The Annals of Statistics
2020, Vol. 48, No. 4, 2208–2229
https://doi.org/10.1214/19-AOS1884

TWO-SAMPLE HYPOTHESIS TESTING FOR INHOMOGENEOUS RANDOM GRAPHS

BY DEBARGHYA GHOSHDASTIDAR[1,*], MAURILIO GUTZEIT[2,†], ALEXANDRA CARPENTIER[2,‡] AND ULRIKE VON LUXBURG[1,**]

[1]*Department of Computer Science, University of Tübingen,* [*]*debarghya.ghoshdastidar@uni-tuebingen.de;* [**]*luxburg@informatik.uni-tuebingen.de*

[2]*Faculty of Mathematics, Otto von Guericke University Magdeburg,* [†]*maurilio.gutzeit@ovgu.de;* [‡]*alexandra.carpentier@ovgu.de*

The study of networks leads to a wide range of high-dimensional inference problems. In many practical applications, one needs to draw inference from one or few large sparse networks. The present paper studies hypothesis testing of graphs in this high-dimensional regime, where the goal is to test between two populations of inhomogeneous random graphs defined on the same set of n vertices. The size of each population m is much smaller than n, and can even be a constant as small as 1. The critical question in this context is whether the problem is solvable for small m.

We answer this question from a minimax testing perspective. Let P, Q be the population adjacencies of two sparse inhomogeneous random graph models, and d be a suitably defined distance function. Given a population of m graphs from each model, we derive minimax separation rates for the problem of testing $P = Q$ against $d(P, Q) > \rho$. We observe that if m is small, then the minimax separation is too large for some popular choices of d, including total variation distance between corresponding distributions. This implies that some models that are widely separated in d cannot be distinguished for small m, and hence, the testing problem is generally not solvable in these cases.

We also show that if $m > 1$, then the minimax separation is relatively small if d is the Frobenius norm or operator norm distance between P and Q. For $m = 1$, only the latter distance provides small minimax separation. Thus, for these distances, the problem is solvable for small m. We also present near-optimal two-sample tests in both cases, where tests are adaptive with respect to sparsity level of the graphs.

1. Introduction. Analysis of random graphs has piqued the curiosity of probabilists since its inception decades ago, but the widespread use of networks in recent times has made statistical inference from random graphs a topic of immense interest for both theoretical and applied researchers. This has caused a fruitful interplay between theory and practice leading to deep understanding of statistical problems that, in turn, has led to advancements in applied research. Significant progress is clearly visible in problems related to network modelling (Albert and Barabási (2002), Lovász (2012)), community detection (Abbe and Sandon (2016), Decelle et al. (2011)), network dynamics (Berger et al. (2005)) among others, where statistically guaranteed methods have emerged as effective practical solutions. Quite surprisingly, the classical problem of hypothesis testing of random graphs is yet to benefit from such joint efforts from theoretical and applied researchers. It should be noted the problem itself is actively studied in both communities. Testing between brain or "omics" networks have surfaced as a crucial challenge in the context of both modelling and decision making (Ginestet et al. (2017), Hyduke, Lewis and Palsson (2013)). On the other hand, phase transitions are

Received June 2018; revised February 2019.
MSC2020 subject classifications. Primary 62H15; secondary 62C20, 05C80, 60B20.
Key words and phrases. Two-sample test, inhomogeneous Erdős–Rényi model, minimax testing.

now known for the problems of detecting high-dimensional geometry or strongly connected groups in large random graphs (Arias-Castro and Verzelen (2014), Bubeck et al. (2016)). However, little progress has been made in the design of consistent tests for general models of random graphs. The present paper takes a step towards addressing this general concern.

While research on testing large random graphs has been limited, hypothesis testing in large dimension is an integral part of modern statistics literature. In fact, Bai and Saranadasa (1996) demonstrated the need for studying high-dimensional statistics through a two-sample testing problem, where one tests between two n-variate normal distributions with different means by accessing m i.i.d. observations from either distributions, where $m \ll n$ (we denote the dimension by n since, in the context of graphs, the number of vertices n governs the dimensionality of the problem). More recent works in this direction provide tests and asymptotic guarantees as $m \to \infty$ and $\frac{n}{m} \to \infty$ (Cai, Liu and Xia (2014), Chen and Qin (2010), Ramdas et al. (2015)). Similar studies also exist in the context of testing whether a n-dimensional covariance matrix is identity, where only $m \ll n$ i.i.d. observations of the data is available (Arias-Castro, Bubeck and Lugosi (2015), Berthet and Rigollet (2013), Ledoit and Wolf (2002)). It is also known that the computational complexity of this problem is closely related to the clique detection problem in random graphs (Berthet and Rigollet (2013)). An extreme version of high-dimensional testing arises in the signal detection literature, where one observes a high-dimensional (Gaussian) signal, and tests the nullity of its mean. Subsequently, asymptotics of the problem is considered for $n \to \infty$ but fixed sample size ($m = 1$ or a constant), and minimax separation rates between the null and the alternative hypotheses are derived (Baraud (2002), Ingster and Suslina (2003), Mukherjee, Pillai and Lin (2015), Verzelen and Arias-Castro (2017)). The signal detection problem has also been extended in the case of matrices in the context of trace regression (Carpentier and Nickl (2015)) and sub-matrix detection (Sun and Nobel (2008)) among others, where the latter generalises the planted clique detection problem.

In practice, the problem of testing random graphs comes in a wide range of flavours. For instance, while dealing with graphs associated with chemical compounds (Shervashidze et al. (2011)) or brain networks of several patients collected at multiple laboratories (Ginestet et al. (2017)), one has access to a large number of graphs (large m). This scenario is more amenable as one can resort to the vast literature of nonparametric hypothesis testing that can even be applied to random graphs. A direct approach to this problem is to use kernel based tests (Gretton et al. (2012)) in conjunction with graph kernels (Kondor and Pan (2016), Vishwanathan et al. (2010)), which does not require any structural assumptions on the network models. However, known guarantees for such tests depend crucially on the sample size, and one cannot conclude about the fidelity of such tests for very small m. A more challenging situation arises when m is small, and unfortunately, this is often the case in network analysis. For example, if the graphs correspond to brain networks collected from patients in a single lab setup ($m < 20$), brain networks of one individual obtained from test-retest MRI scans (Landman et al. (2010)), or molecular interaction networks arising from genomic or proteomic data (Hyduke, Lewis and Palsson (2013)). Test-retest data of a patient provide only $m = 2$ networks, while omics data typically result in one large interaction network, that is, $m = 1$. Hence, designing two-sample tests for small populations for graphs is a problem of immense significance, and yet, practical tests with statistical guarantees are rather limited.

From a theoretical perspective, only a handful of results on testing large random graphs are known. While finding hidden cliques have been a long-standing open problem, Arias-Castro and Verzelen (2014), Verzelen and Arias-Castro (2015) for the first time provide a characterisation of the more basic problem of detecting a planted clique in an Erdős–Rényi graph. Gao and Lafferty (2017) and Lei (2016) consider generalised variants of this problem while designing tests to distinguish a stochastic block model from an Erdős–Rényi graph, or to estimate the number of communities in a stochastic block model, respectively. In a different direction, Bubeck et al. (2016) study the classical problem of testing whether a given graph corresponds to a neighbourhood graph in a high-dimensional space, or it is generated from Erdős–Rényi model. This result is in fact a specific instance of the generic problem of detecting whether a network data has a dependence structure or is unstructured (Bresler and Nagaraj (2018), Daskalakis, Dikkala and Kamath (2018), Ryabko (2017)). The first study on two-sample testing of graphs, under a relatively broad framework is by Tang et al. (2017a), where the authors test between a pair of random dot product graphs which are undirected graphs on a common set of n vertices with mutually independent edges and low-rank population adjacency matrices. A test statistic based on the difference in adjacency spectral embeddings is shown to be asymptotically consistent as $n \to \infty$ provided that the rank of the population adjacencies is fixed and known. Tang et al. (2017b) study a more general problem of comparing two random dot product graphs defined on different vertex sets, where the vertices have latent Euclidean representations. The latent representation can be recovered from the adjacency spectral embedding, and kernel two-sample testing for Euclidean data can be employed to solve the testing problem. Ghoshdastidar et al. (2017) provide a framework to formulate the graph two-sample problem with minimal structural restrictions, and show that this general framework can be used to prove minimax optimality of tests based on triangle counts and spectral properties under special cases.

In this paper, we restrict ourselves to graphs on a common vertex set of size n and sampled from an inhomogeneous Erdős–Rényi (IER) model (Bollobás, Janson and Riordan (2007)), that is, we consider undirected and unweighted random graphs where the edges occur independently, but there is no structural assumption on the population adjacency matrix. We study the problem of testing between two IER models, where m i.i.d. graphs are observed for each model. Apparently, allowing $m \geq 1$ appears to be a slight generalisation of the $m = 1$ case (Tang et al. (2017a)), but we show that in some situations, the testing problem behaves differently in the $m = 1$ and $m > 1$ cases. It is also well established that many graph learning problems have different behaviour in the case of dense and sparse graphs. This is indeed true in the context of testing for geometric structures (Bubeck et al. (2016)) and community detection (Verzelen and Arias-Castro (2015)). Bearing this in mind, we study the two-sample problem at different levels of sparsity of the graph. A formal description of the problem is presented in Section 2.

Given the above framework, one may resort to a variety of testing procedures. A classical approach involves viewing the problem as an instance of closeness testing for high-dimensional discrete distributions (Chan et al. (2014), Daskalakis, Dikkala and Kamath (2018)) and using variants of the χ^2-test or related localisation procedures. On the other hand, exploiting the independence of edges, one may even view the problem as a instance of multiple testing, and may resort to tests based on higher criticism (Donoho and Jin (2004), Donoho and Jin (2015)). A more direct approach may be to simply compare the adjacency spectral embeddings (Tang et al. (2017a)), or other network statistics (Ghoshdastidar et al. (2017)) or even the raw adjacency matrices. Sections 3–5 present a variety of testing problems, which differ in terms of the distance $d(P, Q)$, that is, how we quantify the separation between the two models. We show that some of the above principles are not useful for small m since the associated testing problems are generally unsolvable in these cases. However, in some cases, one can construct uniformly consistent tests that work with a small number of observation, even $m = 1$. Section 6 discusses the practicality of graph two-sample testing and also presents minimax separation under special cases of the IER model, such as Erdős–Rényi or stochastic block models, or under different notions of sparsity in graphs. The detailed proofs are provided in the Supplementary Material (Ghoshdastidar et al. (2020)).

2. Problem statement. In this section, we formally state the generic two-sample graph testing problem studied in this paper. We also present the minimax framework that forms the basis of our theoretical analysis.

We use the notation $\lesssim$ and $\gtrsim$ to denote the standard inequalities but ignoring absolute constants. Further, we use $\asymp$ to denote that two quantities are same up to possible difference in constant scaling. We use $\wedge$ and $\vee$ (or $\bigwedge$ and $\bigvee$) to denote minimum and maximum, respectively. We also need several standard norms and distances. For two discrete distributions, we denote the total variation distance by $\mathrm{TV}(\cdot,\cdot)$ and the symmetric Kullback–Leibler (KL) divergence by $\mathrm{SKL}(\cdot,\cdot)$. The latter is a symmetrised version of KL-divergence (Daskalakis, Dikkala and Kamath (2018)). We use the following quantities for any matrix:

(i) Frobenius norm, $\|\cdot\|_F$, is the root of sum of squares of all entries,
(ii) max norm, $\|\cdot\|_{\max}$, is largest absolute entry of the matrix,
(iii) zero norm, $\|\cdot\|_0$, is the number of nonzero entries,
(iv) operator norm, $\|\cdot\|_{\mathrm{op}}$, is the largest singular value of the matrix, and
(v) row sum norm, $\|\cdot\|_{\mathrm{row}}$, (or, the induced ∞-norm) is the maximum absolute row sum of the matrix.

2.1. *The model and the testing problem.* Throughout the paper, $V = \{1, 2, \dots, n\}$ denotes a set of n vertices, and we consider undirected graphs defined on V. Any such graph can be expressed as $G = (V, E_G)$, where E_G is the set of undirected edges. We use the symmetric matrix $A_G \in \{0,1\}^{n\times n}$ to denote the adjacency matrix of G, where $(A_G)_{ij} = 1$ if $(i,j) \in E_G$, and 0 otherwise. The class of inhomogeneous random graphs, or more precisely inhomogeneous Erdős–Rényi (IER) graphs, on V can be described as follows. Let $\mathbb{M}_n \subset [0,1]^{n\times n}$ be the set of symmetric matrices with zero diagonal, and off-diagonal entries in $[0,1]$. For any $P \in \mathbb{M}_n$, we say that G is an IER graph with population adjacency P, denoted by $G \sim \mathrm{IER}(P)$, if the adjacency matrix A_G is a symmetric random matrix such that $(A_G)_{ij} \sim \mathrm{Bernoulli}_{0-1}(P_{ij})$, and $((A_G)_{ij})_{1\le i<j\le n}$ are independent.

Let $P, Q \in \mathbb{M}_n$. Given m independent observations from each of $\mathrm{IER}(P)$ and $\mathrm{IER}(Q)$, we would like to test between the alternatives

$$\mathcal{H}_0 : P = Q \quad \text{and} \quad \mathcal{H}_1 : d(P,Q) > \rho \tag{1}$$

for some specified distance function d and a threshold $\rho \ge 0$. At this stage, note that the distribution $\mathrm{IER}(P)$ is completely characterised by the expected adjacency matrix P. Hence, $\mathcal{H}_0$ in (1) is similar both under the mean difference alternative and the general difference alternative (Ramdas et al. (2015)). Hence, one may assume d to be either a distance between the distributions $\mathrm{IER}(P)$ and $\mathrm{IER}(Q)$, or a matrix distance between P and Q. Different examples of d are considered in Sections 3–5, which result in specific instances of the testing problems.

We note that the complexity of graph inference problems is often governed by the sparsity of the graphs. To take the effect of sparsity into account, we restrict the problem to models such that $\|P\|_{\max} \vee \|Q\|_{\max} \le \delta$ for some $\delta \in (0,1]$ where δ may decay with m, n. Intuitively, we consider only graphs that are uniformly sparse, that is any edge can occur with probability at most δ. For instance, if $\delta \asymp \frac{1}{n}$, we mostly observe sparse graphs with bounded expected degrees. Such a uniform sparsity restriction is along the lines of a scalar sparsity parameters introduced in some graph estimation problems (Klopp, Tsybakov and Verzelen (2017)). More general notions of sparsity may be considered as discussed later in Section 6. Based on the above considerations, we formally state the following general framework for graph two-sample testing:

$$\mathcal{H}_0 : (P,Q) \in \Omega_0 \quad \text{vs.} \quad \mathcal{H}_1 : (P,Q) \in \Omega_1, \tag{2}$$

where

$$\Omega_0 = \{(P, Q) \in \mathbb{M}_n \times \mathbb{M}_n : P = Q, \|P\|_{\max} \leq \delta\},$$
$$\Omega_1 = \{(P, Q) \in \mathbb{M}_n \times \mathbb{M}_n : d(P, Q) > \rho, \|P\|_{\max} \vee \|Q\|_{\max} \leq \delta\}. \tag{3}$$

Note that the hypotheses are governed by the distance d, the integers m, n, and the positive scalars ρ, δ, where the last two terms may depend on m, n.

2.2. *Minimax framework.* Given the graphs $G_1, \dots, G_m \sim_{\text{iid}} \text{IER}(P)$ and $H_1, \dots, H_m \sim_{\text{iid}} \text{IER}(Q)$, a test Ψ is a binary function of $2m$ adjacency matrices, where $\Psi = 0$ when the test accepts $\mathcal{H}_0$, and $\Psi = 1$ otherwise. The maximum or worst-case risk of a test is given by

$$R(\Psi, n, m, d, \rho, \delta) = \sup_{\theta \in \Omega_0} \mathsf{P}_\theta(\Psi = 1) + \sup_{\theta \in \Omega_1} \mathsf{P}_\theta(\Psi = 0),$$

which is the sum of maximum possible Type-I and Type-II error rates incurred by the test. Here, we use θ to denote any tuple (P, Q). The minimax risk for the problem in (2)–(3) is defined as

$$R^*(n, m, d, \rho, \delta) = \inf_{\Psi} R(\Psi, n, m, d, \rho, \delta). \tag{4}$$

Our aim in this paper is to find the minimax separation ρ^* for a given problem, which is the smallest possible ρ such that $R^*(n, m, d, \rho, \delta) \leq \eta$ for some prespecified $\eta \in (0, 1)$. In the subsequent sections, we consider testing problems where the separation between P and Q is defined in terms of various distance functions. We provide bounds for ρ^* for the different testing problems in terms of the various parameters of the problem (2)–(3). Though our formal bounds are explicit in terms of η, we generally assume η to be a pre-specified constant (e.g., $\eta = 0.05$) and focus on the dependence of ρ^* on n, m and δ. Our aim is to provide upper and lower bounds for ρ^* that are same up to difference in absolute constants and functions of η.

3. Challenges of testing with small sample size. The theme of this paper is to test between two populations of sparse graphs, where the sample size m is much smaller than the number of vertices n. Our main interest is in cases where m is a small constant, or may grow very slowly with n. In this section, we show that in case of some popular distance functions, the testing problem is nearly unsolvable if m is small.

We formalise the notion of unsolvability in the following way. For any instance of the two-sample testing problem (2)–(3), there is a trivial upper bound for ρ^* which is the maximal possible value that can be attained by $d(P, Q)$ (diameter with respect to d). As an example, if $d(P, Q) = \text{TV}(\text{IER}(P), \text{IER}(Q))$, then $\rho^* \leq 1$ trivially. Similarly, if $d(P, Q) = \|P - Q\|_F$, a trivial upper bound is $\rho^* \leq n\delta$. On the other hand, for small m, if there is a lower bound $\rho_\ell \leq \rho^*$ such that ρ_ℓ is equal or close to the trivial upper bound, then there exist model pairs such that $d(P, Q)$ is nearly as large as the diameter and yet cannot be distinguished for small m. Hence, we may conclude that the problem (2)–(3) with the specific choice of d is unsolvable for small m under a worst-case (minimax) analysis.

We present the first instance of such an impossibility result for the case of total variation distance. For any two probability mass functions p and q, defined on the space of undirected n-vertex graphs,

$$\text{TV}(p, q) = \frac{1}{2} \sum_G |p(G) - q(G)|,$$

where the summation is over all unweighted undirected graphs on n vertices. We present the following minimax rate in the small sample regime.

PROPOSITION 3.1 (ρ^* for total variation distance). *Consider the problem in* (2)–(3) *with* $d(P,Q) = \mathrm{TV}(\mathrm{IER}(P), \mathrm{IER}(Q))$ *and* $\delta \in (0,1)$. *Let* $\eta \in (0,1)$ *be the allowable risk. For any* $\delta \ge C' \frac{\ln n}{n^2}$,

$$1 - \frac{1}{n} \le \rho^* \le 1 \quad \textit{for } m \le C\sqrt{\ln(1 + 4(1-\eta)^2)}\frac{n}{\ln n},$$

where $C \in (0,1)$ *and* $C' > 1$ *are absolute constants.*

In particular, for large n and $m \lesssim \frac{n}{\ln n}$, *we have* $\rho^* \approx 1$.

PROOF SKETCH. The main idea is to find an appropriate choice of (P,Q) such that $\mathrm{TV}(\mathrm{IER}(P), \mathrm{IER}(Q)) \ge 1 - \frac{1}{n}$, and yet they cannot be distinguished using $m \lesssim \frac{n}{\ln n}$ samples. For this, we use standard approaches to derive minimax lower bounds (Baraud (2002)). This is described in the present context of two-sample testing of IER graphs in Ghoshdastidar et al. (2020).

In particular for the present proof, the choice of P, Q is the following. Under $\mathcal{H}_0$, we set $P = Q$ such that the model corresponds to Erdős–Rényi (ER) model with edge probability $\frac{\delta}{2}$. Under $\mathcal{H}_1$, we keep P as before whereas each entry of Q is chosen independent and uniformly from $\{\frac{\delta}{2} - \gamma, \frac{\delta}{2} + \gamma\}$. An appropriate choice of $\gamma \le \frac{\delta}{2}$ leads to the result. □

The stated lower bound shows that at least $m \gtrsim \frac{n}{\ln n}$ samples are needed to test for separation in total variation distance, which is beyond the small sample regime that we are interested in. In fact, in the case of constant m, one can improve the stated bound as $1 - e^{-C''n} \le \rho^* \le 1$, where C'' is a constant that depends on m.

An impossibility result also holds in the case of symmetric KL-divergence, which has been effectively used for high-dimensional discrete distributions, particularly Ising models (Daskalakis, Dikkala and Kamath (2018)). For any two probability mass functions p and q, defined on the space of undirected n-vertex graphs, the symmetric KL-divergence is given by

$$\mathrm{SKL}(p,q) = \sum_G p(G)\ln\left(\frac{p(G)}{q(G)}\right) + q(G)\ln\left(\frac{q(G)}{p(G)}\right),$$

where the summation is over all unweighted undirected graphs on n vertices. Note that the above distance is unbounded even for finite n since $\mathrm{SKL}(p,q) = \infty$ if there exists G_0 such that $p(G_0) = 0$ but $q(G_0) \ne 0$. We present the following result that demonstrates the impossibility of testing with respect to the symmetric KL-divergence when sample size m is small.

PROPOSITION 3.2 (ρ^* for symmetric KL-divergence). *Let* $\eta \in (0,1)$ *and consider the problem in* (2)–(3) *with* $d(P,Q) = \mathrm{SKL}(\mathrm{IER}(P), \mathrm{IER}(Q))$ *and any* $\delta \in (0,1)$. *Then*

$$\rho^* = \infty \quad \textit{for } m \le \frac{2}{\delta}\ln\big((1-\eta)n\big).$$

PROOF SKETCH. The basic technique for the proof is similar to Proposition 3.1, but we choose P and Q such that $\mathrm{SKL}(\mathrm{IER}(P), \mathrm{IER}(Q)) = \infty$, and yet the models are indistinguishable for small m.

To be precise, under $\mathcal{H}_0$ we set $P = Q$ corresponding to an ER model. Under $\mathcal{H}_1$, we set Q to be the same as P except for a randomly chosen entry, for which $Q_{ij} = 0 \ne P_{ij}$. This implies that $\mathrm{IER}(P)$ and $\mathrm{IER}(Q)$ do not have a common support, which leads to $\mathrm{SKL}(\mathrm{IER}(P), \mathrm{IER}(Q)) = \infty$. However, if the sample size is small (small m) or the graphs are sparse (small δ), then the models cannot be distinguished. □

The above result shows that testing for separation in symmetric KL-divergence is impossible for $m \lesssim \ln n$ sample even when the graphs are dense ($\delta \asymp 1$). However, the situation is worse in the case of sparse graphs. If $\delta \asymp \frac{1}{n}$, then at least $m \gtrsim n \ln n$ samples are necessary, which is worse than the condition for total variation distance.

Both Propositions 3.1 and 3.2 suggest that achieving a small sample complexity (small m) could be difficult under general difference alternatives, that is, if d corresponds to distributional distances. Hence, subsequent discussions focus only on matrix distances. However, even in this case, the two-sample problem is not necessarily easily solvable for all distances or dissimilarities.

PROPOSITION 3.3 (ρ^* for zero norm/effect rarity). *Let $\eta \in (0, 1)$ and consider the problem in (2)–(3) with $d(P, Q) = \|P - Q\|_0$ and any $\delta \in (0, 1)$. Then*

$$\rho^* = n(n-1) \quad \textit{for all } m < \infty.$$

PROOF SKETCH. The proof is straightforward since the entries P and Q can be arbitrarily close but still be unequal. Hence, the models may not be distinguishable though $\|P - Q\|_0 = n(n-1)$, which is the trivial upper bound. □

Proposition 3.3 may be viewed as a trivial extremity of the rare/weak effect studied in the context of multiple testing (Donoho and Jin (2015)). To put it simply, here we view the problem as testing $P_{ij} = Q_{ij}$ or $P_{ij} \neq Q_{ij}$ for every $i < j$, and the edge independence in IER graphs leads to a problem of multiple independent comparisons. Proposition 3.3 states that if $\min_{P_{ij} \neq Q_{ij}} |P_{ij} - Q_{ij}|$ is arbitrarily small, that is, the individual effects are arbitrarily weak, then they cannot be detected even when the effects are dense $\|P - Q\|_0 = n(n-1)$. A more detailed analysis of the rare/weak effect in the sparse Bernoulli setting may be done by imposing a threshold $\min_{P_{ij} \neq Q_{ij}} |P_{ij} - Q_{ij}|$ that characterises weakness of the effect. We do not discuss further on this effect, and instead, we proceed to other instances of (2)–(3), where the problem can be solved for small m.

4. Testing for separation in Frobenius norm. The previous section focused on impossibility results, where ρ^* is typically large (close to trivial upper bound) when m is small. In this section and the next one, we study two instances of the problem (2)–(3) where tests can be constructed even for small m. Formally, we show that the lower bound for ρ^* can be much smaller than the trivial upper bound, and subsequently, we propose two-sample tests to derive nearly matching upper bounds for ρ^*.

We first quantify the separation in terms of Frobenius norm, that is, $d(P, Q) = \|P - Q\|_F$. This is equivalent to viewing the adjacencies as $\binom{n}{2}$-dimensional Bernoulli vectors, and using two-sample test for high-dimensional vectors—a well-studied problem in the Gaussian case (Chen and Qin (2010)). We state the following bounds for the minimax separation ρ^*.

THEOREM 4.1 (ρ^* for Frobenius norm separation). *Consider the two-sample problem (2)–(3) with $d(P, Q) = \|P - Q\|_F$, any $\delta \in (0, 1)$ and any $\eta \in (0, 1)$. There exist absolute constants $C_1, C_2 \geq 1$ such that*:

1. $\frac{n\delta}{4} \leq \rho^* \leq n\delta$ *for* $m = 1$,
2. $(\frac{1}{4} \wedge \sqrt{\frac{\eta^2 \ell_\eta}{8C_1}}) n\delta \leq \rho^* \leq n\delta$ *for* $m > 1$ *and* $\delta \leq \frac{C_1}{\eta^2 mn}$, *and*
3. $\sqrt{\frac{\ell_\eta}{8} \frac{n\delta}{m}} \leq \rho^* \leq \sqrt{\frac{C_2}{\eta} \frac{n\delta}{m}}$ *for* $m > 1$ *and* $\delta \geq \frac{C_1}{\eta^2 mn}$,

where $\ell_\eta = \sqrt{\ln(1 + 4(1-\eta)^2)}$. *Hence, assuming the allowable risk* η *is fixed, we have* $\rho^* \asymp n\delta$ *for* $m = 1$ *and* $\rho^* \asymp n\delta \wedge \sqrt{\frac{n\delta}{m}}$ *for* $m \geq 2$.

Theorem 4.1 provides a clear characterisation of the minimax separation ρ^* (up to factors of η) when the distance between models is in terms of Frobenius norm. The second and third statements deal with the case of $m > 1$. In the ultra-sparse regime, that is $\delta \lesssim \frac{1}{mn}$, one observes a total of only $O(n)$ edges from the entire population of $2m$ graphs generated from either models. This information is insufficient for testing equality of models, and hence, it is not surprising that $\rho^* \asymp n\delta$, which is the trivial upper bound. On the other hand, when $\delta \gtrsim \frac{1}{mn}$, we find a nontrivial separation rate indicating that the problem is solvable in this case.

The surprising finding of Theorem 4.1 is that $\rho^* \asymp n\delta$ for $m = 1$, which informally means that the problem is not solvable when one observes only $m = 1$ sample from each model. This result is significant since it shows that the problem of testing for separation in Frobenius norm is unsuitable in the setting of comparing between two large networks, for instance, the case of testing between two omics networks.

4.1. *Proof of Theorem* 4.1. We provide an outline of the proof of the above result highlighting the key technical lemmas. We sketch their proofs here, and the detailed proofs can be found in Ghoshdastidar et al. (2020). To prove the lower bounds, we have the following result.

LEMMA 4.2 (Necessary conditions for detecting Frobenius norm separation). *For the testing problem* (2)–(3) *with* $d(P, Q) = \|P - Q\|_F$ *and for any* $\eta \in (0, 1)$, *the minimax risk* (4) *is at least* η *if either of the following conditions hold*:

$$\text{(i)} \quad \rho < \frac{n\delta}{4} \bigwedge \sqrt{\frac{\ell_\eta}{8}\frac{n\delta}{m}}, \quad \textit{or} \quad \text{(ii)} \quad m = 1, \qquad \rho < \frac{n\delta}{4}.$$

PROOF SKETCH. The proof follows the basic approach of Proposition 3.1, and also uses same choice of P, Q. We set $P = Q$ corresponding to ER model with edge probability $\frac{\delta}{2}$ under $\mathcal{H}_0$. For $\mathcal{H}_1$, we set the same P, but each entry of Q is chosen independent and uniformly from $\{\frac{\delta}{2} - \gamma, \frac{\delta}{2} + \gamma\}$. One can easily see that Q is chosen uniformly from a set of $2^{n(n-1)/2}$ matrices, but for each choice of Q, we have $\|P - Q\|_F \approx n\gamma$. Hence, a choice of $\gamma \in (\frac{\rho}{n}, \frac{\delta}{2}]$ implies that the pair of (P, Q) for every choice of Q lies in Ω_1.

Subsequently, we use the techniques of Baraud (2002) to show that for $m \geq 2$ and $\delta \gtrsim \frac{1}{mn}$, there is an appropriate choice $\gamma < \frac{\delta}{2}$ for which the random choice of $(P, Q) \in \Omega_1$ cannot be distinguished from the null case. If $\delta \lesssim \frac{1}{mn}$, then the same situation occurs even for the choice $\gamma = \frac{\delta}{2}$. Finally, for $m = 1$, we observe that the same proof leads to the conclusion that the random choice of $(P, Q) \in \Omega_1$ is indistinguishable from the null case for any choice of $\gamma \leq \frac{\delta}{2}$. In particular, $\gamma = \frac{\delta}{2}$ leads to the claim in (ii).

We elaborate on the distinction between the cases $m = 1$ and $m > 1$. Consider the former case of $m = 1$ under $\mathcal{H}_1$, where we have $G_1 \sim \mathrm{IER}(P) = \mathrm{ER}(\frac{\delta}{2})$ and $H_1 \sim \mathrm{IER}(Q)$ with Q being chosen randomly as described above. Due the uniform choice of Q, one can easily verify that the probability of each edge in H_1 is $\frac{\delta}{2}$, which is same as that of G_1. Hence, although the two graphs are sampled from different generative models with $\|P - Q\|_F > \rho$, they are essentially similar due to the random choice of Q. On the other hand, let $m = 2$, $G_1, G_2 \sim_{\mathrm{iid}} \mathrm{ER}(\frac{\delta}{2})$ and $H_1, H_2 \sim_{\mathrm{iid}} \mathrm{IER}(Q)$ with Q being random as before. Although H_1, H_2 are independent conditioned on the choice of Q, the two graphs are mutually dependent without the knowledge of Q and hence, the population $\{H_1, H_2\}$ does not have the same distribution as $\{G_1, G_2\}$. $\square$

We continue with the proof of Theorem 4.1. The lower bounds in the second and third statements of Theorem 4.1 follow from condition (i) above by accounting for the conditions on δ and noting $C_1 \geq 1$ and $\ell_\eta \leq \sqrt{\ln 5}$. For the upper bounds in first two statements, note that $\rho^* \leq n\delta$ trivially holds since $\|P - Q\|_F \leq n(\|P\|_{\max} \vee \|Q\|_{\max})$. To derive the upper bound for the third case, we construct the following two-sample test. Let $A_{G_1}, \dots, A_{G_m}$ and $A_{H_1}, \dots, A_{H_m}$ be the adjacency matrices of the $2m$ graphs. We define

$$\widehat{\mu} = \sum_{\substack{i,j=1 \\ i<j}}^{n} \Bigg(\sum_{k \leq m/2} (A_{G_k})_{ij} - (A_{H_k})_{ij} \Bigg) \Bigg(\sum_{k > m/2} (A_{G_k})_{ij} - (A_{H_k})_{ij} \Bigg), \tag{5}$$

$$\widehat{\sigma} = \sqrt{\sum_{\substack{i,j=1 \\ i<j}}^{n} \Bigg(\sum_{k \leq m/2} (A_{G_k})_{ij} + (A_{H_k})_{ij} \Bigg) \Bigg(\sum_{k > m/2} (A_{G_k})_{ij} + (A_{H_k})_{ij} \Bigg)}, \tag{6}$$

and consider the test

$$\Psi_F = \mathbf{1}\left\{ \frac{\widehat{\mu}}{\widehat{\sigma}} > \frac{t_1}{\sqrt{\eta}} \right\} \cdot \mathbf{1}\left\{ \widehat{\sigma} > \frac{t_2}{\eta^{3/2}} \right\} \tag{7}$$

for some positive constants t_1, t_2, where $\mathbf{1}\{\cdot\}$ is the indicator function. We state the following guarantee for Ψ_F.

LEMMA 4.3 (Sufficient conditions for detecting Frobenius norm separation). *Consider the testing problem with $d(P, Q) = \|P - Q\|_F$ and the test Ψ_F in (7). There exist absolute constants t_1, t_2, C and C' such that for any $\eta \in (0, 1)$ and $\delta \in (0, 1)$, if*

$$m \geq 2 \quad \textit{and} \quad \rho \geq \sqrt{\frac{C}{\eta} \frac{n\delta}{m}} \bigvee \frac{C'}{\eta^{3/2} m}, \tag{8}$$

then $R(\Psi_F, n, m, d, \rho, \delta) \leq \eta$.

PROOF SKETCH. The proof is based on concentration statements for $\widehat{\mu}$ and $\widehat{\sigma}$ derived via Chebyshev's inequality using

$$\mu = \mathsf{E}[\widehat{\mu}] = \frac{m^2}{8} \|P - Q\|_F^2, \qquad \sigma^2 = \mathsf{E}[\widehat{\sigma}^2] = \frac{m^2}{8} \|P + Q\|_F^2$$

and bounds for $\mathsf{Var}[\widehat{\mu}]$ and $\mathsf{Var}[\widehat{\sigma}^2]$ in terms of $\|P - Q\|_F$ and $\|P + Q\|_F$.

For the type-I-error, that is, under $\mathcal{H}_0$, these concentration bounds lead to large enough choices for t_1 and t_2 such that at least one of the events

$$\left\{ \frac{\widehat{\mu}}{\widehat{\sigma}} > \frac{t_1}{\sqrt{\eta}} \right\} \quad \text{and} \quad \left\{ \widehat{\sigma} > \frac{t_2}{\eta^{3/2}} \right\}$$

has small probability, which bounds the probability of the event $\{\Psi_F = 1\}$. On the other hand, by construction, in order to control the type-II-error rate we want to ensure that both the events

$$\left\{ \frac{\widehat{\mu}}{\widehat{\sigma}} \leq \frac{t_1}{\sqrt{\eta}} \right\} \quad \text{and} \quad \left\{ \widehat{\sigma} \leq \frac{t_2}{\eta^{3/2}} \right\}$$

have small probability under $\mathcal{H}_1$. Now, the requirement in (8) on ρ guarantees that

$$\mu \gtrsim \sigma \gtrsim \frac{1}{\sqrt{8}\eta^{3/2}},$$

which allows us to rewrite these two events as large deviation statements of $\widehat{\mu}$ and $\widehat{\sigma}^2$ from their means as required. $\square$

To conclude the proof of Theorem 4.1, observe that if $\delta \geq \frac{C'^2}{C\eta^2 mn}$, then the $\sqrt{\frac{n\delta}{m}}$ term in the above result dominates and we obtain the upper bound in the third statement of Theorem 4.1 with $C_1 = \frac{C'^2}{C}$. Hence the theorem holds.

4.2. *Further remarks on Theorem 4.1.* As part of the proof of Theorem 4.1, we propose a test in (7) which provides the nontrivial upper bound in the third statement of Theorem 4.1. Though this bound matches the corresponding lower bound up to factors of η, the difference is rather large with respect to η. This is an artefact of the proof of Lemma 4.3 which is based on Chebyshev's inequality, and its effect can also be seen in the two thresholds $\frac{t_1}{\sqrt{\eta}}$ and $\frac{t_2}{\eta^{3/2}}$ defined in (7), which can be very high for small η limiting the practical usefulness of the test. Below, we show that this can be improved using more refined concentration inequalities, but provides a sufficient condition that is weaker by a factor of $\ln n$.

PROPOSITION 4.4 (Improving dependence on η). *Consider the two sample problem of Theorem 4.1 and assume $m \geq 2$. Define the test*

$$\Psi'_F = \mathbf{1}\left\{\frac{\widehat{\mu}}{\widehat{\sigma}} > t_1 \ln\left(\frac{2}{\eta}\right) \ln\left(\frac{n}{\eta}\right)\right\} \cdot \mathbf{1}\left\{\widehat{\sigma} > t_2 \ln^2\left(\frac{2}{\eta}\right) \ln\left(\frac{n}{\eta}\right)\right\}, \tag{9}$$

where $\widehat{\mu}, \widehat{\sigma}$ are as in (5)–(6). There exist constants t_1, t_2, C and C' such that for any $\eta \in (0, 1)$ and $\delta \in (0, 1)$, the test in (9) has a risk at most η whenever

$$\rho \geq C \ln\left(\frac{2}{\eta}\right) \sqrt{\frac{n\delta}{m} \ln\left(\frac{n}{\eta}\right)} \bigvee \frac{C'}{m} \ln^2\left(\frac{2}{\eta}\right) \ln\left(\frac{n}{\eta}\right). \tag{10}$$

PROOF SKETCH. The proof is very close in spirit to that of Lemma 4.3 above, but it is based on a much more involved concentration statement than Chebyshev's inequality (stated in the Supplementary Material). As a result, the test is based on the same test statistic $\frac{\widehat{\mu}}{\widehat{\sigma}}$ and differs from (7) only in the choice of thresholds.

More specifically, due to the fact that $\widehat{\mu}$ and $\widehat{\sigma}^2$ are sums of products of sums with strong independence properties, we can derive concentration inequalities with logarithmic dependence on η for them by repeated application of Bernstein's inequality. However, this comes with the additional $\ln n$ factor which can be seen in equation (10) above. $\square$

A key feature of both tests is that they are adaptive, that is, they do not require specification of the sparsity parameter δ. We highlight the importance of this property in the following remark.

REMARK 4.5 (Adaptivity of proposed tests). The testing problem in (2)–(3) is defined with respect to the sparsity parameter δ, which in turn governs the minimax separation rate ρ^*. It is not hard to convince one that for any P, Q, it is impossible to estimate $\|P\|_{\max} \vee \|Q\|_{\max}$ from few observations (small m), and setting $\delta = 1$ is clearly suboptimal for sparse graph models. Hence, it is desirable to construct tests that do not require knowledge of δ, and both tests in (7) and (9) are adaptive in this sense. Adaptivity of these tests are achieved by estimating $\|P + Q\|_F$, which is a lower bound for $2n\delta$.

5. Testing for separation in operator norm. In this section, we study the two sample testing problem where $d(P, Q) = \|P - Q\|_{\mathrm{op}}$, and provide bounds on the minimax separation ρ^* for all $m \ge 1$. An interesting finding of this section is that one can obtain a nontrivial minimax separation rate even for $m = 1$, that is, the problem is indeed solvable even with a single observation from each model. Our main result is the following.

THEOREM 5.1 (ρ^* for operator norm separation). *Consider the two-sample problem* (2)–(3) *with* $d(P, Q) = \|P - Q\|_{\mathrm{op}}$, *and any* $m \ge 1$, $\delta \in (0, 1)$. *Let* $\eta \in (0, 1)$ *and* $\ell'_\eta = \frac{\ell_\eta}{\sqrt{8}} \wedge \frac{1}{16}$. *There exist constants* $C, C' \ge 1$ *such that*:

1. $\frac{n\delta}{4} \le \rho^* \le n\delta$ *for* $\delta \le \frac{\ell'^2_\eta}{16mn}$, *and*
2. $\ell'_\eta\sqrt{\frac{n\delta}{m}} \le \rho^* \le \sqrt{\frac{n\delta}{m}}(C\sqrt{\ln(\frac{n}{\eta})} \bigvee \frac{4C'}{\ell'_\eta}\ln(\frac{n}{\eta}))$ *otherwise.*

The theorem shows that the problem is not solvable in the ultra-sparse regime, that is, $\delta \lesssim \frac{1}{mn}$. However, beyond this regime there is a nontrivial separation rate, which Theorem 5.1 finds up to a factor of $\ln n$. It is natural to ask whether the additional logarithmic factor is necessary. Later in this section, we refine Theorem 5.1 to remove the $\ln n$ term in the upper bound (see Corollary 5.5). This is achieved by using a nonadaptive test, which has prior knowledge of δ (see Proportion 5.4).

5.1. *Proof of Theorem* 5.1. The lower bounds in the theorem are due to the following necessary condition.

LEMMA 5.2 (Necessary condition for detecting operator norm separation). *For the testing problem* (2)–(3) *with* $d(P, Q) = \|P - Q\|_{\mathrm{op}}$, $\delta \in (0, 1)$ *and* $m \ge 1$, *and for any* $\eta \in (0, 1)$, *the minimax risk* (4) *is at least* η *if*

$$\rho < \frac{n\delta}{4} \bigwedge \ell'_\eta\sqrt{\frac{n\delta}{m}}.$$

PROOF SKETCH. The proof follows the technique of Baraud (2002) to derive minimax lower bounds as used in the previous results in this paper. Hence, we mainly focus on the choice of P, Q used in the present proof. We set $P = Q$ corresponding to ER model with edge probability $\frac{\delta}{2}$ under $\mathcal{H}_0$. For $\mathcal{H}_1$, we set the same P, but Q is randomly chosen in the following way. We partition the vertices randomly into two groups, and set $Q_{ij} = \frac{\delta}{2} + \gamma$ for i, j belonging to the same group, and $Q_{ij} = \frac{\delta}{2} - \gamma$ otherwise. The random choice of Q is due to randomly sampling one of 2^{n-1} possible splits of the vertex set.

One can see that for each choice of Q, we have $\|P - Q\|_{\mathrm{op}} = \gamma(n-1)$. Hence, a choice of $\gamma \in (\frac{\rho}{n-1}, \frac{\delta}{2}]$ implies that the pair of (P, Q) for every choice of Q lies in Ω_1. Now similar to the proof of Lemma 4.2, we show for any $m \ge 1$ and $\delta \gtrsim \frac{1}{mn}$, there is an appropriate choice $\gamma < \frac{\delta}{2}$ for which the random choice of $(P, Q) \in \Omega_1$ cannot be distinguished from the null case. If $\delta \lesssim \frac{1}{mn}$, then the same situation occurs even for the choice $\gamma = \frac{\delta}{2}$, which leads to the claim. □

We now prove the upper bounds in Theorem 5.1. For this, we note that $\|P - Q\|_{\mathrm{op}} \le \|P - Q\|_{\mathrm{row}} \le n\delta$ is the trivial upper bound for ρ^*. To prove the nontrivial upper bound, we consider the following test:

$$\Psi_{\mathrm{op}} = \mathbf{1}\left\{\frac{\|S^-\|_{\mathrm{op}}}{\sqrt{\|S^+\|_{\mathrm{row}}}} > t_1\sqrt{\ln\left(\frac{n}{\eta}\right)}\right\} \cdot \mathbf{1}\left\{\|S^+\|_{\mathrm{row}} > t_2 \ln\left(\frac{n}{\eta}\right)\right\}, \tag{11}$$

where

$$S^- = \sum_{k=1}^{m} A_{G_k} - A_{H_k} \quad \text{and} \quad S^+ = \sum_{k=1}^{m} A_{G_k} + A_{H_k}. \tag{12}$$

We have the following sufficient condition for the two-sample test Ψ_{op}.

LEMMA 5.3 (Sufficient condition for detecting operator norm separation). *Consider the testing problem with $d(P, Q) = \|P - Q\|_{\mathrm{op}}$ and the test Ψ_{op} in (11). There exist absolute constants t_1, t_2, C and C' such that for any $\eta \in (0, 1)$, $\delta \in (0, 1)$ and $m \geq 1$, the risk $R(\Psi_{\mathrm{op}}, n, m, d, \rho, \delta) \leq \eta$ if*

$$\rho \geq C\sqrt{\frac{n\delta}{m}\ln\left(\frac{n}{\eta}\right)} \bigvee \frac{C'}{m}\ln\left(\frac{n}{\eta}\right).$$

PROOF SKETCH. The proof mainly controls the type-I and type-II error rates for Ψ_{op}. We achieve this through the matrix Bernstein inequality (Oliveira (2008), Tropp (2012)), which in the present case guarantees that

$$\|S^- - m(P - Q)\|_{\mathrm{op}} \lesssim \sqrt{m\|P + Q\|_{\mathrm{row}} \ln n}$$

with high probability if $\|P + Q\|_{\mathrm{row}} \gtrsim \frac{\ln n}{m}$. We also use a Bernstein-type concentration result to ensure that $\|S^+\|_{\mathrm{row}} \asymp m\|P + Q\|_{\mathrm{row}}$ for $\|P + Q\|_{\mathrm{row}} \gtrsim \frac{\ln n}{m}$, and $\|S^+\|_{\mathrm{row}} \lesssim \ln n$ otherwise.

We bound the type-I error by noting that under $\mathcal{H}_0$, that is $P = Q$, the above concentration results imply that with high probability:

- $\|S^-\|_{\mathrm{op}} \lesssim \sqrt{\|S^+\|_{\mathrm{row}} \ln n}$ for $\|P + Q\|_{\mathrm{row}} \gtrsim \frac{\ln n}{m}$, and
- $\|S^+\|_{\mathrm{row}} \lesssim \ln n$ for $\|P + Q\|_{\mathrm{row}} \lesssim \frac{\ln n}{m}$.

Hence, we have $\Psi_{\mathrm{op}} = 0$ with high probability for a suitable choices of t_1, t_2.

On the other hand, we control the type-II error in the following way. Under $\mathcal{H}_1$, we have $\|P - Q\|_{\mathrm{op}} > \rho$ with $\rho \geq \frac{C'}{m}\ln(\frac{n}{\eta})$, which implies that

$$\|P + Q\|_{\mathrm{row}} \geq \|P - Q\|_{\mathrm{op}} \gtrsim \frac{\ln n}{m}.$$

So the second indicator in Ψ_{op} is guaranteed to be 1 for C' large enough. This condition also allows us to use the matrix Bernstein inequality which, along with the reverse triangle inequality, gives

$$\begin{aligned}\|S^-\|_{\mathrm{op}} &\geq m\|P - Q\|_{\mathrm{op}} - \|S^- - m(P - Q)\|_{\mathrm{op}} \\ &\gtrsim m\rho - \sqrt{m\|P + Q\|_{\mathrm{row}} \ln n}.\end{aligned}$$

We use the stated assumption $\rho \geq C\sqrt{\frac{n\delta}{m}\ln(\frac{n}{\eta})}$ and the fact $\|S^+\|_{\mathrm{row}} \asymp m\|P + Q\|_{\mathrm{row}}$ to prove that the first indicator in Ψ_{op} is also true with high probability for large enough C. This bounds the type-II error rate. □

To conclude the proof of Theorem 5.1, we note that the upper bound in the second statement is obtained by observing that $\frac{C'}{m} \leq \frac{4C'}{\ell'_\eta}\sqrt{\frac{n\delta}{m}}$ for $\delta \geq \frac{\ell'^2_\eta}{16mn}$. We note that the above analysis of the test Ψ_{op} is based on the matrix Bernstein inequality (Tropp (2012)), which makes the upper bound worse factor of $\ln n$. One may alternatively use other concentration results based on refinements of the trace method (Bandeira and van Handel (2016), Lu and Peng (2013)) to obtain slightly different sufficient conditions in Lemma 5.3. For instance, use of Bandeira and van Handel ((2016), Corollary 3.9) does provide a rate of $\sqrt{\frac{n\delta}{m}}$, but such bounds hold only if $\delta \gtrsim \frac{\mathrm{polylog}(n)}{n}$.

5.2. *An optimal nonadaptive test.* One of the objectives of this work is to determine whether it is possible to test between two large graphs, that is, the case of $m = 1$. Theorem 5.1 provides an affirmative answer to this question, but our proposed test Ψ_{op} (11) is not optimal since the sufficient condition on ρ is worse than the necessary condition by a factor of $\ln n$. As we note above, this is a consequence of our use of matrix Bernstein inequality in the proof of Lemma 5.3, and leads to the question whether one can improve the result by using more sharp concentration techniques known in the case of random graphs. We show that this is indeed true, at least for $m = 1$, and can be shown using concentration of trimmed or regularised adjacency matrices (Le, Levina and Vershynin (2017)).

Assume $m = 1$, and let $G \sim \mathrm{IER}(P)$ and $H \sim \mathrm{IER}(Q)$ be the two random graphs. Also assume that δ is specified. For some constant $c \geq 6$, let A'_G be the adjacency matrix of the graph obtained by deleting all edges in G that are incident on vertices with degree larger than $cn\delta \ln(\frac{2}{\eta})$. Similarly, we obtain A'_H from H. Define a test as

$$\Psi'_{\mathrm{op}} = \mathbf{1}\left\{ \|A'_G - A'_H\|_{\mathrm{op}} > t\sqrt{n\delta}\ln^2\left(\frac{2}{\eta}\right)\right\} \tag{13}$$

for some constant $t > 0$. We have the following guarantee for Ψ'_{op}.

PROPOSITION 5.4 (Optimality for $m = 1$). *Consider the testing problem with $d(P, Q) = \|P - Q\|_{\mathrm{op}}$, $m = 1$ and $\delta > \frac{10}{n}$. There exist absolute constants c, t such that for any $\eta \in (0, 1)$, the maximum risk of Ψ'_{op} is at most η if*

$$\rho \geq 2t\sqrt{n\delta}\ln^2\left(\frac{2}{\eta}\right).$$

Hence, assuming η is fixed, $\rho^* \asymp \sqrt{n\delta}$ for $\delta > \frac{10}{n}$ whereas $\rho^* \asymp n\delta$ for $\delta \leq \frac{10}{n}$.

PROOF SKETCH. The proof differs from that of Lemma 5.3 in the use of a different matrix concentration inequality. We rely on the concentration of the trimmed adjacency matrix (Le, Levina and Vershynin (2017)). In the present context, the result states that for $\delta > \frac{10}{n}$ and $G \sim \mathrm{IER}(P)$ with $\|P\|_{\max} \leq \delta$,

$$\|A'_G - P\|_{\mathrm{op}} \leq C\sqrt{n\delta}\ln^2\left(\frac{2}{\eta}\right)$$

with probability $1 - \frac{\eta}{4}$ for some absolute constant $C > 0$.

For $P = Q$, we have with probability $1 - \frac{\eta}{2}$,

$$\|A'_G - A'_H\|_{\mathrm{op}} \leq \|A'_G - P\|_{\mathrm{op}} + \|A'_H - Q\|_{\mathrm{op}} \leq 2C\sqrt{n\delta}\ln^2\left(\frac{2}{\eta}\right),$$

and hence, setting $t = 2C$ ensures that the type-I error rate is smaller than $\frac{\eta}{2}$. Similarly, it follows that under $\mathcal{H}_1$,

$$\begin{aligned}\|A'_G - A'_H\|_{\mathrm{op}} &\geq \|P - Q\|_{\mathrm{op}} - \|A'_G - P\|_{\mathrm{op}} + \|A'_H - Q\|_{\mathrm{op}} \\ &\geq \rho - 2C\sqrt{n\delta}\ln^2\left(\frac{2}{\eta}\right),\end{aligned}$$

with probability $1 - \frac{\eta}{2}$. Hence, $\|A'_G - A'_H\|_{\text{op}}$ exceeds the threshold for $\rho > 4C\sqrt{n\delta}\ln^2(\frac{2}{\eta})$, and we have a similar bound on the type-II error rate. □

While Ψ'_{op} indeed achieves optimality, it comes at the cost of prior knowledge of δ. As we note in Remark 4.5, this is an unreasonable restriction and hence, the suboptimal test Ψ_{op} (11) may be preferable due to its adaptivity. We do not know if an adaptive optimal operator norm based test can be constructed since existing concentration bounds for sparse IER graphs are typically in terms of the largest edge probability, which is hard to estimate.

5.3. *Improving the bounds in Theorem 5.1.* As discussed above, the nonadaptive test Ψ'_{op} helps in improving the upper bound on ρ^* for $m = 1$. We have not studied whether Ψ'_{op} (13) and Proposition 5.4 extend to the case of $m > 1$. However, the following corollary shows that the results of Sections 4 and 5 can be combined to remove the undesirable $\ln n$ factor in the upper bound of ρ^* in Theorem 5.1 for any $m \geq 1$. For convenience, let the allowable risk η be a constant, which we ignore in the statement of the result.

COROLLARY 5.5 (Improved bounds on ρ^* for operator norm separation). *Consider the two-sample problem* (2)–(3) *with $d(P, Q) = \|P - Q\|_{\text{op}}$, and any $m \geq 1$, $\delta \in (0, 1)$. There exists a constant $C > 0$ such that*

$$\rho^* \asymp n\delta \quad \textit{for } \delta \leq \frac{C}{mn}, \quad \textit{and} \quad \rho^* \asymp \sqrt{\frac{n\delta}{m}} \quad \textit{otherwise}.$$

PROOF. The lower bounds on ρ^* follow from Theorem 5.1, and hence it remains to prove that

$$\rho^* \lesssim n\delta \bigwedge \sqrt{\frac{n\delta}{m}} \quad \text{for all } m \geq 1.$$

We claim that this is a direct consequence of the upper bounds in Proposition 5.4 and Theorem 4.1. To see this, consider the following test:

$$\Psi = \Psi'_{\text{op}} \quad \text{for } m = 1, \quad \text{and} \quad \Psi = \Psi_F \quad \text{for } m \geq 2.$$

For $m = 1$, Proposition 5.4 leads to the above stated upper bound.

For $m \geq 2$, note that given δ and ρ, the two-sample problems under Frobenius and operator norm separation have the same null set Ω_0 (3), whereas the sets under alternative hypotheses are

$$\Omega_1^{\text{op}} = \{(P, Q) : \|P - Q\|_{\text{op}} > \rho, \|P\|_{\max} \vee \|Q\|_{\max} \leq \delta\}, \quad \text{and}$$
$$\Omega_1^F = \{(P, Q) : \|P - Q\|_F > \rho, \|P\|_{\max} \vee \|Q\|_{\max} \leq \delta\}.$$

We use the relation $\|P - Q\|_F \geq \|P - Q\|_{\text{op}}$ to observe that $\Omega_1^{\text{op}} \subset \Omega_1^F$. As a consequence, the maximum risk of Ψ_F under the two different distances can be related as

$$R(\Psi_F, n, m, \|\cdot\|_{\text{op}}, \rho, \delta) \leq R(\Psi_F, n, m, \|\cdot\|_F, \rho, \delta),$$

that is, if ρ is fixed and Ψ_F is used for the problem of testing difference in operator norm, then it achieves a smaller worst-case risk than what it achieves for the the problem of testing difference in Frobenius norm. This implies that the minimax separation for operator norm is at most the minimax separation for Frobenius norm. The claim now follows from the upper bounds in Theorem 4.1 for $m \geq 2$. □

6. Discussion. In this section, we discuss related graph two-sample problems, and comment on practical two-sample tests.

6.1. *A note on the sparsity condition.* The notion of sparsity has no formal definition in the context of random graphs, unlike sparsity in the signal detection literature. The informal definition of a sparse graph is a graph where the number of edges are not arbitrarily large compared to the number of vertices n. The sparsity condition used in (3), that is, $\|P\|_{\max} \leq \delta$, is one approach to define sparse graphs. More precisely, this condition implies that the expected number of edges is at most $n^2\delta$, and in particular, for $\delta \asymp \frac{1}{n}$, we obtain graphs where the number of edges grows linearly with n. However, one can induce sparsity through alternative restrictions on P:

$$\text{(i)} \quad \sum_{ij} P_{ij} \leq \delta_1, \qquad \text{(iii)} \quad \|P\|_{\text{row}} \leq \delta_3,$$

$$\text{(ii)} \quad \|P\|_F \leq \delta_2, \qquad \text{(iv)} \quad \|P\|_0 \leq \delta_4$$

among others. We note the δ_i's are of different order than δ used in (3). The first condition bounds the expected number of edges, while condition (iii) provides graphs with bounded expected degrees. The last restriction is along the lines of the signal detection literature since it implies that at most δ_4 entries in P are nonzero, and results in random graphs with absolutely bounded number of edges (not only in expectation).

It is natural to ask to whether our results also extend to the case where sparsity is controlled through any one of the conditions. We do not provide a complete characterisation in each case, but present two corollaries of Theorems 4.1 and 5.1 which show that some of our results easily extend to alternative notions of sparsity.

COROLLARY 6.1 (ρ^* under Frobenius norm sparsity). *Consider the problem of testing between the following two hypotheses sets*:

$$\Omega_0 = \{P = Q, \|P\|_F \leq \delta_2\}, \quad \textit{and}$$
$$\Omega_1 = \{\|P - Q\|_F > \rho, \|P\|_F \vee \|Q\|_F \leq \delta_2\}.$$

The bounds on minimax separation stated in Theorem 4.1 hold in this case with the substitution $\delta_2 = n\delta$.

PROOF SKETCH. The proof of the corollary follows immediately from that of Theorem 4.1. The substitution $\delta_2 = n\delta$ is a consequence of the relation $\|P\|_F \leq n\|P\|_{\max}$. □

We also have a result in the case of graphs with bounded expected degrees, which can be similarly derived from Theorem 5.1. We do not know if the more precise rates given in Corollary 5.5 can be extended to this setting.

COROLLARY 6.2 (ρ^* under row sum norm sparsity). *Consider the problem of testing between the following two hypotheses sets*:

$$\Omega_0 = \{P = Q, \|P\|_{\text{row}} \leq \delta_3\}, \quad \textit{and}$$
$$\Omega_1 = \{\|P - Q\|_{\text{op}} > \rho, \|P\|_{\text{row}} \vee \|Q\|_{\text{row}} \leq \delta_3\}.$$

The bounds on minimax separation stated in Theorem 5.1 hold in this case with the substitution $\delta_3 = n\delta$.

6.2. *On the practicality of proposed tests*. The theme of this paper has been to explore different separation criteria for which the graph two sample testing problem can be solved for a small population size. In the process of addressing this question, we suggest adaptive tests Ψ_F (7), Ψ'_F (9) and Ψ_{op} (11) that also turn out to be near-optimal for the problems of detecting separation in Frobenius or operator norms. However, the practical applicability of these tests have not been discussed so far.

We note that in practice, one is more interested in the testing problem $P = Q$ vs. $P \neq Q$, and hence, a more basic question that needs to be addressed is—Which separation criterion should be used? The findings of this paper suggest that for $m = 1$, operator norm separation is a possible choice, whereas other distances like total variation distance and Frobenius norm should not be considered. For $m > 1$ but small, we show that one could detect separation in Frobenius norm using Ψ_F (7) or detect separation in operator norm using Ψ_{op} (11). We compare the relative merits of both tests in terms of sample complexity in the following way.

REMARK 6.3 (Comparison between Ψ_F and Ψ_{op}). Consider P and Q such that $P \neq Q$ and $\|P\|_{\max} \vee \|Q\|_{\max} \leq \delta$. Ignoring constants and terms involving η, Lemma 4.2 shows that $m \gtrsim \frac{n\delta}{\|P-Q\|_F^2}$ samples are necessary to distinguish between the two models. On the other hand, Lemma 5.2 shows that $m \gtrsim \frac{n\delta}{\|P-Q\|_{\mathrm{op}}^2}$ samples are needed to distinguish between P, Q. Since $\|P - Q\|_F \geq \|P - Q\|_{\mathrm{op}}$, testing for Frobenius norm separation is easier than testing for separation in operator norm for $m > 1$. In other words, one may expect Ψ_F to have a smaller risk than Ψ_{op}.

However, the tests Ψ_F, Ψ'_F and Ψ_{op} require n to be very large or the graphs to be dense to achieve a small risk, and hence have limited applicability in moderate-sized problems. It is known that the practical applicability of concentration based tests can be improved by using bootstrapped variants (Gretton et al. (2012)), which approximate the null distribution by generating bootstrap samples through random mixing of the two populations. Simulations, not included in this paper, show that permutation based bootstrapping provides a reasonable rejection rate for moderate sample size ($m \geq 10$), but such bootstrapping is not effective for smaller m, for instance $m = 2$. Furthermore, the relative merit Ψ_F (or Ψ'_F) over Ψ_{op}, as suggested by Remark 6.3, could not be verified in case of the bootstrapped variants.

When $m = 1$ and the population adjacency is of low rank, Tang et al. (2017a) suggest an alternative bootstrapping principle based on estimation of the population adjacency P and then drawing bootstrap samples from estimate of P. Simulations in Ghoshdastidar and von Luxburg (2018) show that this procedure works to some extent for dense IER graphs but only when P has a small (known) rank. When the rank is unknown or, more generally, if P does not have a low rank, such bootstrapped tests are not necessarily reliable.

In the related work (Ghoshdastidar and von Luxburg (2018)), we explore alternative possibilities for constructing practical tests derived from Frobenius norm or operator norm based test statistics that work even for $m = 1, 2$. These tests use statistics similar to the ones studied in the present work, but are based on asymptotic null distributions that hold approximately or under stronger assumptions. For instance, Ghoshdastidar and von Luxburg (2018) show that under $\mathcal{H}_0$ and assumptions on the edge density of the graphs, the Frobenius norm based statistic $\frac{\widehat{\mu}}{\widehat{\sigma}}$ (see (5)–(6)) is asymptotically dominated by a standard normal random variable as $n \to \infty$. Based on this, we propose an asymptotic distribution based test that is powerful for all $m \geq 2$ and moderately sparse graphs, and is reliable even in the case of real networks. For $m = 1$, we consider the operator norm of re-scaled version of $A_{G_1} - A_{H_1}$, which approximately follows the Tracy–Widom law under $\mathcal{H}_0$ as $n \to \infty$. The practical applicability of these tests stem from the fact that they do not explicitly rely on concentration inequalities that lead to large thresholds, as in the present paper, nor do they use bootstrapping strategies, which often require large sample sizes or assumptions on the graph model. Thus, our related work provides more practically useful tests whose statistical guarantees hold either approximately or under additional assumptions.[1]

6.3. *Minimax separation under structural assumptions.* The present paper studies the two-sample problem for IER graphs, where the population adjacency matrices do not have any structural restriction. In other words, we study a hypothesis testing problem in a dimension of $\binom{n}{2}$ with a sample size of m. Under this broad framework, Theorem 4.1 shows that the minimax separation in Frobenius norm ρ_F^* is given by

$$\rho_F^* \asymp \begin{cases} n\delta & \text{for } m = 1, \text{ and} \\ n\delta \bigwedge \sqrt{\dfrac{n\delta}{m}} & \text{for } m > 1. \end{cases}$$

Similarly, Corollary 5.5 shows that the minimax separation in operator norm

$$\rho_{\mathrm{op}}^* \asymp n\delta \bigwedge \sqrt{\frac{n\delta}{m}} \quad \text{for all } m \geq 1.$$

It is natural to ask if these rates decrease if we further impose structural assumptions on the population adjacencies thereby effectively reducing the problem dimension. In this section, we provide some initial results in this direction. We impose the structural assumptions by restricting the possible values for the population adjacency matrices. Formally, we define $\widetilde{\mathbb{M}}_n \subset \mathbb{M}_n$ as the set of symmetric matrices in $[0,1]^{n\times n}$, whose diagonal entries are zero and satisfy additional structural assumptions (specified below). Subsequently, the graph two-sample problem, restricted to a special graph class, can be stated as the problem of testing between

$$\mathcal{H}_0 : (P, Q) \in \Omega_0 \cap (\widetilde{\mathbb{M}}_n \times \widetilde{\mathbb{M}}_n) \quad \text{vs} \quad \mathcal{H}_1 : (P, Q) \in \Omega_1 \cap (\widetilde{\mathbb{M}}_n \times \widetilde{\mathbb{M}}_n),$$

where Ω_0 and Ω_1 are as defined in the original problem (2)–(3).

We begin with the most simple case of Erdős–Rényi (ER) graphs, where the restricted set $\widetilde{\mathbb{M}}_n$ corresponds to the symmetric matrices whose diagonal entries are zero and all off-diagonal entries are identical. Note that each distribution is modelled by a single parameter, and hence, we have a hypothesis testing problem in one dimension. We present following result on minimax separation for Frobenius and operator norms in this setting. We simplify the statement by ignoring absolute constants including the allowable risk η, which is assumed to be fixed.

PROPOSITION 6.4 (Minimax separation for testing ER graphs). *Consider the graph two-sample problem restricted to ER graphs with specified n, m and δ. The minimax separation rates for Frobenius norm ρ_F^* and for operator norm ρ_{op}^* satisfy*

$$\rho_F^* \asymp \rho_{\mathrm{op}}^* \asymp n\delta \bigwedge \sqrt{\frac{\delta}{m}} \quad \textit{for all } m \geq 1.$$

In particular, in the sparsity regime $\delta \gtrsim \frac{1}{n^2 m}$, the minimax separation is much smaller than the corresponding rates for IER *graphs.*

[1]The implementations for the practical tests proposed in Ghoshdastidar and von Luxburg (2018) as well as bootstrapped variants of tests studied in the present paper are available at: https://github.com/gdebarghya/Network-TwoSampleTesting

PROOF SKETCH. The proof is relatively simple, and borrows ideas from the Theorem 4.1. Hence, we only provide a brief sketch of the proof.

We consider the two-sample problem with $G_1, \ldots, G_m \sim_{\text{iid}} \text{ER}(p)$ and $H_1, \ldots, H_m \sim_{\text{iid}} \text{ER}(q)$, where $p, q \le \delta$ denote the edge probabilities in either models. Note that the result can be equivalently stated as: the minimax separation between p and q is $\delta \wedge \frac{1}{n}\sqrt{\frac{\delta}{m}}$.

To derive a lower bound, we use the approach stated in the previous results and define p, q as follows. Under $\mathcal{H}_0$, we set $p = q = \frac{\delta}{2}$ whereas under $\mathcal{H}_1$, we let $p = \frac{\delta}{2}$ and q is randomly selected from $\{p - \gamma, p + \gamma\}$ for some $\gamma \le \frac{\delta}{2}$. It turns out that $\gamma \asymp \frac{\delta}{2} \wedge \frac{1}{n}\sqrt{\frac{\delta}{m}}$ is an appropriate choice, which corresponds to the claimed lower bound for minimax separation.

The upper bound is obtained by using a simple test that compares the edge densities estimated from the two population. Define a test

$$\Psi = \mathbf{1}\{|\widehat{p} - \widehat{q}| > t_1\} \cdot \mathbf{1}\{(\widehat{p} + \widehat{q}) > t_2\},$$

where $\widehat{p} = \frac{1}{m\binom{n}{2}} \sum_k \sum_{i<j} (A_{G_k})_{ij}$ and $\widehat{q} = \frac{1}{m\binom{n}{2}} \sum_k \sum_{i<j} (A_{H_k})_{ij}$ are the two edge density estimates, and t_1, t_2 are suitably defined thresholds. The proof strategy of Lemma 4.3 combined with judicious choice of thresholds provides the desired upper bound claim in the result. $\square$

We next increase the complexity of the problem by considering stochastic block model with 2 classes (2-SBM). This class of graphs has been studied in the context of one-sample hypothesis testing, particularly for detecting community structure in graphs and estimating the number of communities in a block model (Bickel and Sarkar (2016), Gao and Lafferty (2017), Lei (2016)). We consider typical 2-SBM graphs, which are characterised by a binary vector denoting the communities of the nodes as well as the *within community* and *inter-community* edge probabilities. Formally, this is represented by the set $\widetilde{\mathbb{M}}_n \subset \mathbb{M}_n$ of matrices which can be transformed to have a 2 classes block structure through row/column permutations. Furthermore, the off-diagonal entries in each matrix can take at most two distinct values. It is easy to observe that each matrix is governed by $n + 2$ parameters, and the problem has a dimension much smaller than $\binom{n}{2}$-dimensional IER problem. However, the following result shows that the minimax separation does not decrease in this case compared to the general IER setting.

PROPOSITION 6.5 (Minimax separation for testing 2-SBM graphs). *Consider the graph two-sample problem restricted to* 2*-SBM graphs with specified* n, m *and* δ. *The minimax separation for operator norm is*

$$\rho^*_{\text{op}} \asymp n\delta \bigwedge \sqrt{\frac{n\delta}{m}} \quad \textit{for all } m \ge 1.$$

For minimax separation in Frobenius norm ρ^*_F, *the above rate hold only for* $m \ge 2$. *For* $m = 1$, *we have loose bounds* $n\delta \wedge \sqrt{n\delta} \lesssim \rho^*_F \lesssim n\delta$.

Hence, *the minimax separation for* 2*-SBM is similar to the corresponding rates for* IER *graphs* (*with possible exception of* ρ^*_F *for* $m = 1$).

PROOF SKETCH. We first note that the testing problem is a restriction of the original IER testing problem, and hence, the upper bounds of ρ^*_F and ρ^*_{op}, derived in Theorem 4.1 and Corollary 5.5, respectively, also hold in this case. Hence, we only need to prove the lower bounds

$$\rho^*_{\text{op}} \gtrsim n\delta \bigwedge \sqrt{\frac{n\delta}{m}} \quad \text{and} \quad \rho^*_F \gtrsim n\delta \bigwedge \sqrt{\frac{n\delta}{m}}$$

for all $m \geq 1$. This lower bound on ρ^*_{op} follows directly from the proof of Lemma 5.2. Recall that the construction used to prove Lemma 5.2 was essentially a case of distinguishing a 2-SBM from an ER, where the latter is also a special case of a 2-SBM. Hence, the same proof works in the present case as well, and the claimed lower bound for ρ^*_{op} holds.

The lower bound for ρ^*_F also follows from the same construction and computing the Frobenius norm distance between the choice of population adjacency matrices used in proof of Lemma 5.2. □

Proposition 6.5 may have important consequences since it apparently implies that for any model that is more complex than the simple 2-SBM, the two-sample problem is as difficult as the general setting of IER graphs. However, a more in-depth study may be required before a strong claim can be made in this context. For instance, one should take into account the fact that the broad literature on graph clustering and stochastic block model often require an additional assumption of balanced community size. It would be interesting to understand whether the presence of balanced communities also simplify the detection/testing problems.

In a broader context, one should note that Propositions 6.4 and 6.5 only provide minimax rates for the Frobenius and operator norm distances for these special classes of IER graphs. On the other hand, Proposition 3.1 and 3.2 demonstrate the limitation of distribution based distances when no restriction is imposed on IER model. Hence, there is a possibility that, under special models, total variation and KL-divergence based tests are as useful or even better than Frobenius or operator norm based tests. Insights into these questions, along with minimax rates for related models such as k-SBM and random dot product graphs, may provide a clear understanding of the problem of testing graphs on a common set of vertices.

6.4. *Extensions.* Several extensions of the two sample problem (2)–(3) can be studied. Earlier in this section, we have discussed the possibility of considering alternative notions of sparsity. Another interesting, and practically significant, extension is to the case of directed graphs. The problem naturally extends to this framework, and the proposed adaptive tests easily tackle this generalisation without any critical modification. For instance, in the case Ψ_F, one merely needs to define $\widehat{\mu}$ and $\widehat{\sigma}$ as a summation over all off-diagonal terms and the thresholds change only by constant factors. The analysis of such tests as well as the minimax lower bounds can be easily derived from our proofs. The same conclusion is true for the case of operator norm separation, particularly, when the upper bounds are derived based on Ψ_{op} and the matrix Bernstein inequality.

In this paper, we only consider the problem of identity testing, that is, $P = Q$ or $d(P, Q) > \rho$. One may also study the more general problem of closeness testing, which ignores small differences between the models, that is, one tests between the hypotheses

$$\Omega_0 = \{d(P, Q) \leq \epsilon, \|P\|_{\max} \vee \|Q\|_{\max} \leq \delta\}, \quad \text{and}$$
$$\Omega_1 = \{d(P, Q) > \rho, \|P\|_{\max} \vee \|Q\|_{\max} \leq \delta\}$$

for some pre-specified $\epsilon < \rho$. The proposed tests, which are primarily based on the principle of estimating $d(P, Q)$ may be easily adapted to this setting by appropriately modifying the test thresholds. However, it is not clear whether the minimax separation bounds in Theorems 4.1 and 5.1 easily extend to this setting as well.

From a practical perspective, one may face a more general problem of two sample graph testing, where the graphs are not defined on a common set of vertices and may even be of different sizes. This situation is generally hard to study, but tests for this problem are often used in many applications, where one typically computes some scalar or vector function from each graph and comments on the difference between two graph populations based on this function (Stam et al. (2007)). We study this principle in a recent work (Ghoshdastidar et al.

(2017)), and propose a formal framework for testing between any two random graphs through the means of a network function $f : \mathcal{G}_{\geq n} \to \mathcal{M}$ that maps the space of graphs on at least n vertices to some metric space $\mathcal{M}$. We argue that if the network function concentrates for some sub-class of random graphs as $n \to \infty$, then one can indeed construct two sample tests based on the network function. However, such a test cannot distinguish between equivalence classes, that is, random graph models that behave identically under the mapping f.

Acknowledgements. We are grateful to the Associate Editor and the anonymous reviewers for their comments and suggestions, which have greatly improved the technical content and the exposition of the paper.

The first author was supported in part by the Deutsche Forschungsgemeinschaft (DFG, FOR1735), the Institutional Strategy of the University of Tübingen (DFG, ZUK 63) and by the Baden-Württemberg Stiftung through the Eliteprogramm for Postdocs.

The second author was supported by the Deutsche Forschungsgemeinschaft (DFG) Emmy Noether grant MuSyAD (CA 1488/1-1).

The third author was supported in part by the DFG Emmy Noether grant MuSyAD (CA 1488/1-1), by the DFG-314838170, GRK 2297 MathCoRe, by the DFG GRK 2433 DAEDALUS (384950143/GRK2433), by the DFG CRC 1294 "Data Assimilation", Project A03 and by the UFA-DFH through the French–German Doktorandenkolleg CDFA 01-18.

The fourth author was supported in part by the DFG FOR1735, the Institutional Strategy of the University of Tübingen (DFG, ZUK 63) and DFG Cluster of Excellence "Machine Learning—New Perspectives for Science" EXC 2064/1, project number 390727645.

SUPPLEMENTARY MATERIAL

Supplement: Appendix to "Two-sample hypothesis testing for inhomogeneous random graphs" (DOI: 10.1214/19-AOS1884SUPP; .pdf). The supplementary material (Ghoshdastidar et al. (2020)) contains the detailed proofs of the lemmas and the propositions stated in Sections 3, 4 and 5.

REFERENCES

ABBE, E. and SANDON, C. (2016). Achieving the KS threshold in the general stochastic block model with linearized acyclic belief propagation. In *Advances in Neural Information Processing Systems* **29**.

ALBERT, R. and BARABÁSI, A.-L. (2002). Statistical mechanics of complex networks. *Rev. Modern Phys.* **74** 47–97. MR1895096 https://doi.org/10.1103/RevModPhys.74.47

ARIAS-CASTRO, E., BUBECK, S. and LUGOSI, G. (2015). Detecting positive correlations in a multivariate sample. *Bernoulli* **21** 209–241. MR3322317 https://doi.org/10.3150/13-BEJ565

ARIAS-CASTRO, E. and VERZELEN, N. (2014). Community detection in dense random networks. *Ann. Statist.* **42** 940–969. MR3210992 https://doi.org/10.1214/14-AOS1208

BAI, Z. and SARANADASA, H. (1996). Effect of high dimension: By an example of a two sample problem. *Statist. Sinica* **6** 311–329. MR1399305

BANDEIRA, A. S. and VAN HANDEL, R. (2016). Sharp nonasymptotic bounds on the norm of random matrices with independent entries. *Ann. Probab.* **44** 2479–2506. MR3531673 https://doi.org/10.1214/15-AOP1025

BARAUD, Y. (2002). Non-asymptotic minimax rates of testing in signal detection. *Bernoulli* **8** 577–606. MR1935648

BERGER, N., BORGS, C., CHAYES, J. T. and SABERI, A. (2005). On the spread of viruses on the Internet. In *Proceedings of the Sixteenth Annual ACM-SIAM Symposium on Discrete Algorithms* 301–310. ACM, New York. MR2298278

BERTHET, Q. and RIGOLLET, P. (2013). Optimal detection of sparse principal components in high dimension. *Ann. Statist.* **41** 1780–1815. MR3127849 https://doi.org/10.1214/13-AOS1127

BICKEL, P. J. and SARKAR, P. (2016). Hypothesis testing for automated community detection in networks. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **78** 253–273. MR3453655 https://doi.org/10.1111/rssb.12117

BOLLOBÁS, B., JANSON, S. and RIORDAN, O. (2007). The phase transition in inhomogeneous random graphs. *Random Structures Algorithms* **31** 3–122. MR2337396 https://doi.org/10.1002/rsa.20168

BRESLER, G. and NAGARAJ, D. (2018). Optimal single sample tests for structured versus unstructured network data. In *Conference on Learning Theory*. *PMLR* **75** 1657–1690.

BUBECK, S., DING, J., ELDAN, R. and RÁCZ, M. Z. (2016). Testing for high-dimensional geometry in random graphs. *Random Structures Algorithms* **49** 503–532. MR3545825 https://doi.org/10.1002/rsa.20633

CAI, T. T., LIU, W. and XIA, Y. (2014). Two-sample test of high dimensional means under dependence. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **76** 349–372. MR3164870 https://doi.org/10.1111/rssb.12034

CARPENTIER, A. and NICKL, R. (2015). On signal detection and confidence sets for low rank inference problems. *Electron. J. Stat.* **9** 2675–2688. MR3432430 https://doi.org/10.1214/15-EJS1087

CHAN, S.-O., DIAKONIKOLAS, I., VALIANT, G. and VALIANT, P. (2014). Optimal algorithms for testing closeness of discrete distributions. In *Proceedings of the Twenty-Fifth Annual ACM-SIAM Symposium on Discrete Algorithms* 1193–1203. ACM, New York. MR3376448 https://doi.org/10.1137/1.9781611973402.88

CHEN, S. X. and QIN, Y.-L. (2010). A two-sample test for high-dimensional data with applications to gene-set testing. *Ann. Statist.* **38** 808–835. MR2604697 https://doi.org/10.1214/09-AOS716

DASKALAKIS, C., DIKKALA, N. and KAMATH, G. (2018). Testing Ising models. In *Proceedings of the Twenty-Ninth Annual ACM-SIAM Symposium on Discrete Algorithms* 1989–2007. SIAM, Philadelphia, PA. MR3775918 https://doi.org/10.1137/1.9781611975031.130

DECELLE, A., KRZAKALA, F., MOORE, C. and ZDEBOROVÁ, L. (2011). Asymptotic analysis of the stochastic block model for modular networks and its algorithmic applications. *Phys. Rev. E* **84** 066106.

DONOHO, D. and JIN, J. (2004). Higher criticism for detecting sparse heterogeneous mixtures. *Ann. Statist.* **32** 962–994. MR2065195 https://doi.org/10.1214/009053604000000265

DONOHO, D. and JIN, J. (2015). Higher criticism for large-scale inference, especially for rare and weak effects. *Statist. Sci.* **30** 1–25. MR3317751 https://doi.org/10.1214/14-STS506

GAO, C. and LAFFERTY, J. (2017). Testing network structure using relations between small subgraph probabilities. Preprint. Available at arXiv:1704.06742.

GHOSHDASTIDAR, D. and VON LUXBURG, U. (2018). Practical methods for graph two-sample testing. In *Advances in Neural Information Processing Systems* **31**.

GHOSHDASTIDAR, D., GUTZEIT, M., CARPENTIER, A. and VON LUXBURG, U. (2017). Two-sample tests for large random graphs using network statistics. In *Conference on Learning Theory*. *PMLR* **65** 954–977.

GHOSHDASTIDAR, D., GUTZEIT, M., CARPENTIER, A. and VON LUXBURG, U. (2020). Supplement to "Two-sample hypothesis testing for inhomogeneous random graphs." https://doi.org/10.1214/19-AOS1884SUPP.

GINESTET, C. E., LI, J., BALACHANDRAN, P., ROSENBERG, S. and KOLACZYK, E. D. (2017). Hypothesis testing for network data in functional neuroimaging. *Ann. Appl. Stat.* **11** 725–750. MR3693544 https://doi.org/10.1214/16-AOAS1015

GRETTON, A., BORGWARDT, K. M., RASCH, M. J., SCHÖLKOPF, B. and SMOLA, A. (2012). A kernel two-sample test. *J. Mach. Learn. Res.* **13** 723–773. MR2913716

HYDUKE, D. R., LEWIS, N. E. and PALSSON, B. (2013). Analysis of omics data with genome-scale models of metabolism. *Mol. BioSyst.* **9** 167–174.

INGSTER, YU. I. and SUSLINA, I. A. (2003). *Nonparametric Goodness-of-Fit Testing Under Gaussian Models*. *Lecture Notes in Statistics* **169**. Springer, New York. MR1991446 https://doi.org/10.1007/978-0-387-21580-8

KLOPP, O., TSYBAKOV, A. B. and VERZELEN, N. (2017). Oracle inequalities for network models and sparse graphon estimation. *Ann. Statist.* **45** 316–354. MR3611494 https://doi.org/10.1214/16-AOS1454

KONDOR, R. and PAN, H. (2016). The multiscale Laplacian graph kernel. In *Advances in Neural Information Processing Systems*.

LANDMAN, B. A., HUANG, A. J., GIFFORD, A., VIKRAM, D. S., LIM, I. A., FARRELL, J. A., BOGOVIC, J. A., HUA, J., CHEN, M. et al. (2011). Multi-parametric neuroimaging reproducibility: A 3-T resource study. *NeuroImage* **54** 2854–2866.

LE, C. M., LEVINA, E. and VERSHYNIN, R. (2017). Concentration and regularization of random graphs. *Random Structures Algorithms* **51** 538–561. MR3689343 https://doi.org/10.1002/rsa.20713

LEDOIT, O. and WOLF, M. (2002). Some hypothesis tests for the covariance matrix when the dimension is large compared to the sample size. *Ann. Statist.* **30** 1081–1102. MR1926169 https://doi.org/10.1214/aos/1031689018

LEI, J. (2016). A goodness-of-fit test for stochastic block models. *Ann. Statist.* **44** 401–424. MR3449773 https://doi.org/10.1214/15-AOS1370

LOVÁSZ, L. (2012). *Large Networks and Graph Limits*. *American Mathematical Society Colloquium Publications* **60**. Am. Math. Soc., Providence, RI. MR3012035 https://doi.org/10.1090/coll/060

LU, L. and PENG, X. (2013). Spectra of edge-independent random graphs. *Electron. J. Combin.* **20** Paper 27. MR3158266

MUKHERJEE, R., PILLAI, N. S. and LIN, X. (2015). Hypothesis testing for high-dimensional sparse binary regression. *Ann. Statist.* **43** 352–381. MR3311863 https://doi.org/10.1214/14-AOS1279

OLIVEIRA, R. I. (2009). Concentration of the adjacency matrix and of the Laplacian in random graphs with independent edges. Preprint. Available at arXiv:0911.0600.

RAMDAS, A., REDDI, S., POCZOS, B., SINGH, A. and WASSERMAN, L. (2015). Adaptivity and computation-statistics tradeoffs for kernel and distance based high dimensional two sample testing. Preprint. Available at arXiv:1508.00655.

RYABKO, D. (2017). Hypotheses testing on infinite random graphs. In *International Conference on Algorithmic Learning Theory*. *Proc*. *Mach*. *Learn*. *Res*. (*PMLR*) **76** 1–12. Proceedings of Machine Learning Research PMLR. MR3776703

SHERVASHIDZE, N., SCHWEITZER, P., VAN LEEUWEN, E. J., MEHLHORN, K. and BORGWARDT, K. M. (2011). Weisfeiler–Lehman graph kernels. *J. Mach*. *Learn*. *Res*. **12** 2539–2561. MR2845672

STAM, C. J., JONES, B. F., NOLTE, G., BREAKSPEAR, M. and SCHELTENS, P. (2007). Small-world networks and functional connectivity in Alzheimer's disease. *Cereb. Cortex* **17** 92–99.

SUN, X. and NOBEL, A. B. (2008). On the size and recovery of submatrices of ones in a random binary matrix. *J. Mach*. *Learn*. *Res*. **9** 2431–2453. MR2460888

TANG, M., ATHREYA, A., SUSSMAN, D. L., LYZINSKI, V., PARK, Y. and PRIEBE, C. E. (2017a). A semi-parametric two-sample hypothesis testing problem for random graphs. *J. Comput*. *Graph*. *Statist*. **26** 344–354. MR3640191 https://doi.org/10.1080/10618600.2016.1193505

TANG, M., ATHREYA, A., SUSSMAN, D. L., LYZINSKI, V. and PRIEBE, C. E. (2017b). A nonparametric two-sample hypothesis testing problem for random graphs. *Bernoulli* **23** 1599–1630. MR3624872 https://doi.org/10.3150/15-BEJ789

TROPP, J. A. (2012). User-friendly tail bounds for sums of random matrices. *Found*. *Comput*. *Math*. **12** 389–434. MR2946459 https://doi.org/10.1007/s10208-011-9099-z

VERZELEN, N. and ARIAS-CASTRO, E. (2015). Community detection in sparse random networks. *Ann*. *Appl*. *Probab*. **25** 3465–3510. MR3404642 https://doi.org/10.1214/14-AAP1080

VERZELEN, N. and ARIAS-CASTRO, E. (2017). Detection and feature selection in sparse mixture models. *Ann*. *Statist*. **45** 1920–1950. MR3718157 https://doi.org/10.1214/16-AOS1513

VISHWANATHAN, S. V. N., SCHRAUDOLPH, N. N., KONDOR, R. and BORGWARDT, K. M. (2010). Graph kernels. *J. Mach*. *Learn*. *Res*. **11** 1201–1242. MR2645450 https://doi.org/10.1093/chemse/bjq147