PhD Candidacy Exam Prospectus

Limit theorems for omnibus embeddings of multiple random graphs

Konstantinos Pantazis

Advisor:
Vince Lyzinski

Contents

1 Introduction **2**

2 Mathematical Background **3**

2.1 Random Dot Product Graph model (RDPG) 3

2.2 Adjacency Spectral Embedding on RDPG models 4

3 Previous work **7**

3.1 CLT for an Omnibus Embedding of JRDPGs 7

4 Present work **10**

4.1 CLT for estimated eigenvectors of Correlated-ER graphs 10

4.2 CLT for a dampened omnibus embedding of RDPGs 12

5 Future work **13**

5.1 CLT for directed omnibus embeddings 13

1 Introduction

Statistical inference across multiple graphs is of great import in many different scientific fields ranging from example, neuroscience to social sciences and machine learning to telecommunications. Inference on random graphs often depends on low-dimensional Euclidean representations of their vertices, typically given by spectral decompositions of the vertices of adjacency matrices. This enables the machinery of classical statistics to be directly applied to the embedded networks. However, simultaneously embedding multiple networks into the same Euclidean space is a non-trivial task due to the nonidentifiability (embeddings are unique up to orthogonal transforms) inherent to many common random graph models. Numerous methods for this simultaneous embedding tasks have been introduced, including those in [8, 1, 11, 21].

Of particular note is the the omnibus embedding from Levin et al., a multi-scale, multiple graph inference method introduced in [8]. In this approach, multiple graphs on the same vertex set are jointly embedded into a single low-dimensional space with a distinct low-dimensional representation for each vertex of each graph. The omnibus embedding both generates consistent estimates of the latent positions within and across all graphs in the Random Dot Product Graph setting (see [2] for a survey on RDPGs) and provides consistent testing (with high empirical power) for distinguishing when graphs are statistically different across the sample.

The layout will be as follows. In Chapter 2, we describe an appropriate generative model for latent position graphs called the Random Dot Product Graph model. We also discuss the Adjacency Spectral Embedding (ASE) of a network, derived from the eigen-decomposition of an adjacency matrix, which is the key tool for our network embeddings. In Chapter 3, we present the omnibus method of Levin et al. [8], and we define the omnibus embedding and state a key Central Limit Theorem for ASE-derived estimates of its latent positions. This result implies that omnibus embeddings can be deployed for subsequent network inference tasks including consistent classification [19, 10], consistent clustering [16, 9], hypothesis testing and community detection [17, 1, 10]. In Chapter 4, a similar central limit theorem with a different omnibus structure is presented, as well as a central limit theorem between estimated latent positions under a dependent-edge random graph model. In Chapter 5, future directions are discussed. Specifically, proving similar limit theorems in the directed network case, i.e., when the omnibus matrix is asymmetric.

2 Mathematical Background

We introduce the notation being used throughout this text. For a positive integer n, we let $[n] = \{1, 2, \cdots, n\}$, and let $J_n \in \mathbb{R}^{n\times n}$ be the matrix with all of its entries identically equal to one. The set of $d \times d$ real orthogonal matrices is denoted by $\mathcal{O}_d$. The Kronecker product is denoted by $\otimes$ symbol. We represent a simple (no self-loops or multiple edges), un-weighted and un-directed graph as the ordered pair $G = (V, E)$, where $V = [n]$ the set of nodes and $E \subset \binom{n}{2}$ the set of edges of the graph. We denote by $A \in \{0, 1\}^{n\times n}$ the adjacency matrix of the graph, i.e., A_{ij} is equal to 1 if there exists an edge between nodes i and j, and 0 otherwise. Where there is no danger of confusion, we will often refer to a graph G and its adjacency matrix A interchangeably. Finally, the symbols $\|\cdot\|_F, \|\cdot\|, \|\cdot\|_{2\to\infty}$ correspond to the Frobenius, spectral and two-to-infinity norms respectively.

2.1 Random Dot Product Graph model (RDPG)

To provide a a theoritically-principled paradigm for graph inference, we focus on a particular class of random graphs, namely latent positions random graphs [6]. In this model, every vertex has associated to it a (unobserved) latent position belonging to some topological (latent) space. More specifically, we are interested in a latent position graph model called d-dimensional Random Dot Product Graph (RDPG) model [22]. In RDPGs, the latent space is a subspace of $\mathbb{R}^d$, thus the latent positions are vectors. Conditional on the latent positions, edges are independently sampled and the probability of an edge between two vertices i and

j is given by the dot product of their associated latent vectors. The formal definition is as follows.

Definition 2.1. Let F be a distribution on $\mathbb{R}^d$, where $x^T y \in [0,1]$ for all $x, y \in \text{supp}(F) \subset \mathbb{R}^d$. Let $X_1, X_2, \cdots, X_n \overset{i.i.d.}{\sim} F$ represent the rows of the matrix $X = [X_1^T | X_2^T | \cdots | X_n^T]^T \in \mathbb{R}^{n \times d}$ and define $P = XX^T \in [0,1]^{n \times n}$. We write that $(A, X) \sim \text{RDPG}(F, n)$ when $A \in \{0,1\}^{n \times n}$ is the adjacency matrix such that for all $i < j$, conditioned on X_i, X_j,

$$A_{ij} \overset{ind.}{\sim} \text{Bernoulli}(X_i^T X_j).$$

We say that A is the adjacency matrix of a random dot product graph with latent positions given by the rows of X.

RDPGs belong to a wider class of random graphs called *latent position* random graphs [6, 4, 14], in which every vertex's latent position belongs to some space $\mathcal{X}$(not necessarily Euclidean), and the probability of any two vertices i, j, p_{ij}, is given by a function $\kappa : \mathcal{X} \times \mathcal{X} \mapsto [0,1]$ (known as the link function) of their associated latent positions (x_i, x_j). Thus, $p_{ij} = \kappa(x_i, x_j)$.and given these latent positions, $A_{ij} \overset{ind}{\sim} \text{Bernoulli}(p_{ij})$.

Note that we often assume that matrix $P \in [0,1]^{n \times n}$ is of low rank $d << n$, and in this low-rank setting, random dot product graphs are a fairly general class of conditionally independent-edge random graphs, containing, for example, Erdős-Rényi [5] and positive semidefinite stochastic block models [7] as submodels.

2.2 Adjacency Spectral Embedding on RDPG models

Oliveira [12] shows that under mild conditions, if $(A, X) \sim \text{RDPG}(F, n)$ then the spectral norm of $A - P$ is comparatively small. In particular, with high probability we have that

$$\|A - P\| \leq C \log n,$$

for some constant $C > 0$. Since we want to estimate the unobserved latent positions of X, we introduce the following spectral decomposition of A,

Definition 2.2 (Adjacency spectral embedding)**.** 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 singular values, obtained by $ASE(A, d) = U_A S_A^{1/2}$, where $S_A \in \mathbb{R}^{d\times d}$ is a diagonal matrix whose entries are the top d eigenvalues of $|A| = (A^T A)^{1/2}$ (in non-decreasing 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 1. Note that the RDPG model is non-identifiable; it is specified only up to orthogonal rotation of its latent positions. To see this, if $X \in \mathbb{R}^{n\times d}$ is a matrix of latent positions and $W \in \mathcal{O}_d$ is orthogonal, then $P = XX^T = XWW^TX^T = (XW)(XW)^T$. As a concequence, the ASE provides a consistent estimate for the true latent positions up to orthogonal transformations.

Under the setting above, Sussman et al. [16] shows that the latent positions of a random dot product graph can be consistently estimated (up to an orthogonal transformation) by the eigen-decomposition of its adjacency matrix, given the latent positions are sampled independently from the same distribution. Explicitly, Theorem 4.1 in [16] states that with probability at least $1 - 2\eta$, there exists an orthogonal matrix $W \in \mathcal{O}_d$ such that

$$\|U_A S_A^{1/2} - WX\|_F \leq C\sqrt{\log\left(\frac{n}{\eta}\right)},$$

where $C > 0$ is some fixed constant.

Theorem 5.1 in Lyzinski et al. [10] gives a bound on the difference between the estimated latent positions obtained by the ASE and the true latent positions using the two-to-infinity norm. Under the same setting, with high probability there exists an orthogonal matrix

$W \in \mathcal{O}_d$ such that

$$\|U_A S_A^{1/2} - WX\|_{2\to\infty} \leq C\frac{\log^2 n}{\sqrt{n}}.$$

The two-to-infinity norm yields finer uniform control than the more common spectral and Frobenius norms, as it ensures all estimated latent positions are uniformly close to the corresponding true latent positions. As we are interested in inference on the estimates of the latent positions (i.e., estimated rows of the data matrix $X \in \mathbb{R}^{n\times d}$), the finer control provided by two-to-infinity norm bounds is often preferential for proving results about the down-stream inference task. Moreover, Cape et al. [3] recently developed technical and theoretical tools for studying the geometry of singular sub-spaces using the two-to-infinity norm, and this machinery provides us with sharp bounds on latent position estimates under a variety of noise conditions, and provides a valuable tool for statistical graph inference in our setting.

It has been suggested in [13] that for two-block stochastic blockmodels , for large regimes of the parameter space, the normalized Laplacian spectral embedding is to be preferred over that of the adjacency matrix for subsequent inference. In [20], the authors used the normalized Laplacian matrix of a RDPG (instead of the adjacency matrix) to show consistency of the latent positions via an analogous embedding called Laplacian spectral embedding (LSE). Further, they proved an analogous limit theorem as in Theorem 1 for the Laplacian spectral embedding in the $m = 1$ case, and they compared how ASE and LSE impact subsequent inference tasks.

3 Previous work

3.1 CLT for an Omnibus Embedding of JRDPGs

The main focus of [8] is multiscale graph inference, and in this section, we illustrate the main tools and results [8] towards this end. We first give a definition of a multiple random graph model called the Joint Random Dot Product Graph (JRDPG) model, and we present a definition of the omnibus multiple graph embedding. Finally, we present a central limit theorem for the latent position estimates presented by the omnibus embedding.

Definition 3.1 (Joint Random Dot Product Graph(JRDPG))**.** Let F be a distribution on $\mathbb{R}^d$, where $x^T y \in [0,1]$ for all $x, y \in \text{supp}(F) \subset \mathbb{R}^d$. Let $X_1, X_2, \cdots, X_n \overset{i.i.d.}{\sim} F$ represent the rows of matrix $X = [X_1^T | X_2^T | \cdots | X_n^T]^T \in \mathbb{R}^{n \times d}$ and define $P = XX^T \in [0,1]^{n \times n}$. A collection of m random graphs (represented by their adjacency matrices $A^{(1)}, A^{(2)}, \cdots, A^{(m)}$) is distributed as a joint random dot product graph, written $(A^{(1)}, A^{(2)}, \cdots, A^{(m)}, X) \sim \text{JRDPG}(F, n, m)$, if $(A^{(k)}, X) \sim \text{RDPG}(F, n)$ for each $k = 1, 2, \cdots, m$. That is, for each $k \in [m]$, $A^{(k)}$ is conditionally independent given X, with edges independently distributed as

$$A_{ij}^{(k)} \sim \text{ Bernoulli}(X_i^T X_j)$$

for all $i < j$, $i, j \in [n]$. Note that the $A^{(i)}$'s are independent graphs, and the same X is used to generate all m graphs.

The omnibus embedding is simply the spectral decomposition of an omnibus matrix M, i.e., M is a block matrix containing multiple adjacency matrices. As stated in Levin et al [8], the omnibus embedding has numerous favorable properties. For multiple-graph hypothesis testing, it allows the comparison of graphs without the need to perform pairwise alignments of the embeddings of different graphs, as in [18]. In other words, the Procrustes alignment in the case of separately embedded graphs is unavoidable due to the non-identifiability, whereas an alignment is incorporated a priori in the case of the omnibus matrix. In addition, the fact that the omnibus embedding produces m distinct estimated latent positions for each vertex, combined with the asymptotic normality of these estimated latent positions (see Theorem 1) provide a test for the incorporated alignment of the omnibus matrix.

Definition 3.2 (Omnibus Embedding)**.** Let $A^{(1)}, A^{(2)}, \cdots, A^{(m)} \in \mathbb{R}^{n\times n}$ be adjacency matrices of a collection of m vertex-aligned, undirected graphs. We define the mn-by-mn omnibus matrix of $A^{(1)}, A^{(2)}, \cdots, A^{(m)}$ by

$$M = \begin{bmatrix} A^{(1)} & \frac{A^{(1)}+A^{(2)}}{2} & \frac{A^{(1)}+A^{(3)}}{2} & \cdots & \frac{A^{(1)}+A^{(m)}}{2} \\ \frac{A^{(2)}+A^{(1)}}{2} & A^{(2)} & \frac{A^{(2)}+A^{(3)}}{2} & \cdots & \frac{A^{(2)}+A^{(m)}}{2} \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ \frac{A^{(m)}+A^{(1)}}{2} & \frac{A^{(m)}+A^{(2)}}{2} & \frac{A^{(m)}+A^{(3)}}{2} & \cdots & A^{(m)} \end{bmatrix},$$

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

$$OMNI(A^{(1)}, A^{(2)}, \cdots, A^{(m)}, d) = ASE(M, d) = U_M S_M^{1/2} \in \mathbb{R}^{mn\times d},$$

where $S_M \in \mathbb{R}^{d\times d}$ is a diagonal matrix whose entries are the top d of eigenvalues of $|M|= (M^T M)^{1/2}$ (in nondecreasing order) and $U_M \in \mathbb{R}^{mn\times d}$ is the matrix whose columns are the orthonormal eigenvectors corresponding to the eigenvalues in S_M.

In Levin et al. [8], the authors prove consistency and asymptotic normality of the estimated latent positions provided by ASE applied to the omnibus matrix M. Below, we provide the asymptotic normality of the residuals which imply the asymptotic distribution of the estimated latent positions under the JRDPG model.

Theorem 1. *(CLT for an Omnibus Embedding of JRDPGs)*
Let $(A^{(1)}, A^{(2)}, \cdots, A^{(m)}, X) \sim \text{JRDPG}(F, n, m)$ and let M denote the omnibus matrix as in Definition 3.2. Let $\tilde{P} = U_{\tilde{P}} S_{\tilde{P}} U_{\tilde{P}}^T$ be the spectral decomposition of $\tilde{P}$, where $\tilde{P} = P \otimes J_m \in$

$\mathbb{R}^{mn\times mn}$, and where $S_{\tilde{P}} \in \mathbb{R}^{d\times d}$ is a diagonal matrix whose diagonal entries are eigenvalues (in non-decreasing order), and $U_{\tilde{P}} \in \mathbb{R}^{mn\times d}$ is the orthogonal matrix whose columns are the eigenvectors corresponding to the eigenvalue in $S_{\tilde{P}}$. If $h = n(s-1)+i$ where $i \in [n], s \in [m]$, then there exists a sequence of orthogonal matrices $\{W_n\}$ such that

$$n^{1/2} W_n^T [U_M S_M^{1/2} - U_{\tilde{P}} S_{\tilde{P}}^{1/2}]_h \xrightarrow{\mathcal{D}} \mathcal{N}(0, \Sigma(X_i)),$$

where $\Sigma(y) = (m+3)\Delta^{-1}\widetilde{\Sigma}(y)\Delta^{-1}/4m$, $\Delta = \mathbb{E}[X_1 X_1^T] \in \mathbb{R}^{d\times d}$ and $\widetilde{\Sigma}(y) = \mathbb{E}[(y^T X_1 - (y^T X_1)^2) X_1 X_1^T]$.

In Arroyo et al. [1], the authors introduce a new model for multi-scale graph inference, the so called common subspace independent-edge (COSIE) multiple random graph model, which describes a heterogeneous collection of networks with a shared latent structure on the vertices but potentially different connectivity patterns for each graph. In particular, the COSIE model describes a class of independent-edge random graphs $G^{(1)}, \cdots, G^{(m)}$ whose Bernoulli adjacency matrices $A^{(i)}$, $1 \le i \le m$, have expectation of the form $P = VR^{(i)}V^T$, where $R^{(i)}$ is a low-dimensional matrix and V is a matrix of orthonormal columns.

Furthermore, by applying a joint adjacency spectral embedding–the multiple adjacency spectral embedding(MASE)–to their model, they show consistent estimation of underlying parameters for each graph, and under mild conditions, MASE estimates satisfy asymptotic normality and yield improvements for graph eigenvalue estimation. As a result, the COSIE model combined with the MASE embedding can be deployed for a number of subsequent network inference tasks, including dimensionality reduction, classification, hypothesis testing and community detection.

The COSIE model and the MASE embedding allow graphs to (possibly) have distinct underlying distributions, whereas the omnibus embedding assumes that all the graphs share the same underlying distribution. On the other hand, the omnibus embedding can identify whole-graph differences as well as vertex-level differences, since a set of estimated latent positions is formed for each vertex ($U_M \in \mathbb{R}^{mn\times d}$), whereas the MASE embedding generates a single estimated latent position for each vertex ($V \in \mathbb{R}^{n\times d}$), making this approach rely on other methods for testing, such as the parametric bootstrap approach [18].

4 Present work

Currently, we are interested in expanding the theoretical results of the previous work using more general structures for the omnibus matrix, as well as proving a limit theorem for dependent-edge random graphs.

4.1 CLT for estimated eigenvectors of Correlated-ER graphs

In this subsection, we first give a model for correlated random graph pairs, and then we state a central limit theorem for a pair of graphs generated under this model. The random correlated Erdos-Renyi model is defined as follows.

Definition 4.1 (Correlated Erdos-Renyi (CorrER) random graph model)**.** Let $\mathcal{A}_n$ denote the set of adjacency matrices corresponding to simple, undirected graphs, i.e. the set of symmetric, hollow $\{0,1\}$-valued $n \times n$ matrices. Define $P \in [0,1]^{n\times n}$ to be the probability matrix and $R \in [0,1]^{n\times n}$ to be a metrix of entry-wise correlations. We say that a pair of adjacency matrices $(A, B) \sim \mathrm{CorrER}(P, R)$ if $A, B \in \mathcal{A}_n$, and for each $u < v$, B_{uv} are independent with $B_{uv} \sim \mathrm{Bernoulli}(P_{uv})$ and A_{uv} are independent with $A_{uv} \sim \mathrm{Bernoulli}(P_{uv})$. Additionally, $\{B_{uv}\}$ and $\{A_{u'v'}\}$ are collectively independent except that for $u, v \in [n]$ with $u < v$, it holds that $\mathrm{correlation}(A_{uv}, B_{uv}) = R_{uv}$.

Our next theorem explores the impact of the correlation matrix R on the relative ASE's of A and B. Note that, in contrast to the previous theorem the following central limit theorem is for the distribution of the difference between estimated latent positions corresponding to the same true latent position, rather than between true and estimated latent positions.

Theorem 2. Let F be a distribution on $\mathbb{R}^d$, where $x^T y \in [0,1]$ for all $x, y \in \text{supp}(F) \subset \mathbb{R}^d$, and suppose $X_1, X_2, \cdots, X_n \overset{i.i.d.}{\sim} F$ represent the rows of matrix $X = [X_1^T | X_2^T | \cdots | X_n^T]^T \in \mathbb{R}^{n \times d}$. Let $\rho \in [0,1]$ and let $(A, B) \sim \text{CorrER}(XX^T, \rho J)$ be correlated Erdos-Renyi random graphs. Let $P = VSV^T$ be the spectral decomposition of P, where $S \in \mathbb{R}^{d \times d}$ is a diagonal matrix whose nonnegative diagonal entries are the d eigenvalues of P (in non-decreasing order) and $V \in \mathbb{R}^{n \times d}$ is the orthogonal matrix whose columns are the eigenvectors corresponding to the eigenvalues in S. Then, for each $i \in [n]$,

$$n^{\frac{1}{2}} [U_A S_A^{1/2} - W^\star U_B S_B^{1/2}]_i \xrightarrow{\mathcal{D}} \mathcal{N}(0, \Sigma(X_i)) \tag{1}$$

where $\Sigma(y) = 2(1-\rho)\Delta^{-1}\,\mathbb{E}[X_1 X_1^T (y^T X_1 - (y^T X_1)^2)]\Delta^{-1}$, $\Delta = \mathbb{E}[X_1 X_1^T]$ is the second moment matrix and $W^\star \in \mathbb{R}^{d \times d}$ is an orthogonal matrix equal to $W_1 W_2^T$, where W_1, W_2 derive from the SVD of the matrix $BVS_B^{-1}U_B B^T$, i.e., $BVS_B^{-1}U_B B^T = W_1 \Sigma_1 W_2^T$.

The omnibus embedding from Section 3 induces the same amount of correlation between any two estimated latent positions independently of m, the number of graphs. To see this, fix some $i \in [n]$ and some $s_1, s_2 \in [m]$. Let $h_1 = n(s_1 - 1) + i$ and $h_2 = n(s_2 - 1) + i$. The h_1, h_2-th rows of $\text{ASE}(M, d)$ correspond to the estimated latent positions for vertex i. Then, the covariance matrix of their difference is given by

$$\Sigma_{omni}(y) = \frac{1}{2}\Delta^{-1}\,\mathbb{E}[X_1 X_1^T (y^T X_1 - (y^T X_1)^2)]\Delta^{-1}.$$

Note that the above covariance matrix does not depend on m. Also, when $m = 2$ and $\rho = \frac{3}{4}$, the covariance matrix in the CorrER case is equal to the covariance matrix in the omnibus case, which makes apparent the correlation between the estimates imputed by the omnibus embedding.

4.2 CLT for a dampened omnibus embedding of RDPGs

As shown in the previous subsection, the omnibus embedding induces correlation across the estimated latent positions of the graphs. The off-diagonal blocks of the omnibus matrix (as described in Definition 3.2) consist of the unweighted pairwise average of the corresponding adjacency matrices, which results to the same amount of mutual dependence across all the graphs. In contrast, in time-series of networks applications, it is often natural to assume that the dependence between two networks in different times t_1, t_2 decays as their difference $|t_1 - t_2|$ gets bigger, and also decays over time. This motivates the definition of the *dampened* omnibus matrix shown below,

Definition 4.2 (dampened omnibus matrix)**.** Let $A^{(1)}, A^{(2)}, \cdots, A^{(m)} \in \mathbb{R}^{n\times n}$ be adjacency matrices of a collection of m undirected graphs. We define the mn-by-mn dampened omnibus matrix of $A^{(1)}, A^{(2)}, \cdots, A^{(m)}$ by

$$M = \begin{bmatrix} A^{(1)} & \frac{A^{(1)}+A^{(2)}}{2} & \frac{A^{(1)}+2A^{(3)}}{3} & \frac{A^{(1)}+3A^{(4)}}{4} & \cdots & \frac{A^{(1)}+(m-1)A^{(m)}}{m} \\ \frac{A^{(1)}+A^{(2)}}{2} & A^{(2)} & \frac{A^{(2)}+2A^{(3)}}{3} & \frac{A^{(2)}+3A^{(4)}}{4} & \cdots & \frac{A^{(2)}+(m-1)A^{(m)}}{m} \\ \frac{A^{(1)}+2A^{(3)}}{3} & \frac{A^{(2)}+2A^{(3)}}{3} & A^{(3)} & \frac{A^{(3)}+3A^{(4)}}{4} & \cdots & \vdots \\ \frac{A^{(1)}+3A^{(4)}}{4} & \frac{A^{(2)}+3A^{(4)}}{4} & \frac{A^{(3)}+3A^{(4)}}{4} & A^{(4)} & \cdots & \vdots \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots \\ \frac{A^{(1)}+(m-1)A^{(m)}}{m} & \frac{A^{(2)}+(m-1)A^{(m)}}{m} & \frac{A^{(3)}+(m-1)A^{(m)}}{m} & & \cdots & A^{(m)} \end{bmatrix}.$$

The proof of the central limit theorem for the dampened omnibus matrix combines ideas from [8] and [3], and the covariance matrix of the estimated latent positions are more subtle, depending also on the block row s in addition to i-th vertex (in $h = n(s-1)+i$).

Theorem 3. Let $(A^{(1)}, A^{(2)}, \cdots, A^{(m)}, X) \sim \mathrm{JRDPG}(F, n, m)$ and let M denote the omnibus matrix as in Definition 4.2. Let $\tilde{P} = U_{\tilde{P}} S_{\tilde{P}} U_{\tilde{P}}^T$ be the spectral decomposition of $\tilde{P}$, where $S_{\tilde{P}} \in \mathbb{R}^{d\times d}$ is a diagonal matrix whose diagonal entries are its eigenvalues (in non-decreasing order), and $U_{\tilde{P}} \in \mathbb{R}^{mn\times d}$ is the orthogonal matrix whose columns are the eigenvectors corresponding to the eigenvalues in $S_{\tilde{P}}$. If $h = n(s-1)+i$, where $i \in [n], s \in [m]$, then there exists a sequence of orthogonal matrices $\{W_n\}_n$ such that

$$n^{1/2} W_n^T [U_M S_M^{1/2} - U_{\tilde{P}} S_{\tilde{P}}^{1/2}]_h \xrightarrow{\mathcal{D}} \mathcal{N}(0, \Sigma(X_i)),$$

where $\Sigma(y) = \frac{1}{m^2}\Phi(s)\Delta^{-1}\,\mathbb{E}[(y^T X_j - (y^T X_j)^2) X_j X_j^T]\Delta^{-1}$, $\Delta = \mathbb{E}[X_1 X_1^T] \in \mathbb{R}^{d\times d}$ and

$$\Phi(s) = (s-1)^2 + (m-s) + 2(s-2)\sum_{q\geq s}\frac{1}{q} + \frac{3s-1}{s^2} + \sum_{q>s}\frac{1}{q^2} + \Big(\sum_{q\geq s}\frac{1}{q}\Big)^2.$$

.

As stated previously, the embedding with the unweighted pairwise average omnibus matrix induces the same amount of correlation between any two estimated latent positions. This lies on the fact that the covariance matrix of the difference of two estimated latent positions remains constant. On the contrary, in the dampened omnibus embedding, the covariance matrix of the difference of two estimated latent positions, for example, $\text{ASE}(M,d)_{h_1}$ and $\text{ASE}(M,d)_{h_2}$, where $h_1 = n(s_1-1)+i$, $h_2 = n(s_2-1)+i$, depends on the block rows s_1, s_2, as well as the number of graphs m. In particular, the induced correlation between the h_1-th and h_2-th estimated rows of dampened omnibus matrix increases when $s_1 \approx s_2 < \lfloor \frac{m}{2} \rfloor$ and reduces when $|s_1 - s_2|$ gets larger or $s_1 \approx s_2 > \lfloor \frac{m}{2} \rfloor$.

5 Future work

5.1 CLT for directed omnibus embeddings

All the previous results have been shown for the undirected omnibus matrix, and we plan to extend these results to the directed case. The omnibus matrix illustrated below, is a directed matrix which captures the dependency's decay over time uniformly between any two time-series of graphs.

Definition 5.1 (Symmetrically decayed omnibus matrix). *Let* $A^{(1)}, A^{(2)}, \cdots, A^{(m)} \in \mathbb{R}^{n\times n}$ *be adjacency matrices of a collection of* m *undirected graphs. We define the* mn*-by-*mn *non-symmetrical omnibus matrix of* $A^{(1)}, A^{(2)}, \cdots, A^{(m)}$ *by*

$$M_{dec} = \begin{bmatrix} A^{(1)} & \frac{A^{(1)}+A^{(2)}}{2} & \frac{2A^{(1)}+A^{(3)}}{3} & \frac{3A^{(1)}+A^{(4)}}{4} & \cdots & \frac{(m-1)A^{(1)}+A^{(m)}}{m} \\ \frac{A^{(1)}+A^{(2)}}{2} & A^{(2)} & \frac{A^{(2)}+A^{(3)}}{2} & \frac{2A^{(2)}+A^{(4)}}{3} & \cdots & \frac{(m-2)A^{(2)}+A^{(m)}}{m-1} \\ \frac{A^{(1)}+2A^{(3)}}{3} & \frac{A^{(2)}+A^{(3)}}{2} & A^{(3)} & \frac{A^{(3)}+A^{(4)}}{2} & \cdots & \vdots \\ \frac{A^{(1)}+3A^{(4)}}{4} & \frac{A^{(2)}+2A^{(4)}}{3} & \frac{A^{(3)}+A^{(4)}}{2} & A^{(4)} & \cdots & \vdots \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots \\ \frac{A^{(1)}+(m-1)A^{(m)}}{m} & \frac{A^{(2)}+(m-2)A^{(m)}}{m-1} & \frac{A^{(3)}+(m-3)A^{(m)}}{m-2} & \cdots & \frac{A^{(m-1)}+A^{(m)}}{2} & A^{(m)} \end{bmatrix}$$

In Sussman et al. [15], it is shown that the latent positions of a random dot product graph can be consistently estimated by the eigen-decomposition of its adjacency matrix in the directed case, however no directed-graph CLT analogues exist for the estimated latent positions. Proving a central limit theorem for the above omnibus matrix will extend our current results which require symmetric omnibus structure.

References

[1] J. Arroyo, A. Athreya, J. Cape, G. Chen, C. E. Priebe, and J. T. Vogelstein. Inference for multiple heterogeneous networks with a common invariant subspace. *arXiv preprint arXiv:1906.10026*, 2019.

[2] Avanti Athreya, Donniell E. Fishkind, Keith Levin, Vince Lyzinski, Youngser Park, Yichen Qin, Daniel L. Sussman, Minh Tang, Joshua T. Vogelstein, and Carey E. Priebe. Statistical inference on random dot product graphs: a survey, 2017.

[3] Joshua Cape, Minh Tang, and Carey E. Priebe. The two-to-infinity norm and singular subspace geometry with applications to high-dimensional statistics, 2017.

[4] P. Diaconis and S. Janson. Graph limits and exchangeable random graphs. *Rendiconti di Matematica, Serie VII*, 28:33–61, 2008.

[5] P. Erdős and A. Rényi. On random graphs, I. *Publicationes Mathematicae*, 6:290–297, 1959.

[6] P. D. Hoff, A. E. Raftery, and M. S. Handcock. Latent space approaches to social network analysis. *Journal of the American Statistical Association*, 97(460):1090–1098, 2002.

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

[8] Keith Levin, Avanti Athreya, Minh Tang, Vince Lyzinski, Youngser Park, and Carey E. Priebe. A central limit theorem for an omnibus embedding of multiple random graphs and implications for multiscale network inference, 2017.

[9] 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(2):2905–2922, 2014.

[10] V. Lyzinski, M. Tang, A. Athreya, Y. Park, and C. E. Priebe. Community detection and classification in hierarchical stochastic blockmodels. *arXiv preprint arXiv:1503.02115*, 2015.

[11] A. M. Nielsen and D. Witten. The multiple random dot product graph model. *arXiv preprint arXiv:1811.12172*, 2018.

[12] R. I. Oliveira. Concentration of the adjacency matrix and of the laplacian in random graphs with independent edges. *arXiv preprint arXiv:0911.0600*, 2009.

[13] Purnamrita Sarkar and Peter J. Bickel. Role of normalization in spectral clustering for stochastic blockmodels. *The Annals of Statistics*, 43(3):962–990, Jun 2015.

[14] Anna L. Smith, Dena M. Asta, and Catherine A. Calder. The geometry of continuous latent space models for network data, 2017.

[15] 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(499):1119–1128, 2012.

[16] D. L. Sussman, M. Tang, and C. E. Priebe. Consistent latent position estimation and vertex classification for random dot product graphs. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 36(1):48–57, 2014.

[17] M. Tang, A. Athreya, D. L. Sussman, V. Lyzinski, and C. E. Priebe. A nonparametric two-sample hypothesis for random dot product graphs. arXiv preprint. http://arxiv.org/abs/1403.7249, 2014.

[18] M. Tang, A. Athreya, D. L. Sussman, V. Lyzinski, and C. E. Priebe. A semiparametric two-sample hypothesis testing for random dot product graphs. arXiv preprint. http://arxiv.org/abs/1403.7249, 2014.

[19] M. Tang, D. L. Sussman, and C. E. Priebe. Universally consistent vertex classification for latent positions graphs. *The Annals of Statistics*, 41(3):1406–1430, 2013.

[20] Minh Tang and Carey E. Priebe. Limit theorems for eigenvectors of the normalized laplacian for random graphs, 2016.

[21] S. Wang, J. Arroyo, J. T. Vogelstein, and C. E. Priebe. Joint embedding of graphs. *IEEE transactions on pattern analysis and machine intelligence*, 2019.

[22] 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.