Bernoulli **30**(3), 2024, 1746–1773
https://doi.org/10.3150/23-BEJ1651

Signal detection in degree corrected ERGMs

YUANZHE XU[a] and SUMIT MUKHERJEE[b]

Columbia University, NY, USA, [a]*yuanzhe.xu@columbia.edu,* [b]*sm3949@columbia.edu*

In this paper, we study sparse signal detection problems in “degree corrected” Exponential Random Graph Models (ERGMs). We study the performance of two tests based on conditionally centered sum of degrees/maximum of degrees, for a wide class of such ERGMs. The performance of these tests match the performance of corresponding uncentered tests in the $\boldsymbol{\beta}$ model (*Ann. Statist.* **46** (2018) 1288–1317). Focusing on the degree corrected two star ERGM, we show that improved detection is possible at “criticality” using a test based on (unconditional) sum of degrees. In this setting we provide matching lower bounds in all parameter regimes, which is based on correlations estimates between degrees under the alternative, and is of possible independent interest.

Keywords: Asymptotic efficiency; auxiliary variables; ERGM; phase transition; signal detection; two star

1. Introduction

Studying network models has a long and rich history in Statistics, with applications across various disciplines such as Social Science, Biology, Neuroscience, Climatology, and Ecology, to name a few. One of the most well known network models is the Exponential Random Graph Model (which are often abbreviated as ERGM). ERGMs originated in the Social Science Literature (c.f. Anderson, Wasserman and Crouch (1999), Frank and Strauss (1986), Holland and Leinhardt (1981), Robins et al. (2007), Wasserman and Faust (1994), Wasserman and Pattison (1996) and the references there-in), and have since then received considerable attention in Statistics and Probability (c.f. Chatterjee and Diaconis (2013), Chatterjee, Diaconis and Sly (2011), Götze, Sambale and Sinulis (2021), Mukherjee, Mukherjee and Sen (2018), Mukherjee, Mukherjee and Yuan (2018), Schweinberger and Stewart (2020), Shalizi and Rinaldo (2013) and references there-in). ERGMs represent exponential families of distributions that are defined on the space of simple labeled graphs with a finite dimensional sufficient statistics, which are usually taken to be subgraph counts. The simplest class of examples under this framework consists of the one parameter ERGM, which admits a one dimensional sufficient statistic. Below we start by introducing such a one parameter ERGM:

Letting $\mathcal{G}_n$ denote the set of all simple labeled graphs G with vertex set $[n] := \{1,2,\dots,n\}$, we consider the following probability mass function on $\mathcal{G}_n$:

$$\mathbb{P}_{n,\theta}(G) := \frac{1}{Z_n(\theta,H)} \exp\left\{\theta \frac{N(H,G)}{n^{\zeta-2}}\right\}. \tag{1}$$

Here

(i) H is a graph of fixed size (such as an edge, triangle, cycle, star, etc.),
(ii) $N(H,G)$ is the number of copies of the graph H in the graph G,
(iii) ζ is the number of vertices in the graph H,
(iv) θ is a real valued parameter,
(v) $Z_n(\theta,H)$ is the normalizing constant.

In particular if the graph H is an edge, then $N(H,G)$ is the number of edges in G. In this case, the model in (1) reduces to an Erdős-Rényi model, where the edges of the graph G are i.i.d from a suitable

1350-7265

Bernoulli distribution. For any other choice of H, the model in (1) is not an Erdős-Rényi model since one allows nontrivial dependence between the edges. An ERGM can thus be thought of as a natural generalization of the Erdős-Rényi model, which allows for growing degrees of dependence between edges through the term $N(H,G)$. It is natural to allow for this dependence while modeling networks, to incorporate features like "friends of friends are more likely to be friends". However, one drawback of ERGMs (or at least the model introduced in (1)) is that the edges of the random graph are still jointly exchangeable, in the sense that permuting the labels of vertices of G does not change the distribution of the graph G. Consequently each coordinate of the degree sequence $(d_1,\ldots,d_n)$ marginally has the same distribution for all $i \in [n]$. This may not be desirable for modeling networks where there are a few vertices of very high degree (see Bhamidi, Steele and Zaman (2015)), when compared to the remaining vertices. Such a feature is often present in social networks, where the vertex corresponding to a popular/famous person has a very high degree compared to the remaining vertices.

One model which captures degree heterogeneity is the β-model of social networks (c.f. (Blitzstein and Diaconis, 2011, Chatterjee, Diaconis and Sly, 2011, Chatterjee and Mukherjee, 2019, Mukherjee, Mukherjee and Sen, 2018, Rinaldo, Petrović and Fienberg, 2013) and references there-in). The β-model is defined by the following p.m.f. on $\mathcal{G}_n$:

$$\mathbb{P}_{n,\boldsymbol{\beta}}(G) := \frac{1}{Z_n(\boldsymbol{\beta})} \exp\Big\{\sum_{i=1}^n \beta_i d_i\Big\}. \tag{2}$$

Here

(i) $(d_1,\ldots,d_n)$ is the degree sequence of the graph G.
(ii) $\boldsymbol{\beta} = (\beta_1,\ldots,\beta_n)^T \in \mathbb{R}^n$ is a vector valued parameter,
(iii) $Z_n(\boldsymbol{\beta})$ is the normalizing constant.

In this model, for each vertex $i \in [n]$ there is a real valued parameter β_i which controls the effect of the i^{th} vertex, and consequently the typical size of the degree d_i. This allows for heterogeneity among the degrees. A large value of β_i results in a large value of the degree of the i^{th} vertex, and vice versa. One drawback of the β-model (2) is that the edges of the graph G are no longer dependent. This is not immediate from (2), but is not hard to check (see for e.g. Chatterjee, Diaconis and Sly (2011)). Thus although the β-model allows for degree heterogeneity, it does not involve dependence between the edges.

A natural way to retain both the dependence between edges and the heterogeneity of the degrees is to consider an exponential family which has both the terms $\theta N(H,G)$ and $\sum_{i=1}^n \beta_i d_i$ in the exponent. Indeed, dependence between edges is present because of the term $\theta N(H,G)$, and degree heterogeneity is present because of the term $\sum_{i=1}^n \beta_i d_i$. Such a model, which we introduce formally below, can be thought of as a degree corrected ERGM.

1.1. Degree corrected ERGMs

As before, let $\mathcal{G}_n$ denote the set of all simple labelled graphs G with vertex set $[n] := \{1,2,\ldots,n\}$. Given a graph $G \in \mathcal{G}_n$, by slight abuse of notation we use G to also denote the adjacency matrix of G, defined as follows:

$$G_{ij} = \begin{cases} 1 & \text{If an edge is present between vertices } i \text{ and } j \text{ in } G, \\ 0 & \text{If no edge is present between vertices } i \text{ and } j \text{ in } G. \end{cases} \tag{3}$$

For any edge $e = (i,j)$ we set $G_e = G_{ij}$ by convention. Thus, we encode presence or absence of edges by $\{0,1\}$. By convention, set $G_{ii} := 0$, and note that G is a symmetric $n \times n$ matrix with 0 on the diagonal, and $\{0,1\}$ entries on the off-diagonals. Let $(d_1,d_2,\cdots,d_n)$ denote the labeled degree sequence of the graph G, defined by

$$d_i := \sum_{j=1}^{n} G_{ij}, 1 \le i \le n.$$

Let H be a fixed connected subgraph with $\zeta := |V(H)| \ge 2$, where $V(H)$ denotes the set of vertices of H and $|V(H)|$ denotes its cardinality. In particular, this means H is not an isolated vertex. Assume that the vertices of H are labeled as $[\zeta] = \{1,2,\ldots,\zeta\}$. Let $\mathcal{I}_n$ denote the the set of all 1-1 maps from $[\zeta]$ to $[n]$. For any $G \in \mathcal{G}_n$, let $N(H,G)$ denote the number of copies of H in G, defined by

$$N(H,G) = \sum_{\iota \in \mathcal{I}_n} \prod_{(i,j)\in E(H)} G_{\iota(i),\iota(j)},$$

where $E(H) := \{(a,b) \in V(H) : (a,b) \text{ is an edge in } H\}$ is the edge set of H. As for illustration, the expression of $N(H,G)$ when H is an edge, a triangle, and a two star (to be denoted by $K_2, K_3, K_{1,2}$ respectively) are given by:

$$N(K_2,G) = \sum_{i\neq j} G_{ij} = 2\sum_{i<j} G_{ij} = \sum_{i=1}^{n} d_i,$$
$$N(K_3,G) = \sum_{i\neq j\neq k} G_{ij}G_{jk}G_{ki} = 6\sum_{i<j<k} G_{ij}G_{jk}G_{ki},$$
$$N(K_{1,2},G) = \sum_{i\neq j\neq k} G_{ij}G_{ik} = 2\sum_{i=1}^{n}\sum_{j<k} G_{ij}G_{ik} = 2\sum_{i=1}^{n}\binom{d_i}{2}.$$

Here, by a two star, we mean a path of length 2, which has 3 vertices and 2 edges. Given a parameter $\theta > 0$ and vector $\boldsymbol{\beta} = (\beta_1,\beta_2,\ldots,\beta_n) \in \mathbb{R}^n$, we subsequently define a probability mass function on $\mathcal{G}_n$ by setting

$$\mathbb{P}_{n,\theta,\boldsymbol{\beta}}(G) := \frac{1}{Z_n(\boldsymbol{\beta},\theta,H)}\exp\Big\{\frac{\theta}{n^{\zeta-2}}N(H,G) + \sum_{i=1}^{n}\beta_i d_i\Big\}, \tag{4}$$

where as usual $Z_n(\boldsymbol{\beta},\theta,H)$ is the normalizing constant. The scaling $n^{\zeta-2}$ ensures that the resulting model is nontrivial as $n\to\infty$ (c.f. Chatterjee and Diaconis (2013)). If $\beta_i = \beta_0$ for some $\beta_0 \in \mathbb{R}$ free of i, then the model in (4) is an Exponential Random Graph Model with two sufficient statistics $N(H,G)$ and $E(G)$, where $E(G) = \frac{1}{2}N(K_2,G)$ is the number of edges in the graph G. In this case the random graph G represents a bivariate exchangeable array. More precisely, for any permutation $\pi \in S_n$ the graph G_π defined by $G_\pi(i,j) := G_{\pi(i),\pi(j)}$ has the same distribution as G, i.e. $G_\pi \stackrel{D}{=} G$. The vector of parameters $\boldsymbol{\beta}$, therefore, measures the individual effects of each vertex, and for a general vector $\boldsymbol{\beta}$ a random graph G from the model (4) is no longer exchangeable. For $\theta > 0$, the term $N(H,G)$ ensures that there is positive dependence among the edges in G, in the sense that conditional on presence of an edge, any other edge is more likely to be present. If $\theta = 0$, the model (4) reduces to the β-model as in (2), in which all edges G_{ij} are independent, with

$$\mathbb{P}_{n,0,\boldsymbol{\beta}}(G_{ij} = 1) = \frac{e^{\beta_i+\beta_j}}{1+e^{\beta_i+\beta_j}}.$$

Thus the model in (4) combines the features of the β-model and traditional ERGMs. We will use the term degree corrected ERGM to refer to the model (4).

1.2. Hypothesis testing problem for $\boldsymbol{\beta}$

Given the model (4), a natural question is to carry out inference regarding the vector $\boldsymbol{\beta}$. In the setting where $\theta = 0$, the problem of estimation of $\boldsymbol{\beta}$ using the MLE $\hat{\boldsymbol{\beta}}_{ML}$ was studied in Chatterjee, Diaconis and Sly (2011), where the authors gave bounds on $||\hat{\boldsymbol{\beta}}_{ML} - \boldsymbol{\beta}||_\infty$. The question of testing of the grand null hypothesis $\boldsymbol{\beta} = \mathbf{0}$ versus non negative sparse alternatives was studied in Mukherjee, Mukherjee and Sen (2018), where the authors show that the form of the consistent test depends on the sparsity level and strength of the signal. Since both these papers assumed $\theta = 0$, the edges of the graph G were independent, which was used significantly in the proofs of the results. A natural question is whether one can extend these results in the presence of dependence between edges. In this paper, we study the question of testing the grand null hypothesis $\boldsymbol{\beta} = \beta_0\mathbf{1}$ against sparse one sided alternatives, when the parameter $\beta_0 \in \mathbb{R}, \theta > 0$ and the graph H are known. Essentially we want to test the null hypothesis that all nodes in the network are equally popular (i.e. have the same β_i), versus the alternative hypothesis that there is a small hub of nodes which are more popular (have a higher value of β_i) compared to the baseline popularity β_0 of the remaining nodes. In section 1.4 we briefly discuss what can go wrong if the parameter β_0 is not assumed to be known. Below we formally introduce the testing problem.

Let $\beta_0 \in \mathbb{R}$ be known. Let G be a graph drawn from the probability distribution (4), and for a known $\theta > 0$ and given $\beta_0 \in \mathbb{R}$ we consider the following hypothesis testing problem:

$$\mathcal{H}_0 : \boldsymbol{\beta} = \beta_0\mathbf{1} \quad \text{vs} \quad \mathcal{H}_1 : \boldsymbol{\beta} \in \Xi(s, A). \tag{5}$$

Here under the null hypothesis we have $\beta_i = \beta_0$ for all $i \in [n]$ and we denote this null probability measure as $\mathbb{P}_{n,\theta,\beta_0}$. The set of vectors $\Xi(s, A)$ in the alternative hypothesis H_1 is defined as

$$\Xi(s, A) := \left\{ \boldsymbol{\beta} = \beta_0\mathbf{1} + \boldsymbol{\mu} : |\mathrm{supp}\boldsymbol{\mu}| \geq s, \text{ and } \min_{i \in \mathrm{supp}\boldsymbol{\mu}} \mu_i \geq A \right\}. \tag{6}$$

In words, under the alternative hypothesis there is a sparse set S of size s, such that $\beta_i \geq \beta_0 + A$ if $i \in S$, and $\beta_i = \beta_0$ if $i \notin S$. Our main goal of this paper is to study the effect of the nuisance parameter θ on the hypothesis testing problem (5). For studying the proposed hypothesis testing problem, here we adopt an asymptotic minimax framework similar to (Mukherjee, Mukherjee and Sen, 2018, Mukherjee, Mukherjee and Yuan, 2018), which is introduced below (see also (Burnašev, 1979, Ingster, 1994, Ingster, Ingster and Suslina, 2003)).

Suppose $T_n : \mathcal{G}_n \mapsto \{0, 1\}$ is a non-randomized test function. If $T_n = 1$, we reject the null hypothesis H_0, and if $T_n = 0$, we do not reject the null hypothesis H_0. Define the risk of test $T_n(G)$ as the sum of type I and type II errors, as follows:

$$R(T_n, \Xi(s, A), \boldsymbol{\beta}) := \mathbb{P}_{n,\theta,\beta_0}(T_n(G) = 1) + \sup_{\boldsymbol{\beta} \in \Xi(s,A)} \mathbb{P}_{n,\theta,\boldsymbol{\beta}}(T_n(G) = 0). \tag{7}$$

Given a sequence of test functions $\{T_n\}_{n\geq1}$ for the testing problem (5), we call $\{T_n\}_{n\geq1}$ as

(i) Asymptotically Powerful, if

$$\lim_{n\to\infty} R(T_n, \Xi(s, A), \boldsymbol{\beta}) = 0; \tag{8}$$

(ii) Asymptotically not Powerful, if

$$\liminf_{n\to\infty} R(T_n, \Xi(s,A), \boldsymbol{\beta}) > 0; \tag{9}$$

(iii) Asymptotically Powerless, if

$$\lim_{n\to\infty} R(T_n, \Xi(s,A), \boldsymbol{\beta}) = 1. \tag{10}$$

By definition, both type I and type II errors converge to 0 for asymptotically powerful tests. Also, if a sequence of tests is asymptotically powerless, then it is also asymptotically not powerful, and so (iii) is a stronger notion than (ii).

1.3. Main results

In this section we present and discuss our main results. To that end, we first consider general degree corrected ERGMs and analyze the performance of two natural tests. We then focus on a particular degree corrected ERGM, where the graph H is a two star. In this setting we show that the general tests studied above attains the "optimal detection boundary" for all configurations (θ, β_0) barring a specific point, which we refer to as the critical point/configuration. At this point, using a slightly different test from the ones studied under the general ERGM framework, we are able to detect much lower signals, compared to the independent case ($\theta = 0$).

1.3.1. General degree corrected ERGMs

In this section, we discuss the hypothesis testing problem (5) in the setting of general degree corrected ERGMs as in (4). Specifically, we will show how signal density and strength (s, A) coordinate together to determine the threshold for testing efficiency. Two natural test statistics for this problem are the sum of degrees $\sum_{i=1}^n d_i$, and the maximum degree $\max_{i\in[n]} d_i$. However, because of dependence, it is very difficult to calibrate the cut-off for these statistics, as they depend on the parameter θ in a nontrivial way. To counter this, we use conditionally centered versions of the sum of degrees, and the maximum degree, similar to what was done in Mukherjee, Mukherjee and Yuan (2018).

Our first theorem studies the performance of a test based on conditionally centered sum of degrees. For stating the result we require a few notations.

Definition 1.1. Let $\mathcal{E} := \{(i,j) : 1 \le i < j \le n\}$ be the set of all edges in the complete graph K_n. For any $e = (i,j) \in \mathcal{E}$, let $N_e(H,G)$ denote the number of copies of H in the graph G which contains the edge e, and let $N_{e,f}(H,G)$ denote the number of copies of H in the graph G which contains both the edges e, f.

Setting $\psi(x) := \frac{e^x}{1+e^x}$ for $x \in \mathbb{R}$, for any $e = (i,j) \in \mathcal{E}$ we have

$$\mathbb{E}_{n,\theta,\boldsymbol{\beta}}\Big(G_e \Big| G_f : f \neq e\Big) = \psi(\theta t_e(H,G) + \beta_i + \beta_j), \tag{11}$$

where $t_e(H,G) := \frac{N_e(H,G)}{n^{\zeta-2}}$.

Since our results are asymptotic in nature, below we introduce some standard notations, to be used in the remainder of the paper.

Definition 1.2. Given two sequences of real numbers $\{a_n\}_{n\geq 1}$ and $\{b_n\}_{n\geq 1}$, we will use the notation $a_n = O(b_n)$ or $a_n \lesssim b_n$ to imply the existence of a positive finite constant c free of n, such that $a_n \leq cb_n$. We use the notation $a_n \gg b_n$ ($a_n \ll b_n$) to imply $\lim_{n\to\infty} \frac{a_n}{b_n} = \infty$ ($\lim_{n\to\infty} \frac{a_n}{b_n} = 0$ respectively).

Theorem 1.1. *With G from the model (4), consider the hypothesis testing problem described in (5) with (θ, β_0) known. If $sA \to \infty$, then for any sequence L_n such that $n \ll L_n \ll nsA$ the conditionally centered sum of degrees test $T_n(G)$ given by*

$$T_n(G) = \begin{cases} 1 & \text{if } \sum_{e\in\mathcal{E}} \left[G_e - \mathbb{E}_{n,\theta,\boldsymbol{\beta}_0}\left(G_e \middle| G_f : f \neq e\right)\right] > L_n, \\ 0 & \text{otherwise}, \end{cases}$$

is asymptotically powerful.

In settings where the signal size s is small, a test based on the conditionally centered maximum of degrees can sometimes detect lower signals. The performance of this test is studied in our second result.

Theorem 1.2. *With G from the model (4), consider the hypothesis testing problem described in (5) with (θ, β_0) known. Then there exists constants κ, C such that if $A \geq \kappa\sqrt{\frac{\log n}{n}}$ and $L_n = C\sqrt{n \log n}$, then the conditionally centered maximum degree test defined by*

$$T_n(G) = \begin{cases} 1 & \text{if } \max_{i\in[n]} \sum_{e\ni i} \left[G_e - \mathbb{E}_{n,\theta,\boldsymbol{\beta}_0}\left(G_e \middle| G_f : f \neq e\right)\right] > L_n, \\ 0 & \text{otherwise}, \end{cases}$$

is asymptotically powerful.

Note that the conditionally centered sum test of Theorem 1.1 requires $A \gg \frac{1}{s}$ to be asymptotically powerful, whereas the conditionally centered max test of Theorem 1.2 requires $A \geq \kappa\sqrt{\frac{\log n}{n}}$ for the same. Comparing the two thresholds $\frac{1}{s}$ and $\kappa\sqrt{\frac{\log n}{n}}$ of these two tests yields that the conditionally centered maximum degree test is better (has a lower detection boundary) for sparser alternatives (more precisely, $s \ll \sqrt{\frac{n}{\log n}}$), whereaas the conditionally centered sum of degrees test is better for denser alternatives ($s \gg \sqrt{\frac{n}{\log n}}$). This is similar to the findings of Mukherjee, Mukherjee and Sen (2018), where it was shown that optimal rate detection is obtained by the sum of degrees if $s = n^b$ with $b > 1/2$ (see (Mukherjee, Mukherjee and Sen, 2018, Theorem 3.1)), and by the maximum degree test if $b < 1/2$ (see (Mukherjee, Mukherjee and Sen, 2018, Theorem 3.3)).

1.3.2. Degree corrected two star ERGM

In Theorems 1.1 and 1.2, there is no effect of the nuisance parameter θ on the detection rate of the tests. To demonstrate that the best possible detection rate can change depending on the value of θ, we study in detail the degree corrected two star ERGM, The two star is the graph $K_{1,2}$, which is a path of length 3. For notational and computational convenience, for the degree corrected two star ERGM our edge variables take values in $\{-1, 1\}$ instead of $\{0, 1\}$. More precisely, given a graph $G \in \mathcal{G}_n$, our adjacency matrix Y is now defined as follows:

$$Y_{ij} = \begin{cases} 1 & \text{if } (i,j) \text{ is an edge in } G, \\ -1 & \text{otherwise}. \end{cases}$$

As before, we set $Y_{ii} = 0$ by convention. Thus Y is a symmetric matrix with $\{-1,1\}$ entries, and 0 on the diagonal. Let $(k_1, k_2, \cdots, k_n)$ denote the labeled "degree sequence" of the graph Y, i,e,

$$k_i := \sum_{j=1}^{n} Y_{ij}, 1 \le i \le n.$$

The following display introduces the degree corrected two star ERGM as a p.m.f. on $\{-1,1\}^{\binom{n}{2}}$:

$$\mathbb{P}_{n,\theta,\boldsymbol{\beta}}(Y) = \frac{1}{Z_n(\boldsymbol{\beta},\theta)} \exp\left\{\frac{\theta}{n-1}\widetilde{N}(K_{1,2},G) + \frac{1}{2}\sum_{i=1}^{n}\beta_i k_i\right\}, \tag{12}$$

where

$$\widetilde{N}(K_{1,2},G) := \sum_{i=1}^{n}\sum_{j<k} Y_{ij}Y_{ik} = \frac{1}{2}\sum_{i=1}^{n} k_i^2 - \frac{n(n-1)}{2}.$$

Having observed Y, consider the same hypothesis testing problem (5) as above. For the sake of clarity of presentation, in this section we parametrize the signal size s and signal strength A by n^b and n^t respectively, where $b \in (0,1)$ and $t < 0$. The detection boundary for this problem demonstrates a phase transition depending on the nuisance parameter θ. Stating this requires the following partitioning of the parameter space for (θ, β_0):

Definition 1.3.

- Let $\Theta_1 = \Theta_{11} \cup \Theta_{12}$, where $\Theta_{11} := (0,1/2) \times \{0\}$, and $\Theta_{12} = \{(\theta,\beta_0) : \theta > 0, \beta_0 \neq 0\}$.
- Let $\Theta_2 := (1/2,\infty) \times \{0\}$.
- Let $\Theta_3 := (1/2, 0)$.
  Note that $\Theta_1 \cup \Theta_2 \cup \Theta_3 = (0,\infty) \times \mathbb{R}$.

Our first lower bound result describes the detection boundary for the degree corrected two star ERGM if $(\theta, \beta_0) \in \Theta_1$. Recall the notions of asymptotically powerful and asymptotically powerless from (8) and (10) respectively.

Theorem 1.3. *Let Y be an observation from from (12), and assume $(\theta,\beta_0) \in \Theta_1$ is known. Consider the hypothesis testing problem described in (5) with $s = n^b$ and $A = n^t$ for $\in (0,1)$ and $t < 0$.*

(a) *If $b \ge \frac{1}{2}$ and $b + t < 0$, all tests are asymptotically powerless.*
(b) *If $b \ge \frac{1}{2}$ and $b + t > 0$, then the conditionally centered sum test of Theorem 1.1 is asymptotically powerful.*
(c) *If $b < \frac{1}{2}$ and $t + \frac{1}{2} \le 0$ then all tests are asymptotically powerless.*
(d) *If $b < \frac{1}{2}$ and $t + \frac{1}{2} > 0$ then the conditionally centered max test of Theorem 1.2 is asymptotically powerful.*

Our second lower bound result describes the detection boundary for the degree corrected two star ERGM if $(\theta, \beta_0) \in \Theta_2$. Recall the definition of asymptotically not powerful from (9).

Theorem 1.4. *Let Y be an observation from from (12), and assume $(\theta,\beta_0) \in \Theta_2$ is known. Consider the hypothesis testing problem described in (5) with $s = n^b$ and $A = n^t$ for $\in (0,1)$ and $t < 0$.*

(a) *If* $b \geq \frac{1}{2}$ *and* $b + t < 0$, *all tests are asymptotically not powerful.*
(b) *If* $b \geq \frac{1}{2}$ *and* $b + t > 0$, *then the conditionally centered sum test of Theorem 1.1 is asymptotically powerful.*
(c) *If* $b < \frac{1}{2}$ *and* $t + \frac{1}{2} \leq 0$ *then all tests are asymptotically not powerful.*
(d) *If* $b < \frac{1}{2}$ *and* $t + \frac{1}{2} > 0$ *then the conditionally centered max test of Theorem 1.2 is asymptotically powerful.*

Note that at a qualitative level, the detection boundary in the regimes Θ_1 and Θ_2 are the same. The only difference is that below the detection boundary, in domain Θ_1 Theorem 1.3 shows that all tests are powerless, and in domain Θ_2 Theorem 1.4 shows that all tests are asymptotically not powerful. On the other hand, something fundamentally different happens in the critical domain Θ_3, which corresponds to the choice $(\theta, \beta_0) = (1/2, 0)$. In this case the optimal testing threshold is significantly lower than the other regimes, and does not depend on whether $b < 1/2$ or $b > 1/2$. Moreover, we note that this improved performance does not follow from either Theorem 1.1 or 1.2. In this case a test based on the unconditional sum of degrees attains the optimal detection boundary, for all values of (s, A). This is explained in our final result below.

Theorem 1.5. *Let* Y *be an observation from from* (12), *and assume* $(\theta, \beta_0) = (\frac{1}{2}, 0)$ *is known. Consider the hypothesis testing problem described in* (5), *with* $s = n^b$ *and* $A = n^t$ *for some* $b \in (0, 1)$ *and* $t < 0$.

(a) *If* $b + t + \frac{1}{2} < 0$, *then all tests are asymptotically powerless.*
(b) *If* $b + t + \frac{1}{2} > 0$, *then the total degree test* $T_n(.)$ *defined by*

$$T_n(G) = \begin{cases} 1 & \text{if } \sum_{i=1}^n k_i > L_n, \\ 0 & \text{otherwise}, \end{cases}$$

is asymptotically powerful for some sequence L_n *satisfying* $L_n \gg n^{3/2}$.

This demonstrates that the much weaker criterion $b + t + \frac{1}{2} > 0$ is enough for detection at criticality, whereas away from criticality we need stronger conditions on b, t. Similar phenomenon of improved detection at criticality have been observed for Ising models Deb et al. (to appear), Mukherjee, Mukherjee and Sen (2018), Mukherjee, Mukherjee and Yuan (2018). Given that the two star ERGM can be viewed as an Ising model, it is thus not surprising that this continues to hold here. A summary of the detection boundary for the degree corrected two star ERGM is given in Figure 1 below.

1.4. Main contributions and future scope

In this paper we introduce degree corrected ERGMs, which combine the above two concepts traditional ERGMs with the $\boldsymbol{\beta}$-model and thereby allowing for not only degree heterogeneity but also dependence between the edges. We study the performance of two tests, which are based on conditionally centered sum of degrees, and conditionally centered maximum degree. The detection rate of these two tests match the performance of the corresponding tests based on unconditionally centered sum of degree and unconditional maximum degree, respectively, in the independent case ($\theta = 0$). To explore the sharpness of these general tests, we subsequently study the degree corrected two star ERGM in detail. Here we show that in all parameter configurations other than $(\theta, \beta_0) = (1/2, 0)$, the optimal detection boundary is attained by one of the conditionally centered tests. At the critical configuration $(\theta, \beta_0) = (1/2, 0)$, we show that the optimal detection rate is significantly improved, and this optimal rate is attained by a test based on the unconditionally centered sum of degrees.

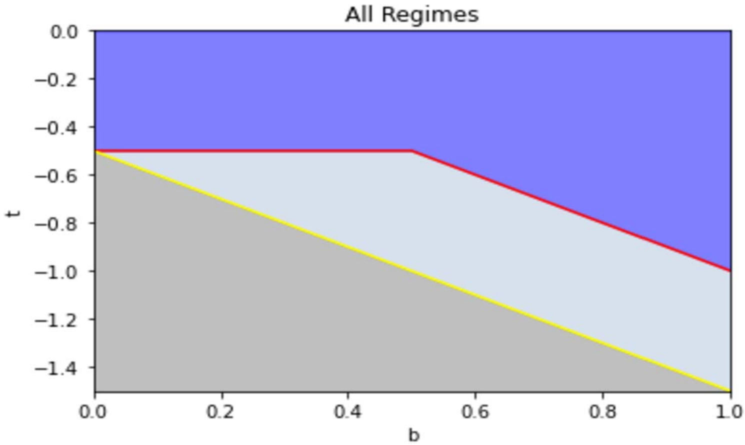

Figure 1. In this figure, we plot (b, t) along X and Y axis respectively, where $s = n^b$ is the size of the signal set, and $A = n^t$ is the magnitude of the signal. The range of b is $(0, 1)$, and the range of t is $(-\infty, 0)$. The deep blue portion of the plot represents the pairs (b, t) where detection is possible in all regimes $\Theta_1 \cup \Theta_2 \cup \Theta_3$. The light blue portion of the plot represents the pairs (b, t) where detection is possible Θ_3, but not for $\Theta_1 \cup \Theta_2$. Finally, the grey portion of the plot represents the pairs (b, t) where detection is impossible in all regimes $\Theta_1 \cup \Theta_2 \cup \Theta_3$. Also note that in $\Theta_1 \cup \Theta_2$ the optimal test depends on whether $b < 1/2$ or $b > 1/2$, whereas in Θ_3 the optimal test does not depend on b.

Throughout this paper we assume that the parameters (θ, β_0) are known. If (θ, β_0) is unknown, it may be possible to estimate (θ, β_0) if the signal (s, A) is small, by ignoring the signals altogether and estimating the parameters via the null model MLE/pseudo-likelihood. However such a strategy is hopeless for all values of (s, A), without the knowledge of (θ, β_0). Indeed, consider the following extreme configuration when $s = n, A = \infty$, in which case the graph G equals K_n with probability 1 for any value of β_0. On the other hand, if $s = A = 0$, but $\theta = \infty$, the observed graph is again K_n with probability 1 for any value of β_0. Thus having observed G, it is impossible to decide whether signal is present or absent, if we are not told the value of θ. It remains to be seen to what extent a partial knowledge of (θ, β_0) can help in our testing problem. Throughout, we also assume that the parameters (s, A) of the alternative hypothesis are also known. In fact, it follows from our proofs that the full knowledge of (s, A) are not required to derive the cut-off for our test statistic. The knowledge of lower bounds on (s, A) suffice for this purpose. Thus, as long as we know that $s \geq s_0, A \geq A_0$, and the parameters (s_0, A_0) fall in the regime where asymptotically powerful testing is possible, our proposed tests will be asymptotically powerful.

The analysis of the conditionally centered sum and maximum of degrees for general degree corrected ERGMs is achieved using concentration results based on the method of exchangeable pairs (Chatterjee (2007)). Focusing on the degree corrected two star ERGM, to verify the improved detection rate at criticality, we introduce a continuous auxiliary variable $\phi \in \mathbb{R}^n$ (similar to Mukherjee, Mukherjee and Yuan (2018)), and show that a suitable function of ϕ is stochastically much larger under the alternative than under the null hypothesis. Using this, we show that the unconditional sum of degrees is (stochastically much larger under the alternative, which gives the improved detection at criticality. The lower bound argument uses the second moment method, which reduces to bounding the correlation between the degrees under the alternative. In the regimes Θ_1 and Θ_3, using GHS inequality (Lebowitz (1974)) we can bound the correlations between the edges under the alternative by the correlation under the null, for which bounds are available from Mukherjee and Xu (2023), using exchangeability of the null model. In the regime Θ_2 we need to do a conditional second moment argument restricted to the set where the degrees are large. In the absence of a conditional GHS inequality, we have to directly bound the conditional correlations between the edges under the alternative. To do this, we make crucial use of the auxiliary variable ϕ and set up a recursive equation involving the correlations between degrees of the graph. This recursion leads to a uniform bound on the correlations which is also a tight upper bound (in terms of rate), and suffices for the second moment argument. It is of interest to see if one can set up similar recursive equations to bound correlation between edges in general (degree corrected) ERGMs, in presence/absence of auxiliary variables. In general, analysis of cubic and higher order ERGMs is a challenging problem. Preliminary calculations suggest that concentration results based on exchangeable pairs may be used to derive lower bounds for very high temperature (θ sufficiently small). However, it is unclear whether a general lower bound which applies to all $\theta > 0$ can be obtained, more so because even the phase transition boundaries are not entirely characterized for such ERGMs (up to the best of our knowledge).

In this paper we focus on the optimal detection rates while studying the detection boundary. A natural follow up question is to study existence of sharp constants (depending on θ, β_0) which controls the detection boundary for the degree corrected two star ERGM. Similar to Mukherjee, Mukherjee and Sen (2018), we expect a sharp phase transition (i.e. existence of a constant which determines the optimal detection boundary) in the regime $b < 1/2$, when $(\theta, \beta_0) \neq (1/2, 0)$. We believe that to attain optimal detection constants, one needs to study a conditionally centered version of the Higher Criticism Test in the regime $1/4 < b < 1/2$, wheres the maximum test should suffice in the regime $b < 1/2$. Going beyond the two star case, it is of interest to find optimal detection rates, both away from, and at, "criticality", for general degree corrected ERGMs. A major challenge in carrying out the lower bound argument beyond the two star case is the absence of tight correlation bounds for general ERGMs, both under the null and alternative hypotheses.

1.5. Outline

The outline of the paper is as follows. In section 2 we verify all the upper bound results of this paper, namely Theorem 1.1, Theorem 1.2, parts (b) and (d) of Theorem 1.3 and Theorem 1.4, and part (b) of Theorem 1.5. In section 3 we prove all the lower bound results, namely parts (a) and (c) of Theorem 1.3 and Theorem 1.4, and part (a) of Theorem 1.5. The proofs of the main results use some supporting lemmas, the proofs of which are deferred to the supplementary file (Xu and Mukherjee (2024)).

2. Proof of Theorems 1.1 and 1.2

We will need the following concentration bound for conditionally centered linear statistics for proving the results of this section. The proof of this lemma is similar to (Deb and Mukherjee, 2023, Lemma 2.1) and (Mukherjee, Mukherjee and Yuan, 2018, Lemma 1).

Lemma 2.1. *Let G be a random graph from the model (4). Then for any arbitrary collection of positive numbers $\{c_e\}_{e \in \mathcal{E}}$ and any $x > 0$ we have*

$$\mathbb{P}_{n,\theta,\boldsymbol{\beta}}\Big(\Big|\sum_{e \in \mathcal{E}} c_e \Big(G_e - \mathbb{E}_{n,\theta,\boldsymbol{\beta}}\big(G_e \big| G_f : f \neq e\big)\Big)\Big| > x\Big) \leq 2\exp\left\{-\frac{x^2}{\lambda \sum_{e \in \mathcal{E}} c_e^2}\right\}, \tag{13}$$

where $\lambda = \lambda(\theta, H)$ is a constant depending only on $\theta > 0$ and the subgraph H.

Proof. Produce an exchangeable pair (G,G') in the following way:

Pick a random vertex pair I uniformly from the set $\mathcal{E}$ with cardinality $N = \binom{n}{2}$. If $I = e$, replace the random variable G_e by G'_e, which is a pick from the conditional distribution given $\{G_f, f \neq e\}$. Let this new graph be denoted by G'. It is easy to verify that (G,G') is indeed an exchangeable pair. Setting $J(G) := \sum_{e\in\mathcal{E}} c_e G_e$, note that

$$\begin{aligned} h(G) := \mathbb{E}_{n,\theta,\boldsymbol{\beta}}\left(J(G) - J(G')\middle|G\right) =& \frac{1}{N}\sum_{e\in\mathcal{E}} c_e\left(G_e - \mathbb{E}_{n,\theta,\boldsymbol{\beta}}\left(G_e\middle|G_f : f \neq e\right)\right) \\ =& \frac{1}{N}J(G) - \frac{1}{N}\sum_{e\in\mathcal{E}} c_e \frac{\exp\left\{\frac{\theta}{n^{\zeta-2}}N_e(H,G) + \beta_e\right\}}{1 + \exp\left\{\frac{\theta}{n^{\zeta-2}}N_e(H,G) + \beta_e\right\}}, \end{aligned}$$

where $N_e(H,G)$ is the number of copies of H in the graph G, which contains the edge e. Using the fact that the derivative of the function $\psi(x) = \frac{e^x}{1+e^x}$ is bounded by $\frac{1}{4}$, this gives

$$\begin{aligned} |h(G) - h(G')| \leq& \frac{|c_I|}{N} + \frac{|\theta|}{4Nn^{\zeta-2}}\sum_{e\in\mathcal{E}} |c_e||N_e(H,G) - N_e(H,G')| \\ \leq& \frac{|c_I|}{N} + \frac{|\theta|}{4Nn^{\zeta-2}}\sum_{e\in\mathcal{E}} |c_e|N_{e,I}(H,K_n), \end{aligned}$$

where $N_{e,f}(H,K_n)$ is the number of copies of H in the complete graph K_n passing through both the edges e and f. Consequently, we have

$$\begin{aligned} &\left|\mathbb{E}_{n,\theta,\boldsymbol{\beta}}\left((h(G) - h(G'))(J(G) - J(G'))\middle|G\right)\right| \\ &\leq \frac{1}{N}\sum_{f\in\mathcal{E}} |c_f|\left[\frac{|c_f|}{N} + \frac{|\theta|}{4Nn^{\zeta-2}}\sum_{e\in\mathcal{E}} |c_e|N_{e,f}(H,K_n)\right] \\ &= \frac{1}{N^2}\sum_{f\in\mathcal{E}} c_f^2 + \frac{|\theta|}{4N^2n^{\zeta-2}}\sum_{e,f\in\mathcal{E}} N_{e,f}(H,K_n)|c_e||c_f| \\ &= \frac{1}{N^2}\sum_{e,f\in\mathcal{E}} B_N(e,f)|c_e||c_f|, \end{aligned}$$

where B_N is a $N \times N$ symmetric matrix defined by:

$$B_N(e,f) := \begin{cases} 1 \text{ if } e = f, \\ \frac{|\theta|}{4n^{\zeta-2}}N_{e,f}(H,K_n) \text{ if } e \neq f. \end{cases}$$

Now for any $e \neq f$ we have

$$N_{e,f}(H,K_n) \lesssim \begin{cases} n^{\zeta-4} & \text{if } e \text{ and } f \text{ have no vertex in common,} \\ n^{\zeta-3} & \text{if } e \text{ and } f \text{ have one vertex in common.} \end{cases}$$

This gives

$$\max_{e\in\mathcal{E}}\sum_{f\in\mathcal{E}}B_N(e,f)\lesssim 1+n^2\frac{1}{n^{\zeta-2}}n^{\zeta-4}+n\frac{1}{n^{\zeta-2}}n^{\zeta-3}\lesssim 1,$$

which in turn implies that the operator norm of the matrix B_N is $O(1)$, and consequently,

$$\left|\mathbb{E}_{n,\theta,\boldsymbol{\beta}}\Big((h(G)-h(G'))(J(G)-J(G'))\Big|G\Big)\right|\lesssim\frac{1}{N^2}\sum_{e\in\mathcal{E}}c_e^2\lesssim\frac{1}{n^4}\sum_{e\in\mathcal{E}}c_e^2.$$

Then by Stein's Method for concentration inequalities as in (Chatterjee, 2007, Theorem 1.5 (ii)), the conclusion of the lemma follows. □

2.1. Proof of Theorem 1.1

To begin, using Lemma 2.1 with $c_e=1$ for all $e\in\mathcal{E}$ gives the existence of a constant λ (depending only on θ,H) such that

$$\mathbb{P}_{n,\theta,\boldsymbol{\beta}_0}\Big(\Big|\sum_{e\in\mathcal{E}}\big(G_e-\mathbb{E}_{n,\theta,\boldsymbol{\beta}_0}(G_e\big|G_f:f\neq e)\big)\Big|>L_n\Big)\leq 2\exp\left\{-\frac{L_n^2}{\lambda\binom{n}{2}}\right\}\longrightarrow 0, \tag{14}$$

where the last limit uses $L_n\gg n$. This shows that type I error converges to 0.

It thus remains to show that type II error converges to 0. To this effect, note that $t_e(H,G)\leq t_e(H,K_n)$ which is bounded, and so therefore there exist a constant $\delta>0$ such that

$$\begin{aligned}
&\mathbb{E}_{n,\theta,\boldsymbol{\beta}}\big(G_e\big|G_f:f\neq e\big)-\mathbb{E}_{n,\theta,\boldsymbol{\beta}_0}\big(G_e\big|G_f:f\neq e\big)\\
&\quad=\psi(\theta t_e(H,G)+\beta_i+\beta_j)-\psi(\theta t_e(H,G)+2\beta_0)\\
&\quad\geq\delta\min\{\beta_i+\beta_j-2\beta_0,1\}.
\end{aligned}\tag{15}$$

Adding this gives

$$\sum_{e\in\mathcal{E}}\Big(\mathbb{E}_{n,\theta,\boldsymbol{\beta}}\big(G_e\big|G_f:f\neq e\big)-\mathbb{E}_{n,\theta,\boldsymbol{\beta}_0}\big(G_e\big|G_f:f\neq e\big)\Big)\geq\delta nsA.$$

Since $L_n\ll nsA$, for all n large we have

$$\begin{aligned}
&\mathbb{P}_{n,\theta,\boldsymbol{\beta}}\Big(\sum_{e\in\mathcal{E}}(G_e-\mathbb{E}_{n,\theta,\boldsymbol{\beta}_0}(G_e|G_f:f\neq e))\leq L_n\Big)\\
\leq&\mathbb{P}_{n,\theta,\boldsymbol{\beta}}\Big(\Big|\sum_{e\in\mathcal{E}}(G_e-\mathbb{E}_{n,\theta,\boldsymbol{\beta}}(G_e|G_f:f\neq e))\Big|\geq L_n\Big)\leq 2\exp\left\{-\frac{L_n^2}{\lambda\binom{n}{2}}\right\},
\end{aligned}$$

where we again invoke Lemma 2.1 in the last line above. This gives

$$\sup_{\beta\in\Xi(s,A)} \mathbb{P}_{n,\theta,\beta}\Big(\Big|\sum_{e\in\mathcal{E}}\big(G_e-\mathbb{E}_{n,\theta,\boldsymbol{\beta}_0}\big(G_e\big|G_f:f\neq e\big)\big)\Big|\leq L_n\Big)\leq 2\exp\left\{-\frac{L_n^2}{\lambda\binom{n}{2}}\right\},$$

which converges to 0 as $L_n \gg n$. This completes the proof of the theorem. □

2.2. Proof of Theorem 1.2

As in the previous theorem, it suffices to show that both type I and type II errors converge to 0. For estimating the type I error, using a union bound gives

$$\begin{aligned}&\mathbb{P}_{n,\theta,\boldsymbol{\beta}_0}\Big(\max_{1\leq i\leq n}\Big|\sum_{e\ni i}(G_e-\mathbb{E}_{n,\theta,\boldsymbol{\beta}_0}\big(G_e\big|G_f:f\neq e\big)\Big|>C\sqrt{n\log n}\Big)\\&\leq\sum_{i=1}^{n}\mathbb{P}_{n,\theta,\boldsymbol{\beta}_0}\Big(\Big|\sum_{e\ni i}(G_e-\mathbb{E}_{n,\theta,\boldsymbol{\beta}_0}\big(G_e\big|G_f:f\neq e\big)\Big|>C\sqrt{n\log n}\Big)\leq 2n\exp\left\{-\frac{C^2 n\log n}{\lambda(n-1)}\right\},\end{aligned}\tag{16}$$

where the last inequality uses Lemma 2.1 with $c_e = 1$ if $e \ni i$, and 0 otherwise. For the choice $C > \sqrt{\lambda}$ the RHS above converges to 0, and so Type I error converges to 0.

For estimating the Type II error, fix vertex i such that $\beta_i \geq A$. Then using (15) gives

$$\sum_{e\ni i}\Big(\mathbb{E}_{n,\theta,\boldsymbol{\beta}}\big(G_e\big|G_f:f\neq e\big)-\mathbb{E}_{n,\theta,\boldsymbol{\beta}_0}\big(G_e\big|G_f:f\neq e\big)\Big)\geq\delta n\min\{A,1\}.$$

Since $A \geq \kappa\sqrt{\frac{\log n}{n}}$, for all n large we have

$$\delta n\min\{A,1\}\geq\delta\kappa\sqrt{n\log n}\geq 2C\sqrt{\log n}$$

for the choice $\kappa = \frac{2C}{\delta}$. This gives

$$\begin{aligned}&\mathbb{P}_{n,\theta,\boldsymbol{\beta}}\Big(\sum_{e\ni i}(G_e-\mathbb{E}_{n,\theta,\beta_0\mathbf{1}}\big(G_e\big|G_f:f\neq e\big)\leq C\sqrt{n\log n}\Big)\\&\leq\mathbb{P}_{n,\theta,\boldsymbol{\beta}}\Big(\Big|\sum_{e\ni i}(G_e-\mathbb{E}_{n,\theta,\boldsymbol{\beta}}\big(G_e\big|G_f:f\neq e\big)\Big|\geq C\sqrt{n\log n}\Big)\leq 2\exp\left\{-\frac{C^2 n\log n}{\lambda(n-1)}\right\},\end{aligned}$$

where the last inequality again uses Lemma 2.1. Consequently,

$$\sup_{\beta\in\Xi(s,A)}\mathbb{P}_{n,\theta,\beta}\Big(\max_{1\leq i\leq n}\Big|\sum_{e\ni i}\big(G_e-\mathbb{E}_{n,\theta,\boldsymbol{\beta}_0}\big(G_e\big|G_f:f\neq e\big)\big)\Big|\leq C\sqrt{n\log n}\Big)\leq 2n\exp\left\{-\frac{C^2 n\log n}{\lambda(n-1)}\right\},$$

which converges to 0 as before for the choice $C > \sqrt{\lambda}$. □

2.3. Proof of parts (b) and (d) of Theorem 1.3 and Theorem 1.4

Part (b) follows by a direct application of Theorem 1.1, on noting that $sA = n^{b+t} \to \infty$ if $b+t>0$. Similarly, part (d) follows by a direct application of Theorem 1.2, on noting that $A = n^t \gg \sqrt{\frac{\log n}{n}}$ if $t > -\frac{1}{2}$. Both Theorem 1.1 and Theorem 1.2 were proved for $\{0,1\}$ valued random variables, but essentially the same proof goes through for $\{-1,1\}$ valued random variables. □

2.4. Proof of Theorem 1.5 part (b)

To prove Theorem 1.5 part (b) (as well as parts (a) and (c) of Theorem 1.4 later), we express the two star model as a mixture of β models by introducing auxiliary variables, as done in Mukherjee and Xu (2023), Park and Newman (2004). Suppose Y be a random graph from the degree corrected two star model (12). Conditional on Y, let $(\phi_1, \cdots, \phi_n)$ be mutually independent components, with

$$\phi_i \sim N\Big(\frac{k_i}{n-1}, \frac{1}{\theta(n-1)}\Big). \tag{17}$$

The joint distribution of (ϕ, Y) is computed in the following Proposition. The proof of this is deferred to the supplementary file.

Proposition 2.1.

(a) *Given ϕ, the random variables $(Y)_{1\le i<j\le n}$ are mutually independent, with*

$$\mathbb{P}_{n,\theta,\boldsymbol{\beta}}(Y_{ij} = 1|\phi) = \frac{e^{\theta(\phi_i+\phi_j)+\frac{1}{2}(\beta_i+\beta_j)}}{e^{\theta(\phi_i+\phi_j)+\frac{1}{2}(\beta_i+\beta_j)} + e^{-\theta(\phi_i+\phi_j)-\frac{1}{2}(\beta_i+\beta_j)}}.$$

(b) *The marginal density of ϕ (w.r.t. Lebesgue measure) is proportional to*

$$f_{n,\theta,\boldsymbol{\beta}}(\phi) := \exp\Big\{-\sum_{i<j} p_{ij}(\phi_i, \phi_j)\Big\}, \tag{18}$$

where $p_{ij}(x,y)$ equals

$$\begin{aligned}&\frac{\theta}{2}(x^2+y^2) - \log\cosh\big[\theta(x+y) + \frac{1}{2}(\beta_i+\beta_j)\big] \\ =&\frac{\theta}{4}(x-y)^2 + q\Big(\frac{x+y}{2}\Big) + \log\cosh\Big(\theta(x+y)\Big) - \log\cosh\Big(\theta(x+y) + \frac{1}{2}(\beta_i+\beta_j)\Big),\end{aligned} \tag{19}$$

with

$$q(x) := \theta x^2 - \log\cosh(2\theta x). \tag{20}$$

We first state the following lemma about the function $q(.)$ introduced in (20) above, the proof of which follows from straightforward calculus (see for e.g. Dembo and Montanari (2010)).

Lemma 2.2. *If $\theta > 1/2$, the equation $q'(x) = 2\theta[x - \theta\tanh(2\theta x)]$ has a unique positive root t (which depends on θ)) on $(0,\infty)$. Further, t is the unique global minimizer of $q(.)$ on $[0,\infty)$.*

We will use the notation t introduced in the above lemma throughout the rest of the paper. We now state the following lemma, which contains the analogues of (Mukherjee and Xu, 2023, Lemma 4.1(a)), and (Mukherjee and Xu, 2023, Lemma 3.3(c)). The proof of this lemma is deferred to the supplementary file.

Lemma 2.3. *Suppose $\boldsymbol{\beta} \in [0, 2n^{-1/2}]^n$, and ℓ is a positive integer. Then there exists a positive finite constant C depending only on ℓ, θ such that the following happens:*

(a) *If $\theta = 1/2$, we have*

$$\max_{1 \le i \le n} \mathbb{E}_{n,\theta,\boldsymbol{\beta}} |\phi_i - \bar{\phi}|^{\ell} \le C n^{-\ell/2}.$$

(b) *If $\theta > \frac{1}{2}$, setting*

$$U := \cap_{i=1}^n V_i, \quad V_i := \{Y : k_i \ge (n-1)t/2\},$$

we have

$$\max_{1 \le i \le n} \mathbb{E}_{n,\theta,\boldsymbol{\beta}} \left(|\phi_i - t|^{\ell} \middle| U \right) \le C n^{-\ell/2}.$$

Proof of Theorem 1.5 part (b). We begin by claiming the existence of a sequence of positive reals $K_n \to \infty$ such that

$$\lim_{n \to \infty} \sup_{\boldsymbol{\beta} \in \Xi(s,A)} \mathbb{P}_{n,\theta,\boldsymbol{\beta}}(\tanh(\bar{\phi}) \le n^{-1/4} K_n) = 0. \tag{21}$$

Given (21), we first finish the proof of the theorem. Using part (a) of Proposition 2.1 we get

$$\begin{aligned}\sum_{i<j} \left[Y_{ij} - \tanh(\bar{\phi}) \right] &= \sum_{i<j} \left[\tanh\left(\frac{\phi_i + \phi_j}{2} + \frac{\beta_i + \beta_j}{2} \right) - \tanh(\bar{\phi}) \right] + O_P(n) \\ &\ge \sum_{i<j} \left[\tanh\left(\frac{\phi_i + \phi_j}{2} \right) - \tanh(\bar{\phi}) \right] + O_P(n) \\ &\gtrsim - \sum_{i<j} \left(\frac{\phi_i + \phi_j}{2} - \bar{\phi} \right)^2 \gtrsim -n \sum_{i=1}^n (\phi_i - \bar{\phi})^2 + O_P(n) = O_P(n),\end{aligned}$$

where the last equality uses Lemma 2.3 part (a). Using (21) along with the above display gives

$$\lim_{n \to \infty} \sup_{\boldsymbol{\beta} \in \Xi(s,A)} \mathbb{P}_{n,\theta,\boldsymbol{\beta}} \left(\sum_{i<j} Y_{ij} \le n^{3/2} K_n \right) = 0,$$

and so Type II error converges to 0. Since

$$\mathbb{P}_{n,\theta,\boldsymbol{\beta}_0} \left(\sum_{i<j} Y_{ij} > n^{\frac{3}{2}} K_n^{\frac{1}{2}} \right) \to 0,$$

using (Mukherjee and Xu, 2023, Theorem 1.1), Type I error converges to 0 as well. This shows that the test which rejects for large values of $\sum_{i<j} Y_{ij}$ is asymptotically powerful.

It thus remains to verify (21). To this end, assume without loss of generality that

$$\beta_i = \begin{cases} A & \text{if } 1 \le i \le s, \\ 0 & \text{if } s+1 \le i \le n, \end{cases} \tag{22}$$

where $A = n^t$. Also if $b + t + 1/2 > 0$, replacing t by $t' := \min(t, -1/2)$ we have

$$b + t' + 1/2 = \min\left(b + t + 1/2, b - \frac{1}{2} + \frac{1}{2}\right) = \min(b + t + 1/2, b) > 0.$$

Since the distribution of $\bar{\phi}$ is stochastically increasing in A, without loss of generality by replacing t by t' if necessary we can assume $t \le -\frac{1}{2}$, which gives $A \le n^{-1/2}$. Using Taylor's series expansion twice, we have

$$\begin{aligned} &\log\cosh\left(\frac{\phi_i + \phi_j}{2} + \frac{\beta_i + \beta_j}{2}\right) - \log\cosh\left(\frac{\phi_i + \phi_j}{2}\right) \\ =&\frac{\beta_i + \beta_j}{2}\tanh\left(\frac{\phi_i + \phi_j}{2}\right) + O(\beta_i + \beta_j)^2 \\ =&\frac{\beta_i + \beta_j}{2}\tanh(\bar{\phi}) + O\Big((\beta_i + \beta_j)|\phi_i + \phi_j - 2\bar{\phi}|\Big) + O(\beta_i + \beta_j)^2. \end{aligned}$$

Summing over $i < j$ and using (18) and (19) we get

$$\begin{aligned} -\log f_{n,\theta,\boldsymbol{\beta}}(\phi) =& -\log f_{n,\theta,\mathbf{0}}(\phi) - \frac{(n-1)sA}{2}\tanh(\bar{\phi}) \\ &+ O\left(nA\sum_{i=1}^{s}|\phi_i - \bar{\phi}| + sA\sum_{i=1}^{n}|\phi_i - \bar{\phi}| + nsA^2\right), \end{aligned} \tag{23}$$

where

$$\begin{aligned} -\log f_{n,\theta,\mathbf{0}}(\phi) :=& \sum_{i<j}\left[\frac{1}{8}(\phi_i - \phi_j)^2 + q\Big(\frac{\phi_i + \phi_j}{2}\Big)\right] \\ =& \frac{n}{8}\sum_{i=1}^{n}(\phi_i - \bar{\phi})^2 + \sum_{i<j} q\Big(\frac{\phi_i + \phi_j}{2}\Big), \end{aligned} \tag{24}$$

with $q(.)$ as in (20). As the notation above suggests, $f_{n,\theta,\mathbf{0}}$ defined above is the (unnormalized) density of ϕ under $\mathcal{H}_0$. Using (23), along with Lemma 2.3 part (a) we have

$$-\log f_{n,\theta,\boldsymbol{\beta}}(\phi) = -\log f_{n,\theta,\mathbf{0}}(\phi) - \frac{nsA}{2}\tanh(\bar{\phi}) - R_n,$$

where

$$\mathbb{E}_{n,\theta,\boldsymbol{\beta}}|R_n| \lesssim \sqrt{n}sA + nsA^2 \lesssim \sqrt{n}sA$$

using $A \le n^{-1/2}$. Thus, for any $K > 2$ fixed and $K_n' := n^{3/4}sA$ we have

$$\begin{aligned} &\mathbb{P}_{n,\theta,\boldsymbol{\beta}}(\tanh(\bar{\phi}) < Kn^{-1/4}) \\ \le&\mathbb{P}_{n,\theta,\boldsymbol{\beta}}(|R_n| > K_n') + \mathbb{P}_{n,\theta,\boldsymbol{\beta}}(\tanh(\bar{\phi}) < Kn^{-1/4}, |R_n| \le K_n') \end{aligned}$$

$$\leq \mathbb{P}_{n,\theta,\boldsymbol{\beta}}(|R_n| > K_n') + e^{K_n'} \frac{\mathbb{E}_{n,\theta,\boldsymbol{\beta}_0} \exp\left[\frac{nsA}{2}\tanh(\bar{\phi})\right] 1\left\{\tanh(\bar{\phi}) < Kn^{-1/4}\right\}}{\mathbb{E}_{n,\theta,\boldsymbol{\beta}_0} \exp\left[\frac{nsA}{2}\tanh(\bar{\phi})\right] 1\left\{|R_n| \leq K_n'\right\}}$$

$$\leq \mathbb{P}_{n,\theta,\boldsymbol{\beta}}(|R_n| > K_n') + \frac{e^{K_n' + \frac{Kn^{3/4}sA}{2} - \frac{nsA\tanh(2Kn^{-1/4})}{2}}}{\mathbb{P}_{n,\theta,\boldsymbol{\beta}_0}\left(\bar{\phi} > 2Kn^{-1/4}, |R_n| \leq K_n'\right)}.$$

On letting $n \to \infty$ and noting that $K_n' = n^{3/4}sA \gg \sqrt{n}sA$ we have

$$\lim_{n\to\infty} \sup_{\boldsymbol{\beta}\in\Xi(s,A)} \mathbb{P}_{n,\theta,\boldsymbol{\beta}}(|R_n| > K_n') = 0, \text{ and } \lim_{n\to\infty} \mathbb{P}_{n,\theta,\boldsymbol{\beta}_0}\left(\bar{\phi} > 2Kn^{-1/4}, |R_n| \leq K_n'\right) = \mathbb{P}(\zeta > 2K) > 0,$$

where ζ has density proportional to $e^{-\zeta^4/12 - \zeta^2/24}$ (c.f. (Mukherjee and Xu, 2023, Lemma 4.2)), and the convergence of the second term uses $K > 2$. Combining the last two displays we have

$$\lim_{n\to\infty} \sup_{\boldsymbol{\beta}\in\Xi(s,A)} \mathbb{P}_{n,\theta,\boldsymbol{\beta}}(\tanh(\bar{\phi}) < Kn^{-1/4}) = 0.$$

Since this holds for every fixed $K > 2$, there exists $K_n \to \infty$ such that

$$\limsup_{n\to\infty} \sup_{\boldsymbol{\beta}\in\Xi(s,A)} \mathbb{P}_{n,\theta,\boldsymbol{\beta}}(\tanh(\bar{\phi}) < K_n n^{-1/4}) = 0.$$

This verifies (21), and hence completes the proof of the theorem. □

3. Proof of parts (a) and (c) Theorems 1.3 and 1.4

With $\Xi(s, A)$ as defined in (6), consider the following subset of $\Xi(s, A)$.

$$\tilde{\Xi}(s, A) := \left\{\boldsymbol{\beta} = \beta_0 \mathbf{1} + \boldsymbol{\mu} : |\text{supp}(\boldsymbol{\mu})| = s, and\ \mu_i = A, i \in \text{supp}(\boldsymbol{\mu})\right\}. \tag{25}$$

Let $\pi(d\boldsymbol{\beta})$ be a prior on $\Xi(s, A)$, which put probability mass $1/\binom{n}{s}$ on each of configurations in $\tilde{\Xi}(s, A)$. And let $\mathbb{Q}_\pi(.) := \int \mathbb{P}_{n,\theta,\boldsymbol{\beta}}(.)\pi(d\boldsymbol{\beta})$ denote the marginal distribution of Y under this prior. To show that all tests for the problem (5) are asymptotically powerless, using the second moment method it suffices to show that

$$\lim_{n\to\infty} \mathbb{E}_{n,\theta,\boldsymbol{\beta}_0} L_\pi(Y)^2 = 1, \text{ where } L_\pi(Y) := \frac{\mathbb{Q}_\pi(Y)}{\mathbb{P}_{n,\theta,\boldsymbol{\beta}_0}(Y)} \tag{26}$$

is the likelihood ratio. The following lemma gives an upper bound to the second moment of $L_\pi(.)$.

Lemma 3.1. *For any (θ, β_0), with $L_\pi(.)$ as defined in (26) we have*

$$\mathbb{E}_{\mathcal{H}_0} L_\pi^2(Y) \leq \exp\left\{A^2 s^2 Cov_{\boldsymbol{\beta}=(\beta_0/2)\mathbf{1}}(k_1, k_2) + \frac{2s^2}{n}(e^{A^2 Var_{\boldsymbol{\beta}=(\beta_0/2)\mathbf{1}}(k_1)} - 1)\right\}, \tag{27}$$

whenever $n > 2s$.

Proof. Define $\Lambda_s := \{S \big| S \subset \{1,2,...,n\}, |S| = s\}$. For any $S \in \Lambda_s$, define a vector $\boldsymbol{\beta}_S$ by setting

$$\beta_{S,i} = \begin{cases} \beta_0 + A & \text{if } i \in S, \\ \beta_0 & \text{if } i \notin S. \end{cases} \tag{28}$$

By symmetry, the normalizing constant $Z_n(\boldsymbol{\beta}_S, \theta)$ is the same for all $S \in \Lambda_s$, which we denote by $Z_n(\boldsymbol{\beta}_{[s]}, \theta)$ for the rest of this proof. Then, a direct calculation gives

$$\begin{aligned}
\mathbb{E}_{\mathcal{H}_0} L_\pi^2(Y) &= \frac{Z_n^2(\beta_0,\theta)}{Z_n^2(\boldsymbol{\beta}_{[s]},\theta)} \frac{1}{\binom{n}{s}^2} \mathbb{E}_{\mathcal{H}_0} \sum_{S_1,S_2 \in \Lambda_s} e^{\sum_{j\in S_1} \frac{A}{2} k_j + \sum_{j \in S_2} \frac{A}{2} k_j} \\
&= \frac{Z_n(\beta_0,\theta)}{Z_n^2(\boldsymbol{\beta}_{[s]},\theta)} \frac{1}{\binom{n}{s}^2} \sum_{S_1,S_2 \in \Lambda} \frac{Z_n(\boldsymbol{\beta}_{S_1} + \boldsymbol{\beta}_{S_2},\theta)}{Z_n(\boldsymbol{\beta}_{S_1} + \boldsymbol{\beta}_{S_2},\theta)} \sum_Y e^{\frac{\theta}{2n}\sum_{i=1}^n k_i^2 + \sum_{j=1}^n \frac{\beta_{S_1,j}+\beta_{S_2,j}}{2} k_j} \\
&= \frac{1}{\binom{n}{s}^2} \sum_{S_1,S_2 \in \Lambda} \frac{Z_n(\beta_0,\theta) Z_n(\boldsymbol{\beta}_{S_1} + \boldsymbol{\beta}_{S_2},\theta)}{Z_n(\boldsymbol{\beta}_{S_1},\theta) Z_n(\boldsymbol{\beta}_{S_2},\theta)} = \frac{1}{\binom{n}{s}^2} \sum_{S_1,S_2 \in \Lambda} e^{R_{S_1,S_2}},
\end{aligned} \tag{29}$$

where

$$\begin{aligned}
R_{S_1,S_2} :=& \log\left(\frac{Z_n(\beta_0,\theta) Z_n(\boldsymbol{\beta}_{S_1} + \boldsymbol{\beta}_{S_2},\theta)}{Z_n(\boldsymbol{\beta}_{S_1},\theta) Z_n(\boldsymbol{\beta}_{S_2},\theta)}\right) \\
=& \log Z_n(\boldsymbol{\beta}_{S_1} + \boldsymbol{\beta}_{S_2},\theta) - \log Z_n(\boldsymbol{\beta}_{S_2},\theta) - \log Z_n(\boldsymbol{\beta}_{S_1},\theta) + \log Z_n(\beta_0,\theta).
\end{aligned}$$

Setting $W = S_1 \bigcap S_2$, note that R_{S_1,S_2} only depends on $|W|$ by symmetry. Thus, without loss of generality we assume that $S_1 = \{1,2,3,...,s\}$ and $S_2 = \{1,2,...,w,s+1,s+2,...,2s-w\}$. Consequently,

$$R_{S_1,S_2} = \sum_{j\in S_1} \Big[\log Z_n(\boldsymbol{\beta}_{[j]} + \boldsymbol{\beta}_{S_2},\theta) - \log Z_n(\boldsymbol{\beta}_{[j-1]} + \boldsymbol{\beta}_{S_2},\theta) - \log Z_n(\boldsymbol{\beta}_{[j]},\theta) + \log Z_n(\boldsymbol{\beta}_{[j-1]},\theta) \Big],$$

where $\boldsymbol{\beta}_{[j]}$ denotes the vector $\boldsymbol{\beta}$ which equals A on first j entries, and β_0 for rest of its entries, The summand in the RHS above equals

$$\begin{aligned}
&\log Z_n(\boldsymbol{\beta}_{[j]} + \boldsymbol{\beta}_{S_2},\theta) - \log Z_n(\boldsymbol{\beta}_{[j-1]} + \boldsymbol{\beta}_{S_2},\theta) - \log Z_n(\boldsymbol{\beta}_{[j]},\theta) + \log Z_n(\boldsymbol{\beta}_{[j-1]},\theta) \\
&= \int_0^A \frac{\partial \log Z_n(\boldsymbol{\beta}_{[j-1]} + \boldsymbol{\beta}_{S_2} + \gamma \mathbf{e_j},\theta)}{\partial \beta_j} d\gamma - \int_0^A \frac{\partial \log Z_n(\boldsymbol{\beta}_{[j-1]} + \gamma \mathbf{e_j},\theta)}{\partial \beta_j} d\gamma \\
&= \int_0^A A \sum_{r\in S_2} \frac{\partial \log Z_n(\boldsymbol{\beta}_{[j-1]} + \boldsymbol{\xi} + \gamma \mathbf{e_j})}{\partial \beta_j \partial \beta_r} |_{\boldsymbol{\xi} \le \boldsymbol{\beta}_{S_2}} d\gamma \\
&= \int_0^A A \sum_{r\in S_2} Cov_{\boldsymbol{\beta} = \boldsymbol{\beta}_{[j-1]} + \boldsymbol{\xi} + \gamma \mathbf{e_j}}(k_j, k_r) d\gamma.
\end{aligned}$$

If $A \to 0$, then $\boldsymbol{\beta} \ge \mathbf{0}$ if $\beta_0 \ge 0$, and $\boldsymbol{\beta} \le \mathbf{0}$ for all n large if $\beta_0 < 0$. Since GHS inequality (Lebowitz, 1974) holds if either $\boldsymbol{\beta} \ge \mathbf{0}$ or $\boldsymbol{\beta} \le \mathbf{0}$ (the second conclusion follows on noting that $Cov_{\boldsymbol{\beta}}(k_r, k_s)$ is same as $= Cov_{\boldsymbol{\beta}}(-k_r,-k_s)$, thereby giving

$$Cov_{\boldsymbol{\beta}=\boldsymbol{\beta}_{[j-1]}+\boldsymbol{\xi}+\gamma\mathbf{e_j}}(k_j,k_r) \le Cov_{\boldsymbol{\beta}=(\beta_0/2)\mathbf{1}}(k_j,k_r).$$

Combining the above two displays, this gives

$$\begin{aligned} R_{S_1,S_2} &\le \sum_{j\in S_1}\int_0^A A\sum_{r\in S_2} Cov_{\boldsymbol{\beta}=(\beta_0/2)\mathbf{1}}(k_j,k_r)d\gamma \\ &= A^2 w Var_{\boldsymbol{\beta}=(\beta_0/2)\mathbf{1}}(k_1) + A^2(s^2-w)Cov_{\boldsymbol{\beta}=(\beta_0/2)\mathbf{1}}(k_1,k_2). \end{aligned}$$

Along with (29), this further gives

$$\mathbb{E}_{\mathcal{H}_0} L_\pi^2(Y) \le \exp\{A^2 s^2 Cov_{\boldsymbol{\beta}=(\beta_0/2)\mathbf{1}}(k_1,k_2)\}\mathbb{E}_W \exp\{A^2 Var_{\boldsymbol{\beta}=(\beta_0/2)\mathbf{1}}(k_1)W\},$$

where W follows Hypergeometric distribution with parameters (n,s,s). Since $2s<n$, we have that W is stochastically dominated by a binomial distribution with parameters $\left(s,\frac{s}{n-s}\right)$ ((Mukherjee, Mukherjee and Sen, 2018, Lemma 6.1)), which gives

$$\mathbb{E}_W \exp\{A^2 Var_{\boldsymbol{\beta}=(\beta_0/2)\mathbf{1}}(k_1)W\} \le \exp\left\{\frac{2s^2}{n}(e^{A^2 Var_{\boldsymbol{\beta}=(\beta_0/2)\mathbf{1}}(k_1)}-1)\right\}.$$

Combining the last two displays, we have verified (27). □

3.1. Proof of parts (a) and (c) of Theorem 1.3

With L_π as in defined in (26), it suffices to show that

$$\lim_{n\to\infty}\mathbb{E}_{\mathcal{H}_0} L_\pi^2(Y) = 1.$$

By (Mukherjee and Xu, 2023, Lemma 3.3(d)) along with (17) we have

$$Var_{\boldsymbol{\beta}=(\beta_0/2)\mathbf{1}}(\sum_{e\in\mathcal{E}} Y_e) \lesssim n^2,$$

which gives the existence of a constant c depending on θ such that

$$Var_{\boldsymbol{\beta}=(\beta_0/2)\mathbf{1}}(k_1)\le cn, \quad Cov_{\boldsymbol{\beta}=(\beta_0/2)\mathbf{1}}(k_1,k_2)\le c.$$

Using this along with Lemma 3.1 gives

$$\mathbb{E}_{\mathcal{H}_0} L_\pi^2(Y) \le \exp\left\{cA^2s^2 + \frac{2s^2}{n}(e^{cA^2n}-1)\right\}. \tag{30}$$

3.1.1. Proof of part (a)

In this regime we have $s=n^b$ and $A=n^t$ with $b\ge\frac{1}{2}$ and $b+t<0$. This gives

$$\max(A^2n, A^2s^2) = \max(n^{2t+1}, n^{2t+2b}) = n^{2t+2b}\to 0,$$

using which the exponent in the RHS of (30) converges to 0. This completes the proof of part (a).

3.1.2. Proof of part (c)

In this regime we have $s = n^b$ and $A = n^t$ with $b < \frac{1}{2}$ and $t \le -\frac{1}{2}$. This gives $A^2 s^2 = n^{2b+2t} \to 0$. Also

$$\frac{s^2}{n} e^{cA^2 n - 1} \le e^{c-1} \frac{s^2}{n} = e^{c-1} n^{2b-1} \to 0.$$

Consequently, the RHS of (30) again converges to 0. This completes the proof of part (c). □

3.2. Proof of Theorem 1.5 part (a)

As before, with L_π defined in (26), it is sufficient to show that

$$\lim_{n\to\infty} \mathbb{E}_{\mathcal{H}_0} L_\pi^2(Y) = 1.$$

To this effect, using (Mukherjee and Xu, 2023, Lemma 4.1(b)) along with (17) we get

$$Var_{\mathcal{H}_0}(\sum_{e\in\mathcal{E}} Y_e) \lesssim n^3.$$

Along with the non-negativity of covariance, this gives

$$Cov_{\mathcal{H}_0}(k_1, k_2) \lesssim n.$$

For getting the optimal bound on $Var_{\mathcal{H}_0}(k_1)$, use (17) to get

$$Var_{\mathcal{H}_0}(k_1) \lesssim n^2 Var_{\mathcal{H}_0}(\phi_1) + n \lesssim n^2 \Big[Var_{\mathcal{H}_0}(\bar{\phi}) + Var_{\mathcal{H}_0}(\phi_1 - \bar{\phi}) \Big] + n \lesssim n,$$

where the last inequality uses (Mukherjee and Xu, 2023, Lemma 4.1). Combining the two displays above along with Lemma 3.1 gives the existence of a constant c free of n, such that

$$\mathbb{E}_{\mathcal{H}_0} L_\pi^2(Y) \le \exp\left\{ cA^2 s^2 n + \frac{2s^2}{n}(e^{cA^2 n} - 1) \right\}. \tag{31}$$

Now, recall that in this regime we have $s = n^b$ and $A = n^t$ with $b + t + \frac{1}{2} < 0$. This allows us to conclude that $A^2 s^2 n = n^{2b+2t+1} \to 0$. Also, noting that $2t + 1 < 0$ we have

$$\frac{s^2}{n}(e^{cA^2 n} - 1) \le n^{2b-1}(e^{cn^{2t+1}} - 1) \lesssim n^{2b+2t+1} \to 0.$$

Along with (31), this gives $\lim_{n\to\infty} \mathbb{E}_{\mathcal{H}_0} L_\pi^2(Y) = 1$. This completes the proof of part (a). □

3.3. Proof of Theorem 1.4 parts (a) and (c)

Restricting the probability measure (12) to the set U (as defined in Lemma 2.3 part (b)), define the probability measure $\mathbb{P}_{n,\boldsymbol{\beta},U}(\cdot)$ by setting

$$\mathbb{P}_{n,\boldsymbol{\beta},U}(Y) = \frac{1}{Z_n^+(\boldsymbol{\beta},\theta)} \exp\left\{ \frac{\theta}{2n} \sum_{i=1}^n k_i^2 + \frac{1}{2} \sum_{i=1}^n \beta_i k_i \right\} 1\{Y \in U\}, \tag{32}$$

where

$$Z_n^+(\boldsymbol{\beta},\theta)=\sum_{Y\in U}\exp\left\{\frac{\theta}{2n}\sum_{i=1}^n k_i^2+\frac{1}{2}\sum_{i=1}^n \beta_i k_i\right\}$$

is the restricted normalizing constant. As before, consider the sub parameter space $\tilde{\Xi}(s,A)$ defined in (25), and let $\pi(d\boldsymbol{\beta})$ be a prior on $\tilde{\Xi}(s,A)$, which puts probability $1/\binom{n}{s}$ on each element in $\tilde{\Xi}(s,A)$. And let $\mathbb{Q}_{\pi,U}(.):=\int \mathbb{P}_{n,\boldsymbol{\beta},U}(.)\pi(d\boldsymbol{\beta})$ denote the mixed alternative distribution of Y. Since (Mukherjee and Xu, 2023, Proposition 3.2) gives $\mathbb{P}_{n,\theta,\boldsymbol{\beta}_0}(U)\to 1/2$, for verifying the absence of asymptotically powerful tests setting

$$L_{\pi,U}(Y):=\frac{\mathbb{Q}_{\pi,U}(Y)}{\mathbb{P}_{\mathcal{H}_0,U}(Y)}, \tag{33}$$

it suffices to show:

$$\mathbb{E}_{\mathcal{H}_0,U}L^2_{\pi,U}(Y)\to 1. \tag{34}$$

Proceeding similar to Lemma 3.1, we get

$$\mathbb{E}_{\mathcal{H}_0,U}L^2_{\pi,U}(Y)=\frac{1}{\binom{n}{s}^2}\sum_{S_1,S_2\in\Lambda}\frac{Z_n^+(0,\theta)Z_n^+(\boldsymbol{\beta}_{S_1}+\boldsymbol{\beta}_{S_2},\theta)}{Z_n^+(\boldsymbol{\beta}_{S_1},\theta)Z_n^+(\boldsymbol{\beta}_{S_2},\theta)}. \tag{35}$$

Setting $R^+_{S_1,S_2}$ as

$$R^+_{S_1,S_2}:=\left(\log Z_n^+(\boldsymbol{\beta}_{S_1}+\boldsymbol{\beta}_{S_2},\theta\log Z_n^+(\boldsymbol{\beta}_{S_2},\theta))-(\log Z_n^+(\boldsymbol{\beta}_{S_1},\theta)-\log Z_n^+(0,\theta)\right),$$

a Taylor's series expansion gives

$$R_{S_1,S_2}=A^2\sum_{i\in S_1}\sum_{j\in S_2}Cov_{\boldsymbol{\delta}=\alpha\mathbf{1}_{S_1}+\gamma\mathbf{1}_{S_2}}(k_i,k_j|U), \tag{36}$$

where $\alpha,\gamma\in(0,A)$ and $\mathbf{1}_S$ denote vector having unit signals at S, and $\boldsymbol{\delta}:=\alpha\mathbf{1}_{S_1}+\gamma\mathbf{1}_{S_2}\in[0,2n^{-1/2}]^n$. We now claim that

Lemma 3.2.

$$\max_{1\le i<j\le n}\sup_{\boldsymbol{\beta}\in[0,2n^{-1/2}]^n}Cov_{\boldsymbol{\beta}}(k_i,k_j|U)\lesssim 1.$$

We defer the proof of Lemma 3.2 to the end of the section. Finally, use Lemma 2.3 part (b) to conclude that

$$\max_{1\le i\le n}Var_{\boldsymbol{\delta}}(k_i|U)\lesssim n. \tag{37}$$

Given Lemma 3.2 along with (37) and (36), we have the existence of a constant C free of n such that

$$R_{S_1,S_2}\le CWA^2n+Cs^2A^2,$$

which along with (35) gives

$$\mathbb{E}_{\mathcal{H}_0,U} L^2_{\pi,U}(Y) \le \exp\{CA^2s^2\}\mathbb{E}_W \exp\{CA^2nW\}, \tag{38}$$

where W follows Hypergeometric distribution with parameters (n,s,s). As before, using the fact that $n > 2s$, W is stochastically dominated by a binomial distribution with parameters $(s, \frac{s}{n-s})$. This gives

$$\mathbb{E}_{\mathcal{H}_0,U} L^2_{\pi,U}(Y) \le \exp\{CA^2s^2 + \frac{2s^2}{n}(e^{CA^2n} - 1)\}. \tag{39}$$

3.3.1. Proof of Theorem 1.4 part (a)

In this regime we have $s = n^b$ and $A = n^t$ with $b \ge \frac{1}{2}$ and $b + t < 0$. This gives $A^2s^2 = n^{2t+2b} \to 0$. Also we have $A^2n = n^{2t+1} \to 0$, and so

$$\frac{s^2}{n}(e^{CA^2n} - 1) \lesssim s^2A^2 = n^{2b+2t} \to 0.$$

Combining the above two displays with (39), we have $\mathbb{E}_{\mathcal{H}_0,U} L^2_{\pi,U}(Y) \to 1$, as desired. The proof of part (a) is complete.

3.3.2. Proof of Theorem 1.4 part (c)

In this regime we have $s = n^b$ and $A = n^t$ with $b < \frac{1}{2}$ and $t + \frac{1}{2} < 0$. This gives

$$A^2s^2 = n^{2t+2b} \le n^{2t+1} \to 0.$$

Also we have $A^2n = n^{2t+1} \to 0$, and so

$$\frac{s^2}{n}(e^{CA^2n} - 1) \lesssim s^2A^2 = n^{2b+2t} \to 0.$$

Combining the above two displays with (39), we have $\mathbb{E}_{\mathcal{H}_0,U} L^2_{\pi,U}(Y) \to 1$, as desired. The proof of part (c) is complete. □

3.4. Proof of Lemma 3.2

Analogous to the definition of U in Lemma 2.3 part (b), define

$$\widetilde{U} := \cap_{i=1}^n \widetilde{V}_i \quad \widetilde{V}_i := \Big\{\phi_i \in \Big[0,2\Big]\Big\}.$$

The following lemma, to be used in the proof of Lemma 3.2, shows that with high probability the sets U and $\widetilde{U}$ occur simultaneously, and so expectations involving U can be transferred to expectations involving $\widetilde{U}$ at a very low cost. This lemma will be used frequently in the rest of this section, sometimes without an explicit mention.

Lemma 3.3. *Suppose $\theta > 1/2$, and $\boldsymbol{\beta} \in [0, 2n^{-1/2}]$. Then we have the following conclusions:*

(a) $\log \mathbb{P}_{n,\theta,\boldsymbol{\beta}}(U\Delta\widetilde{U}) \lesssim -n$.

(b) *For any random variable W such that $\mathbb{E}W^2 \le 1$, we have*

$$\left|\mathbb{E}W1\{U\} - \mathbb{E}W1\{\widetilde{U}\}\right| \le \sqrt{\mathbb{P}_{n,\theta,\boldsymbol{\beta}}(U\Delta\widetilde{U})}.$$

The proof of Lemma 3.3 is deferred to the supplementary file. We now prove a correlation bound for higher order terms, which will be used for proving Lemma 3.2.

Lemma 3.4. *Suppose $\theta > 1/2$, and $\boldsymbol{\beta} \in [0, 2n^{-1/2}]$. Then for any pair of indices $\{i_1, i_2, i_3\}$, which are not necessarily distinct, we have*

$$Cov_{n,\theta,\boldsymbol{\beta}}\Big((\phi_{i_1} - t)(\phi_{i_2} - t), \phi_{i_3} - t|\widetilde{U}\Big) \lesssim n^{-2}.$$

Proof. Setting $M(i_1,i_2,i_3) := Cov_{n,\theta,\boldsymbol{\beta}}\Big((\phi_{i_1} - t)(\phi_{i_2} - t), \phi_{i_3} - t|\widetilde{U}\Big)$, we claim that

$$\max_{1\le i_1,i_2\le n}\left|M(i_1,i_2,i_3) - \frac{\theta^3 \text{sech}^6(2\theta t)}{(n-1)^3}\sum_{j_1\ne i_1, j_2\ne i_2, j_3\ne i_3}\ \sum_{\ell_a\in(i_a,j_a),1\le a\le 3} M(\ell_1,\ell_2,\ell_3)\right| = O(n^{-2}). \quad (40)$$

We first complete the proof of the lemma, deferring the proof of (40). The above display implies the existence of a constant C free of n, such that

$$\max_{1\le i_1,i_2,i_3\le n}\left|M(i_1,i_2,i_3) - \sum_{1\le j_1,j_2,j_3\le n} B_n\Big((i_1,i_2,i_3),(j_1,j_2,j_3)\Big)\right| \le \frac{C}{n^2}, \quad (41)$$

where B_n is a symmetric $n^3 \times n^3$ matrix with non-negative entries, satisfying

$$\sum_{1\le j_1,j_2,j_3\le n} B_n\Big((i_1,i_2,i_3),(j_1,j_2,j_3)\Big) = 8\theta^3\text{sech}^6(2\theta t) = [2\theta\text{sech}^2(2\theta t)]^3 < 1.$$

where the last inequality uses (Mukherjee and Xu, 2023, Lemma 1.2). Thus the matrix $(\mathbf{I} - B_n)^{-1}$ has ℓ_∞ operator norm equal to $(1 - 8\theta^3\text{sech}^6(2\theta t))^{-1} < \infty$, and so (41) gives

$$\max_{1\le i_1,i_2,i_3\le n} |M(i_1,i_2,i_3)| \le C(1 - 8\theta^3\text{sech}^6(2\theta t))^{-1}n^{-2},$$

from which the desired conclusion follows.

It thus remains to verify (40). There are various possibilities depending on which of the indices $\{i,j,\ell\}$ are distinct. Below we argue the case $i_1 = i_2 = i$ and $i_3 = j$, with $\{i,j\}$ distinct, noting that the bound follows by similar calculations for other choices. To this end, setting $k_{i,t} := k_i - (n-1)t$ note that $(\phi_i - t|Y) \sim N\Big(\frac{k_{i,t}}{n-1}, \frac{1}{(n-1)\theta}\Big)$. Consequently, we have

$$\begin{aligned}
&\mathbb{P}_{n,\theta,\boldsymbol{\beta}}(\widetilde{U})Cov_{n,\theta,\boldsymbol{\beta}}\Big((\phi_i - t)^2, \phi_j - t|\widetilde{U}\Big)\\
=&\mathbb{E}_{n,\theta,\boldsymbol{\beta}}\Big[(\phi_i - t)^2(\phi_j - t)1\{\widetilde{U}\}\Big] - \mathbb{E}_{n,\theta,\boldsymbol{\beta}}\Big[(\phi_i - t)^2 1\{\widetilde{U}\}\Big]\mathbb{E}_{n,\theta,\boldsymbol{\beta}}\Big[(\phi_j - t)1\{\widetilde{U}\}\Big]\\
=&\mathbb{E}_{n,\theta,\boldsymbol{\beta}}\Big[(\phi_i - t)^2(\phi_j - t)1\{U\}\Big] - \mathbb{E}_{n,\theta,\boldsymbol{\beta}}\Big[(\phi_i - t)^2 1\{U\}\Big]\mathbb{E}_{n,\theta,\boldsymbol{\beta}}\Big[(\phi_j - t)1\{U\}\Big] + O(e^{-cn})
\end{aligned}$$

$$
\begin{aligned}
&= \mathbb{E}_{n,\theta,\boldsymbol{\beta}}\left[\Big(\frac{k_{i,t}^2}{(n-1)^2} + \frac{1}{(n-1)\theta}\Big)\frac{k_{i,t}}{n-1}1\{U\}\right] \\
&- \mathbb{E}_{n,\theta,\boldsymbol{\beta}}\left[\Big(\frac{k_{i,t}^2}{(n-1)^2} + \frac{1}{(n-1)\theta}\Big)1\{U\}\right]\mathbb{E}_{n,\theta,\boldsymbol{\beta}}\left[\frac{k_{i,t}}{n-1}1\{U\}\right] + O(e^{-cn}) \\
&= \mathbb{P}_{n,\theta,\boldsymbol{\beta}}(U)Cov_{n,\theta,\boldsymbol{\beta}}\Big(\frac{k_{i,t}^2}{(n-1)^2} + \frac{1}{(n-1)\theta}, \frac{k_{j,t}}{n-1}|U\Big) + O(e^{-cn}) \\
&= \frac{1}{(n-1)^3}\mathbb{P}_{n,\theta,\boldsymbol{\beta}}(U)Cov_{n,\theta,\boldsymbol{\beta}}\Big(k_{i,t}^2, k_{j,t}|U\Big) + O(e^{-cn}) \\
&= \frac{1}{(n-1)^3}\mathbb{P}_{n,\theta,\boldsymbol{\beta}}(U)\sum_{a_1,a_2\neq i, b\neq j} Cov\Big(Y_{ia_1,t}Y_{ia_2,t}, Y_{jb,t}|U\Big) + O(e^{-cn}),
\end{aligned} \tag{42}
$$

where $Y_{ij,t} := Y_{ij} - t$, and the change from U to $\widetilde{U}$ uses Lemma 3.3 and incurs the cost $O(e^{-cn})$. Proceeding to estimate the RHS of (42), set $r_{ij,t} := \mathbb{E}(Y_{ij}|\phi)$, and for $a_1 \neq a_2$ note that

$$
\begin{aligned}
&\mathbb{P}_{n,\theta,\boldsymbol{\beta}}(U)Cov_{n,\theta,\boldsymbol{\beta}}\Big(Y_{ia_1,t}Y_{ia_2,t}, Y_{jb,t}|U\Big) \\
&= \mathbb{E}_{n,\theta,\boldsymbol{\beta}}(Y_{ia_1,t}Y_{ia_2,t}Y_{jb,t}1\{U\}) - \mathbb{E}_{n,\theta,\boldsymbol{\beta}}(Y_{ia_1,t}Y_{ia_2,t}1\{U\})\mathbb{E}_{n,\theta,\boldsymbol{\beta}}(Y_{jb}1\{U\}) \\
&= \mathbb{E}_{n,\theta,\boldsymbol{\beta}}\Big[Y_{ia_1,t}Y_{ia_2,t}Y_{jb,t}1\{\widetilde{U}\}\Big] - \mathbb{E}_{n,\theta,\boldsymbol{\beta}}\Big[Y_{ia_1,t}Y_{ia_2,t}1\{\widetilde{U}\}\Big]\mathbb{E}_{n,\theta,\boldsymbol{\beta}}\Big[Y_{jb,t}1\{\widetilde{U}\}\Big] + O(e^{-cn}) \\
&= \mathbb{E}_{n,\theta,\boldsymbol{\beta}}\Big[r_{ia_1,t}r_{ia_2,t}r_{jb,t}1\{\widetilde{U}\}\Big] - \mathbb{E}_{n,\theta,\boldsymbol{\beta}}\Big[r_{ia_1,t}r_{ia_2,t}1\{\widetilde{U}\}\Big]\mathbb{E}_{n,\theta,\boldsymbol{\beta}}\Big[r_{jb}1\{\widetilde{U}\}\Big] + O(e^{-cn}) \\
&= \mathbb{P}_{n,\theta,\boldsymbol{\beta}}(\widetilde{U})Cov_{n,\theta,\boldsymbol{\beta}}(r_{ia_1,t}r_{ia_2,t}, r_{jb,t}|\widetilde{U}) + O(e^{-cn}).
\end{aligned} \tag{43}
$$

In the above display, we have again moved from U to $\widetilde{U}$ at a cost $O(e^{-cn})$, using Lemma 3.3. A one term Taylor's series expansion gives

$$
\begin{aligned}
r_{ij,t} &= \tanh\Big[\theta(\phi_i + \phi_i) + (\beta_i + \beta_j)\Big] - \tanh(2\theta t) \\
&= \Big[\theta(\phi_i - t + \phi_j - t) + \frac{1}{2}(\beta_i + \beta_j)\Big]\operatorname{sech}^2(2\theta t) + \xi_{ij},
\end{aligned} \tag{44}
$$

where

$$
|\xi_{ij}| \lesssim (\phi_i - t)^2 + (\phi_j - t)^2 + \beta_i^2 + \beta_j^2 \lesssim (\phi_i - t)^2 + (\phi_j - t)^2 + n^{-1}.
$$

On taking expectations, $Cov_{n,\theta,\boldsymbol{\beta}}\Big(r_{ia_1,t}r_{ia_2,t}, r_{jb,t}|\widetilde{U}\Big)$ equals

$$
\theta^3\operatorname{sech}^6(2\theta t)\sum_{u\in\{i,a_1\}, v\in\{i,a_2\}, w\in\{j,b\}} Cov_{n,\theta,\boldsymbol{\beta}}\Big((\phi_u - t)(\phi_v - t), (\phi_w - t)|\widetilde{U}\Big) + O(n^{-2}), \tag{45}
$$

where we have used Lemma 2.3 part (b). On the other hand, if $a_1 = a_2 = a$, then using the fact that

$$
Y_{ia,t}^2 = 1 + t^2 - 2tY_{ia} = 1 - t^2 - 2tY_{ia,t},
$$

we have

$$
\begin{aligned}
&-\frac{1}{2t}\mathbb{P}_{n,\theta,\boldsymbol{\beta}}(U)Cov_{n,\theta,\boldsymbol{\beta}}\left(Y_{ia,t}^2,Y_{jb,t}|U\right)\\
=&\mathbb{P}_{n,\theta,\boldsymbol{\beta}}(U)Cov_{n,\theta,\boldsymbol{\beta}}\left(Y_{ia,t},Y_{jb,t}|U\right)\\
=&\mathbb{E}_{n,\theta,\boldsymbol{\beta}}(Y_{ia,t}Y_{jb,t}1\{U\})-\mathbb{E}_{n,\theta,\boldsymbol{\beta}}(Y_{ia,t}1\{U\})\mathbb{E}_{n,\theta,\boldsymbol{\beta}}(Y_{jb}1\{U\})\\
=&\mathbb{E}_{n,\theta,\boldsymbol{\beta}}\left[Y_{ia,t}Y_{jb,t}1\{\widetilde{U}\}\right]-\mathbb{E}_{n,\theta,\boldsymbol{\beta}}\left[Y_{ia,t}1\{\widetilde{U}\}\right]\mathbb{E}_{n,\theta,\boldsymbol{\beta}}\left[Y_{jb,t}1\{\widetilde{U}\}\right]+O(e^{-cn})\\
=&\mathbb{E}_{n,\theta,\boldsymbol{\beta}}\left[r_{ia}r_{jb}1\{\widetilde{U}\}\right]-\mathbb{E}_{n,\theta,\boldsymbol{\beta}}\left[r_{ia}1\{\widetilde{U}\}\right]\mathbb{E}_{n,\theta,\boldsymbol{\beta}}\left[r_{jb}1\{\widetilde{U}\}\right]+O(e^{-cn})\\
=&\mathbb{P}_{n,\theta,\boldsymbol{\beta}}(\widetilde{U})Cov_{n,\theta,\boldsymbol{\beta}}(r_{ia},r_{jb}|\widetilde{U})+O(e^{-cn})=O(n^{-1}),
\end{aligned}
\tag{46}
$$

where the last equality again uses Lemma 2.3 part (b), along with (44). Combining (42), (43), (45) and (46) we have

$$
\begin{aligned}
&Cov_{n,\theta,\boldsymbol{\beta}}\left((\phi_i-t)^2,\phi_j-t|\widetilde{U}\right)\\
=&\frac{\theta^3\mathrm{sech}^6(2\theta t)}{(n-1)^3}\sum_{a_1,a_2\neq i,b\neq j}\ \sum_{u\in\{i,a_1\},v\in\{i,a_2\},w\in\{j,b\}}Cov_{n,\theta,\boldsymbol{\beta}}\left((\phi_u-t)(\phi_v-t),(\phi_w-t)|\widetilde{U}\right)+O(n^{-2}),
\end{aligned}
$$

which verifies (40) for the choice $\{i_1=i_2=i,i_3=j\}$. This completes the proof of the claim. □

Proof of Lemma 3.2. We proceed via a similar argument as in the proof of Lemma 3.4. Setting $M(i_1,i_2):=Cov_{n,\theta,\boldsymbol{\beta}}(k_{i_1},k_{i_2}|U)$ for $1\le i_1,i_2\le n$, we begin by claiming

$$
\max_{1\le i_1,i_2\le n}\left|M(i_1,i_2)-\frac{1}{(n-1)^2}\sum_{j_1\neq i_1,j_2\neq i_2}\ \sum_{u\in\{i_1,j_1\},v\in\{i_2,j_2\}}\left[C_0+C_1(\beta_u+\beta_v)\right]M(u,v)\right|=O(1),
\tag{47}
$$

where

$$
C_0:=\theta^2\mathrm{sech}^4(2\theta t),\quad C_1:=\frac{\theta^2}{2}\mathrm{sech}^2(2\theta t).
\tag{48}
$$

Given (47), and noting that $Var_{n,\theta,\boldsymbol{\beta}}(k_i|U)=O(n)$ by Lemma 2.3 part (b), we conclude

$$
\max_{i_1\neq i_2}\left|M(i_1,i_2)-\sum_{j_1\neq j_2}B_n\Big((i_1,i_2),(j_1,j_2)\Big)M(j_1,j_2)\right|=O(1),
\tag{49}
$$

where B_n is a symmetric $n(n-1)$ matrix with non-negative entries, satisfying

$$
\sum_{j_1\neq j_2}B_n\Big((i_1,i_2),(j_1,j_2)\Big)\le 8(C_0+2A)\overset{A\to 0}{\to}8C_0=8\theta^3\mathrm{sech}^6(2\theta t)=[2\theta\mathrm{sech}^2(2\theta t)]^3<1.
$$

where the last inequality uses (Mukherjee and Xu, 2023, Lemma 1.2). Thus the ℓ_∞ operator norm of $(\mathbf{I}-B_n)^{-1}$ converges to $(1-8\theta^3\mathrm{sech}^6(2\theta t))^{-1}<\infty$, which along with (49) gives

$$
\max_{i_1\neq i_2}M(i_1,i_2)=O(1),
$$

as desired.

It thus remains to verify (47). To this end, for any $i \neq j$, we have

$$
\begin{aligned}
&\mathbb{P}_{n,\theta,\boldsymbol{\beta}}(U)Cov_{n,\theta,\boldsymbol{\beta}}(k_i,k_j|U)\\
=&\mathbb{P}_{n,\theta,\boldsymbol{\beta}}(U)\sum_{a\neq i,b\neq j} Cov_{n,\theta,\boldsymbol{\beta}}(Y_{ia},Y_{jb}|U)\\
=&\sum_{a\neq i,b\neq j}\Big\{\mathbb{E}_{\beta}\Big[Y_{ia}Y_{jb}1\{U\}\Big]-\mathbb{E}_{\beta}\Big[Y_{ia}1\{U\}\Big]\mathbb{E}_{n,\theta,\boldsymbol{\beta}}\Big[Y_{jb}1\{U\}\Big]\Big\}\\
=&\sum_{a\neq i,b\neq j}\Big\{\mathbb{E}_{\beta}\Big[Y_{ia}Y_{jb}1\{\widetilde{U}\}\Big]-\mathbb{E}_{\beta}\Big[Y_{ia}1\{\widetilde{U}\}\Big]\mathbb{E}_{n,\theta,\boldsymbol{\beta}}\Big[Y_{jb}1\{\widetilde{U}\}\Big]\Big\}+O(e^{-cn})\\
=&\sum_{a\neq i,b\neq j}\Big\{\mathbb{E}_{\beta}\Big[r_{ia}r_{jb}1\{\widetilde{U}\}\Big]-\mathbb{E}_{\beta}\Big[r_{ia}1\{\widetilde{U}\}\Big]-\mathbb{E}_{\beta}\Big[r_{jb}1\{\widetilde{U}\}\Big]\Big\}+O(e^{-cn})\\
=&\mathbb{P}_{n,\theta,\boldsymbol{\beta}}(\widetilde{U})Cov_{n,\theta,\boldsymbol{\beta}}(r_{ia},r_{jb}|\widetilde{U})+O(e^{-cn}).
\end{aligned}
\tag{50}
$$

In the above display, $r_{ij} := \tanh\Big[\theta(\phi_i+\phi_a)+\frac{1}{2}(\beta_i+\beta_a)\Big]$. A Taylor's series expansion gives

$$
\begin{aligned}
r_{ij} =&\tanh\Big[\theta(\phi_i+\phi_j)+\frac{1}{2}(\beta_i+\beta_j)\Big]\\
=&\tanh(2\theta t)+\Big[\theta(\phi_i-t+\phi_j-t)+\frac{1}{2}(\beta_i+\beta_j)\Big]\operatorname{sech}^2(2\theta t)\\
&+\frac{1}{2}\Big[\theta(\phi_i-t+\phi_j-t)+\frac{1}{2}(\beta_i+\beta_j)\Big]^2\tanh''(2\theta t)+\xi_{ij},
\end{aligned}
$$

where

$$
|\xi_{ij}| \lesssim |\phi_i-t|^3+|\phi_j-t|^3+|\beta_i|^3+|\beta_j|^3 \lesssim |\phi_i-t|^3+|\phi_j-t|^3+n^{-3/2}.
$$

Using the above display we have

$$
Cov_{n,\theta,\boldsymbol{\beta}}(r_{ia},r_{jb}|\widetilde{U}) = \sum_{u\in\{i,a\},v\in\{j,b\}}\Big[C_0+C_1(\beta_u+\beta_v)\Big]Cov_{n,\theta,\boldsymbol{\beta}}(\phi_u,\phi_v|\widetilde{U})+O(n^{-2}), \tag{51}
$$

where the bound on the error term uses Lemma 2.3 part (b) and Lemma 3.4. In the above display, the constants C_0, C_1 are as in (48).

Finally, we have

$$
\begin{aligned}
&\mathbb{P}_{n,\theta,\boldsymbol{\beta}}(\widetilde{U})Cov_{n,\theta,\boldsymbol{\beta}}(\phi_u,\phi_v|\widetilde{U})\\
=&\mathbb{E}_{n,\theta,\boldsymbol{\beta}}\Big[\phi_u\phi_v1\{\widetilde{U}\}\Big]-\mathbb{E}_{n,\theta,\boldsymbol{\beta}}\Big[\phi_u1\{\widetilde{U}\}\Big]\mathbb{E}_{n,\theta,\boldsymbol{\beta}}\Big[\phi_v1\{\widetilde{U}\}\Big]\\
=&\mathbb{E}_{n,\theta,\boldsymbol{\beta}}\Big[\phi_u\phi_v1\{U\}\Big]-\mathbb{E}_{n,\theta,\boldsymbol{\beta}}\Big[\phi_u1\{U\}\Big]\mathbb{E}_{n,\theta,\boldsymbol{\beta}}\Big[\phi_v1\{U\}\Big]+O(e^{-cn})\\
=&\frac{1}{(n-1)^2}\mathbb{E}_{n,\theta,\boldsymbol{\beta}}\Big[k_uk_v1\{U\}\Big]-\frac{1}{(n-1)^2}\mathbb{E}_{n,\theta,\boldsymbol{\beta}}\Big[k_u1\{U\}\Big]\mathbb{E}_{n,\theta,\boldsymbol{\beta}}\Big[k_v1\{U\}\Big]+O(e^{-cn})\\
=&\frac{\mathbb{P}_{n,\theta,\boldsymbol{\beta}}(U)}{(n-1)^2}Cov_{n,\theta,\boldsymbol{\beta}}(k_u,k_v|U)+O(e^{-cn}).
\end{aligned}
\tag{52}
$$

Combining (50), (51) and (52) we have

$$Cov_{n,\theta,\boldsymbol{\beta}}(k_i,k_j)=\frac{1}{(n-1)^2}\sum_{a\neq i,b\neq j}\sum_{u\in\{i,a\},v\in\{j,b\}}[C_0+C_1(\beta_u+\beta_v)]Cov_{n,\theta,\boldsymbol{\beta}}(k_u,k_v|U)+O(1),$$

from which (47) follows. This completes the proof of the lemma. □

Acknowledgements

The authors thank Rajarshi Mukherjee for helpful discussions throughout the project. The authors also thank the AE and an anonymous referee for helpful comments. SM gratefully thanks NSF (DMS 1712037) for support during this research.

Supplementary Material

Supplement to "Signal detection in degree corrected ERGMs" (DOI: 10.3150/23-BEJ1651SUPP; .pdf). The supplementary material contains the proofs of Proposition 2.1, Lemma 2.3, and Lemma 3.3.

References

Anderson, C.J., Wasserman, S. and Crouch, B. (1999). A p^* primer: Logit models for social networks. *Soc. Netw.* **21** 37–66.
Bhamidi, S., Steele, J.M. and Zaman, T. (2015). Twitter event networks and the superstar model. *Ann. Appl. Probab.* **25** 2462–2502. MR3375881 https://doi.org/10.1214/14-AAP1053
Blitzstein, J. and Diaconis, P. (2011). A sequential importance sampling algorithm for generating random graphs with prescribed degrees. *Internet Math.* **6** 489–522.
Burnašev, M.V. (1979). Minimax detection of an imperfectly known signal against a background of Gaussian white noise. *Teor. Veroyatn. Primen.* **24** 106–118. MR0522240
Chatterjee, S. (2007). Stein's method for concentration inequalities. *Probab. Theory Related Fields* **138** 305–321. MR2288072 https://doi.org/10.1007/s00440-006-0029-y
Chatterjee, S. and Diaconis, P. (2013). Estimating and understanding exponential random graph models. *Ann. Statist.* **41** 2428–2461. MR3127871 https://doi.org/10.1214/13-AOS1155
Chatterjee, S., Diaconis, P. and Sly, A. (2011). Random graphs with a given degree sequence. *Ann. Appl. Probab.* **21** 1400–1435. MR2857452 https://doi.org/10.1214/10-AAP728
Chatterjee, S. and Mukherjee, S. (2019). Estimation in tournaments and graphs under monotonicity constraints. *IEEE Trans. Inf. Theory* **65** 3525–3539. MR3959003 https://doi.org/10.1109/TIT.2019.2893911
Deb, N. and Mukherjee, S. (2023). Fluctuations in mean-field Ising models. *Ann. Appl. Probab.* **33** 1961–2003. MR4583662 https://doi.org/10.1214/22-aap1857
Deb, N., Mukherjee, R., Mukherjee, S. and Yuan, M. (to appear). Detecting structured signals in Ising models. *Ann. Appl. Probab.*
Dembo, A. and Montanari, A. (2010). Gibbs measures and phase transitions on sparse random graphs. *Braz. J. Probab. Stat.* **24** 137–211. MR2643563 https://doi.org/10.1214/09-BJPS027
Frank, O. and Strauss, D. (1986). Markov graphs. *J. Amer. Statist. Assoc.* **81** 832–842. MR0860518
Götze, F., Sambale, H. and Sinulis, A. (2021). Concentration inequalities for polynomials in α-sub-exponential random variables. *Electron. J. Probab.* **26** 48. MR4247973 https://doi.org/10.1214/21-ejp606
Holland, P.W. and Leinhardt, S. (1981). An exponential family of probability distributions for directed graphs. *J. Amer. Statist. Assoc.* **76** 33–65. MR0608176
Ingster, Y.I. (1994). Minimax detection of a signal in l_p metrics. *J. Math. Sci.* **68** 503–515.

Ingster, Y., Ingster, J.I. and Suslina, I. (2003). *Nonparametric Goodness-of-Fit Testing Under Gaussian Models* **169**. Berlin: Springer.

Lebowitz, J.L. (1974). GHS and other inequalities. *Comm. Math. Phys.* **35** 87–92. MR0339738

Mukherjee, R., Mukherjee, S. and Sen, S. (2018). Detection thresholds for the β-model on sparse graphs. *Ann. Statist.* **46** 1288–1317. MR3798004 https://doi.org/10.1214/17-AOS1585

Mukherjee, R., Mukherjee, S. and Yuan, M. (2018). Global testing against sparse alternatives under Ising models. *Ann. Statist.* **46** 2062–2093. MR3845011 https://doi.org/10.1214/17-AOS1612

Mukherjee, S. and Xu, Y. (2023). Statistics of the two star ERGM. *Bernoulli* **29** 24–51. MR4497238 https://doi.org/10.3150/21-bej1448

Park, J. and Newman, M.E.J. (2004). Solution of the two-star model of a network. *Phys. Rev. E (3)* **70** 066146. MR2133810 https://doi.org/10.1103/PhysRevE.70.066146

Rinaldo, A., Petrović, S. and Fienberg, S.E. (2013). Maximum likelihood estimation in the β-model. *Ann. Statist.* **41** 1085–1110. MR3113804 https://doi.org/10.1214/12-AOS1078

Robins, G., Pattison, P., Kalish, Y. and Lusher, D. (2007). An introduction to exponential random graph (p^*) models for social networks. *Soc. Netw.* **29** 173–191.

Schweinberger, M. and Stewart, J. (2020). Concentration and consistency results for canonical and curved exponential-family models of random graphs. *Ann. Statist.* **48** 374–396. MR4065166 https://doi.org/10.1214/19-AOS1810

Shalizi, C.R. and Rinaldo, A. (2013). Consistency under sampling of exponential random graph models. *Ann. Statist.* **41** 508–535. MR3099112 https://doi.org/10.1214/12-AOS1044

Wasserman, S. and Faust, K. (1994). *Social Network Analysis: Methods and Applications* **8**. Cambridge: Cambridge University Press.

Wasserman, S. and Pattison, P. (1996). Logit models and logistic regressions for social networks. I. An introduction to Markov graphs and *p*. *Psychometrika* **61** 401–425. MR1424909 https://doi.org/10.1007/BF02294547

Xu, Y. and Mukherjee, S. (2024). Supplement to "Signal detection in degree corrected ERGMs." https://doi.org/10.3150/23-BEJ1651SUPP

Received October 2022 and revised July 2023