Network two-sample test for block models

Chung K. Nguen Oscar H. Madrid-Padilla Arash A. Amini

Department of Statistics
University of California, Los Angeles
California, CA 90024
{cnguen1, oscar.madrid, aaamini}@stat.ucla.edu

Abstract

We consider the two-sample testing problem for networks, where the goal is to determine whether two sets of networks originated from the same stochastic model. Assuming no vertex correspondence and allowing for different numbers of nodes, we address a fundamental network testing problem that goes beyond simple adjacency matrix comparisons. We adopt the stochastic block model (SBM) for network distributions, due to their interpretability and the potential to approximate more general models. The lack of meaningful node labels and vertex correspondence translate to a graph matching challenge when developing a test for SBMs. We introduce an efficient algorithm to match estimated network parameters, allowing us to properly combine and contrast information within and across samples, leading to a powerful test. We show that the matching algorithm, and the overall test are consistent, under mild conditions on the sparsity of the networks and the sample sizes, and derive a chi-squared asymptotic null distribution for the test. Through a mixture of theoretical insights and empirical validations, including experiments with both synthetic and real-world data, this study advances robust statistical inference for complex network data.

1 Introduction

Network data is pervasive across various fields, including transportation [7], trading [5], social networks [11, 41], neuroscience [32, 31, 6, 30], ecology [35], and politics [4], among others. To address the diverse needs of these applications, recent statistical methods have emerged to handle scenarios where multiple networks are observed [22, 28, 19, 3]. In this paper, we consider the problem of testing whether two sets of networks have been generated from the same probability model. Specifically, in network two-sample testing, we are given two collections of graphs $\{G_{11}, \dots, G_{1N_1}\}$ and $\{G_{21}, \dots, G_{2N_2}\}$ generated as $G_{rt} \sim F_r$ for $r \in \{1, 2\}$, $t \in [N_r] := \{1, \dots, N_r\}$, and would like to test

$$H_0 : F_1 = F_2 \quad \text{against} \quad H_1 : F_1 \neq F_2. \tag{1}$$

We approach the two-sample testing problem in (1) without requiring the existence of vertex correspondence. This means that the nodes in different networks are unlabeled, and as a consequence, the ith node in the tth network of the rth sample has no direct correspondence or meaning in any of the other networks. Additionally, it is possible for the number of nodes, denoted as n_{rt}, to vary across different networks. As such, the problem we are studying is truly a network testing problem, and not immediately reducible to testing the distributions of a collection of adjacency matrices.

To further elaborate on the subtlety of graph versus matrix testing, consider the adjacency matrices $A_{rt} \in \{0, 1\}^{n_{rt} \times n_{rt}}$ associated with each graph G_{rt} for $r \in \{1, 2\}$ and $t \in [N_r]$, by fixing a particular node labeling for each graph. Then, graph-level distribution F_1 induces distributions on the adjacency matrices $A_{1t}, t \in [N_1]$. However, these distributions are not directly comparable since the

39th Conference on Neural Information Processing Systems (NeurIPS 2025).

dimensions of $\{A_{1t}\}$ could be different. Even if the dimensions are the same (i.e., graphs $\{G_{1t}\}$ all have the same number of nodes), the distributions of $\{A_{1t}\}$ depend on the particular labelings chosen for each graph. Changing the labelings, will change the distributions of $\{A_{1t}\}$, although we want to treat all such distributions as the same. In other words, when testing the equality of distributions of the adjacency matrices, the formulation must account for equality up to relabeling of the nodes, and disregard the potential size mismatches, to truly test at the level of graph distributions.

The absence of vertex correspondence introduces a fundamental challenge that elevates this problem beyond simple matrix comparisons. In fact, the unlabeled two-sample testing problem subsumes the classic *graph isomorphism problem* as a special case. If we consider samples of size $N_1 = N_2 = 1$ where the generating distributions are point masses on two fixed graphs, G_1 and G_2, the two-sample test reduces to deciding if G_1 is isomorphic to G_2. Our problem can thus be seen as a "noisy" or statistical version of graph isomorphism, where we must determine if two distributions on graphs are equivalent up to a permutation.

To model the distribution of the adjacency matrices, we adopt the stochastic block model (SBM) framework [17] as a foundational structure. The SBM is one of the most commonly used models in the literature, valued not only for its simplicity and interpretability but also for its role as a universal approximator. It is effectively the network equivalent of a histogram [29]; an SBM with a sufficiently large number of communities, K, can form a step-function approximation to an arbitrary graphon in the L^2 sense [2, 14].

This capability positions our method as a "bin-then-test" strategy, which is analogous to classic non-parametric approaches like Pearson's chi-squared test, where continuous data is binned to form contingency tables. This perspective makes the SBM a natural and powerful starting point, providing a flexible, non-parametric framework for the two-sample testing of networks.

In this work, we assume that the adjacency matrices are generated from SBM distributions. In particular, an adjacency matrix $A_{rt} \in \{0,1\}^{n_{rt} \times n_{rt}}$ is generated as follows. First, we draw a vector of random *community* labels $z_{rt} = (z_{rt1}, \ldots, z_{rtn})^\top \in [K]^n$ with entries that are independent draws from a categorical distribution with parameter $\pi_r = (\pi_{r1}, \ldots, \pi_{rK})$, where $\pi_{rl} > 0$ for all l and $\sum_{l=1}^K \pi_{rl} = 1$. Here, K is the number of communities. Then, given the labels z_{rt}, the entries of A_{rt}, below the diagonal, are independently drawn as

$$(A_{rt})_{ij} \,|\, z_{rt} \;\sim\; \mathsf{Ber}((B_r)_{z_{rti}\, z_{rtj}}), \quad i < j,$$

where $B_r \in [0,1]^{K \times K}$ is a symmetric matrix of probabilities. We assume the networks to be undirected, with no self-loops, hence A_{rt} is extended to a symmetric matrix with zero diagonal entries. Throughout, we write $A_{rt} \sim \mathsf{SBM}_{n_{rt}}(B_r, \pi_r)$ for A_{rt} generated as described above. We often suppress the dependence on n_{rt} and simply write $A_{rt} \sim \mathsf{SBM}(B_r, \pi_r)$.

Given $A_{rt} \sim \mathsf{SBM}(B_r, \pi_r)$, the network testing problem (1) is equivalent to the following

$$\begin{cases} H_0 : \exists P \in \Pi_K, & B_1 = P B_2 P^\top \quad \text{and} \quad \pi_1 = P\pi_2 \\ H_1 : \forall P \in \Pi_K, & B_1 \neq P B_2 P^\top \quad \text{or} \quad \pi_1 \neq P\pi_2 \end{cases} \tag{2}$$

where Π_K is the set of $K \times K$ permutation matrices. The inclusion of permutation matrices P in (2) is related to the subtlety of testing unlabeled network models. The fact that there is no pre-defined node label or vertex correspondence across networks leads to the meaninglessness of community labels in SBM. Consequently, demanding that $B_1 = B_2$ under the null hypothesis is not meaningful. Put another way, SBM parameters (B_r, π_r) are only identifiable up a permutation; see (4).

Before proceeding, we generalize the problem slightly. In (2), the null hypothesis is rejected if the class proportions π_r differ modulo permutation. Given the SBM labels, this scenario is straightforward to test by estimating and sorting class proportions to remove permutation ambiguity. To avoid this relatively trivial case, we treat the class proportions as nuisance parameters. This choice is also practical, as communities are typically defined by connectivity patterns rather than relative sizes. Additionally, differences in class proportions π_r can sometimes be transferred to connectivity matrices B_r via community *refinement* (see Appendix C.1 in the supplementary material). With these considerations in mind, we focus on the following formulation for the remainder of the paper: Given $A_{rt} \sim \mathsf{SBM}(B_r, \pi_{rt})$, we test

$$\begin{cases} H_0 : \exists P \in \Pi_K, & B_1 = P B_2 P^\top \\ H_1 : \forall P \in \Pi_K, & B_1 \neq P B_2 P^\top. \end{cases} \tag{3}$$

Note that here, the class proportions π_{rt} are allowed to vary across networks without violating the null hypothesis.

To develop a test for (3), it is essential to address the significant challenge posed by the fact that matrices B_r are only identifiable up to permutations. A potential approach to constructing a test statistic is to estimate the community labels for each network, using any of the many existing approaches (for instance [4]). Subsequently, one can proceed to estimate B_r by computing block averages of the adjacency matrices, based on these labels. However, the challenge lies in how to effectively combine, and compare, the estimates from different networks when vertex correspondence is not assumed. This becomes particularly daunting, as matching estimated labels across different networks boils down to searching over the space of permutations Π_K, which has $K!$ elements, a non-trivial task even for moderate values of K. It is worth noting that this matching challenge exists even if we observe a single network for each of the two samples, since one has to properly match the communities of the two networks to be able to compare the connectivity matrices B_1 and B_2.

1.1 Summary of results

In this paper, we tackle the core challenges of unlabeled network comparison by developing a practical and theoretically-grounded two-sample test. Our main contributions are:

- **An Efficient and Consistent Matching Algorithm.** We introduce a computationally efficient spectral matching algorithm to solve the fundamental problem of aligning estimated SBM parameters. We prove that this algorithm consistently recovers the correct alignment under standard spectral conditions on the underlying connectivity matrices (formally, the (η, θ)-friendly property).
- **A Rigorous Characterization of the Test Statistic.** We establish the key theoretical properties of our test. Under the null hypothesis, we prove that our statistic converges to a $\chi^2_{K(K+1)/2}$ distribution under mild conditions on network sparsity (average degree $\nu_n = \Omega(\log n)$) and sample size. This provides a practical foundation for accurate p-value calculation.
- **Proof of Consistency and Power.** Under the alternative hypothesis, we show that the test is consistent and powerful. The test statistic grows at a rate of $\Omega(N\nu_n^2)$, demonstrating that its power increases with both the number of networks (N) and their average degree (ν_n).
- **A Scalable and Empirically Validated Test.** We provide a practical, computationally efficient algorithm for the two-sample test, whose scalability stands in contrast to computationally prohibitive methods like subgraph counting. We demonstrate its superior performance and robustness on a wide range of synthetic data, real-world networks, and non-SBM models, including graphons and RDPGs.

1.2 Related work

Two-sample testing for labeled networks with known vertex correspondence has been well studied in the literature. Notable efforts in the labeled case include the following: [16], where the authors develop a test based on the geometric characterization of the space of the graph Laplacians. [15] consider the problem from a minimax testing perspective, focusing on the theoretical characterization of the minimax separation with respect to the number of networks and the number of nodes. [25] develop a test based on an omnibus embedding of multiple networks into a single space. [8, 9] develop a spectral-based test statistic that has an asymptotic Tracy–Widom law under null. [18] introduce a novel approach to testing degree corrected stochastic block models based on interlacing balance measure.

Existing literature on testing (1) with unlabeled networks with no vertex correspondence is limited, but there have been notable efforts. For instance, [22] develop a geometric and statistical framework for making inferences about populations of unlabeled networks. [37] introduce a two-sample test for random dot product graphs based on estimating the maximum mean discrepancy between the latent positions. [27] develop a two-sample test, based on subgraph counts, and characterize the null distribution under a graphon model. While not designed for testing, two algorithms described in [28] solve a closely related problem of clustering network-value data.

As discussed earlier, the construction of our test relies on the subroutine that matches the estimated connectivity matrices of two networks. This type of problem is known as weighted graph matching and can be formulated as a quadratic assignment problem (QAP) which is NP-hard. One of the

standard techniques is to relax the constraint set of permutation matrices to its convex hull [26, 39, 1] or the set of orthogonal matrices [13], and then "round" the solution to the set of permutation matrices. [38] proposes to solve the weighted matching by solving a linear assignment problem on a matrix derived from eigenvectors of the adjacency matrices of the two graphs. [12] show exact recovery of the permutation with high probability for the Gaussian Wigner model. The idea of our main matching algorithm is mentioned in passing in [36, Section 39.4] in the context of testing isomorphism of two graphs, and is attributed to [24].

2 Matching Methodology

In this section, we begin by describing the challenges associated with the matching problem, specifically the task of aligning two matrices, B_1 and B_2, by relabeling the communities. Subsequently, we introduce our proposed matching algorithm, which is at the heart of our test construction.

2.1 Matching challenge

From a statistical perspective, the challenge in testing samples from $\mathsf{SBM}(B, \pi)$ is that the parameters (B, π) are identifiable only up to permutations. That is, for any permutation matrix

$$\mathsf{SBM}(B, \pi) = \mathsf{SBM}(PBP^\top, P\pi) \tag{4}$$

as distributions. To illustrate the challenge more clearly, consider observing the two-sample problem with only two adjacency matrices

$$A_1 \sim \mathsf{SBM}(B_1, \pi_1), \quad A_2 \sim \mathsf{SBM}(B_2, \pi_2).$$

Assume that the null hypothesis holds, where we can assume $B_2 = B_1$. To devise a test, a natural first step is to fit an SBM to A_r, using any consistent community detection algorithm, to obtain the labels, from which we can construct an estimate $\widehat{B}_r = \mathcal{B}(A_r)$ of B_r, for $r = 1, 2$. The operator $\mathcal{B}$ is a shorthand for the procedure that produces such estimate, the details of which are discussed in Section 3. Even though $B_1 = B_2$, in general, we will have

$$\widehat{B}_2 \approx P^* \widehat{B}_1 P^{*\top},$$

for some permuation matrix $P^* \in \Pi_K$, since in each case the communities will be estimated in an unknown (arbitrary) order. There is no way a priori to guarantee the same order of estimated communities in the two cases. This is a manifestation of the permutation ambiguity in (4). To simplify future discussions, let us introduce the following notation for *exact matching*:

$$B_1 \xrightarrow{P^*} B_2 \quad \iff B_2 = P^* B_1 P^{*T}. \tag{5}$$

The above discussion shows that two-sample testing for SBMs requires solving the following subproblem:

Problem 1. Assume that $B_1 \xrightarrow{P^*} B_2$ for $B_1, B_2 \in [0,1]^{K \times K}$. The *noisy graph isomorphism* (or graph matching) problem is to recover a matching permutation $P^* \in \Pi_K$ between B_1 and B_2, using only noisy observations $\widehat{B}_r \approx B_r$ for $r = 1, 2$.

As alluded to in Section 1.2, this problem is in general hard. In particular, finding an optimal matching in Frobenius norm reduces to solving a quadratic assignment problem (QAP), which is NP-hard in general.

2.2 Spectral matching

For a large class of weighted networks, we can avoid solving a QAP to recover a matching. The core idea is that if a connectivity matrix has a unique spectral signature, its eigenvalue decomposition (EVD) contains the information needed to recover any permutation. The following lemma makes this precise.

Lemma 2.1. Consider two $K \times K$ symmetric matrices B_1 and B_2 with EVDs given by $B_r = Q_r \Lambda_r Q_r^\top$ for $r = 1, 2$. If the eigenvalues of B_1 are distinct, then a permutation matrix P^* satisfies $B_2 = P^* B_1 P^{*\top}$ if and only if there exists a diagonal sign matrix S^* such that:

$$\Lambda_2 = \Lambda_1, \tag{6}$$

$$Q_2 = P^* Q_1 S^*. \tag{7}$$

Algorithm 1 Spectral Matching: $\mathcal{M}(\widehat{B}_1 \to \widehat{B}_2)$

Input: Estimated connectivity matrices $\widehat{B}_1$, $\widehat{B}_2$
Output: Estimated sign matrix $\widehat{S}$, estimated permutation matrix $\widetilde{P}$
1: **for** $r = 1, 2$ **do**
2: Perform EVD on $\widehat{B}_r$ to obtain $\widehat{B}_r = \widehat{Q}_r \Lambda_r \widehat{Q}_r^\top$.
3: **end for**
4: Recover the diagonal signs $\widehat{S}_{ii} = \operatorname{sign}\big([\widehat{Q}_2^\top \mathbf{1}]_i / [\widehat{Q}_1^\top \mathbf{1}]_i\big)$ and set $\widehat{S} = \operatorname{diag}(\widehat{S}_{ii})$.
5: Recover the permutation matrix as $\widetilde{P} = \mathrm{LAP}(\widehat{Q}_1 \widehat{S} \widehat{Q}_2^\top)$.
6: **return** $\widehat{S}$, $\widetilde{P}$.

This result provides a direct algebraic path to finding P^*. To ensure this process is robust against the noise present in estimates $\widehat{B}_r$, we require a slightly stronger condition that the eigenvalues are not just distinct, but well-separated. This, along with a technical condition ensuring that no eigenvector is orthogonal to the all-ones vector, guarantees that the matching permutation P^* is **unique** and that our spectral approach is stable. We formalize these combined requirements as the "(η, θ)-friendly" property in Appendix A.1.

Our proposed spectral matching algorithm, based on Lemma 2.1, is outlined in Algorithm 1. It relies on solving the *linear assignment problem* (LAP), defined for a $K \times K$ matrix Q as:

$$\mathrm{LAP}(Q) := \operatorname*{argmax}_{P \in \Pi_K} \operatorname{tr}(PQ). \tag{8}$$

The LAP can be solved efficiently, for example, by the Hungarian algorithm [23].

The intuition behind Algorithm 1 follows directly from Lemma 2.1. From (7), we can left-multiply by $\mathbf{1}^\top$ to get $\mathbf{1}^\top Q_2 = \mathbf{1}^\top P^* Q_1 S^*$. Since $\mathbf{1}^\top P^* = \mathbf{1}^\top$, this simplifies to $Q_2^\top \mathbf{1} = S^* Q_1^\top \mathbf{1}$. This gives us a way to recover the unknown signs: $S^*_{ii} = [Q_2^\top \mathbf{1}]_i / [Q_1^\top \mathbf{1}]_i$. The aforementioned condition that eigenvectors are not orthogonal to the all-ones vector ensures the denominator $[Q_1^\top \mathbf{1}]_i$ is non-zero, making this step well-defined. For robustness in the noisy case, we only take the sign of this ratio, which is exactly Step 3 of the algorithm.

Once the sign matrix $\widehat{S} \approx S^*$ is known, we can rearrange (7) to isolate the permutation matrix: $P^* = Q_2 S^* Q_1^\top$. In the noisy setting with estimates $\widehat{Q}_r$, the product $\widehat{Q}_2 \widehat{S} \widehat{Q}_1^\top$ will be a noisy version of P^*. We find the *closest* valid permutation matrix by projecting it onto the set Π_K, which can be done by solving the LAP:

$$\operatorname*{argmin}_{P \in \Pi_K} \|\widehat{Q}_2 - P \widehat{Q}_1 \widehat{S}\|_F^2 = \operatorname*{argmax}_{P \in \Pi_K} \operatorname{tr}(\widehat{Q}_2^\top P \widehat{Q}_1 \widehat{S}) = \mathrm{LAP}(\widehat{Q}_1 \widehat{S} \widehat{Q}_2^\top).$$

This is precisely Step 4 of the algorithm. In the sequel, we write $\mathcal{M}(\widehat{B}_1 \to \widehat{B}_2)$ to denote the permutation matrix $\widetilde{P}$ returned.

Remark 2.1 (Computational Complexity)**.** A key advantage of our spectral matching approach is its computational efficiency. The complexity of Algorithm 1 is dominated by the EVD and the LAP solution. The LAP on a $K \times K$ matrix can be solved in $O(K^3)$ time using the Hungarian algorithm. This cost depends only on the number of communities K, not the number of nodes n. In practice, the matching step is extremely fast, making our overall test scalable to very large networks.

3 Test construction

We are now ready to describe our main algorithm for constructing an SBM two-sample test. The first step is to estimate the connectivity matrix B for each observed network. Given an adjacency matrix A and a vector of estimated community labels $\widehat{z} \in [K]^n$ (obtained from an algorithm like spectral clustering [4] or Bayesian community detection [35]), we define the block-sum operator $\mathcal{S}$ and block-count operator $\mathcal{C}$:

$$[\mathcal{S}(A, z)]_{k\ell} = \sum_{i,j} A_{ij} 1\{z_i = k, z_j = \ell\}, \tag{9}$$

$$[\mathcal{C}(z)]_{k\ell} = n_k(z)\big(n_\ell(z) - 1\{k = \ell\}\big), \tag{10}$$

Algorithm 2 SBM Two-Sample Test (SBM-TS)

Input: Adjacency matrices A_{rt} and initial label estimates $\widehat{z}_{rt}^{(0)}$, for $t \in [N_r]$ and $r = 1, 2$.

Output: Two-sample test statistic $\widehat{T}_n$.

Match to the first network within each sample r:

for $t = 1, 2, \ldots, N_r$ and $r = 1, 2$ **do**

  Set $\widehat{z}_{rt} = \sigma_{rt} \circ \widehat{z}_{rt}^{(0)}$ for independent random permutation σ_{rt}.

  Set $\widehat{S}_{rt} = \mathcal{S}(A_{rt}, \widehat{z}_{rt})$ and $\widehat{m}_{rt} = \mathcal{C}(\widehat{z}_{rt})$.

  Set $\widehat{B}_{rt} = \widehat{S}_{rt} \oslash \widehat{m}_{rt}$.

  Set $\widehat{P}_{rt} = \mathcal{M}(\widehat{B}_{rt} \to \widehat{B}_{r1})$.

end for

Find the global matching permutation:

Set $\widehat{B}_r = \frac{1}{N_r} \sum_{t=1}^{N_r} \widehat{P}_{rt} \widehat{B}_{rt} \widehat{P}_{rt}^\top$ for $r = 1, 2$.

Set $\widehat{P} = \mathcal{M}(\widehat{B}_2 \to \widehat{B}_1)$.

Align the sums and counts, across all samples, and form aggregate estimates:

Set $\widehat{S}_r = \sum_{t=1}^{N_r} \widehat{P}_{rt} \widehat{S}_{rt} \widehat{P}_{rt}^\top$ and $\widehat{m}_r = \sum_{t=1}^{N_r} \widehat{P}_{rt}\, \widehat{m}_{rt}\, \widehat{P}_{rt}^\top$ for $r = 1, 2$.

Set $\widehat{S}_2' = \widehat{P} \widehat{S}_2 \widehat{P}^\top$ and $\widehat{m}_2' = \widehat{P} \widehat{m}_2 \widehat{P}^\top$.

Let $\widehat{B} = (\widehat{B}_{k\ell})$ where $\widehat{B}_{k\ell} = \dfrac{\widehat{S}_{1k\ell} + \widehat{S}'_{2k\ell}}{\widehat{m}_{1k\ell} + \widehat{m}'_{2k\ell}}$ and set $\widehat{\sigma}^2_{k\ell} = \widehat{B}_{k\ell}(1 - \widehat{B}_{k\ell})$.

Let $\widehat{B}^{(1)} = \widehat{S}_1 \oslash \widehat{m}_1$ and $\widehat{B}^{(2)} := \widehat{S}_2' \oslash \widehat{m}_2'$. (*Stronger estimates*)

Let $\widehat{h}_{k\ell}$ be the harmonic mean of $\widehat{m}_{1k\ell}$ and $\widehat{m}'_{2k\ell}$ and form the test statistic:

$$\widehat{T}_n := \sum_{k \le \ell} \frac{\widehat{h}_{k\ell}}{2\widehat{\sigma}^2_{k\ell}} \big(\widehat{B}^{(1)}_{k\ell} - \widehat{B}^{(2)}_{k\ell}\big)^2. \tag{11}$$

where $n_k(z) = \sum_{i=1}^n 1\{z_i = k\}$. Both operators produce a $K \times K$ matrix. The estimate of the connectivity matrix is then $\mathcal{B}(A, z) = \mathcal{S}(A, z) \oslash \mathcal{C}(z)$, where $\oslash$ is elementwise division. For diagonal blocks ($k = \ell$) a double-counting occurs in both the numerator and denominator which can be avoided, for computational efficiency, by restricting the sums to $i < j$.

3.1 Main algorithm

Our testing procedure, detailed in Algorithm 2, unfolds in three main stages. First, to enable aggregation, we align all networks *within* each sample to a common reference (e.g., the first network). Second, with the intra-sample alignments complete, we compute an average connectivity matrix for each sample and perform a single *global* match between the two samples. Finally, with all networks aligned to a common orientation, we aggregate the block-level edge counts across all networks to form powerful, low-variance estimates of the connectivity matrices and construct the final chi-squared test statistic.

To understand the form of the final statistic, consider an idealized scenario where all matchings are perfect and the community labels are known. Under the null hypothesis ($B_1 = B_2 = B$), the aggregate edge counts $\widehat{S}_{rk\ell}$ would follow Binomial distributions, and by the CLT, the estimates $\widehat{B}^{(1)}_{k\ell}$ and $\widehat{B}^{(2)}_{k\ell}$ are approximately independent normal variables: $\widehat{B}^{(r)}_{k\ell} \overset{d}{\approx} N(B_{k\ell}, \sigma^2_{k\ell}/m_{rk\ell})$ for $r = 1, 2$, where $\sigma^2_{k\ell} = B_{k\ell}(1 - B_{k\ell})$ is the common Bernoulli variance and $m_{rk\ell}$ are the total block pair counts for each sample. The standardized difference of these two estimates is

$$\sqrt{h_{k\ell}/(2\sigma^2_{k\ell})}(\widehat{B}^{(1)}_{k\ell} - \widehat{B}^{(2)}_{k\ell}) \;\overset{d}{\approx}\; N(0, 1)$$

where $h_{k\ell}$ is the harmonic mean of $m_{1k\ell}$ and $m_{2k\ell}$. Squaring this quantity, we see that each term in our test statistic (11) is an estimate of this squared standard normal variable. This provides the core intuition why $\widehat{T}_n$ will converge to $\chi^2_{K(K+1)/2}$ distribution, a result we formalize in Section 4.

Remark 3.1 (Algorithmic Refinements)**.** Two details in Algorithm 2 warrant mention. First, the final estimates $\widehat{B}^{(r)}$ (Step 12) are constructed by pooling edge counts *before* dividing. This produces a lower-variance estimate than averaging the individual $\widehat{B}_{rt}$'s. Second, the randomization of labels in Step 2 is a technical device to ensure the estimated matchings are statistically independent of the data, which simplifies the theoretical analysis (see Lemma E.3 in the appendix).

4 Theory

In this section, we present the main theoretical guarantees for our proposed test. For clarity and accessibility, we state our key results—the asymptotic null distribution and the consistency of the test—under a simplified set of assumptions. These assumptions presume the use of a community detection algorithm with a near-optimal misclassification rate, which is achievable in the sparse regimes we consider (e.g., [42]). This allows us to express the conditions directly in terms of fundamental network parameters like the average expected degree ν_n. We provide the more general, finite-sample versions of these theorems, which hold for any community detection algorithm via its misclassification rate, in Appendix A.

4.1 Matching Consistency

Our test relies on the ability of the spectral matching algorithm to correctly align the estimated SBM parameters. The following result provides an informal guarantee, with the formal statement deferred to Theorem A.1 in the appendix.

Theorem 4.1 (Matching Consistency, Informal)**.** If the underlying connectivity matrices B_r have distinct eigenvalues and their eigenvectors are not orthogonal to the all-ones vector (i.e., they are "friendly" per Definition A.1 in the appendix), our spectral matching algorithm (Algorithm 1) consistently recovers the correct permutation between them from noisy estimates $\widehat{B}_r$, provided the estimation error is sufficiently small.

4.2 Asymptotic Guarantees for the Test

We now present the main results for our two-sample test statistic, $\widehat{T}_n$. We begin with its limiting distribution under the null hypothesis.

Theorem 4.2 (Null Distribution)**.** Assume that Algorithm 2 is applied to two SBM samples with the same connectivity matrix $B = (\nu_n/n)B^0$ for a fixed $K \times K$ matrix B^0. Assume that:

(i) B^0 has distinct eigenvalues, and no eigenvector of it sums to zero.

(ii) The number of networks in each sample grows at most polynomially in the number of nodes n, i.e., $N := \max\{N_1, N_2\} = O(n^a)$ for some $a \geq 0$.

(iii) The average expected degree grows at least logarithmically with the network size, i.e., $\nu_n = \Omega(\log n)$.

Then, the test statistic $\widehat{T}_n$ converges in distribution to a chi-squared distribution:

$$\widehat{T}_n \Rightarrow \chi^2_{K(K+1)/2}.$$

Next, we show that the test is consistent under the alternative hypothesis, meaning its power approaches one as the network size grows.

Theorem 4.3 (Consistency)**.** Assume that Algorithm 2 is applied to two SBM samples with different connectivity matrices $B_r = (\nu_n/n)B^0_r$ for $r = 1, 2$, such that B^0_1 is not a permutation of B^0_2. Under the same conditions on the eigenvalues/eigenvectors of B^0_r and the growth of N_r as in Theorem 4.2, and assuming the average expected degree $\nu_n \to \infty$, the test is consistent. Specifically, with probability approaching 1,

$$\widehat{T}_n = \Omega(N\nu_n^2),$$

where $N = \max\{N_1, N_2\}$. This ensures that for any fixed significance level, the power of the test approaches 1 as $n \to \infty$.

Remark 4.4 (On the Sparsity Condition)**.** One might intuitively expect that aggregating N networks could weaken the sparsity requirement of $\nu_n = \Omega(\log n)$. However, this fundamental dependence on $\log n$ persists even for large N. To see why, consider a sample of N networks of size n. This can be viewed as a single, large, block-diagonal adjacency matrix of size $nN \times nN$, where the off-diagonal blocks corresponding to inter-network connections are unobserved. For a single, fully observed network of size nN, established results for tasks like community detection require the average degree to scale as $\Omega(\log(nN)) = \Omega(\log n + \log N)$. Since our problem has less information (due to the missing off-diagonal blocks), the requirement on the $\log n$ term cannot be removed by increasing N.

Remark 4.5 (Practicality of the Sparsity Condition)**.** It is important to note that the condition $\nu_n = \Omega(\log n)$ still describes a very sparse graph regime. For instance, if the average degree grows as $\log n$, adding one million nodes to a network would only increase the average degree by approximately $\log(10^6) \approx 14$. As we demonstrate empirically in Appendix C.2, the average degree in many real-world networks grows even faster than logarithmically with network size, suggesting this assumption is mild in practice.

5 Experimental results

In this section, we provide experimental results on real and simulated networks and compare our proposed test to two existing approaches. The code for reproducing all experiments in this section is publicly available at `https://github.com/aaamini/sbm-ts`.

5.1 On Model Misspecification and the Choice of K

A key motivation for our SBM-based test is its applicability beyond the strict SBM setting. Because SBMs can approximate general smooth graphons with arbitrary precision by increasing the number of communities K, our test provides a robust method for non-parametric network comparison. Our experiments on data from graphon and RDPG models (Subsections 5.5 and 5.4) provide strong empirical evidence for this robustness.

Theoretical Guidance on the Range of K for Graphons. While a full theoretical analysis under a general graphon model is beyond our scope, we can construct a semi-rigorous argument by combining our test statistic's structure with recent oracle inequalities for graphon estimation [21]. This analysis, detailed in Appendix C.3, shows that the power of the test depends on balancing three competing factors: (i) the model approximation error (bias), which decreases as K grows; (ii) two sources of statistical estimation error (variance), which increase with K. For the test to be consistent, all three error terms must vanish relative to the signal strength. This leads to a sufficient set of conditions on the growth of K:

$$K \to \infty, \quad \text{while} \quad K^2 = o(n\nu_n) \quad \text{and} \quad \log K = o(\nu_n).$$

These conditions provide valuable theoretical guidance: K should grow with the network size to ensure the SBM is a good approximation, but this growth is limited by the network's size and sparsity. This leaves a wide and practical range for choosing K.

Practical Selection of K. In practice, where the optimal K is unknown, we propose a data-driven procedure akin to a parametric bootstrap to select a value that maximizes test power for a given problem scale:

(i) For a candidate K_0 (e.g., $K_0 = 10$), fit a K_0-SBM to one sample, yielding estimates $(\widehat{B}, \widehat{\pi})$.
(ii) Create a synthetic alternative by perturbing the estimate: $\widehat{B}_{alt} = \widehat{B} + \text{Noise}$.
(iii) For a range of values $K \in [2, K_{\max}]$, generate many synthetic datasets under $H_0 : (\widehat{B}, \widehat{B})$ and $H_1 : (\widehat{B}, \widehat{B}_{alt})$.
(iv) For each K, calculate the Area Under the ROC Curve (AUC) of our test in distinguishing these synthetic samples.
(v) Select the K that yields the highest AUC.

In our experiments, this procedure proved effective. For instance, in the graphon experiment of Section 5.5, it correctly suggested a small value of $K = 2$.

Table 1: Mean AUC-ROC for various methods for the general SBM experiment.

K	2	3	15	20
SBM-TS	0.95	0.91	0.83	0.81
ASE	0.62	0.54	0.49	0.48
NCLM	0.70	0.64	0.50	0.50

5.2 Competing methods

We benchmark our method by comparing it against the two test from the literature: (1) Network Clustering based on Log Moments (NCLM) [28] and (2) the Maximum Mean Discrepancy of the Adjacency Spectral Embeddings (ASE-MMD) [37]. Whilst these two methods were not originally designed for two-sample testing (NCLM is a clustering algorithm for networks and ASE-MMD is designed for sample size 1), there is a natural way to extend them into this setting. See section D.1 of supplementary material for a more detailed discussion.

5.3 General SBM with random B

We generate 50 instances of the testing problem $B_1 = B_2 = \rho B^{(i)}$ versus $(B_1, B_2) = (\rho B^{(i)}, \rho B_\varepsilon^{(i)})$, where $B^{(i)}$ is a symmetric matrix whose entries are drawn from $\mathsf{U}(0.2, 0.7)$ and $B_\varepsilon^{(i)} = B^{(i)} + M_\varepsilon^{(i)}$ such that $M_\varepsilon^{(i)}$ has entries $N(0, \varepsilon^2)$ subject to symmetry. We set the sample sizes to $N_r = 100$, the number of vertices to $n_{rt} = 10000$, the noise level to $\varepsilon = 0.05$, and sparsity factor to $\rho = 0.1$.

From Table 1, we can see that SBM-TS exhibits superior performance over the competitors, but its performance somewhat deteriorates as K increases. The primary reason is that significantly increasing K makes it more challenging to satisfy the approximate matching condition (14). Nonetheless, the median area under the ROC curve (AUC) for $K \in \{15, 20\}$ is 0.89 and 0.87 respectively, which shows the strong performance of our proposed test in the cases where (14) holds.

5.4 Random dot product graphs

We consider two experiments around the two sample testing problem

$$H_0 : A, A' \sim \mathsf{RDPG}_n(\mathbb{F}, \rho), \quad \text{against} \quad H_1 : A \sim \mathsf{RDPG}_n(\mathbb{F}, \rho),\ A' \sim \mathsf{RDPG}_n(\mathbb{F}', \rho)$$

We set the number of vertices to $n = 10000$ and the sparsity parameter $\rho = 0.15$. For Experiment 1, we take $\mathbb{F} = \mathbb{F}_1 := N(0, \Sigma)$ and $\mathbb{F}' = N(0, I_2)$ where $\Sigma = \begin{bmatrix} 3 & 2 \\ 2 & 3 \end{bmatrix}$. For Experiment 2, $\mathbb{F} = \mathbb{F}_2 := \frac{1}{2}N(0, \Sigma) + \frac{1}{2}N(0, O\Sigma O^\top)$ while $\mathbb{F}' = N(0, I_2)$ as in Experiment 1. Here, $\mathbb{F}_2$ is a mixture of two Gaussian distributions and $O = \left(\begin{smallmatrix} 0 & 1 \\ -1 & 0 \end{smallmatrix}\right)$. Due to the invariance of the RDPG to an orthogonal transformation, the two experiments are equivalent.

Figure 1c summarizes the results, comparing the performance of SBM-TS to ASE-MMD. While SBM-TS essentially exhibits the ROC of the perfect test, ASE-MMD shows poor performance. The reason is that, as proposed in [37], ASE-MMD does not have a matching mechanism to account for the potential orthogonal transformation mismatch between the estimated latent positions for A and A' under null. This defect is present in the original paper [37] as can be seen from the presence of the orthogonal matrix $\boldsymbol{W}_{n,m}$ in the asymptotic null limit of the statistic in Theorem 5 of [37].

5.5 Graphon

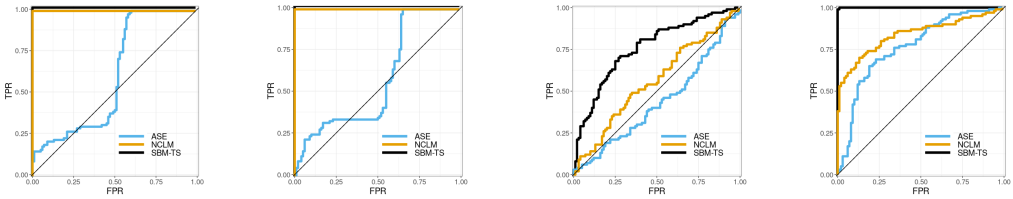
Consider the two-sample testing problem

$$H_0 : A, A' \sim \mathsf{Graphon}(\rho W), \quad \text{against} \quad H_1 : A \sim \mathsf{Graphon}(\rho W),\ A' \sim \mathsf{Graphon}(\rho W_{\varepsilon,\delta})$$

where ρ is a sparsity factor. We take $W(v_1, v_2) = \frac{1}{4}(v_1^2 + v_2^2 + \sqrt{v_1} + \sqrt{v_2})$, the graphon from [2], and let $W_{\varepsilon,\delta}$ be the following perturbation of W,

$$W_{\varepsilon,\delta}(v_i, v_j) = W(v_i, v_j) + \varepsilon \cdot 1\big\{v_i, v_j \in [0.5 - \delta; 0.5 + \delta]\big\}.$$

In this experiment, we let $\varepsilon = 0.05$, $\delta = 0.2$. For a general graphon, spectral clustering does not provide a good SBM fit. To fit a proper K-SBM, we use a Bayesian community detection algorithm [35]. The algorithm of Section 5.1 suggests $K = 2$ as the optimal choice and we set $d = K$ for ASE-MMD. Figure 1f illustrates the resulting ROCs for two cases of (n, ρ), namely, $(10^3, 0.5)$ and $(10^4, 0.05)$, showing a clear advantage for SBM-TS against the ASE-MMD.



(a) $\mathbb{F} = \mathbb{F}_1$ (b) $\mathbb{F} = \mathbb{F}_2$ (d) $n = 1000, \rho = 0.5$ (e) $n = 10000, \rho = 0.05$

(c) RDPG experiments (f) Graphon experiments

Figure 1: ROC curves for the RDPG and Graphon experiments

5.6 COLLAB dataset

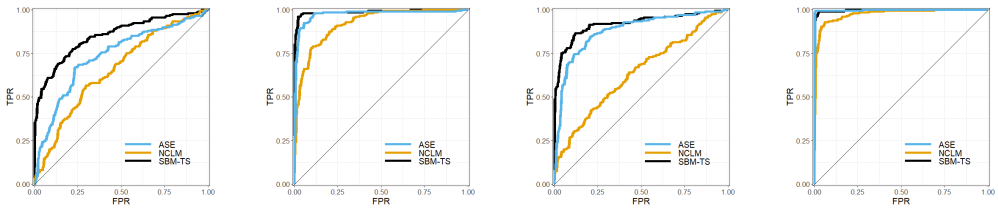
The COLLAB dataset is a scientific collaboration dataset first introduced in [40]. This dataset contains 5000 networks derived from public collaboration data in three scientific fields: $\mathcal{C}_1$) High energy Physics, $\mathcal{C}_2$) Condensed Matter Physics, and $\mathcal{C}_3$) Astrophysics. Each graph corresponds to an ego-network of a researcher and is labeled by their primary field of research. We consider two-sample testing problems of the form (1) with $N_1 = N_2 = m$; the two samples under null will be drawn at random (without replacement) both from class $\mathcal{C}_i$, and under the alternative from classes $\mathcal{C}_i$ and $\mathcal{C}_j$ for $i \neq j$. Two choices of $m \in \{5, 10\}$ and several choices of (i, j) are considered, as detained in Figure 2 where the resulting ROCs are plotted. One observes that SBM-TS has superior or comparable performance to the competing methods in distinguishing the two classes in each case.

6 Discussion

We have introduced a practical and theoretically-grounded framework for the two-sample testing of unlabeled networks, addressing the core challenge of permutation invariance with a scalable spectral matching algorithm.

Our approach is inherently modular, separating the problem into a matching stage and a testing stage. This structure opens several avenues for future work. For instance, while our current method solves the matching problem in two sequential steps, one could explore alternative optimization strategies, such as alternating minimization over permutations and signs. Low-rank approaches that match only the most informative eigenvectors instead of full matrices are another option.

The choice of the stochastic block model was motivated by its ability to approximate more general network models. Our experiments successfully demonstrated that our test is a robust and practical option for non-SBMs, such as those from graphon distributions. A key theoretical next step is to extend our analysis to provide formal guarantees for the test under this broader class of models, solidifying the connection between SBM-based testing and non-parametric network analysis.



(a) $m = 5$, C1 vs. C3. (b) $m = 5$, C2 vs C3. (c) $m = 10$, C1 vs C3. (d) $m = 10$, C2 vs C3.

Figure 2: ROC curves for the COLLAB dataset.

Acknowledgments

This material is based upon work supported by the National Science Foundation under Award No. 1945667.

References

[1] Y. Aflalo, A. Bronstein, and R. Kimmel. On convex relaxation of graph isomorphism. *Proceedings of the National Academy of Sciences*, 112(10):2942–2947, 2015. doi: 10.1073/pnas.1401651112. URL https://www.pnas.org/doi/abs/10.1073/pnas.1401651112.

[2] E. M. Airoldi, T. B. Costa, and S. H. Chan. Stochastic blockmodel approximation of a graphon: Theory and consistent estimation. *Advances in Neural Information Processing Systems*, 26, 2013.

[3] A. Amini, M. Paez, and L. Lin. Hierarchical stochastic block model for community detection in multiplex networks. *Bayesian Analysis*, 19(1):319–345, 2024.

[4] A. A. Amini, A. Chen, P. J. Bickel, and E. Levina. Pseudo-likelihood methods for community detection in large sparse networks. *Annals of Statistics*, pages 1–27, 2013.

[5] M. Barigozzi, G. Fagiolo, and D. Garlaschelli. Multinetwork of international trade: A commodity-specific analysis. *Physical Review E*, 81(4):046104, 2010.

[6] U. Braun, A. Harneit, G. Pergola, T. Menara, A. Schäfer, R. F. Betzel, Z. Zang, J. I. Schweiger, X. Zhang, K. Schwarz, et al. Brain network dynamics during working memory are modulated by dopamine and diminished in schizophrenia. *Nature communications*, 12(1):3478, 2021.

[7] A. Cardillo, J. Gómez-Gardenes, M. Zanin, M. Romance, D. Papo, F. d. Pozo, and S. Boccaletti. Emergence of network features from multiplexity. *Scientific reports*, 3(1):1344, 2013.

[8] L. Chen, N. Josephs, L. Lin, J. Zhou, and E. D. Kolaczyk. A spectral-based framework for hypothesis testing in populations of networks, 2020.

[9] L. Chen, J. Zhou, and L. Lin. Hypothesis testing for populations of networks, 2021.

[10] L. H. Y. Chen, L. Goldstein, and Q.-M. Shao. *Normal Approximation by Stein's Method*. Springer, 2010. ISBN 9783642150074. URL https://openlibrary.org/books/OL37140834M/Normal_Approximation_by_Stein's_Method.

[11] N. Eagle and A. Pentland. Reality mining: sensing complex social systems. *Personal and ubiquitous computing*, 10:255–268, 2006.

[12] Z. Fan, C. Mao, Y. Wu, and J. Xu. Spectral graph matching and regularized quadratic relaxations: Algorithm and theory. In H. D. III and A. Singh, editors, *Proceedings of the 37th International Conference on Machine Learning*, volume 119 of *Proceedings of Machine Learning Research*, pages 2985–2995. PMLR, 13–18 Jul 2020. URL https://proceedings.mlr.press/v119/fan20a.html.

[13] S. Feizi, G. Quon, M. Recamonde-Mendoza, M. Médard, M. Kellis, and A. Jadbabaie. Spectral alignment of graphs. *IEEE Transactions on Network Science and Engineering*, 7(3):1182–1197, 2020. doi: 10.1109/TNSE.2019.2913233.

[14] C. Gao, Y. Lu, and H. H. Zhou. Rate-optimal graphon estimation. *Annals of Statistics*, 43(6):2624–2652, 2015.

[15] D. Ghoshdastidar, M. Gutzeit, A. Carpentier, and U. von Luxburg. Two-sample hypothesis testing for inhomogenous random graphs. *Annals of Statistics*, 48(4):2208–2229, 2020.

[16] C. E. Ginestet, J. Li, P. Balachandran, S. Rosenberg, and E. D. Kolaczyk. Hypothesis testing for network data in functional neuroimaging. *Annals of Applied Statistics*, 11(2):725 – 750, 2017. doi: 10.1214/16-AOAS1015. URL https://doi.org/10.1214/16-AOAS1015.

[17] P. W. Holland, K. B. Laskey, and S. Leinhardt. Stochastic blockmodels: First steps. *Social networks*, 5(2):109–137, 1983.

[18] J. Jin, Z. T. Ke, S. Luo, and Y. Ma. Optimal network pairwise comparison. *Journal of the American Statistical Association*, 120(550):1048–1062, 2025.

[19] N. Josephs, A. A. Amini, M. Paez, and L. Lin. Nested stochastic block model for simultaneously clustering networks and nodes. *arXiv preprint arXiv:2307.09210*, 2023.

[20] R. W. Keener. *Theoretical Statistics: Topics for a Core Course*. Springer Texts in Statistics. Springer, New York, 2010. ISBN 9780387938387.

[21] O. Klopp, A. B. Tsybakov, and N. Verzelen. Oracle inequalities for network models and sparse graphon estimation. *The Annals of Statistics*, 45(1):316 – 354, 2017. doi: 10.1214/16-AOS1454. URL `https://doi.org/10.1214/16-AOS1454`.

[22] E. Kolaczyk, L. Lin, S. Rosenberg, J. Xu, and J. Walters. Averages of unlabeled networks: Geometric characterization and asymptotic behavior, 2019.

[23] H. W. Kuhn. The Hungarian Method for the Assignment Problem. *Naval Research Logistics Quarterly*, 2(1–2):83–97, March 1955. doi: 10.1002/nav.3800020109.

[24] F. T. Leighton and G. l. Miller. Certificates for graphs with distinct eigen values. Orginal Manuscript, 1979.

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

[26] V. Lyzinski, D. E. Fishkind, M. Fiori, J. T. Vogelstein, C. E. Priebe, and G. Sapiro. Graph matching: Relax at your own risk. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 38(1):60–73, 2016. doi: 10.1109/TPAMI.2015.2424894.

[27] P.-A. G. Maugis, S. C. Olhede, C. E. Priebe, and P. J. Wolfe. Testing for equivalence of network distribution using subgraph counts. *Journal of Computational and Graphical Statistics*, 29 (3):455–465, apr 2020. doi: 10.1080/10618600.2020.1736085. URL `https://doi.org/10.1080%2F10618600.2020.1736085`.

[28] S. S. Mukherjee, P. Sarkar, and L. Lin. On clustering network-valued data. *Advances in Neural Information Processing Systems*, 30, 2017.

[29] S. C. Olhede and P. J. Wolfe. Network histograms and universality of blockmodel approximation. *Proceedings of the National Academy of Sciences*, 111(41):14722–14727, 2014.

[30] O. H. M. Padilla, Y. Yu, and C. E. Priebe. Change point localization in dependent dynamic nonparametric random dot product graphs. *Journal of Machine Learning Research*, 23(1): 10661–10719, 2022.

[31] Y. Park, H. Wang, T. Nöbauer, A. Vaziri, and C. E. Priebe. Anomaly detection on whole-brain functional imaging of neuronal activity using graph scan statistics. *Neuron*, 2(3,000):4–000, 2015.

[32] R. Prevedel, Y.-G. Yoon, M. Hoffmann, N. Pak, G. Wetzstein, S. Kato, T. Schrödel, R. Raskar, M. Zimmer, E. S. Boyden, et al. Simultaneous whole-animal 3d imaging of neuronal activity using light-field microscopy. *Nature methods*, 11(7):727–730, 2014.

[33] N. Ross. Fundamentals of Stein's method. *Probability Surveys*, 8(none):210 – 293, 2011. doi: 10.1214/11-PS182. URL `https://doi.org/10.1214/11-PS182`.

[34] D. Schoch. *networkdata: Repository of Network Datasets (v0.1.14)*, 2022. URL `https://doi.org/10.5281/zenodo.7189928`.

[35] L. Shen, A. Amini, N. Josephs, and L. Lin. Bayesian community detection for networks with covariates. *arXiv preprint arXiv:2203.02090*, 2022.

[36] D. Spielman. *Spectral and Algebraic Graph Theory*. 2019.

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

[38] S. Umeyama. An eigendecomposition approach to weighted graph matching problems. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 10(5):695–703, 1988. doi: 10.1109/34.6778.

[39] J. T. Vogelstein, J. M. Conroy, V. Lyzinski, L. J. Podrazik, S. G. Kratzer, E. T. Harley, D. E. Fishkind, R. J. Vogelstein, and C. E. Priebe. Fast approximate quadratic programming for graph matching. *PLOS ONE*, 10(4):1–17, 04 2015. doi: 10.1371/journal.pone.0121002. URL `https://doi.org/10.1371/journal.pone.0121002`.

[40] P. Yanardag and S. Vishwanathan. Deep graph kernels. In *Proceedings of the 21th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, KDD '15, page 1365–1374, New York, NY, USA, 2015. Association for Computing Machinery. ISBN 9781450336642. doi: 10.1145/2783258.2783417. URL https://doi.org/10.1145/2783258.2783417.

[41] Y. Yu, O. H. M. Padilla, D. Wang, and A. Rinaldo. Optimal network online change point localisation. *arXiv preprint arXiv:2101.05477*, 2021.

[42] A. Y. Zhang and H. H. Zhou. Minimax rates of community detection in stochastic block models. *The Annals of Statistics*, 44(5):2252–2280, 2016.

A Full Theoretical Statements

This section provides the formal definitions and the complete versions of the theoretical results presented in Section 4. We follow a top-down approach, beginning with the general theorems expressed via the misclassification rate (Tier 2), then presenting the full finite-sample bounds from which they are derived (Tier 3), and finally providing the complete proofs.

A.1 Formal Definition and Matching Guarantee

We begin with the formal definition of an (η, θ)-friendly matrix, which is a key condition for our spectral matching algorithm.

Definition A.1 ((η, θ)-friendly)**.** Let $B \in [0,1]^{K\times K}$ be a symmetric matrix with eigenvalue decomposition $B = \sum_{k=1}^{K} \lambda_k q_k q_k^\top$. We call B (η, θ)-friendly if:

$$\min_{1\le k\neq \ell\le K} |\lambda_k - \lambda_\ell| > \eta, \qquad \text{(Eigenvalue Separation)} \tag{12}$$

$$\min_{k\in[K]} |q_k^\top \mathbf{1}| > \theta. \qquad \text{(Eigenvector Condition)} \tag{13}$$

We consider connectivity matrices B_1 and B_2 that are (θ, η)-friendly. Recall our shorthand notation for an exact matching, introduced in (47). Similarly, it is helpful to introduce the following graphical mnemonic for approximate matching:

$$B_1 \overset{P}{\rightsquigarrow} B_2 \quad \Longleftrightarrow \quad \|B_2 - PB_1P^\top\|_F \le \frac{\theta\eta}{2\sqrt{2K}}. \tag{14}$$

Then, we have following guarantee for the matching algorithm:

Theorem A.1 (Matching Consistency)**.** Suppose that B_1 and B_2 are $K \times K$ connectivity matrices that are (θ, η)-friendly with $\theta < \frac{1}{4\sqrt{K}}$, and with EVDs given by $B_r = Q_r \Lambda Q_r^\top$, $r = 1, 2$. Moreover, assume $B_1 \overset{P^*}{\longrightarrow} B_2$ for some permutation matrix P^*. For $r = 1, 2$, let $\widehat{B}_r$ be an estimate of B_r satisfying

$$\widehat{B}_r \overset{P_r}{\rightsquigarrow} B_r \tag{15}$$

for some permutation matrices P_r. Let $\widehat{Q}_1, \widehat{Q}_2$, $\widehat{S}$ and $\widetilde{P}$ be as defined in Algorithm 1, and let

$$S_r = \operatorname*{argmin}_{S\in\Psi_K} \|Q_r - P_r\widehat{Q}_r S\|_F, \quad r = 1, 2, \tag{16}$$

where Ψ_K is the set of $K \times K$ diagonal sign matrices. Then, the following holds:

(a) *Sign recovery:* $\widehat{S} = \widetilde{S} := S_1 S^* S_2$ with S^* as in Lemma 2.1.
(b) *Permutation recovery:* $\mathrm{LAP}(\widehat{Q}_1 \widetilde{S} \widehat{Q}_2^\top) = \{\widetilde{P}\}$ where $\widetilde{P} := P_2^\top P^* P_1$, and

$$\|\widehat{B}_2 - \widetilde{P}\widehat{B}_1\widetilde{P}^\top\|_F \le \sum_{r=1}^{2} \|B_r - P_r\widehat{B}_r P_r^\top\|_F. \tag{17}$$

Remark A.2. One can also ask whether our matching algorithm is *self-consistent*? That is, if we reverse the order of the $\widehat{B}_1$ and $\widehat{B}_2$, do we get permutations that are transposes of each other? The

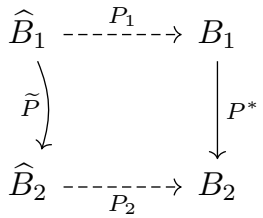


Figure 3: Schematic diagram of permutation recovery in Theorem A.1. The solid and dashed straight arrows correspond to exact and approximate match. The bent arrow represents an application of the matching algorithm $\mathcal{M}$.

answer is yes. Notice that if we reverse the order of $\widehat{B}_1$ and $\widehat{B}_2$, the sign matrix $\hat{S}$ does not change. Moreover,

$$B_1 \xrightarrow{P^*} B_2 \quad \text{implies} \quad B_2 \xrightarrow{(P^*)^\top} B_1.$$

Hence, $\mathrm{LAP}(\widehat{Q}_2\widetilde{S}\widehat{Q}_1^\top) = \{P'\}$ where $P' = P_1^\top (P^*)^\top P_2 = (P_2^\top P^* P_1)^\top = \widetilde{P}^\top$.

Intuituon behind Theorem A.1 To gain a deeper understanding of why our target permutation matrix is $P_2^\top P^* P_1$, let us analyze Figure 3. Within this illustration, each edge symbolizes a matching between matrices based on both the direction of the edge and the permutation matrix associated with it. To clarify the nature of these matchings, we use different types of arrows:

1. Dashed straight arrows represent approximate matching. For instance, the top edge corresponds to $\widehat{B}_1 \overset{P_1}{\rightsquigarrow} B_1$, following the notation from Equation (14).
2. In contrast, the solid straight solid arrow signifies exact matching corresponding to $B_1 \xrightarrow{P^*} B_2$, with the notation from (5).

Considering that the inverse of a permutation matrix is its transpose, we can trace the path from $\widehat{B}_1$ to $\widehat{B}_2$ in Figure 3. We start from $\widehat{B}_1$, progress through B_1, then B_2, and finally arrive at $\widehat{B}_2$. The permutation matrices associated with this path are P_1, P^* and $P_2^\top$ (the inverse of P_2). By multiplying these matrices together, we obtain the desired permutation matrix, which is $P_2^\top P^* P_1$.

A.2 General Results via Misclassification Rate (Tier 2)

The simplified theorems in the main paper (Theorems 4.2 and 4.3) are direct consequences of the more general results presented here. These theorems are applicable to *any* community detection algorithm, with their conditions expressed in terms of the algorithm's misclassification rate, $\widetilde{\varepsilon} := \max_{r,t} \mathrm{Mis}(z_{rt}, \widehat{z}_{rt})$. In what follows, $N = \max\{N_1, N_2\}$.

Theorem A.3 (Null Distribution, Tier 2)**.** Assume Algorithm 2 is applied to two SBM samples generated from $B = (\nu_n/n)B^0$ for a fixed B^0. The null distribution $\widehat{T}_n \Rightarrow \chi^2_{K(K+1)/2}$ holds if:

(i) B^0 has distinct eigenvalues and no eigenvector of it sums to zero.

(ii) The number of networks grows at most polynomially: $N = o(n^\alpha)$ for some $\alpha > 0$.

(iii) The misclassification rate $\widetilde{\varepsilon}$ satisfies both:

$$\sqrt{\nu_n n^{1+\alpha}} \cdot \widetilde{\varepsilon} = o_p(1),$$
$$\widetilde{\varepsilon} \le CK^{-3/2} \quad \text{for some constant } C = C(B^0).$$

Theorem A.4 (Consistency, Tier 2)**.** Assume Algorithm 2 is applied to two SBM samples from $B_1 = (\nu_n/n)B_1^0$ and $B_2 = (\nu_n/n)B_2^0$. The test is consistent, with $\widehat{T}_n = \Omega(N\nu_n^2)$, if:

(i) B_1^0 and $PB_2^0P^\top$ are different for any permutation matrix P.

(ii) The misclassification rate $\widetilde{\varepsilon}$ satisfies both:

$$\widetilde{\varepsilon} = o_p(1),$$
$$\widetilde{\varepsilon} \le CK^{-3/2} \quad \text{for some constant } C = C(B_1^0, B_2^0).$$

Remark A.5 (Connecting Tiers)**.** The Tier 1 results in the main paper follow from the theorems above. Near-optimal community detection algorithms [42] can achieve $\widetilde{\varepsilon} = O(N\exp(-c\nu_n/K))$. Taking $\nu_n = C\log n$ for a sufficiently large C, this rate is small enough to satisfy all conditions on $\widetilde{\varepsilon}$ in Theorems A.3 and A.4 for sufficiently large n (assuming K is fixed). Note that in the main paper, we stated $N = O(n^a)$ while here $N = o(n^\alpha)$, noting that the former implies the latter by taking say $a \le \alpha/2$ for any $\alpha > 0$. The next section details how these Tier 2 conditions are, in turn, derived from our full finite-sample analysis.

A.3 Complete Finite-Sample Bounds (Tier 3)

The Tier 2 results presented above are convenient simplifications of our full, non-asymptotic technical bounds (Tier 3). We state these full bounds below and explain how the Tier 2 conditions arise from them.

We make the following "sparsity" and "size" assumptions

$$\frac{\nu_n}{n} \;\le\; B_{k\ell} \;\le\; \frac{C_1\nu_n}{n}, \tag{18}$$
$$n \;\le\; n_{rt} \le C_2 n \tag{19}$$

for all $k, \ell \in [K]$ and $t \in [N_r], r = 1, 2$. We also write $\pi_k = \mathbb{P}(z_{rti} = k)$, recalling that we are under the null, and set $\pi_{\min} = \min_k \pi_k$.

Let us fix some $\kappa \in (0,1)$ and $\alpha > 0$ and let $\beta = 1/(\bar{\kappa}\pi_{\min})$ where $\bar{\kappa} = 1 - \kappa$. For any N, let

$$\delta_n^{(N)} := \sqrt{\frac{3\beta^2\alpha\log n}{Nn\nu_n}}. \tag{20}$$

For convenience, we assume that (n is large enough so that)

$$\delta_n^{(1)} \le 1, \quad \frac{\log n}{n} \le \frac{\kappa^2\pi_{\min}}{3\alpha}. \tag{21}$$

Moreover, define

$$\gamma_n = C_1\Big(56\beta^3 \cdot \widetilde{\varepsilon} + \delta_n^{(1)}\Big), \quad \text{where} \quad \widetilde{\varepsilon} := \max_{\substack{t \in [N_r] \\ r=1,2}} \mathrm{Mis}(z_{rt}, \widehat{z}_{rt}).$$

Theorem A.6 (Null distribution, Tier 3)**.** Fix K and assume that Algorithm 2 is applied to two SBM samples, with sizes satisfying (19), and the same connectivity matrix B satisfying (18) and $\|B\|_{\max} \le 0.99$. Moreover, suppose that B is (η, θ)-friendly and with probability $1 - o(1)$,

$$\frac{\nu_n}{n}\gamma_n \le \min\Big(\theta, \frac{1}{5\sqrt{K}}\Big) \cdot \frac{\eta}{2\sqrt{2}K}. \tag{22}$$

Let $N = \max\{N_1, N_2\}$ and assume that for some $\kappa \in (0,1)$ and $\alpha > 0$ and $c > 0$, we have

$$\gamma_n = o_p(1), \quad \sqrt{N\nu_n n}\,\widetilde{\varepsilon} = o_p(1), \quad N\nu_n n \to \infty, \quad N = o(n^\alpha),$$

and $\min\{N_1, N_2\} \ge cN$. Then, for $\widehat{T}_n$, the output of Algorithm 2, we have

$$\widehat{T}_n \Rightarrow \chi^2_{K(K+1)/2}.$$

Derivation of Tier 2 from Tier 3 (Null): The conditions in Theorem A.3 ensure that the premises of Theorem A.6 hold. Specifically, since we always have $\delta_n^{(1)} = o(1)$, the condition $\gamma_n = o_p(1)$ is satisfied if $\widetilde{\varepsilon} = o_p(1)$. Assuming $B = (\nu_n/n)B^0$, and B_0 has distinct eigenvalues, with no eigenvector that is orthogonal to the all-ones vector, B will be $((\nu_n/n)\eta_0, \theta_0)$-friendly for some constants $\theta_0 > 0$ and $\eta_0 > 0$. As a result condition (22) holds if $\gamma_n \le \min(\theta_0, \frac{1}{5\sqrt{K}}) \cdot \frac{\eta_0}{2\sqrt{2}K}$,

and since $\gamma_n \asymp \widetilde{\varepsilon}$, the condition holds if $\widetilde{\varepsilon} \lesssim K^{-3/2}$. The condition $\sqrt{\nu_n n^{1+\alpha}} \cdot \widetilde{\varepsilon} = o_p(1)$ is a slightly stronger restatement of the $\sqrt{N\nu_n n} \cdot \widetilde{\varepsilon} = o_p(1)$ requirement, using $N = o(n^\alpha)$. Condition $N\nu_n n \to \infty$ trivially holds given $\nu_n = \Omega(\log n)$.

Next, we show that the test is consistent, in the sense that $\widehat{T}_n \to \infty$ under the alternative hypothesis that the two SBMs are different. For two $K \times K$ matrices B_1, B_2, consider the pseudometric

$$d_F^{\mathrm{ma}}(B_1, B_2) := \min_{P \in \Pi_K} \|B_2 - P B_1 P^\top\|_F, \tag{23}$$

where Π_K is the set of $K \times K$ permutation matrices.

Theorem A.7 (Consistency, Tier 3)**.** Assume that Algorithm 2 is applied to two SBM samples, with sizes satisfying (19), and connectivity matrices B_1 and B_2 satisfying (18). For $r = 1, 2$, let

$$\xi_r := C_1\big(40\beta^3\widetilde{\varepsilon} + \delta_n^{(N_r)}\big)$$

and assume that

$$\sqrt{12}K \max_{r=1,2} \xi_r \le \frac{d_F^{\mathrm{ma}}(B_1, B_2)}{\nu_n/n}. \tag{24}$$

Moreover, let let $\mathcal{G}$ be the event that (22) holds. and assume that $48C_2\beta^4\widetilde{\varepsilon} \le 1$. Here, C_1 and C_2 are the same constants as in (18). Then, with probability at least $1 - 19N_+K^2n^{-\alpha} - P(\mathcal{G}^c)$,

$$\widehat{T}_n \ge \frac{Nn^2}{12\beta^2} \cdot d_F^{\mathrm{ma}}(B_1, B_2)^2.$$

Derivation of Tier 2 from Tier 3 (Consistency): The premise of Theorem A.7 is dominated by the misclassification rate $\widetilde{\varepsilon}$ through the term ξ_r. If we assume $\widetilde{\varepsilon} = o_p(1)$ as in Theorem A.4, then $\xi_r \to 0$. Since $d_F^{\mathrm{ma}}(B_1, B_2)/(\nu_n/n) = d_F^{\mathrm{ma}}(B_1^0, B_2^0) > 0$ is a constant under the alternative, the condition holds for large n, guaranteeing consistency. It follows that $\widehat{T}_n \ge \frac{N\nu_n^2}{12\beta^2} d_F^{\mathrm{ma}}(B_1^0, B_2^0)^2 = \Omega(N\nu_n^2)$.

B Proofs of the main results

B.1 Notation

We often consider a one-to-one correspondence between permutations matrices $P \in \mathbb{R}^{K\times K}$ and permutations σ on the set $[K] := \{1, \dots, K\}$. The correspondence $\sigma \mapsto P_\sigma$ is defined by the following identity: $(P_\sigma v)_i = v_{\sigma(i)}$ for all $v \in \mathbb{R}^K$. It follows that $(P_\sigma^\top v)_i = v_{\sigma^{-1}(i)}$. It is also helpful to note that $[P_\sigma]_{i*} = e_{\sigma(i)}^\top$ and $[P_\sigma^\top]_{*j} = [P_\sigma]_{j*}^\top = e_{\sigma(j)}$. This implies that if B is a $K \times K$ matrix, $[P_\sigma B P_\sigma^\top]_{ij} = B_{\sigma(i),\sigma(j)}$. The linear assignment problem is often written as $\max_\sigma \sum_{i=1}^K B_{i,\sigma(i)}$. The cost function in this case is equivalent to $\mathrm{tr}(BP_\sigma^\top)$. This follows by noting $[BP_\sigma^\top]_{ij} = B_{i*}\, e_{\sigma(j)} = B_{i,\sigma(j)}$.

Let P and Q be matrices with associated permutations σ and τ respectively. Notice that $\sigma \circ \tau$ is the permutation associated with QP, since $(QPv)_i = (Pv)_{\tau(i)} = v_{\sigma(\tau(i)))}$.

For $p \in [1, \infty)$ and a $n \times m$ matrix A, the $\ell_p \to \ell_p$ operator norm of a matrix $A = (a_{ij})$ is $\|A\|_p := \sup_{x\neq 0} \|Ax\|_p / \|x\|_p$. In the special cases $p = 1, \infty$, we have $\|A\|_1 = \max_j \sum_i |a_{ij}|$ and $\|A\|_\infty = \max_i \sum_j |a_{ij}|$. We also write $\|A\|_{\max} = \max_{i,j} |A_{ij}|$.

For label vectors $z, \widehat{z} \in [K]^n$, we write $d_{\mathrm{H}}(z, \widehat{z}) = \sum_{i=1}^n 1\{z_i \neq \widehat{z}_i\}$ for their Hamming distance, $d_{\mathrm{NH}}(z, \widehat{z}) = d_{\mathrm{H}}(z, \widehat{z})/n$ for their the normalized Hamming distance, and $\mathrm{Mis}(z, \widehat{z}) = \min_{\sigma\in\Pi_K} d_{\mathrm{NH}}(z, \widehat{z} \circ \sigma)$ for the corresponding misclassification rate.

B.2 Proof of Theorem A.1 (Matching Consistency)

Recall that $\|\Delta\|_\infty = \max_{i,j} |\Delta_{ij}|$. We need the following lemma on the perturbation of the LAP problem:

Lemma B.1. We have $\mathrm{LAP}(I + \Delta) = \{I\}$ as long as $\|\Delta\|_\infty < 1/2$.

Since B_1 and B_2 are similar matrices, they have the same eigenvalues. If the EVD of B_1 is $B_1 = Q_1 \Lambda Q_1^\top$, then the EVD of B_2 can written as $B_2 = Q_2 \Lambda Q_2^\top$. By Lemma 2.1,

$$Q_2 = P^* Q_1 S^* \tag{25}$$

for some sign matrix S^*.

First, we show assertion (a). Let $\widehat{\Delta}_r := P_r \widehat{Q}_r S_r - Q_r$, so that $\widehat{Q}_r = P_r^\top (Q_r + \widehat{\Delta}_r) S_r$ and

$$\widehat{Q}_r^\top \mathbf{1} = S_r (Q_r + \widehat{\Delta}_r)^\top \mathbf{1}$$

using $P_r \mathbf{1} = \mathbf{1}$. Then, $[\widehat{Q}_r^\top \mathbf{1}]_k = S_{r,kk}(Q_{r,*k}^\top \mathbf{1} + \widehat{\Delta}_{r,*k}^\top \mathbf{1})$ where $Q_{r,*k}$ and $\Delta_{r,*k}$ are the k-th columns of Q_r and Δ_r. From the definition of $\hat{S}_{kk}$,

$$\widehat{S}_{kk} = \operatorname{sign}\left(\frac{[\widehat{Q}_1^\top \mathbf{1}]_k}{[\widehat{Q}_2^\top \mathbf{1}]_k}\right) = \operatorname{sign}\left(\frac{Q_{1,*k}^\top \mathbf{1} + \widehat{\Delta}_{1,*k}^\top \mathbf{1}}{Q_{2,*k}^\top \mathbf{1} + \widehat{\Delta}_{2,*k}^\top \mathbf{1}}\right) S_{1,kk} S_{2,kk}. \tag{26}$$

From (25), it follows that $S_{kk}^* = \operatorname{sign}\big([Q_2^\top \mathbf{1}]_k / [Q_1^\top \mathbf{1}]_k\big)$. Then, to show $\widehat{S}_{kk} = S_{1,kk} S_{kk}^* S_{2,kk}$, it suffices to show that

$$|Q_{r,*k}^\top \mathbf{1}| = \Big|\sum_{j=1}^K Q_{r,jk}\Big| > \Big|\sum_{j=1}^K \widehat{\Delta}_{r,jk}\Big| = |\widehat{\Delta}_{r,*k}^\top \mathbf{1}| \tag{27}$$

for $r = 1, 2$. Since B_r are (θ, η)-friendly, we have $|\sum_{j=1}^K Q_{r,jk}| \geq \theta$ and so it is enough to show that $|\sum_{j=1}^K \widehat{\Delta}_{r,jk}| \leq \|\widehat{\Delta}_{r,*k}\|_1 < \theta$.

The noise matrices $\widehat{\Delta}_r$ are controlled by the Davis–Kahan theorem,

$$\begin{aligned}\|\widehat{\Delta}_{r,*k}\|_2 &= \|Q_{r,*k} - P_r \widehat{Q}_{r,*k} S_{r,kk}\|_2 \\ &\leq \frac{2\sqrt{2}}{\eta} |\!|\!| B_r - P_r \widehat{B}_r P_r^\top |\!|\!|_F.\end{aligned}$$

Then, using assumption (15),

$$\|\widehat{\Delta}_{r,*k}\|_1 \leq \sqrt{K} \|\widehat{\Delta}_{r,*k}\|_2 \leq \frac{2\sqrt{2K}}{\eta} \frac{\theta\eta}{2\sqrt{2}K} \leq \theta, \tag{28}$$

proving assertion (a). Note also that we have shown

$$|\!|\!| \widehat{\Delta}_r |\!|\!|_1 = \max_k \|\widehat{\Delta}_{r,*k}\|_1 \leq \theta. \tag{29}$$

Next, we prove $\mathrm{LAP}(\widehat{Q}_1 \widetilde{S} \widehat{Q}_2^\top) = \{\widetilde{P}\}$. Using (25), and the definitions of $\widehat{\Delta}_r$, we have

$$\begin{aligned}\widehat{Q}_2 &= P_2^\top (Q_2 + \widehat{\Delta}_2) S_2 \\ &= P_2^\top (P^* Q_1 S^* + \widehat{\Delta}_2) S_2 \\ &= P_2^\top (P^* (P_1 \widehat{Q}_1 S_1 - \widehat{\Delta}_1) S^* + \widehat{\Delta}_2) S_2 \\ &= P_2^\top P^* P_1 \widehat{Q}_1 S_1 S^* S_2 - P_2^\top P^* \widehat{\Delta}_1 S^* S_2 + P_2^\top \widehat{\Delta}_2 S_2 \\ &= \widetilde{P} \widehat{Q}_1 \widetilde{S} + \Delta\end{aligned}$$

where we let $\Delta = -P_2^\top P^* \widehat{\Delta}_1 S^* S_2 + P_2^\top \widehat{\Delta}_2 S_2$. We can then write

$$\widehat{Q}_1 \widetilde{S} \widehat{Q}_2^\top = \widehat{Q}_1 \widetilde{S} (\widetilde{P} \widehat{Q}_1 \widetilde{S} + \Delta)^\top = \widetilde{P}^\top (I + \Delta_0). \tag{30}$$

where $\Delta_0 = \widetilde{P} \widehat{Q}_1 \widetilde{S} \Delta^\top$. It is then enough to study

$$\mathrm{LAP}\big(\widetilde{P}^\top (I + \Delta_0)\big) = \mathrm{LAP}(I + \Delta_0) \cdot \widetilde{P}$$

where the equality follows by a change-of-variable argument. The result follows from Lemma B.1 if we show $\|\Delta_0\|_\infty \leq 1/2$.

We note that for permutation and sign matrices, both $\|\!| \cdot |\!\|_\infty$ and $\|\!| \cdot |\!\|_1$ are equal to 1. Using the submultiplicative property of $\|\!| \cdot |\!\|_p$ for $p = 1, 2$, we have

$$\|\!|\Delta^\top|\!\|_\infty = \|\!|\Delta|\!\|_1 \le \|\!|\widehat{\Delta}_1|\!\|_1 + \|\!|\widehat{\Delta}_2|\!\|_1 < 2\theta$$

where we have used (29). Then,

$$\|\!|\Delta_0|\!\|_\infty \le \|\!|\widehat{Q}_1|\!\|_\infty \|\!|\Delta^\top|\!\|_\infty \le 2\theta \|\!|\widehat{Q}_1|\!\|_\infty.$$

Since $\widehat{Q}_1$ has unit-norm rows, that is, $\|\widehat{Q}_{1,k*}\|_2 = 1$ for all k, we obtain

$$\|\!|\widehat{Q}_1|\!\|_\infty = \max_k \|\widehat{Q}_{1,k*}\|_1 \le \sqrt{K}.$$

Putting the pieces together

$$\|\Delta_0\|_\infty \le \|\!|\Delta_0|\!\|_\infty \le 2\theta\sqrt{K} < 1/2$$

where the last inequality is by assumption. This proves $\mathrm{LAP}(\widehat{Q}_1 \widetilde{S} \widehat{Q}_2^\top) = \{\widetilde{P}\}$.

To prove the inequality in part (b), let $D_r := B_r - P_r \widehat{B}_r P_r^\top$. Then, some algebra, using $B_2 = P^* B_1 P^{*T}$, gives

$$\begin{aligned}
\widetilde{P}\widehat{B}_1\widetilde{P}^\top &= P_2^\top P^* P_1 \widehat{B}_1 P_1^\top P^{*T} P_2 \\
&= P_2^\top P^* (B_1 - D_1) P^{*T} P_2 \\
&= P_2 P^* B_1 P^{*T} P_2 - P_2^\top P^* D_1 P^{*T} P_2 \\
&= P_2^\top B_2 P_2 - P_2^\top P^* D_1 P^{*T} P_2 \\
&= P_2^\top D_2 P_2 + \widehat{B}_2 - P_2^\top P^* D_1 P^{*T} P_2,
\end{aligned}$$

and so

$$\begin{aligned}
\|\!|\widetilde{P}\widehat{B}_1\widetilde{P}^\top - \widehat{B}_2|\!\|_F &\le \|\!|P_2^\top D_2 P_2 - P_2^\top P^* D_1 P^{*T} P_2|\!\|_F \\
&\le \|\!|D_1|\!\|_F + \|\!|D_2|\!\|_F.
\end{aligned}$$

The proof is complete.

B.3 Proof of Theorem A.6 (Null distribution, Tier 3)

If B is (η, θ)-friendly, then it is (η, θ')-friendly with $\theta' = \min(\theta, \frac{1}{5\sqrt{K}})$. For simplicity, let us redefine θ to be θ', so that B is (η, θ)-friendly with $\theta < \frac{1}{4\sqrt{K}}$. Let $\mathcal{D}$ be the event that (22) holds. After redefinition, $\mathcal{D}$ is equivalent to

$$\frac{\nu_n}{n}\gamma_n \le \frac{\theta\eta}{2\sqrt{2}K}, \tag{31}$$

and by assumption, we have $P(\mathcal{D}) = 1 - o(1)$. Let B_r^* be some (a priori) fixed version of the connectivity matrix for each of the two groups $r = 1, 2$. By Proposition F.3 and the union bound, there is an event $\mathcal{A}$ with

$$P(\mathcal{A}^c) \le 3N_+K^2 n^{-\alpha} + KN_+ e^{-\kappa^2 n \pi_{\min}/3}$$

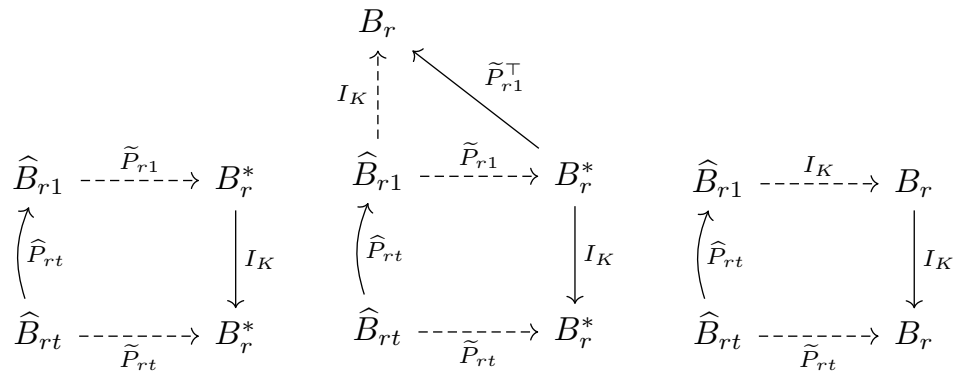
and permutation matrices $\widetilde{P}_{rt} \in \Pi_K$ such that on $\mathcal{A} \cap \mathcal{D}$:

$$\|\widetilde{P}_{rt}\widehat{B}_{rt}\widetilde{P}_{rt}^\top - B_r^*\|_{\max} \;\le\; C_1 \frac{\nu_n}{n}\Big(56\beta^2 \cdot \widetilde{\varepsilon}_{rt} + \delta_n^{(1)}\Big) \;\le\; \frac{\nu_n}{n}\gamma_n \;\le\; \frac{\theta\eta}{2\sqrt{2}K}$$

for all $t \in [N_r]$ and $r = 1, 2$, where $\widetilde{\varepsilon}_{rt} = \mathrm{Mis}(z_{rt}, \widehat{z}_{rt})$. This implies that

$$\mathcal{A} \cap \mathcal{D} \;\subset\; \mathcal{E}_2 := \Big\{\widehat{B}_{rt} \overset{\widetilde{P}_{rt}}{\rightsquigarrow} B_r^* \quad \text{for all } t = 1, \ldots, N_r,\ r = 1, 2\Big\}.$$

According to Lemma E.3, we can assume that, in the above, $\widetilde{P}_{rt}$ is independent of everything else.



(a) Step 5 (b) Simplification by introducing B_r. (c) After conditioning on B_r

Figure 4: Matching to first network in group r. Here, $t > 1$ in the bottom level. Bent arrow is an application of the matching algorithm $\mathcal{M}$. The edges can be reversed in which case the permutation matrix should be replaced with its transpose. Left-side quantities are random, while right-side quantities are deterministic. Dashed and solid arrows correspond to approximate and exact matching, respectively. See the discussion at the end of Section 4.1 for more details on the nature of these diagrams.

Figure 4a illustrates the "matching-to-first" in step 5 of the algorithm. As this diagram shows, by Theorem A.1, on event $\mathcal{E}_2$, we have $\widehat{P}_{rt}^\top = \widetilde{P}_{rt}^\top I_K \widetilde{P}_{r1}$ that is,

$$\mathcal{E}_2 \subset \mathcal{E}_1 := \big\{\widehat{P}_{rt} = \widetilde{P}_{r1}^\top \widetilde{P}_{rt}\big\}.$$

To simplify, let us define $B_r := \widetilde{P}_{r1}^\top B_r^* \widetilde{P}_{r1}$, that is,

$$B_r^* \xrightarrow{\widetilde{P}_{r1}^\top} B_r.$$

Note that B_r is random version of the true B_r^*, due to the randomness of $\widetilde{P}_{r1}$, although, it is independent of everything else. Then, a little algebra shows that on $\mathcal{E}_1 \cap \mathcal{E}_2$,

$$\widehat{B}_{rt} \overset{\widehat{P}_{rt}}{\rightsquigarrow} B_r$$

for all $t \in [N_r]$ and $r = 1, 2$. Figure 4b illustrates the above inequality (the red path). Now, since $B_r, r = 1, 2$ are independent of everything else, we can condition on them and continue with the argument as if they were deterministic. The resulting diagram is shown in Figure 4c; the effect is as if we assumed $\widetilde{P}_{r1} = I_K$ and $B_r = B_r^*$. The above conditioning argument shows that we can do this without loss of generality.

From now on, we work on $(\mathcal{A} \cap \mathcal{D}) \cap \mathcal{E}_1 \cap \mathcal{E}_2 = \mathcal{A} \cap \mathcal{D}$ (by the above argument), on which we have $\widehat{P}_{rt} = \widetilde{P}_{rt}$ as discussed above. Then, from the definition of $\widehat{B}_r$ in step 7, we note

$$\widehat{B}_r - B_r = \frac{1}{N_r} \sum_{t=1}^{N_r} (\widetilde{P}_{rt} \widehat{B}_{rt} \widetilde{P}_{rt}^\top - B_r).$$

On $\mathcal{E}_2$ the $\|\!|\cdot|\!\|_F$ of each term on the RHS is bounded by $\theta\eta/(2\sqrt{2}K)$, and since a norm is a convex function, the same holds for the LHS. That is,

$$\widehat{B}_r \overset{I_K}{\rightsquigarrow} B_r \tag{32}$$

for $r = 1, 2$. Since we are under the null, there is a permutation P^* such that $B_2 \xrightarrow{P^*} B_1$. Combining with (32), we can applyTheorem A.1—with $P_1 = P_2 = I_K$ and the roles of B_1 and B_2 switched—to conclude that $\widehat{P} = P^*$ in step 8.

Remark B.1. We could have assumed $B_1^* = B_2^*$ in the above argument, since we are under the null. However, when passing to $B_r, r = 1, 2$ we could lose the equality among B_1 and B_2 (due to $\widetilde{P}_{r1}, r = 1, 2$ potentially being different). Hence, we do not gain anything by making the assumption $B_1^* = B_2^*$.

Remark B.2. Everything up to and including (32) holds under the alternative $B_1 \neq B_2$ as well. This will be used in the proof of Theorem A.7.

The following arguments are all on $\mathcal{A}$. Let $S_{rt} = \mathcal{S}(A_{rt}, z_{rt})$, $m_{rt} = \mathcal{C}(A_{rt}, z_{rt})$ and

$$\widetilde{B}_r := \frac{S_r}{m_r}, \quad S_r := \textstyle\sum_t S_{rt}, \quad m_r := \textstyle\sum_t m_{rt}$$

Since $\widehat{P}_{rt} = \widetilde{P}_{rt}$, we have

$$\frac{\widehat{S}_r}{\widehat{m}_r} = \frac{\sum_t \widetilde{P}_{rt}\widehat{S}_{rt}\widetilde{P}_{rt}^\top}{\sum_t \widetilde{P}_{rt}\widehat{m}_{rt}\widetilde{P}_{rt}^\top}$$

where $\widehat{S}_r$ and $\widehat{m}_r$ are as defined in step 9. Let $N = \max\{N_1, N_2\}$. Then, from Proposition F.4 and union bound on $r = 1, 2$, there is an event $\mathcal{B}_1$ with

$$P(\mathcal{B}_1^c) \leq 2(C_2 N + 2K^2)n^{-\alpha} + 2NKe^{-\kappa^2 n\pi_{\min}/3} \tag{33}$$

such that on $\mathcal{B}_1$, we have

$$\|(\widehat{S}_r/\widehat{m}_r) - \widetilde{B}_r\|_{\max} \;\leq\; 40\, C_1\beta^3 \frac{\nu_n}{n}\widetilde{\varepsilon} \tag{34}$$

where $\widetilde{\varepsilon} = \max_{r,t} \widetilde{\varepsilon}_{rt}$. Also, since $\widehat{P} = P^*$, we have $\widehat{S}_2' = P^*\widehat{S}_2 P^{*T}$ and $\widehat{m}_2' = P^*\widehat{m}_2 P^{*T}$. Moreover, from Proposition F.4, on $\mathcal{B}_1$,

$$\|\widetilde{B}_r - B_r\|_{\max} \leq C_1 \frac{\nu_n}{n}\delta_n^{(N_r)} \leq C_1 \frac{\nu_n}{n} \tag{35}$$

since $\delta_n^{(N_r)} \leq \delta_n^{(1)} \leq 1$. Using $\|B_r\|_{\max} \leq C_1\nu_n/n$ and triangle inequality, we have

$$\|\widetilde{B}_r\|_{\max} \lesssim \frac{\nu_n}{n}, \quad \|\widehat{S}_r/\widehat{m}_r\|_{\max} \lesssim \frac{\nu_n}{n} \tag{36}$$

where we have treated C_1, C_2 and β^3 as constants and absorbed them into $\lesssim$ symbol; this will do from time to time in the rest of the proof.

Next, we have $\|m_{rt} - \widetilde{P}_{rt}\widehat{m}_{rt}\widetilde{P}_{rt}^\top\|_{\max} \leq 6\beta\, n_{rt}^2\, \widetilde{\varepsilon}_{rt}$. It follows that

$$\|m_r - \widehat{m}_r\|_{\max} \;\leq\; 6\beta \sum_t n_{rt}^2\, \widetilde{\varepsilon}_{rt} \;\leq\; 6C_2^2\beta^2 N_r n^2 \widetilde{\varepsilon} \tag{37}$$

where $m_r = \sum_t m_{rt}$. Let $m_2' = P^* m_2 P^{*T}$. Then, the same bound as above holds for $\|m_2' - \widehat{m}_2'\|_{\max}$. Let h be the elementwise harmonic mean of m_1 and m_2'. Since $\widehat{h}$ is the elementwise harmonic mean of $\widehat{m}_1$ and $\widehat{m}_2'$, we have

$$\|h - \widehat{h}\|_{\max} \leq 12C_2\beta^2 N n^2 \widetilde{\varepsilon} \tag{38}$$

where $N = \max\{N_1, N_2\}$. Since h is elementwise the harmonic mean of $m_r, r = 1, 2$, we have

$$\min_{k,\ell} h_{k\ell} \geq \min_{k,\ell,r} [m_r]_{k\ell} \geq cNn^2/(2\beta^2). \tag{39}$$

The factor 2 is for handling the case $k = \ell$.

Controlling $\widehat{\sigma}$. Note that

$$\widehat{B} = \frac{\widehat{S}_1 + P^*\widehat{S}_2 P^*}{\widehat{m}_1 + P^*\widehat{m}_2 P^{*T}}.$$

Since $B_1 = P^* B_2 P^{*T}$, this is essentially an estimator like $\widehat{B}$ in Proposition F.4 based on an independent sample of size $N_+ := N_1 + N_2$ from $\mathsf{SBM}(B_1, \pi)$. It follows from Proposition F.4 that there is an event $\mathcal{B}_2$ with

$$P(\mathcal{B}_2^c) \leq (C_2 N_+ + 2K^2)n^{-\alpha} + N_+ K e^{-\kappa^2 n\pi_{\min}/3} \tag{40}$$

such that on $\mathcal{B}_2$, we have

$$\|\widehat{B} - B_1\|_{\max} \le C_1 \frac{\nu_n}{n}\Big(40\beta^3\widetilde{\varepsilon} + \delta_n^{(N_+)}\Big) \;\le\; \frac{\nu_n}{n}\gamma_n$$

where the second inequality is by $\delta_n^{(N_+)} \le \delta_n^{(1)}$.

Let $\sigma^2 = (\sigma_{k\ell}^2)$ where $\sigma_{k\ell}^2 = B_{1k\ell}(1-B_{1k\ell})$. The function $f(x) = x(1-x)$ has derivative satisfying $|f'(x)| \le 1$ for $x \in [0,1]$, hence f is 1-Lipschitz there implying

$$\|\widehat{\sigma}^2 - \sigma^2\|_{\max} \le \|\widehat{B} - B_1\|_{\max}.$$

Since $\nu_n/n \le B_{1k\ell} \le 0.99$, we have $\sigma_{k\ell}^2 \ge 0.01\nu_n/n$. Ignoring constants, we have shown

$$\|\widehat{B} - B_1\|_{\max} \le \frac{\nu_n}{n}\gamma_n, \quad \|\widehat{\sigma}^2 - \sigma^2\|_{\max} \le \frac{\nu_n}{n}\gamma_n, \quad \min_{k,\ell}\widehat{\sigma}_{k\ell}^2 \gtrsim \frac{\nu_n}{n}.$$

High probability event. Let $\mathcal{B} = \mathcal{B}_1 \cap \mathcal{B}_2$. Then, the event $\mathcal{A} \cap \mathcal{B}$ has high probability. Indeed, we have

$$P(\mathcal{A}^c) \le 3N_+K^2n^{-\alpha} + KN_+e^{-\kappa^2 n\pi_{\min}/3}$$
$$P(\mathcal{B}^c) \le 3(C_2N_+ + 2K^2)n^{-\alpha} + 3N_+Ke^{-\kappa^2 n\pi_{\min}/3}.$$

Using union bound and further bounding $C_2N_+ + 2K^2 \le 4C_2N_+K^2$ and $e^{-\kappa^2 n\pi_{\min}/3} \le n^{-\alpha}$ (by assumption (21)), we obtain

$$P(\mathcal{A}^c \cup \mathcal{B}^c) \le 19N_+K^2n^{-\alpha} \tag{41}$$

which goes to zero under the assumption $N_+K^2n^{-\alpha} = o(1)$.

Controlling $\widehat{T}_n$. Define

$$D := \frac{S_1}{m_1} - \frac{S_2'}{m_2'}, \quad \widehat{D} := \frac{\widehat{S}_1}{\widehat{m}_1} - \frac{\widehat{S}_2'}{\widehat{m}_2'},$$
$$\widehat{E} := \frac{\sqrt{\widehat{h}}}{\sqrt{2}\widehat{\sigma}}\widehat{D}, \quad E := \frac{\sqrt{h}}{\sqrt{2}\sigma}D$$

where $S_2' = P^*S_2P^{*T}$ and $m_2' = P^*m_2P^{*T}$. Note also that $S_r/m_r = \widetilde{B}_r$. Let $d = K(K+1)/2$. For the rest of the proof, we treat the above matrices as vectors in $\mathbb{R}^d$ by considering only the elements on and above the diagonal, in a particular order (say rowwise).

We have $\widehat{T}_n = \|\widehat{E}\|_F^2$ on $\mathcal{A} \cap \mathcal{B}$ and since $P(\mathcal{A} \cap \mathcal{B}) = 1 - o(1)$, it is enough to establish $\widehat{E} \Rightarrow (N(0,1))^{\otimes d}$; see [20, Theorem 9.15]. Clearly,

$$\widehat{E} \;=\; \frac{\sqrt{\widehat{h}}}{\sqrt{h}} \cdot \frac{\sigma}{\hat{\sigma}} \cdot E \;+\; \frac{\sqrt{\widehat{h}}}{\sqrt{h}} \cdot \frac{\sigma}{\hat{\sigma}} \cdot \frac{\sqrt{h}}{\sqrt{2}\sigma}(\widehat{D} - D). \tag{42}$$

From (34), with high probability, we have that

$$\|\frac{\sqrt{h}}{\sqrt{2}\sigma}(\widehat{D} - D)\|_{\max} \;\lesssim\; \frac{\sqrt{Nn^2}}{(\nu_n/n)^{1/2}}\frac{\nu_n\widetilde{\varepsilon}}{n} \;=\; \sqrt{Nn\nu_n}\widetilde{\varepsilon}. \tag{43}$$

Next we have

$$\Big\|\frac{\sigma^2}{\widehat{\sigma}^2} - \mathbf{1}_d\Big\|_{\max} \le \frac{\|\sigma^2 - \widehat{\sigma}^2\|_{\max}}{\min_{k,\ell}\widehat{\sigma}_{k,\ell}^2} \lesssim \frac{(\nu_n/n)\gamma_n}{\nu_n/n} \le \gamma_n$$

hence

$$\sigma/\widehat{\sigma} = \mathbf{1}_d + o_p(1). \tag{44}$$

Similarly,

$$\Big\|\frac{\widehat{h}}{h} - \mathbf{1}_d\Big\|_{\max} \le \frac{\|\widehat{h} - h\|_{\max}}{\min_{k,\ell} h_{k\ell}} \lesssim \frac{Nn^2\widetilde{\varepsilon}}{Nn^2} \le \widetilde{\varepsilon}$$

and hence

$$\sqrt{\widehat{h}}/\sqrt{h} = \mathbf{1}_d + o_p(1). \tag{45}$$

Lemma B.2. $E := (\sqrt{h}/(\sqrt{2}\sigma))D \Rightarrow (N(0,1))^{\otimes d}$, under the assumptions of Theorem A.6.

Proof. See Section E.1. □

Therefore, from (42)–(45), Lemma B.2, and Slutsky's theorem, we obtain that $\widehat{E} \Rightarrow (N(0,1))^{\otimes d}$, provided that $\sqrt{N\nu_n n}\,\widetilde{\varepsilon} = o(1)$. The proof is complete.

B.4 Proof of Theorem A.7 (Consistency, Tier 3)

We will follow the notation and argument in the proof of Theorem A.6. On event $\mathcal{A} \cap \mathcal{D} \cap \mathcal{B}$ defined there, we have correct matching $\widehat{P}_{rt} = \widetilde{P}_{rt}$—where $\widetilde{P}_{rt}$ is as defined in the proof of Theorem A.6—and (34) and (35) hold. Since $\mathcal{G} \subset \mathcal{D}$, all the above also holds on $\mathcal{A} \cap \mathcal{G} \cap \mathcal{B}$. This is the event we will work on.

The two inequalities (34) and (35) give

$$\|(\widehat{S}_r/\widehat{m}_r) - B_r\|_{\max} \le C_1 \frac{\nu_n}{n}\Big(40\beta^3\widetilde{\varepsilon} + \delta_n^{(N_r)}\Big) = \xi_r \frac{\nu_n}{n}, \quad r = 1, 2.$$

Since, $\widehat{S}_2'/\widehat{m}_2' = \widehat{P}(\widehat{S}_2/\widehat{m}_2)\widehat{P}^\top$, we also have

$$\|(\widehat{S}_2'/\widehat{m}_2') - \widehat{P}B_2\widehat{P}^\top\|_{\max} \le \xi_2 \frac{\nu_n}{n}.$$

Using $(a+b+c)^2 \le 3(a^2+b^2+c^2)$, we obtain

$$\begin{aligned}
\|\!|(\widehat{S}_1/\widehat{m}_1) - (\widehat{S}_2'/\widehat{m}_2')|\!\|_F^2 &\ge \frac{1}{3}\|\!|B_1 - \widehat{P}B_2\widehat{P}^\top|\!\|_F^2 \\
&\quad - \|\!|(\widehat{S}_1/\widehat{m}_1) - B_1|\!\|_F^2 - \|\!|(\widehat{S}_2'/\widehat{m}_2') - \widehat{P}B_2\widehat{P}^\top|\!\|_F^2 \\
&\ge \frac{1}{3}\min_P \|\!|B_1 - PB_2P^\top|\!\|_F^2 - 2K^2\Big(\frac{\nu_n}{n}\Big)^2 \max_{r=1,2}\xi_r^2 \\
&\ge \frac{1}{6}\min_P \|\!|B_1 - PB_2P^\top|\!\|_F^2
\end{aligned}$$

by assumption (24). Next, we note that (37) holds in this case. Let $m_2' = \widehat{P}m_2\widehat{P}^\top$ and let h be the elementwise harmonic mean of m_1 and m_2'. Then, both (38) and (39) hold, irrespective of the specific $\widehat{P}$. On $\mathcal{G}$, we have $48C_2\beta^4\widetilde{\varepsilon} \le 1$ which, combined with the latter two inequalities, yields

$$\min_{k\ell} \widehat{h}_{k\ell} \ge Nn^2/(4\beta^2).$$

Also $\widehat{\sigma}_{k\ell}^2 = \widehat{B}_{k\ell}(1 - \widehat{B}_{k\ell}) \le 1/4$. Then,

$$\widehat{T}_n \ge \frac{Nn^2/(4\beta^2)}{2 \cdot \frac{1}{4}} \|\!|(\widehat{S}_1/\widehat{m}_1) - (\widehat{S}_2'/\widehat{m}_2')|\!\|_F^2. \tag{46}$$

Putting the pieces together, we have the desired inequality on $(\mathcal{A} \cap \mathcal{B}) \cap \mathcal{G}$. Combined with the probability bound (41) for $(\mathcal{A} \cap \mathcal{B})^c$, the proof is complete.

B.5 Proof of Lemma 2.1

We restate the lemma here with the extra uniqueness clause added.

Lemma B.3. Consider two $K \times K$ matrices B_1 and B_2 with EVDs given by $B_r = Q_r \Lambda Q_r^\top$, $r = 1, 2$, for some diagonal matrix Λ with distinct diagonal entries. Then, a permutation matrix P^* satisfies

$$B_1 \xrightarrow{P^*} B_2 \tag{47}$$

if and only if there exists a diagonal sign matrix S^*, such that

$$Q_2 = P^* Q_1 S^*. \tag{48}$$

Moreover if B_1 is (θ, η) friendly, then there is at most one P^* satisfying (47).

Proof. Since P^*Q_1 is an orthogonal matrix, by absorbing P^* into Q_1 and redefining Q_1, we can assume $P^* = I$, without loss of generality. The problem reduces to showing that (*) $Q_2\Lambda Q_2^T = Q_1\Lambda Q_1^T$ iff there is a sign matrix S^* such that $Q_2 = Q_1S^*$. Let $Q = Q_2^TQ_1$ and note that Q is an orthogonal matrix. Multiplying (*) on the left and right by Q_2^T and Q_2, the problem reduces to showing that (**) $\Lambda = Q\Lambda Q^T$ for an orthogonal matrix Q and diagonal matrix Λ iff Q is a diagonal sign matrix (i.e., $Q = S^*$).

Assume (**) holds, the other direction being trivial. Note that changing Λ to $\Lambda + \alpha I$ does not change (**), hence we can shift the diagonal entries of Λ arbitrarily. Let $\Lambda = \mathrm{diag}(\lambda_k)$. Since (λ_k) are distinct, we can shift them so that $\lambda_1 < 0$ and $\lambda_k > 0$ for $k \geq 2$. From (**), looking at the first entries of the two sides, $\lambda_1 = \sum_{k\geq 1} Q_{ik}^2\lambda_k$, hence

$$0 = -(1 - Q_{11}^2)\lambda_1 + \sum_{k\geq 2} Q_{1k}^2\lambda_k.$$

Every term on the RHS is non-negative. It follows that every term has to be zero, implying $Q_{11}^2 = 1$ and $Q_{1k} = 0$ for $k \geq 2$. This proves the assertion for the first row of Q. Repeating the argument for the other rows, the result follows.

Next, we prove the uniqueness. Suppose that there exist permutation matrices P and $\tilde{P}$ such that $PB_1P^T = B_2 = \tilde{P}B_1\tilde{P}^T$. By the argument above, there exist sign matrices S and $\tilde{S}$ such that $PQ_1S = Q_2, = \tilde{P}Q_1\tilde{S}$. Hence,

$$Q_1 = P^T\tilde{P}Q_1\tilde{S}S.$$

The problem then reduces to showing that if $Q = PQS$ where Q is an orthogonal matrix, P a permutation matrix and S a sign matrix, then $P = I$. If $S = I$ (the trivial sign matrix), then $P = QQ^T = I$. So assume $S \neq I$ and $P \neq I$. Then, there is j such that $PQ_{.,j} = -Q_{.,j}$, hence,

$$\mathbf{1}^TQ_{.,j} = \mathbf{1}^TPQ_{.,j} = -\mathbf{1}^TQ_{.,j}$$

where $\mathbf{1}$ is the all-ones vector. This gives $\mathbf{1}^TQ_{.,j} = 0$, contradicting friendliness of B_1. The proof is complete.

□

C Additional Discussion

C.1 Class proportions and community refinements

Consider the following example to illustrate how class-proportion differences can be absorbed into refined community structures. Suppose we start with two SBMs with equal connectivity matrices but differing class proportions:

$$\pi_1 = (0.5, 0.5), \quad \pi_2 = (0.4, 0.6), \quad B_1 = B_2 = \begin{pmatrix} 0.4 & 0.1 \\ 0.1 & 0.3 \end{pmatrix}.$$

To eliminate the discrepancy in proportions, we refine each community into two sub-communities, effectively doubling the number of communities. Specifically, for π_1, we split its first community into sub-communities of proportions $(0.4, 0.1)$ and its second community into $(0.1, 0.4)$, so that the total proportions of the four new sub-communities become $(0.4, 0.1, 0.1, 0.4)$. For π_2, we retain its original first community (proportion 0.4), and split its second community (proportion 0.6) into three sub-communities of sizes 0.1, 0.1, 0.4. These four new sub-communities also have overall proportions $(0.4, 0.1, 0.1, 0.4)$. Hence the refined SBMs now have the same class proportions:

$$\tilde{\pi}_1 = \tilde{\pi}_2 = (0.4, 0.1, 0.1, 0.4).$$

However, the connectivity matrices enlarge to reflect the finer partition:

$$\tilde{B}_1 = \begin{pmatrix} 0.4 & 0.4 & 0.1 & 0.1 \\ 0.4 & 0.4 & 0.1 & 0.1 \\ 0.1 & 0.1 & 0.3 & 0.3 \\ 0.1 & 0.1 & 0.3 & 0.3 \end{pmatrix}, \quad \tilde{B}_2 = \begin{pmatrix} 0.4 & 0.1 & 0.1 & 0.1 \\ 0.1 & 0.3 & 0.3 & 0.3 \\ 0.1 & 0.3 & 0.3 & 0.3 \\ 0.1 & 0.3 & 0.3 & 0.3 \end{pmatrix}.$$

Note that class-proportion differences vanish (since $\tilde{\pi}_1 = \tilde{\pi}_2$), but the matrices $\tilde{B}_1$ and $\tilde{B}_2$ now differ explicitly. This clarifies that proportion differences can equivalently be represented as connectivity differences through community refinement, motivating our focus on testing differences in B_r modulo permutations while treating class proportions as nuisance parameters.

C.2 Empirical Analysis of Average Degree Growth

To support the claim in Remark 4.5 that the $\nu_n = \Omega(\log n)$ condition is mild in practice, we analyzed the average degree growth in several real-world network datasets from the PyTorch Geometric library. We sampled subgraphs of increasing size from each dataset and computed the ratio of the average degree to $\log n$. As shown in Table 2, this ratio tends to grow with n, indicating that the average degree in these networks grows faster than $\log n$. Qualitatively similar behavior is observed across other datasets (`amazon-photo`, `wikics`, `corafull`, `coauthor-cs`, and `planetoid-cora`).

Table 2: Average degree vs. size for subsets of the Amazon Computers dataset. The final column shows that the average degree grows faster than $\log n$.

Dataset	n	n/n_{total}	**avg_degree**	**avg_degree /** $\log n$
amazon-computers	138	0.01	0.435	0.088
amazon-computers	276	0.02	0.623	0.111
amazon-computers	688	0.05	1.674	0.256
amazon-computers	1376	0.1	3.078	0.426
amazon-computers	2751	0.2	6.057	0.765
amazon-computers	6876	0.5	17.121	1.938
amazon-computers	13752	1.0	35.756	3.752

Implementation. The experiment was implemented in a lightweight PyTorch Geometric script (`avg_degree_tables_min.py`, provided in the supplementary material for reproducibility.), which randomly subsamples nodes, constructs the induced subgraph, and computes the mean degree via `torch_geometric.utils.degree`. We fix a random seed for reproducibility and report average values over multiple fractions $n/n_{\text{total}} \in \{0.01, 0.02, 0.05, 0.1, 0.2, 0.5, 1.0\}$.

C.3 Derivation of the Permissible Range of K for Graphons

This section provides the semi-rigorous derivation for the permissible range of K when applying our test to general graphon models, as summarized in Section 5.1.

We aim to find conditions on K that ensure our test has power when the underlying data comes from two different β-smooth graphons, f_1 and f_2. The argument combines oracle inequalities from [21] with the structure of our test statistic. We apply their results with $\rho_n = \nu_n/n$ and $n_0 = n/K$.

Let $\Theta = (\Theta_{ij}) \in [0,1]^{n\times n}$ be the matrix of connection probabilities, or "discrete graphon," where $\Theta_{ij} = \mathbb{E}[A_{ij}]$. Let $\widehat{\Theta}$ be its K-block estimate based on the observed adjacency matrix A, and let $\check{\Theta}$ be the best K-block approximation of Θ. By Proposition 2.1 of [21], the expected squared Frobenius error of the estimate is bounded:

$$\mathbb{E}\|\widehat{\Theta} - \Theta\|_F^2 \lesssim \|\Theta - \check{\Theta}\|_F^2 + (n\log K + K^2)\Big(\frac{\nu_n}{n} + \frac{K\log K}{n}\Big).$$

Furthermore, for a β-smooth graphon, Proposition 2.5 of [21] bounds the approximation error:

$$\|\Theta - \check{\Theta}\|_F^2 \lesssim n^2\big(\frac{\nu_n}{n}\big)^2 K^{-2\beta} = \nu_n^2 K^{-2\beta}.$$

Assuming $\nu_n \gtrsim K\log K$, the estimation error bound simplifies to:

$$\begin{aligned}\mathbb{E}\|\widehat{\Theta} - \Theta\|_F^2 &\lesssim \nu_n^2 K^{-2\beta} + (n\log K + K^2)\frac{\nu_n}{n} \\ &= \nu_n^2 K^{-2\beta} + \nu_n \log K + \frac{\nu_n K^2}{n} =: \Delta_{n,K}.\end{aligned}$$

Now, consider the alternative hypothesis where we have two discrete graphons, Θ_1 and Θ_2. Ignoring the permutation matching step for simplicity (i.e., assuming identity permutation), our test statistic $\widehat{T}_n$ is proportional to the squared difference of the estimated block models. This is, in turn, related to the squared difference of the estimated discrete graphons:

$$\frac{\widehat{T}_n}{N} \gtrsim \Big(\frac{n}{K}\Big)^2 \|\widehat{B}_1 - \widehat{B}_2\|_F^2 = \|\widehat{\Theta}_1 - \widehat{\Theta}_2\|_F^2.$$

Using a triangle inequality, we can bound this difference:

$$\|\widehat{\Theta}_1 - \widehat{\Theta}_2\|_F^2 \geq \frac{1}{3}\|\Theta_1 - \Theta_2\|_F^2 - \|\widehat{\Theta}_1 - \Theta_1\|_F^2 - \|\widehat{\Theta}_2 - \Theta_2\|_F^2.$$

The true difference is $\|\Theta_1 - \Theta_2\|_F^2 \asymp n^2(\nu_n/n)^2\|f_1 - f_2\|_{L^2}^2 = \nu_n^2\|f_1 - f_2\|_{L^2}^2$. The two estimation error terms are of order $\Delta_{n,K}$. For the test to have power, the signal term must dominate the error terms. This leads to the condition:

$$\nu_n^2\|f_1 - f_2\|_{L^2}^2 \gg \Delta_{n,K}.$$

Dividing by ν_n^2, we require $\frac{\Delta_{n,K}}{\nu_n^2} = o(1)$. This yields the final condition:

$$K^{-2\beta} + \frac{\log K}{\nu_n} + \frac{K^2}{n\nu_n} = o(1),$$

A sufficient set of conditions for this to hold is $K \to \infty$ (to handle the first term), while $\log K = o(\nu_n)$ (for the second term) and $K^2 = o(n\nu_n)$ (for the third term). These also ensure the earlier assumption $\nu_n \gtrsim K \log K$. This provides the justification for the claims made in the main text.

D Additional experiments

Here, we provide more details of the experimental setup and report on more experiments. All experiments were conducted on an internal computing cluster with Intel(R) Xeon(R) Platinum 8160 CPUs (48 cores, 2.10GHz).

D.1 Competing methods

In this section, we describe the competing methods used in the experiments in the main text and below. Some of them were not originally designed for two-sampling testing, but have a natural extension to this setting which we outline below. Throughout this section, we use the terms "network" and "adjacency matrix" interchangeably, since the resulting distance measures between adjacency matrices will be invariant to node permutations.

To measure a distance between two adjacency matrices, NCLM constructs a feature vector for each graph based on its so-called log-moments. To be more precise, [28] considers the k-th *graph moment* of a matrix A, $m_k(A) = \mathrm{tr}[(A/n)^k]$, which is the normalized count of closed walks of length k. The feature vector for an adjacency matrix A is then defined as

$$g_J(A) := \big(\log m_j(A),\, j \in [J]\big)$$

where J is some positive integer. The test statistic for comparing two adjacency matrices A_1 and A_2 is naturally given as the ℓ_2-distance between the feature vectors of the two graphs:

$$d_{\text{NCLM}}(A_1, A_2) := \|g_J(A_1) - g_J(A_2)\|_2.$$

Our experiments, reported in Appendix D.6, suggest that a larger value of J improves performance. However, increasing J quickly increases the computation cost. In the results to follow, we have chosen $J = 20$ which provides a reasonable balance between performance and cost.

To form a distance between two adjacency matrices, ASE-MMD first computes the so-called adjacency spectral embedding (ASE) for each matrix. For an adjacency matrix A, consider $|A| = (A^\top A)^{1/2}$ and let $|A| = \sum_{i=1}^n \lambda_i u_i u_i^\top$ be its eigen-decomposition where $\lambda_1 \geq \cdots \geq \lambda_n \geq 0$ are the eigenvalues and $\{u_i\}$ the corresponding orthonormal basis of eigenvectors. Then, the adjacency spectral embedding of $A \in \{0,1\}^{n\times n}$ into $\mathbb{R}^d$ is

$$\hat{X}(A) = U_A\sqrt{S_A} \in \mathbb{R}^{n\times d} \tag{49}$$

where $S_A = \mathrm{diag}(\lambda_1, \lambda_2, \ldots, \lambda_d)$ and U_A is $n \times d$ matrix whose columns are $u_1, u_2, \ldots, u_d$. The rows of $\hat{X}(A)$ define an empirical distribution $P_{\hat{X}(A)} := \frac{1}{n}\sum_{i=1}^n \delta_{\hat{X}_{i*}(A)}$ where δ_x is the point mass as x. ASE-MDD then measures the distance between two adjacency matrices A_1 and A_2 as the maximum mean discrepancy of the corresponding empirical distributions:

$$d_{\text{ASE-MMD}}(A_1, A_2) = \text{MMD}(P_{\hat{X}(A_1)}, P_{\hat{X}(A_2)}).$$

The MMD relies on a positive definite kernel function $\kappa : \mathbb{R}^d \times \mathbb{R}^d \to \mathbb{R}$. Letting $\hat{X} = \hat{X}(A_1)$ and $\hat{Y} = \hat{X}(A_2)$, one has

$$\text{MMD}(P_{\hat{X}}, P_{\hat{Y}}) = \frac{1}{n(n-1)} \sum_{i \neq j} \kappa(\hat{X}_{i*}, \hat{X}_{j*}) - \frac{2}{mn} \sum_{i \neq j} \kappa(\hat{X}_{i*}, \hat{Y}_{j*}) + \frac{1}{n(n-1)} \sum_{i \neq j} \kappa(\hat{Y}_{i*}, \hat{Y}_{j*}).$$

In experiments reported here, we consider a Gaussian kernel and use random Fourier features to approximate the MMD. The bandwidth is set to be $\sigma^2 = 1$. Additional experiments, reported in Appendix D.5, show that the results are not sensitive to the choice of bandwidth.

Adapting to multiple samples. Next, we describe how we adapt the test statistics to the cases where the sample size is greater than 1. Any measure of dissimilarity, $d(\cdot, \cdot)$, between two networks, can be generalized to the two-sample case, by averaging the dissimilarity of pairs of networks from different samples. More specifically, given the two samples $A_{rt},\ t \in [N_r]$ for $r = 1, 2$, we have the two-sample test statistic $\frac{1}{N_1 N_2} \sum_{t=1}^{N_1} \sum_{s=1}^{N_2} d(A_{1t}, A_{2s})$.

A note on network generation. Throughout the experiments, we replace Bernoulli variables with a clipped version. In particular, write $X \sim \overline{\mathsf{Ber}}(p)$ for $p \in \mathbb{R}$ to denote $X = 1\{U < p\}$ for some $U \sim [0, 1]$. When $p \in [0, 1]$, $\overline{\mathsf{Ber}}(p) = \mathsf{Ber}(p)$, but otherwise $\overline{\mathsf{Ber}}(p)$ is naturally clipped to either 0 or 1. When we write $\mathsf{SBM}(B, \pi)$ in experiments, it is based on $\overline{\mathsf{Ber}}(p)$ generation.

D.2 Simple 2-SBM

Our first simulation reproduces the experiment from [37] on a 2-SBM. In particular, we consider testing $H_0 : A, A' \sim \mathsf{SBM}(B_0, \pi)$ versus $H_1 : A \sim \mathsf{SBM}(B_0, \pi),\ A' \sim \mathsf{SBM}(B_\varepsilon, \pi)$, where

$$B_\varepsilon = \begin{bmatrix} 0.5 + \varepsilon & 0.2 \\ 0.2 & 0.5 + \varepsilon \end{bmatrix}, \tag{50}$$

and $\pi = (0.4, 0.6)$. Note that each sample contains a single adjacency matrix. We consider vertex sizes $n \in \{100, 200, 500, 1000\}$ and noise levels $\varepsilon \in \{0.01, 0.02, 0.05, 0.1\}$ and evaluate the power of the our SBMTS using 1000 Monte-Carlo replications. Table 3 summarizes the results alongside the power estimates for ASE-MMD reported in [37]. The results clearly show the superior performance of SBMTS along both dimensions (n, ε).

Table 3: Power estimates for ASE and SBM-TS for a simple 2-block SBM experiment.

	$\varepsilon = 0.01$		$\varepsilon = 0.02$		$\varepsilon = 0.05$		$\varepsilon = 0.1$	
n	SBM-TS	ASE	SBM-TS	ASE	SBM-TS	ASE	SBM-TS	ASE
100	0.047	-	0.161	0.06	0.776	0.09	1	0.27
200	0.15	-	0.599	0.09	1	0.17	1	0.83
500	0.785	-	1	0.01	1	0.43	1	1
1000	1	0.14	1	1	1	1	1	1

D.3 SW–GOT dataset

The Star Wars (SW)–Game of Thrones (GOT) dataset [34] is derived from popular films and television series. We consider 13 networks: six from the original and sequel SW trilogies, and seven from each of the GOT series. In each graph, vertices correspond to characters and edges indicate whether two characters share a scene. Let us denote the SW and GOT networks as class $\mathcal{C}_1$ and $\mathcal{C}_2$, respectively. The set of characters for both classes overlap across multiple networks, but no vertex correspondence is utilized because network vertices are unlabeled and each network has different number of vertices. Sample graphs from this dataset are shown in Figure 10 in Appendix D.7.

Similar to the COLLAB dataset, we consider a two sample testing problem for distinguishing a null of $(\mathcal{C}_1, \mathcal{C}_1)$ vs. an alternative of $(\mathcal{C}_1, \mathcal{C}_2)$. The resulting ROCs are shown in Figure 5 showing a significant advantage for SBM-TS compared to the competitors.

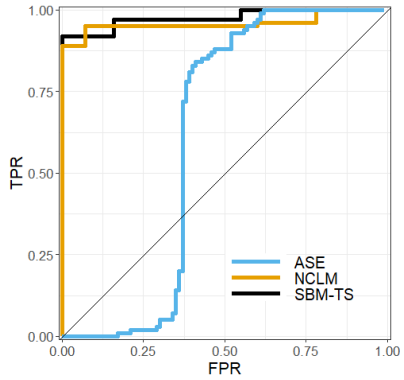
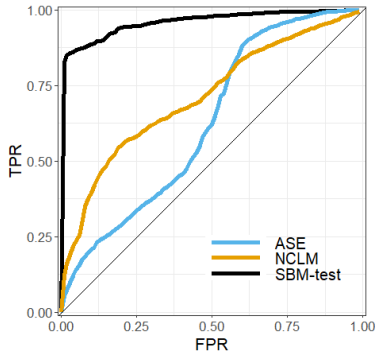


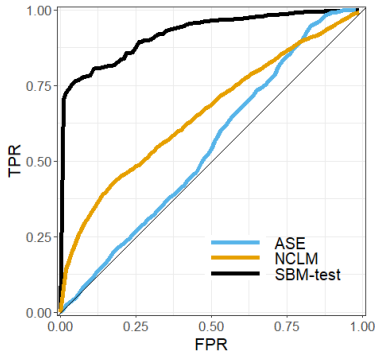
Figure 5: ROC curves for the SW-GOT dataset.

D.4 ROC plots for general SBM with random B

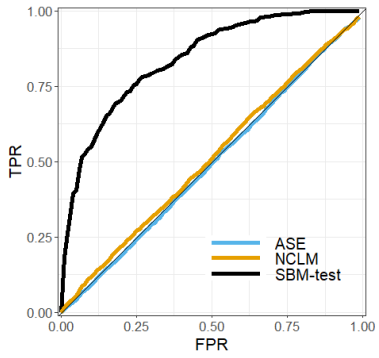
Figure 6 shows the mean ROC curves for the experiment in Section 5.3. The different plots correspond to different values of K. We refer to Section 5.3 for the detailed description of the experiments, where a summary of these curves via their "area under the curve (AUC)" was provided in Table 1.



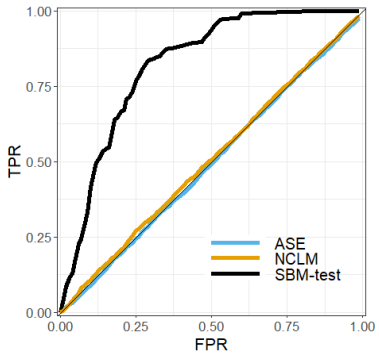
(a) $K = 2$, mean AUC = 0.95.



(b) $K = 3$, mean AUC = 0.91.



(c) $K = 15$, mean AUC $= 0.83$.



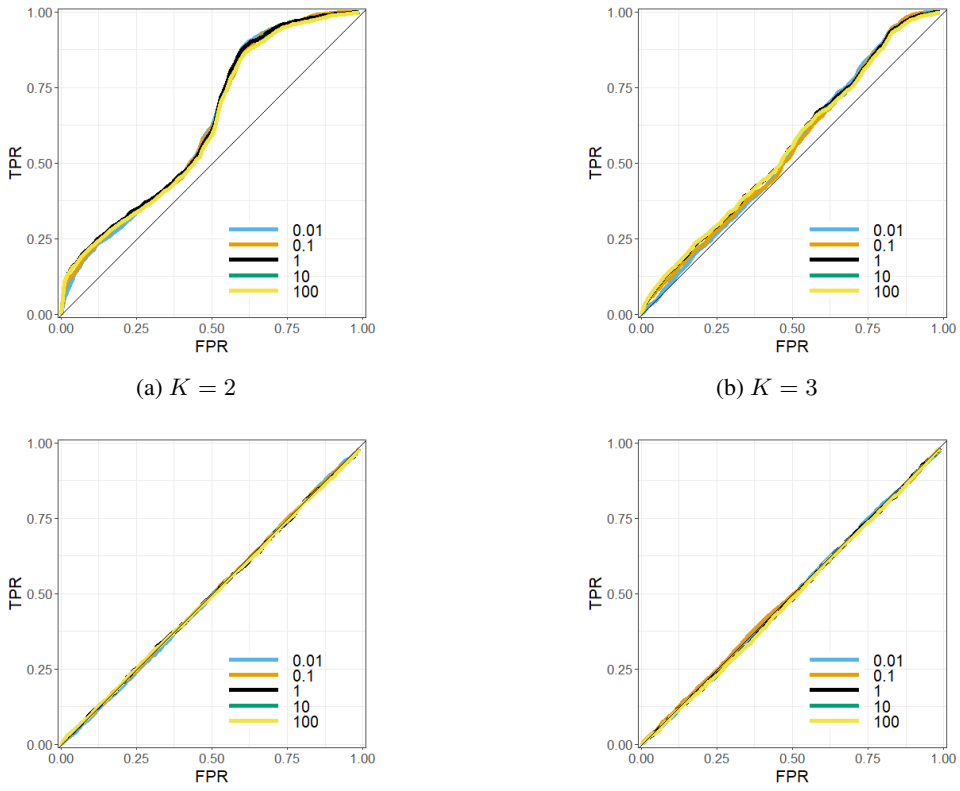
(d) $K = 20$, mean AUC $= 0.81$.

Figure 6: ROC curves for the three methods (SBM-TS, MMD of ASE, and test statistic based on NCLM) averaged over 50 different experiments.

D.5 Bandwidth of ASE-MMD

We have conducted additional experiments in the same setting of Section 5.3, that is, general SBM with random B to determine the effect of bandwidth on the performance of ASE-MMD. The results

are summarized in Figure 7. One observes that the bandwidth does not have a significant bearing on the power of the ASE-MMD test, with ROCs remaining almost the same across the range of $\sigma^2 \in \{0.01, 0.1, 1, 10, 100\}$.



(a) $K = 2$ (b) $K = 3$

(c) $K = 15$ (d) $K = 20$

Figure 7: ROC curves for different choices of the bandwidth σ^2 in the experiment with a general B.

D.6 Number of Log Moments for NCLM

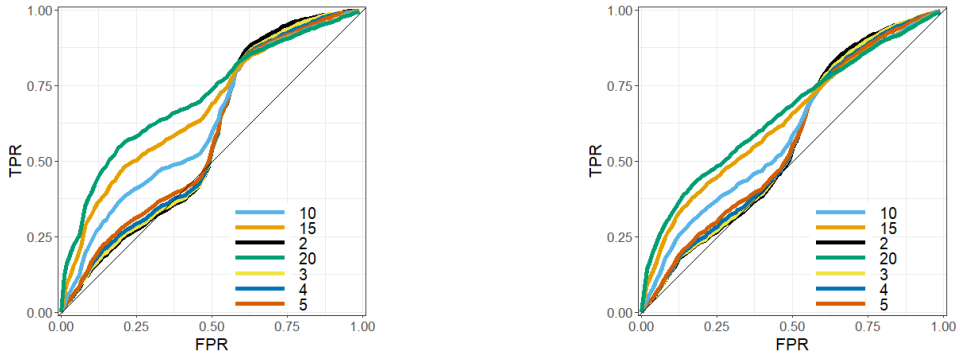
We have performed additional experiments in the same setting of Section 5.3, that is, general SBM with random B to determine the effect of the number J of log-moments on the performance of NCLM. The results are summarized in Figure 8. The general trend is that higher J improves performance with $J = 20$ (the maximum we considered) producing the best results. We also note that this is mainly for small values of K, while for larger $K \in \{15, 20\}$ the test is powerless regradless of the value of J.

D.7 Sample graphs for real-world data

Example of graphs from the COLLAB dataset are shown in Figure 9.

D.8 Empirical runtime of Algorithm 1 With Respect To K

We investigate the mean runtime of Algorithm 1 as a function of the block dimension K. For varying values of K, we generate a random $K \times K$ symmetric matrix B_1 and its random permutation $B_2 = PBP^\top$ and apply Algorithm 1. The results are summarized in Table 4.



(a) $K = 2$ (b) $K = 3$

(c) $K = 15$ (d) $K = 20$

Figure 8: ROC curves for different choices of the number of log-moments in the experiment with a general B.

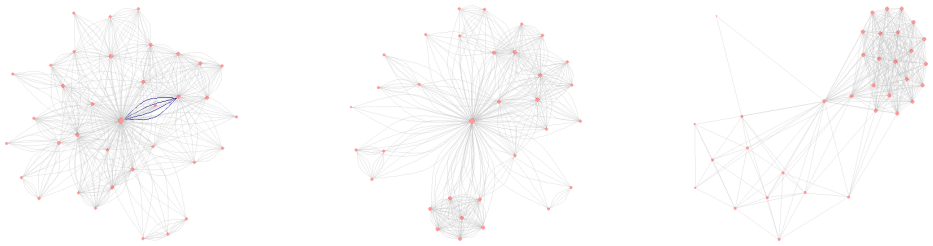


Figure 9: Sample graphs from the COLLAB dataset: High Energy Physics (left), Condensed Matter Physics (middle) and Astrophysics (right).

E Proofs auxiliary results

E.1 Proof of Lemma B.2

Recall that the L^1 Wasserstein distance between the distributions of two random variables Y and Z can be expressed as

$$d_{W_1}(Y, Z) = \sup_{h:\, \|h\|_{\text{Lip}} \leq 1} |E[h(Y)] - E[h(Z)]|$$

Table 4: Mean runtimes (ms) of `matching(B1, B2)` for selected values of K.

	$K = 2$	$K = 10$	$K = 20$	$K = 50$	$K = 100$
Mean runtime (ms)	0.6836	1.0366	1.3055	3.0415	8.3679

Figure 10: Sample graphs from the movie/television dataset: Class 1 (left) and Class 2 (right)

where h ranges over all 1-Lipschitz functions $h : \mathbb{R} \to \mathbb{R}$, that is, h that satisfy $|h(x)-h(y)| \le |x-y|$ for all $x, y \in \mathbb{R}$. See [10, Chapter 4] or [33].

Lemma B.2 follows form the following results:

Proposition E.1. Let $S \sim \mathsf{Bin}(n, p)$ with $p \le 1/2$ and let $W = \frac{\sqrt{n}}{\sigma}(\frac{S}{n} - p)$ where $\sigma = \sqrt{pq}$ and $q = 1 - p$. Let $C \le 0.4785$ be the constant in the Berry-Esseen bound. Then,

$$\sup_{x\in\mathbb{R}} |P(W \le x) - \Phi(x)| \le \frac{C}{\sqrt{np/2}} \tag{51}$$

$$\sup_{h:\|h\|_{\mathrm{Lip}}\le 1} \big|E[h(W)] - E[h(Z)]\big| \le \frac{1}{\sqrt{np/2}} \tag{52}$$

where $Z \sim N(0, 1)$.

Proof. We can write $S = \sum_{i=1}^n B_i$ where B_i are i.i.d. Bernoulli(p) variables. Let $X_i = (B_i - p)/\sigma$ be the standardized versions and note that $W = \frac{1}{\sqrt{n}} \sum_{i=1}^n X_i$. Berry-Esseen bound gives

$$\sup_{x\in\mathbb{R}} |P(W \le x) - \Phi(x)| \ \le \ C\frac{E[|X_1|^3]}{\sqrt{n}}$$

and Corollary 4.1 of [10, page 67] gives $d_{W_1}(W, Z) \le \frac{E[|X_1|^3]}{\sqrt{n}}$. We have

$$E[|X_1|^3] = \frac{1}{\sigma^3} \cdot E[|B_1 - p|^3] = \frac{p^3q + q^3p}{(pq)^{3/2}} = \frac{p^2 + q^2}{\sqrt{pq}} \le \frac{(p + q)^2}{\sqrt{p/2}}$$

which gives the desired inequalities.

□

Lemma E.1. Suppose that for $r \in \{1, 2\}$ and $t \in [N_r]$ the symmetric matrices $A_{rt} \in \{0, 1\}^{n_{rt}\times n_{rt}}$ are independent with independent entries on and above diagonal that satisfy $(A_{rt})_{ij} \sim \mathsf{Ber}(B_{(z_{rt})_i,(z_{rt})_j})$ where $z_{rt} \in [K]^{n_{rt}}$ is deterministic. Let

$$S_r = \sum_{t=1}^{N_r} \mathcal{S}(A_{rt}, z_{rt}), \quad m_r = \sum_{t=1}^{N_r} \mathcal{C}(A_{rt}, z_{rt}),$$

and for all $k, \ell \in [K]$, let $\sigma_{k\ell} = \sqrt{B_{k\ell}(1 - B_{k\ell})}$ and set

$$\xi_{k\ell} \ = \ \frac{\sqrt{\bar{m}_{k\ell}}}{\sqrt{2}\sigma_{k\ell}}\Big(\frac{(S_1)_{k\ell}}{m_{1k\ell}} - \frac{(S_2)_{k\ell}}{m_{2k\ell}}\Big) \tag{53}$$

where $\bar{m}_{k\ell}$ is the harmonic mean of $(m_1)_{k\ell}$ and $(m_2)_{k\ell}$. Assume that $m_{2k\ell} \le c_1 m_{1k\ell}$ for some $c_1 > 0$. Then

$$\sup_{x\in\mathbb{R}} \big|P(\xi_{k\ell} \le x) - P(Z \le x)\big| \le C\sqrt{\frac{n}{\nu_n}}\Big(\frac{1}{\sqrt{m_{1k\ell}}} + \frac{1}{\sqrt{m_{2k\ell}}}\Big).$$

where $C > 0$ is a constant dependent on c_1 and $Z \sim N(0, 1)$.

Proof. Let $\sigma_{kl} = \sqrt{B_{kl}(1-B_{kl})}$. First, notice that by Proposition E.1, it holds that

$$\sup_{x\in\mathbb{R}} \left| P\left(\frac{\sqrt{m_{rkl}}}{\sigma_{kl}} \left(\frac{S_{rkl}}{m_{rkl}} - B_{kl} \right) \le x \right) - \Phi(x) \right| \;\le\; \frac{C}{\sqrt{m_{rkl}\nu_n/n}}$$

for some constant $C > 0$, and for $r = 1, 2$. Hence,

$$\sup_{x\in\mathbb{R}} \left| P\left(\frac{\sqrt{\bar{m}_{k\ell}}}{\sqrt{2}\,\sigma_{kl}} \cdot \frac{S_{rkl}}{m_{rkl}} \le x \right) - \Phi\Big(\sqrt{\frac{2m_{rkl}}{\bar{m}_{k\ell}}} \big(x - \frac{\sqrt{\bar{m}_{k\ell}} B_{kl}}{\sqrt{2}\sigma_{kl}} \big) \Big) \right| \;\le\; \frac{C}{\sqrt{m_{rkl}\nu_n/n}}. \tag{54}$$

Define the random variables

$$X_r \;=\; \frac{\sqrt{\bar{m}_{k\ell}}}{\sqrt{2}\,\sigma_{kl}} \cdot \frac{S_{rkl}}{m_{rkl}}, \quad r = 1, 2,$$

and consider two independent random variable Y_1 and Y_2 with

$$Y_r \;\sim N\left(\frac{\sqrt{\bar{m}_{k\ell}} B_{k\ell}}{\sqrt{2}\sigma_{k\ell}},\, \frac{\bar{m}_{k\ell}}{2m_{rk\ell}} \right).$$

Notice that $Y_1 - Y_2 \;\sim\; N(0, 1)$. Let F_Z denote the CDF of a random variable Z. Then, we can rewrite (54) as

$$\sup_{x\in\mathbb{R}} |F_{X_1}(x) - F_{Y_1}(x)| \;\le\; \frac{C}{\sqrt{m_{rkl}\nu_n/n}}. \tag{55}$$

For any $x, x_2 \in \mathbb{R}$, we have

$$P(X_1 - X_2 \le x \,|\, X_2 = x_2) = P(X_1 \le x + x_2) = F_{X_1}(x + x_2),$$

by independence of X_1 and X_2. Fix $x \in \mathbb{R}$. Since $P(X_1 - X_2 \le x) = E[P(X_1 - X_2 \le x \,|\, X_2)]$, it follows that

$$\begin{aligned}
&|P(X_1 - X_2 \le x) - P(Y_1 - Y_2 \le x)| \\
&\quad = \big| E[F_{X_1}(x + X_2)] - E[F_{Y_1}(x + Y_2)] \big| \\
&\quad = \big| E[F_{X_1}(x + X_2)] - E[F_{Y_1}(x + X_2) + E[F_{Y_1}(x + X_2) - E[F_{Y_1}(x + Y_2)] \big| \\
&\quad \le E[\big| F_{X_1}(x + X_2) - F_{Y_1}(x + X_2) \big|] + \big| E[F_{Y_1}(x + X_2) - E[F_{Y_1}(x + Y_2)] \big| \\
&\quad \le \frac{C}{\sqrt{m_{1kl}\nu_n/n}} + |E[h(X_2)] - E[h(Y_2)]|
\end{aligned}$$

where we have defined $h(z) = F_{Y_1}(x + z)$. Since

$$\begin{aligned}
\|h\|_{\mathrm{Lip}} \le \|h'\|_\infty &= \sqrt{\frac{2m_{2kl}}{\bar{m}_{k\ell}}} \cdot \|\Phi'\|_\infty \\
&= \sqrt{1 + \frac{m_{2kl}}{m_{1kl}}} \cdot \frac{1}{\sqrt{2\pi}} \le \sqrt{1 + c_1} \cdot \frac{1}{\sqrt{2\pi}} =: c_2
\end{aligned}$$

by assumption. Then, from Proposition E.1, $|E[h(X_2)] - E[h(Y_2)]| \le c_2/\sqrt{m_{2kl}\nu_n/n}$

$$|P(X_1 - X_2 \le x) - P(Y_1 - Y_2 \le x)| \le \frac{C}{\sqrt{m_{1kl}\nu_n/n}} + \frac{c_2}{\sqrt{m_{2kl}\nu_n/n}}$$

which gives the desired result. □

Let $\mathcal{I}(\mathbb{R})$ be the set of indicator functions of half-intervals, that is,

$$\mathcal{I}(\mathbb{R}) = \big\{ t \mapsto 1\{t \le x\} : x \in \mathbb{R} \big\}.$$

Lemma E.2. Consider random variables $X_{ni}, i \in [K]$ and Y_n and assume that $\{X_{ni}, i \in [K]\}$ are independent conditional on Y_n. In addition, we have

$$\sup_{h\,\in\,\mathcal{I}(\mathbb{R})} |E[h(X_{ni}) \,|\, Y_n] - E[h(z)]| \cdot 1\{Y_n \in \mathcal{A}_n\} \;\le\; \varepsilon_n, \quad i \in [K]$$

for some sequence of events $\mathcal{A}_n$ and a deterministic sequence of $\varepsilon_n > 0$, and some random variable $Z \sim \mu$. Assume that $\varepsilon_n \to 0$ and $P(Y_n \in \mathcal{A}_n^c) \to 0$ as $n \to \infty$. Then

$$(X_{n1}, \ldots, X_{nK}) \Rightarrow \mu^{\otimes K}.$$

Proof. Let $Z_i, i \in [K]$ be i.i.d. draws from μ. It is enough to show that

$$E\Big[\prod_{i=1}^{K} f_i(X_{ni})\Big] \to E\Big[\prod_{i=1}^{K} f_i(Z_i)\Big] = \prod_i E[f_i(Z_i)] \tag{56}$$

for any collection of $f_1, \dots, f_K \in \mathcal{I}(\mathbb{R})$.

Let us fix one such collection and, for simplicity, write $W_n = \prod_i f_i(X_{ni})$ and $C = \prod_i E[f_i(Z_i)]$. We want to show $E[W_n] \to C$. From the assumption, it follows that

$$\big|E[f_i(X_{ni}) \,|\, Y_n] - E[f_i(Z_i)]\big| \;\le\; \varepsilon_n + 2 \cdot 1\{Y_n \in \mathcal{A}_n^c\}.$$

By conditional independence, $E[W_n \,|\, Y_n] = \prod_i E[f_i(X_{ni}) \,|\, Y_n]$. Then, using $|\prod_{i=1}^K a_i - \prod_{i=1}^K b_i| \le K \max_i |a_i - b_i|$, which holds if $a_i, b_i \in [-1, 1]$ for all i, we have

$$\big|E[W_n \,|\, Y_n] - C\big| \;\le\; K\varepsilon_n + 2K \cdot 1\{Y_n \in \mathcal{A}_n^c\}.$$

Then, we have

$$\begin{aligned} |E[W_n] - C| &= |E[E[W_n \,|\, Y_n]] - C| \\ &\le E\big\{\big|E[W_n \,|\, Y_n]\big\} - C\big| \le K\varepsilon_n + 2KP(Y_n \in \mathcal{A}^c). \end{aligned}$$

Letting $n \to \infty$, the desired result follows from the assumptions. □

Proof of Lemma B.2. Let $z = (z_{rt}, t \in [N_r], r = 1, 2)$. Let $\mathcal{M}(c_1) = \{z : \; m_{2k\ell} \le c_1 m_{1k\ell}\}$. By Lemma E.1, we have

$$\sup_{h \in \mathcal{I}(\mathbb{R})} \big|E[h(\xi_{k\ell}) \,|\, z] - E[h(z)]\big| \cdot 1\{z \in \mathcal{M}(c_1)\} \le C(c_1)\sqrt{\frac{n}{\nu_n}}\Big(\frac{1}{\sqrt{m_{1k\ell}}} + \frac{1}{\sqrt{m_{2k\ell}}}\Big) \tag{57}$$

where $\xi_{k\ell}$ as defined in (53), $Z \sim N(0,1)$ and $C(c_1)$ is some constant dependent on c_1. Let $n_{rtk} := \sum_{i=1}^{n_t} 1\{(z_{rt})_i = k\}$, and consider the event

$$\mathcal{A}_n \;:=\; \Big\{\min_{r,k} \, n_{rtk} \;\ge n_t/\beta, \; \forall t \in [N]\Big\}.$$

Then on $\mathcal{A}_n$, we have $m_{rk\ell} \ge cNn^2/\beta^2$; see also (39). We also have $P(\mathcal{A}_n^c) = o(1)$ under the assumptions of Theorem A.6, by the same argument that controls $\mathcal{E}_1$ in Proposition F.4. Moreover, assuming $N_2 \le N_1$—without loss of generality—on $\mathcal{A}_n$, we have

$$\frac{m_{2k\ell}}{m_{1k\ell}} \le \frac{N_2(C_2 n)^2}{N_1 n^2/\beta^2} \le \beta^2 C_2^2$$

where we have used the size assumption (19). That is, $\mathcal{A}_n \subset \{z \in \mathcal{M}(\beta^2 C_2)\}$. Taking $c_1 = \beta^2 C_2^2$ in (57) and multiplying both sides of the inequality by $1_{\mathcal{A}_n}$, we obtain

$$\sup_{h \in \mathcal{I}(\mathbb{R})} \big|E[h(\xi_{k\ell}) \,|\, z] - E[h(z)]\big| \cdot 1_{\mathcal{A}_n} \le \frac{2\beta\, C(\beta^2 C_2)}{\sqrt{c}} \frac{1}{\sqrt{Nn\nu_n}}.$$

Since $\{\xi_{k\ell}\}$ are independent given z, the result now follows from Lemma E.2 given $Nn\nu_n \to \infty$. □

E.2 Proof of Lemma B.1

For $P \in \Pi_K$, let

$$\begin{aligned} J_1(P) \;&=\; \{i \,:\, P_{ii} \neq 0\}, \\ J_2(P) \;&=\; \{(i,j) \,:\, P_{ij} \neq 0, \; i \neq j\}, \end{aligned}$$

and note that $|J_1(P)| + |J_2(P)| = K$ for any $P \in \Pi_K$. By assumption $\|\Delta\|_\infty < 1/2$, and hence

$$\begin{aligned} 1 + \Delta_{ii} > 1/2 \quad &\text{for} \quad i \in J_1(P) \\ \Delta_{ij} < 1/2 \quad &\text{for} \quad (i,j) \in J_2(P). \end{aligned}$$

We also have

$$\begin{aligned}\operatorname{tr}\big(P^T(I+\Delta)\big) &= \sum_{i,j}(1_{\{i=j\}}+\Delta_{ij})P_{ij}\\ &= \sum_{i\in J_1(P)}(1+\Delta_{ii}) + \sum_{(i,j)\in J_2(P)}\Delta_{ij}.\end{aligned}$$

Let $P \neq I$, so that both $J_2(P)$ and $[K] \setminus J_1(P)$ are nonempty. Then,

$$\begin{aligned}\operatorname{tr}(I+\Delta) - \operatorname{tr}\big(P^T(I+\Delta)\big) &= \sum_{i\in[K]\setminus J_1(P)}(1+\Delta_{ii}) - \sum_{(i,j)\in J_2(P)}\Delta_{ij}\\ &> \frac{1}{2}(K-|J_1(P)|) - \frac{1}{2}|J_2(P)| = 0,\end{aligned}$$

showing that identity is the unique optimal solution.

E.3 Randomization

Let $\mathcal{U}(\Pi_K)$ denote the uniform distribution on the set of permutation matrices Π_K.

Lemma E.3 (Randomization)**.** Assume that there is a permutation matrix $C_{rt} = C_{rt}(A_{rt})$ potentially dependent on the adjacency matrix A_{rt} such that

$$\widetilde{B}_{rt} \overset{C_{rt}}{\rightsquigarrow} B_{rt}$$

where $\widetilde{B}_{rt} = \mathcal{S}(A_{rt}, \widehat{z}_{rt}^{(0)}) \oslash \mathcal{C}(\widehat{z}_{rt}^{(0)})$. Let $\widehat{B}_{rt}$ be constructed as in step step 4. Then,

$$\widehat{B}_{rt} \overset{P_{rt}}{\rightsquigarrow} B_{rt} \tag{58}$$

where $P_{rt} \sim \mathcal{U}(\Pi_K)$ independent of A_{rt} (and hence $\widehat{B}_{rt}$).

Proof. Notice that by construction $\widehat{B}_{rt} = U_{rt}\widetilde{B}_{rt}U_{rt}^\top$ with $U_{rt} \sim \mathcal{U}(\Pi_K)$. Let $P_{rt} = C_{rt}U_{rt}^\top \in \Pi_K$. Then, $C_{rt}\,\widetilde{B}_{rt}\,C_{rt}^\top = P_{rt}\,\widehat{B}_{rt}\,P_{rt}^\top$ and (58) follows. It remains to show independence of P_{rt} and its uniform distribution. Indeed, let $P_0 \in \Pi_K$ and $A_0 \in \{0,1\}^{n\times n}$. We have (showing the dependence of C_{rt} on A_{rt} explicitly)

$$\begin{aligned}P(P_{rt}=P_0, A_{rt}=A_0) &= P(U_{rt}=P_0^\top C_{rt}(A_{rt}), A_{rt}=A_0)\\ &= P(U_{rt}=P_0^\top C_{rt}(A_0), A_{rt}=A_0)\\ &= P(U_{rt}=P_0^\top C_{rt}(A_0))\cdot P(A_{rt}=A_0)\\ &= \frac{1}{|\Pi_K|}\cdot P(A_{rt}=A_0)\end{aligned}$$

where the third line is by the independence of U_{rt} and A_{rt} and the fourth line since $U_{rt} \sim \mathcal{U}(\Pi_K)$. The above shows that the joint distribution factorizes as uniform distribution for P_{rt} and original (marginal) distribution for A_{rt}, which proves the claim. □

F Consistency of connectivity matrix estimation

Given an $n \times n$ adjacency matrix A, we apply a community detection algorithm to obtain label vector $\widehat{z} = (\widehat{z}_1, \ldots, \widehat{z}_n) \in [K]^n$. Let d_{NH} be the normalized Hamming distance, that is, $d_{\text{NH}}(z,\widehat{z}) = d_{\text{H}}(z,\widehat{z})/n$ where $d_{\text{H}}(z,\widehat{z}) = \sum_{i=1}^n 1\{z_i \neq \widehat{z}_i\}$ is the Hamming distance.

Remark F.1. In the proof of consistency below, our results are given in terms of $\|E\|_{\max}$ where E here is a $K \times K$ matrix placeholder for different matrices as given in Propositions F.3 and F.4. However, to simplify the proofs below, when computing upper bounds for $\|E\|_{\max} = \max_{k,\ell\in\{1,\ldots,K\}}|E_{k\ell}|$, we focus on the case $k \neq \ell$ and omit the case $k = \ell$ as it is similar. The only difference in dealing with $k = \ell$ is that the constants would need to be inflated.

We repeated use the following result, which follows for example from Bernstein's inequality:

Proposition F.2. Suppose that $S = \sum_{i=1}^n X_i$ where $X_i \sim \mathsf{Ber}(p_i)$ independently for $i \in [n]$. Assume that $\max_i p_i \le p$. Then, for all $\delta \in [0,1]$,

$$P\Big(\widehat{p} - p \ge \delta p\Big) \le e^{-\delta^2 np/3},$$

and the same inequality holds for $P(p - \widehat{p} \ge \delta p)$.

Proof. By Bernstein inequality, using $|X_i - p_i| \le 1$, for any $u > 0$, we have

$$P(S - E[S] \ge nu) \le \exp\Big(-\frac{nu^2}{2\bar{\sigma}^2 + 2u/3}\Big)$$

where $\bar{\sigma}^2 = \frac{1}{n}\sum_{i=1}^n \mathrm{var}(X_i)$. We have $\bar{\sigma}^2 \le p$ and $E[S] \le np$. Setting $u = \delta p$, we have

$$\begin{aligned} P(S - np \ge \delta np) &\le \exp\Big(-\frac{n\delta^2 p^2}{2p + 2\delta p/3}\Big) \\ &= \exp\Big(-\frac{\delta^2}{2 + 2\delta/3} np\Big) \le \exp(-3\delta^2 np/8) \end{aligned}$$

where the last inequality uses $\delta \le 1$. Replacing $3/8$ with $1/3$ gives a further upper bound. □

Let us write $(A, z) \sim \mathsf{SBM}_n(B, \pi)$ for an n-node draw from a Bayesian SBM, with connectivity B and class prior π.

Proposition F.3. Assume that $(A, z) \sim \mathsf{SBM}_n(B, \pi)$ with B satisfying (18). Let $S = \mathcal{S}(A, z),\ \widehat{S} = \mathcal{S}(A, \widehat{z})$ and $m = \mathcal{C}(z),\ \widehat{m} = \mathcal{C}(\widehat{z})$. Set

$$\widehat{B} = \mathcal{B}(A, \widehat{z}) = \widehat{S}/\widehat{m}, \quad \widetilde{B} = \mathcal{B}(A, z) = S/m.$$

Fix $\kappa \in (0,1)$ and $\alpha > 0$, and let $\beta = 1/(\bar{\kappa}\pi_{\min})$ with $\bar{\kappa} = 1 - \kappa$, and define

$$\delta_n := \sqrt{\frac{3\beta^2 \alpha \log n}{n\nu_n}}.$$

Assume that $\delta_n \le 1$ and $\nu_n \ge 3(1+\alpha)\log n$. For $\sigma \in \Pi_K$, let $\varepsilon(\sigma) = d_{\mathrm{NH}}(z, \sigma \circ \widehat{z})$. Then, with probability at least $1 - 3K^2 n^{-\alpha} - Ke^{-\kappa^2 n\pi_{\min}/3}$, for all $\sigma \in \Pi_K$ such that $12\beta^2\varepsilon(\sigma) \le 1$,

$$\|P_\sigma \widehat{B} P_\sigma^T - \widetilde{B}\|_{\max} \le 56 C_1 \beta^2 \cdot \frac{\nu_n}{n}\varepsilon(\sigma), \tag{59}$$

$$\|\widetilde{B} - B\|_{\max} \le C_1 \frac{\nu_n}{n}\delta_n, \tag{60}$$

$$\|S - P_\sigma \widehat{S} P_\sigma^T\|_{\max} \le 8 C_1 \beta \cdot n\nu_n \varepsilon(\sigma), \tag{61}$$

$$\|m - P_\sigma \widehat{m} P_\sigma^T\|_{\max} \le 6\beta \cdot n^2 \varepsilon(\sigma), \tag{62}$$

$$\min_{k,\ell} m_{k\ell} \ge n^2/\beta^2. \tag{63}$$

of Proposition F.3. By redefining $\widehat{z}$ to be $\sigma \circ \widehat{z}$, one gets $P_\sigma \widehat{B} P_\sigma^T$ in place of $\widehat{B}$. Thus, without loss of generality, we can assume σ to be the identity permutation. Let

$$d_k(z, \widehat{z}) = \sum_i \big|1\{z_i = k\} - 1\{\widehat{z}_i = k\}\big|, \quad d_H(z, \widehat{z}) = \sum_i 1\{z_i \ne \widehat{z}_i\}$$

so that $\sum_k d_k(z, \widehat{z}) = 2 d_H(z, \widehat{z})$. This can verified by writing

$$\big|1\{z_i = k\} - 1\{\widehat{z}_i = k\}\big| = 1\{z_i = k, \widehat{z}_i \ne k\} + 1\{z_i \ne k, \widehat{z}_i = k\}.$$

Let $n_k = \sum_{i=1}^n 1\{z_i = k\}$ and $\widehat{n}_k = \sum_{i=1}^n 1\{\widehat{z}_i = k\}$ and note that $|n_k - \widehat{n}_k| \le d_k(z, \widehat{z})$. Let us also write

$$\rho := \max_k \frac{d_k(z, \widehat{z})}{n_k} \tag{64}$$

and note that $|(\widehat{n}_k/n_k) - 1| \le \rho$.

Let us consider some events. First, consider event

$$\mathcal{E}_0 := \Big\{ \max_j \sum_i A_{ij} \le 2C_1\nu_n \Big\}. \tag{65}$$

Applying Proposition F.2 conditioned on z, with $p = C_1\nu_n/n$, and then taking expectation of both sides to remove the conditioning, we have

$$P\Big(\sum_i A_{ij} \ge C_1\nu_n(1+u) \Big) \le e^{-u^2 C_1\nu_n/3}.$$

Take $u = 1/\sqrt{C_1}$. Then, $P(\mathcal{E}_0^c) \le n^{-\alpha}$ by union bound and assumption $\nu_n/3 \ge (1+\alpha)\log n$.

Next, note that $P(n_k \le (1-\kappa)n\pi_k) \le \exp(-\kappa^2 n\pi_k/3)$ for $\kappa \in (0,1)$, by Proposition F.2. Consider the event

$$\mathcal{E}_1 \ := \ \Big\{ \min_{k=1,\dots K} n_k \ \ge n/\beta \Big\},$$

where $1/\beta = (1-\kappa)\pi_{\min}$. Then, by union bound $P(\mathcal{E}_1^c) \le K\exp(-\kappa^2 n\pi_{\min}/3)$. On $\mathcal{E}_1$,

$$\rho := \max_k \frac{d_k(z,\widehat{z})}{n_k} \le \frac{2d_H(z,\widehat{z})}{\bar{\kappa}\pi_{\min} n} = 2\beta d_{\text{NH}}(z,\widehat{z}). \tag{66}$$

Next, we apply Proposition F.2 conditioned on z, to get

$$\begin{aligned} P\big(|B_{k\ell} - \widetilde{B}_{k\ell}| \ge \delta_n B_{k\ell} \ | \, z \big) \cdot 1_{\mathcal{E}_1} \ &\le \ 2e^{-\delta_n^2 n_k n_\ell B_{k\ell}/3} \cdot 1_{\mathcal{E}_1} \\ &\le \ 2e^{-\delta_n^2 (n^2/\beta^2)(\nu_n/n)/3} \end{aligned}$$

where we have used the lower bound in (18). Take $\delta_n^2 = 3\beta^2\alpha\log n/(n\nu_n)$ and let

$$\mathcal{E}_2 = \Big\{ |B_{k\ell} - \widetilde{B}_{k\ell}| \le \delta_n B_{k\ell} \ \text{ for all } k,\ell \in [K] \Big\}. \tag{67}$$

Then, by union bound, $P(\mathcal{E}_1 \cap \mathcal{E}_2^c) \le P(\mathcal{E}_2^c \,|\, \mathcal{E}_1) \le 2K^2 n^{-\alpha}$. We will work on $\mathcal{E}_0 \cap \mathcal{E}_1 \cap \mathcal{E}_2$ and note that

$$P(\mathcal{E}_0 \cap \mathcal{E}_1 \cap \mathcal{E}_2) \ge 1 - P(\mathcal{E}_0^c) - P(\mathcal{E}_1^c) - P(\mathcal{E}_1 \cap \mathcal{E}_2^c).$$

Note that on $\mathcal{E}_2$,

$$\max_{k,\ell} |B_{k\ell} - \widetilde{B}_{k\ell}| \le \delta_n \cdot C_1\nu_n/n, \tag{68}$$

$$\max_{k,\ell} \widetilde{B}_{k\ell} \le 2C_1\nu_n/n \tag{69}$$

using the upper bound in (18) and assumption $\delta_n \le 1$. This proves (60).

Now, let us establish (61). We have

$$|S_{k\ell} - \widehat{S}_{k\ell}| \le \sum_{i,j} A_{ij} |1\{z_i = k, z_j = \ell\} - 1\{\widehat{z}_i = k, \widehat{z}_j = \ell\}|.$$

Adding and subtracting $1\{\widehat{z}_i = k, z_j = \ell\}$ and expanding by triangle inequality, we get $|S_{k\ell} - \widehat{S}_{k\ell}| \le T_{21} + T_{22}$ where

$$\begin{aligned} T_{21} &:= \sum_{i,j} A_{ij} |1\{z_i = k, z_j = \ell\} - 1\{\widehat{z}_i = k, z_j = \ell\}|, \\ T_{22} &:= \sum_{i,j} A_{ij} |1\{\widehat{z}_i = k, \widehat{z}_j = \ell\} - 1\{\widehat{z}_i = k, z_j = \ell\}|. \end{aligned}$$

Consider T_{22} first. We have

$$\begin{aligned} T_{22} &= \sum_j \Big(\sum_i A_{ij} 1\{\widehat{z}_i = k\} \cdot |1\{\widehat{z}_j = \ell\} - 1\{z_j = \ell\}| \Big) \\ &\le \sum_j \Big(\sum_i A_{ij} \cdot |1\{\widehat{z}_j = \ell\} - 1\{z_j = \ell\}| \Big) \\ &\le 2C_1\nu_n \cdot d_\ell(z,\widehat{z}) \end{aligned}$$

on $\mathcal{E}_0$. Similarly, $T_{21} \le 2C_1\nu_n \cdot d_k(z,\widehat{z})$ on $\mathcal{E}_0$. Then,

$$|S_{k\ell} - \widehat{S}_{k\ell}| \le 2C_1\nu_n \cdot (n_\ell + n_k)\rho \le 4C_1 n\nu_n\rho \tag{70}$$

using $n_k \le n$ for all k. Combined with (66), this proves (61).

Let us define $\widetilde{B}' = \widehat{S}/m$. For $k \neq \ell$, we have $m_{k\ell} = n_k n_\ell$. Then, on $\mathcal{E}_1$,

$$|\widetilde{B}_{k\ell} - \widetilde{B}'_{k\ell}| = \frac{|S_{k\ell} - \widehat{S}_{k\ell}|}{n_k n_\ell} \le 2C_1\nu_n \cdot (n_k^{-1} + n_\ell^{-1})\rho \le 4C_1\beta\frac{\nu_n}{n}\rho$$

using $n_k \ge n/\beta$ for all k. By (69), on $\mathcal{E}_2$, we also have

$$\widetilde{B}'_{k\ell} \le 2C_1\frac{\nu_n}{n}(1 + 2\beta\rho) \le 4C_1\frac{\nu_n}{n}.$$

assuming $2\beta\rho \le 1$. Next we note that

$$\Big|\frac{\widehat{m}_{k\ell}}{m_{k\ell}} - 1\Big| = \Big|\frac{\widehat{n}_k\widehat{n}_\ell}{n_k n_\ell} - 1\Big| \le 3\rho.$$

Assuming that $3\rho \le 1/2$, letting $x = \widehat{m}_{k\ell}/m_{k\ell}$, we have $|1-x| \le 3\rho \le 1/2$, hence $|1-x^{-1|}| \le 6\rho$. It follows that on $\mathcal{E}_2$

$$|\widetilde{B}'_{k\ell} - \widehat{B}_{k\ell}| = \frac{\widehat{S}_{k\ell}}{m_{k\ell}}\Big|1 - \frac{m_{k\ell}}{\widehat{m}_{k\ell}}\Big| \le \widetilde{B}'_{k\ell} \cdot 6\rho \le 24C_1\frac{\nu_n}{n}\rho.$$

By triangle inequality

$$\begin{aligned}\|\widehat{B} - \widetilde{B}\|_{\max} &\le \|\widehat{B} - \widetilde{B}'\|_{\max} + \|\widetilde{B}' - \widetilde{B}\|_{\max} \\ &\le 24C_1\frac{\nu_n}{n}\rho + 4C_1\beta\frac{\nu_n}{n}\rho \\ &\le 28C_1\frac{\nu_n}{n}\beta\rho\end{aligned}$$

using $\beta \ge 1$ (we are weakening the constants in favor of a simpler expression).

For (62), we have $|[m - \widehat{m}]_{k\ell}| \le n_k n_\ell|\frac{\widehat{n}_k\widehat{n}_\ell}{n_k n_\ell} - 1| \le n^2 3\rho$. Putting the pieces together combined with (66) proves the claim. □

F.1 Multi-network extension

Proposition F.4. Assume that $(A_t, z_t) \sim \mathsf{SBM}_{n_t}(B, \pi_t)$ for $t \in [N]$, where B satisfies (18) and n_t satisfy

$$n \le n_t \le C_2 n. \tag{71}$$

Let $S_t = \mathcal{S}(A_t, z_t)$, $\widehat{S}_t = \mathcal{S}(A_t, \widehat{z}_t)$ and $m_t = \mathcal{C}(z_t)$, $\widehat{m}_t = \mathcal{C}(\widehat{z}_t)$. For $\sigma = (\sigma_t) \in \Pi_K^{\otimes N}$, set

$$\widehat{B}(\sigma) := \frac{\sum_t P_{\sigma_t}\widehat{S}_t P_{\sigma_t}^T}{\sum_t P_{\sigma_t}\widehat{m}_t P_{\sigma_t}^T}, \quad \widetilde{B} := \frac{\sum_t S_t}{\sum_t m_t}, \quad \varepsilon_t(\sigma) := d_{\mathrm{NH}}(z_t, \sigma_t \circ \widehat{z}_t)$$

where we interpret the division of matrices as elementwise.

Fix $\kappa \in (0,1)$ and $\alpha > 0$, let $\beta = 1/(\bar{\kappa}\pi_{\min})$ with $\bar{\kappa} = 1 - \kappa$, and define

$$\delta_n := \sqrt{\frac{3\beta^2\alpha\log n}{Nn\nu_n}}.$$

Assume that $\delta_n \le 1$ and $\nu_n \ge 3(1+\alpha)\log n$. Then, with probability at least $1 - (C_2N + 2K^2)n^{-\alpha} - NKe^{-\kappa^2 n\pi_{\min}/3}$, we have, for all $\sigma = (\sigma_t) \in \Pi_K^{\otimes N}$ for which $\max_t \varepsilon_t(\sigma) \le 1/(2\beta^3)$,

$$\|\widehat{B}(\sigma) - \widetilde{B}\|_{\max} \le 40\, C_1\beta^3\frac{\nu_n}{n}\varepsilon, \tag{72}$$

$$\|\widetilde{B} - B\|_{\max} \le C_1\frac{\nu_n}{n}\delta_n, \tag{73}$$

$$\|S_t - P_{\sigma_t}\widehat{S}_t P_{\sigma_t}^T\|_{\max} \le 8C_1\beta \cdot \nu_n n_t\, \varepsilon_t(\sigma), \tag{74}$$

$$\|m_t - P_{\sigma_t}\widehat{m}_t P_{\sigma_t}^T\|_{\max} \le 6\beta \cdot n_t^2\, \varepsilon_t(\sigma), \tag{75}$$

$$\min_{k,\ell}[m_t]_{k\ell} \ge n_t^2/\beta^2, \quad \text{for all } t \in [N] \tag{76}$$

where $\beta = 1/(\bar{\kappa}\pi_{\min})$ with $\bar{\kappa} = 1 - \kappa$.

Note that n_t here is deterministic (size of the t-th network) and different from n_k in the proof of Proposition F.3.

Proof. By redefining $\widehat{z}_t$ to be $\sigma_t(\widehat{z}_t)$, we can assume, without loss of generality, that σ_t is the identity permutation (and hence $P_{\sigma_t} = I_K$) for all t. Note that none of the events we consider below depend on σ, hence the result indeed holds for all σ simultaneously.

First, consider event

$$\mathcal{E}_0 := \Big\{ \max_{t\in[N],\, j\in[n_t]} \sum_{i=1}^{n_t} [A_t]_{ij} \le 2C_1C_2\nu_n \Big\}. \tag{77}$$

Applying Proposition F.2 conditioned on z_t, with $p = C_1\nu_n/n$ and $n_t p \le C_1C_2\nu_n$, and then taking expectation of both sides to remove the conditioning, we have

$$P\Big(\sum_{i=1}^{n_t} [A_t]_{ij} \ge C_1C_2\nu_n(1+u)\Big) \le e^{-u^2 C_1C_2\nu_n/3}.$$

Take $u = 1/\sqrt{C_1C_2} \le 1$. Then, by union bound and assumption $\nu_n/3 \ge (1+\alpha)\log n$, we have $P(\mathcal{E}_0^c) \le C_2 N n^{-\alpha}$.

Next, let $n_{tk} := \sum_{i=1}^{n_t} 1\{(z_t)_i = k\}$, the (true) number of nodes in community k in network t. By Proposition F.2, we have $P(n_{tk} \le \bar{\kappa} n_t \pi_{tk}) \le \exp(-\kappa^2 n_t \pi_{tk}/3)$ for $\kappa \in (0,1)$. Consider the event defined as

$$\mathcal{E}_1 \;:=\; \Big\{ \min_{k=1,\dots K} n_{tk} \;\ge n_t/\beta,\; \forall t \in [N] \Big\},$$

where $\beta = 1/(\bar{\kappa}\pi_{\min})$ with $\bar{\kappa} = 1-\kappa$. Then, by union bound and using $n_t \ge n$,

$$P(\mathcal{E}_1^c) \le NK \exp(-\kappa^2 n \pi_{\min}/3).$$

By the same argument as in the proof of Proposition F.3,and letting $\varepsilon_t = d_{\mathrm{NH}}(z_t, \widehat{z}_t)$, we have on $\mathcal{E}_0 \cap \mathcal{E}_1$,

$$\|S_t - \widehat{S}_t\|_{\max} \le 8C_1\beta\nu_n n_t \varepsilon_t, \quad \|m_t - \widehat{m}_t\|_{\max} \le 6\beta n_t^2 \varepsilon_t, \quad [m_t]_{k\ell} \ge n_t^2/\beta^2. \tag{78}$$

Let us now control $\widetilde{B} = (\sum_t S_t)/(\sum_t m_t)$. Conditioned on z_t, we have

$$[\textstyle\sum_t S_t]_{k\ell} \sim \mathsf{Bin}\big([x \sum_t m_t]_{k\ell}, B_{k\ell}\big).$$

Also note that, for $k \ne \ell$, we have $[\sum_t m_t]_{k\ell} = \sum_t n_{tk} n_{t\ell} \ge \sum_t n_t^2/\beta^2 \ge Nn^2/\beta^2$ on $\mathcal{E}_1$. Then, applying Proposition F.2 conditioned on z_t, we get

$$\begin{aligned} P\big(|B_{k\ell} - \widetilde{B}_{k\ell}| \ge \delta_n B_{k\ell} \mid z_t\big) \cdot 1_{\mathcal{E}_1} \;&\le\; 2e^{-\delta_n^2 [\sum_t m_t]_{k\ell} B_{k\ell}/3} \cdot 1_{\mathcal{E}_1} \\ &\le\; 2e^{-\delta_n^2 (Nn^2/\beta^2)(\nu_n/n)/3} \end{aligned}$$

where we have used the lower bound in (18). Take $\delta_n^2 = 3\beta^2\alpha\log n/(Nn\nu_n)$ and let

$$\mathcal{E}_2 = \Big\{ |B_{k\ell} - \widetilde{B}_{k\ell}| \le \delta_n B_{k\ell} \;\text{ for all } k,\ell \in [K] \Big\}. \tag{79}$$

Then, by union bound, $P(\mathcal{E}_1 \cap \mathcal{E}_2^c) \le P(\mathcal{E}_2^c \,|\, \mathcal{E}_1) \le 2K^2 n^{-\alpha}$. We will work on $\mathcal{E}_0 \cap \mathcal{E}_1 \cap \mathcal{E}_2$ and note that

$$P(\mathcal{E}_0 \cap \mathcal{E}_1 \cap \mathcal{E}_2) \ge 1 - P(\mathcal{E}_0^c) - P(\mathcal{E}_1^c) - P(\mathcal{E}_1 \cap \mathcal{E}_2^c).$$

Note that on $\mathcal{E}_2$,

$$\max_{k,\ell} |B_{k\ell} - \widetilde{B}_{k\ell}| \le \delta_n \cdot C_1\nu_n/n, \tag{80}$$

$$\max_{k,\ell} \widetilde{B}_{k\ell} \le 2C_1\nu_n/n \tag{81}$$

using the upper bound in (18) and assumption $\delta_n \le 1$. This proves (60).

Next we control the deviation $\widehat{B} - \widetilde{B}$. Treating division (and taking absolute values) of matrices as element-wise,

$$\frac{\sum_t \widehat{S}_t}{\sum_t \widehat{m}_t} = \frac{\sum_t S_t + \sum_t (\widehat{S}_t - S_t)}{\sum_t m_t + \sum_t (\widehat{m}_t - m_t)} =: \frac{a+b}{c+d}.$$

We then have

$$\Big|\frac{a+b}{c+d} - \frac{a}{c}\Big| = \Big|\frac{(a/c)+(b/c)}{1+(d/c)} - (a/c)\Big| = \Big|\frac{(b/c)-(a/c)(d/c)}{1+(d/c)}\Big| \le \frac{|(b/c)-(a/c)(d/c)|}{1-|d/c|}.$$

Let $\varepsilon = \max_t \varepsilon_t$. Then, we have

$$\begin{aligned} |[b/c]_{k\ell}| \le \frac{\sum_t |[\widehat{S}_t - S_t]_{k\ell}|}{\sum_t [m_t]_{k\ell}} &\overset{(a)}{\le} 8C_1\beta^3\nu_n \frac{\sum_t n_t\varepsilon_t}{\sum_t n_t^2} \\ &\overset{(b)}{\le} 8C_1\beta^3\frac{\nu_n}{n}\frac{\sum_t n_t\varepsilon_t}{\sum_t n_t} \le 8C_1\beta^3\frac{\nu_n}{n}\varepsilon \end{aligned}$$

where (a) is by (78) and (b) uses assumption $n_t \ge n$. Similarly, we have $|[d/c]_{k\ell}| \le 6\beta^3\varepsilon$. Note that $a/c = \widetilde{B}$ hence elementwise in $[0, 2C_1\nu_n/n]$ on $\mathcal{E}_2$. Assuming that $\beta^3\varepsilon \le 1/2$, we have,

$$\frac{|(b/c)-(a/c)(d/c)|}{1-|d/c|} \le \frac{|b/c|+(a/c)|d/c|}{1-\frac{1}{2}} \le 40\, C_1\beta^3\frac{\nu_n}{n}\varepsilon.$$

(where the final inequality means every element of the LHS is $\le$ RHS) . That is, we have shown

$$\Big|\frac{\sum_t \widehat{S}_t}{\sum_t \widehat{m}_t} - \frac{\sum_t S_t}{\sum_t m_t}\Big| \le 40\, C_1\beta^3\frac{\nu_n}{n}\varepsilon.$$

This proves (72) and finishes the proof. □

NeurIPS Paper Checklist

1. **Claims**

   Question: Do the main claims made in the abstract and introduction accurately reflect the paper's contributions and scope?

   Answer: [Yes]

   Justification: Yes, the main claims made in the abstract and introduction accurately reflect the paper's contributions and scope. The work delivers on its promises by providing a consistent and powerful two-sample test under the stochastic block model, along with theoretical guarantees and empirical validation.

   Guidelines:

   - The answer NA means that the abstract and introduction do not include the claims made in the paper.
   - The abstract and/or introduction should clearly state the claims made, including the contributions made in the paper and important assumptions and limitations. A No or NA answer to this question will not be perceived well by the reviewers.
   - The claims made should match theoretical and experimental results, and reflect how much the results can be expected to generalize to other settings.
   - It is fine to include aspirational goals as motivation as long as it is clear that these goals are not attained by the paper.

2. **Limitations**

   Question: Does the paper discuss the limitations of the work performed by the authors?

   Answer: [Yes]

   Justification: The discussion of the strength of the assumptions for the main results is included in the remarks in the supplementary materials.

   Guidelines:

   - The answer NA means that the paper has no limitation while the answer No means that the paper has limitations, but those are not discussed in the paper.
   - The authors are encouraged to create a separate "Limitations" section in their paper.
   - The paper should point out any strong assumptions and how robust the results are to violations of these assumptions (e.g., independence assumptions, noiseless settings, model well-specification, asymptotic approximations only holding locally). The authors should reflect on how these assumptions might be violated in practice and what the implications would be.
   - The authors should reflect on the scope of the claims made, e.g., if the approach was only tested on a few datasets or with a few runs. In general, empirical results often depend on implicit assumptions, which should be articulated.
   - The authors should reflect on the factors that influence the performance of the approach. For example, a facial recognition algorithm may perform poorly when image resolution is low or images are taken in low lighting. Or a speech-to-text system might not be used reliably to provide closed captions for online lectures because it fails to handle technical jargon.
   - The authors should discuss the computational efficiency of the proposed algorithms and how they scale with dataset size.
   - If applicable, the authors should discuss possible limitations of their approach to address problems of privacy and fairness.
   - While the authors might fear that complete honesty about limitations might be used by reviewers as grounds for rejection, a worse outcome might be that reviewers discover limitations that aren't acknowledged in the paper. The authors should use their best judgment and recognize that individual actions in favor of transparency play an important role in developing norms that preserve the integrity of the community. Reviewers will be specifically instructed to not penalize honesty concerning limitations.

3. **Theory assumptions and proofs**

Question: For each theoretical result, does the paper provide the full set of assumptions and a complete (and correct) proof?

Answer: [Yes]

Justification: The main results and auxiliary lemmas have complete proofs that are provided in the supplementary materials. We also provide meaningful discussion on the underlying set of assumptions.

Guidelines:

- The answer NA means that the paper does not include theoretical results.
- All the theorems, formulas, and proofs in the paper should be numbered and cross-referenced.
- All assumptions should be clearly stated or referenced in the statement of any theorems.
- The proofs can either appear in the main paper or the supplemental material, but if they appear in the supplemental material, the authors are encouraged to provide a short proof sketch to provide intuition.
- Inversely, any informal proof provided in the core of the paper should be complemented by formal proofs provided in appendix or supplemental material.
- Theorems and Lemmas that the proof relies upon should be properly referenced.

4. **Experimental result reproducibility**

Question: Does the paper fully disclose all the information needed to reproduce the main experimental results of the paper to the extent that it affects the main claims and/or conclusions of the paper (regardless of whether the code and data are provided or not)?

Answer: [Yes]

Justification: The experiments section and the corresponding sections of supplementary material provide a clear description of the setup.

Guidelines:

- The answer NA means that the paper does not include experiments.
- If the paper includes experiments, a No answer to this question will not be perceived well by the reviewers: Making the paper reproducible is important, regardless of whether the code and data are provided or not.
- If the contribution is a dataset and/or model, the authors should describe the steps taken to make their results reproducible or verifiable.
- Depending on the contribution, reproducibility can be accomplished in various ways. For example, if the contribution is a novel architecture, describing the architecture fully might suffice, or if the contribution is a specific model and empirical evaluation, it may be necessary to either make it possible for others to replicate the model with the same dataset, or provide access to the model. In general. releasing code and data is often one good way to accomplish this, but reproducibility can also be provided via detailed instructions for how to replicate the results, access to a hosted model (e.g., in the case of a large language model), releasing of a model checkpoint, or other means that are appropriate to the research performed.
- While NeurIPS does not require releasing code, the conference does require all submissions to provide some reasonable avenue for reproducibility, which may depend on the nature of the contribution. For example
  (a) If the contribution is primarily a new algorithm, the paper should make it clear how to reproduce that algorithm.
  (b) If the contribution is primarily a new model architecture, the paper should describe the architecture clearly and fully.
  (c) If the contribution is a new model (e.g., a large language model), then there should either be a way to access this model for reproducing the results or a way to reproduce the model (e.g., with an open-source dataset or instructions for how to construct the dataset).
  (d) We recognize that reproducibility may be tricky in some cases, in which case authors are welcome to describe the particular way they provide for reproducibility. In the case of closed-source models, it may be that access to the model is limited in

some way (e.g., to registered users), but it should be possible for other researchers to have some path to reproducing or verifying the results.

5. **Open access to data and code**

   Question: Does the paper provide open access to the data and code, with sufficient instructions to faithfully reproduce the main experimental results, as described in supplemental material?

   Answer: [Yes]

   Justification: Yes, the paper provides open access to both the code and data, along with sufficient instructions to enable faithful reproduction of the main experimental results.

   Guidelines:

   - The answer NA means that paper does not include experiments requiring code.
   - Please see the NeurIPS code and data submission guidelines (`https://nips.cc/public/guides/CodeSubmissionPolicy`) for more details.
   - While we encourage the release of code and data, we understand that this might not be possible, so "No" is an acceptable answer. Papers cannot be rejected simply for not including code, unless this is central to the contribution (e.g., for a new open-source benchmark).
   - The instructions should contain the exact command and environment needed to run to reproduce the results. See the NeurIPS code and data submission guidelines (`https://nips.cc/public/guides/CodeSubmissionPolicy`) for more details.
   - The authors should provide instructions on data access and preparation, including how to access the raw data, preprocessed data, intermediate data, and generated data, etc.
   - The authors should provide scripts to reproduce all experimental results for the new proposed method and baselines. If only a subset of experiments are reproducible, they should state which ones are omitted from the script and why.
   - At submission time, to preserve anonymity, the authors should release anonymized versions (if applicable).
   - Providing as much information as possible in supplemental material (appended to the paper) is recommended, but including URLs to data and code is permitted.

6. **Experimental setting/details**

   Question: Does the paper specify all the training and test details (e.g., data splits, hyperparameters, how they were chosen, type of optimizer, etc.) necessary to understand the results?

   Answer: [Yes]

   Justification: The experiments in section 5 as well as in the supplementary material are provided with complete details sufficient to reproduce the results.

   Guidelines:

   - The answer NA means that the paper does not include experiments.
   - The experimental setting should be presented in the core of the paper to a level of detail that is necessary to appreciate the results and make sense of them.
   - The full details can be provided either with the code, in appendix, or as supplemental material.

7. **Experiment statistical significance**

   Question: Does the paper report error bars suitably and correctly defined or other appropriate information about the statistical significance of the experiments?

   Answer: [Yes]

   Justification: The experiment section gives clear description of the factors of variability as well as the methods of calculation of the statistical significance of the tests.

   Guidelines:

   - The answer NA means that the paper does not include experiments.

- The authors should answer "Yes" if the results are accompanied by error bars, confidence intervals, or statistical significance tests, at least for the experiments that support the main claims of the paper.
- The factors of variability that the error bars are capturing should be clearly stated (for example, train/test split, initialization, random drawing of some parameter, or overall run with given experimental conditions).
- The method for calculating the error bars should be explained (closed form formula, call to a library function, bootstrap, etc.)
- The assumptions made should be given (e.g., Normally distributed errors).
- It should be clear whether the error bar is the standard deviation or the standard error of the mean.
- It is OK to report 1-sigma error bars, but one should state it. The authors should preferably report a 2-sigma error bar than state that they have a 96% CI, if the hypothesis of Normality of errors is not verified.
- For asymmetric distributions, the authors should be careful not to show in tables or figures symmetric error bars that would yield results that are out of range (e.g. negative error rates).
- If error bars are reported in tables or plots, The authors should explain in the text how they were calculated and reference the corresponding figures or tables in the text.

8. **Experiments compute resources**

   Question: For each experiment, does the paper provide sufficient information on the computer resources (type of compute workers, memory, time of execution) needed to reproduce the experiments?

   Answer: [Yes]

   Justification: The paper indicates the characteristic of the technical cluster used for the experiments.

   Guidelines:

   - The answer NA means that the paper does not include experiments.
   - The paper should indicate the type of compute workers CPU or GPU, internal cluster, or cloud provider, including relevant memory and storage.
   - The paper should provide the amount of compute required for each of the individual experimental runs as well as estimate the total compute.
   - The paper should disclose whether the full research project required more compute than the experiments reported in the paper (e.g., preliminary or failed experiments that didn't make it into the paper).

9. **Code of ethics**

   Question: Does the research conducted in the paper conform, in every respect, with the NeurIPS Code of Ethics `https://neurips.cc/public/EthicsGuidelines`?

   Answer: [Yes]

   Justification: Yes, the research conducted in this paper conforms with the NeurIPS Code of Ethics.

   Guidelines:

   - The answer NA means that the authors have not reviewed the NeurIPS Code of Ethics.
   - If the authors answer No, they should explain the special circumstances that require a deviation from the Code of Ethics.
   - The authors should make sure to preserve anonymity (e.g., if there is a special consideration due to laws or regulations in their jurisdiction).

10. **Broader impacts**

    Question: Does the paper discuss both potential positive societal impacts and negative societal impacts of the work performed?

    Answer: [No]

Justification: No, the paper does not explicitly discuss potential positive or negative societal impacts of the work. The focus is on methodological and theoretical contributions to network inference, without specific commentary on how the techniques might influence or be applied in societal contexts.

Guidelines:

- The answer NA means that there is no societal impact of the work performed.
- If the authors answer NA or No, they should explain why their work has no societal impact or why the paper does not address societal impact.
- Examples of negative societal impacts include potential malicious or unintended uses (e.g., disinformation, generating fake profiles, surveillance), fairness considerations (e.g., deployment of technologies that could make decisions that unfairly impact specific groups), privacy considerations, and security considerations.
- The conference expects that many papers will be foundational research and not tied to particular applications, let alone deployments. However, if there is a direct path to any negative applications, the authors should point it out. For example, it is legitimate to point out that an improvement in the quality of generative models could be used to generate deepfakes for disinformation. On the other hand, it is not needed to point out that a generic algorithm for optimizing neural networks could enable people to train models that generate Deepfakes faster.
- The authors should consider possible harms that could arise when the technology is being used as intended and functioning correctly, harms that could arise when the technology is being used as intended but gives incorrect results, and harms following from (intentional or unintentional) misuse of the technology.
- If there are negative societal impacts, the authors could also discuss possible mitigation strategies (e.g., gated release of models, providing defenses in addition to attacks, mechanisms for monitoring misuse, mechanisms to monitor how a system learns from feedback over time, improving the efficiency and accessibility of ML).

11. **Safeguards**

Question: Does the paper describe safeguards that have been put in place for responsible release of data or models that have a high risk for misuse (e.g., pretrained language models, image generators, or scraped datasets)?

Answer: [NA]

Justification: The paper poses no such risks.

Guidelines:

- The answer NA means that the paper poses no such risks.
- Released models that have a high risk for misuse or dual-use should be released with necessary safeguards to allow for controlled use of the model, for example by requiring that users adhere to usage guidelines or restrictions to access the model or implementing safety filters.
- Datasets that have been scraped from the Internet could pose safety risks. The authors should describe how they avoided releasing unsafe images.
- We recognize that providing effective safeguards is challenging, and many papers do not require this, but we encourage authors to take this into account and make a best faith effort.

12. **Licenses for existing assets**

Question: Are the creators or original owners of assets (e.g., code, data, models), used in the paper, properly credited and are the license and terms of use explicitly mentioned and properly respected?

Answer: [Yes]

Justification: The datasets are both credited in the reference section.

Guidelines:

- The answer NA means that the paper does not use existing assets.
- The authors should cite the original paper that produced the code package or dataset.

- The authors should state which version of the asset is used and, if possible, include a URL.
- The name of the license (e.g., CC-BY 4.0) should be included for each asset.
- For scraped data from a particular source (e.g., website), the copyright and terms of service of that source should be provided.
- If assets are released, the license, copyright information, and terms of use in the package should be provided. For popular datasets, `paperswithcode.com/datasets` has curated licenses for some datasets. Their licensing guide can help determine the license of a dataset.
- For existing datasets that are re-packaged, both the original license and the license of the derived asset (if it has changed) should be provided.
- If this information is not available online, the authors are encouraged to reach out to the asset's creators.

13. **New assets**

    Question: Are new assets introduced in the paper well documented and is the documentation provided alongside the assets?

    Answer: [Yes]

    Justification: Yes, the paper introduces a new algorithm for two-sample testing on networks, and the accompanying implementation is given in the supplementary material. The attached files include implementations for all the tests ensuring reproducibility.

    Guidelines:

    - The answer NA means that the paper does not release new assets.
    - Researchers should communicate the details of the dataset/code/model as part of their submissions via structured templates. This includes details about training, license, limitations, etc.
    - The paper should discuss whether and how consent was obtained from people whose asset is used.
    - At submission time, remember to anonymize your assets (if applicable). You can either create an anonymized URL or include an anonymized zip file.

14. **Crowdsourcing and research with human subjects**

    Question: For crowdsourcing experiments and research with human subjects, does the paper include the full text of instructions given to participants and screenshots, if applicable, as well as details about compensation (if any)?

    Answer: [NA]

    Justification: The paper does not involve crowdsourcing nor research with human subjects.

    Guidelines:

    - The answer NA means that the paper does not involve crowdsourcing nor research with human subjects.
    - Including this information in the supplemental material is fine, but if the main contribution of the paper involves human subjects, then as much detail as possible should be included in the main paper.
    - According to the NeurIPS Code of Ethics, workers involved in data collection, curation, or other labor should be paid at least the minimum wage in the country of the data collector.

15. **Institutional review board (IRB) approvals or equivalent for research with human subjects**

    Question: Does the paper describe potential risks incurred by study participants, whether such risks were disclosed to the subjects, and whether Institutional Review Board (IRB) approvals (or an equivalent approval/review based on the requirements of your country or institution) were obtained?

    Answer: [NA]

Justification: The answer NA means that the paper does not involve crowdsourcing nor research with human subjects.

Guidelines:

- The answer NA means that the paper does not involve crowdsourcing nor research with human subjects.
- Depending on the country in which research is conducted, IRB approval (or equivalent) may be required for any human subjects research. If you obtained IRB approval, you should clearly state this in the paper.
- We recognize that the procedures for this may vary significantly between institutions and locations, and we expect authors to adhere to the NeurIPS Code of Ethics and the guidelines for their institution.
- For initial submissions, do not include any information that would break anonymity (if applicable), such as the institution conducting the review.

16. **Declaration of LLM usage**

Question: Does the paper describe the usage of LLMs if it is an important, original, or non-standard component of the core methods in this research? Note that if the LLM is used only for writing, editing, or formatting purposes and does not impact the core methodology, scientific rigorousness, or originality of the research, declaration is not required.

Answer: [NA]

Justification: The core method development in this research does not involve LLMs as any important, original, or non-standard components.

Guidelines:

- The answer NA means that the core method development in this research does not involve LLMs as any important, original, or non-standard components.
- Please refer to our LLM policy (`https://neurips.cc/Conferences/2025/LLM`) for what should or should not be described.