A SEMIPARAMETRIC TWO-SAMPLE HYPOTHESIS TESTING PROBLEM FOR RANDOM DOT PRODUCT GRAPHS

By Minh Tang, Avanti Athreya, Vince Lyzinski
Daniel L. Sussman and Carey E. Priebe

Johns Hopkins University and Harvard University

Two-sample hypothesis testing for random graphs arises naturally in neuroscience, social networks, and machine learning. In this paper, we consider a semiparametric problem of two-sample hypothesis testing for a class of latent position random graphs. We formulate a notion of consistency in this context and propose a valid test for the hypothesis that two finite-dimensional random dot product graphs on a common vertex set have the same generating latent positions or have generating latent positions that are scaled or diagonal transformations of one another. Our test statistic is a function of a spectral decomposition of the adjacency matrix for each graph and our test procedure is consistent across a broad range of alternatives. In our applications to real biological data, we are able to demonstrate that the test has power even for small sample sizes. We conclude by discussing the relationship between our test and classical generalized likelihood ratio tests.

1. Introduction. The development of a comprehensive machinery for two-sample hypothesis testing for random graphs is of both theoretical and practical importance, with applications in neuroscience, social networks, and linguistics, to name but a few. For instance, testing for similarity across brain graphs is an area of active research at the intersection of neuroscience and machine learning, and practitioners often use classical parametric two-sample tests, such as edgewise t-tests on correlations or Mantel tests, or permutation tests on subgraphs, as approaches to graph comparison [6, 19, 20, 30]. Our goal in this work is to provide a clear setting for a particular two-sample graph testing problem and to exhibit a valid, consistent, tractable test statistic. Our results provide, to the best of our knowledge, the first principled approach to semiparametric two-sample hypothesis testing on graphs.

We focus on a test for the hypothesis that two random dot product graphs on the same vertex set, with known vertex correspondence, have the same generating latent position or have generating latent positions that are scaled or diagonal transformations of one another. This framework includes, as a special case, a test for whether two stochastic blockmodels have the same or related block probability matrices. We use a spectral decomposition of the adjacency matrix to estimate the parameters for each random dot product graph, and our test statistic is a function of an appropriate distance between these estimates.

*Supported in part by National Security Science and Engineering Faculty Fellowship (NSSEFF), Johns Hopkins University Human Language Technology Center of Excellence (JHU HLT COE), and the XDATA program of the Defense Advanced Research Projects Agency (DARPA) administered through Air Force Research Laboratory contract FA8750-12-2-0303.

AMS 2000 subject classifications: Primary 62G10; secondary 62H12,05C80

Keywords and phrases: semiparametric graph inference, two-sample hypothesis testing, consistency, random dot product graph

In the two-sample graph testing problem we address, the parameter dimension grows as the sample size grows. This problem is not precisely analogous to classical two-sample tests for, say, the difference of two parameters belonging to some fixed Euclidean space, in which an increase in data has no effect on the dimension of the parameter. The problem is also not nonparametric, since we view our latent positions as fixed and impose specific distributional requirements on the data—that is, on the adjacency matrices. Indeed, we regard the problem as semiparametric, and we adapt the traditional definition of consistency to this setting. In particular, we have power increasing to one for alternatives in which the difference between the two latent positions grows with the sample size.

As one example of the utility of the test procedures we describe, we consider the problem of matching connectome data from *Caeronabdhitis elegans (C.elegans)*, a hermaphrodite worm whose wiring diagrams have been widely studied [12, 27, 28]. There are a total of 302 neurons in the *C. elegans* brain and there are two different—but related—neuronal networks, characterized by the chemical wiring (chemical synapses) and electrical wiring (gap junctions), with known vertex alignment between the networks. It is of biological relevance to determine the extent to which the two wiring diagrams are similar. This question can be framed in the context of two-sample testing, and we provide one approach to its resolution.

C. elegans is an instance of a pair of graphs with a comparatively small but aligned vertex set, and our numerical results on this specific data indicate that our test procedure provides good power, despite a sample size in the hundreds. Our numerical analysis on other simulated data affirms more broadly that our test has power against a wide class of alternatives for moderate sample sizes. The analysis of much larger data is also a pressing practical problem, and connectome data representing pairs of graphs with known vertex alignment can be on the order of 10^7 vertices and 10^{10} edges; see, for instance [11]. The existence of such large data sets indicates that there are practical problems in which our theoretical guarantees apply.

While it may appear that the requirement of known vertex correspondence between the graphs is a stringent one, the *C. elegans* and connectome data are but two examples of a diverse class of such paired graphs for which subsequent inference is key. Other examples include the comparison of graphs in a time series, such as email correspondence among a group over time, the comparison of document networks in multiple languages, or the comparison of user behavior on different social media platforms.

We conclude the paper with a brief discussion of the applicability of other test statistics, including intuitively appealing tests based on the spectral or Frobenius norm of the difference of adjacency matrices, and a discussion of the connection between our test and classical generalized likelihood ratio tests. Our test statistic is a ratio whose numerator is a distance between the estimated and true latent positions and whose denominator is related to the estimated standard error. As such, it is in the spirit of a Wald test. Although we endeavor to describe the strengths and weakness of several different test statistics, our aim is not to provide a comprehensive analysis of possible tests, but rather to formulate a two-sample test for our random graph model and to prove basic results about our particular choice of test statistic. The specific hypotheses we consider are indicative of the multitude of questions that can arise in the larger context of two-sample hypothesis testing on random graphs.

1.1. *Related Work.* Hypothesis testing on a single graph has a long history, especially when compared to the multiple-graph setting. Problems of clustering and community detection for a graph can be framed as classical parametric hypothesis tests. To touch on several recent results, we note that in [1], the authors translate the problem of community detection into a test for determining whether a graph is Erdös-Renyi or whether it has an unusually dense subgraph. In [22], the authors provide a power analysis of the maximum degree and size invariants for a similar problem and in [4] the authors formulate the problem of determining the number of communities in a network as a hypothesis testing problem involving the number of blocks in a stochastic blockmodel. In contrast, we consider a two-sample problem in a more general setting.

The random dot product graph model generalizes both the stochastic blockmodel (SBM) and degree-corrected SBM. Our results do not directly apply to general latent position models, such as those considered in [13]. Nevertheless, to the extent that latent positions can be estimated accurately in these alternative models—itself a topic of current investigation—a distance between estimated latent positions for two graphs on the same vertex set could be used to derive appropriate hypothesis testing procedures. If the two graphs are not on the same vertex set, or if the vertex correspondence is unknown, other issues arise. Finding the vertex correspondence when one exists, but is unknown, is the problem of "graph matching" and is notoriously difficult [8]. It is possible that graph matching tools can be used as a first step to align the graphs before employing our test, but we do not consider this here. However, an alternate approach to comparing graphs on potentially different vertex sets and with differing numbers of vertices is the subject of our related paper [25]. There, we view the latent positions for the random dot product graph as being i.i.d from some pair of underlying distributions, say F and G, and the graphs comparison translates to the nonparametric test of equality of F and G.

Finally, for the two-sample hypothesis test we consider, one can also construct test statistics using other embedding methods, such as spectral decompositions of normalized Laplacian matrices. To prove the requisite bounds for these Laplacian-based test statistics, however, requires substantial technical machinery, and non-trivial adaptation or generalization of the results in [7, 18, 21], among others. Hence, for simplicity, we focus here on embeddings of the adjacency matrix.

2. Setting. We focus here on two-sample hypothesis testing for the latent position vectors of a pair of *random dot product graphs* (RDPG) [29] on the same vertex set with a known vertex correspondence, i.e., a bijective map φ from the vertex set of one graph to the vertex set of the other graph. We shall assume, without loss of generality, that φ is the identity map. As we have already remarked, the assumption of known vertex correspondence is satisfied in a number of real-world problems. Random dot product graphs are a specific example of *latent position random graphs* [13], in which each vertex is associated with a latent position and, conditioned on the latent positions, the presence or absence of all edges in the graph are independent. The edge presence probability is given by a *link* function, which is a symmetric function of the two latent positions. The link function in a random dot product graph is simply the dot product. Statistical analysis for random graphs has received much recent interest: see [9, 10] for reviews of the pertinent literature.

2.1. *Random Dot Product Graphs.* We begin with a number of necessary definitions and notational conventions. First, we define a random dot product graph on $\mathbb{R}^d$ as follows.

DEFINITION 2.1 (Random Dot Product Graph (RDPG)). Let χ_d^n be defined by

$$\chi_d^n = \{\mathbf{U} \in \mathbb{R}^{n\times d} : \mathbf{U}\mathbf{U}^T \in [0,1]^{n\times n} \text{ and } \operatorname{rank}(\mathbf{U}) = d\}$$

and let $\mathbf{X} = [X_1^\top, \dots, X_n^\top]^T \in \chi_d^n$. Suppose $\mathbf{A}$ is a random adjacency matrix given by

$$\mathbb{P}[\mathbf{A}|\mathbf{X}] = \prod_{i<j} (X_i^T X_j)^{\mathbf{A}_{ij}} (1 - X_i^T X_j)^{1-\mathbf{A}_{ij}}$$

Then we say that $\mathbf{A}$ is the adjacency matrix of a *random dot product graph* with *latent position* $\mathbf{X}$ *of rank* d.

We define the matrix $\mathbf{P} = (p_{ij})$ of edge probabilities by $\mathbf{P} = \mathbf{X}\mathbf{X}^T$. We will also write $\mathbf{A} \sim \text{Bernoulli}(\mathbf{P})$ to represent that the existence of an edge between any two vertices i, j, where $i > j$, is a Bernoulli random variable with probability p_{ij}; edges are independent. We emphasize that the random graphs we consider are undirected and loop-free.

Suppose we are given two adjacency matrices $\mathbf{A}_1$ and $\mathbf{A}_2$ for a pair of random dot product graphs on the same vertex set. Our goal is to develop a consistent, at most level-α test to determine whether or not the two generating latent positions are equal, up to an orthogonal transformation. Indeed, if $\mathcal{O}(d)$ represents the collection of orthogonal matrices in $\mathbb{R}^{d \times d}$ and if $\mathbf{W} \in \mathcal{O}(d)$,then $\mathbf{X}\mathbf{W}\mathbf{W}^T\mathbf{X}^T = \mathbf{P}$, leading to obvious non-identifiability.

2.2. *Hypothesis Testing.* Formally, we state the following two-sample testing problems for random dot product graphs. Let $\mathbf{X}_n, \mathbf{Y}_n \in \chi_d^n$ and define $\mathbf{P}_n = \mathbf{X}_n\mathbf{X}_n^T$ and $\mathbf{Q}_n = \mathbf{Y}_n\mathbf{Y}_n^T$. Given $\mathbf{A} \sim \text{Bernoulli}(\mathbf{P}_n)$ and $\mathbf{B} \sim \text{Bernoulli}(\mathbf{Q}_n)$, we consider the following tests:

(a) *(Equality, up to an orthogonal transformation)*

$$\begin{aligned} & H_0^n\colon \mathbf{X}_n \perp \mathbf{Y}_n \\ \text{against} \quad & H_a^n\colon \mathbf{X}_n \not\perp \mathbf{Y}_n \end{aligned}$$

where $\perp$ denotes that there exists an orthogonal matrix $\mathbf{W} \in \mathbb{R}^{d \times d}$ such that $\mathbf{X}_n = \mathbf{Y}_n\mathbf{W}$.

(b) *(Scaling)*

$$\begin{aligned} & H_0^n\colon \mathbf{X}_n \perp c_n\mathbf{Y}_n \text{ for some constant } c_n > 0 \\ \text{against} \quad & H_a^n\colon \mathbf{X}_n \not\perp c_n\mathbf{Y}_n \text{ for any constant } c_n > 0 \end{aligned}$$

(c) *(Diagonal Transformation)*

$$\begin{aligned} & H_0^n\colon \mathbf{X}_n \perp \mathbf{D}_n\mathbf{Y}_n \text{ for some diagonal } \mathbf{D}_n \text{ with bounded positive entries} \\ \text{against} \quad & H_a^n\colon \mathbf{X}_n \not\perp \mathbf{D}_n\mathbf{Y}_n \text{ for any diagonal } \mathbf{D}_n \text{ with bounded positive entries.} \end{aligned}$$

In fact, throughout this paper, we will consider a sequence of such tests for $n \in \mathbb{N}$. We stress that in our sequential formulation of (a) – (c), the latent positions $\mathbf{X}_n, \mathbf{Y}_n$ need not be related to $\mathbf{X}_{n'}, \mathbf{Y}_{n'}$ for any $n' \neq n$. However, the size of the adjacency matrices $\mathbf{A}$ and $\mathbf{B}$ is quadratic in n and hence the larger n is, the more accurate are our estimates of $\mathbf{X}_n$ and $\mathbf{Y}_n$.

To contextualize our choice of hypotheses, consider the specific case of the stochastic blockmodel [14] and the related degree-corrected stochastic blockmodel [15]. Recall that a stochastic block model on K blocks with block probability matrix $\mathbf{N}$ can be viewed as a random dot product graph whose latent positions are a mixture of K fixed vectors. In (a), we test whether two stochastic blockmodel graphs G_1 and G_2 with fixed block assignments have the same block probability matrices $\mathbf{N}_1 = \mathbf{N}_2$. In (b), we test whether the block probability matrix of one graph is a scalar multiple of the other; i.e. if $\mathbf{N}_1 = c\mathbf{N}_2$. Finally, in (c), we test whether two degree-corrected stochastic blockmodels have the same block probability matrices, but possibly different degree-correction factors.

We describe the test procedures for the above hypothesis tests in more details in the next section. The main idea is that given suitable estimates $\hat{\mathbf{X}}_n$ and $\hat{\mathbf{Y}}_n$ of $\mathbf{X}$ and $\mathbf{Y}$, the associated test statistic is essentially a function of $\min_{\mathbf{W} \in \mathcal{O}_d} \|\hat{\mathbf{X}}_n - \hat{\mathbf{Y}}_n\mathbf{W}\|$.

2.3. *Adjacency spectral embedding and related results.* We now describe the adjacency spectral embedding of [23], which serves as our estimate for the latent positions $\mathbf{X}$ and $\mathbf{Y}$.

DEFINITION 2.2. The *adjacency spectral embedding* of $\mathbf{A}$ into $\mathbb{R}^d$ is given by $\hat{\mathbf{X}} = \mathbf{U_A S_A^{1/2}}$ where

$$|\mathbf{A}| = [\mathbf{U_A}|\widetilde{\mathbf{U}}_\mathbf{A}][\mathbf{S_A} \bigoplus \widetilde{\mathbf{S}}_\mathbf{A}][\mathbf{U_A}|\widetilde{\mathbf{U}}_\mathbf{A}]$$

is the spectral decomposition of $|\mathbf{A}| = (\mathbf{A^T A})^{\mathbf{1/2}}$ and $\mathbf{S_A}$ is the matrix of the d largest eigenvalues of $|\mathbf{A}|$ and $\mathbf{U_A}$ is the matrix whose columns are the corresponding eigenvectors.

Let $\mathbf{X}$ and $\mathbf{Y}$ be two latent positions in $\mathbb{R}^{n\times d}$, and let $\mathbf{A} \sim \text{Bernoulli}(\mathbf{P})$ with $\mathbf{P} = \mathbf{X}\mathbf{X}^T$ and $\mathbf{B} \sim \text{Bernoulli}(\mathbf{Q})$ with $\mathbf{Q} = \mathbf{Y}\mathbf{Y}^T$ represent the associated adjacency matrices of the random dot product graphs with $\mathbf{X}$ and $\mathbf{Y}$, respectively, as their latent positions. We observe that $\mathbf{X}$, $\mathbf{Y}$, and $\mathbf{A}$ and $\mathbf{B}$ all depend on n, but for notational convenience we will suppress this dependence except when imperative for communicating an asymptotic property. Let $\hat{\mathbf{X}}$ and $\hat{\mathbf{Y}}$ denote the corresponding adjacency spectral embeddings of $\mathbf{A}$ and $\mathbf{B}$, respectively. We use $\|\cdot\|_F$ to denote the Frobenius norm of a matrix; $\|\cdot\|$ to denote the spectral norm of a matrix or the Euclidean norm of a vector, depending on the context, and $\|\cdot\|_{2\to\infty}$ to denote the $2\to\infty$ operator norm (that is, the maximum Euclidean norm of the rows of a matrix). Also, we define, for a matrix $\mathbf{M}$ of edge probabilities, the parameters $\delta(\mathbf{M})$, $\gamma_1(\mathbf{M})$, and $\gamma_2(\mathbf{M})$ as follows

$$\delta(\mathbf{M}) = \max_{1\le i\le n} \sum_{j=1}^n M_{ij}$$
$$\gamma_1(\mathbf{M}) = \min_{i,j\in[d+1], i\neq j} \frac{|\mathbf{S_M}(i,i) - \mathbf{S_M}(j,j)|}{\delta(\mathbf{M})}$$
$$\gamma_2(\mathbf{M}) = \min_{1\le i\le d, d+1\le j\le n} \frac{|\mathbf{S_M}(i,i) - \mathbf{S_M}(j,j)|}{\delta(\mathbf{M})}$$

We remark that $\delta(\mathbf{P})$ is simply the maximum expected degree of a graph $\mathbf{A}\sim\text{Bernoulli}(\mathbf{P})$, $\gamma_1(\mathbf{P})$ is the minimum gap between distinct eigenvalues of $\mathbf{P}$, normalized by the maximum expected degree, and lastly, because $\mathbf{P}$ is of rank d, $\gamma_2(\mathbf{P})$ is just $\mathbf{S_P}(d,d)/\delta(\mathbf{P})$. It is immediate that $\gamma_1 \le \gamma_2$.

Throughout this work, our results depend on certain growth conditions, which we state in Assumption 1 below, on the gap between the eigenvalues of $\mathbf{P}_n$ and certain minimum sparsity conditions on $\mathbf{P}_n$ as n increases. These conditions are motivated by established bounds from [2, 3, 17, 24, 26] on the separation between $\mathbf{A}$ and $\mathbf{P}$ and the accuracy of the adjacency spectral embedding in the estimation of the true latent positions. We consolidate these known bounds in the appendix, but in particular they imply

$$\min_{\mathbf{W}\in\mathcal{O}(d)} \|\hat{\mathbf{X}}_n\mathbf{W} - \mathbf{X}_n\|_F = O(d\sqrt{\log n}) \tag{2.1}$$

with high probability.

ASSUMPTION 1. *We assume that there exists $d\in\mathbb{N}$ such that for all n, $\mathbf{P}_n$ is of rank d with d distinct positive eigenvalues. Further, we assume that there exist constants $\epsilon > 0$, $c_0 > 0$ and $n_0(\epsilon, c)\in\mathbb{N}$ such that for all $n\ge n_0$:*

$$\gamma_1(\mathbf{P}_n) > c_0 \tag{2.2}$$
$$\delta(\mathbf{P}_n) > (\log n)^{2+\epsilon} \tag{2.3}$$

Because the parameters $\delta(\mathbf{P})$, $\gamma_1(\mathbf{P})$ and $\gamma_2(\mathbf{P})$ depend on $\mathbf{P}$, they cannot be computed from the adjacency matrices alone. Therefore, we use the corresponding estimates of these quantities, namely $\delta(\mathbf{A})$, $\gamma_1(\mathbf{A})$, and $\gamma_2(\mathbf{A})$. Proposition A.2 of the appendix guarantees the consistency of these estimates, and they also provide a mechanism by which to check whether the conditions in Assumption 1 hold.

We note that a level-α test can easily be generated from Eq. (2.1) itself. However, in the present work, we provide an improved bound for the Frobenius norm of $\hat{\mathbf{X}} - \mathbf{X}\mathbf{W}$ for some orthogonal $\mathbf{W}$. The bound enables us to describe more precisely the class of alternatives over which our test is consistent. Eq. (2.1) requires that, for consistency, the difference between the latent positions $\mathbf{X}_n$ and $\mathbf{Y}_n$ diverge at a particular rate, but we can now expand the class of alternatives to be simply that for which the difference between the latent positions diverges, with no restriction on the rate.

THEOREM 2.3. *Suppose $\mathbf{P} = \mathbf{X}\mathbf{X}^T$ is an $n \times n$ probability matrix of rank d and its eigenvalues are distinct. Suppose also that there exists $\epsilon > 0$ such that $\delta(\mathbf{P}) > (\log n)^{2+\epsilon}$. Let $c > 0$ be arbitrary but fixed. Then there exists a $n_0(c)$ and a universal constant $C \geq 0$ such that if $n \geq n_0$ and $n^{-c} < \eta < 1/2$, then there exists a deterministic $\mathbf{W} \in \mathcal{O}(d)$ such that, with probability at least $1 - 3\eta$,*

$$\Big| \|\hat{\mathbf{X}} - \mathbf{X}\mathbf{W}\|_F - C(\mathbf{X}) \Big| \leq \frac{C d^{3/2} \log(n/\eta)}{C(\mathbf{X}) \sqrt{\gamma_1^7(\mathbf{P}) \delta(\mathbf{P})}} \tag{2.4}$$

where $C(\mathbf{X})$ is a constant depending only on $\mathbf{X}$ and is bounded from above by $\sqrt{d\gamma_2^{-1}(\mathbf{P})}$.

REMARK. We observe that $C(\mathbf{X})$ is a function of the unknown latent position $\mathbf{X}$; the exact expression is given in Lemma A.6. We shall only require an upper bound on $C(\mathbf{X})$ for our subsequent results. Nevertheless, $C(\mathbf{X})$ can be consistently estimated, as $n \to \infty$, by a function $C(\hat{\mathbf{X}})$ of the $\hat{\mathbf{X}}$ as follows

$$C(\hat{\mathbf{X}}) = \sqrt{\operatorname{tr} \mathbf{S}_{\mathbf{A}}^{-1} \hat{\mathbf{X}}^T \hat{\mathbf{D}} \hat{\mathbf{X}} \mathbf{S}_{\mathbf{A}}^{-1}}$$

where $\hat{\mathbf{D}}$ is the diagonal matrix whose diagonal entries are given by

$$\hat{\mathbf{D}}_{ii} = \sum_{j \neq i} \langle \hat{X}_i, \hat{X}_j \rangle (1 - \langle \hat{X}_i, \hat{X}_j \rangle)$$

Using $C(\hat{\mathbf{X}})$ in place of the upper bound for $C(\mathbf{X})$ in Theorem 2.3 may yield, in practice, better performance for the test procedures of Section 3.

As a corollary of Theorem 2.3, we obtain the following.

COROLLARY 2.4. *Let $\{\mathbf{X}_n\}$ be a sequence of latent positions and suppose that the sequence of matrices $\{\mathbf{P}_n\}$ where $\mathbf{P}_n = \mathbf{X}_n \mathbf{X}_n^T$ satisfies Assumption 1. Then there exists a deterministic sequence of orthogonal matrices $\mathbf{W}_n$ such that*

$$\|\hat{\mathbf{X}}_n - \mathbf{X}_n \mathbf{W}_n\|_F - C(\mathbf{X}_n) \xrightarrow{\text{a.s.}} 0$$

We define our test statistic in the next section. However, we reiterate that based on Theorem 2.3, as n grows, the test statistic we construct will not always distinguish between two latent positions $\mathbf{X}_n$ and $\mathbf{Y}_n$ that differ in a constant number of rows, and one of our contributions is to widen the class of alternatives over which consistency can be achieved. First, we adapt the classical notion of consistency to our sequence of hypothesis tests. We state this definition for the case of testing whether the latent positions are equal (up to rotation); its adaptation for the scaling and diagonal tests is clear.

DEFINITION 2.5. Let $\mathbf{X}_n$, $\mathbf{Y}_n$ in $\mathbb{R}^{n\times d}$ and $d_n \in \mathbb{R}$, $n \in \mathbb{N}$, be given. A test statistic T_n and associated rejection region R_n to test the null hypothesis

$$\begin{aligned} H_0^n &: \mathbf{X}_n \perp \mathbf{Y}_n \\ \text{against } H_a^n &: \mathbf{X}_n \not\perp \mathbf{Y}_n \end{aligned}$$

is a *consistent, asymptotically level α test* if for any $\eta > 0$, there exists N_1 such that for all $n > N_1$,

(i) If H_a^n is true, $P(T_n \in R_n) > 1 - \eta$

(ii) If H_0^n is true, $P(T_n \in R_n) \leq \alpha + \eta$

3. Main Results.

3.1. *Equality case.* The first result is concerned with finite sample and asymptotic properties of a test for the null hypothesis $H_0 : \mathbf{X}_n \perp \mathbf{Y}_n$ against the alternative $H_a : \mathbf{X}_n \not\perp \mathbf{Y}_n$, for both the finite sample case of a fixed pair of latent positions $\mathbf{X}_n$ and $\mathbf{Y}_n$ and the asymptotic case of a sequence of latent positions $\{\mathbf{X}_n, \mathbf{Y}_n\}$, $n \in \mathbb{N}$. Our test statistic T_n is a scaled version of

$$\min_{\mathbf{W}\in\mathcal{O}(d)} \|\hat{\mathbf{X}}_n\mathbf{W} - \hat{\mathbf{Y}}_n\|_F.$$

If it is sufficiently small, we do not reject; and if it is larger than a constant for which we provide an upper bound, then we reject.

THEOREM 3.1. *For each fixed n, consider the hypothesis test*

$$H_0^n : \mathbf{X}_n \perp \mathbf{Y}_n \quad \textit{versus} \quad H_a^n : \mathbf{X}_n \not\perp \mathbf{Y}_n$$

where $\mathbf{X}_n$ and $\mathbf{Y}_n \in \mathbb{R}^{n\times d}$ are matrices of latent positions for two random dot product graphs. Let $\hat{\mathbf{X}}_n$ and $\hat{\mathbf{Y}}_n$ be the adjacency spectral embeddings of $\mathbf{A}_n \sim \mathrm{Bernoulli}(\mathbf{P}_n = \mathbf{X}_n\mathbf{X}_n^T)$ and $\mathbf{B}_n \sim \mathrm{Bernoulli}(\mathbf{Q}_n = \mathbf{Y}_n\mathbf{Y}_n^T)$, respectively. Define the test statistic T_n as follows:

$$T_n = \frac{\min_{\mathbf{W}\in\mathcal{O}(d)} \|\hat{\mathbf{X}}_n\mathbf{W} - \hat{\mathbf{Y}}_n\|_F}{\sqrt{d\gamma_2^{-1}(\mathbf{A}_n)} + \sqrt{d\gamma_2^{-1}(\mathbf{B}_n)}}. \tag{3.1}$$

Let $\alpha \in (0,1)$ be given. Then for all $C > 1$, if the rejection region is defined by

$$R := \{t \in \mathbb{R} : t \geq C\},$$

then there exists an $n_1 = n_1(\alpha, C) \in \mathbb{N}$ such that for all $n \geq n_1$, the test procedure with T_n and rejection region R is an at most level α test, i.e., for all $n \geq n_1$, if $\mathbf{X}_n \perp \mathbf{Y}_n$, then

$$\mathbb{P}(T_n \in R) \leq \alpha.$$

Furthermore, consider the sequence of latent positions satisfying Assumption 1 and for which

$$\min_{\mathbf{W}\in\mathcal{O}(d)} \|\mathbf{X}_n\mathbf{W} - \mathbf{Y}_n\| = d_n$$

Suppose $d_n \neq 0$ for infinitely many n. Let $t_1 = \min\{k > 0 : d_k > 0\}$ and sequentially define $t_n = \min\{k > t_{n-1} : d_k > 0\}$. Let $b_n = d_{t_n}$. If $\liminf b_n = \infty$, then this test procedure is consistent in the sense of Definition 2.5 over this sequence of latent positions.

REMARK. This result and its analogues for the scaling and diagonal hypotheses do not require that $\mathbf{A}_n$ and $\mathbf{B}_n$ be independent for any fixed n, nor that the sequence of pairs $(\mathbf{A}_n, \mathbf{B}_n)$, $n \in \mathbb{N}$, be independent. In addition, the requirement that $\liminf b_k = \infty$ can be weakened somewhat. Specifically, there exist constants $C(\mathbf{X}_n)$ and $C(\mathbf{Y}_n)$ such that consistency is achieved as long as

$$\|\mathbf{X}_n\mathbf{W} - \mathbf{Y}_n\|_F \geq C(\mathbf{X}_n) + C(\mathbf{Y}_n) + \epsilon$$

where $\epsilon > 0$ is arbitrary but fixed.

PROOF. For ease of notation, in parts of the proof below we will suppress the dependence of $\mathbf{X}_n$, $\mathbf{Y}_n$, $\mathbf{P}_n$ and $\mathbf{Q}_n$ on n and simply denote these matrices by $\mathbf{X}$, $\mathbf{Y}$, $\mathbf{P}$, and $\mathbf{Q}$, respectively; we will make this dependence explicit when necessary. Suppose that the null hypothesis H_0 is true, so there exists an orthogonal $\widetilde{\mathbf{W}} \in \mathbb{R}^{d\times d}$ such that $\mathbf{X} = \mathbf{Y}\widetilde{\mathbf{W}}$. Let α be given, and let $\eta < \alpha/4$. From (2.4), for all n sufficiently large, there exist orthogonal matrices $\mathbf{W}_X$ and $\mathbf{W}_Y \in \mathcal{O}(d)$ such that with probability at least $1 - \eta$,

$$\begin{aligned}\|\hat{\mathbf{X}} - \mathbf{X}\mathbf{W}_X\|_F &\leq C(\mathbf{X}) + f(\mathbf{X}, \alpha, n)\\ \|\hat{\mathbf{Y}} - \mathbf{Y}\mathbf{W}_Y\|_F &\leq C(\mathbf{Y}) + f(\mathbf{Y}, \alpha, n)\end{aligned}$$

where $f(\mathbf{X}, \alpha, n) \to 0$ as $n \to \infty$ for a fixed $\mathbf{X}$ and α.

Let $\mathbf{W}^* = \mathbf{W}_Y\widetilde{\mathbf{W}}\mathbf{W}_X$. Then there exists a $n_0 = n_0(\alpha)$ such that for all $n > n_0$, with probability at least $1 - \eta$, we have

$$\begin{aligned}\|\hat{\mathbf{X}} - \hat{\mathbf{Y}}\mathbf{W}^*\|_F &\leq \|\hat{\mathbf{X}} - \mathbf{X}\mathbf{W}_X\|_F + \|\mathbf{X}\mathbf{W}_X - \mathbf{Y}\widetilde{\mathbf{W}}\mathbf{W}_X\|_F + \|\mathbf{Y}\widetilde{\mathbf{W}}\mathbf{W}_X - \hat{\mathbf{Y}}\mathbf{W}^*\|_F\\ &\leq \|\hat{\mathbf{X}} - \mathbf{X}\mathbf{W}_X\|_F + \|\mathbf{X}\mathbf{W}_X - \mathbf{Y}\widetilde{\mathbf{W}}\mathbf{W}_X\|_F + \|\left(\mathbf{Y} - \hat{\mathbf{Y}}\mathbf{W}_Y\right)\left(\widetilde{\mathbf{W}}\mathbf{W}_X\right)\|_F\\ &\leq \|\hat{\mathbf{X}} - \mathbf{X}\mathbf{W}_X\|_F + \|\mathbf{Y} - \hat{\mathbf{Y}}\mathbf{W}_Y\|_F\\ &\leq C(\mathbf{X}) + C(\mathbf{Y}) + f(\mathbf{X}, \alpha, n) + f(\mathbf{Y}, \alpha, n)\end{aligned}$$

where we have used the fact that under H_0, $\mathbf{X} = \mathbf{Y}\widetilde{\mathbf{W}}$; the final step follows from this and the unitary invariance of the Frobenius norm. We note that both $C(\mathbf{X})$ and $C(\mathbf{Y})$ are unknown. However, by Theorem 2.3, they can be bounded from above by $(d\gamma_2^{-1}(\mathbf{P}))^{1/2}$ and $(d\gamma_2^{-1}(\mathbf{Q}))^{1/2}$, respectively. Hence for all $n > n_0$, with probability at least $1 - \alpha$, we have

$$\frac{\min_{\mathbf{W}\in\mathcal{O}(d)} \|\hat{\mathbf{X}}_n\mathbf{W} - \hat{\mathbf{Y}}_n\|_F}{\sqrt{d\gamma_2^{-1}(\mathbf{P}_n)} + \sqrt{d\gamma_2^{-1}(\mathbf{Q}_n)}} \leq 1 + r(\alpha, n)$$

where $r(\alpha, n) \to 0$ as $n \to \infty$ for a fixed α. In addition, by Proposition A.2, the terms $\gamma_2^{-1}(\mathbf{P}_n)$ and $\gamma_2^{-1}(\mathbf{Q}_n)$ in the denominator can be replaced by $\gamma_2^{-1}(\mathbf{A}_n)$ and $\gamma_2^{-1}(\mathbf{B}_n)$ for sufficiently large n. Therefore, with probability at least $1 - \alpha$,

$$T_n = \frac{\min_{\mathbf{W}\in\mathcal{O}(d)} \|\hat{\mathbf{X}}_n\mathbf{W} - \hat{\mathbf{Y}}_n\|_F}{\sqrt{d\gamma_2^{-1}(\mathbf{A}_n)} + \sqrt{d\gamma_2^{-1}(\mathbf{B}_n)}} \leq 1 + \widetilde{r}(\alpha, n)$$

where once again, for a fixed α, $\widetilde{r}(\alpha, n) \to 0$ as $n \to \infty$. We can thus take $n_1 = n_1(\alpha, C) = \inf\{n \geq n_0(\alpha): \widetilde{r}(\alpha, n) \leq C - 1\} < \infty$. Then for all $n > n_1$ and $\mathbf{X}_n, \mathbf{Y}_n$ satisfying $\mathbf{X}_n \perp \mathbf{Y}_n$, we conclude

$$\mathbb{P}(T_n \in R) < \alpha.$$

We now prove consistency. Let $\widetilde{\mathbf{W}} = \operatorname{argmin}_{\mathbf{W} \in \mathcal{O}(d)} \|\mathbf{X} - \mathbf{Y}\mathbf{W}\|_F$ and denote by $D(\mathbf{X}, \mathbf{Y}) = \|\mathbf{X} - \mathbf{Y}\widetilde{\mathbf{W}}\|_F$. As before, let $\mathbf{W}^* = \mathbf{W}_Y \widetilde{\mathbf{W}} \mathbf{W}_X$. Note that

$$\begin{aligned} \|\hat{\mathbf{X}} - \hat{\mathbf{Y}}\mathbf{W}^*\|_F &\geq \|\mathbf{Y}\widetilde{\mathbf{W}} - \mathbf{X}\|_F - \|\hat{\mathbf{X}} - \mathbf{X}\mathbf{W}_X + \mathbf{Y}\mathbf{W}_Y\mathbf{W} - \hat{\mathbf{Y}}\mathbf{W}\|_F \\ &\geq D(\mathbf{X}, \mathbf{Y}) - \|\hat{\mathbf{X}} - \mathbf{X}\mathbf{W}_X\|_F - \|\mathbf{Y}\mathbf{W}_Y - \hat{\mathbf{Y}}\|_F \end{aligned} \tag{3.2}$$

Therefore, for all n,

$$\begin{aligned} \mathbb{P}(T_n \notin R) &\leq \mathbb{P}\left(\frac{\|\hat{\mathbf{X}} - \hat{\mathbf{Y}}\mathbf{W}^*\|_F}{\sqrt{d\gamma_2^{-1}(\mathbf{A})} + \sqrt{d\gamma_2^{-1}(\mathbf{B})}} \leq C \right) \\ &\leq \mathbb{P}\left(\frac{\|\mathbf{Y}\widetilde{\mathbf{W}} - \mathbf{X}\|_F - \|\hat{\mathbf{X}} - \mathbf{X}\mathbf{W}_X\|_F - \|\mathbf{Y}\mathbf{W}_Y - \hat{\mathbf{Y}}\|_F}{\sqrt{d\gamma_2^{-1}(\mathbf{A})} + \sqrt{d\gamma_2^{-1}(\mathbf{B})}} \leq C \right) \\ &= \mathbb{P}\left(\|\hat{\mathbf{X}} - \mathbf{X}\mathbf{W}_X\|_F + \|\mathbf{Y}\mathbf{W}_Y - \hat{\mathbf{Y}}\|_F + C' \geq D(\mathbf{X}, \mathbf{Y}) \right) \end{aligned}$$

where $C' = C(\sqrt{d\gamma_2^{-1}(\mathbf{A})} + \sqrt{d\gamma_2^{-1}(\mathbf{B})})$. By Assumption 1, there exists some n_0 and some $c_0 > 0$ such that $\gamma_2(\mathbf{P}_n) \geq c_0$ and $\gamma_2(\mathbf{Q}_n) \geq c_0$ for all $n \geq n_0$. Now, let $\beta > 0$ be given. By the almost sure convergence of $\|\hat{\mathbf{X}} - \mathbf{X}\mathbf{W}_x\|_F$ to $C(\mathbf{X})$, established in Theorem 2.3, and the almost sure convergence of $\gamma_2(\mathbf{A})$ to $\gamma_2(\mathbf{P})$ given in A.2, we deduce that there exists a constant $M_1(\beta)$ and a positive integer $n_0 = n_0(\alpha, \beta)$ so that, for all $n \geq n_0(\alpha, \beta)$,

$$\mathbb{P}(\|\hat{\mathbf{X}} - \mathbf{X}\mathbf{W}_X\|_F + C\sqrt{d\gamma_2^{-1}(\mathbf{A})} \geq M_1/2) \leq \beta/2$$
$$\mathbb{P}(\|\hat{\mathbf{Y}} - \mathbf{Y}\mathbf{W}_Y\|_F + C\sqrt{d\gamma_2^{-1}(\mathbf{B})} \geq M_1/2) \leq \beta/2$$

If $b_n \to \infty$, there exists some $n_2 = n_2(\alpha, \beta, C)$ such that, for all $n \geq n_2$, either $D(\mathbf{X}_n, \mathbf{Y}_n) = 0$ or $D(\mathbf{X}_n, \mathbf{Y}_n) \geq M_1$. Hence, for all $n \geq n_2$, if $D(\mathbf{X}_n, \mathbf{Y}_n) \neq 0$, then $\mathbb{P}(T_n \notin R) \leq \beta$, i.e., our test statistic T_n lies within the rejection region R with probability at least $1 - \beta$, as required. $\square$

3.2. *Scaling Case.* We now consider the case of testing whether the latent positions are related by scaling or diagonal transformation. We state the appropriate theorems below and refer the reader to the appendix for their proofs.

For the scaling case, let $\mathcal{C} = \mathcal{C}(\mathbf{Y}_n)$ denote the class of all positive constants c for which all the entries of $c^2\mathbf{Y}_n\mathbf{Y}_n^T$ belong to the unit interval. We wish to test the null hypothesis $H_0: \mathbf{X}_n \perp c_n\mathbf{Y}_n$ for some $c_n \in \mathcal{C}$ against the alternative $H_a: \mathbf{X}_n \not\perp c_n\mathbf{Y}_n$ for any $c_n \in \mathcal{C}$. In what follows below, we will only write $c_n > 0$, but will always assume that $c_n \in \mathcal{C}$, since the problem is ill-posed otherwise. The test statistic T_n is now a simple modification of the one used in Theorem 3.1: for this test, we compute a Procrustes distance between scaled adjacency spectral embeddings for the two graphs. Specifically, we have

Theorem 3.2. *For each fixed n, consider the hypothesis test*

$$H_0^n: \mathbf{X}_n \perp c_n\mathbf{Y}_n \quad \textit{for some } c_n > 0 \textit{ versus}$$
$$H_a^n: \mathbf{X}_n \not\perp c_n\mathbf{Y}_n \quad \textit{for all } c_n > 0$$

where $\mathbf{X}_n$ *and* $\mathbf{Y}_n \in \mathbb{R}^{n\times d}$ *are latent positions for two random dot product graphs with adjacency matrices* $\mathbf{A}_n$ *and* $\mathbf{B}_n$*, respectively. Define the test statistic* T_n *as follows:*

$$T_n = \frac{\min_{\mathbf{W}\in\mathcal{O}(d)} \|\hat{\mathbf{X}}_n\mathbf{W}/\|\hat{\mathbf{X}}_n\|_F - \hat{\mathbf{Y}}_n/\|\hat{\mathbf{Y}}_n\|_F\|_F}{2\sqrt{d\gamma_2^{-1}(\mathbf{A}_n)}/\|\hat{\mathbf{X}}_n\|_F + 2\sqrt{d\gamma_2^{-1}(\mathbf{B}_n)}/\|\hat{\mathbf{Y}}_n\|_F}. \tag{3.3}$$

Let $\alpha \in (0,1)$ *be given. Then for all* $C > 1$*, if the rejection region is defined by*

$$R := \{t \in \mathbb{R} : t \geq C\},$$

then there exists an $n_1 = n_1(\alpha, C) \in \mathbb{N}$ *such that for all* $n \geq n_1$*, the test procedure with* T_n *and rejection region* R *is an at most level* α *test. Furthermore, consider the sequence of latent position* $\{\mathbf{X}_n\}$ *and* $\{\mathbf{Y}_n\}$*,* $n \in \mathbb{N}$*, satisfying Assumption 1 and for which*

$$\frac{\min_{\mathbf{W}\in\mathcal{O}(d)} \|\mathbf{X}_n\mathbf{W}/\|\mathbf{X}_n\|_F - \mathbf{Y}_n/\|\mathbf{Y}_n\|_F\|_F}{C(\mathbf{X}_n)/\|\mathbf{X}_n\|_F + C(\mathbf{Y}_n)/\|\mathbf{Y}_n\|_F} = d_n. \tag{3.4}$$

Suppose $d_n \neq 0$ *for infinitely many* n*. Let* $t_1 = \min\{k > 0 : d_k > 0\}$ *and sequentially define* $t_n = \min\{k > t_{n-1} : d_k > 0\}$*. Let* $b_n = d_{t_n}$*. If* $\liminf b_n = \infty$*, then this test procedure is consistent in the sense of Definition 2.5 over this sequence of latent positions.*

Remark. We remark that the collection of alternatives in Eq. (3.4) is, effectively, those latent positions $\mathbf{X}_n$ and $\mathbf{Y}_n$ whose scalings onto the sphere remain far enough apart as $n \to \infty$. Indeed, we note that the denominator of our test statistic converges to zero, so we require that the numerator does not become small too quickly. Moreover, since $C(\mathbf{X}_n)$ and $C(\mathbf{Y}_n)$ are bounded by from above by constants, we can replace them by fixed constants (say, 1) in the denominator of Eq. (3.4) and obtain an equivalent class of alternatives.

3.3. *Diagonal Case.* Finally, we consider the case of testing the null hypothesis $H_0{:}\mathbf{X}_n \perp\!\!\!\perp \mathbf{D}_n\mathbf{Y}_n$ for some diagonal matrix $\mathbf{D}_n$. We proceed analogously to the previous case by defining the class $\mathcal{E} = \mathcal{E}(\mathbf{Y}_n)$ to be all positive diagonal matrices $\mathbf{D}_n \in \mathbb{R}^{n\times n}$ such that $\mathbf{D}_n\mathbf{Y}_n\mathbf{Y}_n^T\mathbf{D}_n$ has all entries in the unit interval. Once more, we focus on testing the null hypothesis $H_0{:}\mathbf{X}_n \perp\!\!\!\perp \mathbf{D}_n\mathbf{Y}_n$ for some $\mathbf{D}_n \in \mathcal{E}$ against the alternative $H_a{:}\mathbf{X}_n \not\perp\!\!\!\perp \mathbf{D}_n\mathbf{Y}_n$ for any matrix $\mathbf{D}_n \in \mathcal{E}$. As before, in what follows in this section, we will always assume that $\mathbf{D}_n$ belongs to $\mathcal{E}$, even if this assumption is not explicitly stated. The test statistic T_n in this case is again a simple modification of the one used in Theorem 3.1. However, for technical reasons, our proof of consistency requires an additional condition on the minimum Euclidean norm of each row of the matrices $\mathbf{X}_n$ and $\mathbf{Y}_n$. To avoid certain technical issues, we impose a slightly stronger density assumption on our graphs for this test. These assumptions can be weakened, but at the cost of interpretability. The assumptions we make on the latent positions, which we summarize here, are moderate restrictions on the sparsity of the graphs.

Assumption 2. *We assume that there exists* $d \in \mathbb{N}$ *such that for all* n*,* $\mathbf{P}_n$ *is of rank* d*. Further, we assume that there exist constants* $\epsilon_1 > 0$*,* $\epsilon_2 > 0$*,* $c_0 > 0$ *and* $n_0(\epsilon_1, \epsilon_2, c) \in \mathbb{N}$ *such that for all* $n \geq n_0$*:*

$$\gamma_1(\mathbf{P}_n) > c_0 \tag{3.5}$$

$$\delta(\mathbf{P}_n) > n^{1/2}(\log n)^{\epsilon_1} \tag{3.6}$$

$$\min_i \|X_i\| > \left(\frac{\log n}{\sqrt{\delta(\mathbf{P}_n)}}\right)^{1-\epsilon_2} \tag{3.7}$$

We then have the following result.

THEOREM 3.3. *For each fixed n, consider the hypothesis test*

$$\begin{aligned} &H_0^n\colon \mathbf{X}_n \perp\!\!\!\perp \mathbf{D}_n\mathbf{Y}_n \quad \text{for some diagonal } \mathbf{D}_n \in \mathcal{E} \text{ versus} \\ &H_a^n\colon \mathbf{X}_n \not\perp\!\!\!\perp \mathbf{D}_n\mathbf{Y}_n \quad \text{for any diagonal } \mathbf{D}_n \in \mathcal{E} \end{aligned}$$

where $\mathbf{X}_n$ and $\mathbf{Y}_n \in \mathbb{R}^{n\times d}$ are matrices of latent positions for two random dot product graphs. Let $\hat{\mathbf{X}}_n$ and $\hat{\mathbf{Y}}_n$ be the adjacency spectral embeddings for $\mathbf{X}$ and $\mathbf{Y}$, respectively. For any matrix $\mathbf{Z} \in \mathbb{R}^{n\times d}$, let $\mathcal{D}(\mathbf{Z})$ be the diagonal matrix whose diagonal entries are the Euclidean norm of the rows of $\mathbf{Z}$:

$$\mathcal{D}(\mathbf{Z}) = (\operatorname{diag}(\mathbf{Z}\mathbf{Z}^T))^{1/2}.$$

Denote by $\mathcal{P}(\mathbf{Z})$ the projection of the rows of $\mathbf{Z}$ onto the unit sphere:

$$\mathcal{P}(\mathbf{Z}) = \mathcal{D}^{-1}(\mathbf{Z})\mathbf{Z}$$

where, for simplicity of notation, we write $\mathcal{D}^{-1}(\mathbf{Z})$ for $(\mathcal{D}(\mathbf{Z}))^{-1}$. For any matrix $\mathbf{M} \in \mathbb{R}^{n\times n}$ with positive entries, define $S(\mathbf{M})$ as follows:

$$S(\mathbf{M}) = \frac{170 d^{3/2} \log(n/\eta)}{\sqrt{\gamma_1^7(\mathbf{M})\delta(\mathbf{M})}}. \tag{3.8}$$

We define the test statistic as follows:

$$T_n = \frac{\min_{\mathbf{W}\in\mathcal{O}(d)} \|\mathcal{P}(\hat{\mathbf{X}}_n)\mathbf{W} - \mathcal{P}(\hat{\mathbf{Y}}_n)\|_F}{S(\mathbf{A}_n)\|\mathcal{D}^{-1}(\hat{\mathbf{X}}_n)\|_F + S(\mathbf{B}_n)\|\mathcal{D}^{-1}(\hat{\mathbf{Y}}_n)\|_F}. \tag{3.9}$$

Let $\alpha \in (0,1)$ be given. Then for all $C > 1$, if the rejection region is defined by

$$R := \{t \in \mathbb{R} : t \geq C\},$$

then there exists an $n_1 = n_1(\alpha, C) \in \mathbb{N}$ such that for all $n \geq n_1$, the test procedure with T_n and rejection region R is an at most level-α test. Furthermore, consider the sequence of latent position $\{\mathbf{X}_n\}$ and $\{\mathbf{Y}_n\}$, $n \in \mathbb{N}$, satisfying Assumption 2 and for which

$$D_{\mathcal{P}}(\mathbf{X},\mathbf{Y}) := \frac{\min_{\mathbf{W}\in\mathcal{O}(d)} \|\mathcal{P}(\mathbf{X}_n)\mathbf{W} - \mathcal{P}(\mathbf{Y}_n)\|_F}{S(\mathbf{P}_n)\|\mathcal{D}^{-1}(\mathbf{X})\|_F + S(\mathbf{Q}_n)\|\mathcal{D}^{-1}(\mathbf{Y})\|_F} = d_n \tag{3.10}$$

Suppose $d_n \neq 0$ for infinitely many n. Let $t_1 = \min\{k > 0 : d_k > 0\}$ and sequentially define $t_n = \min\{k > t_{n-1} : d_k > 0\}$. Let $b_n = d_{t_n}$. If $\liminf b_n = \infty$, then this test procedure is consistent in the sense of Definition 2.5 over this sequence of latent positions.

REMARK. If the latent positions of $\mathbf{X}$ and $\mathbf{Y}$ are related by a diagonal transformation, this implies that each row X_i of $\mathbf{X}$ is a scaled version of the corresponding row Y_i of $\mathbf{Y}$; that is, $X_i = c_i Y_i$. Under the null, the angle between the adjacency spectral embeddings $\hat{X}_i$ and $\hat{Y}_i$ should be small. This suggests that we consider a cosine distance between the rows, and the projection in the numerator of our test statistic is essentially just that: namely, it measures the distance between projections of rows of the latent positions on the sphere (see Fig. 1). We make this specific choice of test statistic because it allows us to exploit a series of existing results on bounds for the deviations between true and estimated latent positions. There are several other reasonable choices of test statistic; ours happens to be straightforward to analyze, and the denominator is a natural consequence of a collection of known bounds on the accuracy of the adjacency spectral embedding. Furthermore, this explains our choice for the class of alternatives, which may initially appear rather Byzantine, for which the test satisfies a notion of consistency similar to that given in Definition 2.5. To illustrate, consider the following example of the class of alternatives for which Eq. (3.10) is satisfied. Let $\mathbf{X}_n$ and $\mathbf{Y}_n$ be two matrices of latent positions and suppose that $\min_i \|X_i\| \geq c$ and $\min_i \|Y_i\| \geq c$ for some constant $c > 0$. Then $\delta(\mathbf{P}_n)$ and $\delta(\mathbf{Q}_n)$ are of order $\Theta(n)$. $S(\mathbf{P}_n)$ and $S(\mathbf{Q}_n)$ are of order $\Theta(\log n/\sqrt{n})$. Finally, $\|\mathcal{D}^{-1}(\mathbf{X}_n)\|_F$ and $\|\mathcal{D}^{-1}(\mathbf{Y}_n)\|$ are of order $\Theta(\sqrt{n})$. Therefore, Eq. (3.10) is satisfied provided that $\min_{\mathbf{W}} \|\mathcal{P}(\mathbf{X}_n)\mathbf{W} - \mathcal{P}(\mathbf{Y}_n)\|_F$ diverges to ∞ at a rate faster than $\log n$. We also emphasize that $D_{\mathcal{P}}(\mathbf{X}, \mathbf{Y})$ has the property $D_{\mathcal{P}}(c\mathbf{X}, c\mathbf{Y}) = c^2 D_{\mathcal{P}}(\mathbf{X}, \mathbf{Y})$ for all $c \in (0, 1]$. This suggests that our test procedure may not have much power when the graphs are sparse.

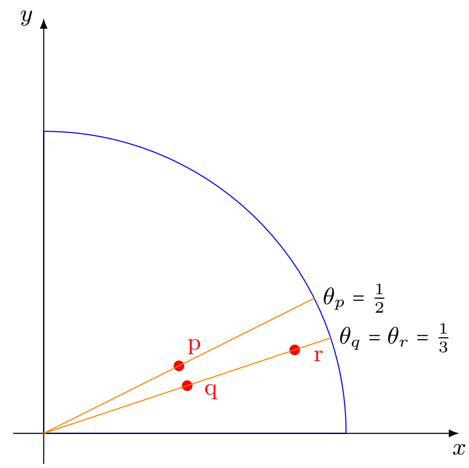

Fig 1: A pictorial example to illustrate the effect of projection. The distance between p and q is originally small, but increases after projection of p to $\theta_p = 1/2$ and q to $\theta_q = 1/3$. The distance between q and r after projection is zero and the distance between p and r after projection decreases.

4. Simulations. In this section, we illustrate the hypothesis tests and test statistics of Section 2 through several simulated data examples. We first consider the problem of testing the null hypothesis $H_0\colon \mathbf{X}_n \perp \mathbf{Y}_n$ against the alternative hypothesis $H_A\colon \mathbf{X}_n \not\perp \mathbf{Y}_n$. We consider random graphs generated according to two stochastic blockmodels with the same block membership probability vector $\boldsymbol{\pi}$ but different block probability matrices. Define $\mathbf{B}_\epsilon$ for $\epsilon \geq 0$ by

$$\mathbf{B}_\epsilon = \begin{bmatrix} 0.5+\epsilon & 0.2 \\ 0.2 & 0.5+\epsilon \end{bmatrix}. \tag{4.1}$$

We then test, for a given $\epsilon > 0$, the hypothesis $H_0\colon \mathbf{X}_n \perp \mathbf{Y}_n^{(\epsilon)}$ against $H_A\colon \mathbf{X}_n \not\perp \mathbf{Y}_n^{(\epsilon)}$ where $\mathbf{X}_n$ corresponds to $\mathbf{B}_0$ and $\mathbf{Y}_n^{(\epsilon)}$ corresponds to $\mathbf{B}_\epsilon$. We evaluate the performance of the test procedure by estimating the level and power of the test statistic for various choices of $n \in \{100, 200, 500, 1000\}$ and $\epsilon \in \{0, 0.05, 0.1, 0.2\}$ through Monte Carlo simulation. The significance level is set to $\alpha = 0.05$ and the rejection regions are specified via one of two approaches, namely (1) bootstrap resampling from the estimated latent positions $\hat{\mathbf{X}}_n$ and $\hat{\mathbf{Y}}_n$ and (2) $\{T > 1\}$ as dictated by the asymptotic theory. The results are given in Table 1 To keep the vertex set fixed and consistent, the block membership vector is sampled once in each Monte Carlo replicate.

	$\epsilon = 0$		$\epsilon = 0.05$		$\epsilon = 0.1$		$\epsilon = 0.2$	
n	bootstrap	theoretical	bootstrap	theoretical	bootstrap	theoretical	bootstrap	theoretical
100	0.08	0	0.11	0	0.31	0	0.99	0.13
200	0.07	0	0.22	0	0.96	0	1	0.98
500	0.06	0	0.97	0	1	0	1	1
1000	0.05	0	1	0	1	1	1	1

TABLE 1

Power estimates for testing the null hypothesis $\mathbf{X}_n \perp \mathbf{Y}_n$ at a significance level of $\alpha = 0.05$. The rejection regions are specified via two methods (1) the asymptotic theoretical rejection region and (2) bootstrap permutation with $B = 200$ bootstrap samples. Each estimate of power is based on 1000 Monte Carlo replicates.

Table 1 indicates that the test has good power and is indeed asymptotically level α. The rejection regions computed using bootstrap resampling are generally less conservative than those specified via the asymptotic theory. Nevertheless, the theoretical rejection regions exhibit power even for moderate values of n such as $n = 200$.

As another example, we consider the problem of detecting the emergence of a new community in a graph. Let $\mathbf{B}_0$ and $\mathbf{B}_1$ be block probability matrices defined by

$$\mathbf{B}_1 \begin{pmatrix} 0.34 & 0.25 \\ 0.25 & 0.25 \end{pmatrix}; \qquad \mathbf{B}_2 = \begin{pmatrix} 0.34 & 0.25 & 0.16 \\ 0.25 & 0.25 & 0.25 \\ 0.16 & 0.25 & 0.34 \end{pmatrix}$$

Graphs generated with block probability matrix $\mathbf{B}_1$ have two blocks of size 400 each while graphs with block probability matrix $\mathbf{B}_2$ have three blocks of size $400 - n_3/2$, $400 - n_3/2$ and $2n_3$. The results are presented in Fig. 2 for various values of $n_3 \in \{0, 4, \ldots, 16\}$.

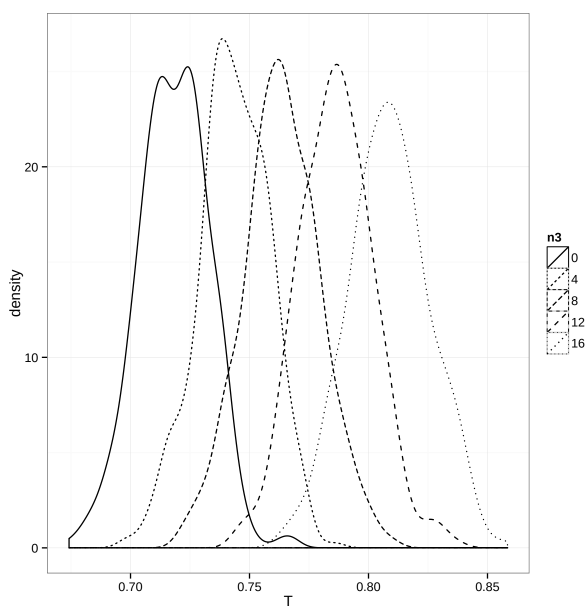

Fig 2: Density estimate (based on 500 Monte Carlo replicates) for the test statistic to detect the emergence of a community of size $n_3 \in \{0, 4, \ldots, 16\}$ in a graph on $n = 800$ vertices. Bootstrap estimates of the critical values for $\alpha = 0.05$ yield power estimates of 0.57 for $n_3 = 4$, 0.92 for $n_3 = 8$, and 1.0 for $n_3 = 12$ and $n_3 = 16$.

	$\epsilon = 0$		$\epsilon = 0.1$		$\epsilon = 0.2$		$\epsilon = 0.4$	
n	bootstrap	theoretical	bootstrap	theoretical	bootstrap	theoretical	bootstrap	theoretical
100	0.08	0	0.08	0	0.19	0	0.87	0
200	0.06	0	0.15	0	0.61	0	1	0
500	0.05	0	0.62	0	1	0	1	1
1000	0.04	0	1	0	1	0	1	1

TABLE 2

Power estimates for testing the null hypothesis $\mathbf{X}_n \perp c_n \mathbf{Y}_n$ for some $c_n > 0$ at a significance level of $\alpha = 0.05$. The rejection regions are specified via two methods (1) the asymptotic theoretical rejection region and (2) bootstrap permutation with $B = 200$ bootstrap samples. Each estimate of power is based on 1000 Monte Carlo replicates.

We next consider the hypothesis test $H_0 \colon \mathbf{X}_n \perp c_n \mathbf{Y}_n$ for some $c_n > 0$ against the alternative $H_A \colon \mathbf{X}_n \not\perp c_n \mathbf{Y}_n$ for any $c_n > 0$. We again employ the model specified in Eq. (4.1). The results are presented in Table 2. We observe that the power of the test is estimated to be roughly 0.19 for $n = 100$ and $\epsilon = 0.2$, which is significantly smaller than the corresponding estimate of 0.99 in Table 1, even though the random graphs models are identical. This is consistent with the notion that the null hypothesis considered in Table 1 is a single element of the hypothesis space in Table 2. For this setup, the theoretical rejection region as specified in Theorem 3.2 exhibits power for moderate values of $n = 500$ and $\epsilon = 0.4$.

C. Elegans wiring diagram. We end this section with a description of how our test procedure applies to the two neuronal networks of *C. elegans*. As we remarked earlier, *C. elegans* has both a network of chemical synapses and a network of electrical gap junctions. The 302 total neurons in *C. elegans* are divided into the 20 that belong to phyrangeal nervous system and the remaining 282 that belong to the somatic nervous system; of the latter, three neurons have no synaptic connection to other neurons. This gives rise to two graphs $\mathbf{A}_c$ and $\mathbf{A}_g$, both on 279 vertices. Graph $\mathbf{A}_c$ has 6393 undirected edges and graph $\mathbf{A}_c$ has 1031 undirected edges. See [27] for more detailed description of the reconstruction of these neuronal networks.

We address the question of whether these two graphs are "similar" by framing it as a two-sample testing problem within our model[1]. Because the two graphs have a significant difference in the number of edges, the null hypothesis is that the generating latent positions are equal up to scaling factor c.

To carry out the test, we embed each graph as a collection of points in $\mathbb{R}^d$ with $d = 6$. The choice of $d = 6$ is selected using the automatic dimension selection procedure of [31]. Denoting by $\hat{\mathbf{X}}_c$ and $\hat{\mathbf{X}}_g$ the resulting embeddings, we compute the test statistic

$$T(\hat{\mathbf{X}}_c, \hat{\mathbf{X}}_g) = \frac{\min_{\mathbf{W} \in \mathcal{O}(d)} \|\hat{\mathbf{X}}_g \mathbf{W} / \|\hat{\mathbf{X}}_g\|_F - \hat{\mathbf{X}}_c / \|\hat{\mathbf{X}}_c\|_F\|_F}{2C(\hat{\mathbf{X}}_g)/\|\hat{\mathbf{X}}_g\|_F + 2C(\hat{\mathbf{X}}_c)/\|\hat{\mathbf{Y}}_c\|_F}.$$

as described in Section 3. We obtain $T(\hat{\mathbf{X}}_c, \hat{\mathbf{X}}_g) = 1.465$. To approximate the p-value, we proceed by random sampling of $B = 1000$ pairs of graphs when the latent positions are given by $\hat{\mathbf{X}}_c$ and random sampling of another $B = 1000$ pairs of graphs when the latent positions are $\hat{\mathbf{X}}_g$. We then compute the associated test statistics for both collections. The approximate p-value associated with the value $T = 1.465$ of the test statistic is smaller than 0.001. We reject the null and conclude that

[1]Reservations about the applicability of this model to *C. elegans* are, of course, legitimate, but we do not address the intricate question of model selection here.

the two neuronal networks are sufficiently distinct (within the collection of networks that can be modeled appropriately by random dot product graphs with latent positions in $\mathbb{R}^6$).

5. Discussion. In summary, we show in this paper that the adjacency spectral embedding can be used to generate simple and intuitive test statistics for the inference problem of testing whether two random dot product graphs on the same vertex set have the same or related generating latent positions. Two-sample graph inference has significant applications in diverse fields. Our test is both a principled and, as our real data examples illustrate, practically viable inference procedure.

Our concentration inequalities allow us to obtain an at most level-α consistent test without specifying the finite-sample or asymptotic distribution of our test statistic. We do not, at present, have a limiting distributional result for our test statistic, and we suspect that such a result would require additional, more restrictive, model assumptions.

The test statistic based on orthogonal Procrustes matching $\min_{\mathbf{W}\in\mathcal{O}(d)} \|\hat{\mathbf{X}}\mathbf{W} - \hat{\mathbf{Y}}\|_F$ is but one of many possible test statistics for testing the hypothesis $\mathbf{X} \perp \mathbf{Y}$. For example, the test statistic $\|\mathbf{A} - \mathbf{B}\|_F$ is intuitively appealing; it is a surrogate measure for the difference $\|\mathbf{X}\mathbf{X}^T - \mathbf{Y}\mathbf{Y}^T\|_F$. Furthermore, $\|\mathbf{A} - \mathbf{B}\|_F^2 = 2\sum_{i<j}(\mathbf{A}_{ij} - \mathbf{B}_{ij})^2$ is a sum of independent Bernoulli random variables; hence it is easily analyzable and may possibly yield more powerful test. However, since $(\mathbf{A}_{ij}-\mathbf{B}_{ij})^2$ is a Bernoulli random variable with parameter $\mathbf{P}_{ij}(1-\mathbf{Q}_{ij})+(1-\mathbf{P}_{ij})\mathbf{Q}_{ij}$, this forces that $(\mathbf{A}_{ij}-\mathbf{B}_{ij})^2 \sim$ Bernoulli$(1/2)$ if $\mathbf{Q}_{ij} = 1/2$, regardless of the value of $\mathbf{P}_{ij}$. Therefore, $\|\mathbf{A}-\mathbf{B}\|_F^2 \sim$ Binomial$(\binom{n}{2}, 1/2)$ whenever $\mathbf{Q} = 1/2\mathbf{J}$ where $\mathbf{J}$ is the matrix of all ones. Thus, $\|\mathbf{A} - \mathbf{B}\|_F$ yields a test that is not consistent for a large class of alternatives.

To relate our test to classical generalized likelihood ratio tests, we note that if we have two independent random dot product graphs with no rank restrictions, the generalized likelihood ratio test statistic reduces to

$$\Lambda = \|\mathbf{A} - \mathbf{B}\|_F^2$$

which is the aforementioned Frobenius norm test statistic. However, computing the generalized likelihood ratio test statistic under rank assumptions is computationally more challenging. We can approximate this quantity by

$$\hat{\Lambda} = \frac{\mathbb{P}(\mathbf{A}|\hat{\mathbf{X}}\hat{\mathbf{X}}^{\mathbf{T}})\mathbb{P}(\mathbf{B}|\hat{\mathbf{X}}\hat{\mathbf{X}}^{\mathbf{T}})}{\mathbb{P}(\mathbf{A}|\hat{\mathbf{Z}})\mathbb{P}(\mathbf{B}|\hat{\mathbf{Z}})}$$

where $\hat{\mathbf{Z}} = \frac{\hat{\mathbf{X}}\hat{\mathbf{X}}^{\mathbf{T}} + \hat{\mathbf{Y}}\hat{\mathbf{Y}}^{\mathbf{T}}}{2}$. The question of how valid this approximation is, and how the limiting distribution of this test statistic is related to ours, is the subject of further research. We emphasize that the likelihood ratio has an independence assumption that we do not require. Also, since $\hat{\mathbf{X}}$ is a consistent estimate for $\mathbf{X}$, our test statistic, which is a scaled version of $||\mathbf{X} - \mathbf{Y}\mathbf{W}||_F$, is in the spirit of a Wald test.

Yet another simple test statistic is based on the spectral norm difference $\|\mathbf{A}-\mathbf{B}\|$; this is once again a surrogate measure for the difference $\|\mathbf{X}\mathbf{X}^T - \mathbf{Y}\mathbf{Y}^T\|$, and such a test statistic may be more robust to model misspecification, e.g. when $\mathbf{A}$ and $\mathbf{B}$ are adjacency matrices of more general latent position random graphs. The concentration bound of [17], which we state in Eq. (A.1) in Proposition A.1, can be used to construct a level-α test for the hypothesis $\mathbf{X} \perp \mathbf{Y}$. This may, however, lead to a test that is consistent for a narrower class of alternatives than that given in Theorem 3.1. Indeed, by Eq. (A.1), the test $\|\mathbf{A}-\mathbf{B}\|$ is consistent when the sequence $((\delta(\mathbf{P}_n) + \delta(\mathbf{Q}_n)) \log n)^{-1/2} \|\mathbf{P}_n - \mathbf{Q}_n\|$

diverges. In addition, we have

$$\begin{aligned}\|\mathbf{P}_n - \mathbf{Q}_n\| &= \|\mathbf{X}_n\mathbf{X}_n^T - \mathbf{Y}_n\mathbf{Y}_n^T\| \\ &= \|(\mathbf{X}_n\mathbf{W}_n - \mathbf{Y}_n)(\mathbf{X}_n\mathbf{W}_n + \mathbf{Y}_n)^T + (\mathbf{X}_n\mathbf{W}_n + \mathbf{Y}_n)(\mathbf{X}_n\mathbf{W}_n - \mathbf{Y}_n)^T\| \\ &\leq 2\|\mathbf{X}_n\mathbf{W}_n - \mathbf{Y}_n\|\|\mathbf{X}_n\mathbf{W}_n + \mathbf{Y}_n\| \leq 2\|\mathbf{X}_n\mathbf{W}_n - \mathbf{Y}_n\|\sqrt{\delta(\mathbf{P}_n) + \delta(\mathbf{Q}_n)}.\end{aligned}$$

for all orthogonal $\mathbf{W}_n$. Therefore, the fact that $(\delta(\mathbf{P}_n) + \delta(\mathbf{Q}_n))^{-1/2}\|\mathbf{P}_n - \mathbf{Q}_n\|$ diverges implies that $\min_{\mathbf{W}\in\mathcal{O}(d)} \|\mathbf{X}_n\mathbf{W}_n - \mathbf{Y}_n\|$ diverges, but not conversely, and hence $\|\mathbf{A} - \mathbf{B}\|$ may be consistent for a narrower class of alternatives. In addition, test statistics based directly on the adjacency matrices are also less flexible. For instance, it is not obvious to us that such test statistics can be easily adapted to test the hypothesis $\mathbf{X} \perp \mathbf{DY}$ for some diagonal matrix $\mathbf{D}$, or to conduct the nonparametric test of equality of the underlying distributions for the latent positions.

Test statistics based on the spectral decomposition of the normalized Laplacian matrices can also be constructed. However, the resulting embedding is an estimate of some transformation of the latent positions rather than the latent positions themselves. More specifically, denote by $\widetilde{\mathbf{X}}_n$ and $\widetilde{\mathbf{Y}}_n$ the spectral decomposition obtained from the normalized Laplacian matrices associated with $\mathbf{A}_n$ and $\mathbf{B}_n$, respectively. Then $\widetilde{\mathbf{X}}_n$ is, up to some orthogonal transformation, "close" to $\mathcal{L}(\mathbf{X}_n)$ where $\mathcal{L}(\mathbf{X}_n)$ is a transformation of $\mathbf{X}_n$, i.e., the i-th row of $\mathcal{L}(\mathbf{X}_n)$ is given by $X_i/\langle X_i, \sum_{j\neq i} X_j\rangle$; similarly, $\widetilde{\mathbf{Y}}_n$ is "close" to $\mathcal{L}(\mathbf{Y}_n)$. The construction of test statistics for testing the hypothesis in Section 2 for $\mathbf{X}_n$ and $\mathbf{Y}_n$ based on the estimates $\widetilde{\mathbf{X}}_n$ and $\widetilde{\mathbf{Y}}_n$ of $\mathcal{L}(\mathbf{X}_n)$ and $\mathcal{L}(\mathbf{Y}_n)$ is certainly possible; however, subtle technical issues regarding assumptions on the sequence of latent positions and speed of convergence of the estimates $\widetilde{\mathbf{X}}_n$ and $\widetilde{\mathbf{Y}}_n$ can arise. In summary, the formulation of the hypotheses and the accompanying test procedures in Section 2 are such that the test statistics are simple functions of the adjacency spectral embeddings of the graphs. Other formulations of comparable two-sample tests could, of course, lead to test statistics that are simple functions of the normalized Laplacian embeddings.

REFERENCES

[1] Arias-Castro, E. and Verzelen, N. (2014). Community detection in dense random networks. *Annals of Statistics* **42** 940–969.

[2] Athreya, A., Lyzinski, V., Marchette, D. J., Priebe, C. E., Sussman, D. L. and Tang, M. (2013). A limit theorem for scaled eigenvectors of random dot product graphs. Arxiv preprint. http://arxiv.org/abs/1305.7388.

[3] Bickel, P. and Sarkar, P. (2013). Role of normalization for spectral clustering in stochastic blockmodels. Arxiv preprint. http://arxiv.org/abs/1310.1495.

[4] Bickel, P. J. and Sarkar, P. (2013). Hypothesis testing for automated community detection in networks. Arxiv preprint at arxiv.org/abs/1311.2694.

[5] Boucheron, S., Lugosi, G. and Massart, P. (2013). *Concentration Inequalities: A nonasymptotic theory of independence*. Oxford University Press.

[6] Bullmore, E. and Sporns, O. (2009). Complex brain networks: Graph theoretical analysis of structural and functional systems. *Nature Rev. Neurosci* **10** 186–198.

[7] Chaudhuri, K., Chung, F. and Tsiatas, A. (2012). Spectral partitioning of graphs with general degrees and the extended planted partition model. In *Proceedings of the 25th conference on learning theory*.

[8] Conte, D., Foggia, P., Sansone, C. and Vento, M. (2004). Thirty years of graph matching in pattern recognition. *International Journal of Pattern Recognition and Artificial Intelligence* **18** 265–298.

[9] Fortunato, S. (2010). Community detection in graphs. *Physics Reports* **486** 75–174.

[10] Goldenberg, A., Zheng, A. X., Fienberg, S. E. and Airoldi, E. M. (2010). A survey of statistical network models. *Foundations and Trends® in Machine Learning* **2** 129–233.

[11] Gray, W. R., Bogovic, J. A., Vogelstein, J. T., Landman, B. A., Prince, J. L. and Vogelstein, R. J. (2012). Magnetic resonance connectome automated pipeline: an overview. *IEEE Pulse* **3** 42–48.

[12] Hall, D. H. and Russell, R. L. (1991). The posterior nervous system of the nematode *Caenorhabdhitis elegans*: Serial reconstruction of identified neurons and complete pattern of synaptic interactions. *Journal of Neuroscience* **11** 1–22.

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

[14] Holland, P. W., Laskey, K. and Leinhardt, S. (1983). Stochastic blockmodels: First steps. *Social Networks* **5** 109–137.

[15] Karrer, B. and Newman, M. E. J. (2011). Stochastic blockmodels and community structure in networks. *Phys. Rev. E* **83**.

[16] Lyzinski, V., Sussman, D. L., Tang, M., Athreya, A. and Priebe, C. E. (2013). Perfect Clustering for Stochastic Blockmodel Graphs via Adjacency Spectral Embedding. Arxiv preprint. http://arxiv.org/abs/1310.0532.

[17] Oliveira, R. I. (2010). Concentration of the adjacency matrix and of the Laplacian in random graphs with independent edges. Arxiv preprint http://arxiv.org/abs/0911.0600.

[18] Qin, T. and Rohe, K. (2013). Regularized spectral clustering under the degree-corrected stochastic blockmodel. arXiv preprint.

[19] Richiardi, J., Achard, S., Bunke, H. and Ville, D. V. D. (2013). Machine Learning With Brain Graphs: predictive modeling approaches for functional imaging in systems neuroscience. *IEEE Signal Processing* **66** 58–70.

[20] Richiardi, J., Eryilmaz, H., Schwartz, S., Vuilleumier, P. and Ville, D. V. D. (2011). Decoding brain states from fMRI connectivity graphs. *NeuroImage* **56** 616–626.

[21] Rohe, K., Chatterjee, S. and Yu, B. (2011). Spectral clustering and the high-dimensional stochastic blockmodel. *Ann. Statist.* **39** 1878–1915.

[22] Rukhin, A. and Priebe, C. E. (2011). A comparative power analysis of the maximum degree and size invariants for random graph inference. *Journal of Statistical Planning and Inference* **141** 1041–1046.

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

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

[25] Tang, M., Athreya, A., Sussman, D. L., Lyzinski, V. and Priebe, C. E. (2014). A nonparametric two-sample hypothesis testing problem for random dot product graphs. Arxiv preprint. http://arxiv.org/abs/1409.2344.

[26] Tang, M., Sussman, D. L. and Priebe, C. E. (2013). Universally consistent vertex classification for latent position graphs. *Annals of Statistics* **31** 1406–1430.

[27] Varshney, L. R., Chen, B. L., Paniagua, E., Hall, D. H. and Chklovskii, D. B. (2011). Structural properties of the *caenorhabditis elegans* neuronal network. *PLOS computational biology* **7**.

[28] White, J. G., Southgate, E., Thomson, J. N. and Brenner, S. (1986). The structure of the nervous system of the nematode *Caenorhabdhitis elegans. Philosophical Transactions of the Royal Society B* **314** 1–340.

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

[30] Zalesky, A., Fornito, A., Harding, I., Cocchi, L., Yücel, M., Pantalis, C. and Bullmore, E. (2010). Whole-brain anatomical networks: does the choice of node matter? *NeuroImage* **50** 970–983.

[31] Zhu, M. and Ghodsi, A. (2006). Automatic dimensionality selection from the scree plot via the use of profile likelihood. *Computational Statistics and Data Analysis* **51** 918–930.

APPENDIX: ADDITIONAL LEMMAS AND PROOFS

Established bounds. We first state several known bounds on the accuracy of the adjacency spectral embedding from [2, 3, 17, 24, 26].

Proposition A.1. *Let $\hat{\mathbf{X}}_n \in \mathbb{R}^{n\times d}$ be the adjacency spectral embedding of the $n \times n$ adjacency matrix $\mathbf{A}_n \sim \mathrm{Bernoulli}(\mathbf{P}_n)$ where $\mathbf{P}_n = \mathbf{X}_n\mathbf{X}_n^T$ is of rank d and its non-zero eigenvalues are distinct. Suppose also that there exists $\epsilon > 0$ such that $\delta(\mathbf{P}_n) \geq (\log n)^{1+\epsilon}$. Let $c > 0$ be arbitrary but fixed. There exists $n_0(c)$ such that if $n > n_0$ and η satisfies $n^{-c} < \eta < 1/2$, then with probability at least $1-\eta$,*

$$\|\mathbf{P}_n - \mathbf{A}_n\| \leq 2\sqrt{\delta(\mathbf{P}_n)\log(n/\eta)} \tag{A.1}$$

and

$$\min_{\mathbf{W} \in \mathcal{O}(d)} \|\hat{\mathbf{X}}_n \mathbf{W} - \mathbf{X}_n\|_F \leq 4\gamma_2^{-1}(\mathbf{P}_n)\sqrt{2d \log{(n/\eta)}} \tag{A.2}$$

Next, we state a simple proposition on the consistency of adjacency-based estimates of $\delta(\mathbf{P}_n), \gamma_1(\mathbf{P}_n)$, and $\gamma_2(\mathbf{P}_n)$. This proposition is a straightforward consequence of Proposition A.1, Hoeffding's inequality, and the Borel-Cantelli lemma, and we omit its proof.

PROPOSITION A.2. *Let $\{\mathbf{X}_n\}$ be a sequence of latent positions and suppose that the sequence of matrices $\{\mathbf{P}_n\}$, where $\mathbf{P}_n = \mathbf{X}_n \mathbf{X}_n^T$, satisfy the condition in Eq.(2.3) in Assumption 1. Let $\{\mathbf{A}_n\}$ be the sequence of adjacency matrices $\mathbf{A}_n \sim \mathrm{Bernoulli}(\mathbf{P}_n)$. Then we have*

$$\frac{\delta(\mathbf{A}_n)}{\delta(\mathbf{P}_n)} \xrightarrow{\text{a.s.}} 1; \quad \frac{\gamma_1(\mathbf{A}_n)}{\gamma_1(\mathbf{P}_n)} \xrightarrow{\text{a.s.}} 1; \quad \frac{\gamma_2(\mathbf{A}_n)}{\gamma_2(\mathbf{P}_n)} \xrightarrow{\text{a.s.}} 1; \tag{A.3}$$

Additional Lemmas. Let $\mathbf{W}$ be such that $\mathbf{U}_{\mathbf{P}}\mathbf{S}_{\mathbf{P}}^{1/2} = \mathbf{X}\mathbf{W}$. We note that such a matrix $\mathbf{W}$ always exists as $\mathbf{U}_{\mathbf{P}}\mathbf{S}_{\mathbf{P}}\mathbf{U}_{\mathbf{P}}^T = \mathbf{P} = \mathbf{X}\mathbf{X}^T$. The proof of Theorem 2.3 proceeds by bounding, in a series of technical lemmas, each of the terms in parentheses in the following decomposition of $\hat{\mathbf{X}} - \mathbf{X}\mathbf{W}$:

$$\begin{aligned}
\hat{\mathbf{X}} - \mathbf{X}\mathbf{W} &= \mathbf{U}_{\mathbf{A}}\mathbf{S}_{\mathbf{A}}^{1/2} - \mathbf{U}_{\mathbf{P}}\mathbf{S}_{\mathbf{P}}^{1/2} \\
&= \mathbf{A}\mathbf{U}_{\mathbf{A}}\mathbf{S}_{\mathbf{A}}^{-1/2} - \mathbf{P}\mathbf{U}_{\mathbf{P}}\mathbf{S}_{\mathbf{P}}^{-1/2} \\
&= (\mathbf{A}\mathbf{U}_{\mathbf{A}}\mathbf{S}_{\mathbf{A}}^{-1/2} - \mathbf{A}\mathbf{U}_{\mathbf{P}}\mathbf{S}_{\mathbf{A}}^{-1/2}) + (\mathbf{A}\mathbf{U}_{\mathbf{P}}\mathbf{S}_{\mathbf{A}}^{-1/2} - \mathbf{A}\mathbf{U}_{\mathbf{P}}\mathbf{S}_{\mathbf{P}}^{-1/2}) + (\mathbf{A}\mathbf{U}_{\mathbf{P}}\mathbf{S}_{\mathbf{P}}^{-1/2} - \mathbf{P}\mathbf{U}_{\mathbf{P}}\mathbf{S}_{\mathbf{P}}^{-1/2})
\end{aligned}$$

We now state these lemmas, beginning with a result from [2].

LEMMA A.3. *If the events in Proposition A.1 occur, then*

$$\|\mathbf{S}_{\mathbf{A}} - \mathbf{S}_{\mathbf{P}}\| \leq \frac{18d \log{(n/\eta)}}{\gamma_1^2(\mathbf{P})} \tag{A.4}$$

$$\|\mathbf{U}_{\mathbf{A}}^T(\mathbf{U}_{\mathbf{A}} - \mathbf{U}_{\mathbf{P}})\|_F \leq \frac{10d \log{(n/\eta)}}{\gamma_1^2(\mathbf{P})\delta(\mathbf{P})} \tag{A.5}$$

We continue with two results from [16]. The first result bounds $\|(\mathbf{A}\mathbf{U}_{\mathbf{A}}\mathbf{S}_{\mathbf{A}}^{-1/2} - \mathbf{A}\mathbf{U}_{\mathbf{P}}\mathbf{S}_{\mathbf{A}}^{-1/2}\|_F$ by viewing it as the difference after one step of the power method for $\mathbf{A}$ when starting at $\mathbf{U}_{\mathbf{P}}$. The second lemma bounds $\|\mathbf{A}\mathbf{U}_{\mathbf{P}}(\mathbf{S}_{\mathbf{A}}^{-1/2} - \mathbf{S}_{\mathbf{P}}^{-1/2})\|_F$.

LEMMA A.4. *If the events in Proposition A.1 occur, then*

$$\|\hat{\mathbf{X}} - \mathbf{A}\mathbf{U}_{\mathbf{P}}\mathbf{S}_{\mathbf{A}}^{-1/2}\|_F \leq \frac{24\sqrt{2}d \log{(n/\eta)}}{\sqrt{\gamma_1^5(\mathbf{P})\delta(\mathbf{P})}} \tag{A.6}$$

LEMMA A.5. *If the events in Proposition A.1 occur, then*

$$\|\mathbf{A}\mathbf{U}_{\mathbf{P}}(\mathbf{S}_{\mathbf{A}}^{-1/2} - \mathbf{S}_{\mathbf{P}}^{-1/2})\|_F \leq \frac{18d^{3/2} \log{(n/\eta)}}{\sqrt{\gamma_1^7(\mathbf{P})\delta(\mathbf{P})}}. \tag{A.7}$$

Our last technical lemma provides a concentration bound for $\|(\mathbf{A}-\mathbf{P})\mathbf{U}_{\mathbf{P}}\mathbf{S}_{\mathbf{P}}^{-1/2}\|_F$. Its proof relies on classical concentration inequalities; the argument we give is straightforward, if tedious.

LEMMA A.6. *Let $\eta > 0$ be arbitrary. Then with probability at least $1-2\eta$, the events in Proposition A.1 occur and furthermore,*

$$\big|\|(\mathbf{A}-\mathbf{P})\mathbf{U}_{\mathbf{P}}\mathbf{S}_{\mathbf{P}}^{-1/2}\|_F^2 - C^2(\mathbf{X})\big| \le \frac{14\sqrt{2d}\log(n/\eta)}{\gamma_2(\mathbf{P})\sqrt{\delta(\mathbf{P})}}. \tag{A.8}$$

where $C^2(\mathbf{X})$ is a constant that depends only on $\mathbf{X}$. In addition, $C^2(\mathbf{X})$ can be written as

$$C^2(\mathbf{X}) = \operatorname{tr} \mathbf{S}_{\mathbf{P}}^{-1/2}\mathbf{U}_{\mathbf{P}}^T\mathbb{E}[(\mathbf{A}-\mathbf{P})^2]\mathbf{U}_{\mathbf{P}}\mathbf{S}_{\mathbf{P}}^{-1/2} = \operatorname{tr} \mathbf{S}_{\mathbf{P}}^{-1/2}\mathbf{U}_{\mathbf{P}}^T\mathbf{D}\mathbf{U}_{\mathbf{P}}\mathbf{S}_{\mathbf{P}}^{-1/2} \le d\gamma_2^{-1}(\mathbf{P})$$

where $\mathbf{D}$ is a diagonal matrix whose diagonal entries are given by

$$\mathbf{D}_{ii} = \sum_{k\neq i}\mathbf{P}_{ik}(1-\mathbf{P}_{ik}).$$

PROOF. Let $Z = \|(\mathbf{A}-\mathbf{P})\mathbf{U}_{\mathbf{P}}\mathbf{S}_{\mathbf{P}}^{-1/2}\|_F^2$. Since our graphs are undirected and loop free, Z is a function of the $\binom{n}{2}$ independent random variables $\{\mathbf{A}_{ij}\}_{i<j}$. Let $\mathbf{A}$ and $\mathbf{A}'$ be two arbitrary adjacency matrices. Denote by $\mathbf{A}^{(kl)}$ the adjacency matrix obtained by replacing the (k,l) and (l,k) entries of $\mathbf{A}$ by those of $\mathbf{A}'$. Let $Z_{kl} = \|(\mathbf{A}^{(kl)}-\mathbf{P})\mathbf{U}_{\mathbf{P}}\mathbf{S}_{\mathbf{P}}^{-1/2}\|_F^2$. The argument we employ is based on the following logarithmic Sobolev concentration inequality for $Z-\mathbb{E}[Z]$ [5, §6.4]. Namely,

THEOREM A.7. *Assume that there exists a constant $v>0$ such that, with probability at least $1-\eta$,*

$$\sum_{k<l}(Z-Z_{kl})^2 \le v.$$

Then for all $t>0$,

$$\mathbb{P}[|Z-\mathbb{E}[Z]|>t] \le 2e^{-t^2/(2v)} + \eta.$$

Let $\mathbf{V} = \mathbf{U}_{\mathbf{P}}\mathbf{S}_{\mathbf{P}}^{-1/2}$. For notational convenience, we denote the i-th row of $\mathbf{V}$ by V_i. We shall also denote the inner product between vectors in Euclidean space by $\langle\cdot,\cdot\rangle$. The i-th row of the product $(\mathbf{A}-\mathbf{P})\mathbf{V}$ is simply a linear combination of the rows of $\mathbf{V}$, i.e.,

$$((\mathbf{A}-\mathbf{P})\mathbf{V})_i = \sum_{j=1}^{n}(\mathbf{A}-\mathbf{P})_{ij}V_j.$$

Hence,

$$Z = \|(\mathbf{A}-\mathbf{P})\mathbf{V}\|_F^2 = \sum_{i=1}^{n}\|((\mathbf{A}-\mathbf{P})\mathbf{V})_i\|^2 = \sum_{i=1}^{n}\sum_{j=1}^{n}\sum_{k=1}^{n}(\mathbf{A}-\mathbf{P})_{ij}(\mathbf{A}-\mathbf{P})_{ik}\langle V_j,V_k\rangle$$

As $\mathbf{A}$ and $\mathbf{A}^{(kl)}$ differs possibly only in the (k,l) and (l,k) entries and that the entries of $\mathbf{A}$ and $\mathbf{A}'$ are binary variables, we have that if $(Z-Z_{kl})$ is non-zero, then

$$\begin{aligned} Z - Z_{kl} &= 2\Big(\sum_{j\neq l}(\mathbf{A}-\mathbf{P})_{kj}\langle V_j,V_l\rangle\Big) + 2\Big(\sum_{j\neq k}(\mathbf{A}-\mathbf{P})_{lj}\langle V_j,V_k\rangle\Big) + (1-2\mathbf{P}_{kl})\langle V_l,V_k\rangle \\ &= 2\sum_{j=1}^{n}(\mathbf{A}-\mathbf{P})_{kj}\langle V_j,V_l\rangle\Big) + 2\sum_{j=1}^{n}(\mathbf{A}-\mathbf{P})_{lj}\langle V_j,V_k\rangle\Big) + c_{kl} \end{aligned}$$

where $c_{kl} = 2(\mathbf{A}-\mathbf{P})_{kl}\langle V_l, V_l\rangle + 2(\mathbf{A}-\mathbf{P})_{lk}\langle V_k, V_k\rangle + (1-2\mathbf{P})_{kl}\langle V_l, V_k\rangle$. We then have

$$(Z - Z_{kl})^2 \le 3(C_{kl}^{(1)} + C_{kl}^{(2)} + c_{kl}^2)$$

where $C_{kl}^{(1)}$ and $C_{kl}^{(2)}$ are given by

$$C_{kl}^{(1)} = 4\sum_{j_1=1}^{n}\sum_{j_2=1}^{n}(\mathbf{A}-\mathbf{P})_{kj_1}(\mathbf{A}-\mathbf{P})_{kj_2}\langle V_l, V_{j_1}\rangle\langle V_l, V_{j_2}\rangle = 4\Big[((\mathbf{A}-\mathbf{P})\mathbf{V}\mathbf{V}^T)_{kl}\Big]^2$$
$$C_{kl}^{(2)} = 4\sum_{j_1=1}^{n}\sum_{j_2=1}^{n}(\mathbf{A}-\mathbf{P})_{lj_1}(\mathbf{A}-\mathbf{P})_{lj_2}\langle V_k, V_{j_1}\rangle\langle V_k, V_{j_2}\rangle = 4\Big[((\mathbf{A}-\mathbf{P})\mathbf{V}\mathbf{V}^T)_{lk}\Big]^2$$

As $C_{kl}^{(1)} = C_{lk}^{(2)}$, $c_{kl} = c_{lk}$, and $C_{kk}^{(1)} > 0$ for all l, k, we thus have

$$\sum_{k<l}(Z - Z_{kl})^2 \le 3\sum_{k<l}\Big(C_{kl}^{(1)} + C_{kl}^{(2)} + c_{kl}^2\Big) \le 3\sum_{k=1}^{n}\sum_{l=1}^{n}C_{kl}^{(1)} + \frac{3}{2}\sum_{k=1}^{n}\sum_{l=1}^{n}c_{kl}^2$$

We now consider each of the term in the above right hand side.

$$\sum_{k=1}^{n}\sum_{l=1}^{n}C_{kl}^{(1)} = 4\sum_{k=1}^{n}\sum_{l=1}^{n}\Big[((\mathbf{A}-\mathbf{P})\mathbf{V}\mathbf{V}^T)_{kl}\Big]^2 = 4\|(\mathbf{A}-\mathbf{P})\mathbf{V}\mathbf{V}^T\|_F^2$$

$$\begin{aligned}
\sum_{k=1}^{n}\sum_{l=1}^{n}c_{kl}^2 &\le 3\sum_{k=1}^{n}\sum_{l=1}^{n}4(\mathbf{A}-\mathbf{P})_{kl}^2(\langle V_l, V_l\rangle^2 + \langle V_k, V_k\rangle^2) + 3\sum_{k=1}^{n}\sum_{l=1}^{n}\langle V_l, V_k\rangle^2 \\
&= 6\sum_{k=1}^{n}4((\mathbf{A}-\mathbf{P})^2)_{kk}\langle V_k, V_k\rangle^2 + 3\sum_{k=1}^{n}\sum_{l=1}^{n}\langle V_l, V_k\rangle^2 \\
&= 24\sum_{k=1}^{n}((\mathbf{A}-\mathbf{P})^2)_{kk}\langle V_k, V_k\rangle^2 + 3\sum_{k=1}^{n}(\mathbf{V}\mathbf{V}^T\mathbf{V}\mathbf{V}^T)_{kk} \\
&= 24\sum_{k=1}^{n}((\mathbf{A}-\mathbf{P})^2)_{kk}\langle V_k, V_k\rangle^2 + 3\|\mathbf{V}\mathbf{V}^T\|_F^2 \\
&\le 24\|(\mathbf{A}-\mathbf{P})^2\|\sum_{k=1}^{n}\langle V_k, V_k\rangle^2 + 3\|\mathbf{V}\mathbf{V}^T\|_F^2 \\
&\le 24\|\mathbf{A}-\mathbf{P}\|^2\|\mathrm{diag}(\mathbf{V}\mathbf{V}^T)\|_F^2 + 3\|\mathbf{V}\mathbf{V}^T\|_F^2
\end{aligned}$$

where the penultimate inequality of the above display follows from the fact that the diagonal elements of $(\mathbf{A}-\mathbf{P})^2$ is majorized by its eigenvalues. We therefore have

$$\begin{aligned}
\sum_{k<l}(Z - Z_{kl})^2 &\le \big(48\|\mathbf{A}-\mathbf{P}\|^2 + \tfrac{9}{2}\big)\|\mathbf{V}\mathbf{V}^T\|_F^2 \\
&\le 49\|\mathbf{A}-\mathbf{P}\|^2\|\mathbf{V}\mathbf{V}^T\|_F^2 \\
&= 49\|\mathbf{A}-\mathbf{P}\|^2\|\mathbf{S}_{\mathbf{P}}^{-1}\|_F^2 \\
&\le 49\|\mathbf{A}-\mathbf{P}\|^2\frac{d}{(\gamma_2(\mathbf{P})\delta(\mathbf{P}))^2}
\end{aligned}$$

By Proposition A.1, for any $\eta > 0$, with probability at least $1-\eta$,

$$\|\mathbf{A}-\mathbf{P}\|^2 \le 4\delta(\mathbf{P})\log(n/\eta)$$

Hence, for all $\eta > 0$, with probability at least $1 - \eta$,

$$\sum_{k<l} (Z - Z_{kl})^2 \leq \frac{196 d \log{(n/\eta)}}{\gamma_2^2(\mathbf{P}) \delta(\mathbf{P})}. \tag{A.9}$$

Denote by $v(\eta)$ the right hand side of the above display. We then have, by Theorem A.7, that for all $t > 0$,

$$\mathbb{P}[|Z - \mathbb{E}[Z]| > t] \leq 2e^{-t^2/(2v(\eta))} + \eta \tag{A.10}$$

Setting t to be

$$t = \frac{14\sqrt{2d} \log{(n/\eta)}}{\gamma_2(\mathbf{P})\sqrt{\delta(\mathbf{P})}}$$

yields $2e^{-t^2/(2v(\eta))} \leq \eta$ as desired.

Finally, we provide a bound for $\mathbb{E}[Z]$ in terms of the parameters $\gamma_2(\mathbf{P})$ and $\delta(\mathbf{P})$. We have

$$\mathbb{E}[Z] = \mathbb{E}[\|(\mathbf{A} - \mathbf{P})\mathbf{V}\|_F^2] = \mathbb{E}[\mathrm{tr}\Big(\mathbf{V}^T(\mathbf{A} - \mathbf{P})^2\mathbf{V}\Big)] = \mathrm{tr}\Big(\mathbf{V}^T\mathbb{E}[(\mathbf{A} - \mathbf{P})^2]\mathbf{V}\Big)$$

We note that

$$\mathbb{E}[((\mathbf{A} - \mathbf{P})^2)_{ij}] = \mathbb{E}\Big[\sum_k (\mathbf{A} - \mathbf{P})_{ik}(\mathbf{A} - \mathbf{P})_{kj}\Big] = \begin{cases} 0 & \text{if } i \neq j \\ \sum_{k \neq i} \mathbf{P}_{ik}(1 - \mathbf{P})_{ik} & \text{if } i = j \end{cases}$$

Hence, $\delta(\mathbf{P})\mathbf{I} - \mathbb{E}[(\mathbf{A} - \mathbf{P})^2]$ is positive semidefinite. We thus have

$$\mathbb{E}[\|(\mathbf{A} - \mathbf{P})\mathbf{V}\mathbf{V}^T\|_F^2] \leq \delta \mathrm{tr} \mathbf{V}^T\mathbf{V} \leq d\gamma_2^{-1}(\mathbf{P}).$$

which establishes the upper bound $C^2(\mathbf{X}) \leq d\gamma_2^{-1}(\mathbf{P})$ as required. □

Proof of Theorem 2.3. From Lemma A.6, we have

$$\Big|\|(\mathbf{A} - \mathbf{P})\mathbf{U}_{\mathbf{P}}\mathbf{S}_{\mathbf{P}}^{-1/2}\|_F^2 - C^2(\mathbf{X})\Big| \leq \frac{14\sqrt{2d}\log{(n/\eta)}}{\gamma_2(\mathbf{P})\sqrt{\delta(\mathbf{P})}}$$

with probability at least $1 - 2\eta$. Now, $a \leq b + c$ implies $\sqrt{a} \leq \sqrt{b} + \frac{c}{2\sqrt{b}}$ and $a \geq b - c \geq 0$ implies $\sqrt{a} \geq \sqrt{b} - \frac{c}{\sqrt{b}}$. Hence

$$-\frac{14\sqrt{2d}\log{(n/\eta)}}{C(\mathbf{X})\gamma_2(\mathbf{P})\sqrt{\delta(\mathbf{P})}} \leq \|(\mathbf{A} - \mathbf{P})\mathbf{U}_{\mathbf{P}}\mathbf{S}_{\mathbf{P}}^{-1/2}\|_F - C(\mathbf{X}) \leq \frac{7\sqrt{2d}\log{(n/\eta)}}{C(\mathbf{X})\gamma_2(\mathbf{P})\sqrt{\delta(\mathbf{P})}}$$

with probability at least $1 - 2\eta$. Applying Lemma A.4 and Lemma A.5 yield

$$-\frac{C_1 d^{3/2}\log{(n/\eta)}}{C(\mathbf{X})\sqrt{\gamma_1^7(\mathbf{P})\delta(\mathbf{P})}} \leq \|\hat{\mathbf{X}} - \mathbf{P}\mathbf{U}_{\mathbf{P}}\mathbf{S}_{\mathbf{P}}^{-1/2}\|_F - C(\mathbf{X}) \leq \frac{C_2 d^{3/2}\log{(n/\eta)}}{C(\mathbf{X})\sqrt{\gamma_1^7(\mathbf{P})\delta(\mathbf{P})}}$$

for some constants $C_1, C_2 > 0$. As $\mathbf{P}\mathbf{U}_{\mathbf{P}}\mathbf{S}_{\mathbf{P}}^{-1/2} = \mathbf{X}\mathbf{W}$ for some $\mathbf{W} \in \mathcal{O}(d)$, Theorem 2.3 follows.

Proof of Theorem 3.2. The proof of this result is almost identical to that of Theorem 3.1. We sketch here the necessary modifications. As before, we suppress dependence on n unless necessary. Let α be given and let $\eta = \alpha/4$. By Theorem 2.3, for n sufficiently large, there exists some orthogonal $\mathbf{W_X} \in \mathcal{O}(d)$ such that, with probability at least $1-\eta$

$$\|\hat{\mathbf{X}}\mathbf{W_X} - \mathbf{X}\|_F \le C(\mathbf{X}) + f(\mathbf{X}, \alpha, n)$$

where for any fixed $\mathbf{X}$ and α, $f(\mathbf{X}, \alpha, n) \to 0$ as $n \to \infty$. From the form of $C(\mathbf{X})$ given in Theorem 2.3, for all n,

$$C(\mathbf{X}_n) \le \sqrt{\frac{d}{\gamma_2(\mathbf{P}_n)}} \tag{A.11}$$

Now, again for n sufficiently large,

$$\begin{aligned}
\|\hat{\mathbf{X}}\mathbf{W_X}/\|\hat{\mathbf{X}}\|_F - \mathbf{X}/\|\mathbf{X}\|_F\|_F &\le \frac{\|\hat{\mathbf{X}}\mathbf{W_X} - \mathbf{X}\|_F}{\|\hat{\mathbf{X}}\|_F} + \|\mathbf{X}\|_F \Big| \frac{1}{\|\hat{\mathbf{X}}\|_F} - \frac{1}{\|\mathbf{X}\|_F} \Big| \\
&\le \frac{\|\hat{\mathbf{X}}\mathbf{W_X} - \mathbf{X}\|_F}{\|\hat{\mathbf{X}}\|_F} + \frac{|\|\hat{\mathbf{X}}\|_F - \|\mathbf{X}\|_F|}{\|\hat{\mathbf{X}}\|_F} \\
&\le \frac{\|\hat{\mathbf{X}}\mathbf{W_X} - \mathbf{X}\|_F}{\|\hat{\mathbf{X}}\|_F} + \frac{|\|\hat{\mathbf{X}}\mathbf{W_X}\|_F - \|\mathbf{X}\|_F|}{\|\hat{\mathbf{X}}\|_F} \\
&\le \frac{2\|\hat{\mathbf{X}}\mathbf{W_X} - \mathbf{X}\|_F}{\|\hat{\mathbf{X}}\|_F} \le \frac{2(C(\mathbf{X}) + f(\mathbf{X}, \alpha, n))}{\|\hat{\mathbf{X}}\|_F}
\end{aligned}$$

with probability at least $1 - \eta$. An analogous bound can also be derived for $\mathbf{Y}$. Under the null hypothesis, $\mathbf{X} \perp c\mathbf{Y}$ for some $c > 0$, so we derive that

$$\min_{\mathbf{W} \in \mathcal{O}(d)} \|\hat{\mathbf{X}}\mathbf{W}/\|\hat{\mathbf{X}}\|_F - \hat{\mathbf{Y}}/\|\hat{\mathbf{Y}}\|_F\|_F \le \frac{2(C(\mathbf{X}) + f(\mathbf{X}, \alpha, n))}{\|\hat{\mathbf{X}}\|_F} + \frac{2(C(\mathbf{Y}) + f(\mathbf{Y}, \alpha, n))}{\|\hat{\mathbf{Y}}\|_F}$$

From Eq.(A.11) and Proposition A.2, we conclude that for n sufficiently large,

$$T_n = \frac{\min_{\mathbf{W} \in \mathcal{O}(d)} \|\hat{\mathbf{X}}\mathbf{W}/\|\hat{\mathbf{X}}\|_F - \hat{\mathbf{Y}}/\|\hat{\mathbf{Y}}\|_F\|_F}{2\sqrt{d\gamma_2^{-1}(\mathbf{A})}/\|\hat{\mathbf{X}}\|_F + 2\sqrt{d\gamma_2^{-1}(\mathbf{B})}/\|\hat{\mathbf{Y}}\|_F} \le 1 + r(\alpha, n)$$

where $r(\alpha, n) \to 0$ as $n \to \infty$ for a fixed α. We can now choose a $n_1 = n_1(\alpha, C)$ for which $r(\alpha, n_1) \le C - 1$. This implies that for all $n \ge n_1$,

$$\mathbb{P}(T_n \in R) \le \alpha$$

which establishes that the test statistic T_n with rejection region R is an at most level-α test. The proof of consistency proceeds in an almost identical manner to that in Theorem 3.1 and we omit the details.

Proof of Theorem 3.3. Our proof relies on a variant of Theorem 2.3, namely the following lemma, which provides a bound on the maximum of the l_2 norm of the rows of $\hat{\mathbf{X}} - \mathbf{X}$. We omit the details of its proof, which is similar to—but simpler than—that of Theorem 2.3; the key step is the use of Hoeffding's inequality to bound the $2 \to \infty$ norm of $(\mathbf{A} - \mathbf{P})\mathbf{U_P}\mathbf{S_P}^{-1/2}$.

LEMMA A.8. *Suppose Assumption 2 holds, and let $c > 0$ be arbitrary. Then there exists a $n_0(c)$ such that for all $n > n_0$ and $n^{-c} < \eta < 1/2$, there exists a deterministic $\mathbf{W} = \mathbf{W}_n \in \mathcal{O}(d)$ such that, with probability at least $1 - 3\eta$,*

$$\|\hat{\mathbf{X}} - \mathbf{X}\mathbf{W}\|_{2\to\infty} = \max_i \|\hat{X}_i - \mathbf{W}X_i\| \le \frac{85 d^{3/2}\log(n/\eta)}{\sqrt{\gamma_1^7(\mathbf{P})\delta(\mathbf{P})}}. \tag{A.12}$$

Continuing with the proof of the theorem, for any $\mathbf{W} \in \mathcal{O}(d)$,

$$\begin{aligned}\|\mathcal{P}(\hat{\mathbf{X}})\mathbf{W} - \mathcal{P}(\mathbf{X})\|_F &= \|\mathcal{D}^{-1}(\hat{\mathbf{X}})\hat{\mathbf{X}}\mathbf{W} - \mathcal{D}^{-1}(\hat{\mathbf{X}})\mathbf{X} + \mathcal{D}^{-1}(\hat{\mathbf{X}})\mathbf{X} - \mathcal{D}^{-1}(\mathbf{X})\mathbf{X}\|_F \\ &\le \|\mathcal{D}^{-1}(\hat{\mathbf{X}})\|_F \|\hat{\mathbf{X}}\mathbf{W} - \mathbf{X}\|_{2\to\infty} + \|(\mathcal{D}^{-1}(\hat{\mathbf{X}}) - \mathcal{D}^{-1}(\mathbf{X}))\mathbf{X}\|_F\end{aligned}$$

The term $\|(\mathcal{D}^{-1}(\hat{\mathbf{X}}) - \mathcal{D}^{-1}(\mathbf{X}))\mathbf{X}\|_F$ can now be written as

$$\begin{aligned}\|(\mathcal{D}^{-1}(\hat{\mathbf{X}}) - \mathcal{D}^{-1}(\mathbf{X}))\mathbf{X}\|_F^2 &= \sum_{i=1}^n \|X_i\|^2 \Big(\frac{1}{\|\hat{X}_i\|} - \frac{1}{\|X_i\|}\Big)^2 \\ &= \sum_{i=1}^n \frac{(\|\mathbf{W}X_i\| - \|\hat{X}_i\|)^2}{\|\hat{X}_i\|^2} \\ &\le \sum_{i=1}^n \frac{\|\mathbf{W}X_i - \hat{X}_i\|^2}{\|\hat{X}_i\|^2} \\ &\le \Big(\|\mathcal{D}^{-1}(\hat{\mathbf{X}})\|_F \|\hat{\mathbf{X}}\mathbf{W} - \mathbf{X}\|_{2\to\infty}\Big)^2\end{aligned}$$

and hence,

$$\|\mathcal{P}(\hat{\mathbf{X}})\mathbf{W} - \mathcal{P}(\mathbf{X})\|_F \le 2\|\hat{\mathbf{X}}\mathbf{W} - \mathbf{X}\|_{2\to\infty}\|\mathcal{D}^{-1}(\hat{\mathbf{X}})\|_F$$

Now, let $\alpha \in (0,1)$ be given and set $\eta = \alpha/4$. Then by Lemma A.8, for n sufficiently large, there exists some orthogonal $\mathbf{W}_{\mathbf{X}}$ such that, with probability at least $1 - \eta$,

$$\|\mathcal{P}(\hat{\mathbf{X}})\mathbf{W}_{\mathbf{X}} - \mathcal{P}(\mathbf{X})\|_F \le \frac{170 d^{3/2}\log(n/\eta)}{\sqrt{\gamma_1^7(\mathbf{P})\delta(\mathbf{P})}}\|\mathcal{D}^{-1}(\hat{\mathbf{X}})\|_F = S(\mathbf{P})\|\mathcal{D}^{-1}(\hat{\mathbf{X}})\|_F. \tag{A.13}$$

An analogous bound holds for $\mathbf{Y}$. Therefore, under the null hypothesis $\mathbf{X} \perp\!\!\!\perp \mathbf{D}\mathbf{Y}$, we have

$$\min_{\mathbf{W}\in\mathcal{O}(d)} \|\mathcal{P}(\hat{\mathbf{X}})\mathbf{W} - \mathcal{P}(\hat{\mathbf{Y}})\|_F \le S(\mathbf{P})\|\mathcal{D}^{-1}(\hat{\mathbf{X}})\|_F + S(\mathbf{Q})\|\mathcal{D}^{-1}(\hat{\mathbf{Y}})\|_F \tag{A.14}$$

and hence,

$$\frac{\min_{\mathbf{W}\in\mathcal{O}(d)} \|\mathcal{P}(\hat{\mathbf{X}})\mathbf{W} - \mathcal{P}(\hat{\mathbf{Y}})\|_F}{S(\mathbf{P})\|\mathcal{D}^{-1}(\hat{\mathbf{X}})\|_F + S(\mathbf{Q})\|\mathcal{D}^{-1}(\hat{\mathbf{Y}})\|_F} \le 1$$

Now, by Proposition A.2, for sufficiently large n, we can replace $S(\mathbf{P})$ and $S(\mathbf{Q})$ with $S(\mathbf{A})$ and $S(\mathbf{B})$ to yield

$$T_n = \frac{\min_{\mathbf{W}\in\mathcal{O}(d)} \|\mathcal{P}(\hat{\mathbf{X}})\mathbf{W} - \mathcal{P}(\hat{\mathbf{Y}})\|_F}{S(\mathbf{A})\|\mathcal{D}^{-1}(\hat{\mathbf{X}})\|_F + S(\mathbf{B})\|\mathcal{D}^{-1}(\hat{\mathbf{Y}})\|_F} \le 1 + r(\alpha, n)$$

where $r(\alpha, n) \to 0$ as $n \to \infty$ for a fixed α. We can therefore choose a $n_1 = n_1(\alpha, C)$ for which $r(\alpha, n_1) \le C - 1$. This implies that for all $n \ge n_1$,

$$\mathbb{P}(T_n \in C) \le \alpha$$

yielding that the test statistic T_n with rejection region R is an at most level-α test.

We now prove consistency of this test procedure. Suppose now that the sequence of latent positions $\{\mathbf{X}_n, \mathbf{Y}_n\}$ is such that $\mathbf{X}_n \not\perp \mathbf{D}_n \mathbf{Y}_n$ for any diagonal $\mathbf{D}_n \in \mathcal{E}$. Denote by $h(\hat{\mathbf{X}}, \hat{\mathbf{Y}})$ the ratio

$$h(\hat{\mathbf{X}}, \hat{\mathbf{Y}}) = \frac{S(\mathbf{P})\|\mathcal{D}^{-1}(\mathbf{X})\|_F + S(\mathbf{Q})\|\mathcal{D}^{-1}(\mathbf{Y})\|_F}{S(\mathbf{A})\|\mathcal{D}^{-1}(\hat{\mathbf{X}})\|_F + S(\mathbf{B})\|\mathcal{D}^{-1}(\hat{\mathbf{Y}})\|_F}$$

and by $f(\hat{\mathbf{X}}, \hat{\mathbf{Y}})$ the ratio

$$f(\hat{\mathbf{X}}, \hat{\mathbf{Y}}) = \frac{\|\mathcal{P}(\hat{\mathbf{X}})\mathbf{W}_{\mathbf{X}} - \mathcal{P}(\mathbf{X})\|_F + \|\mathcal{P}(\hat{\mathbf{Y}})\mathbf{W}_{\mathbf{Y}} - \mathcal{P}(\mathbf{Y})\|_F}{S(\mathbf{A})\|\mathcal{D}^{-1}(\hat{\mathbf{X}})\|_F + S(\mathbf{B})\|\mathcal{D}^{-1}(\hat{\mathbf{Y}})\|_F}$$

Then, for all n,

$$\begin{aligned}
\mathbb{P}(T_n \notin R_n) &\leq \mathbb{P}\Big(\frac{\min_{W \in \mathcal{O}(d)} \|\mathcal{P}(\mathbf{X})\mathbf{W} - \mathcal{P}(\mathbf{Y})\|_F}{S(\mathbf{A})\|\mathcal{D}^{-1}(\hat{\mathbf{X}})\|_F + S(\mathbf{B})\|\mathcal{D}^{-1}(\hat{\mathbf{Y}})\|_F} \leq C + f(\hat{\mathbf{X}}, \hat{\mathbf{Y}})\Big) \\
&\leq \mathbb{P}\Big(h(\hat{\mathbf{X}}, \hat{\mathbf{Y}})(C + f(\hat{\mathbf{X}}, \hat{\mathbf{Y}})) \geq D_{\mathcal{P}}(\mathbf{X}, \mathbf{Y})\Big)
\end{aligned}$$

Now, for a given $\beta > 0$, let $M_1 = M_1(\beta)$ and $n_0 = n_0(\alpha, \beta)$ be such that, for all $n \geq n_0(\alpha, \beta)$,

$$\mathbb{P}(C + f(\hat{\mathbf{X}}, \hat{\mathbf{Y}}) \geq M_1) \leq \beta/2.$$

By Eq. (A.13) and Proposition A.2, $M_1(\beta)$ and $n_0(\alpha, \beta)$ exists for all choice of β.

We now show that there exists, for any $\beta > 0$, some $n_1 = n_1(\beta)$ such that, for all $n \geq n_1(\beta)$,

$$\mathbb{P}(h(\hat{\mathbf{X}}, \hat{\mathbf{Y}}) \geq 4) \leq \beta/2. \tag{A.15}$$

Indeed,

$$h(\hat{\mathbf{X}}, \hat{\mathbf{Y}}) \leq \max\Big\{\frac{S(\mathbf{P})\|\mathcal{D}^{-1}(\mathbf{X})\|_F}{S(\mathbf{A})\|\mathcal{D}^{-1}(\hat{\mathbf{X}})\|_F}, \frac{S(\mathbf{Q})\|\mathcal{D}^{-1}(\mathbf{Y})\|_F}{S(\mathbf{B})\|\mathcal{D}^{-1}(\hat{\mathbf{Y}})\|_F}\Big\}$$

In addition, we have

$$\frac{\|\mathcal{D}^{-1}(\mathbf{X})\|_F}{\|\mathcal{D}^{-1}(\hat{\mathbf{X}})\|_F} = \frac{\sqrt{\sum_{i=1}^n 1/\|X_i\|^2}}{\sqrt{\sum_{i=1}^n 1/\|\hat{X}_i\|^2}} \leq \sqrt{\max_i \frac{\|\hat{X}_i\|^2}{\|X_i\|^2}} = \max_i \frac{\|\hat{X}_i\|}{\|X_i\|} \leq 1 + \frac{\|\hat{\mathbf{X}} - \mathbf{X}\|_{2\to\infty}}{\min_i \|X_i\|}$$

By Lemma A.8 and the condition in Assumption 2 on $\min_i \|X_i\|$, there exist some $n_1(\beta)$ such that for all $n \geq n_1(\beta)$,

$$\mathbb{P}\Big(\frac{\|\hat{\mathbf{X}} - \mathbf{X}\|_{2\to\infty}}{\min_i \|X_i\|} \geq 1\Big) \leq \beta/8.$$

We now apply Proposition A.2 to $S(\mathbf{P})/S(\mathbf{A})$ to conclude that there exist some $n_1(\beta)$ such that for all $n \geq n_1(\beta)$,

$$\mathbb{P}\Big(\frac{S(\mathbf{P})\|\mathcal{D}^{-1}(\mathbf{X})\|_F}{S(\mathbf{A})\|\mathcal{D}^{-1}(\hat{\mathbf{X}})\|_F} \geq 4\Big) \leq \beta/4.$$

The same argument can be applied to the ratio depending on $\hat{\mathbf{Y}}$ and $\mathbf{Y}$. This establishes that there exists some $n_1(\beta)$ such that Eq. (A.15) holds for all $n \geq n_1(\beta)$. Since $D_{\mathcal{P}}(\mathbf{X}_n, \mathbf{Y}_n) \to \infty$, there exists some $n_2 = n_2(\alpha, \beta, C)$ such that for all $n \geq n_2$, $D_{\mathcal{P}}(\mathbf{X}_n, \mathbf{Y}_n) \geq 4M_1(\beta)$. Hence for all $n \geq n_2$,

$\mathbb{P}(T_n \notin R) \leq \beta$, i.e., the test statistic lies within the rejection region with probability at least $1 - \beta$, as required.

Department of Applied Mathematics and Statistics,
Johns Hopkins University,
3400 N. Charles St, Baltimore, MD 21218, USA.
E-mail: mtang10@jhu.edu; dathrey1@jhu.edu; vlyzins1@jhu.edu; cep@jhu.edu

Department of Statistics,
Harvard University,
One Oxford Street, Cambridge, MA, 02138, USA
E-mail: daniellsussman@fas.harvard.edu