A Bootstrap-based Method for Testing Similarity of Matched Networks

Somnath Bhadra
Department of Statistics, University of Florida

Kaustav Chakraborty
Department of Statistics, University of Illinois at Urbana-Champaign

Srijan Sengupta
Department of Statistics, North Carolina State University

Soumendra N. Lahiri
Department of Mathematics and Statistics, Washington University in St. Louis

February 6, 2025

Abstract

This paper studies the matched network inference problem, where the goal is to determine if two networks, defined on a common set of nodes, exhibit a specific form of stochastic similarity. Two notions of similarity are considered: (i) equality, i.e., testing whether the networks arise from the same random graph model, and (ii) scaling, i.e., testing whether their probability matrices are proportional for some unknown scaling constant. We develop a testing framework based on a parametric bootstrap approach and a Frobenius norm-based test statistic. The proposed approach is highly versatile as it covers both the equality and scaling problems, and ensures adaptability under various model settings, including stochastic blockmodels, Chung-Lu models, and random dot product graph models. We establish theoretical consistency of the proposed tests and demonstrate their empirical performance through extensive simulations under a wide range of model classes. Our results establish the flexibility and computational efficiency of the proposed method compared to existing approaches. We also report a real-world application involving the Aarhus network dataset, which reveals meaningful sociological patterns across different communication layers.

Keywords: Two-sample Testing, Multilayer Networks, Parametric Bootstrap, Stochastic Block Model, Random Dot Product Graph
1 Introduction

Consider $n$ entities labeled $1, \ldots, n$, and two undirected networks (without multiple edges or self-loops) represented by $n \times n$ symmetric adjacency matrices $A_1$ and $A_2$. Here, $A_1(i, j) = 1$ if entities $i$ and $j$ interact in the first network, and $A_1(i, j) = 0$ otherwise; similarly for $A_2$. The statistical model assumes $A_1 \sim P_1$ and $A_2 \sim P_2$, where $A_k(i, j) \sim \text{Bernoulli}(P_k(i, j))$ for $1 \leq i < j \leq n$, $k = 1, 2$. This setup aligns with multilayer or multiplex networks studied in the literature (Paul and Chen, 2016; MacDonald et al., 2022; Wang et al., 2023).

Let $\tau(P)$ denote some network feature, i.e., a function of the probability matrix, which represents the model parameter of interest. The inference task of interest is to test whether the networks are similar in terms of $\tau(\cdot)$, i.e., whether $\tau(P_1) = \tau(P_2)$. We call this problem the matched network inference problem to emphasize that the nodes in the two networks are matched, meaning the $i^{th}$ rows of $A_1$ and $A_2$ correspond to the same entity.

This framework encompasses various problems depending on $\tau$. In this paper, we concern ourselves with two inference problems: equality and scaling. Under the equality problem, we want to test whether $P_1 = P_2$, akin to the classical paired sample testing problem. This problem arises in a wide range of scientific problems, such as functional neuroimaging (Ginestet et al., 2017), anatomical brain structures (Bassett et al., 2008), and gene regulatory networks (Zhang et al., 2009).

The scaling problem generalizes the equality problem by testing whether $P_1 = cP_2$ for some (unknown) $c > 0$, with $c = 1$ corresponding to equality. For instance, Gemmetto and Garlaschelli (2017) observed scaling relationships in global trading networks due to shared production or consumption patterns. Scaled relationships between layers are also common in sociology (Nicosia and Latora, 2015). In this paper, we analyzed the Aarhus network of five interaction types (coauthor, leisure, work, lunch, Facebook) among 61
researchers (Rossi and Magnani, 2015). While the test of equality was rejected for all pairs of interactions, the test of scaling was not rejected for a majority of the pairs of interactions. This suggests a proportional relationship that might reflect differences in ease of access, cost, or user preferences for specific modes of interaction. Such findings can lead to deeper sociological insights into communication habits and technological adoption. Note that equality corresponds to $\tau(P) = P$, and scaling to $\tau(P) = P/||P||$, where $|| \cdot ||$ is a matrix norm (e.g., Frobenius norm).

This paper introduces a versatile framework for matched network inference based on a Frobenius norm test statistic. The central idea is to carry out a null-restricted transformation of the estimated models to generate parametric bootstrap resamples, to estimate the sampling distribution of the test statistic under the null. The proposed methodology is highly flexible as it accommodates both equality and scaling problems while remaining applicable across a wide array of network models and estimators. We establish theoretical results under three widely-used models — the stochastic blockmodel, the Chung-Lu model, and the random dot product graph model — that emphasize that the proposed method is consistent under general settings. We also present numerical results under six network models — the three aforementioned models plus the degree-corrected blockmodel, the popularity adjusted blockmodel, and the latent space model — that demonstrate the versatility of the proposed method in practice.

While matched network inference is a relatively new area, some exciting progress has been made in recent years (Tang et al., 2017a,b; Ghoshdastidar et al., 2020, 2017; Ghoshdastidar and von Luxburg, 2018; Li and Li, 2018; Levin et al., 2017; Agterberg et al., 2020; Jin et al., 2024; Wu and Hu, 2024). In Tang et al. (2017a), Tang et al. (2017b), Levin et al. (2017), and Agterberg et al. (2020), the authors studied the problem under the random dot
product graph model and its generalizations. In Li and Li (2018), the authors studied the problem under the stochastic blockmodel. In Ghoshdastidar and von Luxburg (2018), the authors proposed a test based on the spectral norm, and in Ghoshdastidar et al. (2020) and Ghoshdastidar et al. (2017), the authors studied the problem from an information theoretic perspective to derive minimax bounds. The methods most closely related to our work are Tang et al. (2017a), Levin et al. (2017), and Ghoshdastidar and von Luxburg (2018), and we provide a comparative analysis of these methods with our approach.

In related work, recent years have seen substantial growth in theoretical results for bootstrapping network data, with notable contributions from Bhattacharyya and Bickel (2015), Green and Shalizi (2022), Lunde and Sarkar (2022), and Levin and Levina (2019). However, most of the existing work is for non-parametric bootstrap - a notable recent exception being Shao and Le (2024) - while our methodology is based on parametric bootstrap.

The rest of the paper is organized as follows. Section 2 describes the methodology and Section 3 contains theoretical results. Section 4 describes simulation studies with comparisons to existing methods. A case study on the Aarhus network in Section 5 illustrates the practical usefulness of the proposed framework, and Section 6 concludes the paper with discussions and next steps. Open-source R code is publicly available on GitHub.

2 Problem statement and methodology

We begin with a heuristic outline of the proposed method for a generic similarity function, \( \tau \). Then we describe detailed methodology for the two specific versions of \( \tau \) corresponding to the equality and scaling problems in subsections 2.1 and 2.2.

Let \( A_1 \sim P_1 \) and \( A_2 \sim P_2 \), where \( P_1, P_2 \in \mathcal{P} \) and \( \mathcal{P} \) is a known class of network models (e.g., the class of stochastic blockmodels). We want to test \( H_0 : \tau(P_1) = \tau(P_2) \) vs. \( H_1 : \tau(P_1) \neq \tau(P_2) \).
\(\tau(P_1) \neq \tau(P_2)\). Suppose that the model class has a consistent estimator, and we can generate estimators \(\hat{P}_k\) and \(\tau(\hat{P}_k)\) from \(A_k\) for \(k = 1, 2\). Then, a natural test statistic is given by

\[
T(A_1, A_2) = ||\tau(\hat{P}_1) - \tau(\hat{P}_2)||_F,
\]

where \(||M||_F = \sqrt{\sum_i \sum_j |M(i, j)|^2}\) denotes the Frobenius norm of a matrix; if \(\tau\) is a vector instead of a matrix, we use the \(l_2\) norm instead. If the null hypothesis is true, then it is reasonable to expect that \(T(A_1, A_2)\) is small. On the other hand, when \(H_1\) is true and \(||\tau(P_1) - \tau(P_2)||_F\) is large, the test statistic is likely to be larger. Therefore, we should reject when \(T(A_1, A_2)\) is greater than some threshold value. One way to determine the rejection threshold would be to analytically formulate the (asymptotic) sampling distribution of \(T(A_1, A_2)\) under \(H_0\), and compute the upper \((1 - \alpha)\) quantile, where \(\alpha\) is the target level of the test. In Tang et al. (2017a) and Ghoshdastidar and von Luxburg (2018), the authors pursued this approach under their specific model settings to derive rejection thresholds for their test statistics. However, this approach can lead to complications for practitioners, as they must know the asymptotic distribution in order to implement the test.

We propose a parametric bootstrap strategy to estimate the sampling distribution of \(T(A_1, A_2)\) under the null. Bootstrapping is a versatile and well-studied technique for estimating the sampling distribution of a random variable by drawing resamples as a proxy for samples and constructing an empirical distribution (Efron, 1979; Singh, 1981; Beran and Ducharme, 1991; Davison and Hinkley, 1997; Efron and Tibshirani, 1994; Hall, 1993; Shao and Tu, 1995; Lahiri, 2003; Chatterjee and Lahiri, 2011; Hall and Horowitz, 2013; Sengupta et al., 2015). This approach is conceptually appealing as it is simple to implement, and is automatic in nature, such that practitioners can apply it without advanced statistical knowledge. To estimate the sampling distribution of the test statistic under \(H_0\), we first
need to transform $\hat{P}_1$ and $\hat{P}_2$ to their null-restricted counterparts, $\hat{P}_0^1$ and $\hat{P}_0^2$, which satisfy

$$\tau(\hat{P}_0^1) = \tau(\hat{P}_0^2), \quad (1)$$

such that $\hat{P}_0^1$ and $\hat{P}_0^2$ are accurate estimates of $P_1$ and $P_2$ under $H_0$ but not under $H_1$. Next, we generate $B$ matched network pairs $A_{1}^i \sim \hat{P}_0^1, A_{2}^i \sim \hat{P}_0^2$ and compute $T^* = T(A_{1}^i, A_{2}^i)$ for $i = 1, \ldots, B$. The p-value is obtained by comparing the observed value of the test statistic, $T(A_1, A_2)$, with the empirical distribution of $\{T^*_i\}_{i=1}^B$, i.e. $p = \frac{1}{B} \sum_{i=1}^B I[T \leq T^*_i]$, where $I$ is the indicator function, and we reject if $p < \alpha$.

The key challenge is how to transform the estimates $\hat{P}_1$ and $\hat{P}_2$ to their null-restricted counterparts that satisfy (1). There is no universal technique for accomplishing this for a generic $\tau$. However, this transformation can be easily accomplished for the equality and scaling problems, as described in the next two subsections.

### 2.1 Test of equality

Here $\tau(P) = P$ and we want to test $H_0 : P_1 = P_2$ vs. $H_1 : P_1 \neq P_2$. The test statistic is

$$T_{\text{frob}}(A_1, A_2) = \|\hat{P}_1 - \hat{P}_2\|_F. \quad (2)$$

We can transform $\hat{P}_1$ and $\hat{P}_2$ to their null-restricted counterparts by using the pooled estimator, i.e.,

$$\hat{P}_0^1 = \hat{P}_0^2 = \frac{1}{2}(\hat{P}_1 + \hat{P}_2). \quad (3)$$

Clearly this satisfies (1). Crucially, $\hat{P}_0^k$ is a good estimator of $P_k$ for $k = 1, 2$ under the null but it is biased under the alternative. Therefore, resampling from $\hat{P}_0^1$ and $\hat{P}_0^2$ ensures that the bootstrapped distribution mimics the sampling distribution of the test statistic under the null, but not under the alternative. The steps are outlined in Algorithm 1.
2.2 Test of scaling

Here $\tau(P) = \frac{P}{||P||_F}$, and the test statistic is given by

$$T_{\text{scale}}(A_1, A_2) = ||\frac{\hat{P}_1}{\hat{\rho}_1} - \frac{\hat{P}_2}{\hat{\rho}_2}||_F,$$

where $\hat{\rho}_1 = ||\frac{\hat{P}_1}{\hat{\rho}_1}||_F$ and $\hat{\rho}_2 = ||\frac{\hat{P}_2}{\hat{\rho}_2}||_F$. Under the null, there could be a scaling difference between $P_1$ and $P_2$, which means we cannot use a simple pooled estimator to transform $\hat{P}_1$ and $\hat{P}_2$ to their null-restricted counterparts. However, under the null, the scaled probability matrices are equal, i.e., $\frac{P_1}{||P_1||_F} = \frac{P_2}{||P_2||_F}$, and let us denote this as $H = \frac{P_1}{||P_1||_F} = \frac{P_2}{||P_2||_F}$. Then, under the null, $A_1 \sim ||P_1||_F H$ and $A_2 \sim ||P_2||_F H$. Therefore, we can use the pooled version of the scaled estimates to estimate $H$, and rescale the pooled version to obtain the null-restricted counterparts $\hat{P}_1^{(0)}$ and $\hat{P}_2^{(0)}$.

That is, we first compute $\hat{H} = \frac{1}{2} \left( \frac{\hat{P}_1}{\hat{\rho}_1} + \frac{\hat{P}_2}{\hat{\rho}_2} \right)$, and then we rescale it as

$$\hat{P}_1^{(0)} = \hat{\rho}_1 \hat{H}, \quad \hat{P}_2^{(0)} = \hat{\rho}_2 \hat{H}.$$ 

As before, $\hat{P}_1^{(0)}$ and $\hat{P}_2^{(0)}$ now satisfy (1). Furthermore, $\hat{P}_k^{(0)}$ is a good estimator of $P_k$ for $k = 1, 2$ under the null but it is biased under the alternative, which ensures that the resampling distribution approximates the sampling distribution under the null but not under the alternative, as intended. The steps are outlined in Algorithm 1.

**Remark 2.1. Sparse Graph Case:** We now consider the scenario where the networks are sparse. In this case, both $P_1$ and $P_2$ asymptotically converge to a matrix of zeroes, which trivializes the null hypothesis $H_0 : P_1 = P_2$ for the test of equality. Therefore, the case for sparse networks needs to be treated separately. To define sparsity, we represent the probability matrices $P_1$ and $P_2$ as $P_k = \rho_{kn} \tilde{P}_k$ for $k = 1, 2$, where $\rho_{kn}$ are the sparsity factors of the networks, and the entries of $\tilde{P}_k$ are $O(1)$. Without loss of generality, assume that $\rho_{1n}/\rho_{2n}$ does not diverge to infinity. In practice, the sparsity factors are not known a priori,
and the objective is to compare the inherent probability matrices $\tilde{P}_1$ and $\tilde{P}_2$, rather than the sparsity factors themselves. Therefore, we treat the sparsity factors as unknown sequences and test whether $\tilde{P}_2 = \frac{\rho_1}{n} \tilde{P}_1$. Equivalently, this is expressed as testing $H_0: \frac{\|\tilde{P}_1\|_F}{\|P_1\|_F} = \frac{\|\tilde{P}_2\|_F}{\|P_2\|_F}$, which translates to testing the scaled hypothesis $H_0: \frac{\|\tilde{P}_1\|_F}{\|P_1\|_F} = \frac{\|\tilde{P}_2\|_F}{\|P_2\|_F}$.

Input: Data $A_1, A_2$; Network feature $\tau(\cdot)$; Number of bootstraps $B$

Output: p-value for the test of similarity

1: Compute $\tilde{P}_1$ from $A_1$ and $\tilde{P}_2$ from $A_2$
2: Compute the test statistic $T = T(A_1, A_2) = ||\tau(\tilde{P}_1) - \tau(\tilde{P}_2)||_F$ using (2) or (4)
3: Transform $\tilde{P}_1$ and $\tilde{P}_2$ to their “null-restricted” counterparts, $\tilde{P}_1^0$ and $\tilde{P}_2^0$, using (3) or (5).
4: Parametric bootstrap: for $i \leftarrow 1$ to $B$ do
   (a) Generate $A_1^{*i} \sim \tilde{P}_1^{(0)}$, $A_2^{*i} \sim \tilde{P}_2^{(0)}$
   (b) Compute $\tilde{P}_1^{*i}$ from $A_1^{*i}$ and $\tilde{P}_2^{*i}$ from $A_2^{*i}$
   (c) Compute $T^{*i} \leftarrow T(A_1^{*i}, A_2^{*i}) = ||\tau(\tilde{P}_1^{*i}) - \tau(\tilde{P}_2^{*i})||_F$
end
5: The p-value is $p \leftarrow \frac{1}{B} \sum_{i=1}^{B} I[T \leq T^{*i}]$, where $I$ is the indicator function.

Algorithm 1: Test of equality and scaling for matched networks

2.3 Models and estimators

So far, we have used a generic notation, $P$, for the class of models. We now describe six well-known model classes and the corresponding estimators. When implementing the proposed tests, these estimators should be used for steps 1 and 4(b) of Algorithm 1.

- **Stochastic Blockmodel (SBM):** Under an SBM with $K$ communities, $P(i, j) = \omega_{c_i c_j}$, where $\omega$ is a $K$-by-$K$ symmetric matrix of community interaction probabilities, and
$c = \{c_i\}_{i=1}^n$ are the communities of the nodes, with $c_i$ taking its value in $1, \ldots, K$ (Holland et al., 1983). A number of community detection methods (Rohe et al., 2011; Sengupta and Chen, 2015; Zhao et al., 2012; Gao et al., 2017) can be used for estimating the communities $\{\hat{c}_i\}_{i=1}^n$. Once we have $\{\hat{c}_i\}_{i=1}^n$, the model parameters $\omega_{rs}$ are estimated as

$$
\hat{\omega}_{rs} = \frac{\sum_{i,j \mid c_i=c_j=r} A(i,j)}{\hat{n}_r(\hat{n}_r-1)} \quad \text{when} \quad r = s,
$$

and

$$
\hat{\omega}_{rs} = \frac{\sum_{i,j \mid c_i=c_j=r} A(i,j)}{\hat{n}_r \hat{n}_s} \quad \text{when} \quad r \neq s.
$$

Here $\hat{n}_r$ is the size of the estimated $r^{th}$ community. We estimate $P$ as $\hat{P}(i,j) = \hat{c}_i \hat{c}_j$.

- **Chung-Lu (CL) model:** Here $P(i,j) = \theta_i \theta_j$, where $\{\theta_i\}_{i=1}^n$ are the degree parameters (Chung and Lu, 2002; Dasgupta and Sengupta, 2022). We estimate $P$ as $\hat{P}(i,j) = \frac{d_i d_j}{2m}$, where $d_i$ is the degree of the $i^{th}$ node and $m = \sum_{i>j} A(i,j)$.

- **Random dot product graph (RDPG) model:** Under the RDPG (Young and Scheinerman, 2007) with dimension $d$, we have $P = XX'$, where $X_{n \times d}$ is a matrix of rank $d$ such that $[XX'](i,j) \in (0,1)$ for all pairs $(i,j)$. For estimation, we use the adjacency spectral embedding (ASE) of $A$, given by $\hat{X} = U_A S_A^{1/2}$, where $S_A$ is the diagonal matrix of the $d$ largest eigenvalues of $(A'A)^{1/2}$ and $U_A$ is the $n$-by-$d$ matrix whose columns consist of the corresponding eigenvectors (Sussman et al., 2012). We estimate $P$ as $\hat{P} = \hat{X} \hat{X}'$.

- **Latent Space Model (LSM):** Here each node is assumed to have a latent position in $\mathbb{R}^d$ and edge probabilities are determined by pairwise $l_2$ distances between the latent positions, i.e., $\logit(P(i,j)) = \alpha - |z_i - z_j|$, for $1 \leq i < j \leq n$, where $z_i \in \mathbb{R}^d$ is the latent position of the $i^{th}$ and the parameter $\alpha$ controls overall sparsity (Hoff et al., 2002). We estimate $P$ by using the maximum likelihood estimation strategy described in Hoff et al. (2002) as implemented in the R package latentnet (Krivitsky and Handcock, 2008). Note that we consider the null hypothesis to be conditional on latent positions, which is equivalent to the “semiparametric” framework considered in Tang et al. (2017a) but different from the “nonparametric” framework considered in Tang et al. (2017b).
• **Degree Corrected Blockmodel (DCBM):** Under the DCBM with $K$ communities, $P(i,j) = \theta_i \omega_{c_i} \omega_{c_j} \theta_j$, where $\omega$ is a $K$-by-$K$ symmetric matrix of community-community interaction probabilities, and $\{c_i\}_{i=1}^n$ are the communities of the nodes, and $\{\theta_i\}_{i=1}^n$ are degree parameters (Karrer and Newman, 2011). Several community detection methods (Qin and Rohe, 2013; Sengupta and Chen, 2015; Zhao et al., 2012; Gao et al., 2018) can be used for estimating the communities $\{\hat{c}_i\}_{i=1}^n$. The remaining parameters are estimated as $\hat{\omega}_{rs} = \sum_{i,j:\hat{c}_i = r, \hat{c}_j = s} A(i,j)$, and $\hat{\theta}_i = \frac{d_i}{\delta_r}$, where $\delta_r = \sum_{i:\hat{c}_i = r} d_i$ is the degree of the estimated $r^{th}$ community. We estimate $P$ as $\hat{P}(i,j) = \hat{\theta}_i \hat{\omega}_{\hat{c}_i \hat{c}_j} \hat{\theta}_j$.

• **Popularity Adjusted Blockmodel (PABM):** Under the PABM with $K$ communities, $P(i,j) = \theta_{ic} \theta_{jc}$, where $\theta_{ir}$ represents the popularity of the $i^{th}$ node in the $r^{th}$ community, and $\{c_i\}_{i=1}^n$ are the node communities (Sengupta and Chen, 2018). We use the extreme points method of Le et al. (2016) to estimate communities, and estimate the popularity parameters as $\hat{\theta}_{ir} = \frac{\sum_{j:\hat{c}_j = r} A(i,j)}{\sqrt{\sum_{i,j:\hat{c}_i = r, \hat{c}_j = s} A(i,j)}}$. We estimate $P$ as $\hat{P}(i,j) = \hat{\theta}_{ic} \hat{\theta}_{jc}$.

**Remark 2.2.** We omitted the classical Erdős-Rényi model (Erdős and Rényi, 1959), where $P(i,j) = p$ for all pairs $(i,j)$, because the test of equality reduces to the well-known two-sample test of equality of proportions, and the test of scaling is not meaningful since any two Erdős-Rényi models are scaled versions of each other.

## 3 Theoretical results

We now demonstrate the consistency of the proposed methods under three models described in Section 2.3. Similar theoretical results can be proved under some of the other models, but we postpone this to future research. We present the results in two subsections: distributional consistency for the bootstrap methods under the SBM, and testing consistency
under the CL and RDPG models. In both subsections, we analyze the equality and scaling problems under the specified models. All technical proofs are in the supplementary file.

3.1 Bootstrap distributional consistency under the SBM

To prove distributional consistency of the bootstrap, we first derive the limiting null distributions of the relevant test statistics and then use these to show that the proposed bootstrap procedures provide valid approximations in both testing problems. Recall the definitions from Section 2.3. In particular, the vectors $c_{jn}$ and $\hat{c}_{jn}$ are the true and estimated communities (after necessary permutation of the community labels) for the $j$th network, $j = 1, 2$ and $n_r$ and $\hat{n}_r$ are the sizes of the true and estimated $r^{th}$ community for $r = 1, \ldots, K$, where we assume that the number of communities $K$ is known. Note that under the matched network framework null hypotheses, $K$ and $n_r$ is the same for both the networks given by $P_j = \left( (\omega_{j,rs}) \right)$, $j = 1, 2$ We will make use of the following conditions:

(C.1) For $r = 1, \ldots, K$, $\lim_{n \to \infty} \frac{n_r}{n} = \pi^0_r > 0$.

(C.2) There exists $\{a_n\}_{n \geq 1} \subset (0, 1]$ such that $a_n + n^{-2}a_n^{-1} = o(1)$ and $\omega_{j,rs} = a_n\omega_{j,rs}^0$ for all $1 \leq r, s \leq K$, $j = 1, 2$, where $\omega_{j,rs}^0$ does not depend on $n$.

(C.3) $EH(\hat{c}_{jn}, c_{jn}) = o(n^{-1})$ as $n \to \infty$, $j = 1, 2$.

Condition (C.1) says that the $K$ underlying communities are of comparable size. Condition (C.2) ensures that the network has some degree of sparsity. Without this condition, we get a dense network unsuitable for most practical applications (Zhao et al., 2012) and, therefore, we decided to focus on the sparse case. However, we also indicate the limit distribution in the dense case in Remark 3.1 (which is slightly different from the sparse case). Finally, Condition (C.3) gives the requirements for the community finding algo-
rithm. It requires the expected number of the misclassified nodes to decay at a sufficiently fast rate. While this condition is not exactly captured by the popular notions of weak or strong consistency of a community detection algorithm, it can be proved using known bounds for different algorithms proposed in the literature when \( a_n \) goes to zero slowly or is bounded away from zero (as in a dense network). For example, for the two-stage algorithm proposed in Gao et al. (2017, 2018), this condition holds for sparse assortative SBMs when \( a_n \gg n^{-1} \log n \), as implied by the exponential minimax rate of misclassification proved therein. See Remark S1.2 of the Supplementary Materials for more details.

For analyzing the asymptotic properties of the two-sample test, it is pedagogically and notationally simpler to consider the one-sample case first. A straightforward extension of the arguments will be used to prove the limit distributions for the two-sample case. To that end, for the first testing problem \( H_0 : P_1 = P_2 \) against \( H_1 : P_1 \neq P_2 \), write

\[
T_{1n} = \frac{1}{\sqrt{a_n}} T_{\text{frob}}(A_1, A_2) \equiv \frac{1}{\sqrt{a_n}} \| \hat{P}_1 - \hat{P}_2 \|_F.
\]

Let \( \{Z^0_{rs} : 1 \leq r \leq s \leq K\} \) be a collection of iid N(0,1) random variables and set \( Z^0_{rs} = Z^0_{sr} \) if \( r > s \). Write \( Z^0 = ((Z^0_{rs}))_{K \times K} \). Also, let \( Z^0_1, Z^0_2 \) be two independent copies of \( Z^0 \). Then, we have the following result.

**Theorem 3.1.** Suppose that Conditions (C.1)-(C.3) hold. Then,

\[
T_{1n} \rightarrow^d T_{1,\infty}
\]

where \( T_{1,\infty} = \| D^0_1 \odot Z^0_1 - D^0_2 \odot Z^0_2 \|_F \), \( \odot \) denotes the Hadamard product and

\[
D_k = (\sqrt{[1 + \mathbb{1}(r = s)]} \omega^0_{k,rs}), \ k = 1, 2.
\]

**Remark 3.1.** If \( a_n \rightarrow a \neq 0 \) (i.e., \( \{a_n\}_{n \geq 1} \) remains bounded away from zero), then we have a dense network and a version of Theorem 3.1 still holds. In this case, the test statistic
\[ \|\hat{P}_1 - \hat{P}_2\| \to^d \tilde{\Gamma}_1 \] where, with \( \tilde{\tau}^0_{k,r,s} = a\omega^0_{k,r,s} (1 - a\omega^0_{k,r,s}) \) for \( r \neq s \) and \( \tilde{\tau}^0_{k,k,s} = 2a\omega^0_{k,r,s} (1 - a\omega^0_{k,r,s}) \) for \( r = s \), we have

\[
\tilde{\Gamma}_1 = \sum_{r=1}^K \sum_{s=1}^K \left[ \tilde{\tau}^0_{1,r,s} |Z^0_{1,r,s}|^2 + \tilde{\tau}^0_{2,r,s} |Z^0_{2,r,s}|^2 - 2\sqrt{\tilde{\tau}^0_{1,r,s} \tilde{\tau}^0_{2,r,s} Z^0_{1,r,s} Z^0_{2,r,s}} \right].
\]

Next, consider the test of scaling and define

\[
T^2_{2n} = n^2 a_n \left[ T_{\text{scale}}(A_1, A_2) \right]^2 \equiv n^2 a_n \left\| \frac{\hat{P}_1}{\hat{\rho}_1} - \frac{\hat{P}_2}{\hat{\rho}_2} \right\|^2.
\]

The following result establishes the distributional convergence of \( T_{2n} \).

**Theorem 3.2.** Suppose that Conditions (C.1)-(C.3) hold. Then, under the null hypotheses,

\[ T_{2n} \to^d T_{2,\infty} \]

where \( T_{2,\infty} \) is a nondegenerate random variable whose exact definition is complicated and is given in Equation (S10) of the supplementary files.

Theorems 3.1 and 3.2 establish the limiting distributions of the (scaled) test statistics. Next, we establish the validity of the bootstrap procedures. Let \( P_\ast \) and \( E_\ast \) respectively denote the bootstrap probability and the bootstrap expectation. In addition to Conditions (C.1)-(C.3), we will need the following condition for establishing validity of the bootstrap approximation to the null distribution of \( T_{1n} \):

(C.4) \( E_\ast H(e^\ast_{kn}, \hat{e}_{kn}) = o_p(n^{-1}) \) as \( n \to \infty, k = 1, 2 \).

Condition (C.4) requires that an analog of Condition (C.3) be valid in the bootstrap setup. It is a conditional consistency condition on the community detection method that is assumed to have the same level of accuracy when applied to the bootstrap data set, in the weak (in probability) sense. Again, this condition holds for the community detection algorithm of Gao et al. (2017, 2018) for \( a_n \gg n^{-1} \log n \), as the estimated parameters of the SBM lies in the minimaxity class with high probability, as implied by Lemma S1.2 and
Remark S1.1 of the Supplementary Materials below. With this additional condition, we have the following result.

**Theorem 3.3.** Suppose that Conditions (C.1)-(C.4) hold. Also, suppose that the null hypotheses hold for all \( n \geq 1 \). Then,

\[
\sup_{x \geq 0} \left| P_*(T_{1n}^* \leq x) - P(T_{1n} \leq x \mid H_{0,n}) \right| = o_p(1).
\]

Thus, the proposed bootstrap method provides a valid approximation to the null distribution of the test statistic. Next, define the bootstrap version of the test statistic \( T_{2n}^* \) by replacing \((A_1, A_2)\) by \((A_1^*, A_2^*)\), i.e.,

\[
T_{2n}^* = n^2 a_n \left\| \frac{P_1^*}{\rho_1^*} - \frac{P_2^*}{\rho_2^*} \right\|
\]

where \( P_k^* \) is obtained by estimating the SBM model parameters based on \( A_k^* \) as in the original problem and where \( \rho_k^* = \| P_k^* \|, k = 1, 2 \).

**Theorem 3.4.** Suppose that Conditions (C.1)-(C.4) hold. Also, suppose that the null hypotheses hold for all \( n \geq 1 \). Then,

\[
\sup_{x \geq 0} \left| P_*(T_{2n}^* \leq x) - P(T_{2n} \leq x \mid H_{0,n}) \right| = o_p(1).
\]

**Remark 3.2.** As indicated in the statement of Theorem 3.2, the limit distribution of the test statistic in the scale-hypothesis case is rather complicated, involving a nonlinear function of the population parameters and independent standard Gaussian variables. For the validity of the bootstrap, we show that the bootstrap version of the test statistic can be closely approximated by the same nonlinear function but with population parameters replaced by
their estimators and the standard Gaussian random variables by some conditionally weakly convergent sequences of bootstrap random variables. To show weak convergence of this bootstrap stochastic approximation to the same limit, we develop a technical result (a version of the continuous mapping theorem for conditional weak convergence) in Lemma S.4 of the Supplementary materials file that may be of independent interest. Further, as it is evident from the statement of Theorem 3.1 and Equation (S10), the limiting random variables for both $T_{1n}$ and $T_{2n}$ have a continuous distribution on the real line and therefore, both the tests can be calibrated using the bootstrap quantiles to ensure that they attain any prespecified size (i.e., the probability of Type I error), asymptotically. See Remark S1.3 in the Supplementary materials file for more details on using bootstrap calibration and its theoretical justification.

### 3.2 Consistency under the CL and RDPG models

Consider a test statistic $T_n$ and a rejection region of the form $T_n > K$ for testing $H_0$ against $H_1$ at level $\alpha$. Following Tang et al. (2017a), we define a testing method to be consistent if there exists $K > 0$ such that for any $\eta > 0$,

1. If $H_0$ is true, then $\lim_{n \to \infty} P(T_n > K) \leq \alpha + \eta$, and

2. If $H_1$ is true, then $\lim_{n \to \infty} P(T_n > K) > 1 - \eta$

The following theorems show that the proposed testing methods are consistent as per this definition.

#### 3.2.1 Test of equality

First, we consider the CL modeling framework, i.e., we can write $P_1(i, j) = \theta_i \theta_j$ and $P_2(i, j) = \beta_i \beta_j$ for $1 \leq i < j \leq n$. Let $\theta(1) = \min_{1 \leq i \leq n} \theta_i$ and $\beta(1) = \min_{1 \leq i \leq n} \beta_i$. 


Theorem 3.5. Under the CL model, the test of equality is consistent under the following assumptions:

1. For any $\alpha > 0$, $\min \left( \frac{n^\alpha \theta(1)}{\sqrt{\log(n)}}, \frac{n^\alpha \beta(1)}{\sqrt{\log(n)}} \right) \to \infty$.

2. The quantity $\max \left( \frac{\sum_i \theta_i^2}{n \theta^2}, \frac{\sum_i \beta_i^2}{n \beta^2} \right)$ does not diverge to infinity.

3. Under $H_1$, $P_1 \neq P_2$ is in the sense that $\frac{\|P_1 - P_2\|_F}{n^{1/2 + \alpha}} \to \infty$ for any $\alpha > 0$.

Remark 3.3. The first two assumptions enforce regularity conditions. The first assumption ensures that the networks are not too sparse and the expected degrees are not too small. In particular, assumption 1 holds as long as the smallest degree parameter $\theta(1)$ is of the order $1/(\log(n))^k$ for any $k < \infty$. The second assumption ensures that the network parameters do not vary too much. When $H_1$ is true, the third assumption provides a lower limit on the difference of the two models, given by $\|P_1 - P_2\|_F$, such that the they can be told apart.

Next, consider the case where both $P_1$ and $P_2$ belong to the RDPG model class, i.e., $P_1 = X_1X_1'$ and $P_2 = X_2X_2'$ where $X_1$ and $X_2$ are $n$-by-$d$ matrices. For a matrix $M$ with singular values $\sigma_1(M) \geq \sigma_2(M) \geq ...$ and for a fixed $d$, we define the following quantities:

$$\delta(M) = \max_{1 \leq i \leq n} \sum_{j=1}^{n} M_{ji}; \quad \gamma(M) = \frac{\sigma_d(M) - \sigma_{d+1}(M)}{\delta(M)} \leq 1.$$ 

The definition of $\gamma$ depends implicitly on a parameter $d \in \mathbb{N}$. For a matrix $P = XX^T$ of rank $d$, $\delta(P)$ is simply the maximum expected degree of a graph $A \sim Bernoulli(P)$, $\gamma(P)$ is the $d$th largest eigenvalue normalized by $\delta(P)$.

Theorem 3.6. Under the RDPG model, the test of equality is consistent under the following assumptions:

1. $\exists c_0 > 0$ such that $\min(\gamma(P_1), \gamma(P_2)) > c_0$. 

16
2. ∃ ϵ > 0 such that min(δ(P_1), δ(P_2)) > (log n)^{2+ϵ}.

3. Under H_1, P_1 ≠ P_2 is in the sense that \frac{∥P_1 - P_2∥_F}{\Gamma} \to ∞ where \Gamma = \sum_{i=1}^{2} \left[ 3\sqrt{r_i} + (d\gamma^{-1}(P_i))^{1/2} \right].

Remark 3.4. These assumptions are similar to those used in Tang et al. (2017a). The first assumption is an eigengap type condition that ensures that the singular values are well separated. The second assumption ensures that the expected degrees are not too small. The third assumption ensures that under the alternative hypothesis, the two models are sufficiently well separated.

3.2.2 Test of scaling

As before, we start with the CL model. Recall the definitions before Theorem 3.5.

Theorem 3.7. Under the Chung-Lu model, the test of scaling is consistent if the following assumptions hold:

1. For any α > 0, \( \min \left( \frac{n^\alpha \theta_2(1)}{\sqrt{\log(n)}}, \frac{n^\alpha \beta_2(1)}{\sqrt{\log(n)}} \right) \to ∞ \).

2. The quantity \( \max \left( \frac{\sum_i \theta_i^2}{n^\theta}, \frac{\sum_i \beta_i^2}{n^\beta} \right) \) does not diverge to infinity.

3. Under H_1, P_1 ≠ cP_2 for any c is in the sense that \( \frac{∥P_1 - cP_2∥_F}{\rho_1^{1/2+\alpha}} \to ∞ \) for any α > 0, where \( \rho_1 = ∥P_1∥_F \) and \( \rho_2 = ∥P_2∥_F \).

These assumptions are similar to Theorem 3.5 and carry the same interpretations.

Next, under the RDPG model, recall the definitions before Theorem 3.6, which we do not repeat here in the interest of space.

Theorem 3.8. Under the RDPG model, the test of scaling is consistent if the following assumptions hold:
1. ∃ \( c_0 > 0 \) such that \( \min(\gamma(P_1), \gamma(P_2)) > c_0 \).

2. ∃ \( \epsilon > 0 \) such that \( \min(\delta(P_1), \delta(P_2)) > (\log n)^{2+\epsilon} \).

3. Under \( H_1 \), \( P_1 \neq P_2 \) is in the sense that \( \| \frac{1}{\rho_1} P_1 - \frac{1}{\rho_2} P_2 \|_F \to \infty \) where \( \rho_1 = \| P_1 \|_F \) and \( \rho_2 = \| P_2 \|_F \).

The assumptions and their interpretations are similar to those for Theorem 3.6.

**Remark 3.5.** We note that under the RDPG model, the adjacency spectral embedding (ASE) has better concentration properties for sparse graphs with respect to the \( 2 \to \infty \) norm rather than the Frobenius norm (Xie, 2024). In this work, we have used the Frobenius norm as the test statistic, and using the \( 2 \to \infty \) norm instead could lead to sharper consistency results for the test. We consider this to be an important avenue for future research.

### 4 Simulation Study

We now study the numerical performance of the proposed \( T_{\text{frob}} \) (for the test of equality) and \( T_{\text{scale}} \) (for the test of scaling) methods across a range of network models and parametric scenarios. We report two performance metrics: Type I error, i.e., probability of false rejection when \( H_0 \) is true, which should be close to \( \alpha = 5\% \), and power, i.e., probability of true rejection when \( H_1 \) is true, which should be close to 100\%. We used \( n = 100, 200, 300, 400 \), along with 2000 Monte Carlo simulations and \( B = 200 \) bootstrap iterations in each experiment unless specified otherwise. Comparisons with three existing methods: \( T_{\text{ase}} \) proposed by Tang et al. (2017a), \( T_{\text{omni}} \) proposed by Levin et al. (2017), and \( T_{\text{eig}} \) proposed by Ghoshdastidar and von Luxburg (2018) are included wherever appropriate; see Section S4 in the supplementary materials file for more details on these methods. Note that \( T_{\text{ase}} \) applies
to both equality and scaling problems, whereas $T_{omni}$ and $T_{eig}$ apply only to the equality problem.

Since most of the existing work has focused on the RDPG model, we dedicate a full subsection, Section 4.1, to study various scenarios under this model. Next, Section 4.2 covers three other models from Section 2.3. Due to page limits, the following scenarios were moved to Section S5 of the supplementary materials: the sparse RDPG, the SBM, the DCBM, and a model misspecification case study.

4.1 RDPG model

4.1.1 Test of equality

We consider three cases for the test of equality under the RDPG model. Case 1 considers the scenario where, under $H_1$, all latent positions are different. In Case 2, only a small fraction of the latent positions are different under $H_1$. Case 3 (see Section S5 of supplementary materials) considers a sparse RDPG model with higher sample sizes.

Case 1 (all latent positions vary): We generated $A_1 \sim P_1 = X_1 X_1^T$ and $A_2 \sim P_2 = X_2 X_2^T$, where $X_1$ and $X_2$ are two latent matrices of dimension $n \times 2$ (i.e., $d = 2$). The rows of $X_1$ and $X_2$ are generated by sampling with replacement from the rows of $M_1$ and $M_2$, respectively, with probability vector $\pi$, where

$$M_1 = \begin{pmatrix} 0.6 & -0.4 \\ 0.6 & 0.4 \end{pmatrix}; \quad M_2 = \begin{pmatrix} 0.6 + \epsilon & -0.4 - \epsilon \\ 0.6 + \epsilon & 0.4 + \epsilon \end{pmatrix}; \quad \pi = \begin{pmatrix} 0.4 \\ 0.6 \end{pmatrix}. \tag{6}$$

We first generate random variables $z_1, \ldots, z_n$ that take values 1 and 2 with probabilities 0.4 and 0.6, and construct the latent matrices $X_1$ and $X_2$ such that the $j^{th}$ row of $X_k$ is the $z_j^{th}$ row of $M_k$, for $k = 1, 2$ and $j = 1, \ldots, n$. We studied the performance of the methods $T_{frob}, T_{omni}, T_{ase}$ and $T_{eig}$ over values of $\epsilon = 0, 0.02, 0.04, 0.06, 0.08, 0.12$, and this experiment
is performed for network sizes $n = 100, 200, 300, 400$. The null hypothesis $H_0 : P_1 = P_2$ holds true when $\epsilon = 0$, and the alternative $H_1 : P_1 \neq P_2$ holds true when $\epsilon > 0$.

Figure 1 displays the mean rejection rates (with standard error bands) as a function of $\epsilon$. Note that the rejection rate for $\epsilon = 0$ is the type-I error and should be close to 5%. We denote this by a solid black square in Figure 1 and all subsequent plots of rejection rates. We observe that $T_{\text{omni}}$ and $T_{\text{frob}}$ performs about the same, with $T_{\text{omni}}$ being slightly better. Both of them perform much better than $T_{\text{ase}}$ and $T_{\text{eig}}$. In particular, $T_{\text{ase}}$ has very high type-I error as well as low power compared to $T_{\text{omni}}$ and $T_{\text{frob}}$.

**Case 2 (fraction of latent positions vary):** Similar to Case 1, we use the model settings from (6). However, under $H_1$, only 10\% of the rows of $X_2$ are sampled from $M_2$, and the remaining rows of $X_2$ are set to be the same as the corresponding rows of $X_1$. We consider the performance of the methods $T_{\text{frob}}, T_{\text{omni}}, T_{\text{ase}}$ and $T_{\text{eig}}$ over values of $\epsilon = 0, 0.02, 0.04, 0.06, 0.08, 0.12$ for network sizes $n = 100, 200, 300, 400$. The results are displayed in Figure 2. We can see that the proposed test $T_{\text{frob}}$ performs better than $T_{\text{ase}}$, $T_{\text{omni}}$ and $T_{\text{eig}}$, although $T_{\text{omni}}$ catches up with $T_{\text{frob}}$ for $n = 400$.

To summarize, we observe that $T_{\text{frob}}$ and $T_{\text{omni}}$ consistently outperform $T_{\text{ase}}$ and $T_{\text{eig}}$ across all cases. The proposed $T_{\text{frob}}$ has accuracy similar to $T_{\text{omni}}$, with $T_{\text{frob}}$ being more accurate in Case 2 and $T_{\text{omni}}$ being slightly more accurate in Case 1. However, the difference between $T_{\text{frob}}$ and $T_{\text{omni}}$ becomes negligible when the sample size is increased to $n = 400$.

Beyond statistical accuracy, it is also important to consider computational cost. We report the runtimes of the four methods in Figure 3 as a function of $n$. We observe that $T_{\text{frob}}$ is approximately four times faster than $T_{\text{omni}}$ and twice as fast as $T_{\text{ase}}$, as $T_{\text{omni}}$ involves spectral decomposition of a $2n \times 2n$ matrix (Levin et al., 2017) and $T_{\text{ase}}$ requires Procrustes transformation. Thus, $T_{\text{frob}}$ and $T_{\text{omni}}$ have comparable accuracy, and both of them are
much more accurate than $T_{ase}$ and $T_{eig}$. Given its much lower computational cost than $T_{omni}$, the proposed $T_{frob}$ is therefore the preferred method.

### 4.1.2 Test of scaling

As before, we first generate random variables $z_1, \ldots, z_n$ that take values 1 and 2 with probabilities 0.4 and 0.6, and construct the latent matrices $X_1$ and $X_2$ such that the $j^{th}$ row of $X_k$ is $z_j^{th}$ row of $M_k$, for $k = 1, 2$ and $j = 1, \ldots, n$, where

$$M_1 = \begin{pmatrix} 0.6 & -0.4 \\ 0.6 & 0.4 \end{pmatrix}; \quad M_2 = \begin{pmatrix} 0.3 & -0.1 \\ 0.3 & 0.1 \end{pmatrix}. \quad (7)$$

Next, we generate three network adjacency matrices: $A_1 \sim P_1 = X_1 X_1^T$, $A_2 \sim P_2 = (\sqrt{c}X_1)(\sqrt{c}X_1)^T$ for some constant $c > 0$, and $A_3 \sim P_3 = X_2 X_2^T$. Here, the pair $(A_1, A_2)$ satisfies $H_0$ and we performed the test of scaling with $c = 0.5, 0.7, 0.75, 0.8, 0.9$ to obtain Type-1 error rates. We also performed the test using the pair $(A_1, A_3)$ to obtain the power under $H_1$. Table 1 shows that $T_{ase}$ has low power for $n = 100$ and $n = 200$, and Type-1
Figure 2: Rejection rates for RDPG test of equality (Case 2). The solid black square represents the 5% significance level for under $H_0$ ($\epsilon = 0$).

Figure 3: Runtime comparison between various methods

error close to zero in most cases. The proposed method, $T_{scale}$, performed much better with Type I error rates closer to the target value of $\alpha = 5\%$ under $H_0$ and higher power under
H1. The only exception is n = 200, c = 0.9 where the type-I error from the $T_{ase}$ test is closer to $\alpha = 5\%$. Note that $T_{eig}$ and $T_{omni}$ do not apply to the test of scaling.

| n   | $T_{scale}$ | $T_{ase}$ | $T_{scale}$ | $T_{ase}$ | $T_{scale}$ | $T_{ase}$ | $T_{scale}$ | $T_{ase}$ | $T_{scale}$ | $T_{ase}$ |
|-----|-------------|-----------|-------------|-----------|-------------|-----------|-------------|-----------|-------------|-----------|
| 100 | 9.2         | 1.2       | 7.2         | 1.6       | 6.6         | 1.4       | 6.8         | 2.2       | 4.7         | 6.6       |
| 200 | 10.0        | 0.0       | 6.1         | 0.0       | 7.4         | 0.0       | 5.6         | 0.0       | 8.4         | 3.9       |
| 300 | 8.4         | 0.0       | 5.6         | 0.0       | 6.1         | 0.0       | 5.9         | 0.0       | 8.4         | 1.8       |
| 400 | 6.5         | 0.0       | 5.8         | 0.0       | 6.4         | 0.0       | 6.1         | 0.0       | 4.5         | 0.4       |

Table 1: RDPG scaling case: Rejection rates (in percentage) from $T_{scale}$ (the proposed method) and $T_{ase}$ for the scaling case using $B = 200$ bootstrap iterations and averaged over 2000 Monte Carlo simulations. The first five scenarios refer to Type I error rates, and the third scenario refers to the power of the test. Our method performed much better than $T_{ase}$, with Type I error rates closer to the target value of $\alpha = 5\%$ and higher power.

### 4.2 Results for non-RDPG models

#### 4.2.1 Chung-Lu model

For the Chung-Lu model, we sampled one set of parameters $\theta_i \sim \text{Beta}(a = 1, b = 5)$ for $i = 1, \ldots, n$, and used $P_1(i, j) = \theta_i \theta_j$, and used $P_2 = P_1$ to configure the null scenario under the equality case. To configure the alternative scenario, we set $P_2(i, j) = (\theta_i + \epsilon)(\theta_j + \epsilon)$. For comparison, we have implemented the $T_{ase}$ method of Tang et al. (2017a), the $T_{omni}$ method of Levin et al. (2017) and the $T_{frob}$ method of Ghoshdastidar and von Luxburg (2018). As noted earlier, the number of blocks $r$, has to be provided as an input to the $T_{eig}$ method and the authors did not provide a strategy for obtaining $r$. We used four ad-hoc values, $r = 1, 2, 5, 10$, to study the performance of $T_{eig}$. Figure 4 shows the increase in the rejection rate for the equality case in the range $\epsilon = 0, 0.025, 0.05, 0.075, 0.1, 0.15, 0.2$. We can observe that the three methods $T_{ase}$, $T_{omni}$ and $T_{frob}$ perform well. $T_{omni}$ performs the best for smaller values of $n$ but $T_{frob}$ and $T_{ase}$ catch up to $T_{omni}$ for $n = 400$. $T_{eig}$ has a high rejection rate even when null is true for different $r$ values, making it the worst performing method in this case.
For the scaling case, we used the same $P_1$ as the equality case, and set $P_2 = cP_1$ with $c = 0.5, 0.7, 0.75, 0.8, 0.9$, to configure a range of scenarios satisfying $H_0$. We generated another separate set of parameters $\eta_i \sim \text{Beta}(a = 4, b = 3)$ and used $P_2(i, j) = \eta_i \eta_j$ under $H_1$. Note that no existing method applies to the test of scaling under the Chung-Lu model. Table 2 shows that the proposed method, $T_{\text{scale}}$, performed reasonably well, with Type I error rates somewhat higher than $\alpha = 5\%$, and power equal to 1.

Figure 4: Chung-Lu Equality test: Rejection rates for deviating alternatives. $T_{\text{frob}}, T_{\text{ase}},$ and $T_{\text{omni}}$ performed well with Type I error rates close to the target value of $\alpha = 5\%$ and power close to 1. Type I error rates from $T_{\text{eig}}$ were very high for all values of $r = 1, 2, 5, 10$.

4.2.2 Popularity Adjusted Block Model

Here we used parameter configurations from the simulation study of Sengupta and Chen (2018). We consider networks with $K = 2$ equally sized communities. Model parameters are set as $\lambda_{ir} = \alpha \sqrt{\frac{h}{1+h}}$ when $r = c_i$, and $\lambda_{ir} = \beta \sqrt{\frac{1}{1+h}}$ when $r \neq c_i$, where $h$ is the
Table 2: Chung-Lu scaling case: Rejection rates (in percentage) from $T_{scale}$ using $B = 200$ bootstrap iterations and averaged over 2000 Monte Carlo simulations. Scenarios 1 – 5 satisfy $H_0$ and the sixth scenario satisfies the alternative. Our method performed reasonably well with Type I error rates somewhat higher than $\alpha = 5\%$ and power equal to 1.

| $n$  | $P_2 = 0.5P_1$ | $P_2 = 0.7P_1$ | $P_2 = 0.75P_1$ | $P_2 = 0.8P_1$ | $P_2 = 0.9P_1$ | $P_2 \neq cP_1$ |
|------|----------------|----------------|----------------|----------------|----------------|----------------|
| 100  | 11.4           | 8.8            | 8.2            | 9.0            | 8.0            | 100            |
| 200  | 9.1            | 9.8            | 8.6            | 8.3            | 8.6            | 100            |
| 300  | 8.3            | 8.5            | 7.8            | 7.4            | 8.2            | 100            |
| 400  | 7.6            | 7.8            | 7.4            | 8.1            | 7.0            | 100            |

homophily factor. In each community, we designated 50% of the nodes as Category 1 and 50% as Category 2. We set $\alpha = 0.8, \beta = 0.2$ for category 1 nodes and $\alpha = 0.2, \beta = 0.8$ for category 2 nodes. For the equality case, we used $h = 4$ for $P_1$, set $P_1 = P_2$ under the null, and used $h - \epsilon$ for $P_2$ under the alternative, for $\epsilon = 0, 0.025, 0.05, 0.075, 0.1, 0.15, 0.2$, similar to the Chung-Lu case. Since $T_{sig}$ had very high rejection rates under the Chung-Lu model (Figure 4), we skipped it for the PABM and subsequent models. For the test of scaling, we set $h = 4$ for $P_1$, set $P_2 = cP_1$ under $H_0$ with $c = 0.5, 0.7, 0.75, 0.8, 0.9$, and used $h = 2$ for $P_2$ under $H_1$.

The results for the test of equality case are plotted in Figure 5 and the same for the test of scaling case are reported in Table 3. We note that the proposed tests performed quite well. The type I errors were somewhat conservative both for the test of equality and the test of scaling. The power in both cases improved from $n = 100$ to $n = 400$, as expected.

4.2.3 Latent Space Model

Under the latent distance model of Hoff et al. (2002), we used $d = 3$, $\alpha = 3$, and sampled the latent positions $z_1, \ldots, z_n \sim N_3(0, I)$ independently for $P_1$. For the test of equality, we set $P_2 = P_1$ under $H_0$. To configure the alternative scenario for the equality case, we used $\alpha = 3 - \epsilon$ for $\epsilon = 0, 0.025, 0.05, 0.075, 0.1, 0.15, 0.2$, and sampled a different set of $z_1, \ldots, z_n \sim N_3(0, I)$ independently for $P_2$. For the test of scaling, we set $P_2 = cP_1$ under $H_0$ with $c = 0.5, 0.7, 0.75, 0.8, 0.9$, and used the same $P_2$ used in the test of equality under $H_1$. For the scaling case, we used $P_2 = c \times P_1$ under $H_0$, and used $\alpha = 3$ and
sampled another different set of \( z_1, \ldots, z_n \sim N_3(0, I) \) independently to create \( P_2 \) under the alternative model. It has been well documented that estimation under the latent space model is computationally expensive (Raftery et al., 2012; Salter-Townshend and Murphy, 2013). The computational expense for our inferential method is further exacerbated due to bootstrap resampling. Therefore, we used smaller sample sizes, \( n = 30, 40, 50 \), and carried out 500 Monte Carlo iterations for each sample size, and we used \( B = 200 \) bootstrap iterations as before. We note that the computational issue can potentially be resolved by using approximation techniques (Raftery et al., 2012) or variational inference (Salter-Townshend and Murphy, 2013); however, we did not pursue this direction in this work, and we consider this as an important future direction.

The results for the test of equality and the test of scaling are reported in Figure 6 and Table 4, respectively. Our methods work quite well in both cases, with Type I error rates close to the nominal value of 5% and power equal to 100%.

4.3 Computational considerations

The computational costs of implementing the proposed \( T_{\text{frob}} \) and \( T_{\text{scale}} \) methods are primar-
Table 3: PABM scaling case: Rejection rates (in percentage) from $T_{\text{scale}}$ using $B = 200$ bootstrap iterations. Results are averaged over 2000 Monte Carlo simulations.

| $H_0$ is true | $H_1$ is true |
|----------------|----------------|
| $P_2 = 0.5P_1$ | $P_2 = cP_1$ |
| $P_2 = 0.7P_1$ | $P_2 = 0.8P_1$ | $P_2 = 0.9P_1$ |
| $P_2 = 0.75P_1$ | $P_2 = 0.9P_1$ |
| $P_2 \neq cP_1$ |
| $n$ | $T_{\text{scale}}$ | $T_{\text{scale}}$ | $T_{\text{scale}}$ | $T_{\text{scale}}$ | $T_{\text{scale}}$ | $T_{\text{scale}}$ |
| 100 | 1.1 | 1.2 | 1.5 | 2.1 | 2.9 | 34.9 |
| 200 | 0.9 | 1.1 | 1.6 | 1.7 | 2.3 | 99.0 |
| 300 | 0.6 | 0.9 | 0.8 | 1.1 | 0.8 | 100 |
| 400 | 0.0 | 0.2 | 0.4 | 0.6 | 0.8 | 100 |

Figure 6: Rejection rates for the test of equality under the latent space model.
Table 4: Latent space model scaling case: Rejection rates (in percentage) from $T_{\text{scale}}$ using $B = 200$ bootstrap iterations and averaged over 500 Monte Carlo simulations.

|        | $H_0$ is true | $H_1$ is true |
|--------|---------------|---------------|
| $P_2 = 0.5P_1$ | $T_{\text{scale}}$ | $T_{\text{scale}}$ |
|        | $P_2 = 0.7P_1$ | $P_2 = 0.75P_1$ | $P_2 = 0.8P_1$ | $P_2 = 0.9P_1$ | $P_2 \neq cP_1$ |
| $n$    | $T_{\text{scale}}$ | $T_{\text{scale}}$ | $T_{\text{scale}}$ | $T_{\text{scale}}$ | $T_{\text{scale}}$ |
| 30     | 9.2           | 7.2           |
| 40     | 6.2           | 5.6           | 3.6           | 4.2           |
| 50     | 5.4           | 5.8           | 5.2           | 4.6           |

5 Aarhus Computer Science Department Network

This anonymized network dataset was collected by Rossi and Magnani (2015) at the Department of Computer Science at Aarhus University. It includes five kinds of interaction (coauthor, leisure, work, lunch, and Facebook) between $n = 61$ researchers, including professors, postdocs, and Ph.D. students. Importantly, these interactions span both social (e.g., leisure and Facebook) and professional domains (e.g., coauthor and work). Each kind of interaction is represented as an undirected network.

We carried out tests for equality (using $T_{\text{frob}}$) and scaling (using $T_{\text{scale}}$) for each pair of interactions with $\alpha = 5\%$ and $B = 10000$ bootstrap iterations. Since this involves 20 simultaneous tests of hypothesis ($\binom{5}{2}$ pairs each for equality and scaling), we use Bonferroni’s correction for multiple testing and set the significance level for each pairwise test at $\alpha/20 = 0.0025$. Following Han et al. (2015), we used an RDPG model with $d = 4$. The equality hypothesis ($H_0 : P_1 = P_2$) was rejected for all interaction pairs.

From Table 5, we observe that the scaling hypothesis ($H_0 : P_1 = cP_2$ for some $c > 0$) was rejected for three pairs of interaction types: leisure-Facebook, work-Facebook, and lunch-Facebook. The test of scaling was not rejected for the seven remaining pairs of interactions. Notably, the co-author network displayed scaling similarity with every other network, indicating that co-authors are likely to have other kinds of interactions like work, lunch, leisure, and social media, in a proportional manner. This aligns well with our understanding of
how academic interactions often intermingle, and how a co-authorship relation can be
associated with other forms of professional and social interactions. We also observed that
the highest p-value in Table 5 is for the leisure and lunch pair, underscoring the strength
in their scaling similarity, which makes intuitive sense as these two activities are similar.
Thus, the test results reveal interesting similarities across social and professional modes of
interaction while also reflecting differences in communication intensity across these modes.
Such findings are valuable as they simplify the analysis of multilayer networks by reduc-
ing the effective dimensionality of interactions, enabling domain scientists to focus on the
underlying patterns that govern similar behavior across multiple interaction types.

| Interaction type | coauthor | leisure | work    | lunch   | facebook |
|-----------------|----------|---------|---------|---------|----------|
| coauthor        | -        | 0.1325  | 0.1104  | 0.1393  | 0.0538   |
| leisure         | -        | -       | 0.0270  | 0.1956  | 0.0008   |
| work            | -        | -       | -       | 0.0144  | 0.0014   |
| lunch           | -        | -       | -       | -       | 0.0000   |

Table 5: Aarhus network: p-values for the test of scaling (using $T_{scale}$) for all interaction pairs.

6 Discussion

This paper studies the matched network inference problem, where the statistician is given
two independent networks on the same set of entities, and the goal is to determine whether
the two networks are similar. We propose a bootstrap-based testing framework for two key
problems: the equality problem, testing if the networks originate from the same random
graph model, and the scaling problem, testing whether their underlying probability ma-
trices are scaled versions of each other. The proposed methodology works well on a wide
range of random graph models, as demonstrated by our theoretical and empirical results,
and outperforms existing approaches in terms of flexibility, computational efficiency, and
statistical accuracy across a wide range of scenarios.

For future research, an important next step will be to move beyond equality and scal-
ing and develop testing methods for more general and intricate notions of similarity, $\tau(P)$. Examples of network features of interest include expected subgraph counts (Bhattacharyya
and Bickel, 2015), clique numbers (Sengupta, 2018), graph spectra (Van Mieghem, 2010;
Jovanović and Stanić, 2012; Dasgupta and Sengupta, 2022), path lengths (Watts and Strogatz, 1998; Lovekar et al., 2021), and so on. There are two challenges involved in extending the proposed framework to more general $\tau(P)$. First, it needs to be established that the convergence of $\hat{P}$ to $P$ leads to the convergence of $\tau(\hat{P})$ to $\tau(P)$. This hinges on the smoothness properties of $\tau(\cdot)$, and the approximations and expansions analyzed in Zhang and Xia (2022) and Levin and Levina (2019) could be excellent starting points towards such results. The second challenge is how to transform $\tau(\hat{P}_1)$ and $\tau(\hat{P}_2)$ into their null-restricted counterparts which can be used to generate parametric bootstrap resamples to estimate the sampling distribution of the test statistic under the null. We look forward to methodological innovation from the research community that will address these challenges.

Another impactful direction of future research would be extending the proposed framework to dynamic or time-varying networks. Network monitoring has emerged as a highly active area of research in recent years (Woodall et al., 2017; Jeske et al., 2018; Sengupta and Woodall, 2018; Stevens et al., 2021a,b,c,d). In this context, the proposed framework can be used for changepoint analysis and anomaly detection for identifying sudden shifts in connectivity patterns as the network evolves over time, by testing for $H_0: \tau(P_t) = \tau(P_{t+1})$ where $t$ denotes time.

**Supplementary Material**

The supplementary material contains algorithms and proofs of technical results

**References**

Agterberg, J., Tang, M., and Priebe, C. (2020). Nonparametric two-sample hypothesis testing for random graphs with negative and repeated eigenvalues. *arXiv preprint arXiv:2012.09828*.

Athreya, K. B. and Lahiri, S. N. (2006). *Measure theory and probability theory*, volume 19. Springer.

Bassett, D. S., Bullmore, E., Verchinski, B. A., Mattay, V. S., Weinberger, D. R., and Meyer-Lindenberg, A. (2008). Hierarchical organization of human cortical networks in health and schizophrenia. *Journal of Neuroscience*, 28(37):9239–9248.
Beran, R. and Ducharme, G. R. (1991). *Asympotic theory for bootstrap methods in statistics.*

Bhattacharyya, S. and Bickel, P. J. (2015). Subsampling bootstrap of count features of networks. *The Annals of Statistics,* 43(6):2384–2411.

Chakrabarty, S., Sengupta, S., and Chen, Y. (2025). Subsampling based community detection for large networks. *Statistica Sinica,* 35(3):1–42.

Chakraborty, K., Sengupta, S., and Chen, Y. (2025). Scalable estimation and two-sample testing for large networks via subsampling. *Journal of Computational and Graphical Statistics,* 34:1–13.

Chatterjee, A. and Lahiri, S. N. (2011). Bootstrapping lasso estimators. *Journal of the American Statistical Association,* 106(494):608–625.

Chung, F. and Lu, L. (2002). The average distances in random graphs with given expected degrees. *Proceedings of the National Academy of Sciences,* 99(25):15879–15882.

Dasgupta, A. and Sengupta, S. (2022). Scalable estimation of epidemic thresholds via node sampling. *Sankhya A,* 84:321–344.

Davison, A. C. and Hinkley, D. V. (1997). *Bootstrap methods and their application.* Number 1. Cambridge university press.

Efron, B. (1979). Bootstrap Methods: Another Look at the Jackknife. *The Annals of Statistics,* 7(1):1 – 26.

Efron, B. and Tibshirani, R. J. (1994). *An introduction to the bootstrap.* CRC press.

Erdös, P. and Rényi, A. (1959). On random graphs. *Publicationes Mathematicae Debrecen,* 6:290–297.

Ganguly, I., Sengupta, S., and Ghosh, S. (2023). Scalable resampling in massive generalized linear models via subsampled residual bootstrap. *arXiv preprint arXiv:2307.07068.*

Gao, C., Ma, Z., Zhang, A. Y., and Zhou, H. H. (2017). Achieving optimal misclassification proportion in stochastic block models. *The Journal of Machine Learning Research,* 18(1):1980–2024.
Gao, C., Ma, Z., Zhang, A. Y., and Zhou, H. H. (2018). Community detection in degree-corrected block models. *The Annals of Statistics*, 46(5):2153–2185.

Gemmetto, V. and Garlaschelli, D. (2017). Reconstruction of multiplex networks with correlated layers. *arXiv preprint arXiv:1709.03918*.

Ghoshdastidar, D., Gutzeit, M., Carpentier, A., and von Luxburg, U. (2017). Two-sample tests for large random graphs using network statistics. In *Conference on Learning Theory*, pages 954–977.

Ghoshdastidar, D., Gutzeit, M., Carpentier, A., and von Luxburg, U. (2020). Two-sample hypothesis testing for inhomogeneous random graphs. *Annals of Statistics*, 48(4).

Ghoshdastidar, D. and von Luxburg, U. (2018). Practical methods for graph two-sample testing. In *Advances in Neural Information Processing Systems*, pages 3019–3028.

Ginestet, C. E., Li, J., Balachandran, P., Rosenberg, S., and Kolaczyk, E. D. (2017). Hypothesis testing for network data in functional neuroimaging. *The Annals of Applied Statistics*, 11(2):725–750.

Green, A. and Shalizi, C. R. (2022). Bootstrapping exchangeable random graphs. *Electronic Journal of Statistics*, 16(1):1058–1095.

Hall, P. (1993). On edgeworth expansion and bootstrap confidence bands in nonparametric curve estimation. *Journal of the Royal Statistical Society: Series B (Methodological)*, 55(1):291–304.

Hall, P. and Horowitz, J. (2013). A simple bootstrap method for constructing nonparametric confidence bands for functions. *The Annals of Statistics*, 41(4):1892–1921.

Han, Q., Xu, K., and Airoldi, E. (2015). Consistent estimation of dynamic and multi-layer block models. In *International Conference on Machine Learning*, pages 1511–1520.

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(460):1090–1098.

Holland, P., Laskey, K., and Leinhardt, S. (1983). Stochastic blockmodels: first steps. *Social Networks*, 5:109–137.
Jeske, D. R., Stevens, N. T., Tartakovsky, A. G., and Wilson, J. D. (2018). Statistical methods for network surveillance. *Applied Stochastic Models in Business and Industry*, 34(4):425–445.

Jin, J., Ke, Z. T., Luo, S., and Ma, Y. (2024). Optimal network pairwise comparison. *Journal of the American Statistical Association*, (just-accepted):1–25.

Jovanović, I. and Stanić, Z. (2012). Spectral distances of graphs. *Linear Algebra and its Applications*, 436(5):1425–1435.

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

Kleiner, A., Talwalkar, A., Sarkar, P., and Jordan, M. I. (2014). A scalable bootstrap for massive data. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 76(4):795–816.

Krivitsky, P. N. and Handcock, M. S. (2008). Fitting position latent cluster models for social networks with latentnet. *Journal of Statistical Software*, 24.

Lahiri, S. (2003). *Resampling methods for dependent data*. Springer Science & Business Media.

Le, C. M., Levina, E., and Vershynin, R. (2016). Optimization via low-rank approximation for community detection in networks. *Ann. Statist.*, 44(1):373–400.

Levin, K., Athreya, A., Tang, M., Lyzinski, V., Park, Y., and Priebe, C. E. (2017). A central limit theorem for an omnibus embedding of multiple random graphs and implications for multiscale network inference. *arXiv preprint arXiv:1705.09355*.

Levin, K. and Levina, E. (2019). Bootstrapping networks with latent space structure. *arXiv preprint arXiv:1907.10821*.

Li, Y. and Li, H. (2018). Two-sample test of community memberships of weighted stochastic block models. *arXiv preprint arXiv:1811.12593*.

Lovász, L. (2012). *Large networks and graph limits*, volume 60. American Mathematical Society.
Lovekar, K., Sengupta, S., and Paul, S. (2021). Testing for the network small-world property. \textit{arXiv preprint arXiv:2103.08035}.

Lunde, R. and Sarkar, P. (2022). Subsampling sparse graphons under minimal assumptions. \textit{Biometrika}, 110(1):15–32.

MacDonald, P. W., Levina, E., and Zhu, J. (2022). Latent space models for multiplex networks with shared structure. \textit{Biometrika}, 109(3):683–706.

Mukherjee, S. S., Sarkar, P., and Bickel, P. J. (2021). Two provably consistent divide-and-conquer clustering algorithms for large networks. \textit{Proceedings of the National Academy of Sciences}, 118(44):e2100482118.

Nicosia, V. and Latora, V. (2015). Measuring and modeling correlations in multiplex networks. \textit{Physical Review E}, 92(3):032805.

Olhede, S. C. and Wolfe, P. J. (2014). Network histograms and universality of blockmodel approximation. \textit{Proceedings of the National Academy of Sciences}, 111(41):14722–14727.

Paul, S. and Chen, Y. (2016). Consistent community detection in multi-relational data through restricted multi-layer stochastic blockmodel. \textit{Electronic Journal of Statistics}, 10(2):3807–3870.

Politis, D. N. (2024). Scalable subsampling: computation, aggregation and inference. \textit{Biometrika}, 111(1):347–354.

Qin, T. and Rohe, K. (2013). Regularized spectral clustering under the degree-corrected stochastic blockmodel. In \textit{Advances in Neural Information Processing Systems}, pages 3120–3128.

Raftery, A. E., Niu, X., Hoff, P. D., and Yeung, K. Y. (2012). Fast inference for the latent space network model using a case-control approximate likelihood. \textit{Journal of Computational and Graphical Statistics}, 21(4):901–919.

Rohe, K., Chatterjee, S., and Yu, B. (2011). Spectral clustering and the high-dimensional stochastic blockmodel. \textit{The Annals of Statistics}, 39(4):1878–1915.

Rossi, L. and Magnani, M. (2015). Towards effective visual analytics on multiplex and multilayer networks. \textit{Chaos, Solitons & Fractals}, 72:68–76.
Rothe, G. et al. (1981). Some properties of the asymptotic relative pitman efficiency. *The Annals of Statistics*, 9(3):663–669.

Salter-Townshend, M. and Murphy, T. B. (2013). Variational Bayesian inference for the latent position cluster model for network data. *Computational Statistics & Data Analysis*, 57(1):661–671.

Sengupta, S. (2018). Anomaly detection in static networks using egonets. *arXiv preprint arXiv:1807.08925*.

Sengupta, S. and Chen, Y. (2015). Spectral clustering in heterogeneous networks. *Statistica Sinica*, 25:1081–1106.

Sengupta, S. and Chen, Y. (2018). A block model for node popularity in networks with community structure. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 80(2):365–386.

Sengupta, S., Shao, X., and Wang, Y. (2015). The dependent random weighting. *Journal of Time Series Analysis*, 36(3):315–326.

Sengupta, S., Volgushev, S., and Shao, X. (2016). A subsampled double bootstrap for massive data. *Journal of the American Statistical Association*, 111(515):1222–1232.

Sengupta, S. and Woodall, W. H. (2018). Discussion of “Statistical Methods for Network Surveillance” by D. R. Jeske, N.T. Stevens, A. G. Tartakovsky and J. D. Wilson. *Applied Stochastic Models in Business and Industry*, 34(4):446–448.

Shao, J. and Tu, D. (1995). *The jackknife and bootstrap*. Springer Science & Business Media.

Shao, Z. and Le, C. M. (2024). Parametric bootstrap on networks with non-exchangeable nodes. *arXiv preprint arXiv:2402.01866*.

Singh, K. (1981). On the asymptotic accuracy of efron’s bootstrap. *The Annals of Statistics*, pages 1187–1195.

Stevens, N. T., Wilson, J. D., Driscoll, A. R., McCulloh, I., Michailidis, G., Paris, C., Parker, P., Paynabar, K., Perry, M. B., Reisi-Gahrooei, M., Sengupta, S., and Sparks,
R. (2021a). Broader impacts of network monitoring: Its role in government, industry, technology, and beyond. *Quality Engineering*, 33(4):749–757.

Stevens, N. T., Wilson, J. D., Driscoll, A. R., McCulloh, I., Michailidis, G., Paris, C., Parker, P., Paynabar, K., Perry, M. B., Reisi-Gahrooei, M., Sengupta, S., and Sparks, R. (2021b). Foundations of network monitoring: Definitions and applications. *Quality Engineering*, 33(4):719–730.

Stevens, N. T., Wilson, J. D., Driscoll, A. R., McCulloh, I., Michailidis, G., Paris, C., Parker, P., Paynabar, K., Perry, M. B., Reisi-Gahrooei, M., Sengupta, S., and Sparks, R. (2021c). The interdisciplinary nature of network monitoring: Advantages and disadvantages. *Quality Engineering*, 33(4):731–735.

Stevens, N. T., Wilson, J. D., Driscoll, A. R., McCulloh, I., Michailidis, G., Paris, C., Parker, P., Paynabar, K., Perry, M. B., Reisi-Gahrooei, M., Sengupta, S., and Sparks, R. (2021d). Research in network monitoring: Connections with spm and new directions. *Quality Engineering*, 33(4):736–748.

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(499):1119–1128.

Tang, M., Athreya, A., Sussman, D. L., Lyzinski, V., Park, Y., and Priebe, C. E. (2017a). A semiparametric two-sample hypothesis testing problem for random graphs. *Journal of Computational and Graphical Statistics*, 26(2):344–354.

Tang, M., Athreya, A., Sussman, D. L., Lyzinski, V., and Priebe, C. E. (2017b). A nonparametric two-sample hypothesis testing problem for random graphs. *Bernoulli*, 23(3):1599–1630.

Tracy, C. A. and Widom, H. (1996). On orthogonal and symplectic matrix ensembles. *Communications in Mathematical Physics*, 177(3):727–754.

Van Mieghem, P. (2010). *Graph spectra for complex networks*. Cambridge University Press.

Wang, F., Li, W., Padilla, O. H. M., Yu, Y., and Rinaldo, A. (2023). Multilayer random dot product graphs: Estimation and online change point detection. *arXiv preprint arXiv:2306.15286*. 

36
Watts, D. J. and Strogatz, S. H. (1998). Collective dynamics of ‘small-world’ networks. *Nature*, 393:440–442.

Woodall, W. H., Zhao, M. J., Paynabar, K., Sparks, R., and Wilson, J. D. (2017). An overview and perspective on social network monitoring. *IISE Transactions*, 49(3):354–365.

Wu, Q. and Hu, J. (2024). Two-sample test of stochastic block models. *Computational Statistics & Data Analysis*, 192:107903.

Xie, F. (2024). Entrywise limit theorems for eigenvectors of signal-plus-noise matrix models with weak signals. *Bernoulli*, 30(1):388–418.

Young, S. J. and Scheinerman, E. R. (2007). Random dot product graph models for social networks. In *International Workshop on Algorithms and Models for the Web-Graph*, pages 138–149. Springer.

Zhang, B., Li, H., Riggins, R. B., Zhan, M., Xuan, J., Zhang, Z., Hoffman, E. P., Clarke, R., and Wang, Y. (2009). Differential dependency network analysis to identify condition-specific topological changes in biological networks. *Bioinformatics*, 25(4):526–532.

Zhang, Y., Levina, E., and Zhu, J. (2017). Estimating network edge probabilities by neighbourhood smoothing. *Biometrika*, 104(4):771–783.

Zhang, Y. and Xia, D. (2022). Edgeworth expansions for network moments. *The Annals of Statistics*, 50(2):726–753.

Zhao, Y., Levina, E., and Zhu, J. (2012). Consistency of community detection in networks under degree-corrected stochastic block models. *The Annals of Statistics*, 40:2266–2292.
Supplementary materials for
“A Bootstrap-based Method for Testing Similarity of Matched Networks”

S1 Proofs of bootstrap distributional results (Theorems 3.1 - 3.4)

For vectors $c, d \in \Sigma^n$ for some finite alphabet $\Sigma$, define the Hamming distance between $c = (c_1, \ldots, c_n)'$ and $d = (d_1, \ldots, d_n)'$ as

$$H(c, d) = \sum_{i=1}^n \mathbb{I}(c_i \neq d_i),$$

where $\mathbb{I}(\cdot)$ denotes the indicator function. Thus, $H$ counts the number of mismatches among the components of $c$ and $d$. For $n \geq 1$, write $\hat{c}_n = (\hat{c}_1, \ldots, \hat{c}_n)'$ and $c_n = (c_1, \ldots, c_n)'$ for the estimated and the true community assignments (after relabeling as necessary).

Lemma S1.1. For any $1 \leq r, s \leq K$,

$$|\hat{n}_{rs} - n_{rs}| \leq 2nH(\hat{c}_n, c_n) \quad \text{for all} \quad n \geq 1.$$

Proof: Clearly,

$$|\hat{n}_{rs} - n_{rs}| = \left| \sum_{i \neq j} \mathbb{I}(\hat{c}_i = r, \hat{c}_j = s) - \sum_{i \neq j} \mathbb{I}(c_i = r, c_j = s) \right|$$

$$= \left| \sum_{i \neq j} \left\{ \mathbb{I}(\hat{c}_i = r) - \mathbb{I}(c_i = r) \right\} \mathbb{I}(\hat{c}_j = s) \right|$$

$$+ \left| \sum_{i \neq j} \left\{ \mathbb{I}(\hat{c}_j = s) - \mathbb{I}(c_j = s) \right\} \mathbb{I}(c_i = r) \right|$$

$$\leq \sum_{i \neq j} \mathbb{I}(\hat{c}_i \neq c_i) + \sum_{i \neq j} \mathbb{I}(\hat{c}_j \neq c_j),$$

which proves the lemma. \qed

Corollary S1.0.1. Under conditions (C.1) and (C.3),

$$\hat{n}_{rs} = n_{rs}[1 + o_p(1)] \quad \text{for all} \quad 1 \leq r, s \leq K.$$
Lemma S1.2. For any $1 \leq r, s \leq K$, we have the representation:

$$
\left(\hat{\omega}_{rs} - \omega_{rs}\right) = \sqrt{\frac{1 + \mathbb{I}(r = s)}{n_{rs}}} \omega_{rs} \left(1 - \omega_{rs}\right) \cdot Z_{n,rs} + R_{n,rs}
$$

where $R_{n,rs} \leq 4nH(\hat{c}_n, c_n)/\hat{n}_{rs}$ for all $n \geq 1$.

Proof: Write $\tilde{\omega}_{rs} = \sum_{i \neq j} A(i, j) \mathbb{I}(c_i = r, c_j = s)/n_{rs}$. Then, it is easy to check that for $r \neq s$,

$$
(\hat{\omega}_{rs} - \omega_{rs}) = (\tilde{\omega}_{rs} - \omega_{rs}) + (\tilde{\omega}_{rs} - \tilde{\omega}_{rs})
$$

$$
= n_{rs}^{-1} \sum_{i \in C_r, j \in C_s} [A(i, j) - \omega_{rs}] + \sum_{i \in C_r, j \in C_s} A(i, j) \frac{[\hat{n}_{rs} - \hat{n}_{rs}]_{n_{rs}n_{rs}}}{n_{rs}n_{rs}}
$$

$$
+ \hat{n}_{rs}^{-1} \sum_{(i,j) : i \neq j} A(i, j) [\mathbb{I}(\hat{c}_i = r, \hat{c}_j = s) - \mathbb{I}(c_i = r, c_j = s)]
$$

$$
\equiv I_{1n} + I_{2n} + I_{3n}, \text{ say.}
$$

By Lemma S1.1 and the fact that $|A(i, j)| \leq 1$, it follows that $I_{2n} \leq 2nH(\hat{c}_n, c_n)/\hat{n}_{rs}$.

Also, using arguments similar to the proof of Lemma S1.1, one can show that $I_{3n} \leq 2nH(\hat{c}_n, c_n)/\hat{n}_{rs}$. Now set $R_{n,rs} = I_{2n} + I_{3n}$ and $Z_{n,rs} = \left[\omega_{rs}(1 - \omega_{rs})n_{rs}\right]^{1/2} \sum_{i \in C_r, j \in C_s} [A(i, j) - \omega_{rs}]$. Then, by the independence of $\{A(i, j) : 1 \leq i < j \leq n\}$ and the Central Limit Theorem (CLT), the conclusions of the lemma follow for $r \neq s$. The proof of the case $r = s$ is similar. The only notable change is due to the fact for $r = s$, $A(i, j) = A(j, i)$ for all $i \neq j, i, j \in C_r$ and therefore, only half as many variables in the sum are independent. As a result, we write the first term (i.e., $I_{1n}$) as

$$
I_{1n} = n_{rr}^{-1} \sum_{(i,j) : i \neq j} \mathbb{I}(c_i = r = c_j)[A(i, j) - \omega_{rr}]
$$

$$
= 2n_{rr}^{-1} \sum_{(i,j) : i < j} \mathbb{I}(c_i = r = c_j)[A(i, j) - \omega_{rr}]
$$

$$
\equiv \left[2\omega_{rs}(1 - \omega_{rs})/n_{rs}\right]^{1/2} Z_{n,rr},
$$
where, it is now easy to check that $Z_{n,rr}$ satisfies the independence and the asymptotic normality requirements of the lemma. □

**Remark S1.1.** Note that for the expected number of edges to grow with $n$, we must have $a_n n^2 \to \infty$ which we tacitly assumed in (C.2). This, in particular, implies that

\[
\left( \hat{\omega}_{rs} - \omega_{rs} \right) = \sqrt{\frac{[1 + \mathbb{1}(r = s)]\omega_{rs}(1 - \omega_{rs})}{n_{rs}}} \cdot Z_{n,rs}[1 + o_p(1)].
\]

**Remark S1.2.** Here we provide further clarification about the sufficient condition on the sparsity factor $a_n$ for the validity of Condition (C.3) for the algorithm in Gao et al. (2017, 2018). Let $q^0 \equiv \max_{1 \leq r \neq s \leq K} \omega_{rs}^0 < p^0 \equiv \min_{1 \leq r \leq K} \omega_{rr}^0$. Then, the minimax bound on the misclassification rate in Section 2.2 of Gao et al. (2018) yields (in our notation)

\[
EH(\hat{c}_n, c) = O\left( n \exp \left( - C(K) na_n \left[ \sqrt{p^0} - \sqrt{q^0} \right]^2 \right) \right) \quad \text{as } n \to \infty,
\]

for some constant $C(K) \in (0, \infty)$, depending only on $K$ and $\{\pi_r : 1 \leq r \leq K\}$ (and not on $n$). Condition (C.3) now easily follows from this when $na_n \gg \log n$.

**S1.1 Proof of Theorem 3.1**

Note that under $H_0 : P_1 = P_2$,

\[
T_{1n}^2 = a_n^{-1} \| \hat{P}_1 - \hat{P}_2 \|^2_F
= \sum_{(i,j):i \neq j} a_n^{-1} [\hat{P}_1(i,j) - P_1(i,j)]^2 + \sum_{(i,j):i \neq j} a_n^{-1} [\hat{P}_2(i,j) - P_2(i,j)]^2
- 2 \sum_{(i,j):i \neq j} a_n^{-1} [\hat{P}_1(i,j) - P_1(i,j)][\hat{P}_2(i,j) - P_2(i,j)]. \tag{S1}
\]

Consider the first term. Using the fact that for any $i = 1, \ldots, n, \sum_{r=1}^K \mathbb{1}(c_i = r) = 1,$
To prove Theorem 3.2, we will need the following lemma.

where \( \tau \)

we have

Then, by Corollary S1.0.1, Lemma S1.2, and Condition (C.3),

Then, by Corollary S1.0.1, Lemma S1.2, and Condition (C.3),

where \( \tau_{n,rs} = \omega_0^r (1 - a_n \omega_0^{rs}) \) if \( r \neq s \) and \( \tau_{n,rs} = 2 \omega_0^{rs} (1 - a_n \omega_0^{rs}) \) if \( r = s \). Further, by Corollary S1.0.1 and Lemma S1.2,

Now using the identity (S1), the independence of the adjacency matrix \( A_1 \) and \( A_2 \), and relations (S3) and (S4) (and their analogs for \( \hat{P}_2 - P_2 \)), by the continuous mapping theorem, we conclude that

\[
T_{1n}^2 \to^d \sum_{r=1}^{K} \sum_{s=1}^{K} \left( \tau_{1,rs}^0 Z_{1,rs}^0 \right)^2 + \tau_{2,rs}^0 \left( Z_{2,rs}^0 \right)^2 - 2 \sqrt{\tau_{1,rs}^0 \tau_{2,rs}^0 Z_{1,rs}^0 Z_{2,rs}^0},
\]

where \( \tau_{1,rs}^0 = \omega_0^r \) for \( r \neq s \) and \( \tau_{1,rs}^0 = 2 \omega_0^{rs} \) and \( Z_{1,rs}^0 \) is the \( (r,s) \)th component of \( Z_{1}^0 \), and \( \tau_{2,rs}^0, Z_{2,rs}^0 \) are similarly defined for Population 2. This proves the theorem.

\( \square \)

### S1.2 Proof of Theorem 3.2

To prove Theorem 3.2, we will need the following lemma.
Lemma S1.3. Under Conditions (C.1)-(C.3),
\[ \hat{\rho}_k - \rho_k = \sqrt{a_n} \cdot V_k^0 \cdot (1 + o_p(1)) \]
where \( V_k^0 = \frac{\sum_{r=1}^{K} \sum_{s=1}^{K} \pi_{rs}^0 \omega_{rs}^0 Z_{rs}^0}{\sqrt{\sum_{r=1}^{K} \sum_{s=1}^{K} \pi_{rs}^0 [\omega_{rs}^0]^2}} \), \( k = 1, 2 \).

**Proof:** We only consider the case \( k = 1 \). Also, for notational simplicity, we drop the subscript 1 from \( \tau_{1,rs}, \omega_{1,rs}, \ldots \) etc. Note that using Lemma S1.2, Condition (C.2) and the arguments in the derivation of (S4), one can show that
\[ \hat{\rho}_2^2 - \rho_2^2 = \| \hat{P}_1 \|^2_F - \| P_1 \|^2_F \]
\[ = \sum_{r=1}^{K} \sum_{s=1}^{K} \sum_{(i,j); i \neq j} [\hat{P}(i, j)^2 - \omega_{rs}^2] I(c_i = r, c_j = s) \]
\[ = \sum_{r=1}^{K} \sum_{s=1}^{K} n_{rs} [\hat{\omega}_{rs} - \omega_{rs}][\hat{\omega}_{rs} + \omega_{rs}][1 + o_p(1)] \]
\[ = \sum_{r=1}^{K} \sum_{s=1}^{K} n_{rs} \left\{ \frac{\sqrt{a_n}}{n} (\tau_{rs}^0 Z_{rs}^0) \right\} \left[ \sqrt{a_n} \omega_{rs}^0 + O_p \left( \frac{\sqrt{a_n}}{n} \right) \right] \cdot [1 + o_p(1)] \]
\[ = 2n a_n^{3/2} \sum_{r=1}^{K} \sum_{s=1}^{K} \pi_{rs}^0 \tau_{rs} Z_{rs}^0 \omega_{rs}^0 \cdot [1 + o_p(1)]. \]

Since \( \rho_2^2 = \sum_{(i,j); i \neq j} P_1(i, j)^2 = \sum_{r=1}^{K} \sum_{s=1}^{K} \pi_{rs}^0 [\omega_{rs}^0]^2 n^2 a_n^2 (1 + o(1)) \equiv [\rho_2^2]^2 n^2 a_n^2 [1 + o(1)], \)
from Condition (C.2) and the above, it follows that
\[ \hat{\rho}_2^2 = \rho_2^2 + O_p(n a_n^{3/2}) = \rho_2^2 [1 + o_p(1)]. \]

Hence, it follows that
\[ \hat{\rho}_1 - \rho_1 = (\hat{\rho}_1 - \rho_1)(\hat{\rho}_1 + \rho_1) \]
\[ = \frac{2n a_n^{3/2} \left( \sum_{r=1}^{K} \sum_{s=1}^{K} \pi_{rs}^0 \tau_{rs} Z_{rs}^0 \omega_{rs}^0 \right) \cdot [1 + o_p(1)]}{2 \rho_1 (1 + o_p(1))} \]
\[ = \frac{\sqrt{a_n} \left( \sum_{r=1}^{K} \sum_{s=1}^{K} \pi_{rs}^0 \tau_{rs} Z_{rs}^0 \omega_{rs}^0 \right)}{\sqrt{\sum_{r=1}^{K} \sum_{s=1}^{K} \pi_{rs}^0 [\omega_{rs}^0]^2}} \cdot [1 + o_p(1)]. \]

This proves Lemma S1.3. \( \square \)
Proof of Theorem 3.2: Note that under $H_0,$

$$\left\| \frac{\hat{P}_1 - \hat{P}_2}{\hat{\rho}_1} \right\|_F^2 = \left\| \frac{\hat{P}_1 - P_1}{\hat{\rho}_1} + P_2 \left( \left[ \frac{c}{\hat{\rho}_1} - \frac{c}{\hat{\rho}_1} \right] - \left[ \frac{1}{\hat{\rho}_2} - \frac{1}{\hat{\rho}_2} \right] \right) - \frac{\hat{P}_2 - P_2}{\hat{\rho}_2} \right\|_F^2$$

$$= \hat{\rho}_1^{-2}\|\hat{P}_1 - P_1\|_F^2 + \hat{\rho}_2^{-2}\|\hat{P}_2 - P_2\|_F^2 + \|P_2\|^2 \left( \left[ \frac{c}{\hat{\rho}_1} - \frac{c}{\hat{\rho}_1} \right] - \left[ \frac{1}{\hat{\rho}_2} - \frac{1}{\hat{\rho}_2} \right] \right)^2$$

$$+ 2\hat{\rho}_1^{-1} \sum_{(i,j):i\neq j} [\hat{P}_1(i,j) - P_1(i,j)] P_2(i,j) \left( \left[ \frac{c}{\hat{\rho}_1} - \frac{c}{\hat{\rho}_1} \right] - \left[ \frac{1}{\hat{\rho}_2} - \frac{1}{\hat{\rho}_2} \right] \right)$$

$$- 2\hat{\rho}_2^{-1} \sum_{(i,j):i\neq j} [\hat{P}_2(i,j) - P_2(i,j)] P_2(i,j) \left( \left[ \frac{c}{\hat{\rho}_1} - \frac{c}{\hat{\rho}_1} \right] - \left[ \frac{1}{\hat{\rho}_2} - \frac{1}{\hat{\rho}_2} \right] \right)$$

$$- 2[\hat{\rho}_1\hat{\rho}_2]^{-1} \sum_{(i,j):i\neq j} [\hat{P}_1(i,j) - P_1(i,j)][\hat{P}_2(i,j) - P_2(i,j)]$$

$$\equiv I_{1n} + \ldots + I_{6n}, \quad \text{(say).} \quad \text{(S5)}$$

Using arguments in the proof of Theorem 3.1 and Lemma S1.3, one can show that for $k = 1, 2,$

$$I_{kn} = \frac{a_n \sum_r \sum_s \tau_{k,rs}^0 [Z_{k,rs}^0]^2 (1 + o_p(1))}{\hat{\rho}_k^2(1 + o_p(1))}$$

$$= \frac{1}{n^2 a_n} \cdot \frac{\sum_r \sum_s \tau_{k,rs}^0 [Z_{k,rs}^0]^2}{|\rho_k^0|^2} \cdot (1 + o_p(1))$$

$$\equiv \frac{1}{n^2 a_n} \cdot W_k^0 \cdot (1 + o_p(1)), \quad \text{(say).} \quad \text{(S6)}$$

By similar arguments,

$$I_{6n} = \frac{1}{n^2 a_n} \cdot \frac{2 \sum_r \sum_s \sqrt{\tau_{1,rs}^0 \tau_{2,rs}^0 Z_{1,rs}^0 Z_{2,rs}^0}}{\rho_1^0 \rho_2^0} \cdot (1 + o_p(1))$$

$$\equiv \frac{1}{n^2 a_n} \cdot W_6^0 \cdot (1 + o_p(1)), \quad \text{(say).} \quad \text{(S7)}$$

Next consider $I_{3n}.$ Then, using Lemma S1.3 and defining $V_2$ analogously for $P_2,$ it follows that

$$I_{3n} = \rho_2^2 \cdot \left( \left[ \frac{c}{\hat{\rho}_1} - \frac{c}{\hat{\rho}_1} \right] - \left[ \frac{1}{\hat{\rho}_2} - \frac{1}{\hat{\rho}_2} \right] \right)^2$$

$$= \rho_2^2 \cdot \left( \frac{c(\rho_1 - \hat{\rho}_1)}{\rho_1^2(1 + o_p(1))} - \frac{(\rho_2 - \hat{\rho}_2)}{\rho_2^2(1 + o_p(1))} \right)^2$$

$$= \frac{1}{n^2 a_n} \cdot |\rho_2|^2 \cdot \left( \frac{cV_1^0}{|\rho_1|^2} - \frac{V_2^0}{|\rho_2|^2} \right)^2 \cdot (1 + o_p(1))$$

$$\equiv \frac{1}{n^2 a_n} \cdot W_3^0 \cdot (1 + o_p(1)), \quad \text{(say).} \quad \text{(S8)}$$
Also, by Lemmas S1.2 and S1.3, we have for $k = 1, 2$

\[
(\frac{1}{k+1}) I_{(k+3)n} \]

\[
= 2\rho_k^{-1} \sum_{(i,j): i \neq j} \left[ \hat{P}_k(i, j) - P_k(i, j) \right] \hat{P}_2(i, j) \left( \left[ \frac{c}{\rho_1} - \frac{c}{\rho_1} \right] - \left[ \frac{1}{\rho_2} - \frac{1}{\rho_2} \right] \right) \]

\[
= 2\rho_k^{-1} \sum_{r=1}^{K} \sum_{s=1}^{K} n_r^0 n_s^0 \left[ \frac{1}{n_r^0 n_s^0} \left( 1 + \mathbb{I}(r = s) \right) \omega_r^0 \omega_s^0 \right] (1 + o_p(1)) \]

\[
\times \left[ \frac{c(\rho_1 - \hat{\rho}_1)}{(\rho_1^2(1 + o_p(1)))} \right] - \left[ \frac{\rho_2 - \hat{\rho}_2}{(1 + o_p(1))} \right] \]

\[
= -2 \frac{n^2 d_n}{\rho_k^0} \sum_{r=1}^{K} \sum_{s=1}^{K} \sqrt{n_r^0 (1 + \mathbb{I}(r = s)) \omega_r^0 \omega_s^0} \omega_r^0 \omega_s^0 \cdot \left[ \frac{c V_1}{[\rho_1^2]} - \frac{V_2}{[\rho_2^2]} \right] (1 + o_p(1)) \]

\[
\equiv \frac{1}{n^2 d_n} \cdot W_{k+3}^0 \cdot (1 + o_p(1)), \quad \text{(say).} \]

(S9)

Then, from (S5)-(S9), it follows that

\[
T_{2n} \overset{d}{\to} \sqrt{W_1^0 + \ldots + W_6^0}. \quad \text{(S10)}
\]

### S1.3 Proof of Theorem 3.3

It is enough to show that

\[
\sup_{x \geq 0} \left| P_s(T_{1n}^* \leq x) - P(T_{1,\infty} \leq x) \right| = o_p(1).
\]

We proceed as in the proof of Theorem 3.1. Note that

\[
T_{1n}^{*2} = a_n^{-1} \left\| P_1^* - P_2^* \right\|^2 = a_n^{-1} \left\| (P_1^* - \hat{P}) - (P_2^* - \hat{P}) \right\|^2
\]

where $\hat{P} = [\hat{P}_1 + \hat{P}_2]/2$. Restricting attention to $(P_1^* - \hat{P})/\sqrt{a_n}$ first, we note that the bootstrap random variables driving the distribution of $(P_1^* - \hat{P})/\sqrt{a_n}$ are given by

\[
\{ \omega_{1,rs}^* - \bar{\omega}_{1,rs} : 1 \leq r, s \leq K \}
\]

where $\omega_{1,rs}^* = \frac{\sum_{(i,j): c_{rs}^*=c_{rs},i\neq j} A_{i,j}}{\sum_{(i,j): c_{rs}^*=c_{rs},i\neq j}}$ and where $\bar{\omega}_{1,rs} = [\hat{\omega}_{1,rs} + \hat{\omega}_{2,rs}] / 2$. Consider the set $B_n = \{ \hat{c}_{jn} = c_{jn}, j = 1, 2 \}$. Note that $P(B_n^c) \leq P(H(\hat{c}_{1n}, c_{1n}) \geq 1) + P(H(\hat{c}_{2n}, c_{2n}) \geq 1) = o(n^{-1})$, by (C.3). Under $H_0$, the estimated number of communities as well as estimated community memberships coincide on $B_n$. Now conditional on the set $B_n$, repeating the arguments in the proofs of Lemmas S1.1 and S1.2, one can show that on $\{K_1^* = K\}$,

\[
\omega_{1,rs}^* - \bar{\omega}_{1,rs}^* = \sqrt{\frac{1 + \mathbb{I}(r = s)}{\hat{n}_{1,rs}} \bar{\omega}_{1,rs}^*(1 - \bar{\omega}_{1,rs}^*) Z_{1n,rs}^* + R_{1n,rs}^*}. \quad \text{(S11)}
\]
where $Z^*_{1n,rs} = \left[\tilde{\omega}_{1,rs}(1 - \tilde{\omega}_{1,rs})\tilde{n}_{1,rs}\right]^{-1/2} \sum_{i\in\tilde{c}_r,j\in\tilde{c}_s} A_1^*(i,j) - \tilde{\omega}_{1,rs}$ for $r \neq s$, and $Z^*_{1n,rr} = \left[\tilde{\omega}_{1,rr}(1 - \tilde{\omega}_{1,rr})\tilde{n}_{1,rr}/2\right]^{-1/2} \sum_{i\in\tilde{c}_r,j\in\tilde{c}_r,i<j} A_1^*(i,j) - \tilde{\omega}_{1,rr}$ for $r = s$ case, and where $|R^*_{1,n,rs}| \leq 4nH(c^*_{in}, \tilde{c}_n)/n^*_{1,rs}$ for all $r, s$. Note that under $H_0 : P_1 = P_2$, $a_n^{-1}\tilde{\omega}_{rs} \to_P \omega_{1,rs}^0$ for all $r, s$.

Now using a subsequence argument and the multivariate Lindeberg CLT (cf. Chapter 11 of Athreya and Lahiri (2006)), one can show that

$$d_P\left(\left[Z^*_{1,rs}\right]_{1 \leq r, s \leq K}, \left[Z^*_{1,rs}\right]_{1 \leq r, s \leq K}^{A_1, A_2}\right) = o_p(1)$$

where, for random vectors $X, Y$ and a $\sigma$-field $\mathcal{G}$, $d_P(Y, X|\mathcal{G})$ denotes the Prohorov distance between the distribution of $Y$ and the conditional distribution of $X$ given $\mathcal{G}$. Also, using Condition (C.4), it is easy to show that for any $\epsilon > 0$,

$$P_{\epsilon}\left(\max_{1 \leq r, s \leq K} |R^*_{1,n,rs}| > \epsilon\sqrt{a_n}/n\right) = o_p(1).$$

The proof of Theorem 3.3 can now be completed by retracing the arguments in the proof of Theorem 3.1 and using the following lemma. We omit the routine details. \hfill \Box

**Lemma S1.4.** (An extended conditional continuous mapping theorem). For each $n \geq 1$, let $V_n^* \in \mathbb{R}^p, U_n^* \in \mathbb{R}^q$ and $W_n \in \mathbb{R}^r$ be random vectors defined on a probability space $(\Omega, \mathcal{F}, P)$ and let $\mathcal{G}_n$ be a $\sigma$-field such that $W_n$ is $\mathcal{G}_n$-measurable where $p, q, r \geq 1$. Suppose that

(i) $d_P(V^0, V_n^*|\mathcal{G}_n) = o_p(1),$ 

(ii) there exists a nonrandom $u^0 \in \mathbb{R}^q$ such that $P(\|U_n^* - u^0\| > \epsilon|\mathcal{G}_n) = o_p(1)$ for any $\epsilon > 0$, and

(iii) $W_n \to_p w^0$ for some (nonrandom) $w^0 \in \mathbb{R}^r$.

Let $g : \mathbb{R}^{p+q+r} \to \mathbb{R}^s$ be a measurable function such that (a) for any $M > 0$,

$$\sup_{\|x\| \leq M} \|g(x, y, w) - g(x, u^0, w^0)\| \to 0$$

as $y \to u^0$ and $w \to w^0$, and (b) $P(V^0 \in D_g) = 0$, where $D_g$ is the set of discontinuity points of $g(\cdot, u^0, w^0)$. Then,

$$d_P\left(g(V_n^*, U_n^*, W_n), g(V_n^0, u^0, w^0)\right) = o_p(1).$$

(S12)
Proof: Note that (ii) is equivalent to the following seemingly stronger statement (Chapter 3 of Lahiri (2003)):

(ii)': there exists a sequence \( \{ \epsilon_n \} \) with \( \epsilon_n \downarrow 0 \) such that \( P(\|U_n^* - u^0\| > \epsilon_n|\mathcal{G}_n) = o_p(1) \).

We will use (ii)' in place of (ii) in the proof. Using the characterization of convergence in probability in terms of subsequential almost sure convergence, given any subsequence \( \{n_i\} \), it is enough to extract a further subsequence \( \{n_{i_k}\} \subset \{n_i\} \) such that the left side of (S12) converges to zero almost surely along \( \{n_{i_k}\} \). For notational simplicity, we will write \( \{m\} \) for the subsequence \( \{n_{i_k}\} \). Indeed, using (i), (ii)' and (iii), we can extract the subsequence \( \{m\} \subset \{n_i\} \) is such that on a set \( B \in \mathcal{F} \) with \( P(B) = 1 \),

\[
W_m \to u^0, P(\|U_m^* - u^0\| > \epsilon_m|\mathcal{G}_m) = o(1), \quad \text{and} \quad d_P(V^0, V_m) = o(1) \quad \text{(S13)}
\]
as \( m \to \infty \). Given \( \epsilon_0 \in (0, 1) \), there exists \( M_0 \in (1, \infty) \) and \( N_0 \geq 1 \) such that

\[
P(\|V^0\| > M_0 - 1) < \epsilon_0 \quad \text{and} \quad d_P(V^0, V_m^*|\mathcal{G}_m) < \epsilon_0 \quad \text{(S14)}
\]
for all \( m > n_0 \), on \( B \). In particular, by the definition of the Prohorov distance, this implies that on \( B \),

\[
P(\|V_m^*\| > M_0|\mathcal{G}_n) \leq P(\|V^0\| > M_0 - \epsilon_0) + \epsilon_0 \leq 2\epsilon_0
\]

for all \( m > N_0 \). Let \( B_m^* = \{\|V_m^*\| \leq M_0, \|U_m^* - u^0\| \leq \epsilon_m\} \). Also, let \( \iota = \sqrt{-1} \). Then, on \( B \) and for \( m > N_0 \), for any \( t \in \mathbb{R}^s \),

\[
E\left( \exp(\iota t'g(V_n^*, U_n^*, W_n)|\mathcal{G}_n) \right)
= E\left( \exp(\iota t'g(V_n^*, U_n^*, W_n)\mathbb{1}(B_m^*)|\mathcal{G}_n) \right) + E\left( \exp(\iota t'g(V_n^*, U_n^*, W_n)[1 - \mathbb{1}(B_m^*)]|\mathcal{G}_n) \right)
= E\left( \exp(\iota t'g(V_n^*, U_n^*, W_n) - \exp(\iota t'g(V_n^*, u^0, w^0))\mathbb{1}(B_m^*)|\mathcal{G}_n) \right)
+ E\left( \exp(\iota t'g(V_n^*, u^0, w^0)\mathbb{1}(B_m^*)|\mathcal{G}_n) \right)
+ E\left( \exp(\iota t'g(V_n^*, U_n^*, W_n)[1 - \mathbb{1}(B_m^*)]|\mathcal{G}_n) \right)
= E\left( \exp(\iota t'g(V_n^*, u^0, w^0)|\mathcal{G}_n) \right) + R_m, \quad \text{(say)}
\]

where, using (S13) and (S14), on \( B \), one gets

\[
|R_m| \leq 2P(B_m^*|\mathcal{G}_m) + |t|E\left( \left| g(V_n^*, U_n^*, W_m) - g(V_n^*, u^0, w^0) \right| \right| \mathbb{1}(B_m^*)|\mathcal{G}_m) \leq 2\left[ P(\|V_m^*\| > M_0|\mathcal{G}_n) + P(\|U_m^* - u^0\| > \epsilon_m|\mathcal{G}_m) \right]
+ |t| \sup_{\|x\| \leq M_0, \|y\| \leq \epsilon_m, \|z\| \leq \|W_m - u^0\|} \left| g(x, u^0 + y, w^0 + z) - g(x, u^0, w^0) \right|
\leq 4\epsilon_0 + o(1).
\]
Since \( \epsilon_0 \in (0, 1) \) is arbitrary, this implies

\[
E(\exp(it'g(V_n^*, U_n^*, W_n)|\mathcal{G}_n) = E(\exp(it'g(V_0^*, u^0, w^0)|\mathcal{G}_n) + o(1)
\]

as \( m \to \infty \) for all \( t \in \mathbb{R}^s \), on \( B \). Now using condition (b) on \( g \), (S13) and the standard version of the Continuous Mapping Theorem (cf. Chapter 9, Athreya and Lahiri (2006)), one can easily conclude that, on the set \( B \),

\[
E(\exp(it'g(V_n^*, u^0, w^0)|\mathcal{G}_n) = E(\exp(it'g(V_0^*, u^0, w^0)) + o(1) \quad \text{for all } t \in \mathbb{R}^s.
\]

This completes the proof of Lemma S1.4. \( \square \)

**Remark S1.3.** Lemma S1.4 can be used to prove the convergence of functions that are polynomials in \( V_n^* \) and \( U_n^* \) with coefficients that are rational functions of \( W_n \) with a finite limit. In particular, a mixed version of the (Conditional) Slutsky’s Theorem holds:

\[
W_n \odot V_n^* + U_n^* \rightarrow^d w^0 \odot V^0 + u^0, \text{ in probability.}
\]

**Remark S1.4.** Theorem 3.3 establishes that the proposed bootstrap method provides a valid approximation to the null distribution of the test statistic \( T_{1n} \). We can use the bootstrap quantiles to calibrate the test statistic \( T_{1n} \). Specifically, let \( \hat{q}_{1n}(u) \) denote the \( u \)-quantile of the conditional distribution of \( T_{1n} \) given \( A_1, A_2 \) for \( u \in (0, 1) \). Then, we would reject \( H_0 : P_1 = P_2 \) at level \( \alpha_0 \in (0, 1) \) if \( T_{1n} > \hat{q}_{1n}(1 - \alpha_0) \). Since the limit distribution of \( T_{1n} \) under \( H_{0,n} \) is continuous with a positive density on \( (0, \infty) \), it follows that

\[
\hat{q}_{1n}(1 - \alpha_0) \rightarrow_p q_{1,\infty}(1 - \alpha_0)
\]

where \( q_{1,\infty}(u) \) is the \( u \)-quantile of \( T_{1,\infty} \). As a result,

\[
P(T_{1n} > \hat{q}_{1n}(1 - \alpha_0)|H_{0,n})
= P(T_{1n} > q_{1,\infty}(1 - \alpha_0)|H_{0,n}) + o(1)
= P(T_{1,\infty} > q_{1,\infty}(1 - \alpha_0)) + o(1)
= \alpha_0 + o(1), \quad (S15)
\]

and the proposed test attains the desired size \( \alpha_0 \) asymptotically.

**S1.4 Proof of Theorem 3.4**

Let \( \rho_k^0 \) be as defined in the proof of Lemma S1.3, \( k = 1, 2 \) and let \( \gamma_0 = \rho_1^0 / \rho_2^0 \). Using the approximations for \( \hat{\gamma}_n \) from Lemma S1.3, it is easy to check that

\[
\hat{\gamma}_n = \frac{\hat{\rho}_{1n}}{\rho_{2n}} = \frac{\rho_1}{\rho_2}(1 + o_p(1)) \equiv \gamma_0 + o_p(1).
\]
Also, note that using Corollary S1.0.1, Lemma S1.2, the relation above and the identities 
\( \omega_{r,s}^0 = \gamma_0 \omega_{r,s}^0 \) for all \( r, s \) and \( \hat{\rho}_n = \hat{\gamma}_n \rho_2 \) for all \( n \geq 1 \), one gets

\[
\hat{\rho}_{1n}^2 \equiv \| \hat{P}_{1,n} \|^2 = \| \hat{P}_{1,n} + \hat{\gamma}_n \hat{P}_{2,n} \|^2 / 4 \\
= \sum_{(i,j):i \neq j} [\hat{P}_1(i,j)^2 + \hat{\gamma}^2 \hat{P}_2(i,j)^2 + 2\hat{\gamma} \hat{P}_1(i,j) \hat{P}_2(i,j)] / 4 \quad \text{(Dropping the subscript 'n')} \\
= \left[ \hat{\rho}_1^2 + \hat{\gamma}^2 \hat{\rho}_2^2 + 2\hat{\gamma} \sum_{(i,j):i \neq j} \hat{P}_1(i,j) \hat{P}_2(i,j) \right] / 4 \\
= \left[ 2\rho_1^2 + 2\gamma \sum_r \sum_{s} n_{r,s}\omega_{1,r,s}\omega_{2,r,s} \left( 1 + o_p(1) \right) \right] / 4 \\
= \left[ 2(\rho_1^0)^2 n^2 a_n^2 + 2\gamma_0 n^2 a_n^2 \sum_r \sum_{s} n_{r,s} \pi_{r,s}^0 \omega_{1,r,s}^0 \omega_{2,r,s}^0 \left( 1 + o_p(1) \right) \right] / 4 \\
= (\rho_1^0)^2 n^2 a_n^2 \left( 1 + o_p(1) \right). \quad \text{(S16)}
\]

Similarly, \( \hat{\rho}_{2n}^2 \equiv \| \hat{P}_{2,n} \|^2 = (\rho_2^0)^2 n^2 a_n^2 \left( 1 + o_p(1) \right) \). Now retracing the steps in the proof of Theorem 3.2 and using (S17) (and its analog for \( k = 2 \)) and Lemma S1.4, one can complete the proof of Theorem S.9.

\[\square\]

### S2 Proof of Theorem 3.5

First note that, for the Chung-Lu estimate \( \hat{P}^n \) of underlying probability matrix \( P \)

\[
0 \leq \| \hat{P}^n - P \|^2 = \sum_{i,j} (\hat{P}^n(i,j) - P(i,j))^2 \\
= \sum_{i,j} (\hat{\theta}_i \hat{\theta}_j - \theta_i \theta_j)^2 \\
= \sum_{i,j} (\hat{\theta}_i \hat{\theta}_j - \theta_i \theta_j)^2 + \sum_{i,j} \theta_i^2 (\hat{\theta}_j^2 - \theta_j^2) \\
\leq 2 \left[ \sum_{i,j} \theta_i^2 (\hat{\theta}_i - \theta_i)^2 + \sum_{i,j} \theta_i^2 (\hat{\theta}_j - \theta_j)^2 \right] \\
\leq 4n \sum_i (\hat{\theta}_i - \theta_i)^2 \quad \text{since } \hat{\theta}_i, \theta_i < 1 \forall i = 1(1)n
\]

Again note that

\[
E(d_i) = n \theta_i \bar{\theta} - \theta_i^2 = \delta_i \text{ (say) } \forall i = 1(1)n
\]

where \( d_i \) is the degree of the \( i^{th} \) vertex and \( \bar{\theta} = \frac{1}{n} \theta \).

So \( \sum_i \delta_i = (n\bar{\theta})^2 - \sum_i \theta_i^2 \geq (n\bar{\theta})^2 - n\bar{\theta} \) (since \( \theta_i \leq 1 \forall i \)). Since \( n\bar{\theta} \ll (n\bar{\theta})^2 \), \( \exists C \) (a constant) \( > 0 \) with \( \sum_i \delta_i > C(n\bar{\theta})^2 \).

11
For fixed $n$, choose $\epsilon_n = 3\frac{\sqrt{n \log(n)}}{\gamma_n}$

Taking $d_j = \sum_{i \neq j} A(i, j)$ where $A(i, j) \sim \text{Ber}(p(i, j))$ and using Hoeffding concentration inequality, we have

$$\mathbb{P}\left[ \bigcup_i \{|d_i - \delta_i| \geq \epsilon_n\} \right] \leq \sum_i \mathbb{P}\left[ |d_i - \delta_i| \geq \epsilon_n \right]$$

$$\leq \sum_i e^{-\frac{(n/5)(\delta_i/n)(\epsilon_n/n)^2}{2}}$$

$$= \sum_i e^{-\delta_i \epsilon_n^2 / 5n^2}$$

$$= \sum_i e^{-\theta_i \delta_n^2 / 5n}$$

$$\leq ne^{-\gamma_n^2 \epsilon_n^2 / 5n} \quad [\text{since } \theta_i \geq \gamma_n \forall i = 1(1)n]$$

$$= \exp\left\{ - \frac{\gamma_n^2 \epsilon_n^2}{5n} + \log(n) \right\}$$

$$= \exp\left\{ - \frac{4}{5} \log(n) \right\} \to 0$$

Hence under the assumptions,

$$\mathbb{P}\left[ \bigcup_i \{|d_i - \delta_i| \geq \epsilon_n\} \right] \to 0 \text{ as } n \to \infty$$

Observe that

$$\frac{\epsilon_n}{n^{1/2 + \alpha} \theta} = 3 \frac{\sqrt{n \log(n)}}{n^{1/2 + \alpha} \theta \gamma_n}$$

$$= 3 \frac{\sqrt{\log(n)}}{n^{\alpha} \gamma_n \theta}$$

$$\to 0 \quad \text{by assumption (1)}$$

Also observe that

$$\frac{\epsilon_n^2 \sum_i \theta_i^2}{n^{2 + 2\alpha} \theta^4} = 9 \frac{n \log(n)}{\gamma_n^2 n^{1/2 + 2\alpha} \theta^4} \sum_i \theta_i^2$$

$$= 9 \left( \frac{\sum_i \theta_i^2}{n \theta^2} \right) \left( \frac{n \log(n)}{n^{1/2 + 2\alpha} \gamma_n \theta^2} \right)$$

$$= 9 \left( \frac{\sum_i \theta_i^2}{n \theta^2} \right) \left( \frac{\sqrt{n \log(n)}}{n^{1/2 + 2\alpha} \gamma_n \theta} \right)^2$$

Hence from assumption 1, 2

$$\frac{\epsilon_n^2 \sum_i \theta_i^2}{n^{2 + 2\alpha} \theta^4} = o(1)$$

Hence, on the asymptotically $\mathbb{P}$robability-1 set $\bigcap_i \left\{|d_i - \delta_i| < \epsilon_n\right\}$, observe that
\[ \sum_i d_i = (n\bar{\theta})^2 - \sum_i \theta_i^2 + nk_n \text{ and hence } \sqrt{\sum_i d_i} = O\left(n\bar{\theta} + \sqrt{nk_n}\right) \text{ where } k_n \leq O(\epsilon_n) \] and so \( \sum_i d_i = O((n\bar{\theta})^2) \) since \( \theta_i \leq 1 \) and so \( \sum_i \theta_i^2 \leq \sum_i \theta_i = n\bar{\theta} \).

Hence, on the above mentioned set,

\[ |\hat{\theta}_i - \theta_i|^2 \leq \left| \frac{d_i}{\sqrt{\sum_i d_i}} - \theta_i \right|^2 \]

\[ \leq 3\left[ \frac{\epsilon_n^2}{\sum_i d_i} + \frac{3\delta_i^2}{\sqrt{\sum_i d_i}} \right] \]

\[ \leq 3\left[ \frac{\epsilon_n^2}{\sum_i d_i} + \frac{3\delta_i^2}{\sqrt{\sum_i d_i}} \right] + \frac{3}{\sqrt{C}} \left( \frac{n\theta_i\bar{\theta} - \theta_i^2}{n\bar{\theta}} - \theta_i \right)^2 \text{ [since } \sum_i \delta_i \geq C(n\bar{\theta})^2] \]

\[ \leq 3\left[ \frac{\epsilon_n^2}{\sum_i d_i} + \frac{3\delta_i^2}{\sqrt{\sum_i d_i}} \right] + \frac{3}{\sqrt{C}} \left( \frac{n\theta_i\bar{\theta} - \theta_i^2}{n\bar{\theta}} - \theta_i \right)^2 \text{ [since } n\theta_i\bar{\theta} \geq \theta_i^2] \]

\[ \leq 3\left[ \frac{\epsilon_n^2}{\sum_i d_i} + \frac{3\theta_i^2}{\sqrt{\sum_i d_i}} \right] + \frac{3}{\sqrt{C}} \left( \frac{n\theta_i\bar{\theta} - \theta_i^2}{n\bar{\theta}} - \theta_i \right)^2 \text{ [since } \sum_i \delta_i \geq C(n\bar{\theta})^2] \]

\[ \text{(S22)} \]

Hence

\[ T_n^2 = \| \hat{P}^n - P \|_F^2 \leq 4n \sum_i (\hat{\theta}_i - \theta_i)^2 \]

\[ \leq 12n^2 \frac{\epsilon_n^2}{\sum_i d_i} + \frac{12n}{C} \left( \frac{\sum_i \bar{\theta}_i - \sqrt{\sum_i d_i}}{\sqrt{\sum_i d_i}} \right)^2 \sum_i \theta_i^2 + \frac{12n}{\sqrt{C} (n\bar{\theta})^2} \sum_i \theta_i^4 \]

\[ \text{(S23)} \]

Again note that

\[ \left( \frac{\sum_i \bar{\theta}_i - \sqrt{\sum_i d_i}}{\sqrt{\sum_i d_i}} \right)^2 = \frac{(\sum_i \delta_i - \sum_i d_i)^2}{(\sum_i \delta_i + \sqrt{\sum_i d_i})^2} \]

\[ \leq \frac{(n\epsilon_n)^2}{\sum_i \delta_i + \sum_i d_i} \quad \text{[since } (a + b)^2 \geq a^2 + b^2] \]

\[ = O\left(\frac{(n\epsilon_n)^2}{(n\bar{\theta})^2}\right) \quad \text{[since } \sum_i \delta_i = O((n\bar{\theta})^2), \sum_i d_i = O((n\bar{\theta})^2)] \]

\[ = O\left(\frac{\epsilon_n^2}{\theta^2}\right) \]

Hence putting in (S23)

\[ T_n^2 = \| \hat{P}^n - P \|_F^2 \leq O\left( n^2 \frac{\epsilon_n^2}{\sum_i d_i} + n \frac{\epsilon_n^2}{\theta^2} \sum_i d_i + \frac{n}{(n\bar{\theta})^2} \sum_i \theta_i^4 \right) \]

\[ = O\left( n^2 \frac{\epsilon_n^2}{(n\bar{\theta})^2} + n \frac{\epsilon_n^2}{\theta^2(n\bar{\theta})^2} + \frac{n}{(n\bar{\theta})^2} \sum_i \theta_i^4 \right) \]

(S24)
And hence, for any \( \alpha > 0 \),

\[
\frac{1}{n^{1+2\alpha}} T_n^2 = \frac{1}{n^{1+2\alpha}} \| \hat{P}^n - P \|_F^2 \leq O \left( n^{-2\alpha} \left[ \frac{\epsilon_n^2}{n^2} + \frac{\epsilon_n^2}{n^2 \theta^2} \sum_i \theta_i^2 + \frac{1}{(n \theta^2)^2} \sum_i \theta_i^4 \right] \right) \tag{S25}
\]

\[
= o(1) \quad \text{[by (S20) and (S21)]}
\]

Also note that the rejection region of the test is of the form

\[
R = \{ T_n > c_n \}
\]

So, we can write

\[
\mathbb{P}(T_n \notin R) \leq \mathbb{P}(\| \hat{P}_1^n - \hat{P}_2^n \|_F \leq c_n)
\]

\[
\leq \mathbb{P}(\| P_1 - P_2 \|_F - \| \hat{P}_1^n - P_1 \|_F - \| \hat{P}_2^n - P_2 \|_F \leq c_n) \tag{S26}
\]

\[
= \mathbb{P}(\| \hat{P}_1^n - P_1 \|_F + \| \hat{P}_2^n - P_2 \|_F + c_n \geq \| P_1 - P_2 \|_F)
\]

For any \( \alpha > 0 \), taking \( c_n = O(n^{1/2+\alpha}) \) the LHS inside the probability expression above is \( o(1) \), but under alternative, \( P_1 \) and \( P_2 \) are significantly apart in the sense that

\[
\frac{\| P_1 - P_2 \|_F}{n^{1/2+\alpha}} \to \infty \text{ for any } \alpha > 0
\]

(by assumption). Hence

\[
\mathbb{P}(T_n \notin R) \leq \beta_n
\]

for small \( \beta_n \to 0 \) and so

\[
\mathbb{P}(T_n \in R) \geq 1 - \beta_n \to 1.
\]

Hence the test is consistent.

### S3 Proof of Theorem 3.6

Suppose that the null hypothesis \( H_0 \) is true, so \( P_1 = P_2 \). Let \( \alpha \) be given, and let \( \eta < \alpha/4 \). We consider the latent positions \( X_1 \) and \( X_2 \) corresponding to \( P_1 \) and \( P_2 \). From Theorem 1 in Tang et al. (2017a), for all \( n \) sufficiently large, there exists orthogonal matrices \( W_1 \) and \( W_2 \in \mathbb{O}(d) \) such that with probability at least \( 1 - \eta \),

\[
\| \hat{X}_i - X_i \|_F - C(X_i) \leq \frac{Cdlog(n/\eta)}{C(X_i)\sqrt{\gamma^2(X_i)\delta(P_i)}} \Delta g(X_i) \quad \text{for } i = 1, 2,
\]
where $\tilde{X}_i = X_iW_i$ for $i = 1, 2$. Additionally, there exists a constant $K_0 > 0$ such that $K_0 < C(X_i) < (d\gamma^{-1}(P_i))^{1/2}$ for $i = 1, 2$. Now,

$$g(X_i) = \frac{Cd\log(n/\eta)}{C(Z_n)\sqrt{\beta}(P_i)\delta(P_i)} < \frac{Cd\log(n/\eta)}{K_0\sqrt{d\log n}^{2+\epsilon}} \to 0 \quad \text{as} \quad n \to \infty.$$ 

Let, the estimates of $P_1$ and $P_2$ be $\hat{P}_1 = \tilde{X}_1\tilde{X}_1^T$ and $\hat{P}_2 = \tilde{X}_2\tilde{X}_2^T$. Then for $i = 1, 2$,

$$\|\hat{P}_i - P_i\|_F = \|\tilde{X}_i\tilde{X}_i^T - \tilde{X}_i\tilde{X}_i^T\|_F$$

$$= \|\tilde{X}_i\tilde{X}_i^T - \tilde{X}_i\tilde{X}_i^T + \tilde{X}_i\tilde{X}_i^T - \tilde{X}_i\tilde{X}_i^T\|_F$$

$$\leq \|\tilde{X}_i(\tilde{X}_i - \tilde{X}_i)\|_F + \|(\tilde{X}_i - \tilde{X}_i)\tilde{X}_i^T\|_F$$

$$\leq \|\tilde{X}_i\|_F\|\tilde{X}_i - \tilde{X}_i\|_F + \|\tilde{X}_i\|_F\|\tilde{X}_i - \tilde{X}_i\|_F$$

$$= \sqrt{r_i}\|\tilde{X}_i - \tilde{X}_i\|_F + \|\tilde{X}_i\|_F\|\tilde{X}_i - \tilde{X}_i\|_F$$

where $r_i = \text{trace}(P_i)$.

Now, by the Theorem 2.1. from Tang et al. (2017a), both $C(X_1)$ and $C(X_2)$ are bounded above by $(d\gamma^{-1}(P_1))^{1/2}$ and $(d\gamma^{-1}(P_2))^{1/2}$ respectively. Then,

$$\|\tilde{X}_i\|_F \leq \|\tilde{X}_i\|_F + \|\tilde{X}_i - \tilde{X}_i\|_F$$

$$\leq \|\tilde{X}_i\|_F + C(X_i) + g(X_i)$$

$$\leq \sqrt{r_i} + (d\gamma^{-1}(P_i))^{1/2} + g(X_i)$$

$$\leq 2\sqrt{r_i} + (d\gamma^{-1}(P_i))^{1/2} \quad (g(X_i) \leq \sqrt{r_i} \text{ for large } n)$$

Hence,

$$\|\hat{P}_1 - P_1\|_F = \|\tilde{X}_1\tilde{X}_1^T - \tilde{X}_1\tilde{X}_1^T\|_F$$

$$\leq (3\sqrt{r_i} + (d\gamma^{-1}(P_i))^{1/2})\|\tilde{X}_1 - \tilde{X}_1\|_F \quad \text{(S28)}$$

Let, $\Gamma = \Gamma_1 + \Gamma_2$. Then,

$$\frac{1}{\Gamma}\|\hat{P}_1 - \hat{P}_2\|_F = \frac{1}{\Gamma}\|(\hat{P}_1 - P_1) + (P_2 - \hat{P}_2)\|_F \quad \text{(as } P_1 = P_2)$$

$$\leq \frac{1}{\Gamma}\|\hat{P}_1 - P_1\|_F + \|\hat{P}_2 - P_2\|_F$$

$$\leq \frac{1}{\Gamma_1}\|\hat{P}_1 - P_1\|_F + \frac{1}{\Gamma_2}\|\hat{P}_2 - P_2\|_F$$

$$\leq \|\tilde{X}_1 - \tilde{X}_1\|_F + \|\tilde{X}_2 - \tilde{X}_2\|_F \quad \text{(By Equation (S28))}$$

$$\leq C(X_1) + C(X_2) + g(X_1) + g(X_2)$$

15
Hence, with probability at least $1 - \alpha$,

$$\frac{\|\hat{P}_1 - \hat{P}_2\|_F}{\Gamma(\sqrt{d \gamma^{-1}(P_1)} + \sqrt{d \gamma^{-1}(P_2)})} \leq 1 + r(\alpha, n)$$

where $r(\alpha, n) \to 0$ as $n \to \infty$ for a fixed $\alpha$.

Hence,

$$T_n = \|\hat{P}_1 - \hat{P}_2\|_F \leq \Gamma(1 + r(\alpha, n))(\sqrt{d \gamma^{-1}(P_1)} + \sqrt{d \gamma^{-1}(P_2)})$$

$$\implies T_n = \|\hat{P}_1 - \hat{P}_2\|_F \leq 2\Gamma(\sqrt{d \gamma^{-1}(P_1)} + \sqrt{d \gamma^{-1}(P_2)})$$

with probability at least $1 - \alpha$.

Then for $P_1, P_2$ satisfying $P_1 = P_2$, we conclude

$$P(T_n \in R) < \alpha$$

where $R = \{t : t > 2\Gamma(\sqrt{d \gamma^{-1}(P_1)} + \sqrt{d \gamma^{-1}(P_2)})\}$

Now suppose the alternative hypothesis is true. We note that,

$$\|P_1 - P_2\|_F = \|(P_1 - \hat{P}_1) + (\hat{P}_1 - \hat{P}_2) + (\hat{P}_2 - P_2)\|_F$$

$$\leq \|P_1 - \hat{P}_1\|_F + \|\hat{P}_1 - \hat{P}_2\|_F + \|\hat{P}_2 - P_2\|_F$$

$$\implies \|\hat{P}_1 - \hat{P}_2\|_F \geq \|P_1 - P_2\|_F - \|P_1 - \hat{P}_1\|_F - \|P_2 - \hat{P}_2\|_F$$

Therefore, for all $n$,

$$P(T_n \notin R) \leq P(\|\hat{P}_1 - \hat{P}_2\|_F \leq C)$$

$$\leq P(\|P_1 - \hat{P}_1\|_F + \|P_2 - \hat{P}_2\|_F + C \geq \|P_1 - P_2\|_F)$$

Now, let $\beta > 0$ be given. By the convergence of $\|\hat{X}_1 - \tilde{X}_1\|_F$ to $C(X_1)$ in Theorem 1 in Tang et al. (2017a), we deduce that there exists a constant $M_1(\beta)$ and a positive integer $n_0 = n_0(\alpha, \beta)$ so that, for all $n \geq n_0(\alpha, \beta)$,

$$P\left(\|\hat{X}_1 - \tilde{X}_1\|_F + k\sqrt{d \gamma^{-1}(P_1)} \geq M_1/2\right) \leq \beta/2$$

$$P\left(\|\hat{X}_2 - \tilde{X}_2\|_F + k\sqrt{d \gamma^{-1}(P_2)} \geq M_1/2\right) \leq \beta/2$$
By Equation (S28)

\[
P\left(\|\hat{P}_1 - P_1\|_F + k\Gamma\sqrt{d\gamma^{-1}(P_1)} \geq \Gamma M_1/2\right) \leq \beta/2
\]

\[
P\left(\|\hat{P}_2 - P_2\|_F + k\Gamma\sqrt{d\gamma^{-1}(P_2)} \geq \Gamma M_1/2\right) \leq \beta/2
\]

where \(C = k\Gamma(\sqrt{d\gamma^{-1}(P_1)} + \sqrt{d\gamma^{-1}(P_2)})\).

As \(d_n \rightarrow \infty\) there exists some \(n_2 = n_2(\alpha, \beta, C)\) such that for all \(n \geq n_2\), \(\|P_1 - P_2\|_F \geq \Gamma M_1\).

Hence, for all \(n \geq n_2\), \(P(T_n \notin R) \leq \beta\), i.e., our test statistic \(T_n\) lies within the rejection region \(R\) with probability at least \(1 - \beta\), as required.

**Proof of Theorem 3.7**

Let \(\hat{\rho}\) be the analogue of \(\rho\) corresponding to \(\hat{P}\)

Define \(Q = \frac{P}{\hat{\rho}}\), hence \(\hat{Q} = \frac{\hat{P}}{\hat{\rho}}\), i.e.

\[
\hat{Q}(i, j) = \frac{\hat{P}(i, j)}{\hat{\rho}} = n\frac{\hat{P}(i, j)}{\sqrt{\sum_{i,j}(\hat{P}(i, j))^2}}
\]

So, for any \(\alpha > 0\),

\[
\|\hat{Q} - Q\|_F = \|\frac{\hat{P}}{\hat{\rho}} - \frac{P}{\rho}\|_F
\]

\[
= n\left\| \frac{\hat{P}}{\|\hat{P}\|_F} - \frac{P}{\|P\|_F} \right\|_F
\]

\[
\leq n\left[\left\| \frac{\hat{P}}{\|\hat{P}\|_F} - \frac{P}{\|P\|_F} \right\|_F + \left\| P\right\|_F \left| \frac{1}{\|\hat{P}\|_F} - \frac{1}{\|P\|_F} \right| \right] \quad \text{By triangle inequality}
\]

\[
= n\left[\left\| \frac{\hat{P}}{\|\hat{P}\|_F} - \frac{P}{\|P\|_F} \right\|_F + \frac{\left\| P\right\|_F - \left\| \hat{P}\right\|_F}{\|P\|_F} \right]
\]

\[
\leq 2n\left\| \frac{\hat{P}}{\|\hat{P}\|_F} - \frac{P}{\|P\|_F} \right\|_F \quad \text{By triangle inequality}
\]

\[
\leq O\left(\frac{n^{3/2+\alpha}}{\|P\|_F}\right)
\]

The last line of the above equation follows from the calculation of previous (equality) case.
Now observe
\[ \|\hat{P}\|_F = \|P - (P - \hat{P})\|_F \]
\[ \geq \|P\|_F - \|\hat{P} - P\|_F \quad \text{By triangle inequality} \]
\[ \geq (\sum \theta_i^2) - o(n^{1/2+\alpha}) \]
\[ \geq n\theta^2 - o(n^{1/2+\alpha}) \quad \text{since RMS} \geq AM \]
\[ = n\theta^2 \]

Also note that \( \bar{\theta} \geq \gamma_n \) since \( \theta_i \geq \gamma_n \) \( \forall \ i \). So from assumption (1), we get
\[ o(1) = \frac{\sqrt{\log(n)}}{n^{\alpha}\gamma_n \bar{\theta}} \geq \frac{\sqrt{\log(n)}}{n^{\alpha} \bar{\theta}^2} \]. Hence \( \frac{1}{\bar{\theta}^2} \leq o\left(\frac{n^{\alpha}}{\sqrt{\log(n)}}\right) \)

So \( \|Q - \hat{Q}\|_F \leq O\left(\frac{n^{1/2+\alpha}}{\bar{\theta}^2}\right) \leq o\left(\frac{n^{1/2+\alpha}\sqrt{\log(n)}}{n^{\alpha}}\right) = o(n^{1/2} \sqrt{\log(n)}) \leq o(n^{1/2+\alpha}) \) for any \( \alpha > 0 \)

Hence proceeding for the probability calculation part as for the equality case, one can show that the test is consistent and hence the theorem follows.

**Proof of Theorem 3.8**

The proof of this result is almost identical to that of Theorem 3.6. We only describe here the necessary modifications. Let \( \alpha \) be given and let \( \eta = \alpha/4 \). From Theorem 2.1 of Tang et al. (2017a), for sufficiently large \( n \), there exists some orthogonal matrices \( W_1, W_2 \in \mathbb{O}(d) \) such that, with probability \( 1 - \eta \),

\[ \|\hat{X}_1 - \bar{X}_1\|_F \leq C(X_1) + g(X_1) \]
\[ \|\hat{X}_2 - \bar{X}_2\|_F \leq C(X_2) + g(X_2) \]

where \( \bar{X}_1 = X_1W_1 \) and \( \bar{X}_2 = X_2W_2 \); \( g(X_1) \to 0 \) and \( g(X_2) \to 0 \). Now, for \( i = 1, 2 \)

\[ \frac{1}{\|\hat{P}_i\|_F} \hat{P}_i - \frac{1}{\|P_i\|_F} P_i \|_F = \|\hat{Z}_i \hat{Z}_i^T - \hat{Z}_i \hat{Z}_i^T\|_F \]

where \( \hat{Z}_i = \frac{\hat{X}_i}{\|X_i\|_F}, \hat{Z}_i = \frac{\bar{X}_i}{\|X_i\|_F}, \bar{X}_1 = XW \) for any \( W \in \mathbb{O}(d) \)
\[ \| \hat{Z}_i \hat{Z}_i^T - \tilde{Z}_i \tilde{Z}_i^T \|_F \leq \| \hat{Z}_i \|_F \| \hat{Z}_i - \tilde{Z}_i \|_F + \| \tilde{Z}_i \|_F \| \hat{Z}_i - \tilde{Z}_i \|_F \]

\[ = 2 \| \hat{Z}_i - \tilde{Z}_i \|_F \] (Since \( \| \hat{Z}_i \|_F = \| \tilde{Z}_i \|_F = 1 \))

\[ = 2 \frac{\| \hat{X}_i - \tilde{X}_i \|_F}{\| \hat{X}_i \|_F} \]

\[ \leq 2 \frac{\| \hat{X}_i - \tilde{X}_i \|_F}{\| \hat{X}_i \|_F} + 2 \frac{\| \tilde{X}_i \|_F}{\| \hat{X}_i \|_F} \left| 1 - \frac{1}{\| \hat{X}_i \|_F} \right| \]

\[ \leq 2 \frac{\| \hat{X}_i - \tilde{X}_i \|_F}{\| \hat{X}_i \|_F} + 2 \frac{\| \tilde{X}_i \|_F}{\| \hat{X}_i \|_F} \frac{\| \hat{X}_i \|_F}{\| \tilde{X}_i \|_F} \]

\[ \leq 4 \frac{\| \hat{X}_i - \tilde{X}_i \|_F}{\| \hat{X}_i \|_F} \]

\[ \leq 4 \frac{C(X_i) + g(X_i)}{\| \hat{X}_i \|_F} \]

with probability at least \( 1 - \eta \).

Under the null hypothesis, \( P_1 = cP_2 \) for some \( c > 0 \), and by construction, \( \rho_1 = c\rho_2 \).

Then,

\[ \left\| \frac{1}{\rho_1} \hat{P}_1 - \frac{1}{\rho_2} \hat{P}_2 \right\|_F \leq 4 \frac{C(X_1) + g(X_1)}{\| \hat{X}_1 \|_F} + 4 \frac{C(X_2) + g(X_2)}{\| \hat{X}_2 \|_F} \]

Hence, for sufficiently large \( n \),

\[ \left\| \frac{1}{\rho_1} \hat{P}_1 - \frac{1}{\rho_2} \hat{P}_2 \right\|_F \leq \frac{4\sqrt{d^{-1}(A_1)/\| \hat{X}_1 \|_F} + 4\sqrt{d^{-1}(A_2)/\| \hat{X}_2 \|_F}}{4\sqrt{d^{-1}(A_1)/\| \hat{X}_1 \|_F} + 4\sqrt{d^{-1}(A_2)/\| \hat{X}_2 \|_F}} \leq 1 + r(\alpha, n) \]

where \( r(\alpha, n) \to 0 \) as \( n \to \infty \) for a fixed \( \alpha \). Then,

\[ T_n = \left\| \frac{1}{\rho_1} \hat{P}_1 - \frac{1}{\rho_2} \hat{P}_2 \right\|_F \leq 8\left( \sqrt{d^{-1}(A_1)/\rho_1} + \sqrt{d^{-1}(A_2)/\rho_2} \right) \]

Then for \( P_1, P_2 \) satisfying \( P_1 = cP_2 \), we conclude

\[ P(T_n \in R) < \alpha \]

where \( R = \{ t : t > 8\left( \sqrt{d^{-1}(A_1)/\rho_1} + \sqrt{d^{-1}(A_2)/\rho_2} \right) \} \).

The proof of consistency proceeds in an almost identical manner to that in Theorem 3.6, and we omit the details.
S4 Comparison with current methods

Here, we compare the proposed methodology with the current state-of-the-art methods from Tang et al. (2017a) and Ghoshdastidar and von Luxburg (2018). In Tang et al. (2017a), the authors proposed methods for both equality and scaling problems, whereas Ghoshdastidar and von Luxburg (2018) and Levin et al. (2017) only considered the equality problem. This section will focus on the equality problem for brevity, as the insights drawn therein have analogous implications for the scaling test. The ensuing comparison draws from theoretical and conceptual considerations, while a more in-depth simulation-based assessment is in Section 4.

• In Tang et al. (2017a), the authors studied the problem under the RDPG model, proposing

\[ T_{\text{ase}}(A_1, A_2) = \min_{W \in O_n} \|\hat{X}_1 - \hat{X}_2W\|_F, \]  

(S30)

as the test statistic. Here \( \hat{X}_1 \) and \( \hat{X}_2 \) are estimated from \( A_1 \) and \( A_2 \) by using ASE and \( O_n \) is the set of all \( n \)-by-\( n \) orthogonal matrices. Two independent sets of parametric bootstrap iterations are carried out, using the generative models \( \hat{X}_1\hat{X}_1^T \) and \( \hat{X}_2\hat{X}_2^T \), and the p-value is defined as the larger of the two bootstrap p-values. The proposed methodology offers several theoretical and computational advantages compared to the \( T_{\text{ase}} \) method: (i) taking the larger of the two p-values makes the \( T_{\text{ase}} \) test overly conservative, with a nominal type-I error of \( \alpha^2 \) instead of \( \alpha \). In contrast, the proposed test has the correct type-I error calibration; (ii) while the \( T_{\text{ase}} \) test is confined to the RDPG model, the proposed test is versatile as it applies to various statistical models; and (iii) from a computational perspective, the proposed method is cheaper since it requires a single set of bootstrap iterations, compared to two sets for the \( T_{\text{ase}} \) test.

• In Levin et al. (2017), the authors proposed the Omnibus embedding method. Given two adjacency matrices \( A_1 \) and \( A_2 \), the omnibus matrix is given by

\[
M = \begin{pmatrix}
A_1 & \frac{A_1 + A_2}{2} \\
\frac{A_1 + A_2}{2} & A_2.
\end{pmatrix}
\]

Next, let \( S_M \) represent the diagonal matrix of the top \( d \) eigenvalues of \( M \), and let \( U_M \) be the \( (m + n) \times d \) matrix of their eigenvalues. The omnibus embedding is defined as \( \hat{X}_M U_M S_M^{1/2} \), and the test statistic is given by

\[ T_{\text{omni}}(A_1, A_2) = ||\hat{X}_M1 - \hat{X}_M2||_F, \]
where $\hat{X}_{M1}$ is the $n \times d$ matrix consisting of rows 1, \ldots, $n$ of $\hat{X}_M$ and $\hat{X}_{M2}$ is the $n \times d$ matrix consisting of rows $n + 1, \ldots, 2n$ of $\hat{X}_M$.

Note that $T_{\text{omni}}$ further exacerbates the computational issues described in the context of $T_{\text{ase}}$, since it requires spectral decomposition of a $2n \times 2n$ matrix. Furthermore, Levin et al. (2017) did not provide any theoretical consistency results for the test.

- In Ghoshdastidar and von Luxburg (2018), the authors studied the problem under the generic inhomogeneous Erdős-Rényi model. Given $A_1 \sim P_1$ and $A_2 \sim P_2$, they propose estimating $\hat{P}_1$ and $\hat{P}_2$ as stochastic blockmodel approximations with $r$ communities (Lovász, 2012; Olhede and Wolfe, 2014; Zhang et al., 2017). Then, they consider the differenced adjacency matrix, $C$, and construct a scaled version given by

$$\tilde{C}(i, j) = \frac{A_1(i, j) - A_2(i, j)}{\sqrt{(n - 1) \left( \hat{P}_1(i, j)(1 - \hat{P}_1(i, j)) + \hat{P}_2(i, j)(1 - \hat{P}_2(i, j)) \right)}}.$$ 

The test statistic is given by

$$T_{\text{eig}}(A_1, A_2) = n^{2/3} (||\tilde{C}|| - 2), \quad (S31)$$

where $|| \cdot ||$ denotes the spectral norm of a matrix. The test is rejected for large values of $T_{\text{eig}}$ with p-values computed from the standard Tracy-Widom distribution (Tracy and Widom, 1996).

The $T_{\text{eig}}$ test has two advantages: first, it is computationally efficient since no bootstrapping is needed, and second, it is highly versatile as it does not require model specification. However, these advantages come with some associated drawbacks. For the $T_{\text{eig}}$ test the number of communities, $r$, needs to be provided as an input to the algorithm. The authors did not offer a strategy or recommendation for determining this crucial tuning parameter, and, in practice, the results vary quite a bit for different choices of $r$. Furthermore, the asymptotic thresholds do not work very well for finite samples. Finally, the largest singular value may not be sufficiently sensitive to the difference between the null model and the alternative model, making their test less effective. These issues are empirically demonstrated in the rest of this section.

### S4.1 Distributional comparison of test statistics

Consider a point null and a point alternative, and let $F_0$ and $F_1$ be the corresponding distributions of some test statistic. Then, the effectiveness of the test statistic is intrinsically
linked to the separation between \( F_1 \) and \( F_0 \). If the separation between \( F_0 \) and \( F_1 \) for one test statistic is greater than that for another test statistic, then the former is more likely to distinguish the null and the alternative. This gives us a natural means to compare different test statistics, by fixing a null model and an alternative model and comparing the separation between \( F_0 \) and \( F_1 \). Based on this notion, we carried out a comparative analysis of the test statistics \( T_{\text{frob}} \) (the proposed method), \( T_{\text{ase}} \), and \( T_{\text{eig}} \) using simulated network data. We deliberately used model configurations from Tang et al. (2017a) and Ghoshdastidar and von Luxburg (2018) for this analysis. We leave out the \( T_{\text{omni}} \) method from this analysis since (Levin et al., 2017) did not provide theoretical consistency results for this test. Let \( A_1 \sim P_1 \) and \( A_2 \sim P_2 \) where \( P_1, P_2 \) are two-community stochastic blockmodels with 
\[ P_1(i, j) = B(c_i, c_j), \quad \text{and} \quad P_2(i, j) = B_{\epsilon}(c_i, c_j), \]
where 
\[
B = \begin{pmatrix} 0.5 & 0.2 \\ 0.2 & 0.5 \end{pmatrix}, \quad \text{and} \quad B_{\epsilon} = \begin{pmatrix} 0.5 + \epsilon & 0.2 \\ 0.2 & 0.5 + \epsilon \end{pmatrix}.
\]
The community membership vector \((c_1, \ldots, c_n) \in \{1, 2\}^n\) was sampled from the multinomial distribution with \( \pi = (0.4, 0.6) \). Thus, \( P_1 \) and \( P_2 \) share the same community assignment, but have different block probability matrices when \( \epsilon \neq 0 \). We used model parameters \( \epsilon = 0, 0.5, 0.1 \) along with \( n = 100, 200, 300 \). Note that the null hypothesis \( H_0 : P_1 = P_2 \) is satisfied when \( \epsilon = 0 \), and \( H_1 \) is satisfied when \( \epsilon = 0.05, 0.1 \). For each combination of \( \epsilon \) and \( n \), we carried out 100,000 Monte Carlo simulations to ensure that the empirical distributions are accurate proxies for the nominal distributions.

In the interest of space, out of the six scenarios studied, we report results from two scenarios in Figure 7: (a) \( \epsilon = 0 \) vs. \( \epsilon = 0.05 \) with \( n = 300 \), and (b) \( \epsilon = 0 \) vs. \( \epsilon = 0.1 \) with \( n = 200 \). We observe that under both scenarios, \( F_0 \) and \( F_1 \) for \( T_{\text{frob}} \) (the proposed method) are much better separated than \( T_{\text{ase}} \) and \( T_{\text{eig}} \). This implies \( T_{\text{frob}} \) is much more sensitive to the difference between \( H_0 \) and \( H_1 \) than \( T_{\text{ase}} \) and \( T_{\text{eig}} \), and therefore more effective at resolving the hypothesis test. We note that a more formal and rigorous comparison between the test statistics can be carried out by deriving the asymptotic relative efficiencies, and we consider this as a future research direction (Rothe et al., 1981).
Figure 7: Histograms of $T_{frob}$ (top), $T_{ase}$ (middle), and $T_{eig}$ (bottom) for (left) $\epsilon = 0$ vs. $\epsilon = 0.05$ with $n = 300$, and (right) $\epsilon = 0$ vs. $\epsilon = 0.1$ with $n = 200$. The light-colored histogram shows the null sampling distribution or $F_0 (\epsilon = 0)$ and the dark-colored histogram shows the alternative sampling distribution or $F_1 (\epsilon > 0)$. We observe that $F_1$ is well separated from $F_0$ for $T_{frob}$ but not for $T_{ase}$ or $T_{eig}$, which means the proposed test is more effective than the current tests.

S5 Simulation studies

S5.1 Sparse RDPG

In this simulation, we consider a comparison between sparse networks. The probability matrix $P_1$ and $P_2$ is defined similarly as in Case 1, and then we set $P_1^* = n^{-0.25} P_1$ and $P_2^* = n^{-0.25} P_2$. We compute the rejection rates based on $P_1^*$ and $P_2^*$ for the network sizes $n = 300, 400, 1000, 2000$ and $\epsilon = 0, 0.025, 0.05, 0.075, 0.1, 0.2$, and the results are displayed in Figure 8. We can see that the proposed test $T_{frob}$ performs better than $T_{ase}$ or $T_{eig}$. The method $T_{omni}$ works better than $T_{frob}$ for smaller networks, but both the $T_{frob}$ and $T_{omni}$ works similarly for larger networks.

S5.2 Stochastic Block Model

Under the SBM, we set $P_1 = \omega_{c_ic_j}$ and $P_2 = \omega_{c_ic_j}^{(\epsilon)}$ for the test of equality, where

$$\omega = \begin{pmatrix} 0.4 & 0.2 \\ 0.2 & 0.4 \end{pmatrix} \quad \text{and} \quad \omega^{(\epsilon)} = \begin{pmatrix} 0.4 & 0.2 \\ 0.2 & 0.4 + \epsilon \end{pmatrix},$$

with $n = 100, 200, 300, 400$ and $K = 2$ communities of equal size. We used $\epsilon = 0, 0.025, 0.05, \ldots, 0.2$, where $\epsilon = 0$ satisfies $H_0 : P_1 = P_2$ and the non-zero values of $\epsilon$ satisfies $H_1 : P_1 \neq P_2$. The results are plotted in Figure 9. We observe that the proposed test performs well with
low Type-1 error and power increasing to 1 with increasing $\epsilon$ and $n$. Since $T_{eig}$ had very high rejection rates under the Chung-Lu model (Figure 4), we skipped it for the SBM and subsequent models.

For the scaling case, we used $P_2 = c \times P_1$ with $c = 0.5, 0.7, 0.75, 0.8, 0.9$ under the null model, and to configure the alternative scenario we used $P_2 = \omega^{(\epsilon)}_{c_i,c_j}$ with $\epsilon = 0.2$. The results are reported in Table 6, where we again find that the proposed test $T_{scale}$ performs well with low Type-1 errors and high power. We note one exception for $n = 100, c = 0.5$ where the type-1 error is unusually high. We conjecture that this is due to low sample size, as the type-1 error for higher values of $n$ are zero for $c = 0.5$.

**S5.3 Degree Corrected Block Model**

Under the DCBM, we used

$$\omega \propto \begin{pmatrix} 4 & 2 & 1 \\ 2 & 4 & 1 \\ 1 & 1 & 4 \end{pmatrix},$$
$K = 3$ unbalanced communities with the community assignment vector sampled from a multinomial distribution with $\pi = (0.25, 0.25, 0.5)$, and the resultant matrix of $P(i, j)$'s was then scaled to ensure that the expected network density is $\delta = 0.1$. We generated $\theta_i$ for $P_1$ from the $Beta(1, 5)$ distribution (as under the Chung Lu model). Under the DCBM, $P_1$ and $P_2$ can be configured to be different in several ways — by changing the community structure, by changing the $\omega$ matrix, by changing the degree parameters $\theta$, or a combination of all three. In this study we kept the community structure $c$ and the block matrix $\omega$ unchanged between $P_1$ and $P_2$, changing only the generation of $P_2$ as $P_2(i, j) = (\theta_i + \epsilon)(\theta_j + \epsilon)$ (as under Chung Lu model) to see how the effect of change in $\epsilon$ in rejection rate calculation. For the scaling case, we used $P_2 = c \times P_1$ with $c = 0.5, 0.7, 0.75, 0.8, 0.9$ under the null model, and to configure the alternative scenario we generated another separate set of parameters $\eta_i \sim Beta(a = 4, b = 3)$ and used $P_2(i, j) = \eta_i \eta_j$ (similar to the Chung Lu setup). The results for the equality case are plotted in Figure 10 and those for scaling are reported in Table 7. For the scaling case in particular, we observe that the power was 1 but the Type-I error showed a decreasing pattern as $n$ increases. We plan to analyze this more closely in future research.
$H_0$ is true & $H_1$ is true \\
\hline
$P_2 = 0.5P_1$ & $P_2 = 0.7P_1$ & $P_2 = 0.75P_1$ & $P_2 = 0.8P_1$ & $P_2 = 0.9P_1$ & $P_2 \neq cP_1$ \\
\hline
$n$ & $T_{\text{scale}}$ & $T_{\text{scale}}$ & $T_{\text{scale}}$ & $T_{\text{scale}}$ & $T_{\text{scale}}$ & $T_{\text{scale}}$ \\
\hline
100 & 28.30 & 2.25 & 0.95 & 0.30 & 0.00 & 100 \\
200 & 0.00 & 0.00 & 0.00 & 0.00 & 0.00 & 100 \\
300 & 0.00 & 0.00 & 0.00 & 0.00 & 0.00 & 100 \\
400 & 0.00 & 0.00 & 0.10 & 0.05 & 3.45 & 1.00 \\

Table 6: SBM scaling case: Rejection rates (in percentage) from $T_{\text{scale}}$ using $B = 200$ bootstrap iterations and averaged over 2000 Monte Carlo simulations.

$H_0$ is true & $H_1$ is true \\
\hline
$P_2 = 0.5P_1$ & $P_2 = 0.7P_1$ & $P_2 = 0.75P_1$ & $P_2 = 0.8P_1$ & $P_2 = 0.9P_1$ & $P_2 \neq cP_1$ \\
\hline
$n$ & $T_{\text{scale}}$ & $T_{\text{scale}}$ & $T_{\text{scale}}$ & $T_{\text{scale}}$ & $T_{\text{scale}}$ & $T_{\text{scale}}$ \\
\hline
100 & 0.5 & 0.7 & 0.4 & 1.2 & 1.7 & 100 \\
200 & 0.8 & 1.1 & 0.4 & 1.2 & 1.9 & 100 \\
300 & 0.5 & 0.2 & 0.1 & 0.3 & 0.2 & 100 \\
400 & 0.1 & 0.0 & 0.0 & 0.1 & 0.0 & 100 \\

Table 7: DCBM scaling case: Rejection rates (in percentage) from $T_{\text{scale}}$ using $B = 200$ bootstrap iterations and averaged over 2000 Monte Carlo simulations.

S5.4 Model mis-specification

In practice, it might not be known which model class the networks were generated from, and there could be model misspecification. In this simulation study, we consider two cases: when we mis-specify both models as RDPG (and use ASE estimator) and when we mis-specify both models as LSM (and use the LSM estimator).

S5.4.1 Cross-simulation for RDPG

In this simulation, we use the latent distance model of Hoff et al. (2002) to generate the networks, and use RDPG based estimation methods for testing. We used $d = 3$, $\alpha = 3$, and sampled the latent positions $z_1, \ldots, z_n \sim N(0, I)$ independently for $P_1$. For the equality
case, we set $P_2 = P_1$ under the null. To configure the alternative scenario for the equality case, we kept $d$ unchanged, used $\alpha - \epsilon$ instead of $\alpha$, and sampled a second set of latent positions $z_1, \ldots, z_n \sim N(0, I)$ independently for $P_2$. The results based on 2000 Monte Carlo simulations are displayed in Figure 11. We observe that none of the candidate methods work well, including the methods proposed in this paper.

S5.4.2 Cross-simulation for LSM

In this simulation, we consider the networks being generated from two RDPG models. Let $A_1 \sim \text{RDPG}(X_1)$ and $A_2 \sim \text{RDPG}(X_2)$, where $X_1$ and $X_2$ are two latent matrices of dimension $n \times 2$. The rows of $X_1$ and $X_2$ are generated by sampling with replacement from the rows of $M_1$ and $M_2$ respectively with probability vector $\pi$, where

$$
M_1 = \begin{pmatrix} 0.6 & -0.4 \\ 0.6 & 0.4 \end{pmatrix}; \quad M_2 = \begin{pmatrix} 0.6 & -0.4 - \epsilon \\ 0.6 & 0.4 + \epsilon \end{pmatrix}; \quad \pi = \begin{pmatrix} 0.4 \\ 0.6 \end{pmatrix}.
$$

The null hypothesis $H_0 : P_1 = P_2$ holds true when $\epsilon = 0$, and the alternative $H_1 : P_1 \neq P_2$ holds true when $\epsilon > 0$. We will use LSM approach to fit the model and do the hypothesis test based on that estimate.
Figure 11: Model mis-specification: Rejection rates for deviating alternatives.

Figure 12 shows that the power of the test using mis-specified LSM estimate has an increasing trend with $\epsilon$, but has much less power than using RDPG estimation here. Also, to compare between the two different test statistics, $T_{\text{frob}}$ works better than $T_{\text{eig}}$ as rejection rates are higher when using $T_{\text{eig}}$. 
Figure 12: LSM simulation Rejection rates for deviating alternatives for model mis-specification using both $T_{\text{eig}}$ and $T_{\text{frob}}$ using 50 Monte Carlo replicates.