

Optimal Detection of Planted Hubs in Multiplex Networks via a Random Energy Model

Ijay Narang*

Will Perkins[†]

Timothy L. H. Wee[‡]

Abstract

We study the problem of detecting a common planted hub in a multiplex network, formulated as a hypothesis test. The null consists of independent Erdős–Rényi $G(n, m_\ell)$ layers. Under the alternative, random stars of varying sizes k_ℓ are planted around a common latent hub, and background edges are added so that each layer ℓ also contains m_ℓ edges. We identify the critical scaling parameter as $\|\mathbf{a}\|_2$, where the entries of the signal-to-noise ratio vector $\mathbf{a} = (a_\ell)$ are star sizes k_ℓ appropriately scaled by the background degree fluctuations of each layer ℓ . We show that the total variation distance between the null and alternative distributions converges to $\Phi(h)$ when $\|\mathbf{a}\|_2 - \sqrt{2\log n} \rightarrow h$, where Φ is the standard Gaussian cumulative distribution function. Furthermore, we exhibit a computationally efficient test that projects each vertex's degree vector onto the direction of $\mathbf{a}$ and thresholds the maximum. This test attains the optimal error curve, thereby confirming the absence of a statistical-to-computational gap. In a setting where the star sizes (k_ℓ) are unknown to the statistician, we construct an adaptive degree test that also attains the optimal performance asymptotically. In the course of the proofs we show a condensation phase transition in the likelihood ratio that closely resembles that of the random energy model from spin glass theory.

Keywords: planted models, planted star, multiplex, statistical-to-computational gap, optimal detection, random energy model

*School of Computer Science, Georgia Institute of Technology, inarang3@gatech.edu

[†]School of Computer Science, Georgia Institute of Technology, wperkins3@gatech.edu

[‡]School of Mathematics, Georgia Institute of Technology, timothy.wee@gatech.edu

Contents

1	Introduction	3
1.1	Summary of contributions	3
1.2	Problem formulation	4
1.3	Main results	5
1.4	Related work	8
2	Connection to the random energy model	9
3	Condensation of the likelihood ratio	11
4	Optimal testing	14
5	Detection and estimation with unknown star sizes	15
5.1	Hub recovery	17
5.2	Estimating the signal after identifying the hub	17
A	Proofs for Section 2	22
B	Proofs for Section 3	25
B.1	Gaussian estimates for weighted Bernoulli sums	25
B.2	Projected degree estimates	31
C	Proofs for Section 4	32
C.1	Maximum of projected degree	33
C.2	Hub and leaf recovery	35
D	Proofs for Section 5	36
D.1	Proof of Lemma 5.2	36

1 Introduction

Extracting information from networks is a fundamental task that spans many fields including computer science, statistics, and statistical physics. Extensive work has been done to better understand thresholds and algorithms for recovery and detection of planted structures in random graphs. Notable examples include the planted clique problem (Jerrum, 1992; Alon et al., 1998) for which many complexity-theoretic questions remain, as well as the community detection or stochastic block models (Decelle et al., 2011; Bhaskara et al., 2010; Arias-Castro and Verzelen, 2014; Mossel et al., 2015; Abbe, 2018). In recent years, planted combinatorial motifs, often with no intrinsic low rank structure, have received significant attention. These include planted Hamiltonian cycles (Bagaria et al., 2020), hidden nearest-neighbor graphs (Ding et al., 2020), planted matchings and k -factors (Moharrami et al., 2021; Ding et al., 2023; Gaudio et al., 2026; Wee and Mao, 2025; Addario-Berry et al., 2026), and planted trees (Massoulié et al., 2019; Moharrami et al., 2025).

In network science, it is often of interest to identify central nodes, or hubs, which have an outsized influence on the rest of the graph. This has many applications, ranging from page ranking in search engines, to super-spreaders in epidemiological networks (Kirkley, 2024). Such questions have been investigated in the context of preferential or uniform attachment networks, in the form of identifying early vertices or seed graphs (Bubeck et al., 2017; Rácz and Bubeck, 2017; Addario-Berry et al., 2024).

Multiplex networks represent several kinds of relationships among the same entities as separate graph layers. They have become an important framework for studying social, biological, and technological systems (Kivelä et al., 2014; Boccaletti et al., 2014). Examples include social interactions through different communication channels and airport connections operated by different airlines (Reiffers-Masson and Labatut, 2017; Rahmede et al., 2018). In such systems, it is often of practical interest to identify nodes that are persistently important across several network layers.

In this paper, we propose and study a model for detecting planted hubs in multiplex networks. We observe a multiplex graph $\mathbf{G}$ with a fixed number L of layers on the same n vertices. The null consists of independent $G(n, m_\ell)$ graphs. Under the alternative, we pick a hub common to all layers, uniformly at random, and plant a random star of k_ℓ edges centered at this hub. Background edges are added to the graph to bring each layer’s total edge count to m_ℓ . We ask:

- (i) How much background noise can be tolerated in each layer before detection of the common planted hub becomes impossible?
- (ii) Is there a computationally efficient test that attains the optimal detection performance?
- (iii) Can we attain the same test performance when the planted star sizes (k_ℓ) are unknown to the statistician?

1.1 Summary of contributions

Sharp detection asymptotics We give a sharp answer to question (i) by characterizing the asymptotic total variation (TV) distance between the null and alternative distributions. We identify an appropriate signal-to-noise ratio (SNR) vector $\mathbf{a}$ for the layers, whose entries are appropriately scaled star sizes k_ℓ , and show that the parameter $\|\mathbf{a}\|_2$ is the critical scaling parameter. Indeed, we show that TV distance converges to $\Phi(h)$ when $\|\mathbf{a}\|_2 - \sqrt{2 \log n} \rightarrow h \in \mathbb{R}$, where Φ denotes the standard Gaussian cumulative distribution function.

Our results also reveal the phenomenon that a planted hub may still be detectable in the multiplex network with vanishing error, even when every individual layer is asymptotically indis-

tinguishable from its corresponding null model.

We show that a computationally efficient test attains the asymptotically optimal TV curve, thus establishing the absence of a statistical-to-computational gap for this problem. The test statistic involves thresholding the maximum *projected* degrees, which are coefficients of the projections of the vector of layer degrees onto the direction $\mathbf{a}$.

Unknown star sizes In the case when the (k_ℓ) are unknown to the statistician, we address questions (ii) and (iii) by showing that a computationally efficient *adaptive* degree test also attains the optimal TV curve, and asymptotically matches the statistical performance of the projected degree test. Furthermore, we show that the unknown SNR vector $\mathbf{a}$ can also be accurately estimated, conditional on detecting the alternative.

In terms of recovery, we show that both the projected and adaptive procedures asymptotically maximize the probability of correctly identifying the planted hub, over all estimators based on the observed graph and known star sizes.

Likelihood ratio and REM A highlight of our analysis is connecting the likelihood ratio to a partition function resembling that of the *random energy model* (REM) from statistical physics. Unlike the classical REM, the “energy levels” here are dependent, non-Gaussian projected degrees.

To overcome this technical difficulty, we employ techniques from the theory of stable polynomials. In particular, we use the strongly Rayleigh property of the Hypergeometric distribution to obtain cumulant bounds from zero-free regions of the degree generating functions. These bounds yield Gaussian approximations for truncated exponential moments with vanishing relative error. Crucially, the approximations survive conditioning on finitely many other vertices’ degrees, allowing a truncated second moment argument to establish phase transitions of the likelihood ratio, analogous to REM condensation transitions.

Notation We use the standard $\omega(\cdot), o(\cdot), \Theta(\cdot), O(\cdot)$ notation. Write $a \ll b$ to mean $a = o(b)$, and $a \sim b$ to mean $a/b \rightarrow 1$. We say that an event holds with high probability (whp) if it holds with probability $1 - o(1)$ as $n \rightarrow \infty$. The Gaussian cumulative distribution function is denoted by $\Phi(\cdot)$ and its complement by $\Phi^c(\cdot)$. All logarithms are natural.

1.2 Problem formulation

We formally define the models. Fix a positive integer L , the number of layers, and write $N = \binom{n}{2}$.

Given positive integers m_ℓ and k_ℓ , the **planted** distribution $\mathbb{P}_1$ is the distribution over multiplex graphs $\mathbf{G} = (G^{(1)}, \dots, G^{(L)})$ on the common vertex set $[n]$ constructed as follows. First pick a hub vertex $v_* \in [n]$ uniformly at random. Independently in each layer ℓ , add k_ℓ incident edges to v_* uniformly at random, then choose $m_\ell - k_\ell$ edges uniformly over the remaining $N - k_\ell$ unplanted edges.

The **null** distribution $\mathbb{P}_0$ has independent layers $G^{(\ell)} \sim G(n, m_\ell)$: each layer is uniform over all graphs on n vertices with m_ℓ edges. In particular, the edge count in every layer agrees exactly under the two hypotheses.

Given a multiplex graph $\mathbf{G}$, we consider the hypothesis test

$$H_0 : \mathbf{G} \sim \mathbb{P}_0, \quad \text{vs.} \quad H_1 : \mathbf{G} \sim \mathbb{P}_1. \quad (1.1)$$

We will consider (1.1) under both settings when the star sizes $(k_\ell)_{\ell=1}^L$ are known and unknown to the statistician. In both settings, the planted center and the planted edge sets are always hidden.

Let $D_v^{(\ell)}$ be the degree of vertex v in layer ℓ . Its mean μ_ℓ and standard deviation σ_ℓ under the null $\mathbb{P}_0$ are given by

$$\mu_\ell := \mathbb{E}_0 D_v^{(\ell)} = \frac{2m_\ell}{n}, \quad \text{and} \quad \sigma_\ell^2 := \text{Var}_0(D_v^{(\ell)}) = \mu_\ell \left(1 - \frac{2}{n}\right) \frac{N - m_\ell}{N - 1}. \quad (1.2)$$

Throughout the paper, L is fixed and we assume that every layer ℓ satisfies

$$m_\ell = o(N), \quad (\log n)^2 \ll k_\ell \ll \sqrt{n}, \quad \text{and} \quad \frac{k_\ell^2}{\mu_\ell} = O(\log n). \quad (1.3)$$

These assumptions imply $\mu_\ell \gg (\log n)^3$ and $\sigma_\ell^2 \sim \mu_\ell$.

We collect several definitions that will be essential to our results. Define the signal-to-noise ratio (SNR) vector

$$\mathbf{a} = (a_1, \dots, a_L), \quad \text{where} \quad a_\ell = \frac{k_\ell \sigma_\ell}{\mu_\ell}. \quad (1.4)$$

Since $\sigma_\ell^2 \sim \mu_\ell$, we have $a_\ell \sim k_\ell / \sqrt{\mu_\ell}$. We may thus regard a_ℓ as the asymptotic SNR of layer ℓ . It will turn out that the critical order parameter that governs the detection behavior is $\|\mathbf{a}\|_2$.

Define the standardized layer degrees by

$$Y_v^{(\ell)} = \frac{D_v^{(\ell)} - \mu_\ell}{\sigma_\ell}, \quad (1.5)$$

and write $\mathbf{Y}_v = (Y_v^{(1)}, \dots, Y_v^{(L)})$. Define the *projected* degree statistic by

$$S_v = \frac{\langle \mathbf{a}, \mathbf{Y}_v \rangle}{\|\mathbf{a}\|_2}. \quad (1.6)$$

Thus S_v is a scalar weighted aggregate of the standardized degree vector of v .

1.3 Main results

1.3.1 Sharp detection threshold

Theorem 1.1. *Suppose (1.3) holds. Then*

$$\text{TV}(\mathbb{P}_0, \mathbb{P}_1) = \Phi(\|\mathbf{a}\|_2 - \sqrt{2 \log n}) + o(1).$$

We prove Theorem 1.1 in Section 3.

The minimum possible sum of Type-I and Type-II errors is $1 - \text{TV}(\mathbb{P}_0, \mathbb{P}_1)$. At a coarser level, Theorem 1.1 therefore also shows that detection is impossible when $\|\mathbf{a}\|_2 - \sqrt{2 \log n} \rightarrow -\infty$, and achievable with vanishing error when $\|\mathbf{a}\|_2 - \sqrt{2 \log n} \rightarrow +\infty$.

The following immediate corollary gives the scaling window around $\|\mathbf{a}\|_2^2 = 2 \log n$ and describes the shape of the transition.

Corollary 1.2. Suppose (1.3) holds. For $\gamma \in \mathbb{R}$, put

$$\|\mathbf{a}\|_2^2 = 2 \log n - 2\gamma \sqrt{\log n} + o(\sqrt{\log n}). \quad (1.7)$$

Then

$$\text{TV}(\mathbb{P}_0, \mathbb{P}_1) = 1 - \Phi\left(\frac{\gamma}{\sqrt{2}}\right) + o(1).$$

Remark 1.3. Theorem 1.1 reveals a striking feature of multiplex networks: observing several layers together can overcome the fundamental limits of detection from any single layer. Even with a fixed number of layers, where each layer can be asymptotically indistinguishable from its null model, the joint observation can permit detection with vanishing error. The gain comes from combining evidence about the same hidden vertex across layers. For example, fix $1 < r < L$ and suppose that each layer ℓ satisfies

$$a_\ell^2 = \left(\frac{2r}{L} + o(1)\right) \log n.$$

Thus $\|\mathbf{a}\|_2^2 = (2r + o(1)) \log n$. Since $\frac{r}{L} < 1 < r$, we have $a_\ell - \sqrt{2 \log n} \rightarrow -\infty$ for every layer, whereas $\|\mathbf{a}\|_2 - \sqrt{2 \log n} \rightarrow +\infty$. In other words, writing $\mathbb{P}_j^{(\ell)}$ for the marginal law of layer ℓ under H_j for $j = 0, 1$, Theorem 1.1 gives $\text{TV}(\mathbb{P}_0^{(\ell)}, \mathbb{P}_1^{(\ell)}) \rightarrow 0$ for each $\ell \in [L]$, but $\text{TV}(\mathbb{P}_0, \mathbb{P}_1) \rightarrow 1$.

1.3.2 Optimal testing and hub recovery

Denote the likelihood ratio by

$$\Lambda(\mathbf{G}) = \frac{d\mathbb{P}_1}{d\mathbb{P}_0}(\mathbf{G}). \quad (1.8)$$

By the Neyman–Pearson lemma (see e.g. Lehmann and Romano, 2005, Theorem 3.2.1), the test that rejects H_0 when $\Lambda(\mathbf{G}) \geq 1$ minimizes the sum of Type-I and Type-II errors and is hence optimal.

Suppose that the statistician has knowledge of the star sizes $(k_\ell)_{\ell=1}^L$. The SNR vector $\mathbf{a}$ is then known. We will show that projecting the standardized degree vectors in the direction of $\mathbf{a}$, as in (1.6), and thresholding the maximum projected statistic gives a simple test that is asymptotically optimal. For any fixed $\alpha > 0$, define the test

$$\psi_{\text{proj}}(\mathbf{G}) = \mathbf{1}\{\max_v S_v \geq t_n^*\} \quad \text{where} \quad t_n^* = \sqrt{2 \log n - \frac{1}{\alpha} \log \log n}. \quad (1.9)$$

Theorem 1.4. Suppose (1.3) holds. Then for any fixed $\alpha > 1$,

$$\mathbb{P}_0(\psi_{\text{proj}} = 1) = o(1), \quad \text{and} \quad \mathbb{P}_1(\psi_{\text{proj}} = 1) = \Phi(\|\mathbf{a}\|_2 - \sqrt{2 \log n}) + o(1).$$

Furthermore, in the critical window (1.7),

$$\mathbb{P}_j(\{\Lambda \geq 1\} \triangle \{\max_v S_v \geq t_n^*\}) = o(1), \quad \text{for } j \in \{0, 1\}.$$

This theorem is proved in Section 4.

Since the maximum projected degree is efficiently computable, this establishes that the multiplex planted hub detection problem exhibits no statistical-to-computational gap, in a very strong sense: a simple and tractable statistic asymptotically minimizes the sum of Type-I and Type-II errors

throughout the detection scaling window as found in Theorem 1.1. Furthermore, as the second assertion in Theorem 1.4 shows, the $\psi_{\text{proj}}(\mathbf{G})$ and $\Lambda(\mathbf{G})$ threshold tests make the same decision in the critical window, with high probability under both $\mathbb{P}_0$ and $\mathbb{P}_1$.

Theorem 1.4 also reveals an asymmetry in the optimal test, in that it is correct whp under the null, and it is the alternative that gives rise to the error.

Our next result concerns recovery under $\mathbb{P}_1$. We show that maximizing the projected degree (1.6) attains the optimal probability of recovering the hub.

Corollary 1.5. *Suppose (1.3) holds. Let $\hat{v} \in \arg \max_v S_v$. Then*

$$\mathbb{P}_1(\hat{v} = v_*) = \sup_{\tilde{v}} \mathbb{P}_1(\tilde{v} = v_*) + o(1) = \Phi\left(\|\mathbf{a}\|_2 - \sqrt{2 \log n}\right) + o(1), \quad (1.10)$$

where the supremum ranges over all estimators based on the observed multiplex graph $\mathbf{G}$.

The proof is given in Appendix C.

Regarding recovery of the planted star leaves, we instead claim that even with the hub known, a list of candidate leaves comparable in size to the planted star recovers only a vanishing fraction of the planted leaves with high probability. A precise statement is given in Appendix C.

1.3.3 Detection and estimation under unknown star sizes

We retain the model of Section 1.2, the hypotheses in (1.1), and the parameter regime (1.3), and turn to a setting where the planted star sizes

$$(k_\ell)_{\ell=1}^L \quad \text{are unknown to the statistician.} \quad (1.11)$$

As the statistician observes $\mathbf{G}$, the parameters n , L , and $(m_\ell)_{\ell=1}^L$ are assumed known.

The additional challenge in this setting is that the SNR vector $\mathbf{a}$ in (1.4) is unavailable, and thus projected degrees (S_v) in (1.6) cannot be constructed. In this setting, we investigate the detection problem (1.1), as well as estimation of the SNR vector $\mathbf{a}$.

For detection, the alternative test statistic we propose is given as follows. Define

$$R_v = \|(\mathbf{Y}_v)_+\|_2, \quad \text{and} \quad u_n = \sqrt{2 \log n + (2L + 1) \log \log n}, \quad (1.12)$$

where positive parts are taken coordinatewise. Note that the vector $\mathbf{Y}_v = (Y_v^{(1)}, \dots, Y_v^{(L)})$ of standardized degrees in (1.5) does not depend on the star sizes. Define the test

$$\psi_{\text{adap}}(\mathbf{G}) = \mathbf{1}\{\max_{v \in [n]} R_v \geq u_n\}. \quad (1.13)$$

We call this procedure *adaptive* because it is designed to work for every admissible choice of unknown star sizes.

Theorem 1.6. *Suppose (1.3) holds, and we are in the setting of (1.11). Then*

$$\mathbb{P}_0(\psi_{\text{adap}} = 1) = o(1), \quad \text{and} \quad \mathbb{P}_1(\psi_{\text{adap}} = 1) = \Phi\left(\|\mathbf{a}\|_2 - \sqrt{2 \log n}\right) + o(1).$$

The proof is given in Section 5.

In other words, we can indeed attain the optimal error rate implied by Theorem 1.1 for every admissible list of star sizes satisfying (1.3), without knowing those sizes.

We now give an analogous result to Corollary 1.5 for recovery. Let our estimator of the planted hub, which does not require knowledge of the star sizes, be $\hat{v}_{\text{adap}} \in \arg \max_v R_v$. Our next result shows (i) that this estimator attains the optimal recovery probability, even among estimators that know the star sizes, and (ii) if we report $\hat{v}_{\text{adap}}$ only when the test rejects the null and otherwise return no answer, the probability of reporting an incorrect hub tends to zero.

Proposition 1.7. *Suppose (1.3) holds. Then the following hold:*

- (i) $\mathbb{P}_1(\hat{v}_{\text{adap}} = v_*) = \sup_{\tilde{v}} \mathbb{P}_1(\tilde{v} = v_*) + o(1) = \Phi(\|\mathbf{a}\|_2 - \sqrt{2 \log n}) + o(1)$, where the supremum ranges over all such estimators based on the observed $\mathbf{G}$ and star sizes $(k_\ell)_\ell$.
- (ii) $\mathbb{P}_1(\psi_{\text{adap}} = 1, \hat{v}_{\text{adap}} \neq v_*) = o(1)$.

We prove this proposition in Section 5.1.

We next consider estimating the SNR vector $\mathbf{a}$. Let $E_n := \{\psi_{\text{adap}} = 1\}$ be the event that the adaptive procedure does detect the alternative. On E_n , define

$$\hat{\mathbf{a}} := (\mathbf{Y}_{\hat{v}_{\text{adap}}})_+. \quad (1.14)$$

Since $a_\ell = k_\ell \sigma_\ell / \mu_\ell$ by (1.4), this also gives the plug-in estimate for the star sizes $\hat{k}_\ell := (\mu_\ell / \sigma_\ell) \hat{a}_\ell$. Outside E_n , set these estimates to zero.

Our next theorem quantifies the estimation error conditional on E_n in the critical window.

Theorem 1.8. *Suppose (1.3) holds and suppose $\|\mathbf{a}\|_2 - \sqrt{2 \log n} \rightarrow h \in \mathbb{R}$. Then under $\mathbb{P}_1(\cdot \mid E_n)$,*

$$\|\hat{\mathbf{a}} - \mathbf{a}\|_2 = O_{\mathbb{P}}(1). \quad (1.15)$$

Consequently, under the same conditional law, (1.15) gives $\hat{k}_\ell = k_\ell(1 + o_{\mathbb{P}}(1))$ for every layer with $a_\ell \rightarrow \infty$. The norm of the SNR vector can also be estimated as $\|\hat{\mathbf{a}}\|_2 = \|\mathbf{a}\|_2 + O_{\mathbb{P}}(1)$.

When $\|\mathbf{a}\|_2 - \sqrt{2 \log n} \rightarrow +\infty$, i.e., when detection is possible with vanishing error, the bound (1.15) and its consequences above hold under $\mathbb{P}_1$ without conditioning.

When $\|\mathbf{a}\|_2 - \sqrt{2 \log n} \rightarrow -\infty$, i.e., when detection is impossible, the test rejects with probability $o(1)$, and the estimation error $\|\hat{\mathbf{a}} - \mathbf{a}\|_2$ in (1.15) diverges whenever it rejects.

The proof of Theorem 1.8 and details of these consequences are given in Section 5.2.

1.4 Related work

Centrality in multiplex networks Vertex importance across network layers was studied in (Halu et al., 2013; De Domenico et al., 2015). Reiffers-Masson and Labatut (2017) derive opinion centrality by explicitly solving an influence optimization problem. Rahmede et al. (2018) introduce MultiRank, which jointly determines node centralities and layer influences through coupled equations, and apply it to real networks.

Planted subgraph detection General frameworks for understanding arbitrary planted subgraphs address inference thresholds (Mossel et al., 2023), detection with differing signal and background edge densities (Elimelech and Huleihel, 2025), tests of constant polynomial degree in dense Erdős–Rényi graphs (Yu et al., 2025), and recovery of hidden weighted graphs (Hou and Liu, 2026).

A star is a tree, and planted trees have been studied in [Massoulié et al. \(2019\)](#) and [Moharrami et al. \(2025\)](#). The former operates in $G(n, \lambda/n)$ with fixed λ , and the latter considers spanning weighted trees. These works establish a separation between detection and reconstruction in sparse graphs and characterize the asymptotic fraction of planted edges recovered by the minimum spanning tree algorithm in weighted graphs. For $L = 1$, our results describe the sharp threshold behavior of an unweighted planted star model with diverging average degree, a setting not covered by these works.

Error curves in the critical regime Precise testing error curves have been studied for stochastic block models and planted factor graphs ([Mossel et al., 2025](#)), spiked Wigner models ([Moitra and Wein, 2025](#)), and planted matchings ([Wee and Mao, 2025](#)). Beyond graph models, [Addario-Berry et al. \(2010\)](#) study detection of a structured set of nonzero means in a Gaussian vector. Precise results in an abstract planted problem setting, not necessarily involving graph structure, are given in [Perkins \(2013\)](#). A special case of the main result there is analogous to our $L = 1$ case with a known hub and only the degrees being observed. This difference leads to a different optimal statistic than the maximum degree test. See also the remarks in [Perkins \(2013, Section 9\)](#) which suggested studying an unlabeled planted balls and bins problem and served as the original inspiration for this work.

The null likelihood ratio in our critical window converges in probability to a constant strictly between zero and one. This is distinct from the log-Gaussian limits that arise for planted matchings ([Wee and Mao, 2025](#)), spiked models ([El Alaoui et al., 2020](#); [Banerjee and Ma, 2022](#)), subgraph counts ([Janson, 1994](#)), and partition functions in the symmetric perceptron ([Abbe et al., 2022](#)).

Organization Section 2 establishes the likelihood ratio approximation and its connection to the REM. Section 3 proves the likelihood ratio asymptotics and sharp detection thresholds. Section 4 establishes optimality of the projected test, and Section 5 treats detection and estimation with unknown star sizes. Supporting proofs are collected in Appendices A–D.

2 Connection to the random energy model

Our proof begins with a sharp approximation of the likelihood ratio Λ (1.8), which we now spell out:

$$\Lambda(\mathbf{G}) = \frac{1}{n} \sum_{v=1}^n \prod_{\ell=1}^L \frac{\binom{N}{m_\ell} \binom{D_v^{(\ell)}}{k_\ell}}{\binom{n-1}{k_\ell} \binom{N-k_\ell}{m_\ell-k_\ell}}. \quad (2.1)$$

Our first result approximates this likelihood ratio by a sum involving the projected degrees S_v defined in (1.6).

Lemma 2.1. *Suppose (1.3) holds. Then whp under both $\mathbb{P}_0$ and $\mathbb{P}_1$,*

$$\Lambda(\mathbf{G}) = (1 + o(1)) e^{-\|\mathbf{a}\|_2^2/2} \frac{1}{n} \sum_{v=1}^n e^{\|\mathbf{a}\|_2 S_v}. \quad (2.2)$$

We defer the proof to Appendix A.

Thus, the randomness in the likelihood ratio essentially comes from the $\sum_{v=1}^n e^{\|\mathbf{a}\|_2 S_v}$ term. This resembles the partition function of the random energy model (REM) from spin glass theory. The REM was originally formulated by [Derrida \(1981\)](#) as a simplified model of the canonical

Sherrington-Kirkpatrick spin glass (Sherrington and Kirkpatrick, 1975). The energies $\{H_j\}$ in the REM are taken as i.i.d. Gaussian random variables, thus abstracting away any underlying configuration geometry. More precisely, the REM partition function $Z_{\text{REM}}(\beta)$ is given by

$$Z_{\text{REM}}(\beta) = \sum_{j=1}^{2^M} \exp(-\beta H_j), \quad \{H_j\}_{j=1}^{2^M} \stackrel{\text{iid}}{\sim} \mathcal{N}(0, M), \quad (2.3)$$

where M represents the number of sites in the system, and where $\beta > 0$ is an inverse temperature parameter. The motivation for the distribution and scaling of the 2^M energy levels in the REM comes from the Sherrington-Kirkpatrick spin glass in which spins are $\{\pm 1\}$ -valued and their energies are Gaussian random variables (albeit with complex correlation structure).

Despite its apparent simplicity, the REM already exhibits a phase transition as the temperature varies, with critical inverse temperature $\beta_c = \sqrt{2 \log 2}$. As $M \rightarrow \infty$, the following convergence holds in probability:

$$\frac{Z_{\text{REM}}(\beta)}{\mathbb{E} Z_{\text{REM}}(\beta)} \rightarrow \begin{cases} 1 & \text{for fixed } \beta < \beta_c, \\ \Phi(-h) & \text{if } \sqrt{M}(\beta - \beta_c) \rightarrow h \in \mathbb{R}, \\ 0 & \text{for fixed } \beta > \beta_c. \end{cases}$$

The first and third limits are classical (Bovier et al., 2002; Bovier, 2006), the middle limit follows from Kabluchko and Klimovsky (2014, Theorem 4.8 and Lemma 3.3).

In our setting, where the S_v 's are now dependent, we will establish a similar transition (cf. Proposition 3.1):

$$\Lambda(\mathbf{G}) - \Phi(\sqrt{2 \log n} - \|\mathbf{a}\|_2) \xrightarrow{\mathbb{P}_0} 0. \quad (2.4)$$

Thus, as $n \rightarrow \infty$,

$$\Lambda(\mathbf{G}) \xrightarrow{\mathbb{P}_0} \begin{cases} 1 & \text{if } \|\mathbf{a}\|_2 - \sqrt{2 \log n} \rightarrow -\infty, \\ \Phi(-h) & \text{if } \|\mathbf{a}\|_2 - \sqrt{2 \log n} \rightarrow h \in \mathbb{R}, \\ 0 & \text{if } \|\mathbf{a}\|_2 - \sqrt{2 \log n} \rightarrow +\infty. \end{cases}$$

Our results thus show an interesting connection between spin glass phase transitions and detection thresholds for statistical estimation problems. One can compare this to the behavior of the disassortative stochastic block model with low edge density in which the information-theoretic threshold for detection coincides with a condensation phase transition for the corresponding anti-ferromagnetic Ising or Potts model on a random graph (Mossel et al., 2015; Coja-Oghlan et al., 2018).

The main difficulty here is to establish the REM transition for dependent energy levels and connect it to the testing problem. The classical REM techniques do not apply because the projected degrees (S_v) have dependencies. We instead show that conditioning on one vertex's degrees has a negligible effect on the truncated exponential moments at another, hence adapting a truncated second moment argument.

The classical REM has been generalized substantially, for instance to generalized or continuous REMs (Derrida and Gardner, 1986; Olivieri and Picco, 1984; Capocaccia et al., 1987; Ruelle, 1987; Bolthausen and Sznitman, 1998; Bovier and Kurkova, 2004). These models consider Gaussian energies that have a hierarchical correlation structure, in an effort to study replica symmetry breaking. We leave the exploration of further connections between planted subgraph problems and spin glass models such as the GREMs to future work.

3 Condensation of the likelihood ratio

We now state our main result for the likelihood ratio (2.1). It describes the condensation transition under the null, including its behavior throughout the critical window.

Proposition 3.1. *Suppose (1.3) holds, then*

$$\Lambda - \Phi\left(\sqrt{2\log n} - \|\mathbf{a}\|_2\right) \xrightarrow{\mathbb{P}_0} 0. \quad (3.1)$$

In the critical window (1.7), for fixed $\gamma \in \mathbb{R}$,

$$\Lambda \xrightarrow{\mathbb{P}_0} \Phi\left(\frac{\gamma}{\sqrt{2}}\right). \quad (3.2)$$

The proof follows a truncated second moment argument, as in the classical REM. We first state the Gaussian moment estimates needed for the exponential sum in (2.2). All expectations in the next two lemmas are under $\mathbb{P}_0$.

We state the first estimate in terms of arbitrary projection directions $\mathbf{u} = (u_1, \dots, u_L) \in \mathbb{R}^L$ satisfying $\|\mathbf{u}\|_2 = 1$, since this form will be useful later. Define

$$S_v(\mathbf{u}) = \sum_{\ell=1}^L u_\ell \frac{D_v^{(\ell)} - \mu_\ell}{\sigma_\ell}, \quad \text{and} \quad \delta_{\mathbf{u}} = \max_{\ell: u_\ell \neq 0} \frac{|u_\ell|}{\sigma_\ell}.$$

Notice that setting $\mathbf{u} = \mathbf{a}/\|\mathbf{a}\|_2$, we recover the projected degree S_v in (1.6).

The regime (1.3) and variance formula (1.2) imply

$$\mu_\ell \gg (\log n)^3, \quad \sigma_\ell^2 \sim \mu_\ell, \quad \min_{\ell \in [L]} \sigma_\ell \gg (\log n)^{3/2}. \quad (3.3)$$

Lemma 3.2. *Suppose that $\delta_{\mathbf{u}}(\log n)^{3/2} = o(1)$, then, uniformly for $|s|, |t| = O(\sqrt{\log n})$,*

$$\mathbb{E}e^{sS_v(\mathbf{u})} = (1 + o(1))e^{s^2/2}, \quad (3.4)$$

$$\mathbb{E}[e^{sS_v(\mathbf{u})} \mathbf{1}_{\{S_v(\mathbf{u}) \geq t\}}] = (1 + o(1))e^{s^2/2} \Phi^c(t - s), \quad (3.5)$$

$$\mathbb{E}[e^{sS_v(\mathbf{u})} \mathbf{1}_{\{S_v(\mathbf{u}) < t\}}] = (1 + o(1))e^{s^2/2} \Phi(t - s). \quad (3.6)$$

By (3.3), $\min_\ell \sigma_\ell \gg (\log n)^{3/2}$, so the condition on $\delta_{\mathbf{u}}$ holds and the errors are uniform over all unit vectors $\mathbf{u}$, including vectors that depend on n .

The proof is given in Appendix B.

Our next estimate involves two distinct vertices. It follows from a conditional version of the preceding lemma, Lemma B.7, given in Appendix B, which shows that the Gaussian moment estimates remain valid after conditioning on another vertex's layer degrees.

Lemma 3.3. *Suppose (1.3) holds. Then for $|t| = O(\sqrt{\log n})$ and distinct vertices u, v ,*

$$\mathbb{E}_0 \left[e^{\|\mathbf{a}\|_2 S_u + \|\mathbf{a}\|_2 S_v} \mathbf{1}_{\{S_u < t, S_v < t\}} \right] = (1 + o(1))e^{\|\mathbf{a}\|_2^2} \Phi(t - \|\mathbf{a}\|_2)^2. \quad (3.7)$$

Proof of Lemma 3.3. Write $\mathbf{D}_u = (D_u^{(1)}, \dots, D_u^{(L)})$. By Lemma B.7, uniformly in $\mathbf{D}_u$,

$$\mathbb{E}_0 \left[e^{\|\mathbf{a}\|_2 S_v} \mathbf{1}_{\{S_v < t\}} \mid \mathbf{D}_u \right] = (1 + o(1)) e^{\|\mathbf{a}\|_2^2/2} \Phi(t - \|\mathbf{a}\|_2).$$

The result follows by multiplying by $e^{\|\mathbf{a}\|_2 S_u} \mathbf{1}_{\{S_u < t\}}$, taking expectations, and applying Lemma 3.2 to the remaining factor. $\square$

We record a standard Gaussian tail estimate used throughout this paper.

Theorem 3.4 (Mills' ratio; see e.g. Wainwright, 2019, Exercise 2.2). *As $x \rightarrow \infty$,*

$$\Phi^c(x) = \frac{1 + O(x^{-2})}{x\sqrt{2\pi}} e^{-x^2/2}.$$

We are now ready to prove Proposition 3.1.

Proof of Proposition 3.1. By (2.2), whp under $\mathbb{P}_0$, we may write

$$\Lambda = (1 + o(1))W, \quad \text{where} \quad W := \frac{e^{-\|\mathbf{a}\|_2^2/2}}{n} \sum_{v=1}^n e^{\|\mathbf{a}\|_2 S_v}. \quad (3.8)$$

Fix $\alpha > 1$, let t_n^* be as in (1.9), and set

$$B := \left\{ \max_{v \in [n]} S_v \geq t_n^* \right\}.$$

By a union bound, (3.5) with $s = 0$, and Theorem 3.4,

$$\mathbb{P}_0(B) \leq (1 + o(1))n\Phi^c(t_n^*) = O\left((\log n)^{1/(2\alpha)-1/2}\right) = o(1). \quad (3.9)$$

Define the truncated version of W ,

$$\widehat{W} := \frac{e^{-\|\mathbf{a}\|_2^2/2}}{n} \sum_{v=1}^n e^{\|\mathbf{a}\|_2 S_v} \mathbf{1}_{\{S_v < t_n^*\}}.$$

On B^c , truncation removes no summands, so $W = \widehat{W}$. Equations (3.6) and (3.7) give

$$\mathbb{E}_0 \widehat{W} = (1 + o(1))\Phi(t_n^* - \|\mathbf{a}\|_2), \quad (3.10)$$

$$\mathbb{E}_0 \widehat{W}^2 = (1 + o(1)) \left[\frac{e^{\|\mathbf{a}\|_2^2}}{n} \Phi(t_n^* - 2\|\mathbf{a}\|_2) + \frac{n-1}{n} \Phi(t_n^* - \|\mathbf{a}\|_2)^2 \right]. \quad (3.11)$$

We claim that the diagonal term in (3.11) is $o(1)$ throughout the entire range $\|\mathbf{a}\|_2 = O(\sqrt{\log n})$. Indeed, if $2\|\mathbf{a}\|_2 \leq t_n^*$, then

$$\frac{e^{\|\mathbf{a}\|_2^2}}{n} \Phi(t_n^* - 2\|\mathbf{a}\|_2) \leq \frac{e^{(t_n^*)^2/4}}{n} = o(1). \quad (3.12)$$

On the other hand, if $2\|\mathbf{a}\|_2 > t_n^*$, use $\Phi(-x) \leq \frac{1}{2}e^{-x^2/2}$ for $x \geq 0$ to obtain

$$\frac{e\|\mathbf{a}\|_2^2}{n}\Phi(t_n^* - 2\|\mathbf{a}\|_2) \leq \frac{1}{2}\exp\left\{\frac{(t_n^*)^2}{2} - \log n - (\|\mathbf{a}\|_2 - t_n^*)^2\right\} \leq \frac{1}{2}(\log n)^{-1/(2\alpha)} = o(1). \quad (3.13)$$

Using (3.10)–(3.13) and $0 \leq \Phi(t_n^* - \|\mathbf{a}\|_2) \leq 1$, we obtain

$$\mathbb{E}_0(\widehat{W} - \Phi(t_n^* - \|\mathbf{a}\|_2))^2 = \mathbb{E}_0\widehat{W}^2 - 2\Phi(t_n^* - \|\mathbf{a}\|_2)\mathbb{E}_0\widehat{W} + \Phi(t_n^* - \|\mathbf{a}\|_2)^2 = o(1). \quad (3.14)$$

Since $W = \widehat{W}$ on B^c , (3.9) and Markov's inequality applied to (3.14) give, for every $\varepsilon > 0$,

$$\mathbb{P}_0(|W - \Phi(t_n^* - \|\mathbf{a}\|_2)| > \varepsilon) \leq \mathbb{P}_0(B) + \varepsilon^{-2}\mathbb{E}_0(\widehat{W} - \Phi(t_n^* - \|\mathbf{a}\|_2))^2 = o(1). \quad (3.15)$$

Since Φ takes values in $[0, 1]$, (3.15) implies $W = O_{\mathbb{P}_0}(1)$. Thus (3.8) gives $\Lambda - W \xrightarrow{\mathbb{P}_0} 0$, and (3.15) yields

$$\Lambda - \Phi(t_n^* - \|\mathbf{a}\|_2) = (\Lambda - W) + (W - \Phi(t_n^* - \|\mathbf{a}\|_2)) \xrightarrow{\mathbb{P}_0} 0. \quad (3.16)$$

Finally, (1.9) gives $t_n^* - \sqrt{2\log n} = o(1)$. Since $\Phi'(x) \leq 1/\sqrt{2\pi}$,

$$\left|\Phi(t_n^* - \|\mathbf{a}\|_2) - \Phi(\sqrt{2\log n} - \|\mathbf{a}\|_2)\right| \leq \frac{|t_n^* - \sqrt{2\log n}|}{\sqrt{2\pi}} = o(1). \quad (3.17)$$

Equations (3.16) and (3.17) prove (3.1).

In the critical window (1.7),

$$\|\mathbf{a}\|_2 = \sqrt{2\log n} - \frac{\gamma}{\sqrt{2}} + o(1). \quad (3.18)$$

Equations (3.1) and (3.18) imply (3.2). □

We now deduce Theorem 1.1 from this limit.

Proof of Theorem 1.1. We note the following characterization of total variation distance:

$$\mathrm{TV}(\mathbb{P}_0, \mathbb{P}_1) = \mathbb{P}_1(\Lambda \geq 1) - \mathbb{P}_0(\Lambda \geq 1) = \mathbb{E}_0(1 - \Lambda)_+. \quad (3.19)$$

The function $x \mapsto (1 - x)_+$ is bounded by one on $[0, \infty)$ and is Lipschitz. Thus Proposition 3.1 implies

$$\mathrm{TV}(\mathbb{P}_0, \mathbb{P}_1) = 1 - \Phi\left(\sqrt{2\log n} - \|\mathbf{a}\|_2\right) + o(1) = \Phi\left(\|\mathbf{a}\|_2 - \sqrt{2\log n}\right) + o(1),$$

where we use that convergence in probability of the bounded difference implies convergence of its expectation. □

4 Optimal testing

In this section we prove Theorem 1.4, which shows that the optimal likelihood ratio thresholding test reduces to a simpler test thresholding the maximum projected degree statistic $\max_v S_v$.

Our first result determines the behavior of this test statistic under the null.

Lemma 4.1. *Suppose (1.3) holds. For fixed $\alpha > 0$, let t_n^* be as in (1.9). Then under $\mathbb{P}_0$,*

$$\mathbb{P}_0\left(\max_{v \in [n]} S_v \geq t_n^*\right) \longrightarrow \begin{cases} 0, & \alpha > 1, \\ 1, & 0 < \alpha < 1. \end{cases} \quad (4.1)$$

In particular,

$$\max_{v \in [n]} S_v = \sqrt{2 \log n} + o_{\mathbb{P}_0}(1). \quad (4.2)$$

Furthermore, the statements (4.1) and (4.2) continue to hold when one fixed vertex is omitted from the maximum. They also remain valid after replacing the projection direction $\mathbf{a}/\|\mathbf{a}\|_2$ in (1.6) by any deterministic nonnegative unit vector $\mathbf{u}$, possibly depending on n .

The proof is given in Appendix C.

The next lemma describes the maximum projected statistic away from the planted center v_* , as well as the center's own statistic, under $\mathbb{P}_1$. Recall from (1.4) that $\|\mathbf{a}\|_2$ is the effective signal strength.

Lemma 4.2. *Suppose (1.3) holds. For fixed $\alpha > 1$, let t_n^* be as in (1.9). Then*

$$\mathbb{P}_1\left(\max_{v \neq v_*} S_v \geq t_n^*\right) = o(1). \quad (4.3)$$

Moreover,

$$\max_{v \neq v_*} S_v = \sqrt{2 \log n} + o_{\mathbb{P}_1}(1). \quad (4.4)$$

For the planted center,

$$S_{v_*} - \|\mathbf{a}\|_2 \xrightarrow{d} \mathcal{N}(0, 1). \quad (4.5)$$

We defer the proof to Appendix C.

Proof of Theorem 1.4. In this proof, for fixed $\alpha > 1$ defining t_n^* in (1.9), write

$$A := \{\Lambda \geq 1\} \quad \text{and} \quad B := \left\{ \max_v S_v \geq t_n^* \right\},$$

Lemma 4.1 gives $\mathbb{P}_0(B) = o(1)$. By Lemma 4.2,

$$\mathbb{P}_1(B) = \mathbb{P}_1(S_{v_*} \geq t_n^*) + o(1) = \Phi\left(\|\mathbf{a}\|_2 - \sqrt{2 \log n}\right) + o(1),$$

where the last equality follows from (4.5) and $t_n^* = \sqrt{2 \log n} + o(1)$. Comparing with Theorem 1.1 proves asymptotic optimality of the test.

We now turn to the equivalence of the two tests in the critical window (1.7). Proposition 3.1 gives $\Lambda \xrightarrow{\mathbb{P}_0} \Phi(\gamma/\sqrt{2}) < 1$, so $\mathbb{P}_0(A) = o(1)$ and hence $\mathbb{P}_0(A \triangle B) = o(1)$. The exact identity

$$\mathbb{P}_1(A) - \mathbb{P}_0(A) = \text{TV}(\mathbb{P}_0, \mathbb{P}_1)$$

shows that $\mathbb{P}_1(A) = \mathbb{P}_1(B) + o(1)$. Moreover, changing measure on $B \setminus A$ gives

$$\mathbb{P}_1(B \setminus A) = \mathbb{E}_0[\Lambda \mathbf{1}_{B \setminus A}] \leq \mathbb{P}_0(B) = o(1).$$

It follows that $\mathbb{P}_1(A \setminus B) = o(1)$ as well, proving the claimed equivalence under both laws. $\square$

5 Detection and estimation with unknown star sizes

In this section we prove asymptotic optimality of the adaptive test (Theorem 1.6), then establish hub recovery (Proposition 1.7), and estimation of the SNR vector conditional on detection of the alternative in the critical window (Theorem 1.8).

Throughout, (1.11) is in effect so the star sizes are unknown. The projected degree S_v in (1.6) uses the SNR vector $\mathbf{a}$ and is hence unavailable. We use instead the adaptive degree R_v defined in (1.12). An alternative representation for R_v is illuminating: by Cauchy–Schwarz,

$$R_v = \|(\mathbf{Y}_v)_+\|_2 = \sup_{\substack{\mathbf{u} \geq 0 \\ \|\mathbf{u}\|_2 \leq 1}} \langle \mathbf{u}, \mathbf{Y}_v \rangle.$$

Therefore, R_v selects for the strongest directions, among all nonnegative choices of layer weights. The statistic depends only on the observed degrees and their null means and variances, and can be computed in $O(nL)$ time once degrees have been computed.

We first provide two lemmas about the statistical properties of R_v under the null and alternative, respectively. The projected degrees in (1.6) use the same fixed weights $\mathbf{a}/\|\mathbf{a}\|_2$ for every vertex. The adaptive statistic R_v instead chooses the weights separately for each vertex, using the observed degree vector $\mathbf{Y}_v$. To account for this additional maximization, we use the higher threshold $u_n > \sqrt{2 \log n}$ from (1.12), where recall from (4.2) that the null maximum $\max_v S_v$ is $\sqrt{2 \log n} + o_{\mathbb{P}_0}(1)$. The next lemma bounds the maximum at a slightly reduced threshold, as needed for a coupling argument in the proof of Theorem 1.6.

Lemma 5.1. *Suppose (1.3) holds, and let $\sigma_* := \min_{\ell \in [L]} \sigma_\ell$. Then for any deterministic $d_n = O(\sigma_*^{-1})$,*

$$\mathbb{P}_0\{\max_v R_v \geq u_n - d_n\} = o(1),$$

where u_n is defined in (1.12). In particular, $\max_v R_v \leq \sqrt{2 \log n} + o_{\mathbb{P}_0}(1)$.

We note that under (1.3),

$$u_n - \sqrt{2 \log n} = O\left(\frac{\log \log n}{\sqrt{\log n}}\right) = o(1), \quad \text{and} \quad d_n = o(u_n - \sqrt{2 \log n}).$$

Thus $u_n - d_n$ still remains slightly above the background scale $\sqrt{2 \log n}$.

Proof of Lemma 5.1. We prove the stronger bound with threshold $\sqrt{2 \log n} + C \log \log n - d_n$ for any fixed $C > L$. Taking $C = 2L + 1$ will give the stated result. Taylor expansion of the cumulant generating function (as in (B.16)) gives, uniformly over unit vectors $\mathbf{u} \in \mathbb{R}^L$ and $|z| = O(\sqrt{\log n})$,

$$\log \mathbb{E}_0 \exp\{z \langle \mathbf{u}, \mathbf{Y}_v \rangle\} = \frac{z^2}{2} + O\left(\frac{|z|^3}{\sigma_*}\right) = \frac{z^2}{2} + o(1). \quad (5.1)$$

We now approximate all projection directions by a finite set. Put $\varepsilon_n = (\log n)^{-1/2}$ and choose a set $\mathcal{N}_n$ of unit vectors such that every unit vector in $\mathbb{R}^L$ is within Euclidean distance ε_n of some element of $\mathcal{N}_n$, with $|\mathcal{N}_n| \leq (1 + 2/\varepsilon_n)^L$. For every $\mathbf{y} \in \mathbb{R}^L$,

$$\max_{\mathbf{u} \in \mathcal{N}_n} \langle \mathbf{u}, \mathbf{y} \rangle \geq (1 - \varepsilon_n^2/2) \|\mathbf{y}\|_2.$$

Since L is fixed, this set contains only $O((\log n)^{L/2})$ directions. The term $C \log \log n$ in the squared threshold absorbs this factor in the union bound.

Set $x_n = \sqrt{2 \log n + C \log \log n} - d_n$. Since $\|(\mathbf{y})_+\|_2 \leq \|\mathbf{y}\|_2$, a union bound over vertices and the directions in $\mathcal{N}_n$, followed by the Chernoff bound from (5.1), gives

$$\begin{aligned} \mathbb{P}_0\{\max_v R_v \geq x_n\} &\leq n|\mathcal{N}_n| \exp\left\{-\frac{1}{2}(1 - \varepsilon_n^2/2)^2 x_n^2 + o(1)\right\} \\ &= O((\log n)^{(L-C)/2}) = o(1). \end{aligned}$$

By (3.3), $\sigma_* \gg (\log n)^{3/2}$, so $d_n \sqrt{\log n} = o(1)$, and replacing x_n by $(1 - \varepsilon_n^2/2)x_n$ changes its squared value by $O(1)$. $\square$

The next lemma describes the planted hub's standardized degree vector $\mathbf{Y}_{v_*}$ and adaptive statistic R_{v_*} .

Lemma 5.2. *Suppose (1.3) holds. Then under $\mathbb{P}_1$, conditional on the planted center v_* ,*

$$\mathbf{Y}_{v_*} - \mathbf{a} \xrightarrow{d} \mathcal{N}(0, I_L), \quad \text{and} \quad R_{v_*} = \|\mathbf{a}\|_2 + O_{\mathbb{P}_1}(1).$$

If $\|\mathbf{a}\|_2 \rightarrow \infty$, then

$$R_{v_*} - \|\mathbf{a}\|_2 \xrightarrow{d} \mathcal{N}(0, 1). \quad (5.2)$$

The proof is given in Appendix D.

We are now ready to prove Theorem 1.6.

Proof of Theorem 1.6. Lemma 5.1 proves the Type I error bound.

We now consider the Type II error bound. Equivalently, we show that the power satisfies $\mathbb{P}_1(\psi_{\text{adap}} = 1) = \Phi(\|\mathbf{a}\|_2 - \sqrt{2 \log n}) + o(1)$. We couple the null and planted graphs to bound how much planting can increase the degrees of vertices other than the hub. Fix the center v_* and its planted edge sets. Independently in each layer ℓ , sample a uniform graph with m_ℓ edges. Add each planted edge that is absent from this graph. To keep the edge count equal to m_ℓ , delete the same number of edges, chosen uniformly from the edges present outside the planted set. Conditional on the number of planted edges originally present, the original edges outside the planted set form a uniform subset. Uniform deletion therefore leaves a uniform subset of $m_\ell - k_\ell$ edges outside the planted set. Together with the planted edges, these have the required planted distribution.

Each vertex other than the center receives at most one inserted edge in each layer. Denoting the null standardized degree vector in this coupling by $\mathbf{Y}_v^0$, we have coordinatewise

$$\mathbf{Y}_v \leq \mathbf{Y}_v^0 + (\sigma_1^{-1}, \dots, \sigma_L^{-1}), \quad \text{for } v \neq v_*.$$

The map $\mathbf{y} \mapsto \|(\mathbf{y})_+\|_2$ is increasing in each coordinate and has Lipschitz constant 1. Hence

$$\mathbb{P}_1\left(\max_{v \neq v_*} R_v \geq u_n\right) \leq \mathbb{P}_0\left(\max_v R_v \geq u_n - \left(\sum_\ell \sigma_\ell^{-2}\right)^{1/2}\right) = o(1), \quad (5.3)$$

where the last step uses Lemma 5.1 with $d_n = (\sum_\ell \sigma_\ell^{-2})^{1/2} = O(\sigma_*^{-1})$. The rejection event $\{\psi_{\text{adap}} = 1\}$ and the event $\{R_{v_*} \geq u_n\}$ can differ only if $R_v \geq u_n$ for some $v \neq v_*$. Thus (5.3) gives $\mathbb{P}_1(\{\psi_{\text{adap}} = 1\} \triangle \{R_{v_*} \geq u_n\}) = o(1)$, and therefore

$$\mathbb{P}_1(\psi_{\text{adap}} = 1) = \mathbb{P}_1(R_{v_*} \geq u_n) + o(1).$$

If $\|\mathbf{a}\|_2 - \sqrt{2\log n} = O(1)$, then $\|\mathbf{a}\|_2 \rightarrow \infty$, so Lemma 5.2 and $u_n - \sqrt{2\log n} = o(1)$ give

$$\begin{aligned}\mathbb{P}_1(R_{v_*} \geq u_n) &= \mathbb{P}_1(R_{v_*} - \|\mathbf{a}\|_2 \geq u_n - \|\mathbf{a}\|_2) \\ &= \Phi(\|\mathbf{a}\|_2 - u_n) + o(1) = \Phi(\|\mathbf{a}\|_2 - \sqrt{2\log n}) + o(1).\end{aligned}$$

If $\|\mathbf{a}\|_2 - \sqrt{2\log n} \rightarrow \pm\infty$, the bound $R_{v_*} = \|\mathbf{a}\|_2 + O_{\mathbb{P}_1}(1)$ gives power tending to 1 or 0, respectively. Combining the above, we obtain

$$\mathbb{P}_1(\psi_{\text{adap}} = 1) = \Phi(\|\mathbf{a}\|_2 - \sqrt{2\log n}) + o(1).$$

This finishes the proof. $\square$

5.1 Hub recovery

Recall that $\hat{v}_{\text{adap}} \in \arg \max_v R_v$ selects the vertex with the largest statistic. The procedure reports this vertex only when the test rejects.

Proof of Proposition 1.7. We first prove (ii). If the test rejects and $\hat{v}_{\text{adap}} \neq v_*$, then a background vertex has statistic at least u_n . Equation (5.3) shows that this event has probability $o(1)$.

For (i), Theorem 1.6 and (ii) give

$$\begin{aligned}\mathbb{P}_1(\hat{v}_{\text{adap}} = v_*) &\geq \mathbb{P}_1(\psi_{\text{adap}} = 1, \hat{v}_{\text{adap}} = v_*) \\ &= \mathbb{P}_1(\psi_{\text{adap}} = 1) - \mathbb{P}_1(\psi_{\text{adap}} = 1, \hat{v}_{\text{adap}} \neq v_*) \\ &= \Phi(\|\mathbf{a}\|_2 - \sqrt{2\log n}) + o(1).\end{aligned}$$

Corollary 1.5 gives the matching upper bound for every estimator, including those that know the star sizes, so $0 \leq \sup_{\tilde{v}} \mathbb{P}_1(\tilde{v} = v_*) - \mathbb{P}_1(\hat{v}_{\text{adap}} = v_*) = o(1)$. This proves (i). $\square$

5.2 Estimating the signal after identifying the hub

By Lemma 5.2, the standardized degree vector $\mathbf{Y}_{v_*}$ at the hub equals $\mathbf{a}$ plus noise of order one. Hence, we estimate $\mathbf{a}$ using the standardized degree vector at the selected vertex, replacing negative entries by zero. Recall that $E_n = \{\psi_{\text{adap}} = 1\}$ is the event that the test rejects the null hypothesis. Then, on E_n ,

$$\hat{\mathbf{a}} = (\mathbf{Y}_{\hat{v}_{\text{adap}}})_+.$$

Proof of Theorem 1.8. Theorem 1.6 and Proposition 1.7(ii) give

$$\liminf_{n \rightarrow \infty} \mathbb{P}_1(E_n) > 0, \quad \text{and} \quad \mathbb{P}_1(E_n, \hat{v}_{\text{adap}} \neq v_*) = o(1).$$

On $E_n \cap \{\hat{v}_{\text{adap}} = v_*\}$, taking positive parts cannot increase the distance to the nonnegative vector $\mathbf{a}$, so

$$\|\hat{\mathbf{a}} - \mathbf{a}\|_2 \leq \|\mathbf{Y}_{v_*} - \mathbf{a}\|_2.$$

Thus, for any $M > 0$,

$$\mathbb{P}_1(\|\hat{\mathbf{a}} - \mathbf{a}\|_2 > M \mid E_n) \leq \frac{\mathbb{P}_1(E_n, \hat{v}_{\text{adap}} \neq v_*) + \mathbb{P}_1(\|\mathbf{Y}_{v_*} - \mathbf{a}\|_2 > M)}{\mathbb{P}_1(E_n)}.$$

The denominator stays bounded away from zero. The first numerator term tends to zero, and Lemma 5.2 makes the second term arbitrarily small for sufficiently large M . This proves the asserted bounded error under the conditional law. $\square$

Further consequences Under the conditional law in Theorem 1.8, the reverse triangle inequality gives

$$|\|\hat{\mathbf{a}}\|_2 - \|\mathbf{a}\|_2| \leq \|\hat{\mathbf{a}} - \mathbf{a}\|_2 = O_{\mathbb{P}}(1).$$

Since $\|\mathbf{a}\|_2 \sim \sqrt{2\log n} \rightarrow \infty$ in the critical window, this bounded absolute error gives vanishing relative error for the total signal strength. It also gives consistent estimation of the signal direction. Indeed, $\|\hat{\mathbf{a}}\|_2 \geq u_n > 0$ on E_n , and

$$\left\| \frac{\hat{\mathbf{a}}}{\|\hat{\mathbf{a}}\|_2} - \frac{\mathbf{a}}{\|\mathbf{a}\|_2} \right\|_2 \leq \frac{2\|\hat{\mathbf{a}} - \mathbf{a}\|_2}{\|\mathbf{a}\|_2} = O_{\mathbb{P}}\left(\frac{1}{\|\mathbf{a}\|_2}\right) = o_{\mathbb{P}}(1).$$

For an individual layer with $a_\ell \rightarrow \infty$,

$$\frac{\hat{k}_\ell}{k_\ell} = \frac{\hat{a}_\ell}{a_\ell} = 1 + o_{\mathbb{P}}(1).$$

Note that this relative accuracy is guaranteed for layers whose SNR a_ℓ diverges. If a_ℓ stays bounded, the $O_{\mathbb{P}}(1)$ error in (1.15) need not be small relative to a_ℓ , even when the total SNR $\|\mathbf{a}\|_2$ diverges.

When $\|\mathbf{a}\|_2 - \sqrt{2\log n} \rightarrow +\infty$, Theorem 1.6 and Proposition 1.7 give $\mathbb{P}_1(E_n \cap \{\hat{v}_{\text{adapt}} = v_*\}) = 1 - o(1)$. The same bound by $\|\mathbf{Y}_{v_*} - \mathbf{a}\|_2$ therefore holds with probability $1 - o(1)$. By Lemma 5.2, the bounds for $\hat{\mathbf{a}}$ and its norm hold without conditioning, as does the relative star size bound for layers with $a_\ell \rightarrow \infty$.

When $\|\mathbf{a}\|_2 - \sqrt{2\log n} \rightarrow -\infty$, Theorem 1.6 gives $\mathbb{P}_1(E_n) = o(1)$. On E_n , the estimation error diverges because

$$\|\hat{\mathbf{a}} - \mathbf{a}\|_2 \geq \|\hat{\mathbf{a}}\|_2 - \|\mathbf{a}\|_2 \geq u_n - \|\mathbf{a}\|_2 \rightarrow \infty.$$

Statement on AI use

ChatGPT 6.1 was used to assist with mathematical arguments and calculations in extending the $L = 1$ analysis to the multiplex setting. In particular, it proposed the statistic R_v in (1.12) for the case of unknown star sizes, as well as its proof of optimality. The authors rewrote the proofs, and take full responsibility for this work.

Acknowledgments

WP supported in part by NSF grant CCF-2309708.

References

- E. Abbe. Community detection and stochastic block models: recent developments. *Journal of Machine Learning Research*, 18(177):1–86, 2018.
- E. Abbe, S. Li, and A. Sly. Proof of the contiguity conjecture and lognormal limit for the symmetric perceptron. In *2021 IEEE 62nd Annual Symposium on Foundations of Computer Science (FOCS)*, pages 327–338. IEEE, 2022.

- L. Addario-Berry, N. Broutin, L. Devroye, and G. Lugosi. On combinatorial testing problems. *The Annals of Statistics*, 38(5):3063–3092, 2010. doi: 10.1214/10-AOS817. URL <https://doi.org/10.1214/10-AOS817>.
- L. Addario-Berry, C. Fontaine, R. Khanfir, L.-R. Langevin, and S. Têtu. Optimal root recovery for uniform attachment trees and d -regular growing trees. *arXiv preprint arXiv:2411.18614*, 2024. URL <https://arxiv.org/abs/2411.18614>.
- L. Addario-Berry, O. Angel, G. Lugosi, M. Z. Rácz, and T. Schramm. The statistical threshold for planted matchings and spanning trees. *arXiv preprint arXiv:2602.07669*, 2026. URL <https://arxiv.org/abs/2602.07669>.
- N. Alon, M. Krivelevich, and B. Sudakov. Finding a large hidden clique in a random graph. *Random Structures & Algorithms*, 13(3–4):457–466, 1998. doi: 10.1002/(SICI)1098-2418(199810/12)13:3/4<457::AID-RSA14>3.0.CO;2-W.
- E. Arias-Castro and N. Verzelen. Community detection in dense random networks. *The Annals of Statistics*, pages 940–969, 2014.
- V. Bagaria, J. Ding, D. Tse, Y. Wu, and J. Xu. Hidden hamiltonian cycle recovery via linear programming. *Operations research*, 68(1):53–70, 2020.
- D. Banerjee and Z. Ma. Optimal signal detection in some spiked random matrix models: likelihood ratio tests and linear spectral statistics. *The Annals of Statistics*, 50(4):1910–1932, 2022. ISSN 0090-5364,2168-8966. doi: 10.1214/21-aos2150. URL <https://doi.org/10.1214/21-aos2150>.
- A. Bhaskara, M. Charikar, E. Chlamtac, U. Feige, and A. Vijayaraghavan. Detecting high log-densities—an $O(n^{1/4})$ approximation for densest k -subgraph. In *STOC’10—Proceedings of the 2010 ACM International Symposium on Theory of Computing*, pages 201–210. ACM, New York, 2010. ISBN 978-1-60558-817-9.
- S. Boccaletti, G. Bianconi, R. Criado, C. I. del Genio, J. Gómez-Gardeñes, M. Romance, I. Sendiña-Nadal, Z. Wang, and M. Zanin. The structure and dynamics of multilayer networks. *Physics Reports*, 544(1):1–122, 2014. doi: 10.1016/j.physrep.2014.07.001.
- E. Bolthausen and A.-S. Sznitman. On Ruelle’s probability cascades and an abstract cavity method. *Communications in Mathematical Physics*, 197(2):247–276, 1998. doi: 10.1007/s002200050450.
- J. Borcea, P. Brändén, and T. M. Liggett. Negative dependence and the geometry of polynomials. *J. Amer. Math. Soc.*, 22(2):521–567, 2009. ISSN 0894-0347,1088-6834. doi: 10.1090/S0894-0347-08-00618-8. URL <https://doi.org/10.1090/S0894-0347-08-00618-8>.
- A. Bovier. *Statistical Mechanics of Disordered Systems: A Mathematical Perspective*. Cambridge University Press, Cambridge, 2006. ISBN 9780521849913. doi: 10.1017/CBO9780511616808.
- A. Bovier and I. Kurkova. Derrida’s generalized random energy models. I. models with finitely many hierarchies. *Annales de l’Institut Henri Poincaré Probabilités et Statistiques*, 40(4):439–480, 2004. doi: 10.1016/j.anihpb.2003.09.002.
- A. Bovier, I. Kurkova, and M. Löwe. Fluctuations of the free energy in the REM and the p -spin SK models. *The Annals of Probability*, 30(2):605–651, 2002. doi: 10.1214/aop/1023481004.

- S. Bubeck, L. Devroye, and G. Lugosi. Finding adam in random growing trees. *Random Structures & Algorithms*, 50(2):158–172, 2017.
- D. Capocaccia, M. Cassandro, and P. Picco. On the existence of thermodynamics for the generalized random energy model. *Journal of Statistical Physics*, 46(3-4):493–505, 1987. doi: 10.1007/BF01013370.
- A. Coja-Oghlan, F. Krzakala, W. Perkins, and L. Zdeborová. Information-theoretic thresholds from the cavity method. *Advances in Mathematics*, 333:694–795, 2018.
- M. De Domenico, A. Solé-Ribalta, E. Omodei, S. Gómez, and A. Arenas. Ranking in interconnected multilayer networks reveals versatile nodes. *Nature Communications*, 6:6868, 2015. doi: 10.1038/ncomms7868.
- A. Decelle, F. Krzakala, C. Moore, and L. Zdeborová. Asymptotic analysis of the stochastic block model for modular networks and its algorithmic applications. *Physical Review E—Statistical, Nonlinear, and Soft Matter Physics*, 84(6):066106, 2011.
- B. Derrida. Random-energy model: An exactly solvable model of disordered systems. *Physical Review B*, 24(5):2613–2626, 1981. doi: 10.1103/PhysRevB.24.2613.
- B. Derrida and E. Gardner. Solution of the generalised random energy model. *Journal of Physics C: Solid State Physics*, 19(13):2253–2274, 1986.
- J. Ding, Y. Wu, J. Xu, and D. Yang. Consistent recovery threshold of hidden nearest neighbor graphs. In *Conference on Learning Theory*, pages 1540–1553. PMLR, 2020.
- J. Ding, Y. Wu, J. Xu, and D. Yang. The planted matching problem: Sharp threshold and infinite-order phase transition. *Probability Theory and Related Fields*, 187(1-2):1–71, 2023.
- H. Döring, S. Jansen, and K. Schubert. The method of cumulants for the normal approximation. *Probab. Surv.*, 19:185–270, 2022. ISSN 1549-5787. doi: 10.1214/22-ps7. URL <https://doi.org/10.1214/22-ps7>.
- A. El Alaoui, F. Krzakala, and M. Jordan. Fundamental limits of detection in the spiked Wigner model. *The Annals of Statistics*, 48(2):863–885, 2020. ISSN 0090-5364,2168-8966. doi: 10.1214/19-AOS1826. URL <https://doi.org/10.1214/19-AOS1826>.
- D. Elimelech and W. Huleihel. Detecting arbitrary planted subgraphs in random graphs. In N. Haghtalab and A. Moitra, editors, *Proceedings of Thirty Eighth Conference on Learning Theory*, volume 291 of *Proceedings of Machine Learning Research*, pages 1691–1798. PMLR, 2025. URL <https://proceedings.mlr.press/v291/elimelech25a.html>.
- J. Gaudio, C. Sandon, J. Xu, and D. Yang. Phase transitions in planted k -factor recovery. *The Annals of Applied Probability*, 2026. URL <https://arxiv.org/abs/2503.08984>. Accepted for publication; preprint arXiv:2503.08984.
- A. Halu, R. J. Mondragón, P. Panzarasa, and G. Bianconi. Multiplex PageRank. *PLOS ONE*, 8(10):e78293, 2013. doi: 10.1371/journal.pone.0078293.

- Z. Hou and J. Liu. Recovery thresholds for hidden weighted sparse graphs (extended abstract). In S. Hanneke and T. Lattimore, editors, *Proceedings of Thirty Ninth Conference on Learning Theory*, volume 336 of *Proceedings of Machine Learning Research*, pages 3280–3284. PMLR, 2026. URL <https://proceedings.mlr.press/v336/hou26a.html>.
- S. Janson. The numbers of spanning trees, Hamilton cycles and perfect matchings in a random graph. *Combin. Probab. Comput.*, 3(1):97–126, 1994. ISSN 0963-5483,1469-2163. doi: 10.1017/S0963548300001012. URL <https://doi.org/10.1017/S0963548300001012>.
- M. Jerrum. Large cliques elude the metropolis process. *Random Structures & Algorithms*, 3(4):347–360, 1992. doi: 10.1002/rsa.3240030402.
- Z. Kabluchko and A. Klimovsky. Complex random energy model: Zeros and fluctuations. *Probability Theory and Related Fields*, 158(1–2):159–196, 2014. doi: 10.1007/s00440-013-0480-5.
- A. Kirkley. Identifying hubs in directed networks. *Phys. Rev. E*, 109:034310, Mar 2024. doi: 10.1103/PhysRevE.109.034310. URL <https://link.aps.org/doi/10.1103/PhysRevE.109.034310>.
- M. Kivelä, A. Arenas, M. Barthélemy, J. P. Gleeson, Y. Moreno, and M. A. Porter. Multilayer networks. *Journal of Complex Networks*, 2(3):203–271, 2014. doi: 10.1093/comnet/cnu016.
- E. L. Lehmann and J. P. Romano. *Testing Statistical Hypotheses*. Springer Texts in Statistics. Springer, New York, 3rd edition, 2005. doi: 10.1007/0-387-27605-X.
- L. Massoulié, L. Stephan, and D. Towsley. Planting trees in graphs, and finding them back. In *Conference on Learning Theory*, pages 2341–2371. PMLR, 2019.
- M. Moharrami, C. Moore, and J. Xu. The planted matching problem: Phase transitions and exact results. *The Annals of Applied Probability*, 31(6):2663–2720, 2021.
- M. Moharrami, C. Moore, and J. Xu. The planted spanning tree problems: Exact overlap characterization via local weak convergence extended abstract. In *Proceedings of Thirty Eighth Conference on Learning Theory*, volume 291 of *Proceedings of Machine Learning Research*, pages 4166–4167. PMLR, 30 Jun–04 Jul 2025. URL <https://proceedings.mlr.press/v291/moharrami25a.html>.
- A. Moitra and A. S. Wein. Precise error rates for computationally efficient testing. *The Annals of Statistics*, 53(2):854 – 878, 2025. doi: 10.1214/25-AOS2490. URL <https://doi.org/10.1214/25-AOS2490>.
- E. Mossel, J. Neeman, and A. Sly. Reconstruction and estimation in the planted partition model. *Probability Theory and Related Fields*, 162(3):431–461, 2015.
- E. Mossel, J. Niles-Weed, Y. Sohn, N. Sun, and I. Zadik. Sharp thresholds in inference of planted subgraphs. In *The Thirty Sixth Annual Conference on Learning Theory*, pages 5573–5577. PMLR, 2023.
- E. Mossel, A. Sly, and Y. Sohn. Weak recovery, hypothesis testing, and mutual information in stochastic block models and planted factor graphs. In *Proceedings of the 57th Annual ACM Symposium on Theory of Computing*, pages 2062–2073, 2025.

- E. Olivieri and P. Picco. On the existence of thermodynamics for the random energy model. *Communications in Mathematical Physics*, 96(1):125–144, 1984. doi: 10.1007/BF01217351.
- W. Perkins. The forgetfulness of balls and bins. *Random Structures & Algorithms*, 42(2):250–267, 2013.
- M. Z. Rácz and S. Bubeck. Basic models and questions in statistical network analysis. *Stat. Surv.*, 11:1–47, 2017. ISSN 1935-7516. doi: 10.1214/17-SS117. URL <https://doi.org/10.1214/17-SS117>.
- C. Rahmede, J. Iacovacci, A. Arenas, and G. Bianconi. Centralities of nodes and influences of layers in large multiplex networks. *Journal of Complex Networks*, 6(5):733–752, 2018. doi: 10.1093/comnet/cnx050.
- A. Reiffers-Masson and V. Labatut. Opinion-based centrality in multiplex networks: A convex optimization approach. *Network Science*, 5(2):213–234, 2017. doi: 10.1017/nws.2017.7.
- R. Rudzkis, L. Saulis, and V. Statulevičius. A general lemma on probabilities of large deviations. *Litovsk. Mat. Sb.*, 18(2):99–116, 217, 1978. ISSN 0132-2818.
- D. Ruelle. A mathematical reformulation of Derrida’s REM and GREM. *Communications in Mathematical Physics*, 108(2):225–239, 1987. doi: 10.1007/BF01210613.
- D. Sherrington and S. Kirkpatrick. Solvable model of a spin-glass. *Physical Review Letters*, 35(26):1792–1796, 1975. doi: 10.1103/PhysRevLett.35.1792.
- R. Vershynin. *High-Dimensional Probability: An Introduction with Applications in Data Science*, volume 47 of *Cambridge Series in Statistical and Probabilistic Mathematics*. Cambridge University Press, 1st edition, 2018. doi: 10.1017/9781108231596.
- M. J. Wainwright. *High-dimensional statistics: A non-asymptotic viewpoint*, volume 48. Cambridge university press, 2019.
- T. L. H. Wee and C. Mao. Cluster expansion of the log-likelihood ratio: Optimal detection of planted matchings. *arXiv preprint arXiv:2512.14567*, 2025.
- X. Yu, I. Zadik, and P. Zhang. Counting stars is constant-degree optimal for detecting any planted subgraph. *Math. Stat. Learn.*, 8(1-2):105–164, 2025. ISSN 2520-2316,2520-2324. doi: 10.4171/msl/51. URL <https://doi.org/10.4171/msl/51>.

A Proofs for Section 2

In this appendix we collect the remaining proofs for Section 2.

Definition A.1. A random variable X has the **hypergeometric** distribution $\text{Hypergeom}(N, K, n)$ if X is the number of objects with a desired property in a sample without replacement of size n from a population of size N , where K objects in the population have that desired property.

We turn to the proof of Lemma 2.1. We first establish an auxiliary result, that the degrees under the null and alternative concentrate on the corresponding layer's mean, simultaneously across all layers and all vertices.

Lemma A.2. *Suppose (1.3) holds. Then there is a constant C_0 such that, under either $\mathbb{P}_0$ or $\mathbb{P}_1$, with probability $1 - o(1)$,*

$$|D_v^{(\ell)} - \mu_\ell| \leq C_0 \sigma_\ell \sqrt{\log n} + k_\ell, \quad \text{for every } v \in [n], \ell \in [L]. \quad (\text{A.1})$$

Proof of Lemma A.2. We first obtain a tail bound for a hypergeometric variable H with mean $\nu > 0$ (Definition A.1). By Lemma B.3, H has the same distribution as a deterministic integer plus independent Bernoulli variables. Their total variance is at most the sum of their means, which is at most ν . Bernstein's inequality (see e.g. Vershynin, 2018, Theorem 2.8.4) therefore gives

$$\mathbb{P}(|H - \nu| \geq u) \leq 2 \exp \left\{ -\frac{u^2}{2(\nu + u/3)} \right\} \leq 2 \exp \left\{ -c \min \left(\frac{u^2}{\nu}, u \right) \right\}, \quad u > 0, \quad (\text{A.2})$$

for a universal constant $c > 0$. The case $\nu = 0$ is deterministic.

Under $\mathbb{P}_0$, the degree law is $\text{Hypergeom}(N, n-1, m_\ell)$, with mean μ_ℓ and variance $\sigma_\ell^2 \sim \mu_\ell$. Taking $u = C_0 \sigma_\ell \sqrt{\log n}$ in (A.2), and choosing C_0 large enough, makes each failure probability $o(n^{-2})$. The claim for $\mathbb{P}_0$ then follows by a union bound.

Under $\mathbb{P}_1$, condition on the hub and on all the planted edges. Let $s_{v,\ell}$ be the number of planted edges incident to v in layer ℓ . Then

$$D_v^{(\ell)} - s_{v,\ell} \sim \text{Hypergeom}(N - k_\ell, n - 1 - s_{v,\ell}, m_\ell - k_\ell).$$

Writing $p'_\ell = (m_\ell - k_\ell)/(N - k_\ell)$, the hypergeometric mean formula and $0 \leq s_{v,\ell} \leq k_\ell$ give

$$\begin{aligned} \left| \mathbb{E}_1[D_v^{(\ell)} \mid s_{v,\ell}] - \mu_\ell \right| &= \left| s_{v,\ell} + (n - 1 - s_{v,\ell})p'_\ell - \frac{(n-1)m_\ell}{N} \right| \\ &= (1 - p'_\ell) \left| s_{v,\ell} - \frac{(n-1)k_\ell}{N} \right| \leq k_\ell. \end{aligned}$$

To control fluctuations of $D_v^{(\ell)}$ around the conditional mean, we will apply (A.2) to the hypergeometric random variable $D_v^{(\ell)} - s_{v,\ell}$. Its mean satisfies

$$\mathbb{E}_1[D_v^{(\ell)} - s_{v,\ell} \mid s_{v,\ell}] = (n - 1 - s_{v,\ell})p'_\ell \leq (n-1)\frac{m_\ell}{N} = \mu_\ell.$$

Taking $u = C_0 \sigma_\ell \sqrt{\log n}$ in (A.2) and applying a union bound therefore gives, with conditional probability $1 - o(1)$, simultaneously for all $v \in [n]$ and $\ell \in [L]$,

$$\begin{aligned} |D_v^{(\ell)} - \mu_\ell| &\leq \left| D_v^{(\ell)} - \mathbb{E}_1[D_v^{(\ell)} \mid s_{v,\ell}] \right| + \left| \mathbb{E}_1[D_v^{(\ell)} \mid s_{v,\ell}] - \mu_\ell \right| \\ &\leq C_0 \sigma_\ell \sqrt{\log n} + k_\ell. \end{aligned}$$

The estimate is uniform in the hub and planted edges, so the claim follows under $\mathbb{P}_1$. $\square$

Proof of Lemma 2.1. We claim that the SNR vector $\mathbf{a}$ admits the approximation

$$\|\mathbf{a}\|_2^2 = \sum_{\ell=1}^L \frac{k_\ell^2}{\mu_\ell} + o(1) = \sum_{\ell=1}^L \frac{nk_\ell^2}{2m_\ell} + o(1). \quad (\text{A.3})$$

Indeed, (1.2) gives an error in layer ℓ of $O(k_\ell^2/n + (k_\ell^2/\mu_\ell)/n) = o(1)$ by (1.3).

For the rest of the proof, we work on the event that (A.1) holds for every $v \in [n]$ and $\ell \in [L]$. By Lemma A.2, this event has probability $1 - o(1)$ under both $\mathbb{P}_0$ and $\mathbb{P}_1$. The estimates below are uniform over vertices v and layers ℓ on this event.

For each layer ℓ , the corresponding factor in the likelihood ratio (2.1) can be written as

$$\frac{\binom{N}{m_\ell} \binom{D_v^{(\ell)}}{k_\ell}}{\binom{n-1}{k_\ell} \binom{N-k_\ell}{m_\ell-k_\ell}} = C_\ell \frac{(D_v^{(\ell)})_{k_\ell}}{\mu_\ell^{k_\ell}}, \quad \text{where} \quad C_\ell = \frac{(N)_{k_\ell}}{(m_\ell)_{k_\ell}} \frac{\mu_\ell^{k_\ell}}{(n-1)_{k_\ell}},$$

where $(x)_k = x(x-1)\cdots(x-k+1)$ denotes the falling factorial.

Expanding the falling factorials in C_ℓ and using $\mu_\ell = (n-1)m_\ell/N$, $k_\ell^2/n = o(1)$ and $k_\ell^2/m_\ell = O(\log n/n)$, we obtain

$$\log C_\ell = O\left(\frac{k_\ell^2}{n} + \frac{k_\ell^2}{m_\ell}\right) = o(1). \quad (\text{A.4})$$

Set $\Delta_{v,\ell} = D_v^{(\ell)} - \mu_\ell$. By (A.1), for every $0 \leq j < k_\ell$,

$$\left| \frac{\Delta_{v,\ell} - j}{\mu_\ell} \right| \leq \frac{C_0 \sigma_\ell \sqrt{\log n} + 2k_\ell}{\mu_\ell} = O\left(\frac{\log n}{k_\ell}\right) = o(1).$$

We can therefore expand each logarithm using $\log(1+x) = x + O(x^2)$:

$$\begin{aligned} \log \frac{(D_v^{(\ell)})_{k_\ell}}{\mu_\ell^{k_\ell}} &= \sum_{j=0}^{k_\ell-1} \log \left(1 + \frac{\Delta_{v,\ell} - j}{\mu_\ell} \right) \\ &= \frac{k_\ell \Delta_{v,\ell}}{\mu_\ell} - \frac{k_\ell(k_\ell-1)}{2\mu_\ell} + O\left(\frac{k_\ell(\sigma_\ell^2 \log n + k_\ell^2)}{\mu_\ell^2}\right) \\ &= \frac{k_\ell \Delta_{v,\ell}}{\mu_\ell} - \frac{k_\ell^2}{2\mu_\ell} + o(1) \\ &= a_\ell Y_v^{(\ell)} - \frac{a_\ell^2}{2} + o(1), \end{aligned} \quad (\text{A.5})$$

where in the third line we used that the Taylor remainder is $O((\log n)^2/k_\ell) = o(1)$, and $k_\ell/\mu_\ell = O(\log n/k_\ell) = o(1)$, by (1.3), and where the final equality uses $k_\ell \Delta_{v,\ell}/\mu_\ell = a_\ell Y_v^{(\ell)}$ from (1.5) and (1.4), and $k_\ell^2/\mu_\ell = a_\ell^2 + o(1)$ from the error estimate for each layer following (A.3).

Since L is fixed, summing over layers preserves the uniform $o(1)$ error. Combining (A.5) with (A.4) and the definition of S_v in (1.6), the logarithm of each vertex's product in (2.1) is $\|\mathbf{a}\|_2 S_v - \|\mathbf{a}\|_2^2/2 + o(1)$. Exponentiating gives a multiplicative $1 + o(1)$ error, uniform over vertices, that can be taken outside the average in (2.1). This proves (2.2). $\square$

B Proofs for Section 3

In this appendix we prove Lemma 3.2, stated in Section 3, and its conditional version, Lemma B.7. The latter controls the effect of conditioning on the layer degrees of a fixed number of other vertices. All probabilities and expectations involving graph degrees in this appendix are under the null distribution. We use the notation $S_v(\mathbf{u})$ and $\delta_{\mathbf{u}}$ introduced before Lemma 3.2.

B.1 Gaussian estimates for weighted Bernoulli sums

We use the strongly Rayleigh property of the Hypergeometric distribution to bound its cumulants and obtain Gaussian moderate deviations.

B.1.1 Strongly Rayleigh property

For a nonnegative integer-valued random variable Y with finite support, define its probability generating function (PGF) to be $\lambda \mapsto \mathbb{E}[\lambda^Y]$ for $\lambda \in \mathbb{C}$. A polynomial is *real stable* if it has real coefficients and is nonzero whenever all its arguments have positive imaginary part. We say that a distribution on $\{0, 1\}^N$ is *strongly Rayleigh* (cf. Borcea et al., 2009, Definitions 2.9 and 2.10) if its multivariate PGF is real stable.

Fact B.1 (Essentially in Borcea et al., 2009). *The uniform distribution over subsets of a fixed size is strongly Rayleigh. Consequently, the Hypergeometric distribution has a real stable PGF.*

Proof of Fact B.1. Consider the uniform distribution over subsets of size m of $[N]$, i.e. $\mathbb{P}[\mathbf{X} = \mathbf{1}_S] = \binom{N}{m}^{-1}$ for any $S \subset [N]$ with $|S| = m$, where $\mathbf{X} = (X_1, \dots, X_N)$ is a collection of $\{0, 1\}$ -valued random variables with X_i indicating inclusion of the i -th element. Its PGF is

$$g_{\mathbf{X}}(z_1, \dots, z_N) = \frac{1}{\binom{N}{m}} e_m(z_1, \dots, z_N),$$

where $e_m(z_1, \dots, z_N) = \sum_{\substack{S \subseteq [N] \\ |S|=m}} \prod_{i \in S} z_i$ is the m -th elementary symmetric polynomial. Define

$$F(t, z_1, \dots, z_N) = \prod_{i=1}^N (t + z_i) = \sum_{m=0}^N e_m(z_1, \dots, z_N) t^{N-m}.$$

Note that $F(t, z_1, \dots, z_N)$ is real stable. On the other hand

$$e_m(z_1, \dots, z_N) = \frac{1}{(N-m)!} \partial_t^{N-m} F(t, z_1, \dots, z_N) \Big|_{t=0}.$$

By Borcea et al. (2009, Proposition 3.1), real stability is closed under taking derivatives and specializing to real constants, thus e_m inherits the real stability property from $F(t, z_1, \dots, z_N)$.

Suppose without loss of generality that the first K indicators $X_1, \dots, X_K$ are marked as desirable objects. Then $Y = \sum_{i=1}^K X_i \sim \text{Hypergeom}(N, K, m)$. Let $g_Y(\lambda) = \mathbb{E}\lambda^Y$ denote the PGF of Y . We have the relation

$$g_Y(\lambda) = g_{\mathbf{X}}(\underbrace{\lambda, \dots, \lambda}_{K \text{ times}}, 1, \dots, 1).$$

By Borcea et al. (2009, Proposition 3.1, Definition 2.8), since g_Y is the projection and diagonalization of $g_{\mathbf{X}}$ onto the first K coordinates, it is a real stable polynomial. $\square$

A univariate PGF is real stable if and only if it has all real roots. Coupled with the fact that the Hypergeometric PGF has nonnegative coefficients, we arrive at the following.

Corollary B.2. *Let $Y \sim \text{Hypergeom}(N, K, m)$ where $0 \leq K, m \leq N$. Then the PGF of Y has only real nonpositive roots. More precisely, writing $r_0 = \max\{0, m + K - N\}$, its zero root has multiplicity r_0 , and the polynomial $z^{-r_0}g_Y(z)$ has strictly negative roots.*

Our next result gives a Bernoulli factorization for distributions with real stable PGFs.

Lemma B.3. *Let H be a nonnegative integer-valued random variable with finite support whose PGF has only real roots. Then, for some deterministic integers $r_0, J \geq 0$ and independent Bernoulli variables $\xi_1, \dots, \xi_J$,*

$$H \stackrel{d}{=} r_0 + \sum_{j=1}^J \xi_j. \quad (\text{B.1})$$

Proof of Lemma B.3. Write r_0 for the multiplicity of the zero root and $\zeta_1, \dots, \zeta_J$ for the strictly negative roots, counted with multiplicity. Factoring the PGF and using $g_H(1) = 1$ gives

$$g_H(z) = z^{r_0} \prod_{j=1}^J \frac{z - \zeta_j}{1 - \zeta_j} = z^{r_0} \prod_{j=1}^J (1 - p_j + p_j z), \quad p_j = (1 - \zeta_j)^{-1} \in (0, 1).$$

This is the PGF of the claimed sum. □

B.1.2 Cumulant bounds

We begin with a bound on cumulants which leads to asymptotics of the moment and cumulant generating functions. These estimates will be used below to approximate truncated exponential moments. For any random variable Y , we denote by $\kappa_r(Y)$ its r^{th} cumulant.

Lemma B.4. *Let H be a nonnegative integer-valued random variable with finite support and a real-rooted PGF. Then there exists a universal constant $C_0 \geq 1$ such that*

$$|\kappa_r(H)| \leq C_0^r r! \text{Var}(H), \quad r \geq 3. \quad (\text{B.2})$$

In particular, if $\mu = \mathbb{E}H$, $\sigma^2 = \text{Var}(H) > 0$, and $Y = (H - \mu)/\sigma$, then

$$|\kappa_r(Y)| \leq \frac{C_0^r r!}{\sigma^{r-2}}, \quad \text{for all } r \geq 3. \quad (\text{B.3})$$

Furthermore, for $|t| = o(\sigma^{1/3})$ as $\sigma \rightarrow \infty$ we have $\log \mathbb{E}[e^{tY}] = \frac{t^2}{2} + o(1)$ and $\mathbb{E}[e^{tY}] = (1 + o(1))e^{t^2/2}$.

Proof of Lemma B.4. First assume $\sigma^2 = \text{Var}(H) > 0$. Our first step is to establish the following bound for the cumulants:

$$|\kappa_r(Y)| \leq \frac{2}{\sigma^{r-2}} \sum_{q=1}^r (q-1)! S(r, q) \quad (\text{B.4})$$

where $S(r, q)$ denotes Stirling numbers of the second kind. To this end, write $g(z) = \mathbb{E}[z^H]$. By Lemma B.3, with $\{\zeta_j\}_{j=1}^J$ denoting its strictly negative roots and r_0 the multiplicity of its zero root, we have the factorization

$$g(z) = z^{r_0} \prod_{j=1}^J \frac{z - \zeta_j}{1 - \zeta_j}. \quad (\text{B.5})$$

The normalization is deduced from $g(1) = 1$. Furthermore, since $\zeta_j < 0$, by setting $a_j := -1/\zeta_j > 0$, we have

$$\frac{z - \zeta_j}{1 - \zeta_j} = \frac{1 + a_j z}{1 + a_j},$$

and (B.5) becomes

$$g(z) = z^{r_0} \prod_{j=1}^J \frac{1 + a_j z}{1 + a_j}. \quad (\text{B.6})$$

Define the moment and cumulant generating functions $M(t) := \mathbb{E}[e^{tH}]$ and $K(t) := \log M(t)$ respectively. Since H is integer-valued, we have the relations

$$M(t) = g(e^t), \quad K(t) = \log g(e^t), \quad \text{and} \quad \kappa_r(H) = K^{(r)}(0).$$

Plugging (B.6) into $K(t)$ yields

$$K(t) = \log g(e^t) = r_0 t + \sum_{j=1}^J \left(\log(1 + a_j e^t) - \log(1 + a_j) \right) = r_0 t + \sum_{j=1}^J \phi_{a_j}(t), \quad (\text{B.7})$$

where $\phi_a(t) := \log(1 + a e^t) - \log(1 + a)$. The deterministic term $r_0 t$ affects only the first cumulant. For $a > 0$, define

$$\rho := \frac{a}{1 + a} \in (0, 1),$$

so that $\phi_a(t) = \log(1 + \rho(e^t - 1))$. We have the expansions

$$\log(1 + u) = \sum_{q \geq 1} (-1)^{q+1} \frac{u^q}{q}, \quad \text{and} \quad (e^t - 1)^q = q! \sum_{\ell \geq q} S(\ell, q) \frac{t^\ell}{\ell!},$$

where $S(\ell, q)$ denotes Stirling numbers of the second kind. It follows that

$$\begin{aligned} \phi_a(t) &= \sum_{q \geq 1} (-1)^{q+1} \frac{\rho^q}{q} \left(q! \sum_{\ell \geq q} S(\ell, q) \frac{t^\ell}{\ell!} \right) \\ &= \sum_{q \geq 1} (-1)^{q+1} \rho^q (q-1)! \sum_{\ell \geq q} S(\ell, q) \frac{t^\ell}{\ell!}. \end{aligned}$$

Now, fix an integer $r \geq 3$. Reading off the coefficient of $t^r/r!$, we obtain the formula

$$\phi_a^{(r)}(0) = \sum_{q=1}^r (-1)^{q+1} \rho^q (q-1)! S(r, q). \quad (\text{B.8})$$

Now, observe that $\phi_a''(0) = \rho(1 - \rho)$. For $\rho \leq 1/2$, we can write

$$\begin{aligned} |\phi_a^{(r)}(0)| &\leq \rho \sum_{q=1}^r (q-1)! S(r, q) \\ &\leq 2\rho(1 - \rho) \sum_{q=1}^r (q-1)! S(r, q) = 2\phi_a''(0) \sum_{q=1}^r (q-1)! S(r, q). \end{aligned} \quad (\text{B.9})$$

For $\rho > 1/2$, use $\phi_a(t) = t + \phi_{1/a}(-t)$, whose Bernoulli parameter is $1 - \rho < 1/2$. Its derivatives of order $r \geq 2$ have the same absolute values, so (B.9) holds for every $\rho \in (0, 1)$. Putting everything together,

$$\begin{aligned} |\kappa_r(H)| &\leq \sum_{j=1}^J |\phi_{a_j}^{(r)}(0)| \\ &\leq 2 \left(\sum_{q=1}^r (q-1)! S(r, q) \right) \sum_{j=1}^J \phi_{a_j}''(0) = 2\sigma^2 \left(\sum_{q=1}^r (q-1)! S(r, q) \right). \end{aligned} \quad (\text{B.10})$$

Because $Y = (H - \mu)/\sigma$, we have $\kappa_r(Y) = \kappa_r(H)/\sigma^r$, so the desired (B.4) follows for all $r \geq 3$.

Next consider the cumulant generating function $K_Y(t)$ of Y :

$$K_Y(t) = \frac{t^2}{2} + \sum_{r \geq 3} \kappa_r(Y) \frac{t^r}{r!}. \quad (\text{B.11})$$

For notational convenience, define

$$A_r := \sum_{q=1}^r (q-1)! S(r, q)$$

so that (B.4) is $|\kappa_r(Y)| \leq 2A_r/\sigma^{r-2}$. Consider the exponential generating function

$$F(z) := \sum_{r \geq 1} A_r \frac{z^r}{r!}.$$

Using the identity $(e^z - 1)^q = q! \sum_{r \geq q} S(r, q) z^r / r!$ and swapping sums,

$$\begin{aligned} F(z) &= \sum_{r \geq 1} \sum_{q=1}^r (q-1)! S(r, q) \frac{z^r}{r!} = \sum_{q \geq 1} (q-1)! \sum_{r \geq q} S(r, q) \frac{z^r}{r!} \\ &= \sum_{q \geq 1} (q-1)! \cdot \frac{(e^z - 1)^q}{q!} = \sum_{q \geq 1} \frac{(e^z - 1)^q}{q} = -\log(1 - (e^z - 1)) \\ &= -\log(2 - e^z). \end{aligned}$$

The function $F(z) = -\log(2 - e^z)$ is analytic on the disk $|z| < \log 2$, since the nearest singularity occurs at $z = \log 2$ and its complex translates. Therefore the radius of convergence of F is $\log 2$. Thus, fix any $\rho \in (0, \log 2)$ and set $M_\rho := \sup_{|z|=\rho} |F(z)| < \infty$. By Cauchy's estimate,

$$\frac{A_r}{r!} \leq \frac{M_\rho}{\rho^r} \quad \text{for all } r \geq 1,$$

which proves the asserted bound on cumulants, both in (B.2) and in (B.3). Furthermore,

$$\begin{aligned} \left| K_Y(t) - \frac{t^2}{2} \right| &\leq |t|^2 \sum_{r \geq 3} \frac{2A_r}{r!} \left(\frac{|t|}{\sigma} \right)^{r-2} \\ &\leq 2M_\rho |t|^2 \sum_{r \geq 3} \rho^{-r} \left(\frac{|t|}{\sigma} \right)^{r-2} = 2M_\rho |t|^2 \rho^{-2} \sum_{k \geq 1} \left(\frac{|t|}{\rho\sigma} \right)^k. \end{aligned}$$

The logarithm in (B.11) is analytic near zero, and the coefficient bound extends its Taylor series to $|t| < \rho\sigma$. Exponentiating the series and using the identity theorem shows that it agrees there with the branch of $\log M_Y$ that vanishes at zero. Since $|t| = o(\sigma^{1/3})$, we have $|t|/(\rho\sigma) < 1$ and so

$$\left| K_Y(t) - \frac{t^2}{2} \right| \leq \frac{2M_\rho}{\rho^3} \frac{|t|^3/\sigma}{1 - |t|/(\rho\sigma)} = o(1). \quad (\text{B.12})$$

Exponentiating gives

$$M_Y(t) = \mathbb{E}[e^{tY}] = \exp\left(\frac{t^2}{2} + o(1)\right) = \exp\left(\frac{t^2}{2}\right) (1 + o(1)).$$

This finishes the proof when $\sigma^2 > 0$. If $\text{Var}(H) = 0$, then H is deterministic and all its cumulants of order at least two vanish, so (B.2) holds as well. $\square$

B.1.3 Weighted Bernoulli sums and exponential tilting

By Lemma B.3, each layer degree can be represented in distribution as a deterministic shift plus a sum of independent Bernoulli variables. Since the layers are independent, the projected degree $S_v(\mathbf{u})$ has a corresponding representation, with each Bernoulli summand from layer ℓ carrying weight u_ℓ/σ_ℓ . We extend the preceding cumulant bounds to these weighted sums, then use them to obtain Gaussian approximations for their distributions and truncated exponential moments.

In this section, denote

$$X = c + \sum_{j=1}^M w_j \xi_j, \quad (\text{B.13})$$

where ξ_j are independent Bernoulli r.v.'s. Put $\delta = \max_j |w_j|$ and $v = \text{Var}(X) > 0$.

Lemma B.5. *There exists a universal constant C such that*

$$\sup_{x \in \mathbb{R}} \left| \mathbb{P}\left(\frac{X - \mathbb{E}X}{\sqrt{v}} \leq x\right) - \Phi(x) \right| \leq C \frac{\delta}{\sqrt{v}}. \quad (\text{B.14})$$

Proof of Lemma B.5. Lemma B.4 and additivity of cumulants give

$$|\kappa_r(X)| \leq C_0^r r! \sum_j |w_j|^r \text{Var}(\xi_j) \leq C_0^r r! v \delta^{r-2}, \quad r \geq 3, \quad (\text{B.15})$$

where C_0 is universal. For $\mu = \mathbb{E}X$ and $K(u) = \log \mathbb{E}e^{uX}$, summing the cumulant series and its derivatives gives

$$K(u) = \mu u + \frac{vu^2}{2} + O(v\delta|u|^3), \quad (\text{B.16})$$

$$K'(u) = \mu + vu + O(v\delta|u|^2), \quad K''(u) = v + O(v\delta|u|), \quad (\text{B.17})$$

uniformly when $|u|\delta$ is sufficiently small. After normalization, the cumulant bound satisfies the hypotheses of Döring et al. (2022, Theorem 2.4) with $\Delta \asymp \sqrt{v}/\delta$, yielding the claimed error. $\square$

Lemma B.6. *Consider a family of weighted sums X of the form (B.13), whose representations may depend on n . Suppose that the following bounds hold uniformly over the family as $n \rightarrow \infty$:*

$$\delta(\log n)^{3/2} = o(1), \quad \mathbb{E}X = O(\delta), \quad \text{and} \quad \text{Var}(X) = 1 + O(\delta^2).$$

Then (3.4)–(3.6) hold with $S_v(\mathbf{u})$ replaced by X , uniformly over the family and for $|s|, |t| = O(\sqrt{\log n})$.

Proof of Lemma B.6. Write $K(u) = \log \mathbb{E}e^{uX}$. By (B.16), we have $K(s) = s^2/2 + o(1)$, which proves the first claim:

$$\mathbb{E}e^{sX} = e^{K(s)} = (1 + o(1))e^{s^2/2}. \quad (\text{B.18})$$

For the truncated moments, tilt the joint law of $\xi_1, \dots, \xi_M$ in (B.13) by

$$\frac{d\mathbb{P}^{(s)}}{d\mathbb{P}} = e^{sX - K(s)} = \prod_{j=1}^M \frac{e^{sw_j \xi_j}}{\mathbb{E}e^{sw_j \xi_j}}.$$

This product form shows that $\xi_1, \dots, \xi_M$ remain independent Bernoulli variables under $\mathbb{P}^{(s)}$. By (B.17), the mean and variance of X under this law satisfy

$$m_s = K'(s) = s + O(\delta(1 + s^2)), \quad \text{and} \quad \tau_s^2 = K''(s) = 1 + O(\delta^2 + \delta|s|).$$

Thus $(X - m_s)/\tau_s$ is standardized with largest absolute weight $O(\delta)$. Its cumulant bound (B.15) permits Döring et al. (2022, Theorem 2.1), originating from Rudzkis et al. (1978), with $\Delta \asymp \delta^{-1}$. Uniformly for $0 \leq x = O(\sqrt{\log n})$, this gives

$$\frac{\mathbb{P}^{(s)}((X - m_s)/\tau_s \geq x)}{\Phi^c(x)} = \exp\{O(\delta(1 + x^3))\} = 1 + o(1).$$

Reflection gives the corresponding lower tail estimate, while (B.14) controls the complementary tails. Moreover,

$$\frac{t - m_s}{\tau_s} = t - s + O(\delta(1 + \log n)).$$

The logarithmic derivatives of Φ and Φ^c are $O(1 + |x|)$, so this replacement changes either Gaussian tail by a factor $1 + o(1)$. Combining these estimates with (B.18) gives the second claim:

$$\mathbb{E}[e^{sX} \mathbf{1}_{\{X \geq t\}}] = e^{K(s)} \mathbb{P}^{(s)}(X \geq t) = (1 + o(1))e^{s^2/2} \Phi^c(t - s).$$

The same argument applied to $\{X < t\}$ gives the third claim:

$$\mathbb{E}[e^{sX} \mathbf{1}_{\{X < t\}}] = e^{K(s)} \mathbb{P}^{(s)}(X < t) = (1 + o(1))e^{s^2/2} \Phi(t - s).$$

All error bounds are uniform because the estimates depend on the particular sum only through δ and the mean and variance bounds, which the hypotheses control uniformly over the family. $\square$

B.2 Projected degree estimates

Proof of Lemma 3.2. Lemma B.3 and independence of the layers give a representation

$$S_v(\mathbf{u}) \stackrel{d}{=} c + \sum_{\ell,j} \frac{u_\ell}{\sigma_\ell} \xi_{\ell,j},$$

where all the Bernoulli variables on the right are independent. This variable has mean zero, variance one, and largest absolute weight $\delta_{\mathbf{u}}$. In particular, (B.15) gives $|\kappa_r(S_v(\mathbf{u}))| \leq C_0^r r! \delta_{\mathbf{u}}^{r-2}$ for $r \geq 3$. The result follows by Lemma B.6. $\square$

We next show that the preceding estimates are preserved even when conditioning on the full layer degrees of a fixed number of vertices.

Lemma B.7. *Suppose (1.3) holds, and let r be a fixed positive integer. Fix distinct vertices $w_1, \dots, w_r, v$, and let*

$$\mathcal{D} = \{D_{w_i}^{(\ell)} = d_{i\ell} : 1 \leq i \leq r, 1 \leq \ell \leq L\}$$

be an event of positive probability. Then uniformly over $\mathcal{D}$, all unit vectors $\mathbf{u} \in \mathbb{R}^L$, and $|s|, |t| = O(\sqrt{\log n})$,

$$\mathbb{E}[e^{sS_v(\mathbf{u})} \mid \mathcal{D}] = (1 + o(1))e^{s^2/2}, \quad (\text{B.19})$$

$$\mathbb{E}[e^{sS_v(\mathbf{u})} \mathbf{1}_{\{S_v(\mathbf{u}) \geq t\}} \mid \mathcal{D}] = (1 + o(1))e^{s^2/2} \Phi^c(t - s), \quad (\text{B.20})$$

$$\mathbb{E}[e^{sS_v(\mathbf{u})} \mathbf{1}_{\{S_v(\mathbf{u}) < t\}} \mid \mathcal{D}] = (1 + o(1))e^{s^2/2} \Phi(t - s). \quad (\text{B.21})$$

Proof of Lemma B.7. Let $W = \{w_1, \dots, w_r\}$ be the r vertices whose degrees are fixed by $\mathcal{D}$. Once we also fix the edges within W , the degree of v in each layer has an independent Bernoulli representation, by Lemma B.3. We check that its mean and variance differ from their null values by only $O_r(1)$, then apply Lemma B.6 and average over those edges.

Fix the presence or absence of every edge within W in every layer. Write $\mathcal{C}$ for the event combining this choice with $\mathcal{D}$, and suppose $\mathbb{P}(\mathcal{C}) > 0$. The layers remain independent under $\mathcal{C}$, since it conditions each layer separately.

Let us fix a layer ℓ and suppress its index, writing $D_v = D_v^{(\ell)}$, $d_i = d_{i\ell}$, $m = m_\ell$, $\mu = \mu_\ell$, and $\sigma = \sigma_\ell$. Let h_i count the neighbors of w_i inside W in this layer. The number of edges with both endpoints outside W can be computed as

$$m' = m - \sum_{i=1}^r d_i + \frac{1}{2} \sum_{i=1}^r h_i.$$

For each w_i , its $d_i - h_i$ neighbors outside W form a uniform subset of the $n - r$ remaining vertices. The m' edges wholly outside W likewise form a uniform subset of their $\binom{n-r}{2}$ possibilities. These choices are independent: their counts are fixed, their edge classes are disjoint, and every admissible graph is equally likely.

Let ξ_i indicate the edge $\{w_i, v\}$, and let H count edges from v to other vertices outside W . Under $\mathcal{C}$, we have

$$D_v \stackrel{d}{=} \sum_{i=1}^r \xi_i + H, \quad (\text{B.22})$$

with independent summands with distributions

$$\xi_i \sim \text{Bernoulli}\left(\frac{d_i - h_i}{n - r}\right), \quad \text{and} \quad H \sim \text{Hypergeom}\left(\binom{n-r}{2}, n-r-1, m'\right).$$

Hence

$$\mathbb{E}[D_v \mid \mathcal{C}] = \frac{\sum_i (d_i - h_i) + 2m'}{n - r} = \frac{2m - \sum_i d_i}{n - r} = \mu + O_r(1). \quad (\text{B.23})$$

The error is uniform because $d_i \leq n - 1$ and $\mu = o(n)$. For the variance, the r Bernoulli terms contribute at most $r/4$. The hypergeometric term has mean $2m'/(n - r) = \mu + O_r(1)$. Since $m' = m + O_r(n)$ and $\binom{n-r}{2} = N + O_r(n)$, its variance satisfies

$$\begin{aligned} \text{Var}(H \mid \mathcal{C}) &= \frac{2m'}{n - r} \left(1 - \frac{2}{n - r}\right) \frac{\binom{n-r}{2} - m'}{\binom{n-r}{2} - 1} \\ &= (\mu + O_r(1)) \left[\left(1 - \frac{2}{n}\right) \frac{N - m}{N - 1} + O_r(n^{-1}) \right] \\ &= \sigma^2 + O_r(1 + \mu/n) = \sigma^2 + O_r(1), \end{aligned}$$

where the last line uses (1.2) and $\mu = o(n)$. Adding the independent Bernoulli variances therefore gives

$$\text{Var}(D_v \mid \mathcal{C}) = \sigma^2 + O_r(1). \quad (\text{B.24})$$

Applying Lemma B.3 to H and restoring the layers, we obtain the representation (B.13) with $X = S_v(\mathbf{u})$ and each Bernoulli term from layer ℓ weighted by u_ℓ/σ_ℓ . The preceding bounds and layer independence give

$$\mathbb{E}[S_v(\mathbf{u}) \mid \mathcal{C}] = O_{L,r}(\delta_{\mathbf{u}}), \quad \text{Var}(S_v(\mathbf{u}) \mid \mathcal{C}) = 1 + O_{L,r}(\delta_{\mathbf{u}}^2).$$

By (3.3), $\delta_{\mathbf{u}} \leq \sigma_*^{-1} \ll (\log n)^{-3/2}$, where $\sigma_* = \min_{\ell \in [L]} \sigma_\ell$. Thus the hypotheses of Lemma B.6 hold under $\mathcal{C}$, uniformly over the fixed degrees, edges inside W , and unit vectors $\mathbf{u}$. The leading terms in (B.19)–(B.21) do not depend on the edges inside W , so averaging over those edges preserves the uniform relative errors and leaves conditioning only on $\mathcal{D}$. $\square$

For example, taking $r = 1$ and conditioning once gives, for distinct w, v and $|s|, |t| = O(\sqrt{\log n})$,

$$\mathbb{E} \left[e^{sS_w(\mathbf{u}) + tS_v(\mathbf{u})} \mathbf{1}_{\{S_w(\mathbf{u}) < t, S_v(\mathbf{u}) < t\}} \right] = (1 + o(1)) e^{s^2} \Phi(t - s)^2. \quad (\text{B.25})$$

C Proofs for Section 4

We prove Lemmas 4.1 and 4.2 from Section 4, then use these estimates to establish Corollary 1.5 and the leaf recovery limitation stated after it. Throughout this appendix, write

$$a := \|\mathbf{a}\|_2.$$

C.1 Maximum of projected degree

Proof of Lemma 4.1. Define

$$X := \sum_{v=1}^n \mathbf{1}_{\{S_v \geq t_n^*\}}, \quad \text{and} \quad p := \mathbb{P}_0(S_1 \geq t_n^*).$$

By linearity of expectation, Lemma 3.2, and Theorem 3.4, we obtain

$$\mathbb{E}_0 X = np = (1 + o(1)) \frac{1}{\sqrt{4\pi}} (\log n)^{1/(2\alpha)-1/2}.$$

The $\alpha > 1$ case in (4.1) then follows since $\mathbb{P}_0(\max_v S_v \geq t_n^*) = \mathbb{P}_0(X \geq 1) \leq \mathbb{E}_0 X = o(1)$ by Markov's inequality.

If $0 < \alpha < 1$, then $\mathbb{E}_0 X \rightarrow \infty$ instead, and we use the second moment method. Conditioning on vertex 1's layer degrees $(D_1^{(1)}, \dots, D_1^{(L)})$, and applying Lemma B.7 with $s = 0$ gives

$$\mathbb{P}_0(S_1 \geq t_n^*, S_2 \geq t_n^*) = (1 + o(1))p^2.$$

It follows that

$$\frac{\text{Var}_0 X}{(\mathbb{E}_0 X)^2} = \frac{1}{\mathbb{E}_0 X} + o(1) = o(1),$$

The Paley–Zygmund inequality therefore gives $\mathbb{P}_0(X > 0) \geq (\mathbb{E}_0 X)^2 / \mathbb{E}_0[X^2] = 1 - o(1)$, proving the $0 < \alpha < 1$ case of (4.1). Both thresholds differ from $\sqrt{2 \log n}$ by $o(1)$, which proves (4.2). The same calculation applies to a sum over $n - 1$ vertices. The estimates used here, (3.5) and (B.20), are uniform in the choice of the unit vector, proving the last assertion. $\square$

Proof of Lemma 4.2. Recall $\sigma_* = \min_\ell \sigma_\ell$ from Lemma 5.1. We first bound the maximum statistic away from the hub above and below using the same coupling, given as follows. Choose the hub v_* and planted star sets S_ℓ as in the model. Independently of these choices, take independent uniform permutations π_ℓ of all N possible edges. For the upper bound, let $G_0^{(\ell)}$ consist of the first m_ℓ edges under π_ℓ , and let H_ℓ consist of the first $m_\ell - k_\ell$ edges outside S_ℓ . Set $G_1^{(\ell)} = S_\ell \cup H_\ell$. Then $\mathbf{G}_0 \sim \mathbb{P}_0$ and $\mathbf{G}_1 \sim \mathbb{P}_1$.

Observe that the first $m_\ell - k_\ell$ edges outside S_ℓ must occur within the first m_ℓ positions of π_ℓ , since at most k_ℓ positions can be occupied by star edges. Thus $H_\ell \subseteq G_0^{(\ell)}$ almost surely under this coupling. Every vertex other than the hub is incident to at most one star edge in each layer, so $D_v^{(\ell)}(\mathbf{G}_1) \leq D_v^{(\ell)}(\mathbf{G}_0) + 1$ for $v \neq v_*$. Since the projection coefficients are nonnegative, these inequalities give

$$S_v(\mathbf{G}_1) \leq S_v(\mathbf{G}_0) + \varepsilon_n, \quad \text{where} \quad \varepsilon_n := \frac{1}{\|\mathbf{a}\|_2} \sum_\ell \frac{a_\ell}{\sigma_\ell} = O(\sigma_*^{-1}), \quad (\text{C.1})$$

Since $\varepsilon_n \sqrt{\log n} = o(1)$, Lemma 3.2 and Theorem 3.4 imply, for $\alpha > 1$,

$$n\mathbb{P}_0(S_1 \geq t_n^* - \varepsilon_n) = (1 + o(1))n\Phi^c(t_n^*) = o(1).$$

A union bound then proves (4.3), and also the upper bound in (4.4).

For the lower bound, we will consider a smaller “null graph” contained in the planted graph. Let $G_-^{(\ell)}$ consist of the first $m_\ell - k_\ell$ edges of the same permutation π_ℓ , without omitting star edges. Every such edge is either in S_ℓ or among the first $m_\ell - k_\ell$ nonstar edges. Therefore,

$$G_-^{(\ell)} \subseteq G_1^{(\ell)}.$$

Marginally, the layers $G_-^{(\ell)}$ are independent $G(n, m_\ell - k_\ell)$ graphs. Their degree means $\mu_{\ell,-}$ and standard deviations $\sigma_{\ell,-}$ satisfy

$$\mu_{\ell,-} = \mu_\ell - \frac{2k_\ell}{n}, \quad \text{and} \quad \frac{\sigma_{\ell,-}}{\sigma_\ell} = 1 + O\left(\frac{k_\ell}{m_\ell}\right).$$

Using these means and standard deviations, define

$$S_{v,-} := \frac{1}{\|\mathbf{a}\|_2} \sum_\ell a_\ell \frac{D_v^{(\ell)}(\mathbf{G}_-) - \mu_{\ell,-}}{\sigma_{\ell,-}}.$$

The parameter regime (1.3) remains valid with m_ℓ and μ_ℓ replaced by $m_\ell - k_\ell$ and $\mu_{\ell,-}$, respectively. Equations (3.5) and (3.6), applied separately to each layer, show that the maximum absolute standardized degree is $O_{\mathbb{P}}(\sqrt{\log n})$ in each layer. Since $k_\ell/m_\ell = O(\log n/(nk_\ell))$ and $k_\ell/\sigma_\ell = O(\sqrt{\log n})$, changing the centering contributes $O(\sqrt{\log n}/n)$, and changing the scaling contributes $o_{\mathbb{P}}(1)$, uniformly over vertices. Thus the statistic $S_v(\mathbf{G}_-)$, which uses the original means and standard deviations, satisfies

$$\max_{v \in [n]} |S_v(\mathbf{G}_-) - S_{v,-}| = o_{\mathbb{P}}(1).$$

Lemma 4.1, with the unchanged direction $\mathbf{a}/\|\mathbf{a}\|_2$, gives $\max_{v \neq v_*} S_{v,-} = \sqrt{2 \log n} + o_{\mathbb{P}}(1)$. The preceding comparison and $G_-^{(\ell)} \subseteq G_1^{(\ell)}$ therefore give the required lower bound for the planted maximum.

We finally prove (4.5). The degree of the planted hub in layer ℓ has the exact law

$$D_{v_*}^{(\ell)} - k_\ell \sim \text{Hypergeom}(N - k_\ell, n - 1 - k_\ell, m_\ell - k_\ell), \quad (\text{C.2})$$

independently across layers. We compute

$$\mathbb{E}_1 D_{v_*}^{(\ell)} = \mu_\ell + k_\ell + O\left(\frac{k_\ell \mu_\ell}{n} + \frac{k_\ell}{n}\right) = \mu_\ell + k_\ell + o(\sigma_\ell),$$

where in the last equality we used $\frac{k_\ell \sqrt{\mu_\ell}}{n} = \frac{k_\ell}{\sqrt{n}} \sqrt{\frac{\mu_\ell}{n}} = o(1)$. Similarly, the hypergeometric variance, together with $k_\ell = o(\sqrt{n})$, $k_\ell/m_\ell = o(1)$, and $m_\ell = o(N)$, gives

$$\text{Var}_1(D_{v_*}^{(\ell)}) = (1 + o(1))\sigma_\ell^2.$$

Since $\sigma_\ell^2/\mu_\ell = 1 + O(1/n + \mu_\ell/n)$, we also have

$$\frac{k_\ell}{\sigma_\ell} - a_\ell = O\left(\frac{k_\ell}{n\sqrt{\mu_\ell}} + \frac{k_\ell \sqrt{\mu_\ell}}{n}\right) = o(1).$$

Consequently, $\mathbb{E}_1 S_{v_*} = a + o(1)$ and $\text{Var}_1(S_{v_*}) = 1 + o(1)$.

By Lemma B.3, the centered projection is distributed as a weighted sum of independent centered Bernoulli variables, with largest absolute weight at most $1/\sigma_*$. Lemma B.5 shows that the centered and standardized projection converges in distribution to $\mathcal{N}(0, 1)$. The preceding mean and variance estimates therefore prove (4.5). $\square$

C.2 Hub and leaf recovery

Proof of Corollary 1.5. By Lemma 4.2,

$$S_{v_*} - a \xrightarrow{d} \mathcal{N}(0, 1), \quad \text{and} \quad \max_{v \neq v_*} S_v = \sqrt{2 \log n} + o_{\mathbb{P}_1}(1).$$

For every fixed $\varepsilon > 0$, the second relation gives

$$\begin{aligned} \mathbb{P}_1(S_{v_*} > \sqrt{2 \log n} + \varepsilon) - o(1) &\leq \mathbb{P}_1(\widehat{v} = v_*) \\ &\leq \mathbb{P}_1(S_{v_*} \geq \sqrt{2 \log n} - \varepsilon) + o(1). \end{aligned}$$

Applying the Gaussian limit and letting $\varepsilon \downarrow 0$ gives

$$\mathbb{P}_1(\widehat{v} = v_*) = \Phi(a - \sqrt{2 \log n}) + o(1).$$

This gives the success probability of $\widehat{v}$.

It remains to prove that no estimator has a larger limiting success probability. Bayes' rule gives

$$\mathbb{P}_1(v_* = v \mid \mathbf{G}) = \frac{\prod_{\ell=1}^L \binom{D_v^{(\ell)}}{k_\ell}}{\sum_{u=1}^n \prod_{\ell=1}^L \binom{D_u^{(\ell)}}{k_\ell}}.$$

Thus the maximum a posteriori (MAP) estimator maximizes the success probability among all estimators. It selects a vertex maximizing $\prod_{\ell} \binom{D_v^{(\ell)}}{k_\ell}$, or equivalently $\sum_{\ell} \log(D_v^{(\ell)})_{k_\ell}$.

We will next expand $\log(D_v^{(\ell)})_{k_\ell}$ to obtain $\sum_{\ell} \log(D_v^{(\ell)})_{k_\ell} = C_n + aS_v + o(a)$ uniformly over v , where C_n does not depend on v . As in the proof of Lemma 2.1, on the event that (A.1) holds, Taylor's formula gives, uniformly in v ,

$$\log(D_v^{(\ell)})_{k_\ell} = k_\ell \log \mu_\ell + \frac{k_\ell(D_v^{(\ell)} - \mu_\ell)}{\mu_\ell} - \frac{k_\ell(k_\ell - 1)}{2\mu_\ell} + O\left(\frac{k_\ell(\sigma_\ell \sqrt{\log n} + k_\ell)^2}{\mu_\ell^2}\right).$$

All factors are positive on this event, because $(\sigma_\ell \sqrt{\log n} + k_\ell)/\mu_\ell = o(1)$. Since $\sigma_\ell^2 \asymp \mu_\ell$ and $a_\ell^2 \asymp k_\ell^2/\mu_\ell$, the remainder is

$$O\left(\frac{a_\ell^2 \log n + a_\ell^4}{k_\ell}\right).$$

After summing over layers and dividing by $a > 0$, it is at most

$$O\left(\frac{a \log n + a^3}{\min_{\ell} k_\ell}\right) = O\left(\frac{(\log n)^{3/2}}{\min_{\ell} k_\ell}\right) = o(1).$$

Consequently, for a deterministic C_n ,

$$\max_v \left| \sum_{\ell=1}^L \log(D_v^{(\ell)})_{k_\ell} - C_n - aS_v \right| = o(a) \tag{C.3}$$

with probability $1 - o(1)$ under $\mathbb{P}_1$. Here C_n is the sum of the terms in the Taylor expansion that do not depend on v .

Thus, if the MAP estimator selects v_* , (C.3) forces

$$S_{v_*} \geq \max_{v \neq v_*} S_v - o(1).$$

By (4.4) and (4.5), the success probability of the MAP estimator is therefore at most $\Phi(a - \sqrt{2 \log n}) + o(1)$. Together with the lower bound attained by $\hat{v}$, this proves (1.10). $\square$

We now state the leaf recovery limitation mentioned after Corollary 1.5.

Proposition C.1. *Suppose (1.3) holds. For each layer ℓ , let S_ℓ denote its planted leaf set. Then for every fixed $C < \infty$ and $\varepsilon > 0$, any estimator with $|\hat{S}_\ell| \leq Ck_\ell$ satisfies*

$$\mathbb{P}_1(|\hat{S}_\ell \cap S_\ell| \geq \varepsilon k_\ell) = o(1),$$

uniformly over such estimators, even if they use all observed layers and are given the true hub v_ .*

Proof of Proposition C.1. Condition on the observed multiplex graph and on v_* . For every k_ℓ -subset of its neighbors in layer ℓ , the probability of choosing that subset as the planted leaves and then completing the observed graph is the same. The other layers do not alter these probabilities, by conditional independence. Consequently, the posterior distribution of S_ℓ is uniform over the k_ℓ -subsets of the $D_{v_*}^{(\ell)}$ observed neighbors. For any proposed set of size at most Ck_ℓ ,

$$\mathbb{E}_1 \left[\frac{|\hat{S}_\ell \cap S_\ell|}{k_\ell} \mid \mathbf{G}, v_* \right] = \frac{|\hat{S}_\ell \cap N_{G^{(\ell)}}(v_*)|}{D_{v_*}^{(\ell)}} \leq \frac{Ck_\ell}{D_{v_*}^{(\ell)}}.$$

For randomized estimators, also condition on their independent randomness. The degree concentration estimate implies $D_{v_*}^{(\ell)} \geq \mu_\ell/2$ with probability $1 - o(1)$, because $\sigma_\ell \sqrt{\log n} + k_\ell = o(\mu_\ell)$. Conditional Markov's inequality therefore gives, uniformly over the estimators,

$$\mathbb{P}_1 \left(|\hat{S}_\ell \cap S_\ell| \geq \varepsilon k_\ell \right) \leq o(1) + \frac{2Ck_\ell}{\varepsilon \mu_\ell} = o(1),$$

using $k_\ell/\mu_\ell = O(\log n/k_\ell) = o(1)$. $\square$

D Proofs for Section 5

In this appendix we prove Lemma 5.2. We write $a := \|\mathbf{a}\|_2$ and $\mathbf{b} := \mathbf{a}/\|\mathbf{a}\|_2$ throughout.

D.1 Proof of Lemma 5.2

Proof of Lemma 5.2. We first compute the hub's degree means and variances, then establish the Gaussian limit for its standardized degree vector, and finally pass to the norm of its positive part. Throughout the proof, we work under $\mathbb{P}_1$, conditional on v_* . The variables $Y_{v_*}^{(\ell)}$ still use the null means μ_ℓ and standard deviations σ_ℓ from (1.2). Put $s = n - 1$ and $p_\ell = m_\ell/N$. The exact null variance is

$$\sigma_\ell^2 = \mu_\ell(1 - p_\ell) \frac{N - s}{N - 1}.$$

The hub's degree in each layer has the distribution

$$D_{v_*}^{(\ell)} = k_\ell + H_\ell, \quad \text{where} \quad H_\ell \sim \text{Hypergeom}(N - k_\ell, s - k_\ell, m_\ell - k_\ell),$$

independently across layers. A direct calculation gives

$$\mathbb{E}_1 Y_{v_*}^{(\ell)} = \frac{k_\ell(1 - p_\ell)(N - s)}{(N - k_\ell)\sigma_\ell} = a_\ell \frac{N - 1}{N - k_\ell}. \quad (\text{D.1})$$

Consequently the difference from a_ℓ is $o(1)$, uniformly over the finitely many layers. The planted variance satisfies

$$\frac{\text{Var}_1(D_{v_*}^{(\ell)})}{\sigma_\ell^2} = \left(1 - \frac{k_\ell}{m_\ell}\right) \left(1 - \frac{k_\ell}{s}\right) \left(\frac{N}{N - k_\ell}\right)^2 \frac{N - 1}{N - k_\ell - 1} = 1 + o(1).$$

These moment estimates allow us to use the Gaussian approximation from Appendix B. The fact that the Hypergeometric PGF has only real roots and Lemma B.5 show that the distribution function of each centered and standardized H_ℓ differs from Φ by at most $O(\text{Var}_1(H_\ell)^{-1/2})$ at every threshold. Independence and fixed L give the joint Gaussian limit for $\mathbf{Y}_{v_*} - \mathbb{E}_1 \mathbf{Y}_{v_*}$. Since $\mathbb{E}_1 \mathbf{Y}_{v_*} - \mathbf{a} = o(1)$, we obtain

$$\mathbf{Z}_n := \mathbf{Y}_{v_*} - \mathbf{a} \xrightarrow{d} \mathcal{N}(0, I_L).$$

Applying the same approximation to projections gives $\langle \mathbf{b}, \mathbf{Z}_n \rangle \xrightarrow{d} \mathcal{N}(0, 1)$, including when $\mathbf{b}$ varies with n .

It remains to pass from the vector limit to the norm. The map $\mathbf{y} \mapsto \|(\mathbf{y})_+\|_2$ has Lipschitz constant 1, so $R_{v_*} = a + O_{\mathbb{P}_1}(1)$. To obtain the Gaussian fluctuation when $a \rightarrow \infty$, observe that for every nonnegative vector $\mathbf{a}$ and every $\mathbf{z} \in \mathbb{R}^L$,

$$0 \leq \|(\mathbf{a} + \mathbf{z})_+\|_2^2 - \|\mathbf{a}\|_2^2 - 2\langle \mathbf{a}, \mathbf{z} \rangle \leq \|\mathbf{z}\|_2^2.$$

With $\mathbf{z} = \mathbf{Z}_n = O_{\mathbb{P}_1}(1)$, the remainder in this squared norm is bounded in probability, even when some coordinates of $\mathbf{b}$ tend to zero. Taking square roots and using the projection limit gives (5.2):

$$R_{v_*} - a = \langle \mathbf{b}, \mathbf{Z}_n \rangle + o_{\mathbb{P}_1}(1) \xrightarrow{d} \mathcal{N}(0, 1) \quad \text{when } a \rightarrow \infty.$$

□