# Two-Sample Tests for Large Random Graphs Using Network Statistics

**Debarghya Ghoshdastidar**                    DEBARGHYA.GHOSHDASTIDAR@UNI-TUEBINGEN.DE
*Department of Computer Science, University of Tübingen,*
*Sand 14, 72076 Tübingen, Germany*

**Maurilio Gutzeit**                           MGUTZEIT@UNI-POTSDAM.DE
**Alexandra Carpentier**                       CARPENTIER@UNI-POTSDAM.DE
*Department of Mathematics, University of Potsdam,*
*Karl-Liebknecht-Strasse 24-25, D-14476 Potsdam OT Golm, Germany*

**Ulrike von Luxburg**                         LUXBURG@INFORMATIK.UNI-TUEBINGEN.DE
*Department of Computer Science, University of Tübingen,*
*Sand 14, 72076 Tübingen, Germany*

## Abstract

We consider a two-sample hypothesis testing problem, where the distributions are defined on the space of undirected graphs, and one has access to only one observation from each model. A motivating example for this problem is comparing the friendship networks on Facebook and LinkedIn. The practical approach to such problems is to compare the networks based on certain network statistics. In this paper, we present a general principle for two-sample hypothesis testing in such scenarios without making any assumption about the network generation process. The main contribution of the paper is a general formulation of the problem based on concentration of network statistics, and consequently, a consistent two-sample test that arises as the natural solution for this problem. We also show that the proposed test is minimax optimal for certain network statistics.

**Keywords:** Two-sample test; Random graphs; Minimax testing; Concentration inequalities.

## 1. Introduction

Hypothesis testing has received considerable attention in recent times, particularly in the high-dimensional setting, where the number of observations is much smaller than the dimension of data (Gretton et al., 2012; Mukherjee et al., 2015). From this perspective, an extreme situation arises in network analysis, where we observe one large network and need to draw inference based on a single instance of a high-dimensional object. Surprisingly, simple tests are known to work quite well in many of these graph inference problems. For instance, Bubeck et al. (2016) compute the number of signed triangles in a graph to test whether the graph has an underlying geometric structure, or it is purely random (Erdös-Rényi). In the context of community detection, Arias-Castro and Verzelen (2014) use modularity and clique number to detect the presence of dense sub-graphs. Along similar lines, Lei (2016) use the largest few singular values of the adjacency matrix to decide whether a graph has $k$ or more communities. In all these works, the constructed tests exploit the fact that there exist some network statistics, for example number of triangles or the graph spectrum, that help to distinguish between the null and the alternative hypotheses. This principle is not restricted to specific testing problems, but a general approach often used by practitioners from different disciplines in the context of testing and fitting of network models (Rubinov and Sporns, 2010). For

instance, Klimm et al. (2014) write that the Erdös-Rényi (ER) model is a poor fit for the brain network since *"the clustering coefficient (in ER) is smaller and the path length shorter than that of anatomical brain networks"*. However, such a statement is purely qualitative, and arises due to the lack of a formal framework for testing random graphs.

The purpose of the present paper is to formulate a general approach for hypothesis testing for large random graphs, particularly when one does not have access to independent and identically distributed samples generated from the same network model. We focus on the problem of two-sample hypothesis testing, where one observes two random graphs, possibly of different sizes. Based on the two given graphs, the problem is to decide whether the underlying distributions that generate the graphs are same or different. The problem surfaces naturally in several domains, for instance, comparison of brain networks of people with and without Alzheimer (Stam et al., 2007). However, the problem has been previously studied only in a very restrictive setting, where the two graphs are assumed to be random dot product graphs (RDPG) defined on the same set of vertices (Tang et al., 2017). The RDPG model has a semi-parametric characterisation, which allows one to estimate the parameters up to some transformation, and then use these estimates for a two-sample test. Obviously, this approach cannot be directly extended to more general classes of network models. A more critical limitation is that the study cannot be easily extended to compare graphs that do not have a vertex correspondence, and more generally, graphs of different sizes.

One can easily see that the aforementioned hypothesis testing problem is ill-posed in general since the distributions that generate graphs of different size reside in different spaces, and hence, cannot be directly compared. To resolve this issue, we take the practical approach of comparing two models in terms of network statistics. In other words, we compute a function $f$ for both graphs, and decide whether the two graphs are similar or different with respect to $f$. It is obvious that this approach cannot distinguish between two distributions for which $f$ behave in the same way, but this is a cost that is often incurred in practice, where one needs to know the interesting features that distinguish two network models (Stam et al., 2007; Klimm et al., 2014). Special cases of such a problem can be posed, and even solved, quite easily. For example, a simple situation arises if we restrict only to ER graphs, and $f$ corresponds to the edge probability. However, it is not clear how one can pose the problem formally in a more general setting, where one does not fix the network statistic or make any assumption on the network model. In this paper, we tackle this issue by restricting the class of network statistics $f$ that can be used, and then tailor the hypothesis testing problem to the pre-specified network statistic. We restrict the class of statistics $f$ to those that concentrate for large graphs, and subsequently pose the testing problem in terms of the point of concentration of $f$. As a consequence, we also find an uniformly consistent test that arises naturally for this problem, and is also optimal in some cases.

The paper is organised as follows. In Section 2, we present a general assumption on network statistics, and show that two common statistics, based on (i) number of triangles, and (ii) largest $k$ singular values of adjacency matrix, satisfy this assumption. We set up the hypothesis testing problem in Section 3, and present a generic two-sample test for the problem. We also show that if the null and alternative hypotheses are separated by a certain factor, then the test can achieve arbitrarily small error rates for large enough graphs (see Theorem 5). We provide more concrete statements in Section 4, where we restrict the discussion to the above mentioned network statistics, and consider specific network models, namely, edge-independent but inhomogeneous random graphs (Bollobas et al., 2007), and random geometric graphs (Penrose, 2003). Subsequently, in Section 5, we consider inhomogeneous random graphs and the mentioned network statistics, and prove that the separation

condition derived in Theorem 5 matches with the detection boundary. In other words, the proposed test achieves the minimax rate of testing for certain class of random graphs and network statistics. We present proofs of all results in Section 6. Finally, we conclude in Section 7 with some discussion and open questions.

## 2. Network statistic

We denote the set of all undirected unweighted graphs on $n$ vertices by $\mathcal{G}_n$, and let $\mathcal{F}_n$ be the set of all probability distributions on $\mathcal{G}_n$. For convenience, we write $\mathcal{G}_{\geq N}$ instead of $\bigcup_{n \geq N} \mathcal{G}_n$. If $Q$ is a distribution, we use $G \sim Q$ to say that $G$ is generated from $Q$. We also use the notation $o_n(1)$ to denote the class of functions of $n$ that vanish as $n \to \infty$.

Let $(\mathcal{S}, d)$ be a metric space. We call any mapping $f : \mathcal{G}_{\geq 1} \to \mathcal{S}$ a network statistic. Observe that the function $f$ acts on all possible graphs, and maps them to a metric space. Some natural examples include scalar functions such as edge density or number of triangles; multivariate statistics like largest $k$ singular values of adjacency matrix; and even quantities like degree distribution. For the last example, $\mathcal{S}$ corresponds to the space of all distributions on natural numbers endowed with a suitable metric. All the popular network measures (see Rubinov and Sporns, 2010; Bounova and de Weck, 2012) can be put in this framework. However, not all such functions are useful for hypothesis testing. The following assumption provides a way to characterise the useful network statistics.

**Assumption 1 (Statistic that concentrates for large graphs)** *Given a network statistic $f$, we assume that there exist*

- *a set $\mathcal{F}_f \subset \mathcal{F}_{\geq 1}$ (a class of distributions that meet the requirements for concentration of $f$),*

- *a mapping $\mu : \mathcal{F}_f \to \mathcal{S}$ (the point of concentration of $f$ for a particular distribution), and*

- *a mapping $\sigma : \mathcal{F}_f \to \mathbb{R}$ (the deviation of $f$ from $\mu$),*

*such that*

$$\sup_{Q \in \mathcal{F}_f \cap \mathcal{F}_n} \mathsf{P}_{G \sim Q} \big( d(f(G), \mu(Q)) > \sigma(Q) \big) \in o_n(1) . \tag{1}$$

The above assumption implies that if we consider a "good" network statistic $f$, then for any distribution $Q \in \mathcal{F}_n$ and $G \sim Q$, the computed statistic $f(G)$ concentrates about an element in $\mathcal{S}$. The function $\mu$ maps the distribution $Q$ to this point of concentration. A natural choice is $\mu(Q) = \mathsf{E}_{G \sim Q}[f(G)]$ if the expectation exists, where the quantity $\sigma$ would be related to the standard deviation of $f(G)$. However, we show later that a more general definition for $\mu$ and $\sigma$ helps to formulate some common testing problems. To this end, one may note that the concentration (1) may not occur for all distributions in $\mathcal{F}_{\geq 1}$, a typical example being models that are very sparse. Hence, Assumption 1 restricts the class of distributions to $\mathcal{F}_f$ that can be viewed as a subset of distributions for which concentration occurs.

We note that Assumption 1 is quite weak in general since it allows several trivial cases. For example, if $f$ corresponds to the average probability of edges, then setting the deviation $\sigma(\cdot) = 1$ shows that $f$ satisfies Assumption 1 for any arbitrary $\mu : \mathcal{F}_{\geq 1} \to [0, 1]$. Hence, in this case, satisfying Assumption 1 is of no practical significance. We now consider few common statistics to show that the assumption often has interesting and useful consequences.

3

**Example 1 (Average probability of triangle)** *Let $G$ be an undirected graph on $n$ vertices with adjacency matrix $A_G$, then the univariate statistic*

$$f_\Delta(G) = \frac{1}{\binom{n}{3}} \sum_{i<j<k} (A_G)_{ij}(A_G)_{jk}(A_G)_{ik} \tag{2}$$

*provides an estimate of the average probability of occurrence of a triangle. Note that $f_\Delta$ maps $\mathcal{G}_{\geq 1}$ to $\mathcal{S} = [0, 1]$ with metric $d$ being the absolute difference.*

The above function $f_\Delta$ is a normalised version of the number of triangles, where the normalisation makes the statistic independent of the graph size. For instance, any Erdös-Rényi (ER) graph $G$ with edge probability $p$ satisfies $\mathsf{E}_G[f_\Delta(G)] = p^3$ irrespective of the graph size. The following result provides a choice of $\sigma$ such that $f_\Delta$ satisfies Assumption 1 for a very broad class of distributions.

**Lemma 1 ($f_\Delta$ satisfies Assumption 1 under a limited correlation condition)** *Define the quantity $\mu(Q) = \mathsf{E}_{G\sim Q}[f_\Delta(G)]$, and let $\mathcal{F}_f$ be all distributions on graphs such that the presence of any triangle is not correlated with indicators of any non-overlapping edge or triangle, that is,*

$$\mathcal{F}_f = \big\{ Q \in \mathcal{F}_{\geq 1} : \text{for } G \sim Q, \text{ and any } i < j < k \text{ and } i' < j' < k' \text{ with } |\{i, j, k\} \cap \{i', j', k'\}| \leq 1,$$
$$\mathsf{E}_{G\sim Q}[\Delta_{ijk}\Delta_{i'j'k'}] = \mathsf{E}_{G\sim Q}[\Delta_{ijk}]\mathsf{E}_{G\sim Q}[\Delta_{i'j'k'}], \text{ and}$$
$$\mathsf{E}_{G\sim Q}[\Delta_{ijk}(A_G)_{i'j'}] = \mathsf{E}_{G\sim Q}[\Delta_{ijk}]\mathsf{E}_{G\sim Q}[(A_G)_{i'j'}]\big\},$$

*where we use the notation $\Delta_{ijk} = (A_G)_{ij}(A_G)_{jk}(A_G)_{ik}$. Then $f_\Delta$ satisfies Assumption 1 for above $\mu$ and $\mathcal{F}_f$ with $\sigma(Q) = \sqrt{(3D_Q + 1)\mu(Q)\ln n/\binom{n}{3}}$, where $n$ is the size of graph generated from $Q$ and $D_Q$ is the maximum expected degree of any node.*

The above class of distributions encompasses a wide range of real-world network models since we only require that any triangle is uncorrelated from any non-overlapping edge or triangle. This result obviously holds for graphs with independent edges, but also other models such as certain random geometric graphs. For instance, Bubeck et al. (2016) use the following definition for geometric graphs. One samples $n$ random vectors $x_1, \ldots, x_n$ i.i.d. uniform from the unit ball in $\mathbb{R}^r$, which correspond to the vertices of the graph. For a given $p \in [0, 1]$, edge $(i, j)$ is added if $x_i^T x_j \geq \tau_{r,p}$, where the threshold $\tau_{r,p}$ is set such that $\mathsf{P}(x_i^T x_j \geq \tau_{r,p}) = p$. We represent this class of distributions by Geom, and note that any $Q \in$ Geom $\cap \mathcal{F}_n$ is defined by two parameters: the dimension of underlying Euclidean space, $r_Q$, and the edge probability, $p_Q$. Due to Lemma 1, we can say that the statistic $f_\Delta$ satisfies Assumption 1 when we consider distributions from Geom class.

**Lemma 2 ($f_\Delta$ satisfies Assumption 1 for Geom class)** *For any $Q \in$ Geom $\cap \mathcal{F}_n$ with parameters $r_Q, p_Q$, define $\mu(Q) = \mathsf{E}_{G\sim Q}[f_\Delta(G)]$. For any absolute constant $C > 0$, $f_\Delta$ satisfies Assumption 1 for the choice of functions $\sigma(Q) = \sqrt{(3np_Q + 1)\mu(Q)\ln n/\binom{n}{3}}$, and*

$$\mathcal{F}_f = \bigcup_{n\geq 1}\left\{Q \in \text{Geom} \cap \mathcal{F}_n : p_Q \geq \frac{1}{n} \text{ and } C \leq r_Q \leq \frac{(np)^4(\ln\frac{1}{p_Q})^3}{(\ln n)}\right\} .$$

4

We note that the conditions in $\mathcal{F}_f$ stated above are not necessary at this stage of discussion. However, our subsequent discussion on hypothesis testing problem and approach require an estimate of $\sigma$ from the random graph, which in turn imposes few additional constraints on the model. In order to simplify our later discussions based on this setting, we restrict to the smaller set $\mathcal{F}_f$ defined above.

It is easy to see that the choice of $\sigma$ mentioned in Lemmas 1 and 2 is not unique, and using a larger deviation function does not lead to violation of the assumption. However, in some cases, one can even consider a smaller deviation function provided that $\mathcal{F}_f$ is restricted accordingly. Consider the class of inhomogeneous random graphs with independent edges (Bollobas et al., 2007), and denote the set of corresponding distributions by IER. In such graphs, each edge occurs with a different probability, and hence, any $Q \in \mathsf{IER} \cap \mathcal{F}_n$ is characterised by a $n \times n$ symmetric matrix $M_Q$ such that $\mathsf{E}_{G \sim Q}[A_G] = M_Q$. Hence, in this case, $\mu(Q) = \mathsf{E}_{G \sim Q}[f_\Delta(G)] = \frac{1}{\binom{n}{3}} \sum_{i<j<k} (M_Q)_{ij}(M_Q)_{jk}(M_Q)_{ik}$. If $\mathcal{F}_f = \mathsf{IER}$, then Lemma 1 provides a choice of $\sigma$. The following result shows that if we only consider sparse IER graphs, then $f_\Delta$ also satisfies the same assumption for a smaller deviation function.

**Lemma 3 ($f_\Delta$ satisfies Assumption 1 for semi-sparse IER class)** *For any $Q \in \mathsf{IER} \cap \mathcal{F}_n$ with associated matrix $M_Q$, if $\mu(Q) = \mathsf{E}_{G \sim Q}[f_\Delta(G)] = \frac{1}{\binom{n}{3}} \sum_{i<j<k} (M_Q)_{ij}(M_Q)_{jk}(M_Q)_{ik}$, then $f_\Delta$ satisfies Assumption 1 for the choices $\sigma(Q) = 2\sqrt{\mu(Q)\ln n / \binom{n}{3}}$, and*

$$\mathcal{F}_f = \bigcup_{n \geq 1} \left\{ Q \in \mathsf{IER} \cap \mathcal{F}_n : \mu(Q) \geq \frac{\ln n}{n^3} \text{ and } \max_{i,j} \sum_{k \neq i,j} (M_Q)_{ik}(M_Q)_{jk} \leq 1 \right\} .$$

Observe that the condition $\sum_k (M_Q)_{ik}(M_Q)_{jk} \leq 1$ is equivalent to stating that the graphs are sparse enough so that, in the expected sense, no edge appears in more than one triangle. The condition on minimum growth rate of $\mu(Q)$ is not necessary at this stage, and simply ensures that $\mu(Q)$ can be estimated from $G \sim Q$.

Based on Lemma 1, one may also consider a combination of both IER and Geom classes, which allows one to tackle problems where two random graphs are chosen from $\mathsf{IER} \cup \mathsf{Geom}$. This emphasises the flexibility of the present discussion and the forthcoming results in the sense that as long as one can show that a network statistic concentrates, one can apply the framework and result of this paper. We now look at one more common statistic.

**Example 2 (Normalised largest $k$ singular values)** *Let $\lambda_1(A_G) \geq \ldots \geq \lambda_k(A_G)$ be the largest $k$ singular values of adjacency matrix $A_G$. Then the multivariate statistic*

$$f_\lambda(G) = \frac{1}{n}\big(\lambda_1(A_G), \lambda_2(A_G), \ldots, \lambda_k(A_G)\big)^T , \tag{3}$$

*maps every graph to $\mathcal{S} = \mathbb{R}^k$ endowed with any standard metric $d$. For concreteness, we assume $d$ is the Euclidean distance.*

This is yet another statistic whose concentration properties have been well studied, especially for the IER class of random graphs. The scaling of $\frac{1}{n}$ again helps to reduce the dependence on graph size. Alternative ways to achieve this could be to consider the spectrum of normalised adjacency or normalised Laplacian (Chung, 1997).

In the case of IER graphs, Alon et al. (2002) provide the rate of concentration of eigenvalues about their expected values. Based on this, one may set $\mu(Q) = \mathsf{E}_{G \sim Q}[f_\lambda(G)]$, and claim that $f_\lambda$ satisfies Assumption 1. But a more interesting fact, from a practical perspective, is that for $Q \in \mathsf{IER} \cap \mathcal{F}_n$, the statistic $f_\lambda(G)$ also concentrates about $\frac{1}{n}\big(\lambda_1(M_Q), \lambda_2(M_Q), \ldots, \lambda_k(M_Q)\big)^T$. Thus, one can make the following claim for $f_\lambda$ using concentration results of Lu and Peng (2013).

**Lemma 4 ($f_\lambda$ satisfies Assumption 1 for semi-sparse IER class)** *For any $Q \in \mathsf{IER} \cap \mathcal{F}_n$ with associated matrix $M_Q$, define $\mu(Q) = \frac{1}{n}\big(\lambda_1(M_Q), \ldots, \lambda_k(M_Q)\big)^T$, and let $D_Q = \max_i \sum_j (M_Q)_{ij}$. Then $f_\lambda$ satisfies Assumption 1 for the choices $\sigma(Q) = \frac{2.1}{n}\sqrt{kD_Q}$, and*

$$\mathcal{F}_f = \bigcup_{n \geq 1} \left\{ Q \in \mathsf{IER} \cap \mathcal{F}_n : D_Q \geq (\ln n)^{4.1} \right\},$$

*where the constant 2.1 (or 4.1) may be replaced by any value greater than 2 (resp., 4).*

The restricted set $\mathcal{F}_f$ is crucial here. One can show that $f_\lambda$ does not satisfy Assumption 1 over the larger set $\mathsf{IER} \cap \mathcal{F}_{\geq 1}$. In particular, set $k = 1$ and consider any ER distribution $Q \in \mathcal{F}_n$ with edge probability $\frac{c}{n}$ for some constant $c > 0$. Then it is known that $f_\lambda(G) \geq \frac{1}{n}\sqrt{\frac{\ln n}{\ln \ln n}}$ for any $G \sim Q$ (see Krivelevich and Sudakov, 2003), whereas $\mu(Q) \leq \frac{c}{n}$ decays much faster. This shows the importance of restricting the statement of Assumption 1 to a suitable set $\mathcal{F}_f$. Moreover, we note that in the case of $f_\lambda$, one may use an alternative choice of $\sigma$ that provides concentration even for sparser graphs $D_Q \geq (\ln n)^{1.1}$ (Lei and Rinaldo, 2015). Such a choice of $\sigma$ depends on $\max_{ij}(M_Q)_{ij}$, which is difficult to estimate from a single random graph.

## 3. Two-sample hypothesis testing

Based on the discussions about the network statistic, we are now prepared for a formal statement of the hypothesis testing problem under consideration.

**Problem** *Let $f$ be a pre-specified network statistic that satisfies Assumption 1 with the associated quantities $\mathcal{F}_f$ and $\mu$. Let $Q, Q' \in \mathcal{F}_f$, and $G$ and $G'$ be random graphs (of possibly different sizes) generated from $Q$ and $Q'$, respectively. Given $G$ and $G'$, we test the null hypothesis*

$$H_0 : Q, Q' \in \mathcal{F}_f \text{ with } d(\mu(Q), \mu(Q')) \leq \epsilon(Q, Q')$$

*against the alternative hypothesis*

$$H_1 : Q, Q' \in \mathcal{F}_f \text{ with } d(\mu(Q), \mu(Q')) > \rho(Q, Q'),$$

*where $\epsilon, \rho$ are non-negative scalar functions of the distributions $Q, Q'$ so that $\epsilon(Q, Q') \leq \rho(Q, Q')$.*

Setting $\epsilon(\cdot, \cdot) = 0$ restricts $H_0$ to the case where $\mu(Q) = \mu(Q')$. However, as shown later, it is often useful to provide some leeway by allowing $\epsilon$ to be positive. On the other hand, the function $\rho$ plays the role of a separation that is often required in hypothesis testing (see, for instance, Ingster and Suslina, 2000). In the present context, we later show that if $\rho(Q, Q')$ is too small, then there exist hypotheses that cannot be distinguished by any test.

In the rest of the section, we construct a two-sample test, and show that for certain range of $\epsilon$ and $\rho$, the test can achieve arbitrarily small error for large graphs. In Section 5, we use examples to demonstrate that the test is minimax optimal in some cases. Before proceeding, we need few more definitions. For any statistic $f$ and distribution $Q \in \mathcal{F}_f$, we define $\mathcal{B}_f(Q, \epsilon) = \{Q' \in \mathcal{F}_f : d(\mu(Q), \mu(Q')) \leq \epsilon(Q, Q')\}$. Intuitively, one can think of this set as the inverse image of a closed ball in $\mathcal{S}$ centred at $\mu(Q)$ and radius specified by the function $\epsilon$. Similarly, we define a complement of the "ball" for the function $\rho$ as $\overline{\mathcal{B}_f(Q, \rho)} = \{Q' \in \mathcal{F}_f : d(\mu(Q), \mu(Q')) > \rho(Q, Q')\}$. In this terminology, one can see that $H_0$ is true if $Q' \in \mathcal{B}_f(Q, \epsilon)$, while $H_1$ is true for $Q' \in \overline{\mathcal{B}_f(Q, \rho)}$.

### 3.1. A consistent two-sample test

We now construct a test based on concentration of $f$ that is ensured by Assumption 1. To construct the test, we require an additional assumption on $f$, namely, the fact that $\sigma$ as defined in Assumption 1 can be estimated accurately from the graph.

**Assumption 2 ($\sigma(Q)$ can be estimated from $G$)** *Let $\mathcal{F}_f$ and $\sigma$ be as defined in Assumption 1. There exists a function $\widehat{\sigma} : \mathcal{G}_{\geq 1} \to \mathbb{R}$ with the property that for any $\delta > 0$,*

$$\sup_{Q \in \mathcal{F}_f \cap \mathcal{F}_n} \mathsf{P}_{G \sim Q}\big(|\widehat{\sigma}(G) - \sigma(Q)| > \delta \cdot \sigma(Q)\big) \in o_n(1). \tag{4}$$

This assumption rules out some possibilities. For example, one could also state Lemma 4 with $\sigma$ defined in terms of maximum edge probability, $\max_{ij}(M_Q)_{ij}$ instead of maximum expected degree $D_Q$. However, in that case, it would be impossible to estimate $\sigma$ for a single random graph. Thus, such a concentration of $f_\lambda$ stated in terms of $\max_{ij}(M_Q)_{ij}$ is not useful for the test described below.

The two-sample test that we propose is quite straightforward. Given the random graphs $G$ and $G'$, we define the test statistic

$$T(G, G') = \frac{d(f(G), f(G'))}{2\widehat{\sigma}(G) + 2\widehat{\sigma}(G')} , \tag{5}$$

Based on the test statistic in (5), we accept the null hypothesis $H_0$ if $T(G, G') \leq 1$, and reject it if $T(G, G') > 1$. The following result shows that the above test is uniformly consistent for a suitable network measure $f$, and a large enough separation $\rho$.

**Theorem 5 (Proposed test is uniformly consistent)** *Let $f$ be a network statistic satisfying Assumptions 1 and 2, and $\mathcal{F}_f$ and $\sigma$ be as specified in the assumptions. If the functions $\epsilon$ and $\rho$ satisfy*

$$\epsilon(Q, Q') \leq 0.5(\sigma(Q) + \sigma(Q')) \qquad and \qquad \rho(Q, Q') \geq 3.5(\sigma(Q) + \sigma(Q')) \tag{6}$$

*for all $Q, Q' \in \mathcal{F}_f$, then*

$$\sup_{Q \in \mathcal{F}_f \cap \mathcal{F}_{\geq n}} \Bigg( \sup_{Q' \in \mathcal{B}_f(Q, \epsilon) \cap \mathcal{F}_{\geq n}} \underbrace{\mathsf{P}_{G \sim Q, G' \sim Q'}(T(G, G') > 1)}_{\text{Type-I error}} +$$

$$\sup_{Q' \in \overline{\mathcal{B}_f(Q, \rho)} \cap \mathcal{F}_{\geq n}} \underbrace{\mathsf{P}_{G \sim Q, G' \sim Q'}(T(G, G') \leq 1)}_{\text{Type-II error}} \Bigg) \in o_n(1).$$

7

In simple terms, the above theorem states that if the hypothesis testing problem is defined with $\epsilon$ and $\rho$ satisfying (6), and both graphs have at least $n$ vertices, then for any pair of $Q, Q'$, the two-sample test described above achieves an error rate (Type-I + Type-II) that vanishes as $n \to \infty$. Thus, the test is an uniformly consistent test. We note here that unlike the standard literature on hypothesis testing, we consider the asymptotics in the size of the graphs rather than the number of independent observations.

One may use a weaker variant of Assumption 2 , where one needs to estimate some upper bound of $\sigma$. This results in a weaker test that can only distinguish models with a larger separation.

## 4. Examples

We elaborate our discussion on Theorem 5 using the two statistics $f_\Delta$ and $f_\lambda$.

### 4.1. Two-sample testing using $f_\Delta$

Recall that $f_\Delta$ is defined as an estimate of the mean probability of triangle occurrence, and for both IER and Geom classes, $f_\Delta(G)$ concentrates at $\mu(Q) = \mathsf{E}_{G \sim Q}[f_\Delta(G)]$. Since, $f_\Delta$ satisfies Assumption 1, one can pose a two-sample testing problem as described in Section 3, where the set of distributions need to be restricted to a suitable $\mathcal{F}_f$. Let us also simplify the problem by fixing $\epsilon(Q, Q') = 0$, that is, we do not distinguish between two models if they are mapped into the same point by $\mu$. Based on Theorem 5, we claim that when the problem is restricted to semi-sparse IER class, then the statistic in (5) leads to a consistent test for appropriately specified $\rho$.

**Corollary 6 (Consistent test for semi-sparse IER using $f_\Delta$)** *Consider the setting of Lemma 3. Then $f_\Delta$ satisfies Assumption 2 for $\widehat{\sigma}(G) = 2\sqrt{f_\Delta(G) \ln n / \binom{n}{3}}$, where $n$ is number of vertices in $G$. Furthermore, the proposed test is uniformly consistent for the testing problem with*

$$\rho(Q, Q') \geq 7 \left( \sqrt{\frac{\mu(Q) \ln n}{\binom{n}{3}}} + \sqrt{\frac{\mu(Q') \ln n'}{\binom{n'}{3}}} \right)$$

*where $Q \in \mathcal{F}_f \cap \mathcal{F}_n$ and $Q' \in \mathcal{F}_f \cap \mathcal{F}_{n'}$.*

The above result provides a good estimator for the deviation function $\sigma$, and the rest follows immediately from Theorem 5. The sufficient condition on $\rho$ stated above may not seem very intuitive, but we show in the next section that the condition is also necessary (up to logarithmic factor) for two-sample testing of IER graphs using $f_\Delta$. In the case of the Geom class, we obtain a result similar to Corollary 6.

**Corollary 7 (Consistent test for Geom using $f_\Delta$)** *Consider the setting of Lemma 2. For any graph $G$ with $n$ vertices, let $\widehat{p}_G$ denote the estimated edge density, that is, $\widehat{p}(G) = \binom{n}{2}^{-1} \sum_{i<j} (A_G)_{ij}$. Then $f_\Delta$ satisfies Assumption 2 for $\widehat{\sigma}(G) = \sqrt{(3n\widehat{p}(G) + 1) f_\Delta(G) \ln n / \binom{n}{3}}$.*

*As a consequence, the proposed test is uniformly consistent for the testing problem with*

$$\rho(Q, Q') \geq 3.5 \left( \sqrt{\frac{(3np_Q + 1)\mu(Q) \ln n}{\binom{n}{3}}} + \sqrt{\frac{(3n'p_{Q'} + 1)\mu(Q') \ln n'}{\binom{n'}{3}}} \right)$$

*where $Q \in \mathcal{F}_f \cap \mathcal{F}_n$ and $Q' \in \mathcal{F}_f \cap \mathcal{F}_{n'}$.*

As mentioned before, one can also pose a testing problem on IER $\cup$ Geom. In particular, we consider the following two-sample version of the problem studied in (Bubeck et al., 2016). Let there be a sequence of probabilities $(p_n)_{n \geq 1} \subset [0, 1]$, and a sequence of dimensions $(r_n)_{n \geq 1} \subset \mathbb{N}$. For each $n$, consider a set containing exactly two distributions from $\mathcal{F}_n$: (i) an ER distribution with edge probability $p_n$, and (ii) the other from Geom class with parameters $p_n$ and $r_n$. Let two graphs be generated on $n$ vertices from either of these two distributions. The problem is to test whether both are generated from the same model, or different models.

**Corollary 8 (Consistent test for distinguishing between ER and** Geom**)** *Consider above problem with $\frac{\ln n}{n} \leq p_n \leq \frac{1}{\sqrt{n}}$ for all $n$. The proposed test is uniformly consistent if $r_n = o_n\left( (\ln \frac{1}{p_n})^3 \right)$.*

The above result simply implies that the condition for identifiability derived by Bubeck et al. (2016) for the sparse one-sample version of the problem remains unchanged in the two-sample case even if we do not assume any knowledge about the parameters. The restriction on $p_n$ is a consequence of Lemma 3, and the lower bound also helps in accurate estimation of the unknown $p_n$. It is also easy to verify that the above result holds even when the two graphs are of different size but the edge probabilities are same.

### 4.2. Two-sample testing using $f_\lambda$

We now discuss in more detail about testing using the $f_\lambda$ statistic for the case of IER graphs. We state a result below that is along the lines of Corollary 6, but the main objective of this part is to demonstrate there are cases, where one needs to expand the null hypothesis by allowing $\epsilon(Q, Q')$ to be a positive function.

The situation typically arises if we deal with graphs of different sizes. For instance, let $Q, Q'$ correspond to ER models with graph sizes $n, n'$ and edge probabilities $p, p'$, respectively. Further, assume that we consider only the largest singular value, that is, $f_\lambda = \frac{\lambda_1(A_G)}{n}$. As a consequence, $\mu(Q) = p - \frac{p}{n}$, and $\mu(Q') = p' - \frac{p'}{n'}$. Typically, we would like to call the distributions $Q, Q'$ same if $p = p'$, irrespective of the the graph sizes, but in this case, if $n \neq n'$, then one can see that $\mu(Q) \neq \mu(Q')$ even if $p = p'$. This observation suggests that for the case of ER graphs, we should not distinguish between distributions for which $\epsilon(Q, Q') = O\left( \min\left\{ \frac{1}{n}, \frac{1}{n'} \right\} \right)$.

One can easily imagine other scenarios, in particular for stochastic block models, where similar situations arise. The following result presents a general guarantee in this setting.

**Corollary 9 (Consistent test for semi-sparse** IER **using** $f_\lambda$**)** *Consider the setting given in Lemma 4. Then $f_\lambda$ satisfies Assumption 2 for $\widehat{\sigma}(G) = \frac{2.1}{n}\sqrt{k\widehat{D}(G)}$, where $n$ is the number of vertices in $G$ and $\widehat{D}(G)$ is its maximum degree, and $k$ is the number of largest singular values computed in $f_\lambda$. Hence, the proposed test is uniformly consistent for the testing problem with any*

$$\epsilon(Q, Q') \leq \frac{C}{\min\{n, n'\}} \text{ and } \rho(Q, Q') \geq 7.5\sqrt{k}\left( \frac{\sqrt{D_Q}}{n} + \frac{\sqrt{D_{Q'}}}{n'} \right)$$

*where $C > 0$ is any absolute constant, and $Q \in \mathcal{F}_f \cap \mathcal{F}_n$ and $Q' \in \mathcal{F}_f \cap \mathcal{F}_{n'}$.*

We note that based on Theorem 5, the allowable upper limit of $\epsilon$ could be increased, but may not be needed from a practical perspective.

## 5. Minimax optimality

The purpose of this section is to prove that the separation conditions on $\rho$ stated in Corollaries 6 and 9 are necessary for testing between the two alternatives. This implies that the proposed two-sample test is optimal when we restrict to IER graphs, and consider network statistics $f_\Delta$ or $f_\lambda$. Instead of directly stating the converse of Corollaries 6 and 9, we digress a little to study the total variation distance between a particular pair of network models.

Let $(p_n)_{n \geq 1} \subset [0, 1]$ be a sequence of probabilities, and $(\gamma_n)_{n \geq 1} \subset (0, 1)$ be a sequence of small positive values such that $\gamma_n \leq \min\{p_n, 1 - p_n\}$. We will consider two sequences of models on $\mathcal{F}_{\geq 1}$. The first one consists of ER models $(Q_n)_{n \geq 1}$, where $Q_n \in \mathcal{F}_{2n}$ with edge probability $p_n$. Note that we consider only graphs on even numbers of vertices. The second sequence, $(Q'_n)_{n \geq 1}$ consists of mixture distributions defined as follows. For each $n$, let $\ell \in \{-1, +1\}^{2n}$ be a balanced labelling of $2n$ vertices, that is, $\sum_i \ell_i = 0$. For any such labelling, there is a model such that edges between vertices of same label occur with probability $(p_n + \gamma_n)$, while edges between vertices with different labels occur with probability $(p_n - \gamma_n)$. There are $\binom{2n}{n}$ possible balanced labellings, and $Q'_n$ is an uniform mixture of the models obtained with different labellings. In other words, we generate a graph from $Q'_n$ as follows. First we randomly, and uniformly, choose any balanced labelling of $2n$ vertices, and then generate a random graph with two chosen communities, and aforementioned edge probabilities. The following result states that this pair of sequences are asymptotically indistinguishable for small $\gamma_n$.

**Theorem 10 (Distance between $Q_n$ and $Q'_n$)**   *If $\gamma_n = o_n\left(\sqrt{\frac{p_n(1-p_n)}{n}}\right)$, then the total variation distance*

$$\|Q_n - Q'_n\|_{TV} = \frac{1}{2} \sum_G \left|Q_n(G) - Q'_n(G)\right| \in o_n(1) \,,$$

*where the sum is taken over all possible graphs on $2n$ vertices.*

The above result can be of independent interest, particularly when one observes that $Q'_n$ is the uniform distribution on all stochastic block models with two balanced communities, and edge probabilities $(p_n + \gamma_n)$ and $(p_n - \gamma_n)$. For this problem, it is well known that the partitions can be identified if and only if $\gamma_n > C\sqrt{\frac{p_n(1-p_n)\ln n}{n}}$ for some constant $C > 0$. (see, for instance, Chen and Xu, 2016). Theorem 10 shows that without the $\ln n$ factor, one cannot even distinguish the planted graph from a purely random graph. The above detection boundary also matches the fundamental limit of partial recovery of communities in sparse stochastic block models (Mossel et al., 2015). Some results related to Theorem 10 can also be found in Arias-Castro and Verzelen (2014) and Chatterjee (2012) for the problems of testing or estimation of IER models, and in Carpentier and Nickl (2015) and Banks et al. (2016) for the case of signal detection.

Our interest in Theorem 10 stems from the fact that it provides a "hard" instance in the context of testing with $f_\Delta$ or $f_\lambda$. To put this in perspective, let $p_n \leq \frac{1}{2}$, and consider the two-sample testing problem, where the two graphs are generated from either of the above models. Theorem 10 implies that if $\gamma_n = o_n\left(\sqrt{\frac{p_n}{n}}\right)$, then no two-sample test can achieve a low Type-I + Type-II error rate. On the other hand, a simple calculation combined with Corollary 9 shows that the proposed test with $f_\lambda$ statistic is consistent when $\gamma_n \geq 15\sqrt{\frac{p_n}{n}}$ (see Claim 18 for details). Similarly, one can also verify that the same test in combination with $f_\Delta$ is consistent for $\gamma_n \geq 5\sqrt{\frac{p_n \ln n}{n}}$ (see Claim 16). Thus, our

10

general testing principle provides a test based on $f_\lambda$ that can detect separation near the fundamental limit of distinguishability, whereas its combination with $f_\Delta$ is only worse by a logarithmic factor.

The above discussion, when stated formally, provides the following results that, in conjunction with Corollaries 6 and 9, guarantee the minimax (near) optimality of the tests based on $f_\Delta$ and $f_\lambda$.

**Corollary 11 (Minimax separation for testing semi-sparse IER using $f_\Delta$)** *Consider the setting of Lemma 3 and Corollary 6, and without loss of generality, let $n \leq n'$. If*

$$\rho(Q, Q') = o_n\left( \sqrt{\frac{\mu(Q)}{\binom{n}{3}}} + \sqrt{\frac{\mu(Q')}{\binom{n'}{3}}} \right)$$

*for all $Q \in \mathcal{F}_f \cap \mathcal{F}_n$ and $Q' \in \mathcal{F}_f \cap \mathcal{F}_{n'}$, then for any test $\Psi$ and any $n_0$*

$$\sup_{Q \in \mathcal{F}_f \cap \mathcal{F}_{\geq n_0}} \left( \sup_{Q' \in \mathcal{B}_f(Q,\epsilon) \cap \mathcal{F}_{\geq n_0}} \mathsf{P}_{G \sim Q, G' \sim Q'}(\Psi \text{ rejects } H_0) + \right.$$

$$\left. \sup_{Q' \in \overline{\mathcal{B}_f(Q,\rho)} \cap \mathcal{F}_{\geq n_0}} \mathsf{P}_{G \sim Q, G' \sim Q'}(\Psi \text{ accepts } H_0) \right) = 1 \ .$$

To put it simply, if the separation condition is Corollary 6 is not satisfied (ignoring logarithmic difference), then no matter how large the graphs are, one cannot hope to achieve a bounded error rate with any two-sample test. The corresponding optimality result for $f_\lambda$ with $k = 2$ is stated below.

**Corollary 12 (Minimax separation for testing semi-sparse IER using $f_\lambda$)** *Consider the setting of Lemma 4 and Corollary 9, where we consider only the largest two singular values, that is, $k = 2$. Let*

$$\rho(Q, Q') = o_n\left( \frac{\sqrt{D_Q}}{n} + \frac{\sqrt{D_{Q'}}}{n'} \right)$$

*for all $Q \in \mathcal{F}_f \cap \mathcal{F}_n$ and $Q' \in \mathcal{F}_f \cap \mathcal{F}_{n'}$, and $\epsilon(Q, Q')$ be bounded as in Corollary 9. Then for any test $\Psi$ and any $n_0$*

$$\sup_{Q \in \mathcal{F}_f \cap \mathcal{F}_{\geq n_0}} \left( \sup_{Q' \in \mathcal{B}_f(Q,\epsilon) \cap \mathcal{F}_{\geq n_0}} \mathsf{P}_{G \sim Q, G' \sim Q'}(\Psi \text{ rejects } H_0) + \right.$$

$$\left. \sup_{Q' \in \overline{\mathcal{B}_f(Q,\rho)} \cap \mathcal{F}_{\geq n_0}} \mathsf{P}_{G \sim Q, G' \sim Q'}(\Psi \text{ accepts } H_0) \right) = 1 \ .$$

Observe that the above result is sharp since it precisely matches the sufficient condition of Corollary 9. We feel that a similar result can be shown for any $k > 2$ by using a generalisation of Theorem 10 for $k$ community models.

## 6. Proofs

Here, we sequentially present proofs of all the results stated in the previous sections.

**Proof of Lemma 1**

Observe that $\binom{n}{3}f_\Delta(G) = \sum\limits_{i<j<k} \Delta_{ijk}$, and

$$\mathsf{Var}_{G\sim Q}\left(\binom{n}{3}f_\Delta(G)\right) = \sum_{i'<j'<k'}\sum_{i<j<k}\left(\mathsf{E}_{G\sim Q}[\Delta_{ijk}\Delta_{i'j'k'}] - \mathsf{E}_{G\sim Q}[\Delta_{ijk}]\mathsf{E}_{G\sim Q}[\Delta_{i'j'k'}]\right) \ .$$

Consider the inner sum for $(i',j',k') = (1,2,3)$. Under the condition that non-overlapping triangles are uncorrelated, one can see that the terms where at most one of $i,j,k$ is in $\{1,2,3\}$ are zero. So

$$\sum_{i<j<k}\left(\mathsf{E}_{G\sim Q}[\Delta_{ijk}\Delta_{123}] - \mathsf{E}_{G\sim Q}[\Delta_{ijk}]\mathsf{E}_{G\sim Q}[\Delta_{123}]\right)$$

$$= \left(\mathsf{E}_{G\sim Q}[\Delta_{123}] - (\mathsf{E}_{G\sim Q}[\Delta_{123}])^2\right) + \sum_{\substack{1\le i<j\le 3\\k>3}}\left(\mathsf{E}_{G\sim Q}[\Delta_{ijk}\Delta_{123}] - \mathsf{E}_{G\sim Q}[\Delta_{ijk}]\mathsf{E}_{G\sim Q}[\Delta_{123}]\right)$$

$$\le \mathsf{E}_{G\sim Q}[\Delta_{123}] + \sum_{\substack{1\le i<j\le 3\\k>3}}\mathsf{E}_{G\sim Q}[\Delta_{ijk}\Delta_{123}]$$

$$= \mathsf{E}_{G\sim Q}[\Delta_{123}] + \sum_{k>3}\mathsf{E}_{G\sim Q}\left[\Delta_{123}\left((A_G)_{1k}(A_G)_{2k} + (A_G)_{2k}(A_G)_{3k} + (A_G)_{1k}(A_G)_{3k}\right)\right]$$

$$\le \mathsf{E}_{G\sim Q}[\Delta_{123}] + \sum_{k>3}\mathsf{E}_{G\sim Q}\left[\Delta_{123}\left((A_G)_{1k} + (A_G)_{2k} + (A_G)_{3k}\right)\right] \ . \tag{7}$$

We now use the fact that $\Delta_{123}$ is not correlated with any non-overlapping edge to decompose the expectation. Subsequently, summing over $k$ shows that the above quantity can be bounded from above by $\mathsf{E}_{G\sim Q}[\Delta_{123}](1 + 3D_Q)$, where $D_Q = \max_i\sum_k \mathsf{E}[(A_G)_{ik}]$ is the maximum expected degree. Summing over all $i',j',k'$ leads to the conclusion

$$\mathsf{Var}_{G\sim Q}\left(\binom{n}{3}f_\Delta(G)\right) \le \binom{n}{3}\mu(Q)(1+3D_Q) \ ,$$

and by Chebyshev's inequality,

$$\mathsf{P}_{G\sim Q}(|f_\Delta(G) - \mu(Q)| > \sigma(Q)) \le \frac{(3D_Q+1)\binom{n}{3}\mu(Q)}{\binom{n}{3}^2\sigma(Q)^2} = \frac{1}{\ln n} = o_n(1) \ .$$

**Proof of Lemma 2**

This result follows by observing that in a random geometric graph two non-overlapping triangles or edges are uncorrelated. Also, $D_Q = (n-1)p_Q < np_Q$.

**Proof of Lemma 3**

We may proceed similar to the proof of Lemma 1 up to the last but one step of (7), which can be written as

$$\mathsf{E}_{G\sim Q}[\Delta_{123}] + \mathsf{E}_{G\sim Q}[\Delta_{123}]\sum_{k>3}\left((M_Q)_{1k}(M_Q)_{2k} + (M_Q)_{2k}(M_Q)_{3k} + (M_Q)_{1k}(M_Q)_{3k}\right) \ ,$$

and is smaller that $4\mathsf{E}_{G\sim Q}[\Delta_{123}]$ since $Q \in \mathcal{F}_f$ defined in Lemma 3. Thus, $\mathsf{Var}_{G\sim Q}\left(\binom{n}{3}f_\Delta(G)\right) \le 4\mu(Q)\binom{n}{3}$, and the result follows due Chebyshev's inequality and the stated choice of $\sigma$.

**Proof of Lemma 4**

Let $M_Q$ and $D_Q$ in the statement of the lemma, and let $d$ be the Euclidean metric in $\mathbb{R}^k$. Then, from Weyl's inequality, we have

$$d(f_\lambda(G), \mu(Q)) = \frac{1}{n}\sqrt{\sum_{i \leq k}(\lambda_i(A_G) - \lambda_i(M_Q))^2} \leq \frac{\sqrt{k}}{n}\|A_G - M_Q\| \, ,$$

where $\|\cdot\|$ is the spectral norm. Lu and Peng (2013) show that if $D_Q \geq C(\ln n)^4$ holds for a large constant $C > 0$, then $\|A_G - M_Q\| \leq (2 + \delta)\sqrt{D_Q}$ with probability $1 - o_n(1)$ for any $\delta > 0$. This immediately leads to the choices of $\mathcal{F}_f$ and $\sigma(Q)$ stated in the lemma.

**Proof of Theorem 5**

Let $Q \in \mathcal{F}_f$ and $Q' \in \mathcal{B}_f(Q, \epsilon)$, where $\epsilon(Q, Q') \leq 0.5(\sigma(Q) + \sigma(Q'))$. Observe that

$$\begin{aligned}
&\mathsf{P}_{G\sim Q, G'\sim Q'}(T(G, G') > 1) \\
&= \mathsf{P}_{G\sim Q, G'\sim Q'}\big(d(f(G), f(G')) > 2(\widehat{\sigma}(G) + \widehat{\sigma}(G'))\big) \\
&\leq \mathsf{P}_{G\sim Q, G'\sim Q'}\big(d(f(G), f(G')) > 1.5(\sigma(G) + \sigma(G'))\big) \\
&\quad + \mathsf{P}_{G\sim Q, G'\sim Q'}\big(1.5(\sigma(G) + \sigma(G')) > 2(\widehat{\sigma}(G) + \widehat{\sigma}(G'))\big) \\
&\leq \mathsf{P}_{G\sim Q, G'\sim Q'}\big(d(f(G), f(G')) > 1.5(\sigma(G) + \sigma(G'))\big) \\
&\quad + \mathsf{P}_{G\sim Q}\big(\sigma(G) - \widehat{\sigma}(G) > \tfrac{1}{4}\sigma(G)\big) + \mathsf{P}_{G'\sim Q'}\big(\sigma(G') - \widehat{\sigma}(G') > \tfrac{1}{4}\sigma(G')\big) \, .
\end{aligned}$$

Due to Assumption 2, the second and third terms in the bound are both $o_n(1)$. To bound the the first term, we note that

$$\begin{aligned}
d(f(G), f(G')) &\leq d(f(G), \mu(Q)) + d(f(G'), \mu(Q')) + d(\mu(Q), \mu(Q')) \\
&\leq d(f(G), \mu(Q)) + d(f(G'), \mu(Q')) + 0.5(\sigma(Q) + \sigma(Q')) \, .
\end{aligned}$$

Using this fact, we have

$$\begin{aligned}
\mathsf{P}_{G\sim Q, G'\sim Q'}&\big(d(f(G), f(G')) > 1.5(\sigma(G) + \sigma(G'))\big) \\
&\leq \mathsf{P}_{G\sim Q, G'\sim Q'}\big(d(f(G), \mu(Q)) + d(f(G'), \mu(Q')) > \sigma(G) + \sigma(G')\big) \\
&\leq \mathsf{P}_{G\sim Q}\big(d(f(G), \mu(Q)) > \sigma(G)\big) + \mathsf{P}_{G'\sim Q'}\big(d(f(G'), \mu(Q')) > \sigma(G')\big) \, ,
\end{aligned}$$

where both terms in the bounds are $o_n(1)$ due to Assumption 1. Thus, we have established that $\mathsf{P}_{G\sim Q, G'\sim Q'}(T(G, G') > 1) \in o_n(1)$.

We tackle the other case, that is $Q' \in \overline{\mathcal{B}_f(Q, \rho)}$, in a similar way. Here, we have

$$\begin{aligned}
\mathsf{P}_{G\sim Q, G'\sim Q'}&(T(G, G') \leq 1) \\
&\leq \mathsf{P}_{G\sim Q, G'\sim Q'}\big(d(f(G), f(G')) \leq 2.5(\sigma(G) + \sigma(G'))\big) \\
&\quad + \mathsf{P}_{G\sim Q, G'\sim Q'}\big(2.5(\sigma(G) + \sigma(G')) \leq 2(\widehat{\sigma}(G) + \widehat{\sigma}(G'))\big) \, ,
\end{aligned}$$

where the second term can be shown to be $o_n(1)$ using Assumption 2. For the first term, we recall $\rho(Q, Q') \geq 3.5(\sigma(Q) + \sigma(Q'))$, and use triangle inequality in the opposite direction, that is,

$$\begin{aligned}
3.5(\sigma(Q) + \sigma(Q')) &\leq d(\mu(Q), \mu(Q')) \\
&\leq d(f(G), \mu(Q)) + d(f(G'), \mu(Q')) + d(f(G), f(G'))
\end{aligned}$$

Hence, the first term can be bounded as

$$\mathsf{P}_{G\sim Q, G'\sim Q'}\big(d(f(G), f(G')) \leq 2.5(\sigma(G) + \sigma(G')))\big)$$
$$\leq \mathsf{P}_{G\sim Q, G'\sim Q'}\big(d(f(G), \mu(Q)) + d(f(G'), \mu(Q')) > \sigma(G) + \sigma(G')\big)\ ,$$

which is again $o_n(1)$, thereby leading to a similar conclusion for Type-II error.

## Proof of Corollary 6

Note that we only need to show that $f_\Delta$ satisfies Assumption 2 for $\widehat{\sigma}(G) = 2\sqrt{f_\Delta(G)\ln n/\binom{n}{3}}$. The rest of the claim is an immediate consequence of Theorem 5. To prove the first part, we deal with the upper and lower tail separately. For the upper tail, we observe that for any $\delta > 0$,

$$\mathsf{P}_{G\sim Q}(\widehat{\sigma}(G) > (1+\delta)\sigma(Q)) \leq \mathsf{P}_{G\sim Q}\left(f_\Delta(G) > (1+\delta)^2\mu(Q)\right)$$
$$\leq \mathsf{P}_{G\sim Q}\left(f_\Delta(G) - \mu(Q) > \delta\mu(Q)\right)$$
$$\leq \frac{\mathsf{Var}_{G\sim Q}(f_\Delta(G))}{\delta^2\mu^2(Q)} \in o_n(1)\ .$$

The last step uses the bound on $\mathsf{Var}_{G\sim Q}(f_\Delta(G))$ derived in proof of Lemma 3 and the fact $\mu(Q) \geq \frac{\ln n}{n^3}$. We can similarly bound the lower tail probability.

## Proof of Corollary 7

As in the previous proof, we only need to show that $f_\Delta$ satisfies Assumption 2 in this setting, where $\widehat{\sigma}(G) = \sqrt{(3n\widehat{p}(G) + 1)f_\Delta(G)\ln n/\binom{n}{3}}$. We bound the upper tail probability as

$$\mathsf{P}_{G\sim Q}(\widehat{\sigma}(G) > (1+\delta)\sigma(Q)) \leq \mathsf{P}_{G\sim Q}\left((3n\widehat{p}(G) + 1)f_\Delta(G) > (1+\delta)^2(3np_Q + 1)\mu(Q)\right)$$
$$\leq \mathsf{P}_{G\sim Q}\left((3n\widehat{p}(G) + 1) > (1+\delta)(3np_Q + 1)\right) + \mathsf{P}_{G\sim Q}\left(f_\Delta(G) > (1+\delta)\mu(Q)\right)$$

using union bound. Observe that the first term is at most

$$\mathsf{P}_{G\sim Q}\left(\widehat{p}(G) > (1+\delta)p_Q\right) \leq \frac{\mathsf{Var}_{G\sim Q}(\widehat{p}(G))}{\delta^2 p_Q^2} \leq \frac{1}{\delta^2 p_Q\binom{n}{2}} = o_n(1)\ .$$

Here, we use Chebyshev's inequality followed by the fact that $\mathsf{Var}_{G\sim Q}(\widehat{p}(G)) = \binom{n}{2}^{-1}p(1-p)$ since the edges of the geometric graph are still pairwise independent. Finally, we use the condition $p_Q \geq \frac{1}{n}$ for all $Q \in \mathcal{F}_f$ stated in Lemma 2.

For the second term, we use the variance bound stated in proof of Lemma 2 to write

$$\mathsf{P}_{G\sim Q}(f_\Delta(G) > (1+\delta)\mu(Q)) \leq \frac{\mathsf{Var}_{G\sim Q}(f_\Delta(G))}{(\delta\mu(Q))^2} \leq \frac{3np_Q + 1}{\delta^2\binom{n}{3}\mu(Q)}\ . \tag{8}$$

Bubeck et al. (2016, Lemma 1) showed that if $p_Q \geq \frac{1}{4}$ and $r_Q \geq C_p$ for some $C_p > 0$ depending on $p_Q$, then $\mu(Q) \geq p_Q^3$. Taking $C = \sup\limits_{1/4 \leq p \leq 1} C_p$, we can say that if $p_Q \geq \frac{1}{4}$ and $r_Q \geq C$, then $\mu(Q) \geq p_Q^3$. Note that this gives rise to the $C$ mentioned in Lemma 2. Hence, under this regime, the bound in (8) is $o_n(1)$.

For $p_Q < \frac{1}{4}$, which includes the case of $p_Q$ decaying with $n$, Bubeck et al. (2016) showed that $\mu(Q) \geq cp_Q^3(\ln \frac{1}{p_Q})^{3/2}r_Q^{-1/2}$ for some absolute constant $c > 0$. In this case, one can verify that if $np_Q \geq 1$ and the upper bound on $r_Q$ (see Lemma 2) hold, then the bound in (8) is also $o_n(1)$. This takes care of the upper tail probability. For the other part, we have

$$\mathsf{P}_{G\sim Q}(\widehat{\sigma}(G) < (1-\delta)\sigma(Q)) \leq \mathsf{P}_{G\sim Q}\left(\widehat{p}(G) < (1-\delta)p_Q\right) + \mathsf{P}_{G\sim Q}\left(f_\Delta(G) < (1-\delta)\mu(Q)\right),$$

where the terms can be again bounded as above to show that the probability is $o_n(1)$.

**Proof of Corollary 8**

Let $\sigma_E$ and $\sigma_G$ be the deviations defined in Lemmas 3 and 2, respectively. A natural choice for $\sigma$ in combined setting is $\sigma = \max\{\sigma_E, \sigma_G\}$, which can also be estimated accurately from the random graph. In other words, $f_\Delta$ satisfies Assumption 2 in the present setting as well. Also note that $p_n$ decays at the appropriate rate so that the ER graph lies is $\mathcal{F}_f$ defined in Lemma 3. Also the decaying $p_n$ corresponds to the sparser regime discussed in proof of Corollary 7.

Let $Q, Q'$ be the distributions of the two random graphs. If $Q = Q'$, then Theorem 5 directly provides a bound on Type-I error rate as $\epsilon(Q, Q') = 0$. Thus, we only need to check the case, where $Q \neq Q'$, or more precisely, we need to verify that if $r_n = o_n\left((\ln \frac{1}{p_n})^3\right)$, then we get $\rho(Q, Q') > 3.5(\sigma(Q) + \sigma(Q'))$. Assume that $Q$ corresponds to ER graph, and $Q$ is the Geom graph, that is, $\mu(Q') = p_n^3$ and $\mu(Q) \geq cp_n^3(\ln \frac{1}{p_n})^{3/2}r_n^{-1/2}$. It is now easy to verify that under prescribed condition on $r_n$, all $\mu(Q'), \sigma(Q')$ and $\sigma(Q)$ are $o_n(\mu(Q))$. Hence, the separation condition is eventually satisfied.

**Proof of Corollary 9**

Following the proof of Corollary 6, note that we only need to prove concentration of $\widehat{\sigma}$. The rest follows from Theorem 5. The only difference lies in the stated upper bound on $\epsilon$, which follows from the conditions $D_Q \geq (\ln n)^{4.1}$ and $D_{Q'} \geq (\ln n')^{4.1}$.

We now show that $\widehat{\sigma}(G) = \frac{2.1}{n}\sqrt{k\widehat{D}(G)}$ concentrates about $\sigma(Q)$. Recall that $Q \in \mathsf{IER} \cap \mathcal{F}_n$ is characterised by the matrix $M_Q \in [0,1]^{n\times n}$, and $D_Q$ is the maximum row sum of $M_Q$. Let us denote the row sums of $M_Q$ by $D_1, D_2, \ldots, D_n$. Also for $G \sim Q$, denote the degree of vertex-$i$ by $\deg_i$. Hence, we have $\widehat{D}(G) = \max_i \deg_i$, and $\mathsf{E}_{G\sim Q}[\deg_i] = D_i$. For $\delta \in (0,1)$, we bound the upper tail probability by

$$\mathsf{P}_{G\sim Q}\left(\widehat{\sigma}(G) > (1+\delta)\sigma(Q)\right) \leq \mathsf{P}_{G\sim Q}\left(\widehat{D}(G) > (1+\delta)D_Q\right)$$
$$= \mathsf{P}_{G\sim Q}\left(\bigcup_i \{\deg_i > (1+\delta)D_Q\}\right)$$
$$\leq \sum_i \mathsf{P}_{G\sim Q}\left(\deg_i > (1+\delta)D_Q\right) \qquad (9)$$

using the union bound. Consider the cases $D_i \geq (\ln n)^{1.1}$ and $D_i < (\ln n)^{1.1}$ separately. In the former case,

$$\mathsf{P}_{G\sim Q}\left(\deg_i > (1+\delta)D_Q\right) \leq \mathsf{P}_{G\sim Q}\left(\deg_i - D_i > \delta D_i\right) \leq e^{-\delta^2 D_i/3},$$

where we use the Bernstein inequality at the last step. Since, $D_i \geq (\ln n)^{1.1} \geq \frac{6}{\delta^2} \ln n$ for large $n$, the above probability is bounded by $\frac{1}{n^2}$. On the other hand, if $D_i \leq (\ln n)^{1.1}$, we use the Markov inequality to write

$$
\begin{aligned}
\mathsf{P}_{G \sim Q} \left( \deg_i > (1+\delta)D_Q \right) &\leq \mathsf{P}_{G \sim Q} \left( e^{\deg_i} > e^{D_Q} \right) \\
&\leq e^{-D_Q} \prod_j \mathsf{E}_{G \sim Q} \left[ e^{(A_G)_{ij}} \right] ; , \\
&\leq e^{-D_Q} \prod_j \left( 1 + e(M_Q)_{ij} \right) \leq e^{eD_i - D_Q} ,
\end{aligned}
$$

where we use independence of the edges for second inequality, and the fact $(1+x) \leq e^x$ in the last step. Since $D_i < (\ln n)^{1.1}$ and $D_Q \geq (\ln n)^{4.1}$, the above bound is eventually smaller than $\frac{1}{n^2}$. From above arguments, we can see that each term in (9) is at most $\frac{1}{n^2}$, and hence the sum is $o_n(1)$.

To prove the lower tail bound, assume without loss of generality that $D_Q = D_1$. We can see

$$
\begin{aligned}
\mathsf{P}_{G \sim Q} \left( \widehat{\sigma}(G) < (1-\delta)\sigma(Q) \right) &\leq \mathsf{P}_{G \sim Q} \left( \widehat{D}(G) < (1-\delta)D_Q \right) \\
&= \mathsf{P}_{G \sim Q} \left( \bigcap_i \{ \deg_i < (1-\delta)D_Q \} \right) \\
&\leq \mathsf{P}_{G \sim Q} \left( \deg_1 < (1-\delta)D_Q \right) \\
&= \mathsf{P}_{G \sim Q} \left( D_1 - \deg_1 > \delta D_1 \right) \leq e^{-\delta^2 D_1/3} ,
\end{aligned}
$$

where the last bound is due to Bernstein inequality, and is $o_n(1)$ under the condition on $D_Q$. Hence, $f_\lambda$ satisfies Assumption 2, and the claim follows.

**Proof of Theorem 10**

For ease of notation, we drop the subscript $n$ from $Q_n, Q'_n, p_n$ and $\gamma_n$, but we recall that the subscript $n$ corresponds to graphs of size $2n$, and in the case of $Q'_n$, there are exactly $n$ vertices labelled $+1$, and the rest $-1$. We first define the quantity

$$
L(Q, Q') = \sum_G \frac{(Q'(G))^2}{Q(G)} ,
$$

and use Cauchy-Schwarz inequality to bound the total variation distance by

$$
\|Q - Q'\|_{TV} \leq \frac{1}{2} \sqrt{\sum_G Q(G)} \sqrt{\sum_G Q(G) \left( \frac{Q'(G)}{Q(G)} - 1 \right)^2} = \frac{1}{2} \sqrt{L(Q, Q') - 1} ,
$$

where we use the fact $\sum_G Q(G) = \sum_G Q'(G) = 1$. If we can show $L(Q, Q') \leq e^{o_n(1)}$, then the abound bound implies $\|Q - Q'\|_{TV} \in o_n(1)$, which is the claim. Hence, the rest of the proof is about deriving the stated upper bound for $L(Q, Q')$. For this, observe that given a balanced labelling $\ell \in \{-1, +1\}^{2n}$, the conditional distribution of $Q'$ is given by

$$
Q'(G|\ell) = \prod_{i<j} (p + \gamma \ell_i \ell_j)^{(A_G)_{ij}} (1 - p - \gamma \ell_i \ell_j)^{1-(A_G)_{ij}} ,
$$

16

where $A_G$ is the adjacency matrix of $G$. As a consequence, the distribution $Q'$ is given by

$$Q'(G) = \frac{1}{\binom{2n}{n}} \sum_{\ell} \prod_{i<j} (p + \gamma\ell_i\ell_j)^{(A_G)_{ij}} (1 - p - \gamma\ell_i\ell_j)^{(A_G)_{ij}} \,,$$

where the sum is over all balanced labellings. Hence, one can compute $L(Q, Q')$ as

$$L(Q, Q') =$$

$$\sum_G \frac{1}{\binom{2n}{n}^2} \sum_{\ell,\ell'} \prod_{i<j} \left( \frac{(p + \gamma\ell_i\ell_j)(p + \gamma\ell'_i\ell'_j)}{p} \right)^{(A_G)_{ij}} \left( \frac{(1 - p - \gamma\ell_i\ell_j)(1 - p - \gamma\ell'_i\ell'_j)}{1 - p} \right)^{1 - (A_G)_{ij}}$$

Interchanging the sums and summing the products over all possible graphs $G$, we obtain

$$L(Q, Q') = \frac{1}{\binom{2n}{n}^2} \sum_{\ell,\ell'} \prod_{i<j} \left( 1 + \frac{\gamma^2}{p(1-p)} \ell_i\ell_j\ell'_i\ell'_j \right) \,.$$

Note that due to the symmetric nature of $\ell$ and $\ell'$, if we fix an $\ell'$ and sum over $\ell$, then the sum remains same irrespective of the value of $\ell'$. Hence, we may fix $\ell'$ to the vector with the first $n$ coordinates as $+1$, and the rest $-1$, and consider the average only over $\ell$, that is,

$$L(Q, Q') = \frac{1}{\binom{2n}{n}} \sum_{\ell} \prod_{\substack{i<j\leq n \\ n<i<j}} \left( 1 + \frac{\gamma^2}{p(1-p)} \ell_i\ell_j \right) \prod_{i\leq n<j} \left( 1 - \frac{\gamma^2}{p(1-p)} \ell_i\ell_j \right)$$

$$\leq \frac{1}{\binom{2n}{n}} \sum_{\ell} \exp \left( \sum_{\substack{i<j\leq n \\ n<i<j}} \frac{\gamma^2}{p(1-p)} \ell_i\ell_j - \sum_{i\leq n<j} \frac{\gamma^2}{p(1-p)} \ell_i\ell_j \right)$$

$$= \mathsf{E}_\ell \left[ \exp \left( \sum_{\substack{i<j\leq n \\ n<i<j}} \frac{\gamma^2}{p(1-p)} \ell_i\ell_j - \sum_{i\leq n<j} \frac{\gamma^2}{p(1-p)} \ell_i\ell_j \right) \right] \,.$$

The inequality follows from the relation $1 + x \leq e^x$, and in the subsequent step, we view $\ell$ as a random labelling chosen uniformly from all balanced labellings. We define the quantity $S_k = \sum_{i=1}^{k} \ell_i$, and observe that $S_{2n} = 0$, which also implies $\sum_{j>n} \ell_j = -S_n$. Using this relation, the above bound simplifies to

$$L(Q, Q') \leq \exp \left( -\frac{\gamma^2 n}{p(1-p)} \right) \mathsf{E}_{\ell_1,\dots,\ell_n} \left[ \exp \left( \frac{2\gamma^2}{p(1-p)} S_n^2 \right) \right] \tag{10}$$

We now observe that $\ell_1, \dots, \ell_{2n}$ are conditional Bernoulli's with the constraint $S_{2n} = 0$, or exactly $n$ of them can be one. Hence, they can be generated using the following procedure.

- $\ell_1$ takes values $+1$ or $-1$ with equal probability, and

- for $k \geq 1$, if $N_k$ is the number of $+1$'s observed in the first $k$ coordinates, then $\ell_{k+1}$ takes value $+1$ with probability $\frac{n - N_k}{2n - k}$.

17

Based on this observation, we make the following claim, that we prove at the end of this section.

**Claim 13** *For any $1 \leq k < n$ and any $x > 0$ such that $|2xS_k| \leq 1$, we have*

$$\mathsf{E}_{\ell_{k+1}|\ell_1,\ldots,\ell_k}\left[\exp\left(xS_{k+1}^2\right)\right] \leq \exp(x)\exp\left((x+2ex^2)S_k^2\right)$$

Now, define the sequence $(x_k)_{k=0,\ldots,n-1}$ such that $x_0 = \frac{2\gamma^2}{p(1-p)}$, and $x_{k+1} = x_k + 2ex_k^2$ for $k \geq 0$. Observe that under the condition on $\gamma$ in Theorem 10, we eventually have that $x_0 \in [0, \frac{1}{8en}]$, and due to this, the sequence $(x_k)_k$ satisfies the following property.

**Claim 14** *If $x_0 \in [0, \frac{1}{8en}]$ and $x_{k+1} = (x_k + 2ex_k^2)$ for $k = 0,\ldots,n-2$, then*

$$x_k \leq \left(1 + \frac{k}{n-1}\right)x_0 \leq 2x_0$$

*for every $k = 0,\ldots,\leq n-1$. As a consequence, $\sum_{k=0}^{n-1} x_k \leq 2nx_0$.*

We return to (10), and use Claim 13 to bound $L(Q,Q')$ iteratively as

$$\begin{aligned}
L(Q,Q') &\leq \exp(-nx_0/2)\mathsf{E}_{\ell_1,\ldots,\ell_{n-1}}\left[\mathsf{E}_{\ell_n|\ell_1,\ldots,\ell_{n-1}}\left[\exp\left(x_0S_n^2\right)\right]\right] \\
&\leq \exp(-nx_0/2)\exp(x_0)\mathsf{E}_{\ell_1,\ldots,\ell_{n-2}}\left[\mathsf{E}_{\ell_{n-1}|\ell_1,\ldots,\ell_{n-2}}\left[\exp\left(x_1S_{n-1}^2\right)\right]\right] \\
&\leq \exp(-nx_0/2)\exp\left(\sum_{i=0}^{k-1}x_i\right)\mathsf{E}_{\ell_1,\ldots,\ell_{n-k-1}}\left[\mathsf{E}_{\ell_{n-k}|\ell_1,\ldots,\ell_{n-k-1}}\left[\exp\left(x_kS_{n-k}^2\right)\right]\right] \\
&\leq \exp(-nx_0/2)\exp\left(\sum_{i=0}^{n-2}x_i\right)\mathsf{E}_{\ell_1}\left[\exp\left(x_{n-1}S_1^2\right)\right] \;,
\end{aligned}$$

where the condition in Claim 13 is satisfied at every step since $|2x_kS_{n-k-1}| \leq 4x_0|S_n| \leq 1$ as we have observed that $x_0 \leq \frac{1}{8en}$ eventually for large $n$. Finally, note that $S_1^2 = \ell_1^2 = 1$, and hence, using Claim 14 we have the bound

$$L(Q,Q') \leq e^{2nx_0} = e^{4n\gamma^2/p(1-p)} = e^{o_n(1)}$$

for $\gamma = o_n\left(\sqrt{\frac{p(1-p)}{n}}\right)$. Thus, we have the stated result. We conclude this proof with the proof of the two intermediate claims.

**Proof of Claim 13.** Recall the generation process for $\ell_k$, and observe the $S_k = N_k - (k - N_k) = 2N_k - k$. Hence, for every $k \geq 0$, the Bernoulli variable $\ell_{k+1}$ takes the value $+1$ with probability $\frac{n-N_k}{2n-k} = \frac{1}{2} - \frac{S_k}{2(2n-k)}$, and the value $-1$ with probability $\frac{1}{2} + \frac{S_k}{2(2n-k)}$. We evaluate the conditional expectation as follows

$$\begin{aligned}
\mathsf{E}_{\ell_{k+1}|\ell_1,\ldots,\ell_k}\left[\exp\left(xS_{k+1}^2\right)\right] &= \mathsf{E}_{\ell_{k+1}|\ell_1,\ldots,\ell_k}\left[\exp\left(x(S_k^2 + 2S_k\ell_{k+1} + 1)\right)\right] \\
&= \exp\left(x + xS_k^2\right)\mathsf{E}_{\ell_{k+1}|\ell_1,\ldots,\ell_k}\left[\exp\left(2xS_k\ell_{k+1}\right)\right] \\
&= \exp\left(x + xS_k^2\right)\left(\exp(2xS_k)\left(\frac{1}{2} - \frac{S_k}{2(2n-k)}\right) + \exp(-2xS_k)\left(\frac{1}{2} + \frac{S_k}{2(2n-k)}\right)\right) \\
&= \exp\left(x + xS_k^2\right)\left(\cosh(2xS_k) - \frac{S_k}{(2n-k)}\sinh(2xS_k)\right) \;.
\end{aligned}$$

18

One can verify that $z\sinh(z)$ is always positive, while $\cosh(z) \le 1 + \frac{ez^2}{2} \le \exp\left(\frac{ez^2}{2}\right)$ for all $|z| \le 1$. As a consequence, the second term in above expression is positive, and can be ignored for an upper bound, whereas the first term is at most $\exp(2ex^2 S_k^2)$. Hence, the claim. ∎

**Proof of Claim 14.** We prove the claim by induction. Assume $x_k \le \left(1 + \frac{k}{n-1}\right)x_0$. Then

$$x_{k+1} = x_k + 2ex_k^2$$
$$\le x_0\left(1 + \frac{k}{n-1}\right) + 2ex_0^2\left(1 + \frac{k}{n-1}\right)^2$$
$$\le x_0\left(1 + \frac{k}{n-1}\right) + \frac{1}{4(n-1)} \cdot 4x_0$$
$$= \left(1 + \frac{k+1}{n-1}\right)x_0 ,$$

where the second inequality uses the facts $2ex_0 \le \frac{1}{4(n-1)}$ and $(1 + \frac{k}{n-1}) \le 2$ for $k < n$. ∎

**Proof of Corollary 11**

We begin by noting that both sequence of models in Theorem 10 belong to $\mathcal{F}_f$ if the sequences $p_n, \gamma_n$ satisfy $\frac{\ln 2n}{2n} \le p_n - \gamma_n < p_n + \gamma_n \le \frac{1}{\sqrt{2n}}$. As in Theorem 10, let $Q_n$ and $Q'_n$ respectively denote the ER and the labelled graph models. Due to condition on $\rho$ given in statement of Corollary 11, we can define a sequence $t_n \in o_n(1)$ such that

$$\rho(Q_n, Q'_n) = t_n \cdot \left(\sqrt{\frac{\mu(Q_n)}{n^3}} + \sqrt{\frac{\mu(Q'_n)}{n^3}}\right) , \tag{11}$$

where $Q_n, Q'_n$ are the above mentioned distributions. We make the following claim.

**Claim 15** *Let $p_n \in [\frac{\ln 2n}{2n}, \frac{1}{\sqrt{2n}}]$. There exists $\tau_n \in o_n(1)$ that satisfies both the following conditions: (i) $\gamma_n = \tau_n\sqrt{\frac{p_n(1-p_n)}{n}}$, and (ii) $Q'_n \in \overline{\mathcal{B}_f(Q_n, \rho)}$ where $\rho$ is given by (11).*

The above claim shows that, in the present scenario, we can define the sequences $Q_n, Q'_n$ such that $\|Q_n - Q'_n\|_{TV} \in o_n(1)$, and yet the pair satisfies the alternative hypothesis $H_1$. This is the trick we use to prove that no non-trivial two-sample test that can distinguish between these two models.

At this stage, we follow the proof technique of Collier (2012) for lower bounding the error rate of a two-sample test by the error rate of a suitably defined one-sample test. In our case, the one-sample test is: Given a graph $G$, identify whether $G$ is sampled from $Q_n$ or $Q'_n$. Now, let $\Psi$ be any test for our two-sample testing problem. One can use $\Psi$ to construct a one-sample test for the above problem by comparing $G$ with a randomly generated graph from $Q_n$. Collier (2012) argues that this construction leads to a test whose Type-I and Type-II errors are both smaller than that of the of the two-sample test $\Psi$. Subsequently, using a standard testing lower bound (see Baraud, 2002, Section 7.1), one can lower bound the total error rate by $1 - \|Q_n - Q'_n\|_{TV}$. Since, our $\mathcal{F}_f$ contain $Q_n, Q'_n$

for all large $n \geq n_0$, hence we can choose $n$ large enough to get the lower bound arbitrarily close to 1. Hence, we get the claimed supremum of 1 for any test $\Psi$.

While above arguments conclude the proof of Corollary 11, for completeness, we also add the proof of the fact that for larger $\gamma_n$, the proposed two-sample test can indeed distinguish between $Q_n$ and $Q'_n$. In particular, let $\rho(Q_n, Q'_n) = 7\sqrt{\frac{\ln 2n}{\binom{2n}{3}}}(\sqrt{\mu(Q_n)} + \sqrt{\mu(Q'_n)})$. We prove the following.

**Claim 16** *Let $p_n \in [\frac{\ln 2n}{2n}, \frac{1}{\sqrt{2n}}]$. If $\gamma_n \geq 5\sqrt{\frac{p_n \ln n}{n}}$, then $Q'_n \in \overline{\mathcal{B}_f(Q_n, \rho)}$ for above mentioned $\rho$. As a consequence, Corollary 6 implies that proposed test consistently distinguishes $Q_n$ from $Q'_n$.*

We now prove the claims.

**Proof of Claim 15.** For convenience, we drop the subscript $n$. One can easily verify that for the specified models and network statistic $f_\Delta$, $\mu(Q) = p^3$ and $\mu(Q') = p^3 + \gamma^3 - \frac{3}{2n-1}(p^2\gamma + p\gamma^2)$. Since $p \in o_n(1)$, we can abuse our notation to write $\gamma = \tau\sqrt{\frac{p}{n}}$ dropping the factor of $(1-p)$. Now assume $\tau > \sqrt{12p}$, then we compute

$$|\mu(Q) - \mu(Q')| = \left| \frac{\tau^3 p^{3/2}}{n^{3/2}} - \frac{3\tau p^{5/2}}{(2n-1)n^{1/2}} - \frac{3\tau^2 p^2}{(2n-1)n} \right| > \frac{\tau^3 p^{3/2}}{2n^{3/2}} \, ,$$

using the fact that $\tau > \sqrt{12p}$ ensures that the second and third terms are smaller that $\frac{1}{4}$ of the first term. On the other hand, we can bound $\rho$ in (11) from above by

$$\rho(Q, Q') = \frac{t}{n^{3/2}} \left( p^{3/2} + \sqrt{p^3 + \frac{\tau^3 p^{3/2}}{n^{3/2}} - \frac{3\tau p^{5/2}}{(2n-1)n^{1/2}} - \frac{3\tau^2 p^2}{(2n-1)n}} \right)$$

$$\leq \frac{t p^{3/2}}{n^{3/2}} \left( 1 + \sqrt{1 + \frac{\tau^3}{(pn)^{3/2}}} \right) < \frac{3t p^{3/2}}{n^{3/2}}$$

since $\tau \leq 1$ and $np \geq 1$. Hence, we can conclude that if $\tau > \max\{\sqrt{12p}, \sqrt[3]{6t}\}$, then $Q' \in \mathcal{B}_f(Q, \rho)$. Since $p, t \in o_n(1)$, this is satisfied by some $\tau \in o_n(1)$. ∎

**Proof of Claim 16.** Using above computation for $\mu(Q), \mu(Q')$, we have

$$\rho(Q, Q') \leq 7\sqrt{\frac{\ln 2n}{\binom{2n}{3}}} \left( \sqrt{p^3} + \sqrt{p^3 + \gamma^3} \right) \leq \frac{60 p^{3/2} \sqrt{\ln n}}{n^{3/2}}$$

since $\gamma \leq p$ due to definition of $Q'_n$. On the other hand, using the fact that $\gamma \geq 5\sqrt{\frac{p \ln n}{n}}$, we have

$$|\mu(Q) - \mu(Q')| = \left| \gamma^3 - \frac{3}{2n-1}(p^2\gamma + p\gamma^2) \right|$$

$$\geq \gamma^3 \left( 1 - \frac{3}{2n-1}\left( \frac{pn}{25 \ln n} + \frac{\sqrt{np}}{5\sqrt{\ln n}} \right) \right)$$

$$= \gamma^3(1 - o_n(1)) \, ,$$

which is at least $\frac{\gamma^3}{2} > \frac{60 p^{3/2}\sqrt{\ln n}}{n^{3/2}}$ for large $n$. Hence, we have $Q' \in \overline{\mathcal{B}_f(Q, \rho)}$. ∎

20

**Proof of Corollary 12**

The proof follows the line of the previous proof, where we note that for $Q_n, Q'_n$ defined in Theorem 10, we may write $\rho$ as

$$\rho(Q_n, Q'_n) = \frac{t_n}{2n} \left( \sqrt{(2n-1)p_n} + \sqrt{(2n-1)p_n - \gamma_n} \right)$$

for some $t_n \in o_n(1)$. We use the fact that $D_{Q_n} = (2n-1)p_n$ and $D_{Q'_n} = (2n-1)p_n - \gamma_n$, and also observe that both distributions are in $\mathcal{F}_f$ if $p_n \geq \frac{(\ln 2n)^{4.1}}{2n}$. We claim the following.

**Claim 17** *Let $p_n \in [\frac{(\ln 2n)^{4.1}}{2n}, \frac{1}{2}]$. There exists $\tau_n \in o_n(1)$ that satisfies both the following conditions: (i) $\gamma_n = \tau_n \sqrt{\frac{p_n(1-p_n)}{n}}$, and (ii) $Q'_n \in \overline{\mathcal{B}_f(Q_n, \rho)}$ where $\rho$ is given above.*

Hence, as in previous proof we have sequences $Q_n, Q'_n$ with vanishing total variation distance, and yet satisfying $H_1$. Now, we can use arguments similar to proof of Corollary 11 to arrive at the result. Note here, the restricting $p_n$ to be smaller than $\frac{1}{2}$ is not a limitation in this case since the result follows as long as we can find some suitable sequence $p_n$.

As in previous subsection, we also give a proof of the fact that for larger $\gamma_n$, the proposed test can indeed distinguish between $Q_n$ and $Q'_n$. This result, stated below, applies only for $p_n \leq \frac{1}{2}$.

**Claim 18** *Let $p_n \in [\frac{(\ln 2n)^{4.1}}{2n}, \frac{1}{2}]$. If $\gamma_n \geq 15 \sqrt{\frac{p_n}{n}}$, then $Q'_n \in \overline{\mathcal{B}_f(Q_n, \rho)}$ for $\rho$ mentioned in Corollary 9. As a consequence, Corollary 9 implies that proposed test is consistent in distinguishing between $Q_n$ and $Q'_n$.*

**Proof of Claim 17.** We drop the subscript $n$ for convenience. Note that for the $\rho$ defined above, we have $\rho(Q, Q') \leq t\sqrt{\frac{2p}{n}}$. Now, let $M_Q, M_{Q'}$ be the parameter matrix for $Q$ (ER), and $Q'$ (two community model), respectively, and recall that $\mu$ is the vector of the largest two singular values of the parameter matrix, scaled by the graph size $2n$. Hence, $\mu(Q) = \left( \frac{(2n-1)p}{2n}, \frac{p}{2n} \right)$.

On the other hand, $M_{Q'}$ has eigenvalues $((2n-1)p - \gamma)$ and $((2n-1)\gamma - p)$ each with multiplicity 1, and the remaining eigenvalues are $(-p-\gamma)$. Since $\gamma \leq p$, we have $((2n-1)p-\gamma) \geq ((2n-1)\gamma - p)$. Also, if we let $\tau \geq 4\sqrt{\frac{2p}{n}}$, then one can verify that $\gamma \geq \tau\sqrt{\frac{p}{2n}} \geq \frac{2(p+\gamma)}{n}$, which implies that the largest two singular values of $M_{Q'}$ are first two eigenvalues. Hence, we can write $\mu(Q') = \left( \frac{(2n-1)p}{2n} - \frac{\gamma}{2n}, \frac{(2n-1)\gamma}{2n} - \frac{p}{2n} \right)$. If $d$ is the Euclidean distance metric for $f_\lambda$, we have

$$d(\mu(Q), \mu(Q')) = \|\mu(Q) - \mu(Q')\| = \sqrt{\left( \frac{\gamma}{2n} \right)^2 + \left( \gamma - \frac{2p+\gamma}{2n} \right)^2} > \gamma - \frac{2p+\gamma}{2n} \geq \frac{\gamma}{2},$$

where the last inequality holds for $\tau \geq 4\sqrt{\frac{2p}{n}}$. Combining this with the bound on $\rho$, we can conclude that $Q' \in \overline{\mathcal{B}_f(Q, \rho)}$ if $\tau > \max\left\{ 4t, 4\sqrt{2p/n} \right\}$. Since, both terms are $o_n(1)$, we can easily choose a $\tau \in o_n(1)$ satisfying this condition. ∎

**Proof of Claim 18.** Define $\rho(Q, Q') = \frac{7.5}{2n} \left( \sqrt{kD_Q} + \sqrt{kD_{Q'}} \right)$, which is the lower bound for $\rho$ given by Corollary 9. Using $k = 2$ and the above computation for $\mu(Q), \mu(Q')$, we have $\rho(Q, Q') \leq$

$7.5\sqrt{p/n}$. On the other hand, the previously obtained lower bound $\|\mu(Q) - \mu(Q')\| \geq \frac{\gamma}{2}$ still holds. Hence, $Q' \in \overline{\mathcal{B}_f(Q, \rho)}$ if $\gamma \geq 15\sqrt{p/n}$. ∎

## 7. Discussion

The main message of this paper is that two-sample testing is possible in the context of network comparison, where one may not have multiple observations from the same distribution. If one has access to a large population of networks generated from the same model, one may still use standard kernel based test statistics (Gretton et al., 2012) in conjunction with graph kernels (Kondor and Pan, 2016). However, a common situation in practice is where one has exactly two large networks, such as Facebook and LinkedIn connection networks or two brain networks, and needs to decide whether they are similar or different. The present paper concludes that it is indeed possible to address this problem statistically even when the graphs are defined on different entities. However, a formal treatment requires certain considerations:

- There exists a network statistic $f$ that concentrates for large graphs (see Assumption 1).

- If network statistic $f$ is used for comparison, the underlying testing problem merely compares between the point of concentration of $f$ for different models.

- It is often practical to ignore small separations between models, for example, if the comparison is between graphs of different sizes. We characterise this in terms of $\epsilon$.

- On the other hand, similar to the signal detection literature (Ingster and Suslina, 2000; Baraud, 2002), one cannot always hope to distinguish between models with arbitrarily small separation (see Theorem 10 and subsequent corollaries).

In addition, if there exists an accurate estimator $\widehat{\sigma}$ for the deviation of $f$ from its point of concentration, then we show that:

- A general principle provides an uniformly consistent two-sample test for large graphs (Theorem 5). The test does not require any knowledge of the underlying distribution class apart from function $\widehat{\sigma}$.

- For specific network statistics, $f_\Delta$ and $f_\lambda$, and for IER graphs, the test is near-optimal in the minimax sense (see Sections 4 and 5).

- The test is also applicable for other network models (Corollary 7), and even for comparing graphs generated from different distribution classes (Corollary 8).

Hence, we conclude that two-sample testing of large networks is possible whenever one has access to a network statistic that satisfies Assumption 1. In addition, if Assumption 2 is satisfied, then one may also use the proposed two-sample test. The discussion of this paper also leads to some interesting questions for further research. We state two important problems:

- **Concentration of other popular network statistics**
  In this paper, we have only considered triangle and spectrum based statistics as running examples. The natural question one can ask is which other popularly used network statistics

concentrate for generic model classes. For instance, concentration of functions like clustering coefficient and modularity in combination with our results will help to theoretically validate various claims about properties of brain networks.

- **Bootstrapped variant of proposed two-sample test**

  The proposed test primarily relies on concentration of the test statistic (5) under the null and alternative hypotheses. It is often observed that the practical performance of concentration based tests can be improved by using bootstrapped variants (Gretton et al., 2012; Tang et al., 2017). In the present context, we feel that bootstrapping can help to achieve low error rate even for smaller and sparser graphs.

  However, bootstrapping is a challenging problem in the present setting. Gretton et al. (2012) consider a large population problem, where random mixing of the two population helps to estimate the null distribution for the test statistic. Tang et al. (2017) deal with the two graph setting, but the assumption that the graphs are generated from RDPG model allows parameter estimation, which in turn, aids in generating bootstrapped samples from the estimated models. It would be interesting to come up with bootstrapping procedures without such assumptions.

## Acknowledgments

## References

N. Alon, M. Krivelevich, and V. H. Vu. On the concentration of eigenvalues of random symmetric matrices. *Israel Journal of Mathematics*, 131:259–267, 2002.

E. Arias-Castro and N. Verzelen. Community detection in dense random networks. *Annals of Statistics*, 42(3):940–969, 2014.

J. Banks, C. Moore, N. Verzelen, R. Vershynin, and J. Xu. Information-theoretic bounds and phase transitions in clustering, sparse PCA, and submatrix localization. *ArXiv preprint*, arXiv:1607.05222v2:1–36, 2016.

Y. Baraud. Non-asymptotic minimax rates of testing in signal detection. *Bernoulli*, 8(5):577–606, 2002.

B. Bollobas, S. Janson, and O. Riordan. The phase transition in inhomogeneous random graphs. *Random Structures and Algorithms*, 31(3):122, 2007.

G. Bounova and O. de Weck. Overview of metrics and their correlation patterns for multiple-metric topology analysis on heterogeneous graph ensembles. *Physical Review E*, 85:016117, 2012.

S. Bubeck, J. Ding, R. Eldan, and M. Z. Rácz. Testing for high-dimensional geometry in random graphs. *Random Structures & Algorithms*, 49(3):503–532, 2016.

A. Carpentier and R. Nickl. On signal detection and confidence sets for low rank inference problems. *Electronic Journal of Statistics*, 9(2):2675–2688, 2015.

S. Chatterjee. Matrix estiamtion via universal singular value thresholding. *Annals of Statistics*, 43 (1):177–214, 2012.

Y. Chen and J. Xu. Statistical-computational tradeoffs in planted problems and submatrix localization with a growing number of clusters and submatrices. *Journal of Machine Learning Research*, 17(27):1–57, 2016.

F. R. K. Chung. *Spectral graph theory*, volume 92. American Mathematical Society, 1997.

O. Collier. Minimax hypothesis testing for curve registration. In *AISTATS*, 2012.

A. Gretton, K. M. Borgwardt, M. J. Rasch, B. Schölkopf, and A. Smola. A kernel two-sample test. *Journal of Machine Learning Research*, 13:723–733, 2012.

Y. I. Ingster and I. A. Suslina. Minimax nonparametric hypothesis testing for ellipsoids and Besov bodies. *ESAIM: Probability and Statistics*, 4:53–135, 2000.

F. Klimm, D. S. Bassett, J. M. Carlson, and P. J. Mucha. Resolving structural variability in network models and the brain. *PloS Computational Biology*, 10(3):e1003491, 2014.

R. Kondor and H. Pan. The multiscale Laplacian graph kernel. In *NIPS*, 2016.

M. Krivelevich and B. Sudakov. The largest eigenvalue of sparse random graphs. *Combinatorics, Probability and Computing*, 12(1):61–72, 2003.

J. Lei. A goodness-of-fit test for stochastic block models. *Annals of Statistics*, 44(1):401–424, 2016.

J. Lei and A. Rinaldo. Consistency of spectral clustering in stochastic block models. *Annals of Statistics*, 43(1):215–237, 2015.

L. Lu and X. Peng. Spectra of edge-independent random graphs. *The Electronic Journal of Combinatorics*, 20(4):P27, 2013.

E. Mossel, J. Neeman, and A. Sly. Reconstruction and estimation in the planted partition model. *Probability Theory and Related Fields*, 162(3–4):431–461, 2015.

R. Mukherjee, N. S. Pillai, and X. Lin. Hypothesis testing for high-dimensional sparse binary regression. *Annals of Statistics*, 43(1):352–381, 2015.

M. Penrose. *Random geometric graphs*, volume 5. Oxford University Press, 2003.

M. Rubinov and O. Sporns. Complex network measures of brain connectivity: Uses and interpretations. *Neuroimage*, 52:1059–1069, 2010.

C. J. Stam, B. F. Jones, G. Nolte, M. Breakspear, and P. Scheltens. Small-world networks and functional connectivity in Alzheimer's disease. *Cerebral Cortex*, 17(1):92–99, 2007.

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