Two-Sample Tests for Inhomogeneous Random Graphs in L_r Norm: Optimality and Asymptotics

Sayak Chatterjee* University of Pennsylvania
Dibyendu Saha* University of Wisconsin-Madison
Soham Dan IBM Research
Bhaswar B. Bhattacharya University of Pennsylvania

Abstract

In this paper we study the two-sample problem for inhomogeneous Erdős-Rényi (IER) random graph models in the L_r norm, in the high-dimensional regime where the number of samples is smaller or comparable to the size of the graphs. Given two symmetric matrices $P, Q \in [0,1]^{n\times n}$ (with zeros on the diagonals), the two-sample problem for IER graphs (with respect to the L_r norm $\|\cdot\|_r$) is to test the hypothesis $H_0 : P = Q$ versus $H_1 : \|P-Q\|_r \geq \varepsilon$, given a sample of m graphs from the respective distributions. In this paper, we obtain the optimal sample complexity for testing in the L_r norm, for all integers $r \geq 1$. We also derive the asymptotic distribution of the optimal tests under H_0 and develop a method for consistently estimating their variances. This allows us to efficiently implement the optimal tests with precise asymptotic level and establish their asymptotic consistency. We validate our theoretical results by numerical experiments for various natural IER models.

1 INTRODUCTION

A network consists of a set of distinct elements represented by nodes (or vertices) and connections between the nodes represented by links (or edges). These include social networks (where the nodes are individuals/organizations and the edges between the nodes represent their social interaction), telecommunication networks (collection of terminal nodes which are linked together to ensure communication between the terminals), biological networks (for example, an ecosystem can be modeled as a network of interacting species or a protein can be modeled as a network of amino acids), among others. The amount of data collected from these networks have exploded in recent years, presenting statisticians with unique challenges and exciting new opportunities. Although network analysis has been an area of active interest in statistics and machine learning, most classical approaches for graph testing are applicable in the relatively low-dimensional setting, where the sample size (number of graphs) is larger than the size of the graphs (number of vertices). However, in the modern high-dimensional regime [19] the number of samples could be potentially much smaller or comparable to the size of the graph, such as in brain connectivity or protein interaction networks. Consequently, inferential methods for graph-valued data is an active and emerging discipline (see [4, 6, 12, 22, 23, 28, 29, 37] and the references therein).

In this paper we study the 2-sample problem for network data, where the goal is to test whether 2 network models are equal or different, given samples from the respective distributions. This problem arises naturally in a variety of applications. For example, Zhang et al. [38] study the topological changes in gene regulatory networks for 2 different treatments of breast cancer and Bassett et al. [5] study the difference of anatomical brain structures between healthy individuals and schizophrenic patients. Another important field of application is functional neuroimaging data, where the regions of interest in brain are considered as the vertices of the network and functional connectivity between two such regions are represented by the edges of the network. Towards this, Ginestet et al. [16] considered the problem of testing equality of Fréchet means of the graph Laplacians based on i.i.d. samples of networks from 2 distributions. They derived a central limit theorem for the sample Fréchet mean and proposed a Wald-type 2-sample test for the problem. Recently, Maugis et al. [26] (see also [7]) proposed 2-sample tests based on subgraph counts for graphs generated from graphon models [25].

The aforementioned works assume that the size of the networks is fixed and the number of samples increase to infinity. In this paper we are interested in the high-dimensional regime, where the number of samples is much smaller or comparable to the size of the graph. Specifically, we study the two sample problem for inhomogeneous random graph (IER) models in this regime.

Proceedings of the 26th International Conference on Artificial Intelligence and Statistics (AISTATS) 2023, Valencia, Spain. PMLR: Volume 206.

*The first two authors contributed equally to the paper.

Definition 1.1. [3] Given a symmetric matrix $P^{(n)} \in [0,1]^{n\times n}$ with zeroes on the diagonal, a graph G is said to be an *inhomogeneous Erdős-Rényi* (IER) random graph with *edge probability* $P^{(n)} = ((p_{ij})) \in [0,1]^{n\times n}$, denoted as $G \sim \mathrm{IER}(P^{(n)})$, if its symmetric adjacency matrix $A(G) = ((a_{ij}(G))) \in \{0,1\}^{n\times n}$ have independent entries satisfying:

$$a_{ij}(G) \sim \mathrm{Ber}(p_{ij}) \text{ for all } 1 \le i < j \le n.$$

Note that $P^{(n)}$ is a symmetric matrix with elements in $[0,1]$ and zeros on the diagonal. The IER model includes several popular network models, such as stochastic block models [23], the β-model, the Chung-Lu model [10], [8], and random dot product graphs [31].

Given independent samples $G_1, G_2, \ldots, G_m \sim \mathrm{IER}(P^{(n)})$ and $H_1, H_2, \ldots, H_m \sim \mathrm{IER}(Q^{(n)})$, where $P^{(n)} = ((p_{ij}))$ and $Q^{(n)} = ((q_{ij}))$, the two-sample problem for IER graphs is to test the hypothesis

$$H_0 : P = Q \quad \text{versus} \quad H_1 : \|P^{(n)} - Q^{(n)}\|_r \ge \varepsilon, \quad (1)$$

where $\|\cdot\|_r$ is the L_r norm:

$$\|P^{(n)} - Q^{(n)}\|_r = \left(\sum_{1\le i,j\le n} |p_{ij} - q_{ij}|^r \right)^{1/r}.$$

This problem has been studied recently in a series of papers by Ghoshdastidar et al. [13, 14, 15]. In [15] the authors obtained non-asymptotic minimax rates for the testing problem (1) for the Frobenius (L_2) norm and the operator norm, among others. Practical implementations of the optimal tests (for the L_2 norm and the operator norm) using their asymptotic distributions were derived in [15]. In [13] the authors proposed tests for comparing 2 networks based on network statistics, such as triangle counts and largest singular values. The related problem of goodness-of-fit testing in the IER model was studied recently in [9, 11]. In particular, [9] derived local minimax rates for the goodness-of-fit problem in the L_r norm, for $1 \le r \le 2$.

In this paper, we study the 2-sample problem (1) for general L_r norms, for $r \ge 1$. Specifically, we obtain the optimal sample complexity for 2-sample testing in the L_r norm, for all integers $r \ge 1$. The results adapt and extend the classical techniques of Ingster [20] Ingster and Suslina [21] for Gaussian models to the case of IER graphs. We complement these minimax results by deriving the asymptotic null distribution of the optimal tests and consistent estimates of the null variances, for all integers $r \ge 2$. Consequently, we can implement the optimal test with exact asymptotic level (probability of Type I error) which is agnostic to the knowledge of the separation parameter ε. We summarize our results below.

1.1 Summary of Results

1.1.1 Optimal Sample Complexities

Given i.i.d. samples $G_1, G_2, \ldots, G_m$ from $\mathrm{IER}(P^{(n)})$, and $H_1, H_2, \ldots, H_m$ from $\mathrm{IER}(Q^{(n)})$, a test is a binary function $\phi_{m,n} : (\boldsymbol{G}_m; \boldsymbol{H}_m) := (G_1, G_2, \ldots, G_m; H_1, H_2, \ldots, H_m) \to \{0,1\}$, which is 1 when H_0 is rejected and 0 otherwise. The worst-case risk of a test function $\phi_{m,n}$ for the testing problem (1) is defined as:

$$\mathcal{R}(P^{(n)}, Q^{(n)}, \phi_{m,n}) = \mathbb{P}_{H_0}(\phi_{m,n} = 1) + \sup_{\|P^{(n)} - Q^{(n)}\|_r \ge \varepsilon} \mathbb{P}_{H_1}(\phi_{m,n} = 0), \quad (2)$$

which is the sum of the Type I error and the maximum possible Type II error of the test function $\phi_{m,n}$. We are interested in the asymptotic regime where the risk (2) transitions from 0 to 1. This is formalized in the following definition:

Definition 1.2. Given $G_1, G_2, \ldots, G_m$ i.i.d. samples from $\mathrm{IER}(P^{(n)})$, and $H_1, H_2, \ldots, H_m$ i.i.d. samples from $\mathrm{IER}(Q^{(n)})$, where $m = m_n$ can depend on n, a sequence of test functions $\phi_{m,n}$ is said to be *asymptotically powerless* for (1), if there exists a sequence of symmetric matrices $(P^{(n)}, Q^{(n)}) \in [0,1]^{n\times n} \times [0,1]^{n\times n}$ such that $\lim_{n\to\infty} \mathcal{R}(P^{(n)}, Q^{(n)}, \phi_{m,n}) = 1$. On the other hand, a sequence of test functions $\phi_{m,n}$ is said to be *asymptotically powerful* for (1), if for all symmetric matrices $(P^{(n)}, Q^{(n)}) \in [0,1]^{n\times n} \times [0,1]^{n\times n}$, $\lim_{n\to\infty} \mathcal{R}(P^{(n)}, Q^{(n)}, \phi_{m,n}) = 0$.

As in the case for the Gaussian sequence model [21], the optimal sample complexity depends on whether $1 \le r < 2$ or $r \ge 2$. Specifically, we obtain the following results (asymptotic notations are as defined in Section 1.3):

- For any integer $r \ge 2$, the *optimal sample complexity* for the testing problem (1) is, $n^{2/r}/\varepsilon^2$. This means that there is a (computationally efficient) test which is asymptotically powerful for (1) when the sample size $m \gg n^{2/r}/\varepsilon^2$. On the other hand, and all tests are asymptotically powerless when the sample size $m \ll n^{2/r}/\varepsilon^2$ (Theorem 2.1).
- For any $1 \le r < 2$, the optimal sample complexity for the testing problem (1) is $n^{(4/r)-1}/\varepsilon^2$ (Theorem 2.3).
- We also obtain the optimal sample complexity for testing in the L_∞ norm up to a logarithmic factor (Theorem 2.4).

1.1.2 Asymptotic Distribution and Consistency

While optimality results provide mathematical insights on the structure of the 'best' tests, the rejection regions of the optimal tests often require knowledge of the separation parameter, and they can be conservative in practice.

To handle this issue, in Section 3 we propose an implementation of the optimal test statistic using its asymptotic null distribution. First, we show that for any integer $r \geq 2$, the optimal test statistic for the L_r norm when normalized by its variance under H_0 has a central limit theorem (CLT) under H_0 (Theorem 3.1). Moreover, we can consistently estimate the null variance of the test statistic and, consequently, obtain an asymptotically level α test. We also derive conditions for consistency (power converging to 1) for this test, which matches the minimax separation rate in the worst case (Theorem 3.2). This improves and extends the results in [14] where an upper bound on the null variance of the optimal test was estimated for the L_2 norm. We illustrate the finite-sample performances of the proposed tests in simulations. Our experiments show that asymptotic approximations are accurate even for moderate sized networks and for a wide range of network sparsity (Section 4). The codes for all the experiments can be found in the Github repository `https://github.com/sdan2/Lp-graph-testing`.

1.2 Related Work on Testing Based on a Single Observed Network

In this paper we study the 2-sample problem based on multiple i.i.d. networks from an IER model. There is parallel line of work where one observes a single network from the respective distributions and asymptotic properties are derived as the size of the network grows. In this direction, Tang et al. [34, 35] considered testing whether or not 2 random dot product graphs are generated from the same latent positions (see also [1, 36] and the references therein for more recent results); and Li and Li [24] proposed a test for the equality of community memberships in weighted stochastic block models. For general IER models [14, 15] considered tests based on the operator norm of the difference of the adjacency matrices of the observed networks. Recently, for graphon models, [33] proposed a test for network 2-sample inference based on subgraph counts and [32] proposed a test based on a novel graph-distance.

1.3 Asymptotic Notation

Throughout we will use the following standard asymptotic notations. For two positive sequences $\{a_n\}_{n\geq 1}$ and $\{b_n\}_{n\geq 1}$, $a_n = O(b_n)$ means $a_n \leq C_1 b_n$ and $a_n = \Theta(b_n)$ means $C_2 b_n \leq a_n \leq C_1 b_n$, for positive constants C_1, C_2 and n large enough. Moreover, $a_n = o(b_n)$ will mean $a_n/b_n \to 0$, as $n \to \infty$. Similarly, $a_n \lesssim b_n$ means $a_n = O(b_n)$; $a_n \asymp b_n$ means $a_n \lesssim b_n \lesssim a_n$; $a_n \ll b_n$ means $a_n = o(b_n)$; and $a_n \gg b_n$ means $b_n = o(a_n)$.

1.4 Organization

The rest of the paper is organized as follows: The optimal sample complexities for testing in the L_r norms are discussed in Section 2. The asymptotic properties of the test statistics are presented in Section 3. Empirical performances of the tests are illustrated in Section 4. Proofs of the results are given in the supplementary materials.

2 OPTIMAL SAMPLE COMPLEXITIES

In this section, we obtain the optimal sample complexities for the two-sample problem (1) for the norms described above. Throughout we will assume that the sample size m is divisible by r. In particular, suppose that $m = Kr$, for some $L \geq 1$ and partition the set $[m] := \{1, 2, \ldots, m\}$ into r disjoint sets $B_1, B_2, \ldots, B_r$ each with K consecutive elements, that is, $B_\ell := \{K(\ell-1)+1, K(\ell-1)+2, \ldots, K\ell\}$, for $1 \leq \ell \leq r$. Denoting the adjacency matrices of the graphs G_s and H_s by $A(G_s) = ((a_{ij}(G_s))), B(H_s) = ((b_{ij}(H_s)))$, respectively, define

$$\Delta^{(\ell)}_{m,n,r}(i,j) := \sum_{s\in B_\ell}(a_{ij}(G_s) - b_{ij}(H_s)) = \sum_{s=K(\ell-1)+1}^{K\ell}(a_{ij}(G_s) - b_{ij}(H_s)),$$

for $1 \leq \ell \leq r$. To test separation in the L_r norm we consider the following statistic:

$$T_{m,n,r} := \sum_{1\leq i<j\leq n}\prod_{\ell=1}^{r}\Delta^{(\ell)}_{m,n,r}(i,j), \tag{3}$$

when r is even, and

$$T_{m,n,r} := \sum_{1\leq i<j\leq n}\left(\prod_{\ell=1}^{r-1}\Delta^{(\ell)}_{m,n,r}(i,j)\right)\left|\Delta^{(r)}_{m,n,r}(i,j)\right|, \tag{3}$$

when r is odd. Note that

$$\mathbb{E}(\Delta^{(\ell)}_{m,n,r}(i,j)) = K(p_{ij} - q_{ij})$$

and $\{\Delta^{(\ell)}_{m,n,r}(i,j)\}_{1\leq\ell\leq r}$ is a collection of independent random variables. Hence, when r is even,

$$\mathbb{E}(T_{m,n,r}) = K^r\sum_{1\leq i<j\leq n}(p_{ij} - q_{ij})^r = \frac{K^r}{2}\|P^{(n)} - Q^{(n)}\|_r^r, \tag{4}$$

that is, $T_{m,n,r}$ is an unbiased estimate of $\frac{K^r}{2}\|P^{(n)} - Q^{(n)}\|_r^r$. In particular, this implies, $\mathbb{E}_{H_0}(T_{m,n,r}) = 0$ when $P^{(n)} = Q^{(n)}$, and $\mathbb{E}_{H_1}(T_{m,n,r}) \geq \frac{K^r\varepsilon^r}{2}$ when $\|P^{(n)} - Q^{(n)}\|_r \geq \varepsilon$. In case r is odd, because of parity issues $T_{m,n,r}$ is no longer unbiased for $\frac{K^r}{2}\|P^{(n)} - Q^{(n)}\|_r^r$. Nevertheless, since

$$\mathbb{E}(T_{m,n,r})$$

$$= K^r \sum_{1 \le i < j \le n} (p_{ij} - q_{ij})^{r-1} \mathbb{E} \left| \Delta_{m,n,r}^{(r)}(i,j) \right|, \qquad (5)$$

we still get $\mathbb{E}_{H_0}(T_{m,n,r}) = 0$ when $P^{(n)} = Q^{(n)}$, and

$$\mathbb{E}_{H_1}(T_{m,n,r}) \ge K^r \sum_{1 \le i < j \le n} |p_{ij} - q_{ij}|^{r-1} \left| \mathbb{E}(\Delta_{m,n,r}^{(r)}(i,j)) \right|$$
$$= \frac{K^r}{2} \|P^{(n)} - Q^{(n)}\|_r^r \ge \frac{K^r \varepsilon^r}{2},$$

when $\|P^{(n)} - Q^{(n)}\|_r \ge \varepsilon$. Therefore, a natural test for detecting separation in the L_r norm would be to reject H_0 for 'large' values of the statistic $T_{m,n,r}$. In following theorem we show that such a test attains the optimal sample complexity for the testing problem (1) when $r \ge 2$.

Theorem 2.1. *Fix an integer $r \ge 2$ and consider the testing problem* (1) *under the L_r norm. Then the following hold:*

(a) The test $\phi_{m,n,r} = \mathbf{1}\{T_{m,n,r} \ge \frac{1}{4}\left(\frac{m\varepsilon}{r}\right)^r\}$, where $T_{m,n,r}$ is as defined in (3) *above, is asymptotically powerful for* (1)*, whenever $m \gg n^{2/r}/\varepsilon^2$.*

(b) On the other hand, all tests are asymptotically powerless for (1)*, whenever $m \ll n^{2/r}/\varepsilon^2$.*

The proof of the above result is given in Section 1 of the supplementary materials. The upper bound (Theorem 2.1 (a)) proceeds by a variance calculation which shows that $T_{m,n,r}$ concentrates about its expected value (both under H_0 and H_1 as in (1)), hence, $\phi_{m,n,r}$ can detect differences in the L_r norm, whenever $m \gg n^{2/r}/\varepsilon^2$. For the lower bound (which entails finding 'hard' instances of $P^{(n)}$ and $Q^{(n)}$ satisfying (1)), we choose $Q^{(n)} = ((q_{ij}))$ to be the matrix corresponding to the Erdős-Rényi graph $\mathrm{ER}(n, 1/2)$, that is, $q_{ij} = \frac{1}{2}$, for all $1 \le i \ne j \le n$ and $P^{(n)} = ((p_{ij}))$ to be $\frac{1}{2} + \delta$ in $O(n)$ locations and $\frac{1}{2}$ in the remaining locations, where the parameter δ is chosen (depending on n, r, and ε) such that the L_r norm between $P^{(n)}$ and $Q^{(n)}$ is ε. Then a second-moment calculation of the likelihood ratio shows that detecting these distributions is impossible when $m \ll n^{2/r}/\varepsilon^2$.

Although Theorem 2.1 assumes $r \ge 2$ is an integer, it will be evident from the proof that the lower bound holds for all $r \ge 2$ (see Proposition 1.1). In other words, for any $r \ge 2$, all tests are asymptotically powerless for (1), whenever $m \ll n^{2/r}/\varepsilon^2$. To get an upper bound for a non-integer $r \ge 2$, note that by Hölder's inequality,

$$\|P^{(n)} - Q^{(n)}\|_r \le n^{(2/r)-(2/\lceil r \rceil)} \|P^{(n)} - Q^{(n)}\|_{\lceil r \rceil}.$$

Therefore, $\|P^{(n)} - Q^{(n)}\|_r \ge \varepsilon$ implies that $\|P^{(n)} - Q^{(n)}\|_{\lceil r \rceil} \ge \varepsilon / n^{(2/r)-(2/\lceil r \rceil)}$. Consequently, by Theorem 2.1 (a), the test which rejects H_0 when

$$T_{m,n,\lceil r \rceil} \ge \frac{1}{4}\left(\frac{m\hat{\varepsilon}}{r}\right)^r, \qquad \text{where } \hat{\varepsilon} := \varepsilon / n^{(2/r)-(2/\lceil r \rceil)},$$

is asymptotically powerful whenever $m \gg n^{2/r}/\hat{\varepsilon}^2 = n^{(4/r)-(2/\lceil r \rceil)}/\varepsilon^2$. We summarize this result in the following corollary:

Corollary 2.2. *Fix $r \ge 2$ not an integer and consider the testing problem* (1)*. Then the following hold:*

(a) The test $\phi_{m,n,r} = \mathbf{1}\{T_{m,n,\lceil r \rceil} \ge \frac{1}{4}\left(\frac{m\hat{\varepsilon}}{r}\right)^r\}$, where $\hat{\varepsilon} := \varepsilon / n^{(2/r)-(2/\lceil r \rceil)}$, is asymptotically powerful for (1)*, whenever $m \gg n^{(4/r)-(2/\lceil r \rceil)}/\varepsilon^2$.*

(b) On the other hand, all tests are asymptotically powerless for (1)*, whenever $m \ll n^{2/r}/\varepsilon^2$.*

Remark 2.1. By fixing m and varying the separation parameter ε, the above results can be restated in terms of the minimax separation radius. In particular, Theorem 2.1 shows that for testing in the L_r norm the minimax separation radius is $n^{1/r}/\sqrt{m}$, that is, if $\varepsilon \gg n^{1/r}/\sqrt{m}$ the test $\phi_{m,n,r}$ is asymptotically powerful. On the other hand, if $\varepsilon \ll n^{1/r}/\sqrt{m}$ all tests are asymptotically powerless. For example, if $r = 2$ the minimax separation radius is $\sqrt{n/m}$, which also follows from the results in [15].

Next, we consider testing separation in the L_r norm for $1 \le r < 2$. Note that by Hölder's inequality,

$$\|P^{(n)} - Q^{(n)}\|_r \le n^{(2/r)-1} \|P^{(n)} - Q^{(n)}\|_2. \qquad (6)$$

Therefore, $\|P^{(n)} - Q^{(n)}\|_r \ge \varepsilon$ implies that $\|P^{(n)} - Q^{(n)}\|_2 \ge \varepsilon / n^{(2/r)-1}$. Consequently, by Theorem 2.1 (a), the test which rejects H_0 when

$$T_{m,n,2} \ge \frac{1}{4}\left(\frac{m\tilde{\varepsilon}}{r}\right)^r, \qquad \text{where } \tilde{\varepsilon} := \varepsilon / n^{(2/r)-1}, \qquad (7)$$

is asymptotically powerful whenever $m \gg n^{2/r}/\tilde{\varepsilon}^2 = n^{(4/r)-1}/\varepsilon^2$. The following theorem shows that this sample complexity is indeed optimal in the regime $1 \le r < 2$. In this case the lower bound is attained by a 2-sided perturbation. Specifically, for $1 \le i < j \le n$, we choose $q_{ij} = \frac{1}{2}$ and $p_{ij} = q_{ij} + \gamma_{ij}\delta$, where $\{\gamma_{ij}\}$ are i.i.d. ± 1 with probability $\frac{1}{2}$ and δ (depending on n, r, and ε) is such that the L_r norm between $P^{(n)}$ and $Q^{(n)}$ is ε. A second-moment calculation of the likelihood ratio shows that detecting these distributions is impossible when $m \ll n^{(4/r)-1}/\varepsilon^2$ and $1 \le r < 2$. The proof of this result is given in Section 2 of the supplementary materials.

Theorem 2.3. *Fix $1 \le r < 2$ and consider the testing problem* (1)*. Then the following hold:*

(a) The test $\phi_{m,n,r} = \mathbf{1}\{T_{m,n,2} \ge \frac{1}{4}\left(\frac{m\tilde{\varepsilon}}{r}\right)^r\}$, where $\tilde{\varepsilon} := \varepsilon / n^{(2/r)-1}$, is asymptotically powerful for (1)*, whenever $m \gg n^{(4/r)-1}/\varepsilon^2$.*

(b) On the other hand, all tests are asymptotically powerless for (1)*, whenever $m \ll n^{(4/r)-1}/\varepsilon^2$.*

Next, we consider the problem of testing in the L_∞ norm. In this case we have the following result:

Theorem 2.4. *Consider the testing problem* (1) *under the L_∞ norm. Then the following hold:*

(a) There exists an asymptotically powerful test for (1), *whenever $m \gg \log n/\varepsilon^2$.*

(b) On the other hand, all tests are asymptotically powerless for (1), *whenever $m \ll 1/\varepsilon^2$.*

For testing in the L_∞ norm a natural test statistic is:

$$T_{m,n,\infty} := \sup_{1 \le i < j \le n} \left| \sum_{s=1}^{m} (a_{ij}(G_s) - b_{ij}(H_s)) \right|. \quad (8)$$

The test which rejects H_0 for large values of $T_{m,n,\infty}$ attains the optimal sample complexity in the L_∞ norm up to a log-factor. The proof is given in Section 3 of the supplementary materials.

Remark 2.2. The results above give the optimal sample complexity for testing in the L_r norm for all integers $r \ge 1$. In fact, combining Theorems 2.1 and 2.3, the optimal sample complexity for all integers $r \ge 1$ can expressed as $\Theta(n^{\lambda(r)}/\varepsilon^2)$, where $\lambda(r) = \max\{4/r - 1, 2/r\}$. Moreover, $\lim_{r\to\infty} \lambda(r) = 0$, which matches the sample complexity of testing in the L_∞ norm up to a log factor. We conjecture that $\Theta(n^{\lambda(r)}/\varepsilon^2)$ is, in fact, the optimal sample complexity for all $r \ge 1$ (not necessarily an integer). However, our current upper and lower bounds for non-integers $r \ge 2$ (Corollary 2.2) do not match by a factor depending on fractional part of r in the exponent.

3 ASYMPTOTIC PROPERTIES

In this section we derive the asymptotic properties of the test statistic $T_{m,n,r}$ as the size of the graphs $n \to \infty$, where the sample size $m > 1$ is also allowed to depend on n. We begin with the distribution of $T_{m,n,r}$ under H_0.

Theorem 3.1. *Fix an integer $r \ge 2$ and suppose the sequence of matrices $\{P^{(n)}\}_{n\ge1}$ satisfies $\lim_{n\to\infty} \|P^{(n)}\|_r = \infty$ and $\sup_{n\ge1} \|P^{(n)}\|_\infty < 1$. Then under H_0, as $n \to \infty$.*

$$Z_{m,n,r} := \frac{T_{m,n,r}}{\sqrt{\mathrm{Var}_{H_0}(T_{m,n,r})}} \xrightarrow{D} N(0,1), \quad (9)$$

where $\mathrm{Var}_{H_0}(T_{m,n,r})$ denotes the variance of $T_{m,n,r}$ under H_0.

The proof of the above result is given in Section 4 of the supplementary materials. The proof uses the Berry-Esseen theorem to establish the following quantitative bound:

$$\sup_{x\in\mathbb{R}} |\mathbb{P}_{H_0}(Z_{m,n,r} \le x) - \Phi(x)| \lesssim \frac{1}{\sqrt{\|P^{(n)}\|_r^r}}, \quad (10)$$

where $\Phi(x)$ is the distribution function of $N(0,1)$. Since $\lim_{n\to\infty} \|P^{(n)}\|_r = \infty$ by assumption, (10) implies the result in (9).

In order to use Theorem 3.1 to construct a test for (1), we need to consistently estimate $\mathrm{Var}_{H_0}(T_{m,n,r})$. To this end, a direct computation shows that (see (4.3) in the supplementary materials)

$$\begin{aligned} \sigma^2_{m,n,r} &:= \mathrm{Var}_{H_0}(T_{m,n,r}) \\ &= \left(\frac{2m}{r}\right)^r \sum_{1\le i<j\le n} p_{ij}^r (1-p_{ij})^r. \end{aligned} \quad (11)$$

To estimate the variance we split the sample into $2r$ parts and use the first r parts to estimate p_{ij}^r and the last r parts to estimate $(1-p_{ij})^r$ and then take their product. Formally, suppose $m = 2rC$, for some $C \ge 1$, and define

$$\hat{P}_{ij}^{(\ell)} = \frac{1}{C} \sum_{s=C(\ell-1)+1}^{C\ell} a_{ij}(G_s), \quad (12)$$

for $1 \le \ell \le 2r$. Then we can estimate $\sigma^2_{m,n,r}$ as follows:

$$\hat{\sigma}^2_{m,n,r} = \left(\frac{2m}{r}\right)^r \sum_{1\le i<j\le n} \left\{ \prod_{\ell=1}^{r} \hat{P}_{ij}^{(\ell)} \prod_{\ell=r+1}^{2r} (1 - \hat{P}_{ij}^{(\ell)}) \right\}. \quad (13)$$

Note that $\mathbb{E}_{H_0}(\hat{\sigma}^2_{m,n,r}) = \sigma^2_{m,n,r}$, that is, $\hat{\sigma}^2_{m,n,r}$ is an unbiased estimate of $\sigma^2_{m,n,r}$. The following result shows that $\hat{\sigma}^2_{m,n,r}$ consistently estimates $\sigma^2_{m,n,r}$ and, consequently, the test which rejects H_0 when $|T_{m,n,r}| > z_{\frac{\alpha}{2}} \hat{\sigma}_{m,n,r}$ is an asymptotically level α test.

Theorem 3.2. *Fix an integer $r \ge 2$ and suppose the sequence of matrices $\{P^{(n)}\}_{n\ge1}$ satisfies $\lim_{n\to\infty} \|P^{(n)}\|_r = \infty$ and $\sup_{n\ge1} \|P^{(n)}\|_\infty < 1$. Then*

$$\frac{\hat{\sigma}^2_{m,n,r}}{\sigma^2_{m,n,r}} \xrightarrow{P} 1, \quad (14)$$

As a consequence, $\hat{Z}_{m,n,r} := \frac{T_{m,n,r}}{\hat{\sigma}_{m,n,r}} \xrightarrow{D} N(0,1)$ under H_0 and the test which rejects H_0 when $|\hat{Z}_{m,n,r}| > z_{\frac{\alpha}{2}}$ is asymptotically level α, that is,

$$\lim_{n\to\infty} \mathbb{P}_{H_0}(|\hat{Z}_{m,n,r}| > z_{\frac{\alpha}{2}}) = \alpha. \quad (15)$$

Moreover, if $\{P^{(n)}\}_{n\ge1}$ and $\{Q^{(n)}\}_{n\ge1}$ is such that $\|P^{(n)} - Q^{(n)}\|_r^2 \gg \frac{1}{m}\left(\|P^{(n)}\|_r + \|Q^{(n)}\|_r\right)$, then

$$\lim_{n\to\infty} \mathbb{P}_{H_1}(|\hat{Z}_{m,n,r}| > z_{\frac{\alpha}{2}}) = 1. \quad (16)$$

The proof of Theorem 3.2 is given in Section 5 of the supplementary materials. The results show that the statistic $T_{m,n,r}$, which attain the optimal sample complexity for the L_r norm, can be rescaled to obtain a CLT under H_0 and the null varaince can be consistently estimated. One of the advantages of the variance estimation is that it allows us to choose a cut-off for which the probability of Type I error is *equal* to α. This is in contrast to the result in [14, Theorem 1] where an upper bound on the variance is estimated for the L_2 statistic $T_{m,n,2}$, leading to a test which is a conservative, that is, the probability of Type I error is asymptotically *bounded above* by α, instead of being equal to α. Moreover, (16) shows that the test based on $\hat{Z}_{m,n,r}$ is consistent whenever the 'normalized signal strength':

$$\frac{\|P^{(n)} - Q^{(n)}\|_r^2}{\|P^{(n)}\|_r + \|Q^{(n)}\|_r} \gg \frac{1}{m}.$$

In particular, if $\|P^{(n)} - Q^{(n)}\|_r = \varepsilon$, then this implies that the test based on $\hat{Z}_{m,n,r}$ is asymptotically powerful whenever

$$m \gg \frac{\|P^{(n)}\|_r + \|Q^{(n)}\|_r}{\varepsilon^2},$$

which matches the minimax sample complexity obtained in Theorem 2.1 in the worst case, since $\max\{\|P^{(n)}\|_r, \|Q^{(n)}\|_r\} \leq n^{2/r}$. Therefore, we have a practical test for detecting separation in the L_r norm which attains the optimal sample complexity and does not require knowledge of the separation parameter, for any integer $r \geq 2$.

4 NUMERICAL RESULTS

In this section we implement the tests described above for various common IER models. Specifically, we will consider the following 3 models:

- *Erdős-Rényi Model*: This is the basic random graph model where every edge is present independently with probability $p \in [0,1]$. We denote this model by $\mathrm{ER}(n,p)$.

- *Planted Bisection Model*: In this case the edge-probability matrix $P = ((p_{ij}))$ has a 2-block structure:

$$p_{ij} := \begin{cases} a & \text{if } 1 \leq i \neq j \leq \frac{n}{2} \text{ or } \frac{n}{2} < i \neq j \leq n, \\ b & \text{if } 1 \leq i \leq \frac{n}{2} \text{ and } \frac{n}{2} < j \leq n \text{ or} \\ & \frac{n}{2} < i \leq n \text{ and } 1 \leq j \leq \frac{n}{2} \\ 0 & \text{if } i = j, \end{cases}$$

  where $a, b \in [0,1]$. We denote a random graph on n vertices from this model as $\mathrm{PB}(n,a,b)$. This is a special case of the well-known stochastic block model [27], where the vertex set is divided into two equal-sized communities and the edges are added independently depending on the community membership.

- *β-Model*: This is another popular IER model [8, 17, 28, 30], where the edge-probability matrix $P = ((p_{ij}))$ is given by

$$p_{ij} = \frac{e^{\beta_i + \beta_j}}{1 + e^{\beta_i + \beta_j}},$$

  for $1 \leq i \neq j \leq n$, where $\boldsymbol{\beta} = (\beta_1, \beta_2, \ldots, \beta_n)$ is an n-dimensional parameter vector. The β-model is a simple version of a collection of exponential models actively used for analyzing network data, and is a close analogue to the Bradley-Terry model for rankings [18].

Figure 1 shows the histogram of the L_4 statistic $\hat{Z}_{m,n,4}$ and limiting normal density (as in Theorem 3.2) under H_0 with $n = 100$ and $m = 32$ for the Erdős-Rényi model $\mathrm{ER}(n, \frac{1}{2})$ (Figure 1(a)), the (sparse) planted bisection model $\mathrm{PB}(n, 8/n, 2/n)$ (Figure 1 (b)), and the β-model, for a uniformly chosen unit vector $\boldsymbol{\beta} \in \mathbb{R}^n$ (Figure 1 (c)). In all the 3 cases, the histograms closely follow the limiting normal distributions, showing that the asymptotic approximations are accurate even for moderate sized networks and relatively small sample sizes. Moreover, the asymptotic approximations are valid for wide range of sparsity, from dense graphs with $\Theta(n^2)$ edges (as in $\mathrm{ER}(n, \frac{1}{2})$ and the β-model) to sparse graphs with $\Theta(n)$ edges (as in $\mathrm{PB}(n, 8/n, 2/n)$).

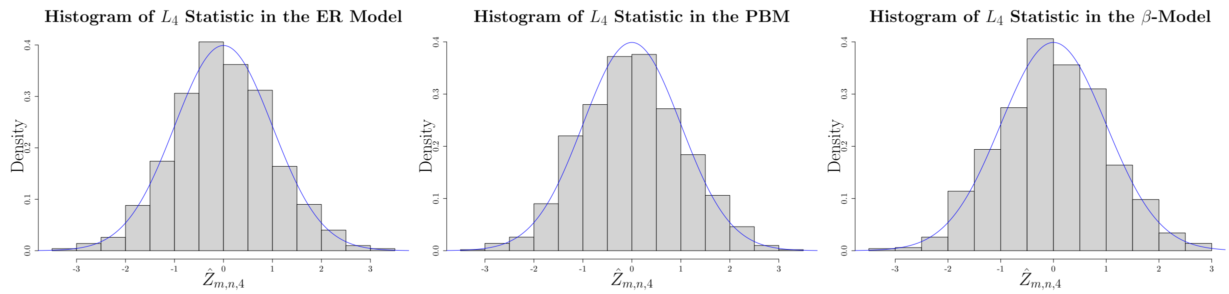
Note that Theorem 2.1 shows that given a separation parameter ε, the sample complexity for testing in the L_r norm decreases with r. In other words, for a fixed sample size m the optimal test for the L_r norm has a smaller minimax separation radius for larger r. For example, the minimax separation radius for the L_2 norm is $\sqrt{n/m}$ whereas for the L_4 norm it is $n^{1/4}/\sqrt{m}$. To illustrate this we consider the following 3 simulation settings:

- We consider $G_1, G_2, \ldots, G_m$ i.i.d. from $\mathrm{ER}(n, 0.2)$ and $H_1, H_2, \ldots, H_m$ i.i.d. from $\mathrm{ER}(n, 0.2 + \varepsilon/n^{2/r})$. This ensures the separation in the L_r norm is $\asymp \varepsilon$.

- We consider $G_1, G_2, \ldots, G_m$ i.i.d. from $\mathrm{PBM}(n, a/n^{2/r}, b/n^{2/r})$ and $H_1, H_2, \ldots, H_m$ i.i.d. from $\mathrm{PBM}(n, a/n^{2/r} + \varepsilon/n^{2/r}, b/n^{2/r} - \varepsilon/n^{2/r})$ (with probabilities truncated to be within $[0,1]$ if required).

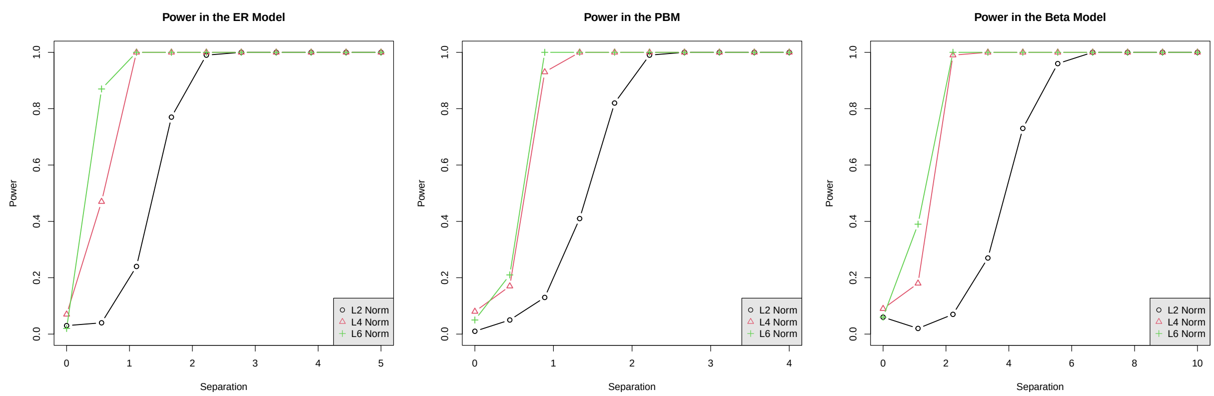
- We consider $G_1, G_2, \ldots, G_m$ i.i.d. from the β-model with $\beta_1, \beta_2, \ldots, \beta_n$ i.i.d. $N(0,1)$ and $H_1, H_2, \ldots, H_m$ from the β-model with $\beta_i + \varepsilon/n^{2/r}$, for $1 \leq i \leq n$.

Figure 2 shows the empirical power (over 100 iterations) for the test based on $\hat{Z}_{m,n,r}$ for the above 3 cases with $n = 50$ and $m = 48$ over a grid of 10 values of ε for $r \in \{2, 4, 6\}$. In all the 3 cases, we observe that the test for the L_6 norm has power reaching 1 faster than the test for the L_4 norm which, in turn, reaches power 1 faster than the



(a) (b) (c)

Figure 1: Histogram of the L_4 statistic $\hat{Z}_{m,n,4}$ with $m = 32$ and $n = 100$ and the asymptotic normal density (in blue) for (a) the Erdős-Rényi model $\mathrm{ER}(n, \frac{1}{2})$, (b) the (sparse) planted bisection model $\mathrm{PB}(n, 8/n, 2/n)$ and (c) the β-model with $\|\boldsymbol{\beta}\|_2 = 1$.



(a) (b) (c)

Figure 2: Power of tests based on $\hat{Z}_{m,n,r}$ as a function of the separation ε with $n = 50$ and $m = 48$ for (a) the Erdős-Rényi model, (b) the planted bisection model, and (c) the β-model.

test for the L_2 norm, as expected from the result in Theorem 2.1.

It is also natural to wonder under what kinds of alternatives does the test for the L_r norm have more power than the tests based on the other norms. The proof of Theorem 2.1 reveals that tests for higher norms are powerful when the matrices $P^{(n)}$ and $Q^{(n)}$ differ on a 'small' of set of coordinates. For example, in the extreme case where $P^{(n)}$ and $Q^{(n)}$ differ on 1 (or a very small number) of coordinates the test for the L_∞ norm is likely to be powerful. On the other hand, if $P^{(n)}$ and $Q^{(n)}$ differ on a positive fraction of the coordinates, the test based on the L_2 norm will tend to perform better. To illustrate this we consider $G_1, G_2, \ldots, G_m$ i.i.d. from $\mathrm{ER}(n, 0.2)$ and $H_1, H_2, \ldots, H_m$ i.i.d. from $\mathrm{IER}(Q^{(n)})$ where the elements of $Q^{(n)}$ equal $0.2 + \varepsilon$ in only 10 locations and 0.2 in the remaining locations. In this setting, Figure 3 shows the power of the tests for the L_2, L_4, and L_∞ norms (with cut-off chosen using the permutation method) with $n = 100$ and $m = 36$. Clearly, in this case the L_∞ norm test outperforms the L_4 test which outperforms the L_2 norm test, as the separation increases.

Acknowledgements

The authors thank the anonymous referees for their valuable comments. B. B. Bhattacharya was supported by NSF CAREER grant DMS 2046393, NSF grant DMS 2113771, a Sloan Research Fellowship, and Wharton Dean's Research Fund.

References

[1] J. Agterberg, M. Tang, and C. Priebe, Nonparametric two-sample hypothesis testing for random graphs with negative and repeated eigenvalues, `arXiv:2012.09828`, 2020.

[2] S. Bhadra, K. Chakraborty, S. Sengupta, and S. Lahiri, A bootstrap-based inference framework for testing similarity of paired networks, `arXiv:1911.06869`, 2019.

[3] B. Bollobas, S. Janson, and O. Riordan, The phase transition in inhomogeneous random graphs, *Random Structures and Algorithms*, Vol. 31 (1), 3–122, 2007.

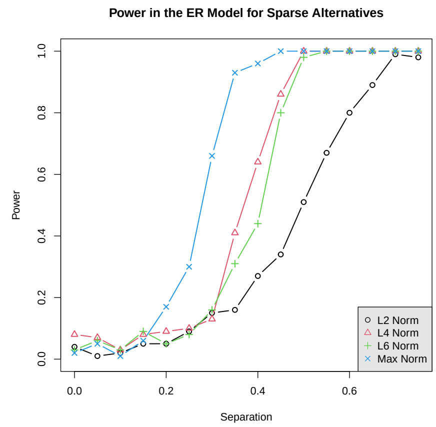


Figure 3: Power of the different tests in the Erdős-Rényi model as a function of the separation ε with $n = 100$ and $m = 36$, when $P^{(n)}$ and $Q^{(n)}$ differ in 10 coordinates.

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

[5] D. S. Bassett, E. Bullmore, B. A. Verchinski, V. S. Mattay, D. R. Weinberger, and A. Meyer-Lindenberg, Hierarchical organization of human cortical networks in health and schizophrenia, *The Journal of Neuroscience*, Vol. 28(37), 9239–9248, 2008.

[6] P. J. Bickel and P. Sarkar, Hypothesis testing for automated community detection in networks, *Journal of the Royal Statistical Society Series B: Statistical Methodology*, Vol. 78(1), 253–273, 2016.

[7] G. Bravo-Hermsdorff, L. M. Gunderson, P. A. Maugis, and C. E. Priebe, A principled (and practical) test for network comparison, `arXiv:2107.11403`, 2021.

[8] S. Chatterjee, P. Diaconis, and A. Sly, Random graphs with a given degree sequence, *Annals of Applied Probability*, Vol. 21(4), 1400–1435, 2011.

[9] J. Chhor and A. Carpentier, Sharp local minimax rates for goodness-of-fit testing in large random graphs, multivariate Poisson families and multinomials, `arXiv:2012.13766`, 2020.

[10] F. Chung and L. Lu, Connected components in random graphs with given expected degree sequences, *Annals of Combinatorics*, Vol. 6, 125–145, 2002.

[11] S. Dan and B. B. Bhattacharya, Goodness-of-fit tests for inhomogeneous random graphs, *Proceedings of the 37th International Conference on Machine Learning*, Vol. 119, 2335-2344, 2020.

[12] A. Elliott, E. Leicht, A. Whitmore, G. Reinert, and F. Reed-Tsochas, A nonparametric significance test for sampled networks. *Bioinformatics*, Vol. 34(1), 64-71, 2017.

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

[14] D. Ghoshdastidar and U. von Luxburg, Practical methods for graph two-sample testing, *Neural Information Processing Systems (NeurIPS)*, 3019-3028, 2018.

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

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

[17] P. Holland and S. Leinhardt, An exponential family of probability distributions for directed graphs, *Journal of the American Statistical Association*, Vol. 76, 33–65, 1981.

[18] D. R. Hunter, MM algorithms for generalized Bradley–Terry models, *Annals of Statistics*, Vol. 32, 384–406, 2004.

[19] D. R. Hyduke, N. E. Lewis, and B. Palsson, Analysis of omics data with genome-scale models of metabolism, *Molecular BioSystems*, Vol. 9 (2), 167–174, 2013.

[20] Y. I. Ingster. The minimax test of nonparametric hypothesis on a distribution density in metrics l_p, *Teoriya Veroyatnostei i ee Primeneniya*, Vol. 31 (2), 384–389, 1986.

[21] Y. I. Ingster and I. A. Suslina, *Nonparametric Goodness-of-Fit Testing Under Gaussian Models*, Lecture Notes in Statistics, Springer, 2003.

[22] E. D. Kolaczyk, *Topics at the Frontier of Statistics and Network Analysis: (Re) visiting the Foundations*, Cambridge University Press, 2017.

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

[24] Y. Li, and H. Li, Two-sample test of community memberships of weighted stochastic block models, `arXiv:1811.12593`, 2018.

[25] L. Lovász, *Large networks and graph limits*, AMS, Providence, RI, 2012.

[26] P. A. Maugis, S.C. Olhede, C. E. Priebe, and P. J. Wolfe, Testing for equivalence of network distribution using subgraph counts, *Journal of Computational and Graphical Statistics*, Vol. 29(3), 455–465, 2020.

[27] E. Mossel, J. Neeman, and A. Sly, Consistency thresholds for the planted bisection model, *Electronic Journal of Probability*, Vol. 21, 1–24, 2016.

[28] R. Mukherjee, S. Mukherjee, S. Sen, Detection thresholds for the beta model in sparse graphs, *Annals of Statistics*, Vol. 46, 1288–1317, 2018.

[29] L. Ospina-Forero, C. M. Deane, and G. Reinert, Assessment of model fit via network comparison methods based on subgraph counts, *Journal of Complex Networks*, Vol. 2, 226–253, 2019.

[30] J. Park and M. E. J. Neuman, Statistical mechanics of networks, *Phys. Rev. E (3)*, Vol. 70, 066117, 13, 2004.

[31] P. Rubin-Delanchy, J. Cape, M. Tang, C. E Priebe, A statistical interpretation of spectral embedding: the generalised random dot product graph, `arXiv:1709.05506`, 2021.

[32] M. Sabanayagam, L. C. Vankadara, and D. Ghoshdastidar, Graphon based clustering and testing of networks: Algorithms and theory, *International Conference on Learning Representations*, 2022.

[33] M. Shao, D. Xia, Y. Zhang, Q. Wu, and S. Chen, Higher-order accurate two-sample network inference and network hashing, `arXiv:2208.07573`, 2022.

[34] M. Tang, A. Athreya, D. L. Sussman, V. Lyzinski, and C. E. Priebe, A nonparametric two-sample hypothesis testing problem for random graphs, *Bernoulli*, Vol. 23, 1599–1630, 2017.

[35] 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*, Vol. 26 (2), 344–354, 2016.

[36] Y. Wang, M. Tang, and S. N. Lahiri, Two-sample testing on latent distance graphs with unknown link functions, `arXiv:2008.01038`, 2020.

[37] P. Wills and F. G. Meyer, Metrics for graph comparison: a practitioner?s guide. *Plos one*, Vol. 15(2), e0228728, 2020.

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