

DETECTION THRESHOLDS FOR THE β -MODEL ON SPARSE GRAPHS

BY RAJARSHI MUKHERJEE, SUMIT MUKHERJEE¹ AND SUBHABRATA SEN²

University of California, Berkeley, Columbia University and Microsoft Research

In this paper, we study sharp thresholds for detecting sparse signals in β -models for potentially sparse random graphs. The results demonstrate interesting interplay between graph sparsity, signal sparsity and signal strength. In regimes of moderately dense signals, irrespective of graph sparsity, the detection thresholds mirror corresponding results in independent Gaussian sequence problems. For sparser signals, extreme graph sparsity implies that all tests are asymptotically powerless, irrespective of the signal strength. On the other hand, sharp detection thresholds are obtained, up to matching constants, on denser graphs. The phase transitions mentioned above are sharp. As a crucial ingredient, we study a version of the higher criticism test which is provably sharp up to optimal constants in the regime of sparse signals. The theoretical results are further verified by numerical simulations.

1. Introduction. Real life networks are ubiquitous in the natural and social sciences. Over the last decade, a huge volume of data has been collected on the structure of large networks and dynamics on these systems. Specialists from a wide array of disciplines such as physics, biology, social science, computer science and statistics have studied these complex systems using specialized techniques. This area has seen an explosion of research in the recent years, and the resulting confluence of diverse ideas has enriched this field of study.

From a statistical perspective, one avenue of research has focused on designing random graph models, which exhibit some properties of real world networks such as a heavy-tailed degree distribution [Barabási and Albert (1999)], a “small world” characteristic [Watts and Strogatz (1998)], etc. There has also been some preliminary work on testing for goodness-of-fit of these models to real data [Bickel, Chen and Levina (2011)].

A parallel line of research has focused on the design and analysis of algorithms for finding structural patterns or “motifs” in network data. Finding these structures is of interest to practitioners in diverse research areas as they often represent functional relationships between the vertices. For example, in a social network setup,

Received August 2016; revised May 2017.

¹Supported in part by NSF Grant DMS-1712037.

²Supported in part by the William R. and Sara Hart Kimball Stanford Graduate Fellowship.
MSC2010 subject classifications. 62G10, 62G20, 62C20.

Key words and phrases. Detection boundary, sparse random graphs, beta model, higher criticism, sparse signals.

a group of vertices with a high edge density may represent a community of individuals who share some common characteristics, for example, a common profession. Guarantees for relevant statistical algorithms are usually derived by analyzing their typical performance on specific random graphs. Prominent examples include finding a single planted community or finding two equal sized communities in a graph. The simpler problem of detecting the presence of these structural patterns is itself statistically challenging, as they are often rare, and thus one needs to design procedures which work in a low signal to noise ratio setup. This has been accomplished for the single community case in Arias-Castro and Verzelen (2013), Verzelen and Arias-Castro (2015). This body of work connects to the broader statistical program of global testing against structured alternatives [Ingster and Suslina (2003), Donoho and Jin (2004), Arias-Castro, Donoho and Huo (2005), Arias-Castro et al. (2008), Addario-Berry et al. (2010), Hall and Jin (2010), Ingster, Tsybakov and Verzelen (2010), Arias-Castro, Candès and Plan (2011), Cai and Yuan (2014), Arias-Castro and Wang (2015), Mukherjee, Pillai and Lin (2015)].

It is well known in social science that vertices of a network are often differentially “attractive,” with the more popular vertices having a higher tendency to form edges. Finding the vertices with higher “attractiveness” is of interest in a number of different contexts—they often represent highly influential nodes and practitioners might wish to screen these vertices for subsequent study. In an online social network, these vertices might actually represent spam accounts which should be detected and removed.

In this paper, we formulate and address the question of detecting differential attractiveness of vertices in networks. While this question is statistically simpler, we expect that mathematically probing the limits of detection can also provide nontrivial information about the corresponding estimation question. The detection problem is of independent interest in a different context—assume that given a network dataset, one wishes to fit a stochastic block model to capture the community structure present in the data. It is well known that differential attractiveness of vertices can confound and affect model fits in this case [Karrer and Newman (2011), Yan et al. (2014)]. Detecting the presence of this structural pattern is therefore imperative in this scenario. We believe that similar testing questions could be useful even in such setups.

1.1. *Framework.* Formally, we observe a labeled graph $\mathcal{G} = (V, E)$. We define $|V| = n$ and we are mainly interested in the asymptotic regime where the graph size $n \rightarrow \infty$. The graph \mathcal{G} can be equivalently described by the adjacency matrix $\mathbf{Y} = (Y_{ij} : 1 \leq i, j \leq n)$, where $Y_{ij} = 1$ if $\{i, j\} \in E$ and $Y_{ij} = 0$ otherwise. We note that trivially, $Y_{ij} = Y_{ji}$ while $Y_{ii} = 0$, as we consider graphs without loops.

Next, we specify a statistical model for the graphs in this context. For any $\boldsymbol{\beta} := (\beta_1, \dots, \beta_n) \in \mathbb{R}^n$ and $\lambda \in [1, n]$, we define a probability distribution on the space

of labeled graphs by setting

$$(1.1) \quad \mathbb{P}_{\beta, \lambda}(\mathbf{Y}) = \prod_{1 \leq i < j \leq n} p_{ij}^{Y_{ij}} (1 - p_{ij})^{1 - Y_{ij}}, \quad p_{ij} := \frac{\lambda}{n} \frac{e^{\beta_i + \beta_j}}{1 + e^{\beta_i + \beta_j}}.$$

Thus under the distribution $\mathbb{P}_{\beta, \lambda}$, the edges are mutually independent with $Y_{ij} \sim \text{Bin}(1, p_{ij})$. The dependence of $\mathbb{P}_{\beta, \lambda}$ on n will be suppressed throughout. β_i should be interpreted as the ‘‘attractiveness’’ of vertex i , while λ controls the sparsity of the graph obtained. We will assume throughout λ is known and suppress the dependence of λ on n . We note that the model (1.1) is intimately related to the p_1 -model of Holland and Leinhardt (1981) and the β -model of Chatterjee, Diaconis and Sly (2011); see Section 1.2 for the background on these models and connections to (1.1). We discuss more on the choice of the model (1.1) and the knowledge of λ in Section 5.

In the context of the above model, we will formulate our problem as a goodness-of-fit type global null hypothesis testing problem against a structured hypothesis. To this end, we define the parameter space

$$(1.2) \quad \Xi(s, A) := \{\beta \in \mathbb{R}_+^n : |S(\beta)| = s, \beta_i \geq A, i \in S(\beta)\},$$

where $S(\beta) := \{1 \leq i \leq n : \beta_i \neq 0\}$ and $\mathbb{R}_+ = [0, \infty)$. The vertices $i \in S(\beta)$ should be interpreted as the popular vertices. Since we expect such vertices to be rare, mathematically we consider the following sequence of hypothesis testing problems:

$$(1.3) \quad H_0 : \beta = \mathbf{0} \quad \text{vs.} \quad H_1 : \beta \in \Xi(s_n, A_n) \subset \mathbb{R}_+^n \setminus \{\mathbf{0}\}$$

for any pair of sequences s_n, A_n . Throughout, we parameterize signal sparsity $s_n = n^{1-\alpha}$ with $\alpha \in (0, 1)$. A statistical test for H_0 versus H_1 is a measurable $\{0, 1\}$ valued function of the data \mathbf{Y} , with 1 denoting the rejection of the null hypothesis H_0 and 0 denoting the failure to reject H_0 . The worst case risk of a test $T_n(\mathbf{Y})$ is defined as

$$(1.4) \quad \text{Risk}_n(T_n, \Xi(s_n, A_n)) := \mathbb{P}_{\mathbf{0}, \lambda}(T_n = 1) + \sup_{\beta \in \Xi(s_n, A_n)} \mathbb{P}_{\beta, \lambda}(T_n = 0).$$

A sequence of tests T_n corresponding to a sequence of model-problem pairs (1.1)–(1.3) is said to be asymptotically powerful (resp., asymptotically powerless) against $\Xi(s_n, A_n)$ if

$$\limsup_{n \rightarrow \infty} \text{Risk}_n(T_n, \Xi(s_n, A_n)) = 0$$

(respectively $\liminf_{n \rightarrow \infty} \text{Risk}_n(T_n, \Xi(s_n, A_n)) = 1$).

1.2. *Background on the β -model for random graphs.* The study of degree sequences of network data has a long and rich history [Holland and Leinhardt (1981), Fienberg and Wasserman (1981), Robins et al. (2007), Goodreau (2007), Barvinok and Hartigan (2013)]. The simplest model for networks based on the degree sequence is an exponential family with the degree sequence as its sufficient statistic. This model is a special case of the p_1 model of Holland and Leinhardt (1981), and will be called the β -model, following recent terminology of Chatterjee, Diaconis and Sly (2011). We will not survey the vast literature on the p_1 model and its applications and instead refer the reader to Blitzstein and Diaconis (2010) for detailed references. The undirected model can be equivalently described as follows: given a vector $\boldsymbol{\beta} = (\beta_1, \dots, \beta_n) \in \mathbb{R}^n$, we form a graph on the vertex set V with $|V| = n$ and edges are added independently with probability

$$(1.5) \quad p_{ij} = \frac{\exp(\beta_i + \beta_j)}{1 + \exp(\beta_i + \beta_j)}.$$

We note that the models (1.1) and (1.5) are identical except for the leading λ/n factor. This factor is introduced to control the sparsity of the graph. The original β -model leads to dense graphs, where a graph on n vertices has $O_p(n^2)$ edges. However, it is widely accepted that real networks are seldom dense, and thus to address this issue, we introduce an additional parameter λ .

The original model (1.5) has been studied widely in recent years. It is known to be the maximum entropy distribution given the degree distribution [see Blitzstein and Diaconis (2010)]. Lauritzen (2002, 2008) characterized these models as natural models for *weakly summarized* exchangeable binary arrays, that is, arrays with distributions determined by row and column totals. Statistical analysis of the model (1.5) has also received a lot of attention—the existence and consistency of MLEs in these models were examined by Chatterjee, Diaconis and Sly (2011) and Rinaldo, Petrović and Fienberg (2013), and normal fluctuation for the MLE was established in Yan and Xu (2013). Perry and Wolfe (2012) also study a general model which includes the β -model as a special case. Finite sample analysis of these models using Markov bases has also been explored in Petrović, Rinaldo and Fienberg (2010), Hara and Takemura (2010), Ogawa, Hara and Takemura (2013), Yan and Xu (2013). In a different direction, Karwa and Slavković (2016) study differentially private parameter estimation in the β model, while Hillar and Wibisono (2013) and Yan, Zhao and Qin (2015) generalize the β model to weighted graphs. We also refer the reader to Yan, Qin and Wang (2016) for asymptotic results on a general family of network models, which includes the β model as a special case.

1.3. *Notation.* For any $n \in \mathbb{N}$, we let $[n] = \{1, \dots, n\}$. For any $i \in [n]$, we denote the degree of vertex i by $d_i := \sum_{j=1}^n Y_{ij} = \sum_{j \neq i} Y_{ij}$. Throughout, $\text{Bin}(n, p)$ will stand for a generic binomial random variable with $n \in \mathbb{N}$ trials and success probability $p \in [0, 1]$. The results in this paper are mostly asymptotic in nature,

and thus require some standard asymptotic notation. If a_n and b_n are two sequences of real numbers, then $a_n \gg b_n$ (and $a_n \ll b_n$) implies that $a_n/b_n \rightarrow \infty$ (resp., $a_n/b_n \rightarrow 0$) as $n \rightarrow \infty$. Similarly, $a_n \gtrsim b_n$ (and $a_n \lesssim b_n$) implies that $\liminf a_n/b_n = C$ for some $C \in (0, \infty]$ (and $\limsup a_n/b_n = C$ for some $C \in [0, \infty)$). Alternatively, $a_n = o(b_n)$ will also imply $a_n \ll b_n$ and $a_n = O(b_n)$ will imply that $\limsup a_n/b_n = C$ for some $C \in [0, \infty)$. We write $a_n = \Theta(b_n)$ if both $a_n = O(b_n)$ and $b_n = O(a_n)$. We write $a_n \sim b_n$ if $\lim \frac{a_n}{b_n} \rightarrow 1$. For any fixed tuple v of real numbers, $C(v)$ will denote a constant depending on elements of v only. Also, throughout we drop the subscript n whenever it is understood that s, A, λ are allowed to vary with n .

2. Tests. The tests used in this paper for the purpose of asymptotically sharp detection are all based on the degree vector (d_1, \dots, d_n) . All the tests aim to capture the idea that the coordinates of the degree vector are stochastically increasing according to the corresponding coordinates of β .

Total Degree Test: This test is based on the total degree in the observed graph, that is, $\sum_{i=1}^n d_i$. The test rejects when the observed total degree is large. The calibration of this test can be achieved by looking at the behavior of $\sum_{i=1}^n d_i$ under the null hypothesis in (1.3). More precisely, by the total degree test we mean a testing procedure which rejects when $\sum_{i=1}^n d_i - \frac{\lambda(n-1)}{2}$ is large (see the proof of Theorem 3.1).

Maximum Degree Test: This test is based on the maximum degree in the observed graph, that is, $\max_{i=1}^n d_i$. The distribution of $\max_{i=1}^n d_i$ under the null hypothesis for $\lambda \gg \log^3 n$ is standard [Bollobás (2001)]. However, for the theoretical calibration of an asymptotically powerful test, we simply need suitable control over the tail of the maximum degree (see the proof of Theorem 3.3). In the rest of the paper, by the maximum degree test we shall mean a testing procedure that rejects when the observed maximum degree is large.

Higher Criticism Test: This test is based on suitably scanning over centered and scaled survival statistics of the degree vector. More precisely, for any $t > 0$ we let

$$HC(t) := \sum_{i=1}^n (\mathcal{I}(D_i > t) - \mathbb{P}_{\beta=0,\lambda}(D_i > t)),$$

$$D_i = \frac{d_i - \frac{\lambda}{2n}(n-1)}{\sqrt{(n-1)\frac{\lambda}{2n}(1 - \frac{\lambda}{2n})}}, \quad i = 1, \dots, n.$$

We then construct a version of the higher criticism test as follows. Define

$$HC := \sup \left\{ GHC(t) := \frac{HC(t)}{\sqrt{\text{Var}_{\beta=0,\lambda}(HC(t))}}, t \in \{\sqrt{2r \log n} : r \in (0, 5)\} \cap \mathbb{N} \right\}.$$

By the higher criticism test, we then mean a testing procedure that rejects when the observed value of HC defined above is large [see the proof of Theorem 3.2.ii(a)].

Indeed this test is designed along the tradition of tests introduced and studied in recent history of sparse signal detection in sequence and regression models [Donoho and Jin (2004), Hall and Jin (2010), Arias-Castro, Candès and Plan (2011), Arias-Castro and Wang (2015), Mukherjee, Pillai and Lin (2015), Barnett, Mukherjee and Lin (2017)]. However, in spite of the similarity in philosophy of construction of the test, sharp analysis of the test turns out to be subtle due to the presence of dependence among the degree sequence. To our best knowledge, this is the first instance of sharp analysis of the higher criticism test for a dependent binomial sequence. This is one of the main technical contributions of the paper, since tackling these dependence issues is fundamentally different from analyzing a weakly dependent Gaussian sequence [Arias-Castro, Candès and Plan (2011)]. A glance at our proof reveals that the case of sparser graphs, with λ behaving polylogarithmically in n is relatively easier. This is intuitively reasonable, since sparse graphs implies “weaker” dependence among the degrees. Since our results are valid for any $\lambda \gg \log n$, more care is needed to attend to the dependence structure in the problem.

3. Main results. We divide the main results according to signal sparsity $\alpha \in (0, 1)$. Our first theorem corresponds to the dense regime of signal sparsity, that is, $\alpha \leq \frac{1}{2}$.

THEOREM 3.1. *Suppose $\alpha \leq \frac{1}{2}$ and let*

$$C_{\text{dense}}(\alpha) = \frac{1}{2} - \alpha.$$

(i) *The total degree test is asymptotically powerful if*

$$(3.1) \quad \tanh(A) \geq \frac{n^{-r}}{\sqrt{\lambda}}, \quad r < C_{\text{dense}}(\alpha).$$

(ii) *All tests are asymptotically powerless if*

$$(3.2) \quad \tanh(A) \leq \frac{n^{-r}}{\sqrt{\lambda}}, \quad r > C_{\text{dense}}(\alpha).$$

We note that one can actually state a slightly stronger result which dictates that all tests are asymptotically powerless when $\lim_{n \rightarrow \infty} s \tanh(A) \sqrt{\frac{\lambda}{n}} = 0$ while the total degree test is asymptotically powerful whenever there exists a diverging sequence $t_n \rightarrow \infty$ such that $s \tanh(A) \sqrt{\frac{\lambda}{n}} \gg t_n$.

Our next result characterizes the detection thresholds in the sparser regime, that is, $\alpha > \frac{1}{2}$. Unlike the denser regime ($\alpha \leq \frac{1}{2}$), the results here depend on the level of graph sparsity λ . In particular, for small λ 's (very sparse graphs) all tests turn out to be asymptotically powerless irrespective of the signal strength.

THEOREM 3.2. Suppose $\alpha \in (1/2, 1)$, $\theta := \lim_{n \rightarrow \infty} \frac{\lambda}{2n}$, and let

$$C_{\text{sparse}}(\alpha) = \begin{cases} 16(1 - \theta)\left(\alpha - \frac{1}{2}\right), & \frac{1}{2} < \alpha < \frac{3}{4}, \\ 16(1 - \theta)(1 - \sqrt{1 - \alpha})^2, & \alpha \geq \frac{3}{4}. \end{cases}$$

(i) Assume $\lambda \ll \log n$. Then all tests are asymptotically powerless irrespective of A .

(ii) Assume $\lambda \gg \log n$.

(a) The higher criticism test is asymptotically powerful if

$$\tanh(A) \geq \sqrt{\frac{C^* \log n}{\lambda}}, \quad C^* > C_{\text{sparse}}(\alpha).$$

(b) All tests are asymptotically powerless if

$$\tanh(A) \leq \sqrt{\frac{C^* \log n}{\lambda}}, \quad C^* < C_{\text{sparse}}(\alpha).$$

REMARK 1. The detection limits presented above are reminiscent of those in Mukherjee, Pillai and Lin (2015). However, where results in Mukherjee, Pillai and Lin (2015) in essence correspond to $\lambda = n$, Theorem 3.2 presents a richer class of phase transitions by taking into account a continuous family of parameters $\theta := \lim \lambda/2n$. However, the proof of such a result is simpler than that presented here due to the lack of dependence structure in traditionally studied regression models. Of course, the essence of our proof rests on understanding and carefully isolating the dependence structure present, and thereby reducing the calculations to a simpler independent binomial sequence problem. The subtlety of the proof relies on keeping track of the errors accumulated in such reductions. Finally, although the proof of the lower bound [Theorem 3.2(ii)(b)] follows a general scheme of bounding a truncated second moment, deciding on the truncation event is subtle.

REMARK 2. The result that all tests are asymptotically powerless when $\lambda \ll \log n$ is intuitive since according to our parameterization $p_{ij} \in [\frac{\lambda}{2n}, \frac{\lambda}{n}]$ regardless of the membership of the vertices in S or S^c . In order to achieve nontrivial detection boundary for $\lambda \ll \log n$, one might need to work with a different parameterization of graph sparsity, which allows for varying graph density between the null and alternative hypotheses.

Theorem 3.2 establishes the sharp optimality of the higher criticism test in the sparse signal regime. Another commonly studied test of practical interest is the maximum degree test introduced in Section 2. Our next theorem characterizes the detection limits of the maximum degree test the proof of which is deferred to Appendix D in Mukherjee, Mukherjee and Sen (2018).

THEOREM 3.3. *Suppose $\alpha \in (1/2, 1)$, $\theta := \lim_{n \rightarrow \infty} \frac{\lambda}{2n}$, and let*

$$C_{\max}(\alpha) = 16(1 - \theta)(1 - \sqrt{1 - \alpha})^2.$$

(i) *Assume $\lambda \gg \log n$. Then the maximum degree test is asymptotically powerful if*

$$\tanh(A) \geq \sqrt{\frac{C^* \log n}{\lambda}}, \quad C^* > C_{\max}(\alpha).$$

(ii) *Assume $\lambda \gg \log^3 n$. Then the maximum degree test is asymptotically powerless if*

$$\tanh(A) \leq \sqrt{\frac{C^* \log n}{\lambda}}, \quad C^* < C_{\max}(\alpha).$$

The results in Theorem 3.3 establish that for $\lambda \gg \log^3 n$, the Maximum Degree Test is sharp optimal iff $\alpha > 3/4$, and loses out to the Higher Criticism Test in the regime of $\alpha \in (1/2, 3/4)$. Although this result is parallel to those observed in sparse detection problems for independent Gaussian and binomial sequence models [Arias-Castro, Candès and Plan (2011), Mukherjee, Pillai and Lin (2015)], the proof of this fact is substantially more involved. This is in particular true about the proof of the lower bound part in Theorem 3.3(ii). Also note that Theorem 3.3(i) hold as soon as $\lambda \gg \log n$. However, in order to prove the lower bound part in the theorem above we crucially make use of the null distribution of $\max_{i=1}^n d_i$, which is readily available for $\lambda \gg \log^3 n$ [Bollobás (2001)]. The situation becomes highly subtle for $\log n \ll \lambda \lesssim \log^3 n$. In particular, we were able to argue that for $\log n \ll \lambda \lesssim \log^3 n$, if one considers the Maximum Degree Test that rejects when $\max_{i=1}^n d_i > np_n + \sqrt{\delta_n np_n q_n \log n}$, where $p_n = \lambda/2n$, $q_n = 1 - p_n$, and δ_n is some sequence of real numbers, then such tests are asymptotically powerless as soon as $C^* < C_{\max}(\alpha)$ defined above, if $\limsup \delta_n \neq 2$. Lowering the requirement of $\lambda \gg \log^3 n$ to $\lambda \gg \log n$ in this case requires a second moment argument along with the Paley–Zygmund inequality. We refer to Appendix E in Mukherjee, Mukherjee and Sen (2018). for more details. On the other hand, the case when $\limsup \delta_n = 2$ is extremely challenging, and the result of the testing problem depends on the rate of convergence of δ_n to 2 along subsequences. We leave this effort to future ventures.

4. Simulation results. We now present the results of some numerical experiments in order to demonstrate the behavior of the various tests in finite samples. To put ourselves in context of our asymptotic analysis, we chose to work with $n = 100$. Since our theoretical results depend both on the signal sparsity and graph

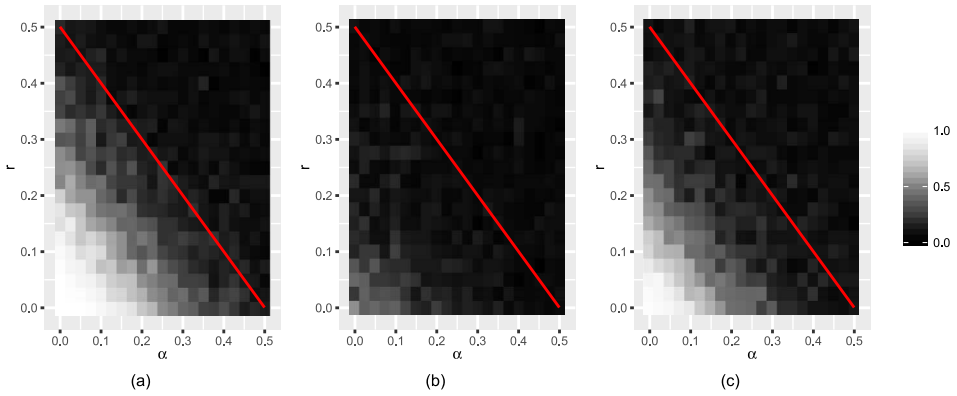


FIG. 1. The power of testing procedures in the dense signal setup. (a) shows the power of the Total Degree test, (b) plots the max-degree while (c) plots the power of the GHC statistic. The theoretical detection threshold is drawn in red.

sparsity, we divide our simulation results accordingly. In each of the situations, we compare the power of the tests (Total Degree, Maximum degree and Higher Criticism, resp.) by fixing the levels at 5%. In particular, we generate the test statistics 100 times under the null and take the 95%-quantile as the cut-off value for rejection. The power against different alternatives are then obtained empirically from 100 repeats each.

Our first set of simulations corresponds to the dense regime, that is, $\alpha \leq \frac{1}{2}$. In Figure 1, we plot the power of each of the three tests for a range of signal sparsity-strength pairs (α, r) , where $\alpha \in (0, \frac{1}{2})$ with increments of size 0.025 and the signal strength is given by $A = \frac{n-r}{\sqrt{\lambda}}$ with $r \in (0, \frac{1}{2})$ with increments of the size 0.025 and $\lambda = 25$. In addition, we plot the theoretical detection boundary given by $r = \frac{1}{2} - \alpha$ in red. As dictated by our theoretical results, the phase transitions are clear in the simulations as well. In particular, we observe that the Total Degree Test performs better in the dense regime. The Higher Criticism Test seems to have some power; the Maximum Degree Test fails have any power in this regime of sparsity.

Our second set of simulations corresponds to the sparse regime, that is, $\alpha > \frac{1}{2}$. In this case, following the theoretical predictions, we divide our simulations based on the graph sparsity λ as well. In Figure 2, we plot the power of each of the three tests for a range of signal sparsity-strength pairs (α, r) for three different values of λ , namely $\lambda = 2, 10, \text{ or } 25$, respectively. Specifically, we choose for signal sparsity index varying between $\alpha \in (\frac{1}{2}, 1)$ with increments of size 0.025, the signal strength to be $A = \sqrt{\frac{C \log n}{\lambda}}$ with $C \in (0, 16)$ with increments of size 0.5. In addition, we plot the theoretical detection boundary given by $r = C_{\text{sparse}}(\alpha) := 16(1 - \theta)\{(\alpha - 1/2)\mathcal{I}(\alpha < 3/4) + (1 - \sqrt{1 - \alpha})^2\mathcal{I}(\alpha \geq 3/4)\}$ with $\theta = \lambda/2n$, in red. To distinguish between the performance of the Higher Criticism Test and

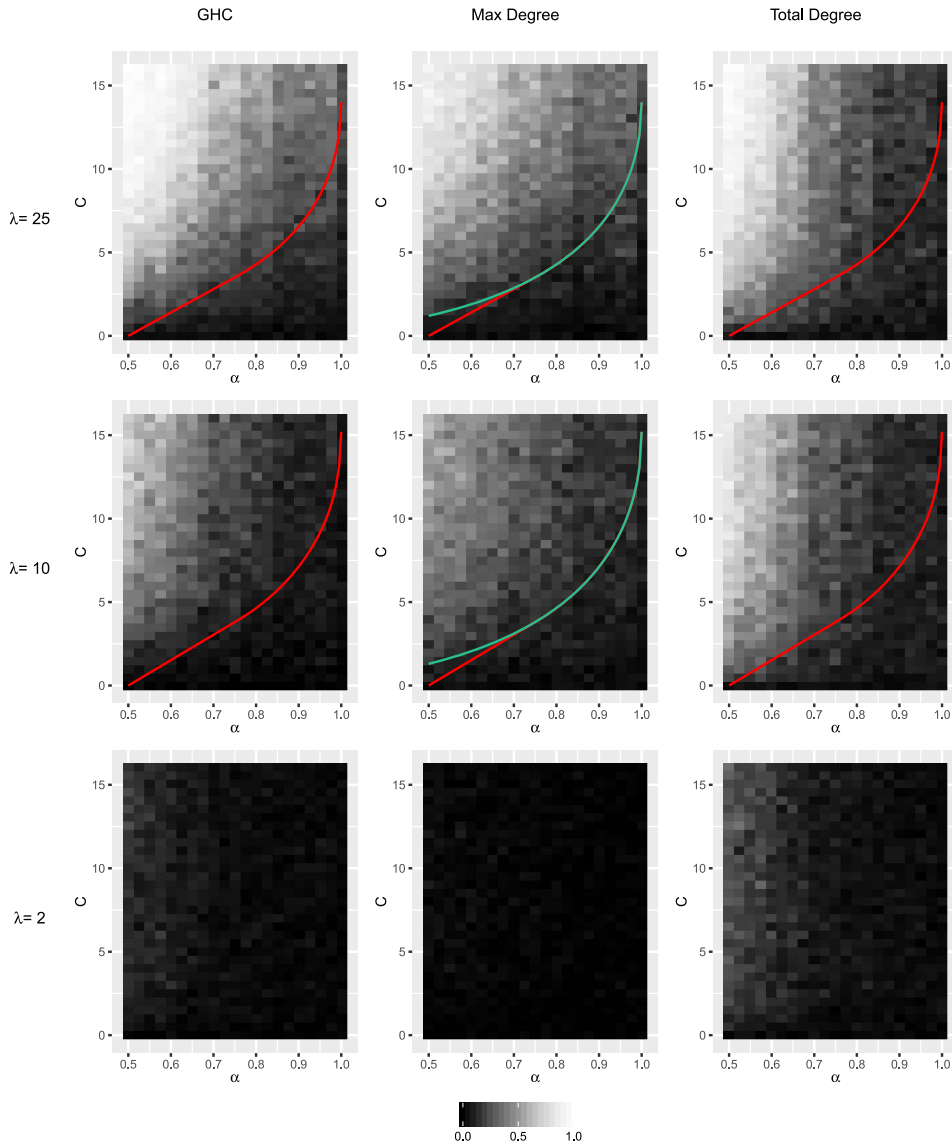


FIG. 2. The power of the testing procedures in the sparse signal setup. The theoretical detection thresholds are drawn in red, while the thresholds for the Maximum Degree Test are drawn in cyan.

the Maximum Degree Test, we also plot the theoretical detection boundary of the Maximum Degree Test in cyan. The simulation results seem to match the theoretical predictions. In particular, when is $\lambda < \log n = 3$, none of the tests seems to have any power in detecting any signal presented. In contrast for λ much larger than $\log n$, the empirical performance of the Higher Criticism Test and Maximum degree Test also follows the theoretical guarantees.

5. Discussion. In this paper, we formulate and study the problem of detecting the presence of differentially attractive nodes in networks through a version of the β -model on sparse graphs. Our mathematical results are sharp in all regimes of signal strength as well as graph and signal sparsity. The model can be further generalized, where instead of using an expit function, the connection probabilities between node i and j are given by $\frac{\lambda}{n}\psi(\beta_i + \beta_j)$ for some function ψ , which is the distribution function of a symmetric random variable, that is, $\psi(x) + \psi(-x) = 1$, and satisfies some reasonable smoothness type regularity conditions. This, in particular, will include the probit link $\psi(x) = \int_{-\infty}^x e^{-t^2/2}/\sqrt{2\pi} dt$. Thereafter, similar to Mukherjee, Pillai and Lin (2015), we believe that one can obtain parallel results here as well, with the sharp constants in the problem changing according to the function ψ .

All of our results assume the knowledge of the graph sparsity parameter λ . For unknown λ , one needs to include supremum over λ in the definition of risk in (1.4). Following intuitions from Arias-Castro and Verzelen (2013), we believe that the nature of detection thresholds for unknown λ remains the same for $\alpha > \frac{1}{2}$. However, it might change for the dense signal regime $\alpha \leq \frac{1}{2}$, where it might be more prudent to use a degree variance type test instead of the total degree test. For the sparse regime ($\alpha > \frac{1}{2}$), we intuitively believe that as a first step, one can choose a vertex at random and estimate λ by the proportion of its neighbors. Since $s \ll \sqrt{n}$ in this regime of signal sparsity, a randomly chosen vertex belongs to the set of vertices with corresponding $\beta_i = 0$ with high probability, and as a result it is not too hard to show that the resulting estimate is consistent in ratio scale. One can then delete this randomly chosen vertex from the network and consider the detection problem on the remaining $n - 1$ vertices, and use the estimated λ as a plug-in. It indeed remains a question to study how the performance of the corresponding testing procedures change under this plug-in principle. Although the analysis of the Maximum Degree Test should not be too difficult to carry out using techniques similar to those carried out here, the analysis of the Higher Criticism Test is extremely subtle and is beyond the scope of the current paper.

Although we only study the sparse detection problem in this paper, the rich phase transitions observed hint at similar complexities in further inferential questions. Of particular interest is the subset selection problem, which corresponds to the identification of vertices of differential attractiveness. We plan to address sharp analyses of such inferential questions in future papers.

6. Proofs of main results. In this section, we prove the main results in this paper.

6.1. *Some properties of binomial distribution.* In this section, we collect some properties of the binomial distribution, which will be used throughout this paper. The first lemma is a simple change of measure argument and the proof is omitted.

LEMMA 6.1. *If $X \sim \text{Bin}(n, p)$, then for any $a > 0$ and Borel subset of B of \mathbb{R} :*

$$\mathbb{E}(a^X \mathbf{1}_{X \in B}) = (ap + (1 - p))^n \mathbb{P}\left(\text{Bin}\left(n, \frac{ap}{ap + (1 - p)}\right) \in B\right).$$

The next lemma is crucial and the proof is deferred to Appendix A in Mukherjee, Mukherjee and Sen (2018).

LEMMA 6.2.

(a) *Suppose $X_n \sim \text{Bin}(n, p_n)$ such that $n \min(p_n, 1 - p_n) \gg \log n$, and $\{C_n\}_{n \geq 1}$ is a sequence of reals converging to $C > 0$. Then we have:*

(i)

$$\mathbb{P}(X_n = np_n + C_n \sqrt{np_n(1 - p_n) \log n}) = \frac{1}{\sqrt{np_n}} n^{-\frac{C^2}{2} + o(1)}$$

and

(ii)

$$\lim_{n \rightarrow \infty} \mathbb{P}(X_n \geq np_n + C_n \sqrt{np_n(1 - p_n) \log n}) = n^{-\frac{C^2}{2} + o(1)}.$$

(b) *Suppose further that $Y_n \sim \text{Bin}(b_n, p'_n)$ is independent of X_n , and*

$$\limsup_{n \rightarrow \infty} p_n < 1, \quad \limsup_{n \rightarrow \infty} \frac{b_n}{\sqrt{n}} < \infty, \quad \lim_{n \rightarrow \infty} \frac{p'_n}{p_n} = 1.$$

Then we have:

(i)

$$\mathbb{P}(X_n + Y_n = np_n + C_n \sqrt{np_n(1 - p_n) \log n}) = \frac{1}{\sqrt{np_n}} n^{-\frac{C^2}{2} + o(1)}$$

and

(ii)

$$\mathbb{P}(X_n + Y_n \geq np_n + C_n \sqrt{np_n(1 - p_n) \log n}) = n^{-\frac{C^2}{2} + o(1)}.$$

6.2. *Proof of Theorem 3.1.* We prove here a slightly stronger result, which dictates that all tests are asymptotically powerless when $\lim_{n \rightarrow \infty} s \tanh(A) \sqrt{\frac{\lambda}{n}} = 0$ while the Total Degree test is asymptotically powerful whenever there exists a diverging sequence $t_n \rightarrow \infty$ such that $s \tanh(A) \sqrt{\frac{\lambda}{n}} \gg t_n$. We begin by stating the following elementary lemma.

LEMMA 6.3. *For any $\beta \in \mathbb{R}_+^n, \lambda > 0$, we have*

$$\text{Var}_{\beta, \lambda} \left(\sum_{i=1}^n d_i \right) \leq 2n\lambda.$$

PROOF. The above bound follows on noting that

$$\begin{aligned} \text{Var}_{\beta,\lambda}(d_i) &\leq \sum_{j \neq i} \mathbb{E}_{\beta,\lambda} Y_{ij} \leq \lambda, \\ \text{Cov}_{\beta,\lambda}(d_i, d_j) &= \text{Var}_{\beta,\lambda}(Y_{ij}) \leq \mathbb{E}_{\beta,\lambda} Y_{ij} \leq \frac{\lambda}{n}. \end{aligned} \quad \square$$

PROOF OF THEOREM 3.1(i). Let ϕ_n be the test defined by

$$\begin{aligned} \phi_n &= 1 && \text{if } \sum_{i=1}^n d_i - \frac{(n-1)\lambda}{2} > K_n \\ &= 0 && \text{otherwise,} \end{aligned}$$

where $K_n := \frac{\lambda s}{8} \tanh(\frac{A}{2})$. We define the subparameter space:

$$(6.1) \quad \tilde{\Xi}(s, A) := \{\beta \in \mathbb{R}_+^n : |S(\beta)| = s, \beta_i = A, i \in S(\beta)\}.$$

Since the distribution of $\sum_{i=1}^n d_i$ is stochastically increasing in A , it suffices to prove that

$$(6.2) \quad \limsup_{n \rightarrow \infty} \mathbb{P}_{\beta=0,\lambda} \left(\sum_{i=1}^n d_i - \frac{(n-1)\lambda}{2} > K_n \right) = 0,$$

$$(6.3) \quad \limsup_{n \rightarrow \infty} \sup_{\beta \in \tilde{\Xi}(s,A)} \mathbb{P}_{\beta,\lambda} \left(\sum_{i=1}^n d_i - \frac{(n-1)\lambda}{2} > K_n \right) = 1.$$

For proving (6.2), an application of Chebyshev’s inequality along with Lemma 6.3 gives

$$\mathbb{P}_{\beta=0,\lambda} \left(\sum_{i=1}^n d_i - \frac{(n-1)\lambda}{2} > K_n \right) = \mathbb{P}_{\beta=0,\lambda} \left(\sum_{i=1}^n d_i - \sum_{i=1}^n \mathbb{E}_{\beta=0,\lambda} d_i > K_n \right) \leq \frac{2n\lambda}{K_n^2},$$

which goes to 0 using (3.1) by choice of K_n , thus proving (6.2). Turning to prove (6.3), for any $\beta \in \tilde{\Xi}(s, A)$ we have

$$\begin{aligned} &\sum_{i=1}^n \mathbb{E}_{\beta,\lambda} d_i - \frac{(n-1)\lambda}{2} \\ &= \frac{\lambda}{n} s(s-1) \left(\frac{e^{2A}}{1+e^{2A}} - \frac{1}{2} \right) + \frac{\lambda}{n} 2s(n-s) \left(\frac{e^A}{1+e^A} - \frac{1}{2} \right) \\ &= \frac{\lambda}{2n} s(s-1) \tanh(A) + \frac{\lambda}{n} s(n-s) \tanh\left(\frac{A}{2}\right) \\ &\geq \frac{\lambda}{2n} \tanh\left(\frac{A}{2}\right) [s(s-1) + s(n-s)] \\ &= \frac{\lambda s(n-1)}{2n} \tanh\left(\frac{A}{2}\right) \geq \frac{\lambda s}{4} \tanh\left(\frac{A}{2}\right). \end{aligned}$$

This immediately gives

$$\begin{aligned} & \mathbb{P}_{\beta,\lambda} \left(\sum_{i=1}^n d_i - \frac{(n-1)\lambda}{2} \leq K_n \right) \\ &= \mathbb{P}_{\beta,\lambda} \left(\sum_{i=1}^n (d_i - \mathbb{E}_{\beta,\lambda} d_i) \leq - \sum_{i=1}^n \mathbb{E}_{\beta,\lambda} d_i + \frac{(n-1)\lambda}{2} + K_n \right) \\ &\leq \mathbb{P}_{\beta,\lambda} \left(\sum_{i=1}^n (d_i - \mathbb{E}_{\beta,\lambda} d_i) \leq -\frac{\lambda s}{4} \tanh\left(\frac{A}{2}\right) + K_n \right) \\ &= \mathbb{P}_{\beta,\lambda} \left(\sum_{i=1}^n (d_i - \mathbb{E}_{\beta,\lambda} d_i) \leq -K_n \right) \leq \frac{2n\lambda}{K_n^2}, \end{aligned}$$

where the last step uses Chebyshev’s inequality Lemma 6.3. This converges to 0 as before, thus verifying (6.3). \square

PROOF OF THEOREM 3.1(ii). Recall the subparameter space $\tilde{\Xi}(s, A_n)$ from (6.1). Let $\pi(d\beta)$ be a prior on $\tilde{\Xi}(s, A_n)$, which puts mass $\frac{1}{\binom{n}{s}}$ on each of the configurations in $\tilde{\Xi}(s, A)$, and let $\mathbb{Q}_\pi(\cdot) := \int \mathbb{P}_{\beta,\lambda}(\cdot) \pi(d\beta)$ denote the marginal distribution of \mathbf{Y} where

$$\mathbf{Y} | \beta \sim \mathbb{P}_{\beta,\lambda}, \quad \beta \sim \pi.$$

Then, setting

$$L_\pi(\mathbf{Y}) := \frac{\mathbb{Q}_\pi(\mathbf{Y})}{\mathbb{P}_{\beta=0,\lambda}(\mathbf{Y})}$$

denote the likelihood ratio, it suffices to show that [Ingster and Suslina (2003)]

$$(6.4) \quad \lim_{n \rightarrow \infty} \mathbb{E}_{\beta=0,\lambda} L_\pi^2 = 1.$$

To this effect, by a direct calculation we have

$$(6.5) \quad \mathbb{E}_{\beta=0,\lambda} L_\pi^2 = \frac{1}{\binom{n}{s}^2} \sum_{S_1, S_2 \subset [n]: |S_1|=|S_2|=s} \prod_{1 \leq i < j \leq n} T_{S_1, S_2}^{ij}(A),$$

where setting $f(A) := \frac{e^A}{1+e^A}$ we define $T_{S_1, S_2}^{ij}(A)$ via the following case by case definition:

- $i, j \in S_1 \cap S_2$.

In this case, set

$$T_{S_1, S_2}^{ij}(A) := \frac{2\lambda f(2A)^2}{n} + \frac{(1 - \frac{\lambda f(2A)}{n})^2}{1 - \frac{\lambda}{2n}}.$$

There are $\binom{Z}{2}$ such terms, where $Z = |S_1 \cap S_2|$.

- $i \in S_1 \cap S_2, j \in (S_1 - S_2) \cup (S_2 - S_1)$ or vice versa.

In this case, set

$$T_{S_1, S_2}^{ij}(A) := \frac{2\lambda f(2A)f(A)}{n} + \frac{(1 - \frac{\lambda f(2A)}{n})(1 - \frac{\lambda f(A)}{n})}{1 - \frac{\lambda}{2n}}.$$

There are $2Z(s - Z)$ such terms.

- $i \in S_1 - S_2, j \in S_2 - S_1$ or vice versa.

In this case, set

$$T_{S_1, S_2}^{ij}(A) := \frac{2\lambda f(A)^2}{n} + \frac{(1 - \frac{\lambda f(A)}{n})^2}{1 - \frac{\lambda}{2n}}.$$

There are $(s - Z)^2$ such terms.

- $i \in S_1 \cap S_2, j \in (S_1 \cup S_2)^c$ or vice versa.

In this case, set

$$T_{S_1, S_2}^{ij}(A) := \frac{2\lambda f(A)^2}{n} + \frac{(1 - \frac{\lambda f(A)}{n})^2}{1 - \frac{\lambda}{2n}}.$$

There are $Z(n - 2s + Z)$ such terms.

- For all other cases, set $T_{S_1, S_2}^{ij}(A) := 1$.

Since $\alpha \geq \frac{1}{2}$, (3.2) implies

$$(6.6) \quad \lim_{n \rightarrow \infty} A\sqrt{\lambda} = 0.$$

Using the fact that $f(0) = \frac{1}{2}, f'(0) = \frac{1}{4}$ and all derivatives of f are uniformly bounded, a two-term Taylor expansion around $A = 0$ gives the uniform bound

$$T_{S_1, S_2}^{ij}(A) \leq 1 + CA^2 \frac{\lambda}{n} \quad \forall A > 0,$$

where C is a finite positive constant. Thus, setting

$$N(z) := \{(S_1, S_2) \subset [n] : |S_1| = |S_2| = s, |S_1 \cap S_2| = z\}$$

using (6.5) we have

$$\begin{aligned} & \mathbb{E}_{\beta=0, \lambda} L_{\pi}^2 \\ & \leq \frac{1}{\binom{n}{2}^2} \sum_{z=0}^s \sum_{(S_1, S_2) \in N(z)} \left(1 + CA^2 \frac{\lambda}{n}\right)^{\binom{s}{2} + 2z(s-z) + (s-z)^2 + z(n-2s+z)} |N(z)| \\ & \leq \frac{1}{\binom{n}{2}^2} \sum_{z=0}^s \sum_{(S_1, S_2) \in N(z)} \exp \left\{ CA^2 \frac{\lambda}{n} \left(\frac{z^2}{2} + s^2 + z(n-2s) \right) \right\} |N(z)| \\ & \leq e^{\frac{3CA^2 s^2 \lambda}{2n}} \mathbb{E}_Z e^{CA^2 \lambda \frac{n-2s}{n} Z}, \end{aligned}$$

where Z is a hypergeometric distribution with parameters (n, s, s) and \mathbb{E}_Z refers to the expectation w.r.t. Z . If $2s \geq n$, then we have $\mathbb{E}_Z e^{CA^2\lambda \frac{n-2s}{n} Z} \leq 1$, giving

$$\mathbb{E}_{\beta=0,\lambda} L_\pi^2 \leq e^{\frac{3CA^2s^2\lambda}{2n}},$$

the right-hand side of which converges to 1 on using (3.2). Thus assume without loss of generality that $2s < n$, in which case Z is stochastically dominated by a binomial distribution with parameters $(s, \frac{s}{n-s})$, which gives

$$\begin{aligned} \mathbb{E} e^{CA^2\lambda \frac{n-2s}{n} Z} &\leq \mathbb{E}_Z e^{CA^2\lambda Z} \leq \left[1 + \frac{s}{n-s} (e^{CA^2\lambda} - 1) \right]^s \\ &\leq \exp \left\{ \frac{2s^2}{n} (e^{CA^2\lambda} - 1) \right\}. \end{aligned}$$

Combining, we have the bound

$$\mathbb{E}_{\beta=0,\lambda} L_\pi^2 \leq \exp \left\{ \frac{3CA^2s^2\lambda}{2n} + \frac{2s^2}{n} (e^{CA^2\lambda} - 1) \right\},$$

the right-hand side of which converges to 1 on using (6.6) along with (3.2). \square

6.3. *Proof of Theorem 3.2.* Throughout the proof, we drop the suffix of λ_n, s_n, A_n , noting that they all depend on n .

PROOF OF THEOREM 3.2(i). We proceed exactly as in the proof of the lower bound in Theorem 3.1. We use the uniform prior π on $\tilde{\Xi}(s, A)$. Then, setting $L_\pi(\mathbf{Y})$ to denote the likelihood ratio, we shall establish that

$$\lim_{n \rightarrow \infty} \mathbb{E}_{\beta=0,\lambda} L_\pi^2 = 1.$$

Similar to the proof for Theorem 3.1, we express

$$(6.7) \quad \mathbb{E}_{\beta=0,\lambda} L_\pi^2 = \frac{1}{\binom{n}{s}^2} \sum_{S_1, S_2 \subset [n]: |S_1|=|S_2|=s} \prod_{1 \leq i < j \leq n} T_{S_1, S_2}^{ij}(A),$$

where T_{S_1, S_2}^{ij} are exactly the same as the earlier definition. As before, we set $f(x) = e^x / (1 + e^x)$. For $0 \leq \mu \leq 1$, and constants $c_1, c_2 > 0$, we consider the function:

$$h(x) = 4\mu f(c_1x) f(c_2x) + \frac{(1 - 2\mu f(c_1x))(1 - 2\mu f(c_2x))}{1 - \mu}.$$

Direct computation yields

$$(6.8) \quad \begin{aligned} h'(x) &= \frac{4\mu}{1 - \mu} [c_1 f'(c_1x) f(c_2x) + c_2 f(c_1x) f'(c_2x)] \\ &\quad - \frac{2\mu}{1 - \mu} [c_1 f'(c_1x) + c_2 f'(c_2x)] \\ &\geq 0, \end{aligned}$$

since $f(x) \geq 1/2$ for $x \geq 0$. Next, we observe that $f(x) \uparrow 1$ as $x \rightarrow \infty$. Thus

$$h(x) \leq 4\mu + \frac{(1 - 2\mu)^2}{1 - \mu}.$$

The expressions derived for $T_{S_1, S_2}^{ij}(A)$ in the proof of the lower bound in Theorem 3.1 imply that each term is exactly of the form h , with $\mu = \lambda/2n$ and appropriate c_1, c_2 . Thus, using the upper bound on h , we have

$$\begin{aligned} \mathbb{E}_{\beta=0, \lambda} L_{\pi}^2 &\leq \mathbb{E}_Z \left[\left(\frac{2\lambda}{n} + \frac{(1 - \lambda/n)^2}{1 - \lambda/2n} \right)^{\binom{Z}{2} + 2Z(s-Z) + Z(n-2s+Z) + (s-Z)^2} \right] \\ (6.9) \qquad &\leq \mathbb{E}_Z \left[\left(\frac{2\lambda}{n} + \frac{(1 - \lambda/n)^2}{1 - \lambda/2n} \right)^{7s^2/2 + nZ} \right], \end{aligned}$$

where $\mathbb{E}_Z[\cdot]$ denotes expectation with respect to Z and Z has a hypergeometric distribution with parameters (n, s, s) , respectively. Next, we note that

$$\frac{2\lambda}{n} + \frac{(1 - \lambda/n)^2}{1 - \lambda/2n} = 1 + \frac{\lambda}{2n} + \frac{(\frac{\lambda}{2n})^2}{1 - \frac{\lambda}{2n}} \leq 1 + \frac{\lambda}{n}$$

for n sufficiently large as $\lambda \ll \log n$. Thus plugging the bounds back into (6.9), we have, for n sufficiently large,

$$(6.10) \qquad \mathbb{E}_{\beta=0, \lambda} L_{\pi}^2 \leq \left(1 + \frac{\lambda}{n}\right)^{7s^2/2} \mathbb{E}_Z \left[\left(1 + \frac{\lambda}{n}\right)^{nZ} \right].$$

We first observe that $(1 + \lambda/n)^{7s^2/2} \leq \exp(\frac{7s^2\lambda}{2n}) \rightarrow 1$ as $n \rightarrow \infty$, using $s = n^{1-\alpha}$ for some $\alpha > 1/2$ and $\lambda \ll \log n$. To bound the second term in (6.10), we note that $(1 + \lambda/n) > 1$ and, therefore,

$$(6.11) \qquad \mathbb{E}_Z \left[\left(1 + \frac{\lambda}{n}\right)^{nZ} \right] \leq \mathbb{E}_U \left[\left(1 + \frac{\lambda}{n}\right)^{nU} \right],$$

where $U \sim \text{Bin}(s, \frac{s}{n-s})$ and \mathbb{E}_U denotes the expectation with respect to U . Finally, we have

$$\begin{aligned} &\mathbb{E}_U \left[\left(1 + \frac{\lambda}{n}\right)^{nU} \right] \\ &= \left[1 + \frac{s}{n-s} \left(\left(1 + \frac{\lambda}{n}\right)^n - 1 \right) \right]^s \leq \exp\left(\frac{s^2}{n-s} \left(\left(1 + \frac{\lambda}{n}\right)^n - 1 \right) \right) \\ &\leq \exp\left(\frac{s^2}{n-s} e^{\lambda} \right) \leq \exp\left(\frac{s^2}{n-s} e^{c \log n} \right) \leq \exp(n^{1-2\alpha+c}) \end{aligned}$$

for any constant $c > 0$, arbitrarily small and n sufficiently large. Thus for any $\alpha > 1/2$, we can choose c sufficiently small such that $\alpha > 1/2 + c$. This implies

that $\mathbb{E}_U[(1 + \lambda/n)^{nU}] \rightarrow 1$ as $n \rightarrow \infty$. Using (6.11) and plugging this bound back into (6.10) gives the desired conclusion. \square

PROOF OF THEOREM 3.2(ii)(a). Recall the version of the higher criticism test introduced in Section 2. We will reject the null hypothesis if $HC > \sqrt{\log n}$. By virtue of centering and scaling of individual $HC(t)$ under the null, we have by union bound and Chebyshev’s inequality:

$$\begin{aligned} & \mathbb{P}_{\beta=0,\lambda}(HC \geq \sqrt{\log n}) \\ & \leq \sum_t \mathbb{P}_{\beta=0,\lambda}(GHC(t) > \sqrt{\log n}) \leq \frac{\sqrt{10 \log n}}{\log n} \rightarrow 0 \quad \text{as } n \rightarrow \infty. \end{aligned}$$

This controls the Type I error of this test. It remains to control the Type II error. We will establish as usual that the noncentrality parameter under the alternative beats the null and the alternative variances of the statistic. We consider alternatives as follows. Let $\mathbb{P}_{\beta,\lambda}$ be such that $\beta_i = A$ for $i \in S$ and $\beta_i = 0$ otherwise, where $A = \sqrt{C^* \frac{\log n}{\lambda}}$ with $16(1 - \theta) \geq C^* > C_{\text{sparse}}(\alpha)$, $\theta = \lim \frac{\lambda}{2n}$, $|S| = s = n^{1-\alpha}$, $\alpha \in (1/2, 1)$. The case of higher signals can be handled by standard monotonicity arguments and are therefore omitted. The following lemma studies the behavior of this statistic under this class of alternatives.

LEMMA 6.4. Let $t = \lfloor \sqrt{2r \log n} \rfloor$ with $r = \min\{1, \frac{C^*}{4(1-\theta)}\}$. Then:

- (a) $\mathbb{E}_{\beta,\lambda}(GHC(t)) \gg \sqrt{\log n}$.
- (b) $\mathbb{E}_{\beta,\lambda}^2(GHC(t)) \gg \text{Var}_{\beta,\lambda}(GHC(t))$.

The Type II error of the HC statistic may be controlled immediately using Lemma 6.4. This is straightforward; however, we include a proof for the sake of completeness. For any alternative considered above, we have, using Chebychev’s inequality and Lemma 6.4,

$$\begin{aligned} & \mathbb{P}_{\beta,\lambda}[HC > \sqrt{\log n}] \\ & \geq \mathbb{P}_{\beta,\lambda}[GHC(t) \geq \sqrt{\log n}] \geq 1 - \frac{\text{Var}_{\beta,\lambda}(GHC(t))}{(\mathbb{E}_{\beta,\lambda}(GHC(t)) - \sqrt{\log n})^2} \rightarrow 1 \end{aligned}$$

as $n \rightarrow \infty$. This completes the proof, modulo that of Lemma 6.4. \square

We describe the proof of Lemma 6.4 next. This necessitates a detailed understanding of the mean and variance of the $HC(t)$ statistics introduced in Section 2. Due to centering, $HC(t)$ has mean 0 under the null hypothesis. Our next proposition estimates the variances of the $HC(t)$ statistics under the null and the class of alternatives introduced above. We also lower bound the expectation of the $HC(t)$ statistics under the alternative. This is the most technical result of this paper and we defer the proof to the Appendix C in Mukherjee, Mukherjee and Sen (2018).

PROPOSITION 6.5. Fix $\theta = \lim_{n \rightarrow \infty} \frac{\lambda}{2n}$. For $t = \lfloor \sqrt{2r \log n} \rfloor$ with $r > \frac{C^*}{16(1-\theta)}$, we have

$$(6.12) \quad \lim_{n \rightarrow \infty} \frac{\log \text{Var}_{\beta=0, \lambda}(HC(t))}{\log n} = 1 - r,$$

$$(6.13) \quad \lim_{n \rightarrow \infty} \frac{\log \mathbb{E}_{\beta, \lambda}(HC(t))}{\log n} \geq 1 - \alpha - \frac{1}{2} \left(\sqrt{2r} - \sqrt{\frac{C^*}{8(1-\theta)}} \right)^2,$$

$$(6.14) \quad \lim_{n \rightarrow \infty} \frac{\log \text{Var}_{\beta, \lambda}(HC(t))}{\log n} = \max \left\{ 1 - \alpha - \frac{1}{2} \left(\sqrt{2r} - \sqrt{\frac{C^*}{8(1-\theta)}} \right)^2, 1 - r \right\}.$$

PROOF OF LEMMA 6.4. We first look at the proof of Part (a). Using (6.12) and (6.13), we have

$$\begin{aligned} \mathbb{E}_{\beta, \lambda}(GHC(t)) &\geq n^{(f_0(r)+o(1))}, \\ f_0(r) &= \frac{1}{2} - \alpha - \frac{r}{2} - \frac{C^*}{16(1-\theta)} + \sqrt{\frac{rC^*}{4(1-\theta)}}. \end{aligned}$$

Thus it suffices to show that $f_0(r) > 0$ for $r = \min\{1, \frac{C^*}{4(1-\theta)}\}$. Now, $C^* > C_{\text{sparse}}(\alpha)$ implies that $C^* > 4(1-\theta)$ for $\alpha \geq 3/4$. Thus in this case, $r = 1$. Therefore, we have, for $\alpha \geq 3/4$,

$$f_0(r) = f_0(1) = 1 - \alpha - \left(1 - \sqrt{\frac{C^*}{16(1-\theta)}} \right)^2 > 0,$$

as $16(1-\theta) \geq C^* > C_{\text{sparse}}(\alpha)$. For $\alpha < 3/4$, we consider two cases. First, consider the case where $C^* > 4(1-\theta)$. As $\alpha < 3/4$, this implies that $C^* > 16(1-\theta)(1-\sqrt{1-\alpha})^2$. In this case, $r = 1$ and the same argument outlined above still goes through. Finally, consider the case $C^* < 4(1-\theta)$. In this case, $r = \frac{C^*}{4(1-\theta)} \leq 1$. We have

$$f_0(r) = f_0\left(\frac{C^*}{4(1-\theta)}\right) = 1/2 - \alpha + \frac{C^*}{16(1-\theta)} > 0 \quad \text{since } C^* > C_{\text{sparse}}(\alpha).$$

Next, we turn to the proof of Part (b). We use (6.13) and (6.14) to get

$$\frac{\mathbb{E}_{\beta, \lambda}^2(GHC(t))}{\text{Var}_{\beta, \lambda}(GHC(t))} = \frac{\mathbb{E}_{\beta, \lambda}^2(HC(t))}{\text{Var}_{\beta, \lambda}(HC(t))} \geq \min\{n^{f_1(r)+o(1)}, n^{f_2(r)+o(1)}\},$$

$$f_1(r) = 1 - \alpha - \frac{1}{2} \left(\sqrt{2r} - \sqrt{\frac{C^*}{8(1-\theta)}} \right)^2,$$

$$f_2(r) = 1 - 2\alpha - \left(\sqrt{2r} - \sqrt{\frac{C^*}{8(1-\theta)}} \right)^2 + r.$$

Therefore, it suffices to show that both $f_1(r)$ and $f_2(r)$ are strictly positive for $r = \min\{1, \frac{C^*}{4(1-\theta)}\}$. To this end, we again note, as in Part (a), that $r = 1$ for $\alpha \geq 3/4$. In this case,

$$f_1(r) = f_1(1) = 1 - \alpha - \left(1 - \sqrt{\frac{C^*}{16(1-\theta)}}\right)^2 > 0$$

as $C_{\text{sparse}}(\alpha) < C^* \leq 16(1-\theta)$. Also, $f_2(r) = f_2(1) = 2f_1(1) > 0$.

It remains to establish the desired proposition for $\alpha < 3/4$. Again, we split the argument into two cases. If $C^* > 4(1-\theta)$, $\alpha < 3/4$ implies that $C^* > 16(1-\theta) \times (1 - \sqrt{1-\alpha})^2$. In this case, $r = 1$ and the argument outlined above still goes through. Finally, we consider the case $C^* < 4(1-\theta)$. In this case, $r = \frac{C^*}{4(1-\theta)}$. We have

$$f_2(r) = f_2\left(\frac{C^*}{4(1-\theta)}\right) = \frac{1}{8(1-\theta)} \left(C^* - 16(1-\theta)\left(\alpha - \frac{1}{2}\right)\right) > 0.$$

Finally, we prove that

$$f_1(r) = f_1\left(\frac{C^*}{4(1-\theta)}\right) = 1 - \alpha - \frac{C^*}{16(1-\theta)} > 0.$$

The validity of the above display is established by contradiction. Suppose, if possible, $f_1(r) \leq 0$. Then one must have $C^* \geq 16(1-\theta)(1-\alpha) \geq 4(1-\theta)$ as $\alpha \leq 3/4$. This is a contradiction to the assumption that $C^* < 4(1-\theta)$. This completes the proof. \square

PROOF OF THEOREM 3.2(ii)(b). Recall the definition of the subparameter space $\tilde{\Xi}(s, A_n)$ from the proof of Theorem 3.1. As in the proof of Theorem 3.1, let $\pi(d\beta)$ be a prior on $\tilde{\Xi}(s, A_n)$, which puts mass $\frac{1}{\binom{n}{s}}$ on each of the configurations in $\tilde{\Xi}(s, A)$, where $A = \sqrt{C^* \frac{\log n}{\lambda}}$ with $C^* < C_{\text{sparse}}(\alpha)$, $\theta = \lim \frac{\lambda}{2n}$, $s = n^{1-\alpha}$, $\alpha \in (1/2, 1)$. Let $\mathbb{Q}_\pi(\cdot) := \int \mathbb{P}_{\beta, \lambda}(\cdot) \pi(d\beta)$ denote the marginal distribution of \mathbf{Y} where

$$\mathbf{Y} | \beta \sim \mathbb{P}_{\beta, \lambda}, \quad \beta \sim \pi,$$

and let

$$L_\pi(\mathbf{Y}) := \frac{\mathbb{Q}_\pi(\mathbf{Y})}{\mathbb{P}_{\beta=0, \lambda}(\mathbf{Y})}$$

denote the likelihood ratio. We introduce some notation for ease of exposition. For any function $h : 2^{[n]} \rightarrow \mathbb{R}$, we denote

$$\mathbb{E}_S h(S) = \frac{1}{\binom{n}{s}} \sum_{\substack{S \subseteq [n], \\ |S|=s}} h(S).$$

Further, we set $\Sigma(S, S) = \sum_{i < j, i, j \in S} Y_{ij}$ and $\Sigma(S, S^c) = \sum_{i \in S, j \in S^c} Y_{ij}$. For $\beta \in \tilde{\Xi}(s, A)$, $S := \text{Support}(\beta) \subseteq [n]$, we define

$$\begin{aligned} L_S &:= \prod_{i < j} \left(\frac{2e^{\beta_i + \beta_j}}{1 + e^{\beta_i + \beta_j}} \right)^{Y_{ij}} \left(\frac{1 - \frac{\lambda}{n} \frac{e^{\beta_i + \beta_j}}{1 + e^{\beta_i + \beta_j}}}{1 - \frac{\lambda}{2n}} \right)^{1 - Y_{ij}} \\ &= \left(\frac{2e^{2A}}{1 + e^{2A}} \right)^{\Sigma(S, S)} \left(\frac{1 - \frac{\lambda}{n} \frac{e^{2A}}{1 + e^{2A}}}{1 - \frac{\lambda}{2n}} \right)^{\binom{s}{2} - \Sigma(S, S)} \\ &\quad \times \left(\frac{2e^A}{1 + e^A} \right)^{\Sigma(S, S^c)} \left(\frac{1 - \frac{\lambda}{n} \frac{e^A}{1 + e^A}}{1 - \frac{\lambda}{2n}} \right)^{s(n-s) - \Sigma(S, S^c)}. \end{aligned}$$

Armed with this notation, we note that $L_\pi(\mathbf{Y}) = \mathbb{E}_S(L_S)$. Next, for $S \subseteq [n]$ and $i \in S$, we define the event

$$\Gamma_{S,i} := \left\{ \frac{\sum_{j \in S^c} Y_{ij} - (n-s)\frac{\lambda}{2n}}{\sqrt{(n-s)\frac{\lambda}{2n}\left(1 - \frac{\lambda}{2n}\right)}} \leq \sqrt{2 \log n} \right\}.$$

Subsequently, for any $S \subseteq [n]$, we introduce the truncation event,

$$\Gamma_S := \bigcap_{i \in S} \Gamma_{S,i}.$$

Let $\tilde{L} := \frac{1}{\binom{n}{s}} \sum_{S \subseteq [n], |S|=s} L_S \mathbf{1}_{\Gamma_S} = \mathbb{E}_S L_S \mathbf{1}_{\Gamma_S}$. The thesis follows provided we establish the following [see, e.g., [Arias-Castro and Verzelen \(2013\)](#)]:

(6.15)
$$\mathbb{E}_{\beta=0, \lambda}(\tilde{L}) = 1 + o(1),$$

(6.16)
$$\mathbb{E}_{\beta=0, \lambda}(\tilde{L})^2 = 1 + o(1).$$

(6.15) is established in Section 6.4 while (6.16) is established in Section 6.5. This completes the proof.

6.4. Proof of (6.15).

$\mathbb{E}_{\beta=0, \lambda}(\tilde{L}) = \mathbb{E}_{\beta=0, \lambda} \mathbb{E}_S L_S \mathbf{1}_{\Gamma_S} = \mathbb{E}_S \mathbb{E}_{\beta=0, \lambda} L_S \mathbf{1}_{\Gamma_S} = 1 - \mathbb{E}_S \mathbb{E}_{\beta=0, \lambda} L_S \mathbf{1}_{\Gamma_S^c}$. Therefore, it suffices to show that $\mathbb{E}_S \mathbb{E}_{\beta=0, \lambda} L_S \mathbf{1}_{\Gamma_S^c} = o(1)$. Now,

(6.17)
$$\mathbb{E}_{\beta=0, \lambda} L_S \mathbf{1}_{\Gamma_S^c} \leq \sum_{i \in S} \mathbb{E}_{\beta=0, \lambda} L_S \mathbf{1}_{\Gamma_{S,i}^c}$$

and

$$\begin{aligned} \mathbb{E}_{\beta=0, \lambda} L_S \mathbf{1}_{\Gamma_{S,i}^c} &= \mathbb{E}_{\beta=0, \lambda} \left(\left(\frac{2e^A}{1 + e^A} \right)^{\sum_{j \in S^c} Y_{ij}} \left(\frac{1 - \frac{\lambda}{n} \frac{e^A}{1 + e^A}}{1 - \frac{\lambda}{2n}} \right)^{(n-s) - \sum_{j \in S^c} Y_{ij}} \mathbf{1}_{\Gamma_{S,i}^c} \right) \\ &= \mathbb{P} \left(\frac{\text{Bin}(n-s, \frac{\lambda}{n} \frac{e^A}{1 + e^A}) - \frac{(n-s)\lambda}{2n}}{\sqrt{(n-s)\frac{\lambda}{2n}\left(1 - \frac{\lambda}{2n}\right)}} > \sqrt{2 \log n} \right). \end{aligned}$$

The last equality in the above display follows by noting that

$$\left(\frac{2e^A}{1+e^A}\right)^{\sum_{j \in S^c} Y_{ij}} \left(\frac{1 - \frac{\lambda}{n} \frac{e^A}{1+e^A}}{1 - \frac{\lambda}{2n}}\right)^{(n-s) - \sum_{j \in S^c} Y_{ij}}$$

is the Radon–Nikodym derivative of $\text{Bin}(n - s, \frac{\lambda}{n} \frac{e^A}{1+e^A})$ with respect to $\text{Bin}(n - s, \frac{\lambda}{2n})$, which is in turn the distribution of $\sum_{j \in S^c} Y_{ij}$ under $\mathbb{P}_{\beta=0, \lambda}$. Denoting $W \sim \text{Bin}(n - s, \frac{\lambda}{n} \frac{e^A}{1+e^A})$, we have

$$\mathbb{E}_{\beta=0, \lambda} L_S \mathbf{1}_{\Gamma_{S,i}^c} = \mathbb{P}\left(\frac{W - \frac{(n-s)\lambda}{n} \frac{e^A}{1+e^A}}{\sqrt{(n-s)\frac{\lambda}{n} \frac{e^A}{1+e^A} (1 - \frac{\lambda}{n} \frac{e^A}{1+e^A})}} > \eta_n\right),$$

where

$$\eta_n = \frac{(n-s)\frac{\lambda}{n} (\frac{1}{2} - \frac{e^A}{1+e^A}) + \sqrt{2 \log n (n-s)\frac{\lambda}{2n} (1 - \frac{\lambda}{2n})}}{\sqrt{(n-s)\frac{\lambda}{n} \frac{e^A}{1+e^A} (1 - \frac{\lambda}{n} \frac{e^A}{1+e^A})}}.$$

Now noting that $\frac{e^A}{1+e^A} = \frac{1}{2} + \frac{A}{4} + O(A^2)$, it is easy to see that

$$\eta_n = \sqrt{2 \log n} \left(1 - \sqrt{\frac{C^*}{16(1-\theta)}} + a_n\right),$$

where $a_n \rightarrow 0$ as $n \rightarrow \infty$, and the sequence is independent of specific $S \subseteq [n]$ with $|S| = s$. Since $\lambda \gg \log n$, we have using part Lemma 6.2, Part (b)(ii),

$$\begin{aligned} & \mathbb{E}_{\beta=0, \lambda} L_S \mathbf{1}_{\Gamma_{S,i}^c} \\ (6.18) \quad &= \mathbb{P}\left(\frac{W - \frac{(n-s)\lambda}{n} \frac{e^A}{1+e^A}}{\sqrt{(n-s)\frac{\lambda}{n} \frac{e^A}{1+e^A} (1 - \frac{\lambda}{n} \frac{e^A}{1+e^A})}} > \sqrt{2 \log n} \left(1 - \sqrt{\frac{C^*}{16(1-\theta)}} + a_n\right)\right) \\ & \leq n^{-(1 - \sqrt{\frac{C^*}{16(1-\theta)}})^2 + o(1)}. \end{aligned}$$

Therefore, combining (6.17) and (6.18) along with the independence of a_n from specific S 's, we have

$$\begin{aligned} \mathbb{E}_S \mathbb{E}_{\beta=0, \lambda} L_S \mathbf{1}_{\Gamma_S^c} & \leq \mathbb{E}_S n^{1-\alpha - (1 - \sqrt{\frac{C^*}{16(1-\theta)}})^2 + o(1)} \\ & = n^{1-\alpha - (1 - \sqrt{\frac{C^*}{16(1-\theta)}})^2 + o(1)} = o(1), \end{aligned}$$

since $1 - \alpha - (1 - \sqrt{\frac{C^*}{16(1-\theta)}})^2 < 0$ whenever $C^* < C_{\text{sparse}}(\alpha)$. This completes the verification of (6.15).

6.5. *Proof of (6.16).* For any function $h : 2^{[n]} \times 2^{[n]} \rightarrow \mathbb{R}$, define

$$\mathbb{E}_{S_1, S_2} h(S_1, S_2) = \frac{1}{\binom{n}{s}^2} \sum_{\substack{(S_1, S_2) \subseteq [n] \times [n], \\ |S_1| = |S_2| = s}} h(S_1, S_2).$$

This allows us to express the second moment of the truncated likelihood ratio as follows:

$$\begin{aligned} \mathbb{E}_{\beta=0, \lambda} (\tilde{L})^2 &= \mathbb{E}_{\beta=0, \lambda} (\mathbb{E}_S L_S \mathbf{1}_{\Gamma_S})^2 \\ &= \mathbb{E}_{S_1, S_2} (\mathbb{E}_{\beta=0, \lambda} (L_{S_1} L_{S_2} \mathbf{1}_{\Gamma_{S_1} \cap \Gamma_{S_2}})). \end{aligned}$$

Before proceeding, we introduce some notation. For any set $T \subseteq [n]$, the set $\{(i, j) \in T\}$ denotes all pairs $i < j$ such that both i and j are in T . Recall that for any $T \subseteq [n]$, $\Sigma(T, T) = \sum_{(i, j) \in T} Y_{ij}$ while for any two sets $S, T \subseteq [n]$, $\Sigma(S, T) = \sum_{i \in S, j \in T} Y_{ij}$. For ease of notation, we will use $\Sigma(S) = \Sigma(S, S)$ if there is no scope for confusion. For S_1, S_2 subsets of $[n]$, let $Z = |S_1 \cap S_2|$. Now if $S_1 := \text{Support}(\beta) \subseteq [n]$ and $S_2 := \text{Support}(\beta') \subseteq [n]$ for $\beta, \beta' \in \tilde{\Xi}(s, A)$, we have

$$\begin{aligned} (6.19) \quad L_{S_1} L_{S_2} &= \prod_{i < j} \left(\frac{4e^{\beta_i + \beta_j + \beta'_i + \beta'_j}}{(1 + e^{\beta_i + \beta_j})(1 + e^{\beta'_i + \beta'_j})} \right)^{Y_{ij}} \\ &\times \left(\left(\frac{1 - \frac{\lambda}{n} e^{\beta_i + \beta_j}}{1 - \frac{\lambda}{2n}} \right) \left(\frac{1 - \frac{\lambda}{n} e^{\beta'_i + \beta'_j}}{1 - \frac{\lambda}{2n}} \right) \right)^{1 - Y_{ij}}. \end{aligned}$$

Further, for $i \in S_1 \cap S_2$, define

$$(6.20) \quad C_i = \left\{ \sum_{j \in (S_1 \cup S_2)^c} Y_{ij} \leq (n-s) \frac{\lambda}{2n} + \sqrt{2(n-s) \log n \frac{\lambda}{2n} \left(1 - \frac{\lambda}{2n}\right)} \right\}.$$

Finally, we set $\mathcal{C} := \bigcap_{i \in S_1 \cap S_2} C_i$. Thus, we have

$$\begin{aligned} &\Gamma_{S_1} \cap \Gamma_{S_2} \\ &= \left\{ \begin{array}{l} \sum_{j \in S_1^c} Y_{ij} \leq (n-s) \frac{\lambda}{2n} + \sqrt{2 \log n} \sqrt{(n-s) \frac{\lambda}{2n} \left(1 - \frac{\lambda}{2n}\right)}, \forall i \in S_1, \\ \sum_{j \in S_2^c} Y_{ij} \leq (n-s) \frac{\lambda}{2n} + \sqrt{2 \log n} \sqrt{(n-s) \frac{\lambda}{2n} \left(1 - \frac{\lambda}{2n}\right)}, \forall i \in S_2 \end{array} \right\} \\ &\subseteq \left\{ \sum_{j \in S_1^c \cap S_2^c} Y_{ij} \leq (n-s) \frac{\lambda}{2n} + \sqrt{2 \log n} \sqrt{(n-s) \frac{\lambda}{2n} \left(1 - \frac{\lambda}{2n}\right)}, \forall i \in S_1 \cap S_2 \right\} \\ &= \mathcal{C}. \end{aligned}$$

This implies that $\mathbb{E}_{\beta=0,\lambda}(L_{S_1}L_{S_2}\mathbf{1}_{\Gamma_{S_1}\cap\Gamma_{S_2}}) \leq \mathbb{E}_{\beta=0,\lambda}(L_{S_1}L_{S_2}\mathbf{1}_{\mathcal{C}})$. We note that the event \mathcal{C} is a function of the edges $\{Y_{ij} : i \in S_1 \cap S_2, j \in (S_1 \cup S_2)^c\}$. Further, it is easy to see from (6.19) that

$$(6.21) \quad L_{S_1}L_{S_2} = \Upsilon \left(\frac{4e^{2A}}{(1+e^A)^2} \right)^{\Sigma(S_1 \cap S_2, (S_1 \cup S_2)^c)} \times \left(\frac{1 - \frac{\lambda}{n} \frac{e^A}{1+e^A}}{(1 - \frac{\lambda}{2n})} \right)^{2(Z(n-2s+Z) - \Sigma(S_1 \cap S_2, (S_1 \cup S_2)^c))},$$

where Υ is independent of $\{Y_{ij} : i \in S_1 \cap S_2, j \in (S_1 \cup S_2)^c\}$. We next make the following observation.

LEMMA 6.6. $\mathbb{E}_{\beta=0,\lambda}(\Upsilon) = 1 + o(1)$, uniformly over Z .

The proof is similar to that of the lower bound in Theorem 3.1, and thus will be deferred to Appendix B.1. Therefore, we have, setting $T_i := \sum_{j \in (S_1 \cup S_2)^c} Y_{ij}$ for $i \in S_1 \cap S_2$, using Lemma 6.6 and (6.21) along with the independence of the collections $\{Y_{ij} : j \in (S_1 \cup S_2)^c\}, i \in S_1 \cap S_2$,

$$(6.22) \quad \begin{aligned} &\mathbb{E}_{\beta=0,\lambda}(L_{S_1}L_{S_2}\mathbf{1}_{\Gamma_{S_1}\cap\Gamma_{S_2}}) \\ &\leq \mathbb{E}_{\beta=0,\lambda}(L_{S_1}L_{S_2}\mathbf{1}_{\mathcal{C}}) \\ &= (1 + o(1)) \left\{ \mathbb{E}_{\beta=0,\lambda} \left(\left(\frac{4e^{2A}}{(1+e^A)^2} \right)^{T_i} \left(\frac{1 - \frac{\lambda}{n} \frac{e^A}{1+e^A}}{(1 - \frac{\lambda}{2n})} \right)^{2((n-2s+Z)-T_i)} \mathbf{1}_{\mathcal{C}_i} \right) \right\}^Z \\ &= (1 + o(1)) g(A)^{Z(n-2s+Z)} (\mathbb{E}_{\beta=0,\lambda}(f(A)^{T_i} g(A)^{-T_i} \mathbf{1}_{\mathcal{C}_i}))^Z, \end{aligned}$$

where we define

$$f(x) := \frac{4e^{2x}}{(1+e^x)^2}, \quad \left(1 - \frac{\lambda}{2n}\right)^2 g(x) := \left(1 - \frac{\lambda}{n} \frac{e^x}{1+e^x}\right)^2.$$

Applying Lemma 6.1 with $\psi(A, \lambda, n) = \frac{f(A)}{g(A)} \frac{\lambda}{2n} + 1 - \frac{\lambda}{2n}$, we immediately have

$$(6.23) \quad \begin{aligned} &\mathbb{E}_{\beta=0,\lambda}(f(A)^{T_i} g(A)^{-T_i} \mathbf{1}_{\mathcal{C}_i}) \\ &= \psi(A, \lambda, n)^{n-2s+Z} \\ &\quad \times \mathbb{P}\left(X \leq (n-s) \frac{\lambda}{2n} + \sqrt{2 \log n} \sqrt{(n-s) \frac{\lambda}{2n} \left(1 - \frac{\lambda}{2n}\right)}\right), \end{aligned}$$

where $X \sim \text{Bin}(n - 2s + Z, \frac{f(A)\lambda}{(2n)g(A)\psi(A,\lambda,n)})$. Combining (6.22) and (6.23), we have

$$\begin{aligned}
 & \mathbb{E}_{\beta=0,\lambda}(L_{S_1}L_{S_2}\mathbf{1}_{\Gamma_{S_1}\cap\Gamma_{S_2}}) \\
 (6.24) \quad & \leq (1 + o(1)) \left[\{g(A)\psi(A, \lambda, n)\}^{n-2s+Z} \right. \\
 & \quad \left. \times \mathbb{P}\left(X \leq (n - s)\frac{\lambda}{2n} + \sqrt{2 \log n} \sqrt{(n - s)\frac{\lambda}{2n}\left(1 - \frac{\lambda}{2n}\right)}\right) \right]^Z.
 \end{aligned}$$

We will bound the right-hand side of (6.24) to complete the proof. First, using Taylor’s expansion, we have

$$(6.25) \quad f^*(x) = g(x)\psi(x, \lambda, n) = 1 + \frac{\lambda}{8n} \frac{x^2}{(1 - \frac{\lambda}{2n})} + \frac{\frac{\lambda}{2n}}{(\frac{\lambda}{2n} - 1)} x^3 f_*(\xi(x))$$

for some $|\xi(x)| \leq |x|$ and

$$f_*(\xi) = \frac{2(e^\xi(11e^\xi - 11e^{2\xi} + e^{3\xi} - 1))}{3(e^\xi + 1)^5}.$$

Next, we analyze the success probability for the distribution of X in (6.24):

$$\delta_n := \frac{f(A)\frac{\lambda}{2n}}{g(A)\psi(A, \lambda, n)} = \frac{\frac{2\lambda}{n} \frac{e^{2A}}{(1+e^A)^2}}{f^*(A)} = \frac{\frac{2\lambda}{n} \frac{e^{2A}}{(1+e^A)^2}}{1 + \frac{\lambda A^2}{8n(1-\frac{\lambda}{2n})} + A^3 f_*(\xi(A))}.$$

Moreover, using Taylor’s expansion for f , we have

$$\frac{e^{2x}}{(1 + e^x)^2} = \frac{1}{4} + \frac{x}{4} + \frac{x^2}{16} + x^3 g_*(\zeta(x))$$

for some $|\zeta(x)| \leq |x|$ and

$$g_*(\zeta) = \frac{e^{2\zeta}(e^{2\zeta} - 7e^\zeta + 4)}{3(e^\zeta + 1)^5}.$$

Therefore, combining everything, we have

$$\delta_n = \frac{2\lambda}{n} \cdot \frac{\frac{1}{4} + \frac{A}{4} + \frac{A^2}{16} + A^3 g_*(\zeta(A))}{1 + \frac{\lambda A^2}{8n(1-\frac{\lambda}{2n})} + A^3 f_*(\xi(A))} = \frac{\lambda}{2n} (1 + A(1 + b_n A))$$

for $\{b_n\}_{n \geq 1}$ such that $b_n \leq C$ for sufficiently large n and some universal constant C . Now we make use of the following lemma.

LEMMA 6.7. *There exists a sequence $\kappa_n(Z)$ such that*

$$\limsup_{n \rightarrow \infty} \sup_{Z \leq s} \kappa_n(Z) = 0$$

and

$$\begin{aligned} & \frac{(n-s)\frac{\lambda}{2n} - (n-2s+Z)\delta_n + \sqrt{2\log n}\sqrt{(n-s)\frac{\lambda}{2n}\left(1-\frac{\lambda}{2n}\right)}}{\sqrt{(n-2s+Z)\delta_n(1-\delta_n)}} \\ &= \sqrt{2\log n}\left(1 - \sqrt{\frac{C^*}{4(1-\theta)}} + \kappa_n(Z)\right). \end{aligned}$$

The proof will be deferred to Appendix B. Applying Lemma 6.7 along with Lemma 6.2, Part (a)(ii), we have

$$\begin{aligned} & \mathbb{P}\left(X \leq (n-s)\frac{\lambda}{2n} + \sqrt{2\log n}\sqrt{(n-s)\frac{\lambda}{2n}\left(1-\frac{\lambda}{2n}\right)}\right) \\ &= \mathbb{P}\left(\frac{X - (n-2s+Z)\delta_n}{\sqrt{(n-2s+Z)\delta_n(1-\delta_n)}} \leq \sqrt{2\log n}\left(1 - \sqrt{\frac{C^*}{4(1-\theta)}} + \kappa_n(Z)\right)\right) \\ &\leq \mathbf{1}_{C^* \leq 4(1-\theta)} + \exp\left(-\log n\left(1 - \sqrt{\frac{C^*}{4(1-\theta)}} + \kappa_n(Z)\right)^2\right) \mathbf{1}_{C^* > 4(1-\theta)} \\ &:= p_n(C^*, Z). \end{aligned}$$

Therefore, by (6.23) and (6.25)

$$\mathbb{E}_{\beta=0,\lambda}(L_{S_1}L_{S_2}\mathbf{1}_{\Gamma_{S_1}\cap\Gamma_{S_2}}) \leq (1+o(1))f^*(A)^{Z(n-2s+Z)}(p_n(C^*, Z))^Z.$$

Using (6.25), we note that f_* is uniformly bounded by a universal constant, and thus for sufficiently large n , $f^*(A) \geq 1$. Consequently, n sufficiently large,

$$f^*(A)^{Z(n-2s+Z)} \leq f^*(A)^{nZ}.$$

Hereafter, we divide our analysis into two cases according to the value of $C^*/(4(1-\theta))$.

6.5.1. *Case I: $C^* \leq 4(1-\theta)$.* In this case, $p_n(C^*, Z) = 1$. Therefore, using (6.25), we have

$$\begin{aligned} & \mathbb{E}_{\beta=0,\lambda}(L_{S_1}L_{S_2}\mathbf{1}_{\Gamma_{S_1}\cap\Gamma_{S_2}}) \\ &\leq (1+o(1))(f^*(A))^{nZ} \\ &\leq (1+o(1))\exp\left(Zn\left(\frac{\lambda}{8n}\frac{A^2}{\left(1-\frac{\lambda}{2n}\right)} + \frac{\frac{\lambda}{2n}}{\left(\frac{\lambda}{2n}-1\right)}A^3f_*(\xi(A))\right)\right). \end{aligned}$$

The uniform universal upper bound on f_* implies that for n sufficiently large,

$$f^*(A) - 1 = \frac{\lambda}{8n}\frac{A^2}{\left(1-\frac{\lambda}{2n}\right)} + \frac{\frac{\lambda}{2n}}{\left(\frac{\lambda}{2n}-1\right)}A^3f_*(\xi(A)) \geq 0.$$

Now, we note that $Z \sim \text{Hypergeometric}(n, s, s) \stackrel{d}{\lesssim} W \sim \text{Bin}(s, \frac{s}{n-s})$ where $\stackrel{d}{\lesssim}$ denotes stochastic ordering. As a result, for sufficiently large n

$$\begin{aligned} \mathbb{E}_{S_1, S_2} \mathbb{E}_{\beta=0, \lambda} (L_{S_1} L_{S_2} \mathbf{1}_{\Gamma_{S_1} \cap \Gamma_{S_2}}) &\leq (1 + o(1)) \mathbb{E} \exp(Wn(f^*(A) - 1)) \\ &= (1 + o(1)) \exp(n^{1-2\alpha + \frac{C^*}{8(1-\theta)} + o(1)}). \end{aligned}$$

We note that $1 - 2\alpha + \frac{C^*}{8(1-\theta)} < 0$ if $C^* < 16(1 - \theta)(\alpha - \frac{1}{2})$ and in this case, we have the desired result. This is indeed true when $\alpha \leq \frac{3}{4}$. Therefore, we are left with the region when $\alpha > \frac{3}{4}$ and $16(1 - \theta)(\alpha - \frac{1}{2}) \leq C^* < 16(1 - \theta) \times (1 - \sqrt{1 - \alpha})^2$. In this region, $C^* > 4(1 - \theta)$, which corresponds to our next subsection.

6.5.2. *Case II: $C^* > 4(1 - \theta)$.* As deduced in the last subsection, it remains to consider the case $C^* > 4(1 - \theta)$, which immediately implies $\alpha > \frac{3}{4}$.

In this case, $p_n(C^*, Z) = \exp(-\log n(1 - \sqrt{\frac{C^*}{4(1-\theta)}} + \kappa_n(Z))^2)$. Therefore,

$$\begin{aligned} &\mathbb{E}_{\beta=0, \lambda} (L_{S_1} L_{S_2} \mathbf{1}_{\Gamma_{S_1} \cap \Gamma_{S_2}}) \\ &\leq (1 + o(1))(f^*(A))^{nZ} \exp\left(-Z \log n \left(1 - \sqrt{\frac{C^*}{4(1-\theta)}} + \kappa_n(Z)\right)^2\right) \\ &\leq (1 + o(1)) \exp(Zn(f^*(A) - 1)) \\ &\quad \times \exp\left(-Z \log n \left(1 - \sqrt{\frac{C^*}{4(1-\theta)}} + \kappa_n(Z)\right)^2\right). \end{aligned}$$

We note that $\limsup_{n \rightarrow \infty} \sup_{Z \leq s} \kappa_n(Z) = 0$ by Lemma 6.7, and $\lim_{n \rightarrow \infty} \frac{\lambda}{2n} = \theta$ with $\lambda \gg \log n$. Thus we have, using (6.25), for any $\varepsilon > 0$, there exists an n_ε such that for all $n \geq n_\varepsilon$

$$\mathbb{E}_{S_1, S_2} \mathbb{E}_{\beta=0, \lambda} (L_{S_1} L_{S_2} \mathbf{1}_{\Gamma_{S_1} \cap \Gamma_{S_2}}) \leq (1 + o(1)) \mathbb{E}_{S_1, S_2} \exp\{Z \log n(\psi(C^*) + \varepsilon)\}$$

with $\psi(\cdot)$ defined as follows. For $x > 0$, we set

$$\psi(x) = \frac{x}{8(1-\theta)} - \left(1 - \sqrt{\frac{x}{4(1-\theta)}}\right)^2.$$

We note that, $\psi'(x) > 0$ whenever $0 < x < 16(1 - \theta)$. As a result, for $4(1 - \theta) < C^* < 16(1 - \theta)$,

$$\psi(C^*) \geq \psi(4(1 - \theta)) = \frac{1}{2} > 0.$$

Therefore, for any $\varepsilon > 0$, for all $n \geq n_\varepsilon$, we have

$$\begin{aligned} & \mathbb{E}_{S_1, S_2} \mathbb{E}_{\beta=0, \lambda} (L_{S_1} L_{S_2} \mathbf{1}_{\Gamma_{S_1} \cap \Gamma_{S_2}}) \\ & \leq (1 + o(1)) \mathbb{E} \exp\{Z \log n (\psi(C^*) + \varepsilon)\} \\ & \leq (1 + o(1)) \mathbb{E} \exp\{W \log n (\psi(C^*) + \varepsilon)\} \\ & \leq (1 + o(1)) \exp\{n^{1-2\alpha + \frac{C^*}{8(1-\theta)} - (1 - \sqrt{\frac{C^*}{4(1-\theta)}})^2 + \varepsilon}\}, \end{aligned}$$

where $W \sim \text{Bin}(s, \frac{s}{n-s})$. The second inequality above follows from the fact that $Z \stackrel{d}{\lesssim} W$. Now, we note that

$$\begin{aligned} & 1 - 2\alpha + \frac{C^*}{8(1-\theta)} - \left(1 - \sqrt{\frac{C^*}{4(1-\theta)}}\right)^2 \\ & = 2 \left\{ (1 - \alpha) - \left(1 - \sqrt{\frac{C^*}{16(1-\theta)}}\right)^2 \right\}. \end{aligned}$$

Since $C^* < 16(1-\theta)(1 - \sqrt{1-\alpha})^2$, we have that $(1 - \alpha) < (1 - \sqrt{\frac{C^*}{16(1-\theta)}})^2$.

Therefore, there exists $\delta > 0$ such that $1 - 2\alpha + \frac{C^*}{8(1-\theta)} - (1 - \sqrt{\frac{C^*}{4(1-\theta)}})^2 < -\delta$.

Choosing $\varepsilon = \delta/2$, we have for all $n \geq n_{\delta/2}$

$$n^{1-2\alpha + \frac{C^*}{8(1-\theta)} - (1 - \sqrt{\frac{C^*}{4(1-\theta)}})^2 + \varepsilon} \leq n^{-\delta/2},$$

which therefore completes the proof of (6.16). \square

SUPPLEMENTARY MATERIAL

Supplement to “Detection thresholds for the β model on sparse graphs” (DOI: [10.1214/17-AOS1585SUPP](https://doi.org/10.1214/17-AOS1585SUPP); .pdf). The supplementary material contain the proofs of additional technical results.

REFERENCES

- ADDARIO-BERRY, L., BROUTIN, N., DEVROYE, L. and LUGOSI, G. (2010). On combinatorial testing problems. *Ann. Statist.* **38** 3063–3092. [MR2722464](#)
- ARIAS-CASTRO, E., CANDÈS, E. J. and PLAN, Y. (2011). Global testing under sparse alternatives: ANOVA, multiple comparisons and the higher criticism. *Ann. Statist.* **39** 2533–2556. [MR2906877](#)
- ARIAS-CASTRO, E., DONOHO, D. L. and HUO, X. (2005). Near-optimal detection of geometric objects by fast multiscale methods. *IEEE Trans. Inform. Theory* **51** 2402–2425. [MR2246369](#)
- ARIAS-CASTRO, E. and VERZELEN, N. (2013). Community detection in random networks. Available at [arXiv:1302.7099](https://arxiv.org/abs/1302.7099).
- ARIAS-CASTRO, E. and WANG, M. (2015). The sparse Poisson means model. *Electron. J. Stat.* **9** 2170–2201. [MR3406276](#)
- ARIAS-CASTRO, E., CANDÈS, E. J., HELGASON, H. and ZEITOUNI, O. (2008). Searching for a trail of evidence in a maze. *Ann. Statist.* **36** 1726–1757. [MR2435454](#)

- BARABÁSI, A.-L. and ALBERT, R. (1999). Emergence of scaling in random networks. *Science* **286** 509–512. [MR2091634](#)
- BARNETT, I., MUKHERJEE, R. and LIN, X. (2017). The generalized higher criticism for testing SNP-set effects in genetic association studies. *J. Amer. Statist. Assoc.* **112** 64–76. [MR3646553](#)
- BARVINOK, A. and HARTIGAN, J. A. (2013). The number of graphs and a random graph with a given degree sequence. *Random Structures Algorithms* **42** 301–348. [MR3039682](#)
- BICKEL, P. J., CHEN, A. and LEVINA, E. (2011). The method of moments and degree distributions for network models. *Ann. Statist.* **39** 2280–2301. [MR2906868](#)
- BLITZSTEIN, J. and DIACONIS, P. (2010). A sequential importance sampling algorithm for generating random graphs with prescribed degrees. *Internet Math.* **6** 489–522. [MR2809836](#)
- BOLLOBÁS, B. (2001). *Random Graphs*, 2nd ed. *Cambridge Studies in Advanced Mathematics* **73**. Cambridge Univ. Press, Cambridge. [MR1864966](#)
- CAI, T. T. and YUAN, M. (2014). Rate-optimal detection of very short signal segments. Available at [arXiv:1407.2812](#).
- CHATTERJEE, S., DIACONIS, P. and SLY, A. (2011). Random graphs with a given degree sequence. *Ann. Appl. Probab.* **21** 1400–1435. [MR2857452](#)
- DONOHO, D. and JIN, J. (2004). Higher criticism for detecting sparse heterogeneous mixtures. *Ann. Statist.* **32** 962–994. [MR2065195](#)
- FIENBERG, S. E. and WASSERMAN, S. (1981). Categorical data analysis of single sociometric relations. *Sociol. Method.* **12** 156–192.
- GOODREAU, S. M. (2007). Advances in exponential random graph (p^*) models applied to a large social network. *Soc. Netw.* **29** 231–248.
- HALL, P. and JIN, J. (2010). Innovated higher criticism for detecting sparse signals in correlated noise. *Ann. Statist.* **38** 1686–1732. [MR2662357](#)
- HARA, H. and TAKEMURA, A. (2010). Connecting tables with zero-one entries by a subset of a Markov basis. In *Algebraic Methods in Statistics and Probability II. Contemp. Math.* **516** 199–213. Amer. Math. Soc., Providence, RI. [MR2730750](#)
- HILLAR, C. and WIBISONO, A. (2013). Maximum entropy distributions on graphs. Available at [arXiv:1301.3321](#).
- 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. and SUSLINA, I. A. (2003). *Nonparametric Goodness-of-Fit Testing Under Gaussian Models. Lecture Notes in Statistics* **169**. Springer, New York. [MR1991446](#)
- INGSTER, Y. I., TSYBAKOV, A. B. and VERZELEN, N. (2010). Detection boundary in sparse regression. *Electron. J. Stat.* **4** 1476–1526. [MR2747131](#)
- KARRER, B. and NEWMAN, M. E. J. (2011). Stochastic blockmodels and community structure in networks. *Phys. Rev. E* (3) **83** 016107. [MR2788206](#)
- KARWA, V. and SLAVKOVIĆ, A. (2016). Inference using noisy degrees: Differentially private β -model and synthetic graphs. *Ann. Statist.* **44** 87–112. [MR3449763](#)
- LAURITZEN, S. L. (2002). Rasch models with exchangeable rows and columns. Research Report Series, No. R-02-2005, Dept. Mathematical Sciences, Aalborg Univ.
- LAURITZEN, S. L. (2008). Exchangeable Rasch matrices. *Rend. Mat. Appl.* (7) **28** 83–95. [MR2463441](#)
- MUKHERJEE, R., MUKHERJEE, S. and SEN, S. (2018). Supplement to “Detection thresholds for the β -model on sparse graphs.” DOI:[10.1214/17-AOS1585SUPP](#).
- MUKHERJEE, R., PILLAI, N. S. and LIN, X. (2015). Hypothesis testing for high-dimensional sparse binary regression. *Ann. Statist.* **43** 352–381. [MR3311863](#)
- OGAWA, M., HARA, H. and TAKEMURA, A. (2013). Graver basis for an undirected graph and its application to testing the beta model of random graphs. *Ann. Inst. Statist. Math.* **65** 191–212. [MR3011620](#)

- PERRY, P. O. and WOLFE, P. J. (2012). Null models for network data. Available at arXiv:1201.5871.
- PETROVIĆ, S., RINALDO, A. and FIENBERG, S. E. (2010). Algebraic statistics for a directed random graph model with reciprocation. In *Algebraic Methods in Statistics and Probability II. Contemp. Math.* **516** 261–283. Amer. Math. Soc., Providence, RI. [MR2730754](#)
- RINALDO, A., PETROVIĆ, S. and FIENBERG, S. E. (2013). Maximum likelihood estimation in the β -model. *Ann. Statist.* **41** 1085–1110. [MR3113804](#)
- 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.
- VERZELEN, N. and ARIAS-CASTRO, E. (2015). Community detection in sparse random networks. *Ann. Appl. Probab.* **25** 3465–3510. [MR3404642](#)
- WATTS, D. J. and STROGATZ, S. H. (1998). Collective dynamics of ‘small-world’ networks. *Nature* **393** 440–442.
- YAN, T., QIN, H. and WANG, H. (2016). Asymptotics in undirected random graph models parameterized by the strengths of vertices. *Statist. Sinica* **26** 273–293. [MR3468353](#)
- YAN, T. and XU, J. (2013). A central limit theorem in the β -model for undirected random graphs with a diverging number of vertices. *Biometrika* **100** 519–524. [MR3068452](#)
- YAN, T., ZHAO, Y. and QIN, H. (2015). Asymptotic normality in the maximum entropy models on graphs with an increasing number of parameters. *J. Multivariate Anal.* **133** 61–76. [MR3282018](#)
- YAN, X., SHALIZI, C., JENSEN, J. E., KRZAKALA, F., MOORE, C., ZDEBOROVÁ, L., ZHANG, P. and ZHU, Y. (2014). Model selection for degree-corrected block models. *J. Stat. Mech. Theory Exp.* **2014** P05007.

R. MUKHERJEE
DEPARTMENT OF BIostatISTICS
UNIVERSITY OF CALIFORNIA, BERKELEY
101 HAVILAND HALL
BERKELEY, CALIFORNIA 94720
USA
E-MAIL: rmukherj@berkeley.edu

S. MUKHERJEE
DEPARTMENT OF STATISTICS
COLUMBIA UNIVERSITY
1255 AMSTERDAM AVENUE
NEW YORK, NEW YORK 10027
USA
E-MAIL: sm3949@columbia.edu

S. SEN
MICROSOFT RESEARCH NEW ENGLAND
1 MEMORIAL DRIVE
CAMBRIDGE, MASSACHUSETTS 02142
USA
E-MAIL: subse@microsoft.com