Connectome Smoothing via Low-rank Approximations

Runze Tang, Michael Ketcha, Alexandra Badea, Evan D. Calabrese,
Daniel S. Margulies, Joshua T. Vogelstein, Carey E. Priebe *Senior Member, IEEE*, Daniel L. Sussman

Abstract— In brain imaging and connectomics, the study of brain networks, estimating the mean of a population of graphs based on a sample is a core problem. Often, this problem is especially difficult because the sample or cohort size is relatively small, sometimes even a single subject, while the number of nodes can be very large with noisy estimates of connectivity. While the element-wise sample mean of the adjacency matrices is a common approach, this method does not exploit underlying structural properties of the graphs. We propose using a low-rank method which incorporates dimension selection and diagonal augmentation to smooth the estimates and improve performance over the naïve methodology for small sample sizes. Theoretical results for the stochastic blockmodel show that this method offers major improvements when there are many vertices. Similarly, we demonstrate that the low-rank methods outperform the standard sample mean for a variety of independent edge distributions as well as human connectome data derived from magnetic resonance imaging, especially when sample sizes are small. Moreover, the low-rank methods yield "eigen-connectomes", which correlate with the lobe-structure of the human brain and superstructures of the mouse brain. These results indicate that low-rank methods are an important part of the toolbox for researchers studying populations of graphs in general, and statistical connectomics in particular.

Index Terms— networks, connectome, low-rank, estimation

I. INTRODUCTION

Estimating a population mean based on a sample of brain networks, as represented by their structural connectomes estimated from DTMRI, is a key challenge. Generally, estimation of a population mean based on samples is at the core of

R. Tang and C. E. Priebe are with the Department of Applied Math & Statistics, The Johns Hopkins University, Baltimore, MD

M. Ketcha and J. T. Vogelstein are with the Department of Biomedical Engineering, The Johns Hopkins University, Baltimore, MD

A. Badea and E. D. Calabrese are with the Department of Radiology, and Department of Biomedical Engineering, Duke University, Durham, NC

D. S. Margulies is with Max Planck Research Group for Neuroanatomy & Connectivity, Max Planck Institute for Human Cognitive and Brain Sciences, Leipzig, Germany

D. L. Sussman is with the Department of Mathematics & Statistics, Boston University, Boston, MA; e-mail: sussman@bu.edu

statistics. For a distribution of graphs, we define the mean graph as the expectation of the adjacency matrix, regardless of the data distribution. The sample mean, motivated by its intuitive appeal, the law of large numbers, and the central limit theorem, has its place as an important statistic for mean graph estimation.

However, the mean of a population of graphs is a high-dimensional object, consisting of $O(N^2)$ parameters for graphs with N nodes (vertices). When the number of samples M is much smaller than N^2, or even N, estimating such a high-dimensional estimands using naive unbiased methods can lead to inaccurate estimates with very high variance. Such small-M large-N occur frequently with new higher-resolution technologies as well as when the focus is on a smaller harder-to-sample subpopulation. With MRI technologies improving in resolution and newer imaging modalities such as electron microscopy becoming more prominent, ever larger connectomes will be under study. Similarly, in animal studies, studies of rare diseases, and when estimating means for small subpopulations, the number of sample graphs M may be quite small.

Furthermore, using poor estimates can lead to low power and accuracy for subsequent inference tasks. By exploiting a bias-variance trade-off, it is often fruitful to develop estimators which have some bias but greatly reduced variance.

When these estimators are biased towards low-dimensional structures which well approximate the full dimensional population mean, major improvements in estimation can be realized [1]. Furthermore, a more parsimonious representation of the mean improves interpretability and allows for novel exploratory analyses.

In statistical connectomics, [2], [3] propose a way to test if there is a difference between the distributions for two groups of networks. While hypothesis testing is the end goal of their work, estimation is a key intermediate step which may be improved by accounting for underlying structure in the mean matrix. Thus, improving the estimation procedures for the mean graph is not only important by itself, but also can be applied to help improve other statistical inference procedures. Indeed, for the task of community detection, [3] propose the use of low-rank mean graphs to find clusters across networks.

The entry-wise sample mean is a reasonable estimator if one is unwilling to take any additional structure into account. However, with only a small sample size, the entry-wise sample mean does not perform very well due to high variance. Intuitively, an estimator incorporating known structure in the true distribution of the graphs, assuming the estimator is com-

putationally tractable, is preferable to the entry-wise sample mean.

We propose an estimator based on a low-rank structure of a family of random graphs. Section III-B discusses details about this estimator. These estimates can improve performance since they have much lower overall variance than naive entry-wise sample means, thereby offsetting the bias towards the low-rank structure in terms of overall error. Additionally, [4] has argued that even many full-rank statistical models may be well approximated by low-rank structures.

The proposed method builds on the idea of approximating a matrix with a low-rank matrix [5] but also incorporates aspects specific to networks such as bounded entries and missing diagonal entries. Selection of the rank for the approximation is a notoriously difficult problem and our work considers two methods for automatic rank selection.

Other matrix decompositions such as non-negative matrix factorization [6] and non-negative SVD [7] aim to decompose a matrix into non-negative factors. This often has an intuitive appeal when some true latent factors are themselves non-negative, such as for dictionary learning tasks. As our primary motivation is estimating the mean graph, non-negativity constraints are not needed and we are still able to use lower-dimensional factors for further exploration. Alternatively, from the motivation of clustering networks themselves, [8] has used non-negative matrix factorization of of vectorized adjacency matrices as a preliminary dictionary learning step.

The use of low-rank and spectral methods is not new to connectomics, neuroscience, and biology [9]. [10] employs low-rank shrinkage for estimation of functional connectivity. Connectome harmonics [11] have been investigated to understand the dynamics of neural networks [12]. The Laplacian spectrums of connectomes have been used for comparison across species [13]. Our work specifies when and how these types of methods can be used to improve performance of mean graph estimation.

Importantly, low-rank methods also produce parsimonious and interpretable representations of the data as represented by the random dot product graph model. The proposed algorithm can be viewed as yielding estimates of low-rank latent structure in the population, regardless of whether the true mean graph is low-rank or not. As we demonstrate, the interpretations of these latent parameter estimates illustrate hypotheses which relate the structure of the connectome to well established anatomical structures (see Fig. 6 and Section V-B) and suggest a basis for mapping structural hierarchies in brain organization.

II. STATISTICAL CONNECTOME MODELS

In connectomics, brain imaging data for each subject can be processed to output a graph, where each vertex represents a well-defined anatomical region present in each subject. For structural brain imaging, such as diffusion tensor MRI, an edge may represent the presence of anatomical connections between the two regions as estimated using tractography algorithms [14]. For functional brain imaging, such as fMRI, an edge between two regions may represent the presence of correlated brain activity between the two regions.

This work considers the scenario of observing M graphs, represented as adjacency matrices, $A^{(1)}, A^{(2)}, \ldots, A^{(M)}$, each having N vertices with $A^{(m)} \in \{0, 1\}^{N \times N}$ for each m. We assume there is a known correspondence for vertices in different graphs, so that vertex i in graph m corresponds to vertex i in graph m' for any i, m, m'. The graphs we consider are undirected and unweighted with no self-loops, so each $A^{(m)}$ is a binary symmetric matrix with zeros along the diagonal.

We assume that the graphs are sampled independently and identically from some distribution andthe mean graph is the expectation of each adjacency matrix.

Definition 2.1 (Mean Graph): Suppose $A^{(1)}, \ldots, A^{(M)} \overset{iid}{\sim} \mathcal{G}$ for some random graph distribution $\mathcal{G}$, with $A^{(m)} \in \{0, 1\}^{N \times N}$ for each m. The *mean graph* is $\mathbb{E}[A^{(m)}]$, which we will denote by P.

Note, this definition requires no assumption on the distribution $\mathcal{G}$.

A. Independent Edge Model

The most general model we consider is the independent edge model (IEM) with parameter given by a mean graph $P \in [0, 1]^{N \times N}$ [15]. An edge exists between vertex i and vertex j with probability P_{ij}, and each edge is present independently of all other edges. Note that the IEM is a generalization of the Erdős-Rényi random graphs, where each edge is present with probability p independently of all other edges [16], [17].

The assumption of independent edges, while not necessary to define the mean graph, is important for our theory. In practice the independent edge assumption may not hold, as brain graphs will have complex spatial and functional dependencies. On the other hand, if $\mathbb{E}[A^{(m)}]$ has low-rank structure then our methods below can still improve performance. Regardless of the structure P, the low-rank method below will provide an estimate of the low-rank structure of P.

B. Random Dot Product Graph

The random dot product graph model (RDPG) [18], [19], assigns latent positions to each vertex in a graph where the probability of an edge being present between two nodes is the dot product of their latent vectors [20].

Formally, let $\mathcal{X} \subset \mathbb{R}^d$ be a set such that $x, y \in \mathcal{X}$ implies $x^\top y = \sum_{i=1}^d x_i y_i \in [0, 1]$. Let $X_1, \ldots, X_N \in \mathcal{X}$ be iid column vectors representing the N latent positions and let $X = [X_1 | \cdots | X_N]^\top \in \mathbb{R}^{N \times d}$. Note, the entries of X are not required to be non-negative and for these purposes this introduce no interpretability issues. We assume that X is the same for all observed graphs.

A random adjacency matrix A is said to be an RDPG if for each adjacency matrix $a \in \{0, 1\}^{n \times n}$,

$$\Pr[A = a | X] = \prod_{i<j} (X_i^\top X_j)^{a_{ij}} (1 - X_i^\top X_j)^{1-a_{ij}}.$$

Conditioned on the latent positions, the RDPG is an independent edge model with probability matrix given by the outer product of the latent position matrix with itself, $P = XX^\top$. This imposes that P is positive-semidefinite and $\text{rank}(P) = \text{rank}(X) \leq d$.

C. Stochastic Blockmodel as an RDPG

One of the most common structures for graphs is that vertices tend to cluster into communities. Vertices of the same community behave similarly, connecting to similar sets of nodes.

This structural property is captured by the stochastic blockmodel (SBM) [21], where each vertex is assigned to a block and the probability that an edge exists between two vertices depends only on their respective block memberships.

The SBM is parameterized by the number of blocks K (generally much less than the number of vertices N), the block probability matrix $B \in [0,1]^{K \times K}$, and the vector of iid block memberships $\tau \in \{1, \ldots, K\}^N$. For each $i \in [N]$, $\tau_i = k$ means vertex i is a member of block k. Denote as ρ_k the probability $\Pr[\tau_i = k]$ for each k. We assume that the τ vector is the same for all observed graphs.

Conditioned on τ, $A_{ij} \overset{ind}{\sim} \mathrm{Bern}(B_{\tau_i, \tau_j})$. To ensure that the SBM can be considered as an RDPG, we always impose that the B matrix for the SBM is positive semidefinite (see Appendix G).

To better describe complex network structures, many generalizations of the SBM incorporate the variation of vertices within blocks. For example, [22] proposed mixed membership stochastic blockmodels, and [23] proposed degree-corrected SBM. These generalizations aim to capture variations among vertices while maintaining parts of the original community structure. The RDPG is useful in this regard since any SBM with degree-correction and mixed-membership can be represented as an RDPG and visa versa [24], [25].

III. ESTIMATORS

A. Element-wise Sample Mean $\bar{A}$

The most natural estimator to consider is the element-wise sample mean. This estimator, defined as $\bar{A} = \frac{1}{M} \sum_{m=1}^{M} A^{(m)}$, is the maximum likelihood estimator (MLE) for the mean graph P if the graphs are sampled from an IEM distribution. It is unbiased so $\mathbb{E}[\bar{A}] = P$ with entry-wise variance $\mathrm{Var}(\bar{A}_{ij}) = P_{ij}(1 - P_{ij})/M$. Moreover, for the independent edge model, $\bar{A}$ is the uniformly minimum-variance unbiased estimator, so it has the smallest variance among all unbiased estimators. Similarly, it enjoys the many asymptotic properties of the MLE as $M \to \infty$ for fixed N. However, if graphs with a large number of vertices are of interest, $\bar{A}$ is not consistent for P as the number of vertices N becomes large for fixed M, while our estimator $\hat{P}$ from Section III-B is consistent for low-rank P.

Additionally, $\bar{A}$ does not exploit any low-dimensional structure. If the graphs are distributed according to an RDPG or SBM, then $\bar{A}$ is no longer the maximum likelihood estimator since it is not guaranteed to satisfy the properties of the mean graph for that model. In the RDPG case, $\mathrm{rank}(\bar{A})$ will with high probability not be equal to d, and for the SBM case, $\bar{A}$ will not have the structure of an SBM mean graph, with K distinct rows and columns. Hence, in either case $\bar{A}$ will be outside the parameter space and idadmissible.

The performance can be especially poor when the sample size M is small, such as when $M \ll N$. For example,

when $M = 1$, $\bar{A}$ is simply the binary adjacency matrix $A^{(1)}$, which is an inaccurate estimate for an arbitrary P compared to estimates which exploit underlying structure, such as the low-rank structure of the RDPG model.

B. Low-Rank Estimator $\hat{P}$

Motivated by the low-rank structure of the RDPG mean matrix, we propose the estimator $\hat{P}$ based on the spectral decomposition of $\bar{A}$, yielding a low rank approximation of $\bar{A}$. Low-rank methods for matrices can be viewed as being closely related to ideas such as principal components analysis. Viewed from a regularization perspective, the low-rank approximation finds an estimate with many zero-eigenvalues which can be viewed as analogous to an L_0 constraint on the eigenvalues of the solutions. These estimates are also related to methods employing nuclear norm penalization [26].

This estimator is similar to the estimator proposed by [26] but incorporates additional adjustments which serve to improve the performance for the specific task of estimating the mean graph. Additionally, we consider an alternative dimension selection technique. Details of the dimension selection procedures and the diagonal augmentation are in Appendices A and B, respectively. To summarize, the overall strategy to compute $\hat{P}$ is described in Algorithm 2. A key component of this algorithm is the low-rank approximation from Algorithm 1.

Algorithm 1 Rank-d approximation of a matrix.

Input: Symmetric matrix $A \in \mathbb{R}^{N \times N}$, dimension $d \leq N$.
Output: $\mathrm{lowrank}_d(A) \in \mathbb{R}^{N \times N}$
1: Compute the algebraically largest d eigenvalues of A, $s_1 \geq s_2 \geq \ldots \geq s_d$ and corresponding orthonormal eigenvectors $u_1, u_2, \ldots, u_d \in \mathbb{R}^N$;
2: $\hat{S} \leftarrow \mathrm{diag}(s_1, \ldots, s_d)$ and $\hat{U} \leftarrow [u_1, \ldots, u_d]$;
3: Return $\hat{U}\hat{S}\hat{U}^\top$;

The first step is to calculate the sample mean $\bar{A}$. In Step 2 to Step 4, the algorithm augments the diagonal of $\bar{A}$ based on [27], selects the dimension $\hat{d}$ to embed (see Appendix A), and computes the low-rank approximation $\tilde{P}^{(0)}$ based on the embedding. Then in Step 5 and Step 6, the algorithm augments the diagonal again based on [28] (see Appendix B) which yields an improved low-rank estimate $\tilde{P}^{(1)}$. Finally, Step 7 thresholds the matrix entries to ensure all elements are between 0 and 1.

For a given dimension d we consider the estimator $\mathrm{lowrank}_d(\bar{A})$ defined as the best rank-d positive-semidefinite approximation of $\bar{A}$. Let $\hat{S}$ be a diagonal matrix with non-increasing entries along the diagonal corresponding to the largest d eigenvalues of $\bar{A}$ and let $\hat{U}$ have columns given by the corresponding eigenvectors. Similarly, let $\tilde{S}$ be the diagonal matrix with non-increasing entries along the diagonal corresponding to the remaining $N - d$ eigenvalues of $\bar{A}$ and let $\tilde{U}$ have columns given by the corresponding eigenvectors. Since the graphs are symmetric, the eigen-decomposition can be computed as $\bar{A}$ as $\hat{U}\hat{S}\hat{U}^\top + \tilde{U}\tilde{S}\tilde{U}^\top = [\hat{U}|\tilde{U}](\hat{S} \oplus \tilde{S})[\hat{U}|\tilde{U}]^T$.

Algorithm 2 Algorithm to compute $\hat{P}$

Input: Adjacency matrices $A^{(1)}, A^{(2)}, \cdots, A^{(M)}$, with each $A^{(m)} \in \{0,1\}^{N \times N}$

Output: Estimate $\hat{P} \in [0,1]^{N \times N}$

1: $\bar{A} \leftarrow \left(\sum_{m=1}^{M} A^{(m)} \right) / M$;
2: $D^{(0)} \leftarrow \text{diag}(\bar{A}\mathbf{1})/(N-1)$;
3: $\hat{d} \leftarrow \text{dimselect}(\bar{A} + D^{(0)})$; (see Appendix A)
4: $\tilde{P}^{(0)} \leftarrow \text{lowrank}_{\hat{d}}(\bar{A} + D^{(0)})$; (see Algorithm 1)
5: $D^{(1)} \leftarrow \text{diag}(\tilde{P}^{(0)})$;
6: $\tilde{P}^{(1)} \leftarrow \text{lowrank}_{\hat{d}}(\bar{A} + D^{(1)})$; (see Algorithm 1)
7: $\hat{P} \leftarrow \min(\max(\tilde{P}^{(1)}, 0), 1)$.

The d-dimensional adjacency spectral embedding (ASE) of $\bar{A}$ is given by $\hat{X} = \hat{U}\hat{S}^{1/2} \in \mathbb{R}^{N \times d}$. For an RDPG, the rows of $\hat{X}$ are estimates of the latent vectors for each vertex [29]. Using the adjacency spectral embedding, the low-rank approximation of $\bar{A}$ is $\hat{X}\hat{X}^{\top} = \hat{U}\hat{S}\hat{U}^{\top}$. Algorithm 1 gives the steps to compute this low-rank approximation for a general symmetric matrix A.

To compute the estimator $\hat{P}$, the rank d must be specified; there are various ways of dealing with dimension selection. In this paper, we explore an elbow selection method proposed in [30] and the universal singular value thresholding (USVT) method [26]. Appendix A discusses details of these methods.

Moreover, when the adjacency matrices are hollow, with zeros along the diagonal, there is a missing data problem that leads to inaccuracies if $\hat{P}$ is computed based only on $\bar{A}$. To compensate for this issue, we use an iterative method developed in [28]. Appendix B discusses details of the iterative method.

IV. THEORETICAL RESULTS

To estimate the mean of a collection of graphs, we compare the two estimators from Section III: the entry-wise sample mean $\bar{A}$ and the low-rank $\hat{P}$ motivated by the RDPG. The mean squared errors (MSE) for our estimators are $\text{MSE}(\hat{P}_{ij}) = \mathbb{E}[\hat{P}_{ij} - P]^2$ and $\text{MSE}(\bar{A}) = \mathbb{E}[\bar{A}_{ij} - P]^2$. The relative efficiency for two estimators is the ratio of their MSE, $\text{RE}(\bar{A}_{ij}, \hat{P}_{ij}) = \frac{\text{MSE}(\hat{P}_{ij})}{\text{MSE}(\bar{A}_{ij})}$, with values above 1 indicating $\bar{A}$ should be preferred while values below 1 indicate $\hat{P}$ should be preferred. Relative efficiency is a useful metric for comparing estimators because it will frequently be invariant to the scale of the noise in the problem and hence is comparable across different settings.

A. SBM

In this section, entry-wise relative efficiency is computed to analyze the performance of these two estimators under the SBM. The asymptotic relative efficiency is defined as $\lim_{N \to \infty} \text{RE}$. We also define the scaled relative efficiency, $N \cdot \text{RE}(\bar{A}_{ij}, \hat{P}_{ij})$ which normalizes the relative efficiency so that the asymptotic scaled relative efficiency is non-zero and finite. Somewhat surprisingly, the results indicate that the

asymptotic relative efficiency will not depend on the sample size M.

For this asymptotic framework, the proportion of vertices in block k, $|\{i : \tau_i = k\}|/N$, will converge to ρ_k as $N \to \infty$ by the law of large numbers.

Denote the block probability matrix as $B = \nu\nu^{\top} \in [0,1]^{K \times K}$. By definition, the mean of the collection of graphs generated from this SBM is $P \in [0,1]^{N \times N}$, where $P_{ij} = B_{\tau_i, \tau_j}$. After observing M graphs on N vertices $A^{(1)}, \cdots, A^{(M)}$ sampled independently from the SBM conditioned on τ, the two estimators can be calculated, $\bar{A}$ and $\hat{P}$.

Lemma 4.1: For the above setting, for any $i \neq j$, if $\text{rank}(B) = K = d$; for large enough N,

$$\mathbb{E}[(\text{lowrank}_d(\bar{A})_{ij} - P_{ij})^2] \approx \frac{1/\rho_{\tau_i} + 1/\rho_{\tau_j}}{MN} P_{ij}(1 - P_{ij}),$$

and

$$\lim_{N \to \infty} N \cdot \text{Var}(\text{lowrank}_d(\bar{A})_{ij}) = \frac{1/\rho_{\tau_i} + 1/\rho_{\tau_j}}{M} P_{ij}(1 - P_{ij}).$$

Theorem 4.2: In the same setting as in Lemma 4.1, for any $i \neq j$, if $\text{rank}(B) = K = d$, then for large enough N:

$$\text{RE}(\bar{A}_{ij}, \text{lowrank}_d(\bar{A})_{ij}) \approx \frac{1/\rho_{\tau_i} + 1/\rho_{\tau_j}}{N}, \quad (1)$$

and the asymptotic relative efficiency (ARE) is

$$\text{ARE}(\bar{A}_{ij}, \text{lowrank}_d(\bar{A})_{ij}) = \lim_{N \to \infty} \text{RE}(\bar{A}_{ij}, \text{lowrank}_d(\bar{A})_{ij}) = 0.$$

Note these theorems are stated for $\text{lowrank}_d(\bar{A})$ rather than $\hat{P}$. This allows us to employ theoretical developments which apply specifically to the rank-d approximation of $\bar{A}$ [31]. While we assume that $\text{rank}(P)$ is known for this theory, asymptotically accurate methods to estimate d in the SBM and RDPG settings have been established [32]. In the large N setting, the other elements of Algorithm 2, notably the diagonal augmentations and thresholding, result in negligible deviations of from $\text{lowrank}_d(\bar{A})$ and hence do not impact the asymptotic theory. Nonetheless, these components do improve finite sample performance, see Appendix C.

Note that ρ_{τ_i} represents the probability that a vertex is assigned to the same block as vertex i, i.e. τ_i-th block. The proofs of these results are provided in Appendix G.

This theorem indicates that under the SBM, $\hat{P}$ is a much better estimate of the mean of the collection of graphs P than $\bar{A}$, especially when N is large. Note that a relative efficiency less than 1 indicates that $\hat{P}$ should be preferred over $\bar{A}$, so under the above assumptions, as $N \to \infty$, $\hat{P}$ performs far better than $\bar{A}$. Note that even though the RE could be greater than 1 for some N, eventually the RE will go to 0 as N increases. The result shows that the relative efficiency is of order $O(N^{-1})$ and $N \cdot \text{RE}(\bar{A}_{ij}, \hat{P}_{ij})$ (denoted as scaled RE) converges to $1/\rho_{\tau_i} + 1/\rho_{\tau_j}$ when $N \to \infty$.

An important aspect of Theorem 4.2 is that the ARE does not depend on the number of graphs M, so the larger the graphs are, the better $\hat{P}$ is relative to $\bar{A}$, regardless of M. For smaller values N, the impact of M may be more substantial. In the SBM and RDPG cases, $\hat{P}$ should still offer substantial improvements over $\bar{A}$. The theory in this case is more difficult

to analyze as the impacts of low-rank approximations in the small N setting are less well understood (see Section IV-C).

The asymptotic results are for a number of vertices going to infinity with a fixed number of graphs, a setting which will be very useful in future connectomics analysis, as the collection of larger and larger brain networks grow from small sample sizes. For example, [33] recently reported a high resolution magnetic resonance microscopy based estimate of the mouse brain using a single mouse.

The approximate formula Eq. 1 indicates that the sizes of the blocks can greatly impact the relative efficiency. As an example, consider a 2-block SBM with large but fixed number of nodes N. If each of the blocks contain half the vertices, then for each pair of vertices, the relative efficiency is approximately $4/N$. If the first block gets larger, with $\rho_1 \to 1$, then the RE for estimating P_{ij} with $\tau_i = \tau_j = 1$ will tend to its minimum of $2/N$. On the other hand as $\rho_1 \to 1$, if $\tau_i = 1$ and $\tau_j = 2$, then, since $\rho_2 = 1 - \rho_1$, the relative efficiency for estimating such an edge pair will be approximately 1 and the same will hold if $\tau_i = \tau_j = 2$. Note that the maximum value for the relative efficiency of two vertices from different blocks in a two-block model is achieved when $\rho_1 = 1/N$ and $\rho_2 = (N-1)/N$ in which case the relative efficiency is $N/(N-1) \approx 1$. (Values of ρ_s below $1/N$ correspond to graphs where no vertices are typically in that block, so the effective minimum that can be considered for ρ_s is $1/N$.) Note, $N \cdot \text{RE}(\bar{A}_{ij}, \hat{P}_{ij})$ achieves its minimum for i and j from different blocks when $\rho_k = 1/K$ for all k.

Finite sample simulations illustrating these results are in Appendix H.

B. RDPG

If instead of assuming that the graphs follow an SBM distribution, we assume that the graphs are distributed according to an RDPG distribution, similar gains in relative efficiency can be realized. While there is no compact analytical formula for the relative efficiency of $\hat{P}$ versus $\bar{A}$ in the general RDPG case, using the same ideas as in Theorem 4.2, we can show that $\text{RE}(\bar{A}_{ij}, \hat{P}_{ij}) = O(1/N)$.

Proposition 4.3: Suppose that $A^{(1)}, A^{(2)}, \ldots, A^{(M)}$ are iid from an RDPG distribution with common latent positions $X_1, \ldots, X_n$, which are drawn iid from a fixed distribution. As the number of vertices $N \to \infty$, it holds for any $i \neq j$ that $\text{RE}(\bar{A}_{ij}, \text{lowrank}_d(\bar{A})_{ij}) = O(1/N)$, where again the asymptotic relative efficiency in N does not depend on M. The proof of this proposition closely follows the proofs of Lemma 4.1 and Theorem 4.2.

C. Generalizations

When low-rank assumptions hold for the mean graph P, our theory and subsequent simulations show that $\hat{P}$ will significantly outperform $\bar{A}$. However, if P is full-rank or nearly full-rank, then we do not have such guarantees.

If the graphs are distributed according to an SBM or an RDPG, the relative efficiency is approximately invariant to the number of graphs M when N is large. If on the other hand, the graphs are generated according to a full-rank independent edge model, then the relative efficiency can change more dramatically as M changes. For larger M, more of the eigenvectors of $\bar{A}$ will concentrate around the eigenvectors of the mean graph. This leads to the fact that the optimal embedding dimension for estimating the mean will increase, making $\bar{A}$ and the optimal low-rank approximation more similar. As a result, $\text{RE}(\bar{A}, \hat{P})$ will increase as M increases for full-rank models, with $\text{RE}(\bar{A}, \hat{P})$ possibly ≥ 1 since it is not guaranteed that $\hat{P}$ will choose the optimal dimension.

As $M \to \infty$ with N fixed, the optimal embedding dimension will itself tend to N. If the dimension selection method also tends to N, such as for the USVT method, then for M large, $\hat{P}$ and $\bar{A}$ will coincide. Note, this relies on also including negative eigenvalues in the estimate. In the general case, for fixed M and N, whether $\hat{P}$ or $\bar{A}$ has better performance is a difficult theoretical question beyond the scope of this manuscript.

Another challenge in some network contexts is sparsity of the graph as N grows. Many of the theoretical underpinnings of our proofs have been extended to this setting [34], which enable the extension of our results to the sparse setting for average degrees as small as $\theta(\log^4 N)$.

V. HUMAN CONNECTOMES

In practice, observed graphs do not follow the independent edge model, let alone an RDPG or SBM, but the mean of a population of graphs is still of interest. To demonstrate that the estimator $\hat{P}$ is useful in such cases, its performance on structural connectomic data is tested. The graphs are based on diffusion tensor MR images of the SWU4 dataset collected and available at the Consortium for Reliability and Reproducibility [35] (see Appendix E for dataset details). The dataset contains 454 brain scans, each of which was processed to yield an undirected, unweighted graph with no self-loops, using the pipeline described in [36], [37]. The vertices of the graphs represent regions in the brain defined according to an atlas. Here, three atlases were used: the JHU atlas with 48 vertices [38], the Desikan atlas with 70 vertices [39], and the CPAC200 atlas with 200 vertices [40]. An edge exists between two vertices whenever there is at least one white-matter tract connecting the corresponding two regions of the brain.

Our goal is to estimate the mean graph of the population P, defined as the entry-wise mean of all the 454 graphs. Fig. 1 (a) shows a heat map of the population mean graph P. Darker pixels indicate a higher proportion of graphs having an edge between the given vertices. Fig. 1(b) depicts the entry-wise sample mean $\bar{A}$ when the sample size is $M = 5$ in the SWU4 dataset example. While $\bar{A}$ is a reasonable estimate of P, there are some vertex-pairs with very inaccurate estimates. The upper triangular area of the heat map for $\bar{A}$ depicts the 18 vertex-pairs which have an absolute estimation error larger than 0.4. When the sample size is small, the performance of $\bar{A}$ degrades due to its high variance. Such phenomena are most obvious when the sample size decreases from $M = 5$ to $M = 1$. Fig. 1(c) shows the heat map of $\bar{A}$ based on sample size $M = 1$. Since there is only one observed graph, $\bar{A}$ is binary and thus very bumpy. Similarly, when the same absolute
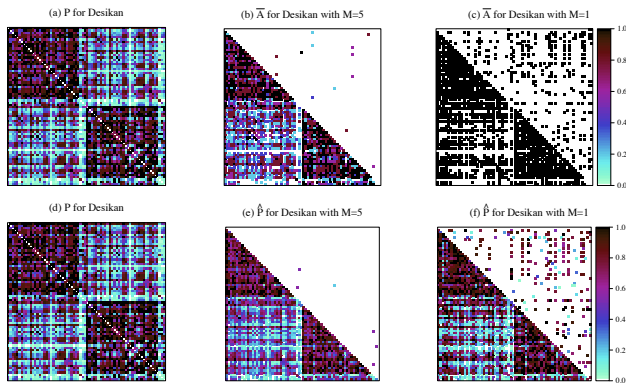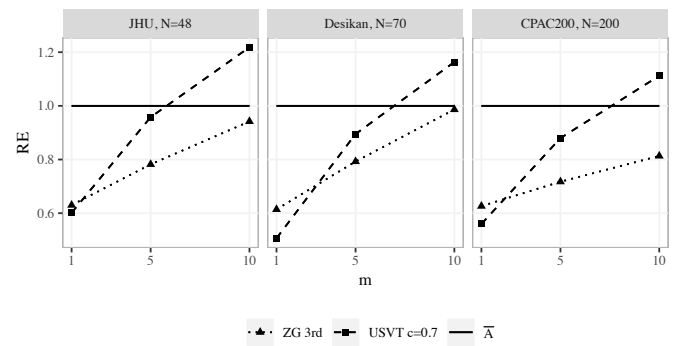
Fig. 1: The heat map of the population mean for the 454 human connectome graphs from the SWU4 dataset and the Desikan atlas are depicted in Panel (a) (also in Panel (d)). Darker pixels indicate a higher proportion of graphs having an edge between the given vertices. The two heat maps in the second column indicate the sample mean for 5 sampled graphs (Panel (b)), and the low-rank estimate $\hat{P}$ for the same 5 graphs with rank $d = 11$ (Panel (e)), computed using Algorithm 2. The mean squared errors (MSE) for this sample are $MSE(\hat{P}) = 0.015$ as compared to $MSE(\bar{A}) = 0.016$, a 3% relative improvement. To highlight the improvements, the upper triangular areas of the heat maps for $\bar{A}$ and $\hat{P}$ show the 18 edges for $\bar{A}$ and 6 edges for $\hat{P}$ which have absolute estimation error larger than 0.4. In the third column, two heat maps using a sample size $M = 1$ ($\hat{P}$ is calculated with a dimension $d = 12$) show a smoothing effect in the heat map of $\hat{P}$ (MSE= 0.049), which leads to a 53% relative improvement compared to $\bar{A}$(MSE= 0.104). Similarly, the same absolute estimation error threshold of 0.4 highlights 504 edges for $\bar{A}$ and 234 edges for $\hat{P}$.

Fig. 2: $\bar{A}$ and $\hat{P}$ were computed on samples of graphs from each atlas: JHU, Desikan, and CPAC200, with different sample sizes M and different dimension selection procedures, ZG and USVT. For each of the two methods for computing $\hat{P}$, relative efficiencies were estimated with respect to the sample mean $\bar{A}$. Confidence intervals all had lengths less than 0.015, and hence were omitted for clarity. When $M = 1$ or 5, $\hat{P}$ always provides substantial improvements (RE $< .7$) compared to $\bar{A}$. For $M = 10$, $\hat{P}$ using USVT has worse performance than $\bar{A}$ but using ZG still improves upon $\bar{A}$.

estimation error threshold is 0.4, 504 (out of 2415) edges in the upper triangular area are highlighted.

When we use the same random sample size of $M = 5$ as in Fig. 1, the plot of $\hat{P}$ in Panel (e) shows a finer gradient of values which results in a 3% relative improvement in estimation of the true probability matrix, P. $\bar{A}$ has mean squared error of 0.016 and $\hat{P}$ has mean squared error of 0.015. The upper triangular area of the heat map for $\hat{P}$ depicts the 6 edges which have absolute estimation error larger than 0.4, whereas 18 edges are highlighted for $\bar{A}$ based on the same threshold.

The smoothing effect is even more obvious when $M = 1$, as in Fig. 1(f). $\hat{P}$ smooths the estimate, especially for edges across the two hemispheres, in the lower left and corresponding upper right block (which is not shown in the heat map). Based on the calculations, $\hat{P}$, with mean squared error 0.049, outperforms $\bar{A}$, with mean squared error 0.104, a 53% relative improvement in estimation. Similarly, the same absolute estimation error threshold of 0.4 highlights 234 edges for $\hat{P}$, less than 50% as many as $\bar{A}$.

A cross validation on the 454 graphs serves to evaluate the performance of the two estimators. Specifically, for each atlas, each Monte Carlo replicate consists of sampling M

graphs out of the 454, and computing the low-rank estimator $\hat{P}$ and the sample mean $\bar{A}$ on the M graphs. These estimates are compared to the sample mean P for all 454 adjacency matrices.

To evaluate performance, the average of the ratios of the mean squared error across all vertex pairs is computed. 1000 cross-validation simulations on each of the three atlases are run for sample sizes of $M = 1, 5, 10$. For $M = 1$, only the 454 distinct possibilities are considered. To determine the rank for $\hat{P}$, we employed Zhu and Ghodsi's method [30] and USVT [26] (see Appendix A).

Fig. 2 shows the estimated relative efficiencies between $\bar{A}$ and $\hat{P}$. For each atlas and each sample size, both dimension selection methods have similar overall performance. Confidence intervals for the estimated relative efficiencies, calculated by assuming a normal distribution, all have lengths less than 0.015. All relative efficiencies are significantly different from 1.

The largest improvements using $\hat{P}$ occur when M is small and N is large, where the RE are smaller than 1. On the other hand, once $M = 10$, $\bar{A}$ tends to do nearly as well or better than $\hat{P}$, except for the larger atlas using ZG.

For the sample with size $M = 5$ from Fig. 1, Fig. 3 shows the values for the absolute estimation error $|\bar{A} - P|$ and $|\hat{P} - P|$, as well as $|\bar{A} - \hat{P}|$. The lower triangular sections show the absolute differences while the upper triangular matrix highlights vertex pairs with absolute differences larger than 0.4. There are 18 edges for $\bar{A}$ and only 6 edges for $\hat{P}$ being highlighted in the figure. Note that approximately 13% of all pairs of vertices are adjacent in all 454 graphs and hence $\bar{A}$ will always have zero error for those pairs of vertices. Nonetheless, $\hat{P}$ typically outperforms $\bar{A}$.
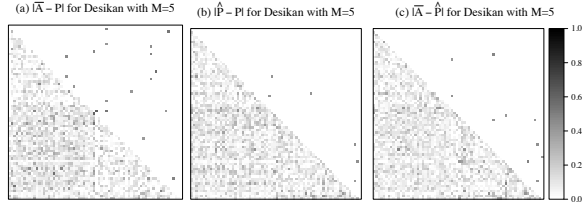
Fig. 3: For the SWU4 data with the Desikan atlas, using a sample of $M = 5$ graphs, $\bar{A}$ and $\hat{P}$ are computed (Fig. 1 (b) and (e)). Here, the lower triangular matrices show the absolute estimation error $|\bar{A} - P|$, $|\hat{P} - P|$ and $|\bar{A} - \hat{P}|$. The embedding dimension for $\hat{P}$ is $d = 11$ selected by the ZG method. The upper triangular matrix highlights the edges with absolute differences larger than 0.4, with 18 edges from $\bar{A}$ and only 6 edges from $\hat{P}$ being highlighted. Overall, $\hat{P}$ provides improved performance over $\bar{A}$ for this sample size.

A. Challenges of the SWU4 Dataset

While $\hat{P}$ performs well when the sample size M is small and the number of vertices N is large, the SWU4 dataset itself does not strictly adhere to the low-rank assumptions of our theory. Whether the dataset has strong or weak low-rank structure was investigated. In Fig. 4, the relative error $\|\text{lowrank}_d(P) - P\|_F^2 / \|P\|_F^2$ of using a rank-d approximation of P (see Algorithm 1) is plotted as solid curves. The rate at which this curve tends to zero provides an indication of the relative increase in error when using $\text{lowrank}_d(\bar{A})$ as compared to $\bar{A}$, when M is large. For all three atlases, substantial errors remain for any low-rank $(< n/2)$ approximation. This can be compared to the dashed lines which show how these error increases would behave if P was truly low-rank where the ranks are selected by Zhu and Ghodsi's method, 13 for JHU, 8 for Desikan, and 37 for CPAC200.
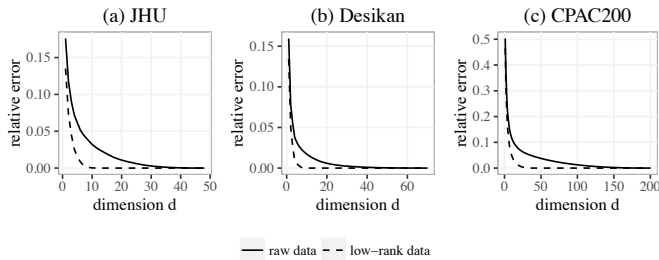

Fig. 4: The solid curves show the relative error $\|\text{lowrank}_d(P) - P\|_F^2 / \|P\|_F^2$ of using a rank-d approximation of P for three different atlases. The values of this curve indicate the increase in relative error when using $\text{lowrank}_d(\bar{A})$ as compared to $\bar{A}$, when M is large. The relative error increases decay relatively slowly , indicating that P is not well approximated by a low-rank matrix. If P were actually low-rank, 13, 8, and 37, respectively, the relative error is plotted with the dashed curves.

While these challenges can negatively impact the performance of low-rank procedures when estimating P, we can also view $\hat{P}$ as estimating latent low-rank structure in P. For

such an estimand, $\hat{P}$ will provide excellent performance, even for large M.

To illustrate this, we considered estimating the d^* approximation of P, denoted as P_{d^*}, where d^* is chosen according the Zhu and Ghodsi method applied to the eigenvalues of P. For the SWU4 dataset with the CPAC 200 atlas, Figure 5 shows the estimated relative efficiency, on a log-scale, of $\bar{A}$ compared to $\hat{P}$, for estimating P, red solid line, and P_{d^*}, dashed blue line. The relative efficiency is estimated based on 100 monte carlo replicates of sampling M graphs for $M \in \{1, 2, 5, 10, 20, 50, 100\}$. Recalling that relative efficiencies below one favor $\hat{P}$, for large M, $\hat{P}$ has poor performance compared to $\bar{A}$ for estimating P but excellent relative efficiency for estimating P_{d^*}.
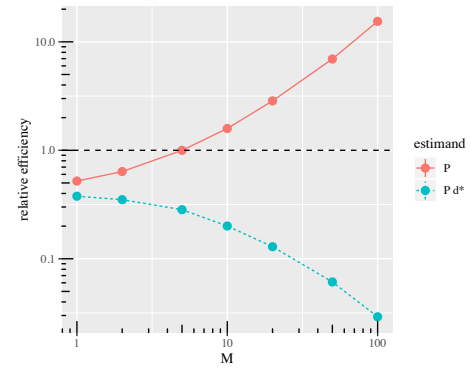
Fig. 5: For the CPAC 200 atlas, the estimated relative efficiency (RE) of $\bar{A}$ compared to $\hat{P}$ on a log-scale, where values below one favor $\hat{P}$. The red solid line is the RE for estimating P and the blue dashed line is the RE for estimating P_{d^*} The horizontal axis denotes the sample size M on a log-scale. While $\hat{P}$ is a relatively poor estimator of P for large M, the reverse is true for P_{d^*}, where $\hat{P}$ is clearly superior to $\bar{A}$, even for large M.

This dataset has other challenges for low-rank methods. First, there are a large number of negative eigenvalues which $\hat{P}$ will not capture.

Low-rank methods can include large negative eigenvalues, however, for low sample sizes excluding negative eigenvalues improved performance for SWU4. See Appendix C for a comparison of these and other parameters in the CPAC200 atlas.

Second, approximately 12.8% of the entries of P are exactly equal to 1. For these edges, $\bar{A}$ will have exactly zero error, while $\hat{P}$ will be a less accurate estimate.

Despite these challenges, when the sample size is relatively small, such as $M = 1$ or $M = 5$, and for a larger number of vertices, $\hat{P}$ gives a better estimate than $\bar{A}$ for the SWU4 dataset. (See Appendix H for a similar synthetic data analysis.) Importantly, this improvement is robust to the embedding dimension as illustrated in Appendix A.1.

B. Eigen-connectomes and Lobe Structure

In addition to yielding potentially improved estimation, low-rank methods simultaneously provide convenient interpretations.

Using the mean graph P for the Desikan atlas, the average of all 454 graphs, we estimated latent positions $\hat{X} \in \mathbb{R}^{N \times d}$. Here, $N = 70$ is the number of vertices and $d = 8$ is the dimension selected by the Zhu and Ghodsi's method [30]. Fig. 6 shows the first 4 dimensions of $\hat{X}$ in the brain space. The value of $\hat{X}_{ij}$ determines the color of the i-th brain region for the j-th dimension; i.e. the j-th entry of the estimated latent vector for the i-th region. Red represents a positive value while blue represents a negative one, and the darker the color, the smaller the magnitude of the $\hat{X}_{ij}$.

The 1st dimension, depicted in Panel (a) of the figure, correlates strongly with degrees for each vertex. The 2nd dimension, panel (b), shows a distinction between the left and right hemisphere. Similarly, the other dimensions qualitatively correspond to different lobes. For example, the red color corresponds to the Frontal and Temporal lobes in the 3rd dimension, while the light blue roughly matches the Occipital lobe in the 4th dimension.

These "eigen-connectomes" demonstrate noteworthy similarity to the structural connectome harmonics of [11]. While prior work has established the utility of eigen-connectomes as a basis set for modeling of functional dynamics, the current study demonstrates the correspondence of these dimensions with lobular divisions. The connection between eigenvectors corresponding to each dimension and lobes will be explored more rigorously in Section V-B. Future work integrating the connections between structural features with the modeling of dynamics may provide insight into the spatial distribution of large-scale cortical hierarchies [12]. Additionally, such representations enable the use of techniques from multivariate analysis to further study the mean graph.

As an example, the brain can be divided into lobes, originally based purely on anatomical considerations, now widely recognized to also play a functional role [41]. While different anatomists partition brain regions differently, there is general agreement on four cortical lobes per hemisphere: frontal, parietal, occipital, and temporal [42], [39], [43].

For the Desikan atlas, there are 70 regions (35 regions for each hemisphere), with each region belonging to a single lobe (see Appendix E).

One might hypothesize that properties of regions within a lobe are more similar than across lobes, as regions within lobes would be expected to share more functional roles. To test whether the embedded latent positions X preserve this property or not, we propose a test statistic T to be the average differences between vertices within the same lobe minus the average differences between vertices across different lobes, i.e.

$$T(X,l) = \sum_{i \neq j} \frac{\mathbf{1}_{l(i)=l(j)}\|X_i - X_j\|_2}{\sum_{k \neq l} \mathbf{1}_{l(k)=l(l)}} - \frac{\mathbf{1}_{l(i) \neq l(j)}\|X_i - X_j\|_2}{\sum_{k \neq l} \mathbf{1}_{l(k) \neq l(l)}}, \quad (2)$$

where $l(i)$ denotes the lobe assignment for vertex i. If the latent positions X and the lobe assignment l are independent, then $T(X,l)$ will be close to zero. A small test statistic $T(X,l)$ indicates that latent positions of the regions within the same lobe are closer compared to the ones across the lobes.

However, the anatomical geometry might contribute to the

(a) 1st dimension

(b) 2nd dimension

(c) 3rd dimension
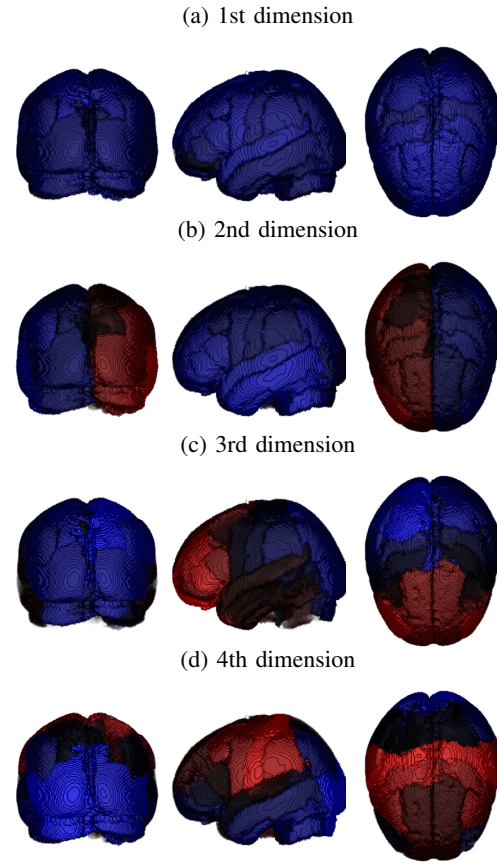
(d) 4th dimension

Fig. 6: Each panel depicts the values of the latent positions for the regions in the Desikan atlas for the mean graph of the SWU4 dataset. The color of the i-th brain region for the j-th dimension is determined by the value of $\hat{X}_{ij}$, i.e. the j-th element of the estimated latent vector for the i-th region. Red represents a positive value while blue represents the negative one, with brighter color indicating larger magnitudes. The 1st dimension, depicted in Panel (a), is relatively flat across the entire brain. In Panel (b), there is a distinction of the left and right hemispheres as conveyed in the 2nd dimension. Similarly, the other dimensions appear to correspond closely with the anatomical lobe structures of the brain.

dependence between X and l, with spatially proximal vertices having similar connectivity patterns. Hence, a small test statistic $T(X,l)$ is evidence that the low-rank methods preserve the lobe structure only if we also condition on anatomy geometry: $H_0 : X$ and l are conditionally independent given anatomical geometry, $H_A : X$ and l are conditionally dependent given anatomical geometry. The test of conditional independence has less power compared to the test of unconditional independence which is performed with a random permutation of lobe labels (which yield a p-value $< 10^{-6}$).

To test under the anatomical geometry conditions, the lobe assignments $l(i)$ were randomly modified so that the number of regions in each lobe remain the same and the regions within the same modified lobe are still spatially connected. In particular, we performed a sequence of randomized flips, where a flip is a swap of two pairs of vertices which preserves

the number of regions in each lobe and maintains the constraint that lobes are spatially contiguous. The lobes are flipped a limited number of times in order to study how the number of flips impacts the test statistic. Appendix D discusses the flipping procedure.

1000 simulations with the test statistics of $T(X, l')$ are performed, each with a fixed number of flips. The number of flips varies from 1 to 10 (Fig. 7). In the violin plot, the dashed line indicates the value of $T(X, l)$ based on the true lobe assignment. The p-value is less than 0.05 if the number of flips is larger than 7. Hence, latent positions in the same lobe are more similar to each other, even after accounting for the fact that geometrically proximal regions may also have similar latent positions.

When the number of flips is small, this test has very little power, with the null distribution being only a small deviation from the original lobes. When the number of flips gets large, eventually the contiguity of the lobes breaks down and the empirical p-values continue to get smaller.
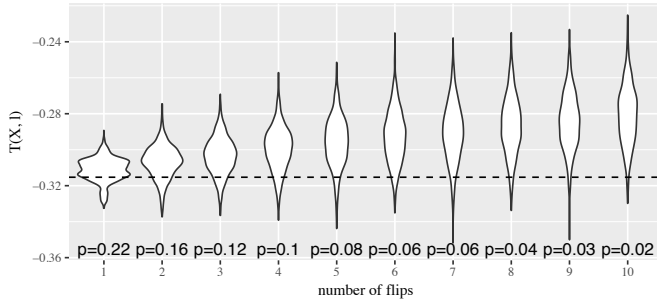

Fig. 7: Violin plot for 1000 simulations of $T(X, l')$ with l' randomly drawn from the set of valid flips, preserving lobe size and contiguity. The dashed line represents the statistic $T(x, l)$ for the true lobes. As the null hypotheses move further from the original lobe assignment, with the number of flips growing, the 95% central region of the null shifts away from $T(x, l)$ under the true lobe assignment. The p-value for each test is provided at the bottom of the figure.

VI. APPLICATION TO A MOUSE CONNECTOME

As a further application of low-rank methods, an MRI-DTI mouse brain connectome [33] with M=1 specimen was evaluated. The data acquisition protocol is described in the appendix, Appendix F, and resulted in a 296 node weighted, directed graphs with vertices again corresponding to regions in the brain. The 296 regions were organized into a multilevel, hierarchical structure. Analysis of the fine-grained and the first level of the hierarchy partitioned the label set into eight superstructures, with four in each hemisphere: forebrain, midbrain, hindbrain, and white matter.

The original matrix $W \in (\mathbb{R}^+)^{296 \times 296}$ is a weighted adjacency matrix with W_{ij} denoting the number of tracts passing through ROIs i and j. As the original weights had very heavy tails, these weights were transformed by setting $A_{ij} = \log(W_{ij} + 1)$. This resulted in the weighted adjacency matrix in Figure 8 (a). Note that Algorithm 2 does not strictly require the entries to be binary and hence is applicable in this weighted setting by only thresholding the elements of $\hat{P}$ to be non-negative.
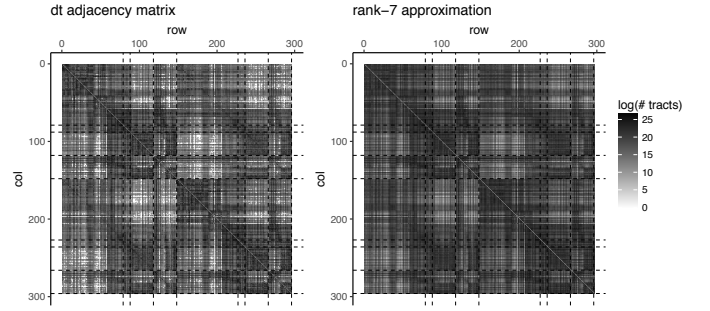
Fig. 8: The left panel shows the weighted adjacency matrix with weights transformed by the transformation $w \mapsto \log(w + 1)$. Higher weights are shown as darker pixels. The right panel shows the rank-7 approximation of this matrix where the rank was chosen using method in [30]. In both panels, dashed lines show the division between the eight different superstructures.

Using the procedure described in Algorithm 2, the dimension selection procedure [30] resulted in a rank-7 approximation which is shown in the right panel of Figure 8. Since the sample size is only one, cross-validation cannot be employed, but visually it appears that the rank-7 approximation captures many of the features in the original matrix.

As with the human data, we studied the relationship between the structure of the graph and the eight superstructures. Panel (a) of Figure 9 shows entries grouped by superstructure of the four scaled singular vectors of the weighted adjacency matrix A corresponding to the largest singular values. The points are colored according to the four superstructures and the shapes are determined by the hemisphere. The ordering of the points groups together nodes in the same superstructure and hemisphere. The second and fourth vectors have structure which correlates closely with the four superstructures and the two hemispheres, respectively. Additionally, the first vector appears to separate the midbrain from the other three superstructures.

The right panel of Figure 9 shows a scatter plot of the entries of the fourth and second vectors along with the class boundaries for the eight-class quadratic discriminant analysis classifier. This classifier achieves a training error rate of $87/296 \approx 0.29$. The error rate is particularly high for the white matter, with $58/60$ vertices being classified incorrectly, meaning that ignoring the white matter, $29/236 \approx 0.12$ vertices were misclassified. Panel (c) shows the normalized confusion matrix for the eight classes, indicating that the forebrain and hindbrain classes are well separated while the white matter and midbrain have more substantial overlap. This matches with the general structure of the white matter which is not defined at the first hierarchical level of the atlas according to spatial structure, while the fore-, mid-, and hindbrain superstructures are.

Finally, the same permutation analysis as Fig. 7 was also performed for the mouse connectome as shown in Fig. 14 in Appendix D. This test again indicates that the eight superstructures are significant even after accounting for the spatial

structure of the regions.

VII. DISCUSSION

Motivated by the RDPG model, via a low-rank approximation to the entry-wise MLE, our methodology takes advantage of low-rank structure of graphs. We give a closed form for the asymptotic relative efficiency between the entry-wise MLE $\bar{A}$ and its low-rank approximation $\text{lowrank}_d(\bar{A})$ in the case of a SBM, showing theoretically that low-rank methods can provide substantial improvements. Our theoretical results are all shown for the large N regime in terms of $\text{lowrank}_d(\bar{A})$, however in practice we have observed that our proposed estimator $\hat{P}$ has even better performance, and asymptotically the difference between them will typically be small under dense low-rank models. Moreover, our estimator outperforms the entry-wise MLE in a cross validation analysis of the SWU4 brain graphs and in low- and full-rank simulation settings when M is small. These results illustrate that $\hat{P}$ performs well even when the low-rank assumption is violated and that $\hat{P}$ is robust and can be applied in practice.

The low-rank methods could also be applied for functional MRI studies by applying them to appropriate correlation matrices or other network estimates. We have performed preliminary explorations in that setting where $\hat{P}$ again performs well for estimating latent structure. Since static fMRI graphs likely have lower noise than structural connectomes, the estimator $\bar{A}$ generally performed well for estimating the population mean graph.

As we have shown in Figure 5, $\hat{P}$ is an excellent estimate of the low-rank latent structure of P. As other authors have noted [4], many random network models will enjoy strong low-rank structure in their mean graph which will be readily captured by $\hat{P}$.

For the human connectome data, the largest improvements using the low-rank method occurred when the number of graphs M was small, while it provided only minor improvements, or even slightly degraded performance, when M was large. However, even in large scale studies, low-rank methods will be useful for estimating graph means for subpopulations, e.g. the population of females over 60 with some college education. Using the element-wise sample mean for such small strata, which may have fewer than ten subjects, will frequently result in a degradation of performance. Similarly, [44] proposed a Bayesian nonparametric approach for modeling the population distribution of network-valued data which reduces dimensionality via a mixture model and our methods could be easily adapted to those ideas.

While the low-rank methods considered in this paper may perform well, further refinements of these methods which account for the particular traits of connectomics data would be useful to improve estimation further. For example, we assume that the adjacency matrix is observed without contamination. However, when heavy-tailed noise is present, robust methods may be necessary. Rank-based methods and robust likelihood methods could be very useful in that case [45], [46]. [47] considers this problem and proposes the use of L^q likelihood to improve both the low-rank and full rank methods for estimating P.

Another issue that arose in the analysis of the connectome dataset was the presence of structural ones in the mean graph for the population. These structural ones appear since edges between certain regions of the brain are present in all members of the healthy population. For these always-present edges, the low-rank methods will have non-zero error while the sample mean will always have zero error. Detecting and incorporating structural ones and zeros could yield methods that share the best elements of both methods considered here.

While in this paper the focus is the estimation of the mean graph P exploiting the low-rank structure, many future directions are quite interesting, such as fitting the SBM or clustering the vertices to detect different brain regions. For example, [48] introduced a region-extraction approach based on a sparse penalty with dictionary learning; [49] performed independent component analysis of fMRI data to draw group inferences; [50] consider a joint embedding model for feature extractions.

REFERENCES

[1] G. V. Trunk, "A problem of dimensionality: A simple example," *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, no. 3, pp. 306–307, 1979.

[2] C. E. Ginestet, P. Balanchandran, S. Rosenberg, and E. D. Kolaczyk, "Hypothesis testing for network data in functional neuroimaging," *arXiv preprint arXiv:1407.5525*, 2014.

[3] S. Bhattacharyya and S. Chatterjee, "Spectral clustering for multiple sparse networks: I," May 2018.

[4] M. Udell, C. Horn, R. Zadeh, and S. Boyd, "Generalized low rank models," *Foundations and Trends® in Machine Learning*, vol. 9, no. 1, pp. 1–118, 2016.

[5] C. Eckart and G. Young, "The approximation of one matrix by another of lower rank," *Psychometrika*, vol. 1, no. 3, pp. 211–218, Sep. 1936.

[6] M. W. Berry, M. Browne, A. N. Langville, V. P. Pauca, and R. J. Plemmons, "Algorithms and applications for approximate nonnegative matrix factorization," *Computational statistics & data analysis*, vol. 52, no. 1, pp. 155–173, Sep. 2007.

[7] C. Boutsidis and E. Gallopoulos, "SVD based initialization: A head start for nonnegative matrix factorization," *Pattern recognition*, vol. 41, no. 4, pp. 1350–1362, Apr. 2008.

[8] N. H. Lee, I.-J. Wang, Y. Park, C. E. Priebe, and M. Rosen, "Automatic dimension selection for a non-negative factorization approach to clustering multiple random graphs," *Stat*, vol. 1050, p. 24, 2014.

[9] A. Banerjee and J. Jost, "Spectral plots and the representation and interpretation of biological data," *Theory in biosciences = Theorie in den Biowissenschaften*, vol. 126, no. 1, pp. 15–21, Aug. 2007.

[10] M. Rahim, B. Thirion, and G. Varoquaux, "Population-Shrinkage of covariance to estimate better brain functional connectivity," in *Medical Image Computing and Computer Assisted Intervention MICCAI 2017*. Springer International Publishing, 2017, pp. 460–468.

[11] S. Atasoy, I. Donnelly, and J. Pearson, "Human brain networks function in connectome-specific harmonic waves," *Nature communications*, vol. 7, p. 10340, Jan. 2016.

[12] D. S. Margulies, S. S. Ghosh, A. Goulas, M. Falkiewicz, J. M. Huntenburg, G. Langs, G. Bezgin, S. B. Eickhoff, F. X. Castellanos, M. Petrides, E. Jefferies, and J. Smallwood, "Situating the default-mode network along a principal gradient of macroscale cortical organization," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 113, no. 44, pp. 12 574–12 579, Nov. 2016.

[13] S. C. de Lange, M. A. de Reus, and M. P. van den Heuvel, "The laplacian spectrum of neural networks," *Frontiers in computational neuroscience*, vol. 7, p. 189, Jan. 2014.

[14] W. Gray, J. Bogovic, J. Vogelstein, B. Landman, J. Prince, and R. Vogelstein, "Magnetic resonance connectome automated pipeline: An overview," *IEEE Pulse*, vol. 2, no. 3, pp. 42–48, 2012.

[15] B. Bollobás, S. Janson, and O. Riordan, "The phase transition in inhomogeneous random graphs," *Random Structures & Algorithms*, vol. 31, no. 1, pp. 3–122, 2007.

[16] E. N. Gilbert, "Random graphs," *Annals of Mathematical Statistics*, vol. 30, no. 4, pp. 1141–1144, 1959.
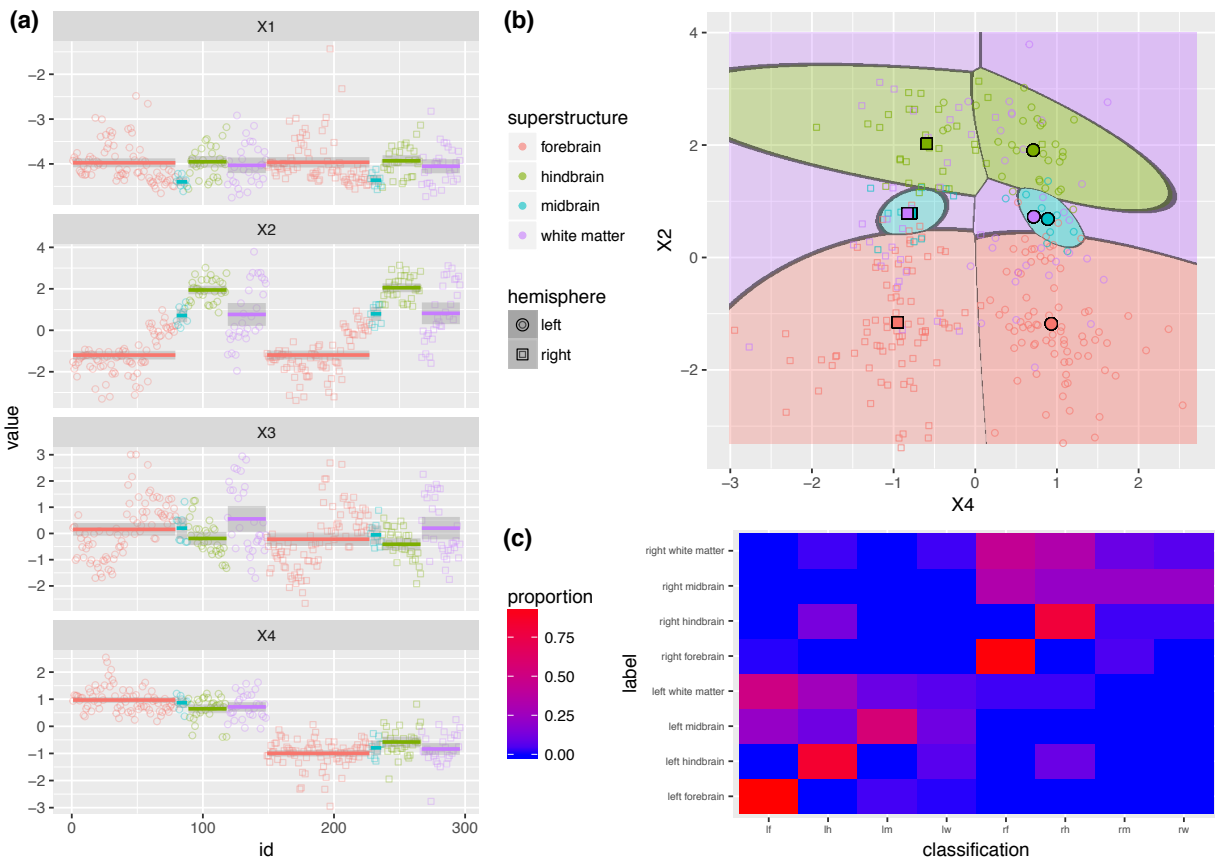
Fig. 9: Panel (a) shows entries of the the 4 scaled singular vectors of A corresponding to largest singular values. The ordering of the entries is the same as the ordering in Figure 8, where vertices in the same superstructure and hemispheres are grouped together. Colors of points denote the superstructures and shapes denote the hemispheres. Panel (b) is a scatter plot of the the fourth (horizontal axis) and second (vertical axis) singular vectors. The background coloring is determined by learning a classifier, based on a mixture of Gaussians, to classify the eight different superstructures. Panel (c) shows the normalized confusion matrix for the mixture of Gaussians learned in panel (b). Each entry corresponds to the proportion of nodes with a given label that were predicted to be each other label, with true and predicted labels given by the row and column, respectively.

[17] P. Erdős and A. Rényi, "On random graphs, I," *Publicationes Mathematicae Debrecen*, vol. 6, pp. 290–297, 1959.

[18] S. J. Young and E. R. Scheinerman, "Random dot product graph models for social networks," in *Algorithms and models for the web-graph*. Springer, 2007, pp. 138–149.

[19] C. L. M. Nickel, "Random dot product graphs a model for social networks," Ph.D. dissertation, Johns Hopkins University, 2008.

[20] P. D. Hoff, A. E. Raftery, and M. S. Handcock, "Latent space approaches to social network analysis," *Journal of the american Statistical association*, vol. 97, no. 460, pp. 1090–1098, 2002.

[21] P. W. Holland, K. B. Laskey, and S. Leinhardt, "Stochastic blockmodels: First steps," *Social networks*, vol. 5, no. 2, pp. 109–137, 1983.

[22] E. M. Airoldi, D. M. Blei, S. E. Fienberg, and E. P. Xing, "Mixed membership stochastic blockmodels," *Journal of Machine Learning Research*, vol. 9, no. Sep, pp. 1981–2014, 2008.

[23] B. Karrer and M. E. Newman, "Stochastic blockmodels and community structure in networks," *Physical Review E*, vol. 83, no. 1, p. 016107, 2011.

[24] 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*, vol. 8, no. 2, pp. 2905–2922, 2014.

[25] P. Rubin-Delanchy, C. E. Priebe, and M. Tang, "Consistency of adjacency spectral embedding for the mixed membership stochastic blockmodel," May 2017.

[26] S. Chatterjee, "Matrix estimation by universal singular value thresholding," *The Annals of Statistics*, vol. 43, no. 1, pp. 177–214, 2015.

[27] D. Marchette, C. Priebe, and G. Coppersmith, "Vertex nomination via

attributed random dot product graphs," in *Proceedings of the 57th ISI World Statistics Congress*, vol. 6, 2011, p. 16.

[28] E. R. Scheinerman and K. Tucker, "Modeling graphs using dot product representations," *Computational Statistics*, vol. 25, no. 1, pp. 1–16, 2010.

[29] D. L. Sussman, M. Tang, and C. E. Priebe, "Consistent latent position estimation and vertex classification for random dot product graphs," *IEEE transactions on pattern analysis and machine intelligence*, vol. 36, no. 1, pp. 48–57, 2014.

[30] M. Zhu and A. Ghodsi, "Automatic dimensionality selection from the scree plot via the use of profile likelihood," *Computational Statistics & Data Analysis*, vol. 51, no. 2, pp. 918–930, 2006.

[31] A. Athreya, C. E. Priebe, M. Tang, V. Lyzinski, D. J. Marchette, and D. L. Sussman, "A limit theorem for scaled eigenvectors of random dot product graphs," *Sankhya A*, vol. 78, no. 1, pp. 1–18, 2016.

[32] D. E. Fishkind, D. L. Sussman, M. Tang, J. T. Vogelstein, and C. E. Priebe, "Consistent adjacency-spectral partitioning for the stochastic block model when the model parameters are unknown," 2012.

[33] E. Calabrese, A. Badea, G. Cofer, Y. Qi, and G. A. Johnson, "A diffusion MRI tractography connectome of the mouse brain and comparison with neuronal tracer data," *Cereb. Cortex*, vol. 25, no. 11, pp. 4628–4637, Nov. 2015.

[34] M. Tang and C. E. Priebe, "Limit theorems for eigenvectors of the normalized laplacian for random graphs," *arXiv preprint arXiv:1607.08601*, 2016.

[35] X.-N. Zuo, J. S. Anderson, P. Bellec, R. M. Birn, B. B. Biswal, J. Blautzik, J. C. Breitner, R. L. Buckner, V. D. Calhoun, F. X. Castellanos *et al.*, "An open science resource for establishing reliability and reproducibility in functional connectomics," *Scientific data*, vol. 1, 2014.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TMI.2018.2885968, IEEE Transactions on Medical Imaging

12

[36] G. Kiar, K. J. Gorgolewski, D. Kleissas, W. G. Roncal, B. Litt, B. Wandell, R. A. Poldrack, M. Wiener, R. J. Vogelstein, R. Burns *et al.*, "Science in the cloud (sic): A use case in mri connectomics," *Giga Science*, vol. 6, no. 5, pp. 1–10, 2017.

[37] G. Kiar, W. Gray Roncal, D. Mhembere, E. Bridgeford, R. Burns, and J. T. Vogelstein, "ndmg: Neurodata's mri graphs pipeline," http://m2g.io, Aug. 2016.

[38] K. Oishi, A. V. Faria, P. C. van Zijl, and S. Mori, *MRI atlas of human white matter*. Academic Press, 2010.

[39] R. S. Desikan, F. Ségonne, B. Fischl, B. T. Quinn, B. C. Dickerson, D. Blacker, R. L. Buckner, A. M. Dale, R. P. Maguire, B. T. Hyman *et al.*, "An automated labeling system for subdividing the human cerebral cortex on mri scans into gyral based regions of interest," *Neuroimage*, vol. 31, no. 3, pp. 968–980, 2006.

[40] S. Sikka, B. Cheung, R. Khanuja, S. Ghosh, C. Yan, Q. Li, J. Vogelstein, R. Burns, S. Colcombe, C. Craddock *et al.*, "Towards automated analysis of connectomes: The configurable pipeline for the analysis of connectomes (c-pac)," in *5th INCF Congress of Neuroinformatics, Munich, Germany*, vol. 10, 2014.

[41] T. Vanderah and D. J. Gould, *Nolte's The Human Brain: An Introduction to its Functional Anatomy, 7e*, 7th ed. Elsevier, 1 Jun. 2015.

[42] B. Fischl, "Freesurfer," *Neuroimage*, vol. 62, no. 2, pp. 774–781, 2012.

[43] D. H. Salat, R. L. Buckner, A. Z. Snyder, D. N. Greve, R. S. Desikan, E. Busa, J. C. Morris, A. M. Dale, and B. Fischl, "Thinning of the cerebral cortex in aging," *Cerebral cortex*, vol. 14, no. 7, pp. 721–730, 2004.

[44] D. Durante, D. B. Dunson, and J. T. Vogelstein, "Nonparametric bayes modeling of populations of networks," *Journal of the American Statistical Association*, no. just-accepted, 2016.

[45] P. J. Huber and E. M. Ronchetti, *Robust statistics*, 2nd ed., ser. Wiley Series in Probability and Statistics. John Wiley & Sons, Inc., Hoboken, NJ, 2009.

[46] Y. Qin and C. E. Priebe, "Maximum L_q-likelihood estimation via the expectation-maximization algorithm: A robust estimation of mixture models," *Journal of the American Statistical Association*, vol. 108, no. 503, pp. 914–928, 2013.

[47] R. Tang, M. Tang, J. T. Vogelstein, and C. E. Priebe, "Robust estimation from multiple graphs under gross error contamination," Jul. 2017.

[48] A. Abraham, E. Dohmatob, B. Thirion, D. Samaras, and G. Varoquaux, "Extracting brain regions from rest fmri with total-variation constrained dictionary learning," in *MICCAI-16th International Conference on Medical Image Computing and Computer Assisted Intervention-2013*. Springer, 2013.

[49] V. D. Calhoun, T. Adali, G. D. Pearlson, and J. Pekar, "A method for making group inferences from functional mri data using independent component analysis," *Human brain mapping*, vol. 14, no. 3, pp. 140–151, 2001.

[50] S. Wang, J. T. Vogelstein, and C. E. Priebe, "Joint embedding of graphs," *arXiv preprint arXiv:1703.03862*, 2017.