APPENDIX A
METHODS

A.1 Algorithm for low-rank approximation

Here we provide the algorithm for computing a low-rank positive semidefinite approximation of a matrix.

Algorithm 2 Rank-d approximation of a matrix.

Input: Symmetric matrix $A \in \mathbb{R}^{N \times N}$, dimension $d \leq N$.
Output: $\text{lowrank}_d(A) \in \mathbb{R}^{N \times N}$

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: $\hat{S} \leftarrow \text{diag}(s_1, \ldots, s_d)$ and $\hat{U} \leftarrow [u_1, \ldots, u_d]$;
3: Return $\hat{U}\hat{S}\hat{U}^\top$;

A.2 Choosing Dimension

Often in dimensionality reduction techniques, the choice for dimension d, relies on analyzing the set of the ordered eigenvalues, looking for a "gap" or "elbow" in the scree-plot. Zhu and Ghodsi (2006) present an automated method for finding this gap in the scree-plot that takes only the ordered eigenvalues as an input and uses Gaussian mixture modeling to find these gaps. The mixture modeling results in multiple candidate dimensions or elbows, and our analysis indicated that underestimating the dimension is much more harmful than overestimating the dimension. For this reason, the 3rd elbow was employed in the experiments performed for this work. While Zhu and Ghodsi (2006) only defines the 1st elbow, we define the s-th elbow as in Algorithm 3.

Algorithm 3 Algorithm to compute the Zhu and Ghodsi's elbow

Input: The number of Zhu and Ghodsi's elbow s, with eigenvalues $\lambda_1, \ldots, \lambda_N$
Output: The s-th Zhu and Ghodsi's elbow

1: Calculate the 1st elbow d_1 based on $\lambda_1, \ldots, \lambda_N$ according to Zhu and Ghodsi (2006)
2: **for** i = 2 to s **do**
3: Calculate the i-th elbow d_i based on $\lambda_{d_{i-1}+1}, \ldots, \lambda_N$ according to Zhu and Ghodsi (2006)
4: **end for**

Universal Singular Value Thresholding (USVT) is a simple estimation procedure proposed in Chatterjee (2015) that can work for any matrix that has "a little bit of structure". In the current setting, it selects the dimension d as the number of singular values that are greater than a constant c times $\sqrt{N/M}$. The specific constant c must be selected carefully based on the mean and variance of the entries, and since overestimating the dimension was not overly harmful, we chose a relatively small value of $c = 0.7$.

Overall, selecting the appropriate dimension is a challenging task and numerous methods could be applied successfully depending on the setting. On the other hand, in our setting, many dimensions will yield nearly optimal mean squared errors and the two methods did not pick drastically different dimensions. Thus efforts to ensure the selected dimension is in the appropriate range are more important than finding the best dimension.

A.2.1 Exploration of Dimension Selection Procedures

To further investigate the impact of the dimension selection procedures, we also considered all possible dimensions for $\hat{P}$ by ranging d from 1 to N. $\hat{\text{MSE}}$ of $\bar{A}$ and $\hat{P}$ was plotted in Fig. 15. The horizontal axis gives dimension d, which only impacts $\hat{P}$, which is why estimated MSE of $\bar{A}$ is shown as flat. When d is small, $\hat{P}$ underestimates the dimension and throws away important information, which leads to relatively poor performance. When $d = N$, $\hat{P}$ is equal to $\bar{A}$, so that the curve for $\hat{\text{MSE}}$ for $\hat{P}$ ends at $\hat{\text{MSE}}(\bar{A})$. In the figure, a triangle denotes the 3rd elbow found by the Zhu and Ghodsi method, and a square denotes the dimension selected by USVT with threshold 0.7. Both dimension selection algorithms tend to select dimensions which nearly minimize the mean squared error.

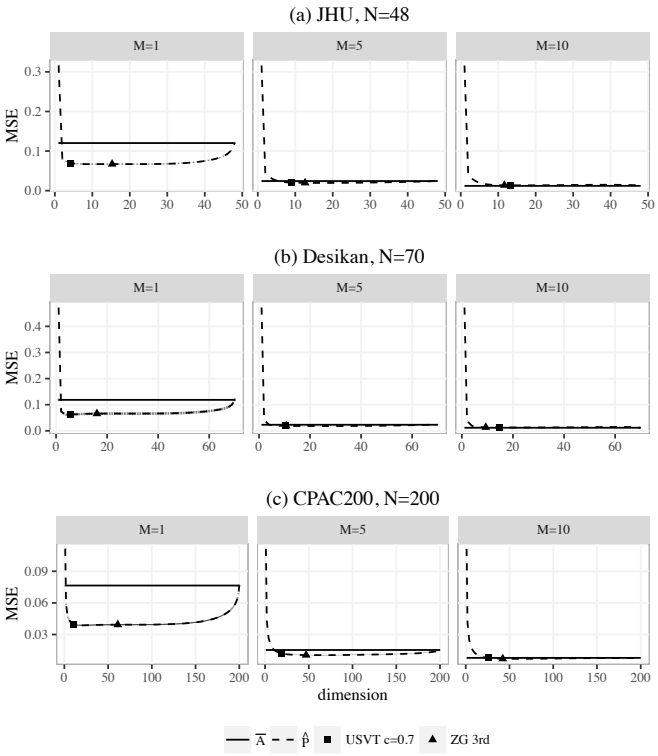

Fig 15: **Comparison of $\hat{\text{MSE}}$ of $\hat{P}$ and $\bar{A}$ for three atlases at three sample sizes for the CoRR data.** These plots show the mean squared error for $\bar{A}$ (solid line) and $\hat{P}$ (dashed line) for three datasets (JHU, Desikan, and CPAC200) while embedding the graphs into different dimensions and with different sample sizes M. The average dimensions chosen by the 3rd elbow of Zhu and Ghodsi is denoted by a triangle and those chosen by USVT with threshold equaling 0.7 is denoted by a square. Vertical intervals, visible mainly in the $N = 48, 70$ and $M = 1$ plots, represent the 95% confidence interval for the mean squared errors. When M is small, $\hat{P}$ outperforms $\bar{A}$ with a flexible range of the embedding dimension including the average of the dimensions selected by Zhu and Ghodsi and USVT.

When M is 1 or 5, $\bar{A}$ has large variance which leads to large MSE. Meanwhile, $\hat{P}$ reduces the variance by taking advantages of inherent low-rank structure of the mean graph. Such smoothing effect is especially obvious while there is only 1 observation. When $M = 1$, all weights of the graph are either 0 or 1, leading to a very bumpy estimate $\bar{A}$. In this case, $\hat{P}$ smooths the connectomes estimate and improves the performance. Additionally, there is a large range of dimensions where the performance for $\hat{P}$ is superior to $\bar{A}$. With a larger M, the performance of $\bar{A}$ improves so that its performance is frequently superior but nearly identical to $\hat{P}$.

A.3 Graph Diagonal Augmentation

The graphs examined in this work have no self-loops and thus the diagonal entries of the adjacency matrix and the mean graph are all zero. However, when computing the low-rank approximation, these structural zeros lead to increased errors in the estimation of the mean graph. While this problem has been investigated in the single graph setting, with multiple graphs, the problem is exacerbated since the variance of the other entries is lower, so the relative impact of the bias in the diagonal entries is higher. Moreover, the sum of eigenvalues of the hollow matrix will be zero, leading to an indefinite matrix, which violates the positive semi-definite assumption. So it is important to remedy the situation that we don't observe the diagonal entries.

Marchette et al. (2011) proposed the simple method of imputing the diagonals to be equal to the average of the non-diagonal entries for the corresponding row, or in equivalently the degree of the vertex divided by $n - 1$. Earlier, Scheinerman and Tucker (2010) proposed using an iterative method to impute the diagonal entries. In this work, these two ideas are combined by first using the row-average method (see Step 3 of Algorithm 1) and then using one step of the iterative method (see Step 6 of Algorithm 1). Note that when computing errors, the diagonal entries are omitted since these are known to be zero.

A.4 Dataset Description

A.4.1 Human Connectomes

The original dataset is from the Emotion and Creativity One Year Retest Dataset provided by Qiu, Zhang and Wei from Southwest University available at the Consortium for Reliability and Reproducibility (CoRR) Zuo et al. (2014); Gorgolewski et al. (2015). It is composed of 235 subjects, all of whom were college students. Each subject underwent two sessions of anatomical, resting state DTI scans, spaced one year apart. Due to incomplete data, only 454 scans are available.

When deriving MR connectomes, the NeuroData team parcellates the brain into groups of voxels as defined by anatomical atlases Kiar et al. (2016). The atlases are defined either physiologically by neuroanatomists (Desikan and JHU), or are generated using an automated segmentation algorithm (CPAC200). Once the voxels in the original image space are grouped into regions, an edge is placed between two regions when there is at least one white-matter tract, derived using a tractography algorithm, connecting the corresponding two parts of the brain Garyfallidis et al. (2014). The resulting graphs are undirected, unweighted, and have no self-loops.

A.4.2 Mouse Connectome

Images of the fixed specimen were acquired on a 9.4 T small animal magnet using a 3D diffusion weighted imaging sequence. 120 unique diffusion directions were acquired using a b value of 4000 s/mm2, interleaved with 11 non-diffusion weighted scans. Images were acquired in 235 hours, and reconstructed at 43 micron resolution. The mouse brain was labeled with 296 regions of interest, 148 per hemisphere Anderson et al. (2017).

To construct a structural connectome of the mouse brain fiber data was reconstructed (max 4 fiber orientations/voxel), then probabilistic tractography was performed using FSL Behrens et al. (2007), (5000 samples per voxel, 21 μm step size, 45 degrees curvature threshold). The 296 seed regions had connectivity estimates produced by counting the number of fibers that originate from one region and fall onto all other regions. This was normalized by the volume of the seed region and resulted in a 296x296 weighted, directed graph.

Appendix B
Proofs for Theory Results

B.1 Outline for Main Theorems

Here the proof of Lemma 4.1 is outlined, which provides the approximate MSE of $\hat{P}$ in the stochastic blockmodel case. The result depends on using the asymptotic results (see Theorem B.1) for the distribution of eigenvectors from Athreya et al. (2016) which extend to the multiple graph setting in a straightforward way.

The first key observation is that since $\bar{A}$ is computed from iid observations each with expectation P, $\bar{A}$ is unbiased for P and $\mathrm{Var}(A_{ij}) = \frac{1}{M} P_{ij}(1 - P_{ij})$. The results of Athreya et al. (2016) provide a central limit theorem for estimates of the latent position in an RDPG model for a single graph. Theorem B.1 describes important details. Since the variance of each entry is scaled by $1/M$ in $\bar{A}$, the analogous result for $\bar{A}$ is that the estimated latent positions will follow an approximately normal distribution with variance scaled by $1/M$ compared to the variance for a single graph.

Since $\hat{P}_{ij} = \hat{X}_i^\top \hat{X}_j$ is a noisy version of the dot product of $\nu_s^\top \nu_t$ from Section 2.3 and each $\hat{X}_i$ is approximately independent and normal, we can use common results for the variance of the inner product of two independent multivariate normals Brown and Rutemiller (1977). After simplifications that occur in the stochastic blockmodel setting, we can derive that the variance of $\hat{P}_{ij}$ converges to $\left(1/\rho_{\tau_i} + 1/\rho_{\tau_j}\right) P_{ij}(1 - P_{ij})/(N \cdot M)$ as $N \to \infty$. Since the variance of $\bar{A}_{ij}$ is $P_{ij}(1 - P_{ij})/M$, the relative efficiency between $\hat{P}_{ij}$ and $\bar{A}_{ij}$ is approximately $(\rho_{\tau_i}^{-1} + \rho_{\tau_j}^{-1})/N$ when N is sufficiently large.

B.2 Proof Details

Here the proofs are presented of the results in Section 4.1. To keep the ideas clear and concise, some details are removed,

which are only slight changes to previous works. We assume the block memberships τ_i are drawn iid from a categorical distribution with block membership probabilities given by $\rho \in [0,1]^K$ where $\sum_i \rho_i = 1$. We will also assume that for a given N, the block memberships are fixed for all graphs.

We denote matrix of between-block edge probabilities by $B = \nu\nu^\top \in [0,1]^{K\times K}$ which we assume has rank K and is positive definite. By definition, the mean of the collection of graphs generated from this SBM is P, where $P_{ij} = B_{\tau_i,\tau_j}$.

We observe M graphs on N vertices $A^{(1)}, \cdots, A^{(M)}$ sampled independently from the SBM conditioned on τ. Define $\bar{A} = \frac{1}{M}\sum_{t=1}^M A^{(t)}$. Let $\hat{U}\hat{S}\hat{U}^\top$ be the best rank-d positive semidefinite approximation of $\bar{A}$, then we define $\hat{P} = \hat{X}\hat{X}^\top$, where $\hat{X} = \hat{U}\hat{S}^{1/2}$.

The proofs presented here will rely on a central limit theorem developed in Athreya et al. (2016). The theorem was modified slightly to account for the multiple graph setting and is presented in the special case of the stochastic blockmodel.

Theorem B.1 (Corollary of Theorem 1 in Athreya et al. (2016))**.** *In the setting above, let $X = [X_1, \ldots, X_N]^\top \in \mathbb{R}^{N\times d}$ have row i equal to $X_i = \nu_{\tau_i}$ (recall that τ_i are drawn from $[K]$ according to the probabilities ρ). Then there exists an orthogonal matrix W such that for each row i and j and any $z \in \mathbb{R}^d$, conditioned on $\tau_i = s$ and $\tau_j = t$,*

$$\begin{aligned}&\Pr\left\{\sqrt{N}(W\hat{X}_i - \nu_s) \le z, \sqrt{N}(W\hat{X}_j - \nu_t) \le z'\right\}\\ =&\Phi(z, \Sigma(\nu_s)/M)\Phi(z', \Sigma(\nu_t)/M) + o(1)\end{aligned} \tag{3}$$

where $\Sigma(x) = \Delta^{-1}\mathbb{E}[X_jX_j^\top(x^\top X_j - (x^\top X_j)^2)]\Delta^{-1}$ and $\Delta = \mathbb{E}[X_1X_1^T]$ is the second moment matrix, with all expectations taken unconditionally. The function Φ is the cumulative distribution function for a multivariate normal with mean zero and the specified covariance, and $o(1)$ denotes a function that tends to zero as $N \to \infty$.

The proof of this result follows very closely the proof of the result in the original paper with only slight modifications for the multiple graph setting.

We now prove a technical lemma which yields the simplified form for the variance under the stochastic blockmodel.

Lemma B.2. *In the same setting as Theorem 4.2, for any $1 \le s, t \le K$:*

$$\nu_s^\top\Sigma(\nu_t)\nu_s = \frac{1}{\rho_s}\nu_s^\top\nu_t\ (1 - \nu_s^\top\nu_t).$$

Proof. Under the stochastic blockmodel with parameters (B,ρ), we have $X_i \overset{iid}{\sim} \sum_{k=1}^K \rho_k\delta_{\nu_k}$, where $\nu = [\nu_1, \cdots, \nu_K]^\top$ satisfies $B = \nu\nu^\top$. Without loss of generality, it can be assumed that $\nu = US$ where $U = [u_1, \cdots, u_K]^\top$ is orthonormal in columns and S is a diagonal matrix. Here it can be concluded that $\nu_s^\top = u_s^\top S$. Defining $R = \text{diag}(\rho_1, \cdots, \rho_K)$, allows

$$\Delta = \mathbb{E}[X_1X_1^\top] = \sum_{k=1}^K \rho_k\nu_k\nu_k^\top = \nu^\top R\nu = SU^\top RUS.$$

Thus

$$\begin{aligned}\nu_s^\top\Sigma(\nu_t)\nu_s =& \sum_{k=1}^K \nu_s^\top\Delta^{-1}\rho_k\nu_k\nu_k^\top\Delta^{-1}\nu_s(\nu_t^\top\nu_k)(1-\nu_t^\top\nu_k)\\ =&\sum_{k=1}^K \rho_k(u_s^\top U^\top R^{-1}Uu_k)^2(\nu_t^\top\nu_k)(1-\nu_t^\top\nu_k)\\ =&\sum_{k=1}^K \rho_k(e_s^\top R^{-1}e_k)^2(\nu_t^\top\nu_k)(1-\nu_t^\top\nu_k)\\ =&\sum_{k=1}^K \rho_k\delta_{sk}\rho_s^{-2}(\nu_t^\top\nu_k)(1-\nu_t^\top\nu_k)\\ =&\frac{1}{\rho_s}\nu_t^\top\nu_s(1-\nu_t^\top\nu_s)\end{aligned}$$

□

Lemma B.3 (Lemma 4.1)**.** *In the same setting as above, for any i, j, conditioning on $X_i = \nu_{\tau_i}$ and $X_j = \nu_{\tau_j}$:*

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

And for N large enough, conditioning on $X_i = \nu_{\tau_i}$ and $X_j = \nu_{\tau_j}$:

$$\mathbb{E}[(\hat{P}_{ij} - P_{ij})^2] \approx \frac{1/\rho_{\tau_i} + 1/\rho_{\tau_j}}{MN}P_{ij}(1-P_{ij}).$$

Proof. Conditioned on $X_i = \nu_k$, we have by Theorem B.1,

$$\mathbb{E}[W\hat{X}_i] = \nu_k + o(1)$$

and

$$N\cdot\text{Cov}(W\hat{X}_i, W_n\hat{X}_i) = \Sigma(\nu_k)/M.$$

Also, Corollary 3 in Athreya et al. (2016) says $\hat{X}_i$ and $\hat{X}_j$ are asymptotically independent. Thus, conditioning on $X_i = \nu_s$ and $X_j = \nu_t$, we have $\lim_{N\to\infty}\mathbb{E}[\hat{X}_i^\top\hat{X}_j] = \lim_{N\to\infty}\mathbb{E}[(W_N\hat{X}_i)^\top]\mathbb{E}[W_N\hat{X}_j] = \nu_s^\top\nu_t = P_{ij}$.

Since $\hat{P}_{ij} = \hat{X}_i^\top\hat{X}_j$ is a noisy version of the dot product of $\nu_s^\top\nu_t$, combined with Lemma B.2 and the results above, by Equation 5 in Brown and Rutemiller (1977), conditioning on $X_i = \nu_s$ and $X_j = \nu_t$:

$$\mathbb{E}[\hat{X}_i^\top\hat{X}_j] = \mathbb{E}[(W_N\hat{X}_i)^\top]\mathbb{E}[W_N\hat{X}_j] = \nu_s^\top\nu_t + o(1) = P_{ij} + o(1)$$

and

$$\begin{aligned}&N\cdot\text{Var}(\hat{P}_{ij})\\ =&\frac{1}{M}\left(\nu_s^\top\Sigma(\nu_t)\nu_s + \nu_t^\top\Sigma(\nu_s)\nu_t^\top\right) + \frac{1}{M^2N}\left(tr(\Sigma(\nu_s)\Sigma(\nu_t))\right) + o(1)\\ =&\frac{1}{M}\left(\nu_s^\top\Sigma(\nu_t)\nu_s + \nu_t^\top\Sigma(\nu_s)\nu_t^\top\right) + o(1)\\ =&\frac{1/\rho_s + 1/\rho_t}{M}P_{ij}(1-P_{ij}) + o(1).\end{aligned}$$

Since $\hat{P}_{ij} = \hat{X}_i^\top\hat{X}_j$ is asymptotically unbiased for P_{ij}, when n is large enough:

$$\mathbb{E}[(\hat{P}_{ij} - P_{ij})^2] = \text{Var}(\hat{P}_{ij}) \approx \frac{1/\rho_s + 1/\rho_t}{MN}P_{ij}(1-P_{ij}) + o(1).$$

□

The proof for Theorem 4.2 is now a simple application of the above lemmas to the ratio of the mean squared errors for $\bar{A}$ and $\hat{P}$.

APPENDIX C
FLIPPING PROCEDURE FOR PERMUTATION TEST

Here the details of the flipping procedure are described for the permutation test mentioned in Section 4.5. As mentioned before, there are 10 lobes and 70 regions based on the Desikan atlas. We say two regions are adjacent if they share a common boundary. Such spatial adjacency is denoted by an adjacency matrix S for the 70 regions, where $S_{ij} = 1$ means region i and region j contain a pair of voxels, v_i and v_j, which are spatially adjacent. If this is true, then region j is defined as a neighbor of region i. The lobe i.d. for region i is denoted by l_i.

Now a uniform 1-flip can be defined by:

1) Selecting a pair of adjacent regions (region i_1 and region j_1) across the boundary of lobes uniformly, i.e. $S_{i_1 j_1} = 1$ and $l(i_1) \neq l(j_1)$;
2) Uniformly selecting another pair of adjacent regions (region i_2 and region j_2 where $i_1 \neq i_2$ and $j_1 \neq j_2$) across the same boundary of lobes uniformly, i.e. $S_{i_2 j_2} = 1$ and $l(i_1) = l(i_2)$ and $l(j_1) = l(j_2)$;
3) Reassigning region j_1 to lobe l_{i_1} and reassign region i_2 to lobe l_{j_2}.

By this definition, after a uniform 1-flip, the number of regions in each lobe stays the same, where only two regions are changed to a different lobe.

Then we can define a uniform k-flip naturally as sequentially performing uniform 1-flip k times. Note that after a uniform k-flip, the number of regions in each lobe still stays the same.

In the permutation test, a uniform k-flip was applied and the test statistic $T(X, l)$ was calculated based on the lobe assignment after flipping. The p-value is computed as the proportion of uniform k-flips with a T value smaller than the T value for the true lobe assignments.

APPENDIX D
STOCHASTIC BLOCKMODEL PARAMETER SETTING

Here the parameters in the stochastic blockmodel example depicted in Figure 3 are given. It is a 5-block SBM with

$$B = \begin{bmatrix} 0.90 & 0.27 & 0.05 & 0.10 & 0.30 \\ 0.27 & 0.67 & 0.02 & 0.26 & 0.14 \\ 0.05 & 0.02 & 0.44 & 0.25 & 0.33 \\ 0.10 & 0.26 & 0.25 & 0.70 & 0.18 \\ 0.30 & 0.14 & 0.33 & 0.18 & 0.58 \end{bmatrix},$$

$$\rho = \begin{bmatrix} 0.22 & 0.39 & 0.05 & 0.16 & 0.18 \end{bmatrix}.$$

APPENDIX REFERENCES

Robert J Anderson, James J Cook, Natalie A Delpratt, John C Nouls, Bin Gu, James O McNamara, Brian B Avants, G Allan Johnson, and Alexandra Badea. An HPC pipeline with validation framework for small animal multivariate brain analysis (SAMBA). *arxiv 1709.10483*, September 2017.

Avanti Athreya, Carey E Priebe, Minh Tang, Vince Lyzinski, David J Marchette, and Daniel L Sussman. A limit theorem for scaled eigenvectors of random dot product graphs. *Sankhya A*, 78(1):1–18, 2016.

T E J Behrens, H Johansen Berg, S Jbabdi, M F S Rushworth, and M W Woolrich. Probabilistic diffusion tractography with multiple fibre orientations: What can we gain? *NeuroImage*, 34(1):144–155, January 2007.

Gerald G Brown and Herbert C Rutemiller. Means and variances of stochastic vector products with applications to random linear models. *Management Science*, 24(2):210–216, 1977.

Sourav Chatterjee. Matrix estimation by universal singular value thresholding. *The Annals of Statistics*, 43(1):177–214, 2015.

Eleftherios Garyfallidis, Matthew Brett, Bagrat Amirbekian, Ariel Rokem, Stefan van der Walt, Maxime Descoteaux, Ian Nimmo-Smith, and Dipy Contributors. Dipy, a library for the analysis of diffusion MRI data. *Frontiers in neuroinformatics*, 8:8, February 2014.

Krzysztof J Gorgolewski, Natacha Mendes, Domenica Wilfling, Elisabeth Wladimirow, Claudine J Gauthier, Tyler Bonnen, Florence JM Ruby, Robert Trampel, Pierre-Louis Bazin, Roberto Cozatl, et al. A high resolution 7-tesla resting-state fmri test-retest dataset with cognitive and physiological measures. *Scientific data*, 2, 2015.

Gregory Kiar, William Gray Roncal, Disa Mhembere, Eric Bridgeford, Randal Burns, and Joshua T. Vogelstein. ndmg: Neurodata's mri graphs pipeline. http://m2g.io, August 2016.

David Marchette, Carey Priebe, and Glen Coppersmith. Vertex nomination via attributed random dot product graphs. In *Proceedings of the 57th ISI World Statistics Congress*, volume 6, page 16, 2011.

Edward R Scheinerman and Kimberly Tucker. Modeling graphs using dot product representations. *Computational Statistics*, 25(1):1–16, 2010.

Mu Zhu and Ali Ghodsi. Automatic dimensionality selection from the scree plot via the use of profile likelihood. *Computational Statistics & Data Analysis*, 51(2):918–930, 2006.

Xi-Nian Zuo, Jeffrey S Anderson, Pierre Bellec, Rasmus M Birn, Bharat B Biswal, Janusch Blautzik, John CS Breitner, Randy L Buckner, Vince D Calhoun, F Xavier Castellanos, et al. An open science resource for establishing reliability and reproducibility in functional connectomics. *Scientific data*, 1, 2014.