiparametric two-sample hypothesis testing problem for random dot product graphs. *Journal of Computational and Graphical Statistics*, 2016a. To appear.

M. Tang, A. Athreya, D. L. Sussman, V. Lyzinski, and C. E. Priebe. A nonparametric two-sample hypothesis testing problem for random dot product graphs. *Bernoulli*, 23: 1599–1630, 2017.

R. Tang, M. Ketcha, J. T. Vogelstein, C. E. Priebe, and D. L. Sussman. Laws of large graphs. arXiv preprint at [http://arxiv.org/abs/1609.01672](http://arxiv.org/abs/1609.01672), 2016b.

T. Tao and V. Vu. Random matrices: Universal properties of eigenvectors. *Random Matrices: Theory and Applications*, 1(01), 2012.

J. A. Tropp. An introduction to matrix concentration inequalities. *Foundations and Trends in Machine Learning*, 8(1–2):1 – 230, 2015.

S. Young and E. Scheinerman. Random dot product graph models for social networks. In *Proceedings of the 5th international conference on algorithms and models for the web-graph*, pages 138–149, 2007.

Y. Yu, T. Wang, and R. J. Samworth. A useful variant of the Davis-Kahan theorem for statisticians. *Biometrika*, 102:315–323, 2015.
SUPPLEMENTARY MATERIAL

We collect here the technical proofs supporting our main result, Theorem 1. We consider the (transposed) $h$-th row of the matrix $\sqrt{n} \left( U_M S_M^{1/2} V^T W_n - U \tilde{P} S_{\tilde{P}}^{1/2} \right)$, where $V, W_n \in \mathbb{R}^{d \times d}$ are orthogonal transformations. We follow the reasoning of Theorem 18 in Lyzinski et al. (2017), decomposing this matrix as $\sqrt{n} \left( U_M S_M^{1/2} V^T W_n - U \tilde{P} S_{\tilde{P}}^{1/2} \right) = \sqrt{n}(N + H)$, where $N, H \in \mathbb{R}^{mn \times d}$. To prove our central limit theorem, we show that the (transposed) $h$-th row of $\sqrt{n} H$ converges in probability to 0 and that the (transposed) $h$-th row of $\sqrt{n} N$ converges in distribution to a mixture of normals. We note that Lemma 2, Observation 2, Lemma 3, Propositions 1 and 2, Lemma 4, and Lemma 5 provide the groundwork for establishing our consistency result, Lemma 1, and, thereafter, for showing that the $h$-th row of $\sqrt{n} H$ converges in probability to 0. Next, Lemma 6 establishes that the $h$-th row of $\sqrt{n} N$ converges in distribution to a mixture of normals; the proof of Theorem 1 then follows from Slutsky’s Theorem.

We begin with a standard matrix concentration inequality, reproduced from Tropp (2015).

**Theorem 2.** (Matrix Bernstein; Tropp, 2015, Theorem 1.6.2) Consider independent random Hermitian matrices $H^{(1)}, H^{(2)}, \ldots, H^{(k)} \in \mathbb{R}^{n \times n}$ with $E[H^{(i)}] = 0$ and $\|H^{(i)}\| \leq L$ with probability 1 for all $i$ for some fixed $L > 0$. Define $H = \sum_{i=1}^{k} H^{(i)}$, and let $v(H) = \|E[H^2]\|$. Then for all $t \geq 0$,$\Pr[\|H\| \geq t] \leq 2n \exp\left\{ \frac{-t^2/2}{v(H) + Lt/3} \right\}.$

We will apply this matrix Bernstein inequality to the omnibus matrix to obtain a bound on $\|M - \hat{P}\|$, from which it will follow by Weyl’s inequality (Horn and Johnson 1985) that the eigenvalues of $M$ are close to those of $EM$.

**Lemma 2.** Let $M \in \mathbb{R}^{mn \times mn}$ be the omnibus matrix of $A^{(1)}, A^{(2)}, \ldots, A^{(m)}$, where $(A^{(1)}, A^{(2)}, \ldots, A^{(m)}, X) \sim \text{JRDPG}(F,n,m).$ Then $\|M - EM\| \leq Cmn^{1/2} \log^{1/2} mn$ w.h.p.

**Proof.** Condition on some $P = XX^T$, so that

$$EM = \hat{P} = \begin{bmatrix} P & P & \ldots & P \\ P & P & \ldots & \vdots \\ \vdots & \vdots & \ddots & \vdots \\ P & \ldots & \ldots & P \end{bmatrix} \in \mathbb{R}^{mn \times mn}. \quad (10)$$

We will apply Theorem 2 to $M - EM$. For all $q \in [m]$ and $i, j \in [n]$, let $e_{ij} = e_i e_j^T + e_j e_i^T$.  

and define block matrix $E_{q,i,j} \in \mathbb{R}^{mn \times mn}$ with blocks of size $n$-by-$n$ by

$$E_{q,i,j} = \begin{bmatrix}
0 & \ldots & 0 & e^{ij} & 0 & \ldots & 0 \\
0 & \ldots & 0 & e^{ij} & 0 & \ldots & 0 \\
\vdots & \ddots & \vdots & \vdots & \ddots & \vdots & \vdots \\
0 & \ldots & 0 & e^{ij} & 0 & \ldots & 0 \\
e^{ij} & \ldots & e^{ij} & 2e^{ij} & e^{ij} & \ldots & e^{ij} \\
0 & \ldots & 0 & e^{ij} & 0 & \ldots & 0 \\
\vdots & \ddots & \vdots & \vdots & \ddots & \vdots & \vdots \\
0 & \ldots & 0 & e^{ij} & 0 & \ldots & 0
\end{bmatrix},$$

where the $e^{ij}$ terms appear in the $q$-th row and $q$-th column. Using this definition, we have

$$M - EM = \sum_{q=1}^{m} \sum_{1 \leq i < j \leq n} \frac{A^{(q)}_{ij} - P_{ij}}{2} E_{q,i,j},$$

which is a sum of $m\binom{n}{2}$ independent zero-mean matrices, with $\| (A^{(q)}_{ij} - P_{ij}) E_{q,i,j} / 2 \| \leq \sqrt{m + 1}$ for all $q \in [m]$ and $i, j \in [n]$.

To apply Theorem 2, it remains to consider the variance term $v(M - EM)$. We note first that, letting $D_{ij} = e_i e_j^T + e_j e_i^T \in \mathbb{R}^{n \times n}$, we have

$$E_{q,i,j} E_{q,i,j} = \begin{bmatrix}
D_{ij} & \ldots & D_{ij} & D_{ij} & D_{ij} & \ldots & D_{ij} \\
\vdots & \ddots & \vdots & \vdots & \ddots & \vdots & \vdots \\
D_{ij} & \ldots & D_{ij} & D_{ij} & D_{ij} & \ldots & D_{ij} \\
\vdots & \ddots & \vdots & \vdots & \ddots & \vdots & \vdots \\
D_{ij} & \ldots & D_{ij} & D_{ij} & D_{ij} & \ldots & D_{ij}
\end{bmatrix},$$

where the $(m + 3)D_{ij}$ term appears in the $q$-th entry on the diagonal. Using the fact that the maximum row sum is an upper bound on the spectral norm (Horn and Johnson 1985),

$$v(M - EM) = \left\| \sum_{q=1}^{m} \sum_{1 \leq i < j \leq n} \frac{(A^{(q)}_{ij} - P_{ij})^2}{4} E_{q,i,j} E_{q,i,j} \right\| \leq \frac{(m + 1)^2(n - 1)}{4}. \quad (11)$$

Applying this upper bound on $v(M - EM)$ in Theorem 2 with $t = 12(m + 1)\sqrt{(n - 1)\log mn}$, we obtain

$$\Pr \left[ \|M - EM\| \geq 12(m + 1)\sqrt{(n - 1)\log mn} \right] \leq 2m^{-3}n^{-2}.$$ 

Integrating over all $X$ yields the result.

**Observation 1.** Let $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_d > 0$ denote the top $d$ eigenvalues of $P = XX^T$, and let $\tilde{P}$ be as in Equation (10). Then $\sigma(\tilde{P}) = \{m\lambda_1, m\lambda_2, \ldots, m\lambda_d, 0, \ldots, 0\}$. 

20
Proof. This is immediate from the structure of \( \tilde{P} \), as defined in Equation (10). \( \square \)

**Observation 2.** Let \( F \) be an inner product distribution on \( \mathbb{R}^d \) and let \( X_1, X_2, \ldots, X_n, Y \) \( i.i.d. \) \( F \). With probability at least \( 1 - d^2/n^2 \), it holds for all \( i \in [d] \) that \( |\lambda_i(P) - n\lambda_i(\mathbb{E}YY^T)| \leq 2d\sqrt{n\log n} \). Further, we have for all \( i \in [d] \), \( \lambda_i(P) \leq Cmn\delta \) with high probability.

Proof. A slightly looser version of this bound appeared in [Athreya et al. (2016)](Athreya). We include a proof of this improved result for the sake of completeness.

Note that for \( 1 \leq i \leq d \), we have \( \lambda_i(P) = \lambda_i(XX^T) = \lambda_i(X^TX) \). Hoeffding’s inequality applied to \((X^TX - n\mathbb{E}YY^T)_{ij}\) yields, for all \( i, j \in [d] \),

\[
\Pr\left[ |(X^TX)_{ij} - n\mathbb{E}YY^T|_{ij} \geq 2\sqrt{n \log n} \right] \leq \frac{2}{n^2}.
\]

A union bound over all \( i, j \in [d] \) implies that \( \|X^TX - n\mathbb{E}YY^T\|_F^2 \leq 4d^2n\log n \) with probability at least \( 1 - 2d^2/n^2 \). Upper bounding the spectral norm by the Frobenius norm, we have \( \|X^TX - n\mathbb{E}YY^T\|_2 \leq 2d\sqrt{n\log n} \) with probability at least \( 1 - 2d^2/n^2 \), and Weyl’s inequality (Horn and Johnson, 1985) thus implies \( |\lambda_i(P) - n\lambda_i(\mathbb{E}YY^T)| \leq 2d\sqrt{n\log n} \) for all \( 1 \leq i \leq d \), from which Observation 1 and the reverse triangle inequality yield

\[
\lambda_i(\tilde{P}) = m\lambda_i(P) \geq m\lambda_d(P) \geq m|n\lambda_d(\mathbb{E}YY^T) - 2d\sqrt{n\log n}| \geq Cmn
\]

for suitably large \( n \). \( \square \)

The next several lemmas follow the reasoning in [Lyzinski et al. (2017)](Lyzinski), in particular Proposition 16, Lemma 17 and Theorem 18, and thus they are stated here without proof.

**Lemma 3.** (Adapted from Lyzinski et al., 2017 Prop. 16) Let \( \tilde{P} = UP_\Sigma U^T \) be the eigendecomposition of \( \tilde{P} \), where \( U_\Sigma \in \mathbb{R}^{mn \times d} \) has orthonormal columns and \( S_\Sigma \in \mathbb{R}^{d \times d} \) is diagonal and invertible. Let \( S_M \in \mathbb{R}^{d \times d} \) be the diagonal matrix of the top \( d \) eigenvalues of \( M \) and \( U_M \in \mathbb{R}^{mn \times d} \) be the matrix with orthonormal columns containing the top \( d \) corresponding eigenvectors, so that \( U_M S_M U_M^T \) is our estimate of \( \tilde{P} \), as described above. Let \( V_1 \Sigma V_2^T \) be the SVD of \( U_\Sigma U_M^T \). Then

\[
\|U_\Sigma U_M - V_1 \Sigma V_2^T\|_F \leq \frac{C \log mn}{n} \ w.h.p.
\]

It will be helpful to have the following two propositions, both of which follow from standard applications of Hoeffding’s inequality.

**Proposition 1.** With notation as above,

\[
\|U_\Sigma (M - \tilde{P})\|_F \leq C\sqrt{mn(m + \log mn)} \ w.h.p.
\]

**Proposition 2.** With notation as above,

\[
\|U_\Sigma (M - \tilde{P}) U_\Sigma \|_F \leq C\sqrt{m \log mn} \ w.h.p.
\]
In what follows, we let $V = V_1V_2^T$, where $V_1$ and $V_2$ are as defined in Lemma 3, i.e., $V_1 \Sigma V_2^T$ is the SVD of $U_\tilde{P}U_M$. The following lemma shows that the matrix $V$ “approximately commutes” with several diagonal matrices that will be of import in later computations.

**Lemma 4.** (Adapted from Lyzinski et al., 2017, Lemma 17) Let $V = V_1V_2^T$ be as defined above. Then

$$\|V_S - S_p V\|_F \leq Cm \log mn \text{ w.h.p. , }$$  \hspace{1cm} (12)

$$\|V_{S}^{1/2} - S_{p}^{1/2} V\|_F \leq \frac{Cm^{1/2} \log mn}{n^{1/2}} \text{ w.h.p.}$$  \hspace{1cm} (13)

and

$$\|V_{S}^{-1/2} - S_{p}^{-1/2} V\|_F \leq C(mn)^{-3/2} \text{ w.h.p.}$$  \hspace{1cm} (14)

To prove our central limit theorem, we require somewhat more precise control on certain residual terms, which we establish in the following key lemma.

**Lemma 5.** Define

$$R_1 = U_\tilde{P}U_\tilde{P}^T U_M - U_\tilde{P} V$$

$$R_2 = VS_{M}^{1/2} - S_{p}^{1/2} V$$

$$R_3 = U_M - U_\tilde{P}U_\tilde{P}^T U_M + R_1 = U_M - U_\tilde{P} V$$

Then the following convergences in probability hold:

$$\sqrt{n} \left[(M - \tilde{P})U_\tilde{P}(VS_{M}^{-1/2} - S_{p}^{-1/2} V)\right] \overset{P}{\to} 0,$$  \hspace{1cm} (15)

$$\sqrt{n} \left[U_\tilde{P}U_\tilde{P}^T(M - \tilde{P})U_\tilde{P}VS_{M}^{-1/2}\right] \overset{P}{\to} 0,$$  \hspace{1cm} (16)

$$\sqrt{n} \left[(I - U_\tilde{P}U_\tilde{P}^T)(M - \tilde{P})R_3 S_{M}^{-1/2}\right] \overset{P}{\to} 0,$$  \hspace{1cm} (17)

and with high probability,

$$\|R_1 S_{M}^{1/2} + U_\tilde{P} R_2\|_F \leq \frac{Cm^{1/2} \log mn}{n^{1/2}}.$$

**Proof.** We begin by observing that

$$\|R_1 S_{M}^{1/2} + U_\tilde{P} R_2\|_F \leq \|R_1\|_F \|S_{M}^{1/2}\| + \|R_2\|_F.$$

Lemma 3 and the trivial upper bound on the eigenvalues of $M$ ensures that

$$\|R_1\|_F \|S_{M}^{1/2}\| \leq \frac{Cm^{1/2} \log mn}{n^{1/2}} \text{ w.h.p. },$$
Combining this with Equation \[13\], we conclude that

\[
\|R_1 S_M^{1/2} + U_P R_2\|_F \leq \frac{C m^{1/2} \log mn}{n^{1/2}} \text{ w.h.p.}
\]

We will establish \[15\], \[16\] and \[17\] order. To see \[15\], observe that

\[
\sqrt{n}\|(M - \hat{P}) U_P (VS_M^{-1/2} - S_P^{-1/2} V)\|_F \leq \sqrt{n}\|(M - \hat{P}) U_P \|\|VS_M^{-1/2} - S_P^{-1/2} V\|_F,
\]

and application of Proposition \[1\] and Lemma \[2\] imply that with high probability

\[
\sqrt{n}\|(M - \hat{P}) U_P (VS_M^{-1/2} - S_P^{-1/2} V)\|_F \leq C \sqrt{\frac{\log mn}{mn^2}},
\]

which goes to 0 as \(n \to \infty\).

To show the convergence in \[16\], we recall that \(U_P S_P^{1/2} = ZW^T\), and observe that since the rows of the latent position matrix \(Z\) are necessarily bounded in Euclidean norm by 1, and since the top \(d\) eigenvalues of \(P\) are of order \(mn\), it follows that

\[
\|U_P\|_2 \to \infty \leq C (mn)^{-1/2} \text{ w.h.p.} \quad (18)
\]

Next, Proposition \[2\] and Observation \[2\] imply that

\[
\|(U_P U_P^T (M - \hat{P}) U_P VS_M^{-1/2})_h\| \leq \|U_P\|_2 \to \infty \|U_P^T (M - \hat{P}) U_P \|\|S_M^{-1/2}\|
\]

\[
\leq C \log^{1/2} \frac{mn}{m^{1/2} n} \text{ w.h.p.,}
\]

which implies \[16\].

Finally, to establish \[17\], we must bound the Euclidean norm of the vector

\[
\left[(I - U_P U_P^T) (M - \hat{P}) R_3 S_M^{-1/2}\right]_h,
\]

where, as defined above, \(R_3 = U_M - U_P V\). Let \(B_1\) and \(B_2\) be defined as follows:

\[
B_1 = (I - U_P U_P^T) (M - \hat{P}) (I - U_P U_P^T) U_M S_M^{-1/2}
\]

\[
B_2 = (I - U_P U_P^T) (M - \hat{P}) U_P (U_P^T U_M - V) S_M^{-1/2}
\]

Recalling that \(R_3 = U_M - U_P V\), we have

\[
(I - U_P U_P^T) (M - \hat{P}) R_3 S_M^{-1/2} = (I - U_P U_P^T) (M - \hat{P}) (U_M - U_P U_P^T U_M) S_M^{-1/2} + (I - U_P U_P^T) (M - \hat{P}) (U_P U_P^T U_M - U_P V) S_M^{-1/2}
\]

\[
= B_1 + B_2.
\]

We will bound the Euclidean norm of the \(h\)-th row of each of these two matrices on the right-hand side, from which a triangle inequality will yield our desired bound on the quantity in
Equation (19). Recall that we use $C$ to denote a positive constant, independent of $n$ and $m$, which may change from line to line.

Let us first consider $B_2 = (I - U_P U_P^T)(M - \hat{P})U_P(U_P^T U_M - V)S_M^{-1/2}$. We have

$$\|B_2\|_F \leq \|(I - U_P U_P^T)(M - \hat{P})U_P\|\|U_P^T U_M - V\|_F \|S_M^{-1/2}\|.$$  

By submultiplicativity of the spectral norm and Lemma 2, $\|(I - U_P U_P^T)(M - \hat{P})U_P\| \leq Cmn^{1/2}\log^{1/2} mn$ with high probability. From Lemma 3 and Observation 2, respectively, we have with high probability

$$\|U_P^T U_M - V\|_F \leq Cn^{-1}\log mn \quad \text{and} \quad \|S_M^{-1/2}\| \leq C(mn)^{-1/2}.

Thus, we deduce that with high probability,

$$\|B_2\|_F \leq \frac{Cm^{1/2}\log^{3/2} mn}{n}$$  

which follows that $\|\sqrt{n}B_2\|_F \xrightarrow{P} 0$, and hence $\|\sqrt{n}(B_2)_{h}\| \xrightarrow{P} 0$.

Turning our attention to $B_1$, and recalling that $U_M^T U_M = I$, we note that

$$\|(B_1)_{h}\| = \left\|[\begin{smallmatrix} I - U_P U_P^T & (M - \hat{P}) & (I - U_P U_P^T) U_M S_M^{-1/2} \end{smallmatrix}]\right\|_h$$

$$= \left\|[\begin{smallmatrix} (I - U_P U_P^T)(M - \hat{P}) & (I - U_P U_P^T) U_M U_M^T U_M S_M^{-1/2} \end{smallmatrix}]\right\|_h$$

$$\leq \|U_M S_M^{-1/2}\| \left\|[\begin{smallmatrix} I - U_P U_P^T & (M - \hat{P}) & (I - U_P U_P^T) U_M U_M^T \end{smallmatrix}]\right\|_h.$$  

Let $\epsilon > 0$ be a constant. We will show that

$$\lim_{n \to \infty} \Pr \left[ \|\sqrt{n}(B_1)_{h}\| > \epsilon \right] = 0. \tag{22}$$  

For ease of notation, define

$$E_1 = (I - U_P U_P^T)(M - \hat{P})(I - U_P U_P^T) U_M U_M^T.$$  

We will show that

$$\lim_{n \to \infty} \Pr \left[ \sqrt{n} \|[E_1]_{h}\| > n^{1/4} \right] = 0, \tag{23}$$  

which will imply (22) since, by Observation 2, $\|U_M S_M^{-1/2}\| \leq C(mn)^{-1/2}$ w.h.p.

Let $Q \in \mathbb{R}^{mn \times mn}$ be any permutation matrix. We observe that

$$QU_P U_P^T Q \hat{P} Q^T = Q \hat{P} Q^T,$$  

24
and thus $Q U_p U_p^T Q^T$ is a projection matrix for $Q P Q^T$ if and only if $U_p U_p^T$ is a projection matrix for $P$. A similar argument applies to the matrix $U_M U_M^T$. Combining this with the exchangeability structure of the matrix $M - P$, it follows that the Frobenius norms of the rows of $E_1$ are equidistributed. This row-exchangeability for $E_1$ implies that $mn \mathbb{E} \| (E_1)_h \|^2 = \| E_1 \|_F^2$. Applying Markov’s inequality,

$$\Pr \left[ \| \sqrt{n} [E_1]_h \| > t \right] \leq \frac{n \mathbb{E} \| (I - U_p U_p^T)(M - P)(I - U_p U_p^T)U_M U_M^T \|_h^2}{t^2} \leq \frac{C m \log mn}{mt^2} \tag{24}$$

We will proceed by showing that with high probability,

$$\| (I - U_p U_p^T)(M - P)(I - U_p U_p^T)U_M U_M^T \|_F \leq C m \log mn, \tag{25}$$

whence choosing $t = n^{1/4}$ in (24) yields that

$$\lim_{n \to \infty} \Pr \left[ \| \sqrt{n} [I - U_p U_p^T(M - P)(I - U_p U_p^T)U_M U_M^T]_h \| > n^{1/4} \right] = 0,$$

and (22) will follow. We have

$$\| (I - U_p U_p^T)(M - P)(I - U_p U_p^T)U_M U_M^T \|_F \leq \| M - P \| \| U_M - U_p U_p^T U_M \| \| U_M \|$$

Theorem 2 implies that the first term in this product is at most $C m n^{1/2} \log^{1/2} mn$ with high probability, and the final term in this product is, trivially, at most 1. To bound the second term, we will follow reasoning similar to that in Lemma 4, combined with the Davis-Kahan theorem. The Davis-Kahan Theorem ([Davis and Kahan, 1970] [Bhatia, 1997]) implies that for a suitable constant $C > 0$,

$$\| U_M U_M^T - U_p U_p^T \| \leq \frac{C \| M - P \|}{\lambda_d(P)}.$$

By Theorem 2 in Yu et al. (2015), there exists orthonormal $W \in \mathbb{R}^{d \times d}$ such that

$$\| U_M - U_p W \|_F \leq C \| U_M U_M^T - U_p U_p^T \|_F.$$

We observe further that the multivariate linear least squares problem

$$\min_{T \in \mathbb{R}^{d \times d}} \| U_M - U_p T \|_F^2$$

is solved by $T = U_p^T U_M$. Thus, combining all of the above,

$$\| U_M - U_p U_p^T U_M \|_F^2 \leq \| U_M - U_p W \|_F^2 \leq C \| U_M U_M^T - U_p U_p^T \|_F^2 \leq C \| U_M U_M^T - U_p U_p^T \|_F^2 \leq C \frac{\| M - P \|}{\lambda_d(P)} \leq \frac{C \log^{1/2} mn}{n^{1/2}} \text{ w.h.p.}$$
Thus, we have
\[
\left\| (I - U_P U_P^T) (M - \tilde{P}) (I - U_P U_P^T) U_M U_M^T \right\|_F \leq \left\| M - \tilde{P} \right\|_F \left\| U_M - U_P U_P^T U_M \right\|_F \\
\leq C m \log mn \text{ w.h.p. ,}
\]
which implies (25), as required, and thus the convergence in (17) is established, completing the proof.

We are now ready to prove Lemma 1 on the consistency of the omnibus embedding; that is, we can now prove that there exists an orthogonal matrix $\tilde{W} \in \mathbb{R}^{d \times d}$ such that with high probability,
\[
\| U_M S_M^{1/2} - U_P S_P^{1/2} \tilde{W} \|_{2 \to \infty} \leq \frac{C m^{1/2} \log mn}{\sqrt{n}}.
\]

**Proof of Lemma 1.** Observe that
\[
U_M S_M^{1/2} - U_P S_P^{1/2} \tilde{V} = (M - \tilde{P}) U_P S_P^{-1/2} \tilde{V} + (M - \tilde{P}) U_P (V S_M^{-1/2} - S_P^{-1/2} \tilde{V}) \\
- U_P U_P^T (M - \tilde{P}) U_P V S_M^{-1/2} \\
+ (I - U_P U_P^T) (M - \tilde{P}) R_3 S_M^{-1/2} + R_3 S_M^{1/2} + U_P R_2.
\]  
(26)

With $B_1$ and $B_2$ defined in Equation (20), the arguments in the proof of Lemma 5 imply that with high probability
\[
\| (M - \tilde{P}) U_P (V S_M^{-1/2} - S_P^{-1/2} \tilde{V}) \| \leq C m^{1/2} n^{-1} \log^{1/2} mn \\
\| U_P U_P^T (M - \tilde{P}) U_P V S_M^{-1/2} \|_F \leq C n^{-1/2} \log^{1/2} mn \\
\| (I - U_P U_P^T) (M - \tilde{P}) R_3 S_M^{-1/2} \|_F \leq \| B_1 \|_F + \| B_2 \|_F \\
\leq C n^{-1/2} m^{1/2} \log mn + C m^{1/2} n^{-1} \log^{3/2} mn
\]

As a consequence, there exists an orthogonal matrix $\tilde{W}$ such that
\[
\| U_M S_M^{1/2} - U_P S_P^{1/2} \tilde{W} \|_F \leq \| (M - \tilde{P}) U_P S_P^{-1/2} \|_F + \frac{C m^{1/2} \log mn}{\sqrt{n}} \text{ w.h.p.}
\]

From this, we deduce that
\[
\max_i \| (U_M S_M^{1/2} - U_P S_P^{1/2} \tilde{W})_i \| \leq \frac{1}{\lambda_d(\tilde{P})} \max_i \| (M - \tilde{P}) U_P)_i \| + \frac{C m^{1/2} \log mn}{\sqrt{n}} \text{ w.h.p.}
\]

Standard application of Hoeffding’s inequality as in Proposition 1 shows that with high probability,
\[
\max_i \| (M - \tilde{P}) U_P)_i \| \leq C \left( m^{1/2} + \log^{1/2} mn \right).
\]
The desired bound follows from Observation 1 applied to $\lambda_d(\tilde{P})$. 

26
Lemma 6. Fix some discussed previously, the true latent positions $X \in \mathbb{R}^{n \times 1}$ where we have used the fact that $Z$. Hence for each $\Sigma$, we denote the matrix of these “true” latent positions by $Z = (X^T, X^T, \ldots, X^T)^T \in \mathbb{R}^{mn \times d}$, so that $Z = Z^*W$ for some suitably-chosen orthogonal matrix $W$.

Recall that we denote the matrix of these “true” latent positions by $Z = (X^T, X^T, \ldots, X^T)^T \in \mathbb{R}^{mn \times d}$, so that $Z = Z^*W$ for some suitably-chosen orthogonal matrix $W$.

Lemma 6. Fix some $i \in [n]$ and some $s \in [m]$ and let $h = m(s-1) + i$. Conditional on $X_i = x_i \in \mathbb{R}^d$, there exists a sequence of $d$-by-$d$ orthogonal matrices $\{W_n\}$ such that

$$n^{1/2} W_n^T [(M - \hat{P})U P^{-1/2}] \xrightarrow{h} N(0, \Sigma(x_i)),$$

where $\Sigma(x_i) \in \mathbb{R}^{d \times d}$ is a covariance matrix that depends on $x_i$.

Proof. For each $n = 1, 2, \ldots$, choose orthogonal $W_n \in \mathbb{R}^{d \times d}$ so that $X = X^*W_n$ (and hence $Z = Z^*W_n$, as well). At least one such $W_n$ exists for each value of $n$, since, as discussed previously, the true latent positions $X$ are specified only up to some rotation $X = U P^{1/2}W = X^*W$. We have

$$n^{1/2} W_n^T [(M - \hat{P})U P^{-1/2}] = n^{1/2} W_n^T [MZ P^{-1} - \hat{P} Z P^{-1}] \xrightarrow{h} N(0, \Sigma(x_i)),$$

where we have used the fact that $P = mS_P$.

Recalling the structure of $Z = Z^*W_n$ (see Equation (5)) and recalling that $X_j = (X_j)^T$, we have

$$n^{1/2} W_n^T [(M - \hat{P})U P^{-1/2}] = \frac{n^{1/2} W_n^T S_P^{-1} W_n}{m} \left( \frac{1}{m} \sum_{j=1}^m \left( A_{ij}^{(s)} + A_{ij}^{(q)} \right) - P_{ij} \right) X_j$$

$$= \frac{n^{1/2} W_n^T S_P^{-1} W_n}{m} \left( \sum_{j \neq i} \left( \frac{m+1}{2} (A_{ij}^{(s)} - P_{ij}) + \frac{1}{m} \sum_{q \neq s} A_{ij}^{(q)} - P_{ij} \right) X_j \right)$$

$$= \left( n^{1/2} W_n^T S_P^{-1} W_n \right) \left( \sum_{j \neq i} \left( \frac{m+1}{2} (A_{ij}^{(s)} - P_{ij}) + \frac{1}{m} \sum_{q \neq s} A_{ij}^{(q)} - P_{ij} \right) X_j \right)$$

$$- n^{1/2} W_n^T S_P^{-1} W_n P_{ii} X_i \frac{1}{n^{1/2}}.$$
Conditioning on \( X_i = x_i \in \mathbb{R}^d \), we first observe that

\[
\frac{P_{ii}}{n^{1/2}} X_i = \frac{x_i^T x_i}{n^{1/2}} x_i \to 0 \text{ a.s.}
\]

(27)

further, the scaled sum

\[
\begin{align*}
\frac{n^{-1/2}}{} \sum_{j \neq i} & \left( \frac{m+1}{2m} (A_{ij}^{(s)} - P_{ij}) + \frac{1}{m} \sum_{q \neq s} A_{ij}^{(q)} - P_{ij} \right) X_j \\
& = \frac{n^{-1/2}}{} \sum_{j \neq i} \left( \frac{m+1}{2m} (A_{ij}^{(s)} - x_j^T x_i) + \frac{1}{m} \sum_{q \neq s} A_{ij}^{(q)} - x_j^T x_i \right) X_j
\end{align*}
\]

is a sum of \( n - 1 \) independent 0-mean random variables, each with covariance matrix given by

\[
\tilde{\Sigma}(x_i) = \frac{m+3}{4m} \mathbb{E} \left[ (x_i^T X_j - (x_i^T X_j)^2) X_j X_j^T \right].
\]

The multivariate central limit theorem thus implies that

\[
\frac{n^{-1/2}}{} \sum_{j \neq i} \left( \frac{m+1}{2m} (A_{ij}^{(s)} - x_j^T x_i) + \frac{1}{m} \sum_{q \neq s} A_{ij}^{(q)} - x_j^T x_i \right) X_j \overset{\mathcal{L}}{\to} \mathcal{N}(0, \tilde{\Sigma}(x_i)).
\]

(28)

By the strong law of large numbers,

\[
\frac{1}{n} X^T X - \Delta \to 0 \text{ a.s.}
\]

However, we also have

\[
\frac{1}{n} (X^*)^T X^* - W_n \Delta W_n^T = W_n \left( \frac{1}{n} X^T X - \Delta \right) W_n^T \to 0 \text{ a.s. ,}
\]

and \( S_P = S_P^{1/2} U_P U_P S_P^{1/2} = (X^*)^T X^* \). Thus,

\[
\frac{1}{n} S_P - W_n \Delta W_n^T \to 0 \text{ a.s.}
\]

Since all matrices involved are order \( d \), which is fixed in \( n \), the convergences in the preceding three equations can be thought of either as element-wise or under any matrix norm. In particular, we have \( \|\frac{1}{n} S_P - W_n \Delta W_n^T\| \to 0 \), whence Weyl’s inequality [Horn and Johnson, 1985] implies that the eigenvalues of \( S_P/n \) converge to those of \( \Delta \). Since both \( S_P/n \) and \( \Delta \) are diagonal, this implies that \( S_P/n \to \Delta \). We note that in the case where \( \Delta \) has distinct diagonal entries, this implies that \( W_n \to I \) as in Athreya et al. (2016), though in the case where \( \Delta \) has repeated eigenvalues, no such convergence is guaranteed. Thus we have shown
that \( nW_n^T S_{P}^{-1} W_n \to \Delta^{-1} \) almost surely. Combining this fact with (28), the multivariate version of Slutsky’s theorem yields

\[
\lim_{n \to \infty} \Pr \left[ \sqrt{n}W_n^T \left[(M - \tilde{P})U_P S_{P}^{-1/2}\right]_h \leq x \right] = \int_{\text{supp } F} \Phi \left( x, \Sigma(x) \right) dF(y).
\]

Now consider \( U_M S_{M}^{1/2} V^T - U_P S_{P}^{1/2} \). From Equation (29), we have

\[
\left(U_M S_{M}^{1/2} V^T - U_P S_{P}^{1/2}\right) W_n = (M - \tilde{P})U_P S_{P}^{-1/2} W_n + HV^T W_n,
\]
where
\[
H = (M - \tilde{P})U_P(\sqrt{V}S_{M}^{-1/2} - S_{P}^{-1/2}V) - U_PU_P^T(M - \tilde{P})U_P\sqrt{V}S_{M}^{-1/2} \\
+ (I - U_PU_P^T)(M - \tilde{P})R_3S_{M}^{-1/2} + R_1S_{M}^{-1/2} + U_PR_2.
\] (30)

Since $V^TW_n$ is unitary, it suffices to show that the $h$-th row of $H$, as defined in (30), when multiplied by $\sqrt{n}$, goes to 0 in probability, from which Slutsky’s Theorem will yield our desired result. This is precisely the content of Lemma 5, except that we need to establish the following convergence in probability:
\[
\sqrt{n} \left[ \left( R_1S_{M}^{1/2} + U_PR_2 \right) \right]_h \overset{P}{\to} 0.
\] (31)

We recall that by Lemma 3 and Equation (18),
\[
\|R_1\|_{2 \to \infty} \leq \|U_P\|_{2 \to \infty}\|U_P^TU_M - V\| \leq \frac{C\log mn}{n^{3/2}} \text{ w.h.p.}
\]
Combining this with Observation 2 and Lemma 4 along with Equation (18) again,
\[
\|\left( R_1S_{M}^{1/2} + U_PR_2 \right)\|_h \leq \|R_1\|_{2 \to \infty}\|S_{M}^{1/2}\| + \|U_P\|_{2 \to \infty}\|R_2\| \\
\leq \frac{C\log mn}{n} \text{ w.h.p. ,}
\]
from which the convergence in (31) follows, completing the proof. \qed