Is it easier to count communities than find them?

Cynthia Rush*[1], Fiona Skerman†[2], Alexander S. Wein‡[3], and Dana Yang§[4]

[1]Department of Statistics, Columbia University.
[2]Department of Mathematics, Uppsala University.
[3]Department of Mathematics, University of California, Davis.
[4]Department of Statistics and Data Science, Cornell University.

April 21, 2025

Abstract

Random graph models with community structure have been studied extensively in the literature. For both the problems of detecting and recovering community structure, an interesting landscape of statistical and computational phase transitions has emerged. A natural unanswered question is: might it be possible to infer properties of the community structure (for instance, the number and sizes of communities) even in situations where actually finding those communities is believed to be computationally hard? We show the answer is no. In particular, we consider certain hypothesis testing problems between models with different community structures, and we show (in the low-degree polynomial framework) that testing between two options is as hard as finding the communities.

In addition, our methods give the first computational lower bounds for testing between two different "planted" distributions, whereas previous results have considered testing between a planted distribution and an i.i.d. "null" distribution.

1 Introduction

Questions of detecting and recovering community structure in random graph models have been studied extensively in the literature. Popular models include the *planted dense subgraph model* [ACV13, HWX15], where an Erdős–Rényi base graph is augmented by adding one or more "communities" — subsets of vertices with a higher-than-average connection probability between them — and the *stochastic block model* (see [Abb17, Moo17] for a survey). There are by now a multitude of results identifying sharp conditions based on the problem parameters, e.g. edge probabilities and

*Email: *cynthia.rush@dcolumbia.edu.* Part of this work was supported by NSF CCF #1849883 and part of the work was done while visiting the Simons Institute for the Theory of Computing, supported by a Google Research Fellowship.

†Email: *fiona.skerman@math.uu.se.* Partially supported by the Wallenberg AI, Autonomous Systems and Software Program WASP and the project AI4Research at Uppsala University. Part of this work was done while visiting the Simons Institute for the Theory of Computing, supported by a Simons-Berkeley Research Fellowship.

‡Email: *aswein@ucdavis.edu.* Partially supported by a Sloan Research Fellowship and NSF CAREER Award CCF-2338091. Part of this work was done at Georgia Tech, supported by NSF grants CCF-2007443 and CCF-2106444. Part of this work was done while visiting the Simons Institute for the Theory of Computing, supported by a Simons-Berkeley Research Fellowship.

§Email: *dana.yang@cornell.edu.* Part of this work was done while visiting the Simons Institute for the Theory of Computing, supported by a Simons-Berkeley Research Fellowship.

number/sizes of communities, under which it is possible (or impossible) to recover (exactly or approximately) the hidden partition of vertices, given a realization of the graph as input. Notably, many settings are believed to exhibit a *statistical-computational gap*; that is, there exists a "possible but hard" regime of parameters where it is *statistically* possible to recover the communities (typically by brute-force search) but there is no known *computationally efficient*, meaning polynomial-time, algorithm for doing so. It may be that this hardness is inherent, meaning no poly-time algorithm exists, which is suggested by a growing body of "rigorous evidence" including reductions from the *planted clique* problem [BBH18, HWX15] and limitations of known classes of algorithms [BMR21, DKMZ11, HS17, SW22].

Despite all this progress, one question that remains relatively unexplored is the following: in the aforementioned "hard" regime, even though it seems hard to recover the communities, might it still be possible to learn *something* about the community structure (e.g., the number or sizes of communities)? After all, in some models it has already been established that *detecting* the presence of a dense subgraph (i.e., distinguishing the planted subgraph model from an appropriate Erdős–Rényi "null" model) appears to be strictly easier than actually recovering which vertices belong to it [BBH18, CX16, HWX15, SW22]. Existing detection-recovery gaps of this nature often occur due to a "trivial" test for detection (e.g., the total edge count), and the motivation for our work is to understand more precisely which properties of the community structure can be inferred in the hard regime, and which ones cannot.

A simple testing problem. One of the simplest inference tasks on the community structure is to detect the number of communities. Let us consider a toy problem of testing between two graph models: under $\mathbb{P}$ the graph contains one community of expected size k, while under $\mathbb{Q}$ the graph contains two communities each of expected size $k/2$. The community membership of each vertex is independent in both models (k/n under $\mathbb{P}$ and $k/(2n)$, $k/(2n)$ under $\mathbb{Q}$) and vertices cannot be members of more than one community. Suppose any pair of vertices from the same community are connected independently with probability $2q$ and $3q$ under $\mathbb{P}$ and $\mathbb{Q}$, respectively, and all the other pairs of vertices are connected independently with probability q under both models. Such a parameterization matches the expected degrees of the nodes under the two distributions, so that a simple test based on the total edge count fails to distinguish between $\mathbb{P}$ and $\mathbb{Q}$. One natural test is to threshold the number of triangles. It is easy to derive that the expected number of triangles under $\mathbb{P}$ and $\mathbb{Q}$ scale as different constant multiples of q^3k^3, and the variance of the number of triangles is of order $\Theta(n^3q^3)$ under both models. Thus, the simple triangle counting algorithm consistently distinguishes $\mathbb{P}$ and $\mathbb{Q}$ if $q^3k^3 \gg \sqrt{n^3q^3}$, i.e. $qk^2/n \gg 1$.

It is intriguing that the condition for the triangle counting algorithm to succeed coincides with the conjectured computational barrier for the more difficult task of finding all members of the community under the model $\mathbb{P}$ [SW22]. In other words, in the entire "hard" regime where one cannot efficiently locate the planted community, the triangle counting algorithm fails to even tell whether the graph contains one or two communities. In this paper, we show that this statement extends beyond the simple triangle counting algorithm to all low-degree tests. Our main result is given in the following (informal) theorem statement.

Theorem 1.1 (Informal)**.** *If $q(k^2/n \vee 1) \leqslant 1/\mathrm{polylog}(n)$, then no low-degree test consistently tests between the graph models with one and two planted communities.*

Moreover, the informal result of Theorem 1.1 extends to a much wider class of testing problems than those for which it is stated. We find that, whenever recovery is computationally hard, all low-degree tests fail to distinguish models with different numbers of planted communities of possibly different sizes. In other words, inferring the community structure is just as hard as finding members of the planted communities themselves. We show a similar phenomenon for graphs with Gaussian weights. See Theorems 2.4 and 2.5 for the formal statements. It is important to note that our results apply even in regimes where it is easy to distinguish $\mathbb{P}$ (or $\mathbb{Q}$) from an Erdős–Rényi graph; that is, one cannot recover our results simply by arguing that both $\mathbb{P}$ and $\mathbb{Q}$ are hard to distinguish from Erdős–Rényi.

We additionally give a few other related results.

Connections between detection and recovery in the low-degree framework. We show a connection between detection and recovery. For a given recovery problem in a planted model there is an equivalent testing problem where one tests between two planted models. In the other direction, for a testing problem between two planted models there is an equivalent recovery problem if the likelihood ratio of the signals exists. This equivalence is in a strong sense: one is low-degree hard if and only if the other one is low-degree hard. See Section 2.3 for a full statement of these results.

Alternative proof strategy for hardness of recovery. We give a reduction showing that if there were an algorithm that successfully *recovers* a planted community, one could turn this into an algorithm for testing one community versus two. Therefore, the "hard" regime for recovery contains the "hard" regime for testing community structure. We prove this in Theorem 7.2 — see Section 7.

Although the reduction seems straightforward, there are some technical challenges: we suppose an algorithm recovers the planted community in the one-community case, but we cannot control how the algorithm behaves when there are two planted communities.

Our reduction provides an alternative method for establishing detection-recovery gaps, rather than studying the recovery problem directly. For problems where recovery of the planted structure is strictly harder than detecting its presence, it is not viable to deduce optimal hardness of recovery from a planted-versus-null testing problem. However, we demonstrate that it is possible to attain the sharp recovery threshold via reduction from a planted-versus-planted problem, as long as the two planted distributions are appropriately chosen.

1.1 Related Work and Open Problems

The low-degree testing framework. Unfortunately, it seems to be beyond the current reach of computational complexity theory to prove that no polynomial-time algorithm can distinguish two random graph models, even under an assumption like P $\neq$ NP. Nonetheless, a popular heuristic — the *low-degree testing framework* [BHK+19, Hop18, HKP+17, HS17] (see [KWB22] for a survey) — gives us a rigorous basis on which to form conjectures about hardness of such problems. Specifically, we will study the power of *low-degree tests*, a class of methods that includes tests based on edge counts, triangle counts, and other small subgraph counts. Strikingly, low-degree tests tend to be as powerful as all known polynomial-time algorithms for testing problems that are (informally speaking) of the flavor that we consider in this paper; see [Hop18, KWB22] for discussion. In this paper, we will prove *low-degree hardness*, meaning failure of all low-degree tests (to be defined formally in Section 2.1), for certain testing problems; this can be viewed as an apparent barrier to fast algorithms that we believe is unlikely to be overcome by known techniques, and perhaps indicates fundamental computational hardness.

Planted-versus-planted testing. We emphasize that there is a key difference between our work and existing hardness results for high-dimensional testing. The testing problems we consider are between two different "planted" distributions, each with a different type of planted structure. In contrast, previous low-degree hardness results for testing (e.g., [Hop18, HKP+17, HS17, KWB22] and many others) have always considered testing between "planted" and "null," where the null distribution has i.i.d. or at least independent entries. On a technical level, planted-versus-null problems are more tractable to analyze because we can explicitly construct a basis of orthogonal polynomials for the null distribution, which enables easy representation of the so-called 'advantage' (see (2.4)), the main quantity to bound when using the low-degree polynomial approach, but this strategy seems more difficult to implement for planted-versus-planted problems.

The idea of planted-versus-null testing goes beyond the low-degree framework. Other forms of average-case lower bounds typically also, either explicitly or implicitly, leverage an easy-to-analyze null distribution; this includes reductions from planted clique (e.g., [BR13, BBH18]), sum-of-squares lower bounds (e.g., [BHK+19, KMOW17]), and statistical query lower bounds (e.g., [DKS17, FGR+17]). In fact, these frameworks seem to struggle in settings where there is not a simple null distribution in the hypothesis test.

Our work overcomes this barrier that has limited the use of the above methods: we demonstrate for the first time that low-degree hardness results can be proven for planted-versus-planted problems. We give some general-purpose formulas (Propositions 2.8 and 2.9) that can be used to analyze a wide variety of such problems in random graphs or random matrices, not limited to just the specific models studied in this paper. The proof techniques are inspired by [SW22], which studies estimation problems rather than testing. On a technical level, the core challenge in our analysis is to bound certain recursively-defined quantities called r_α (defined in (2.5)). These are analogous to the cumulants that appear in [SW22], and while the r_α are not cumulants, they enjoy a number of similar convenient properties (see Section 3) that are important for the analysis.

Open problems. A natural next step is to investigate whether our method yields sharp computational thresholds for other problems that exhibit detection-recovery gaps. For example, the problem of parameter estimation in sparse high-dimensional linear regression likely has a detection-recovery gap (see [BAH+22]) and can potentially be related to a testing problem between two planted models, e.g. between a sparse linear regression and a mixture of two sparse linear regressions.

Another open question is whether our computational hardness result can be shown in ways beyond the low-degree testing framework, such as by using the sum-of-squares framework, statistical query framework, or reduction from the planted clique problem. In particular, if the problem of testing community structure can be reduced from planted clique, this would yield a reduction from planted clique to planted dense subgraph *recovery*, which is an open problem (see [BBH18]).

2 Main results

2.1 Low-degree testing

We begin by explaining what it means for a low-degree test to distinguish two high-dimensional distributions.

Definition 2.1. *Suppose $\mathbb{P}_n$ and $\mathbb{Q}_n$ are distributions on $\mathbb{R}^N$ for some $N = N_n$. A* degree-D test *is a multivariate polynomial $f_n : \mathbb{R}^N \to \mathbb{R}$ of degree at most D (really, a sequence of polynomials, one for each problem size n). Such a test f is said to* strongly separate $\mathbb{P}$ *and* $\mathbb{Q}$ *if, in the limit*

$n \to \infty$,

$$\sqrt{\max\left\{\operatorname*{Var}_{\mathbb{Q}}[f], \operatorname*{Var}_{\mathbb{P}}[f]\right\}} = o\left(|\mathbb{E}_{\mathbb{P}}[f] - \mathbb{E}_{\mathbb{Q}}[f]|\right),$$

and weakly separate $\mathbb{P}$ *and* $\mathbb{Q}$ *if*

$$\sqrt{\max\left\{\operatorname*{Var}_{\mathbb{Q}}[f], \operatorname*{Var}_{\mathbb{P}}[f]\right\}} = O\left(|\mathbb{E}_{\mathbb{P}}[f] - \mathbb{E}_{\mathbb{Q}}[f]|\right).$$

Strong separation is a natural sufficient condition for success of a polynomial-based test because it implies (by Chebyshev's inequality) that $\mathbb{P}$ and $\mathbb{Q}$ can be distinguished by thresholding f's output, with both type I and II errors $o(1)$. Weak separation also implies non-trivial testing, i.e. better than a random guess; see [BAH$^+$22, Prop. 6.1]. In this paper, we characterize the limits of low-degree tests. For upper bounds, in the "easy" regime, we show that a constant-degree test achieves strong separation, implying a poly-time algorithm for testing with $o(1)$ error probability. For lower bounds, in the "hard" regime, we show that for some $D = \omega(\log n)$, no degree-D test can achieve even weak separation. Because many known algorithms can be implemented as degree-$O(\log n)$ polynomials (e.g., spectral methods; see Section 4.2.3 of [KWB22]), we treat this as "evidence" that no polynomial-time algorithm achieves non-trivial testing power, i.e. better than a random guess. Our results, in fact, often rule out much higher degree tests (e.g., $D = n^{\Omega(1)}$), depending on how far the parameters lie from the critical threshold.

2.2 Model formulation

We consider the problem of testing between two random graph models, both of which contain planted communities but with different community structures. We focus on testing between two *additive Gaussian models* where the edge weights are Gaussian, and between two *binary observation models* where the edges are unweighted and the diagonal is set to zero to ensure no self-loops in the graph.

Definition 2.2 (Additive Gaussian model)**.** *Given the number of vertices n, total community size k, signal strength $\lambda > 0$, number of communities M, and vector of community proportions $x \in [0,1]^M$ with $\sum_{\ell=1}^M x_\ell = 1$, define the additive Gaussian model $\mathbb{P} = \mathbb{P}_{\text{Gaussian}}(n, k, \lambda, M, x)$ as follows. Under $\mathbb{P}$, independently for each $i \in [n] := \{1, 2, \ldots, n\}$, the community label σ_i is sampled such that $\sigma_i = \ell$ with probability $x_\ell k/n$ for each $\ell \in [M]$ and $\sigma_i = \star$ (a symbol indicating membership in none of the communities) with probability $1 - k/n$. For each $i, j \in [n]$ with $i \leqslant j$, the edge weight Y_{ij} is sampled from*

$$Y_{ij} \sim \begin{cases} \mathcal{N}\left(\frac{\lambda}{x_\ell}, 1\right), & \sigma_i = \sigma_j = \ell \text{ for some } \ell \in [M], \\ \mathcal{N}(0,1), & \text{otherwise.} \end{cases}$$

For $i > j$, the edge weight Y_{ij} is defined to be Y_{ji}.

Notice that with the above definition, each community $\ell \in \{1, 2, \ldots, M\}$ is expected to be of size $x_\ell k$ and the expected number of vertices which do not belong to any community is $n - k$. The choice of mean λ/x_ℓ ensures that on average, the vertices in one community have the same weighted degree (row sum of Y) as the vertices in any other community.

Definition 2.3 (Binary observation model)**.** *Given the number of vertices n, total community size k, edge probability parameters $q, s \geqslant 0$, number of communities M, and vector of community proportions $x \in \mathbb{R}^M$ with $\sum_{\ell=1}^M x_\ell = 1$, define the Binary observation model $\mathbb{P} = \mathbb{P}_{\text{Binary}}(n, k, q, s, M, x)$*

as follows. The community labels $\{\sigma_i\}_{i\in[n]}$ *are sampled the same way as in the additive Gaussian model. Given the community labels, for each pair of vertices* $i, j \in [n]$ *with* $i < j$*, the edge weight* Y_{ij} *is sampled from*

$$Y_{ij} \sim \begin{cases} \text{Bernoulli}\left(q + \frac{s}{x_\ell}\right), & \sigma_i = \sigma_j = \ell \text{ for some } \ell \in [M], \\ \text{Bernoulli}\,(q), & \text{otherwise.} \end{cases}$$

For $i > j$*, the edge weight* Y_{ij} *is defined to be* Y_{ji} *and the diagonal entries set to zero* $Y_{ii} = 0$*.*

For example, if we want to model two communities of equal sizes, we can choose $M = 2$ and $x_1 = x_2 = \frac{1}{2}$. The communities are then both expected to be of size $k/2$. If we also set $s = q$ we have an in-community connection probability of $3q$ and every other pair of nodes is connected with probability q as in the toy model discussed in the Introduction.

The two models introduced in Definitions 2.2 and 2.3 only differ in the edge weight distributions, as the community labels follow the same distribution under both models. Alternatively, we can write S_ℓ for the set of vertices in community ℓ, so that $\sigma_i = \ell$ if and only if $i \in S_\ell$. Note that by definition, each vertex i can belong to at most one community. In other words, the communities $\{S_\ell\}_{\ell\in[M]}$ are disjoint.

With the other parameters fixed, we consider testing between model $\mathbb{P}$ with M planted communities and community proportions $x \in [0, 1]^M$, and the model $\mathbb{Q}$ with M' planted communities and community proportions $x' \in [0, 1]^{M'}$ for some $M' \neq M$. In short, for both Gaussian and Bernoulli edge weight models, we establish a "hard" regime where the distributions $\mathbb{P}$ and $\mathbb{Q}$ cannot be weakly separated by low-degree tests. We consider the regime $n \to \infty$ and allow all the parameters k, λ, M, x to depend on n; thus, our results can apply to a growing number of communities, although our main focus is on the case where M, M' are fixed so that our upper and lower bounds match.

Theorem 2.4 (Additive Gaussian model)**.** *Given parameters* $n, k, \lambda, M, M', x, x'$*, define distributions* $\mathbb{P} = \mathbb{P}_{\text{Gaussian}}(n, k, \lambda, M, x)$ *and* $\mathbb{Q} = \mathbb{P}_{\text{Gaussian}}(n, k, \lambda, M', x')$*. Assume that* $M \min_\ell x_\ell \geqslant C$ *and* $M' \min_\ell x'_\ell \geqslant C$ *for some constant* $C > 0$*. Write* $\widetilde{M} = |M - M'|$ *and* $\widehat{M} = \max\{M, M'\}$*. We have:*

- *If* $D^5 \widehat{M}^2 \lambda^2 (k^2/n \vee 1) = o(1)$*, then no degree-D test weakly separates* $\mathbb{P}$ *and* $\mathbb{Q}$*.*
- *If* $\widetilde{M}^2 \lambda^2 k^2/n = \omega(1)$ *and* $\widetilde{M}^2 k/\widehat{M}^2 = \omega(1)$*, then there exists a degree-1 test that strongly separates* $\mathbb{P}$ *and* $\mathbb{Q}$*.*

In the regime $k^2 \geqslant n$, $\widehat{M} = O(1)$, and $D \leqslant \text{polylog}(n)$, Theorem 2.4 precisely characterizes (up to logarithmic factors) the computational threshold for low-degree testing. This threshold coincides with the conjectured computational threshold for *recovering* a single planted community, which has been established in the low-degree polynomial framework [SW22, Theorem 2.5]. We focus on the $k^2 \geqslant n$ regime in this paper, as this is where there is a conjectured detection-recovery gap, but we suspect that when $\widehat{M}$ is constant, $\lambda^2(k^2/n \vee 1) \sim 1$ is the computational threshold across the entire parameter regime. The optimal test when $k^2 < n$ should be based on the maximum diagonal entry, and while this is not a polynomial, it should be possible to approximate it by one (similar to Section 4.1.1 of [SW22]).

Theorem 2.5 (Binary observation model)**.** *Given parameters* n, k, q, s, M, M', x, x'*, define distributions* $\mathbb{P} = \mathbb{P}_{\text{Binary}}(n, k, q, s, M, x)$ *and* $\mathbb{Q} = \mathbb{P}_{\text{Binary}}(n, k, q, s, M', x')$*. Assume that* $M \min_\ell x_\ell \geqslant C$ *and* $M' \min_\ell x'_\ell \geqslant C$ *for some constant* $C > 0$ *and that* $q + s/(\min_\ell x_\ell) \leqslant \tau_1$ *for some constant* $\tau_1 < 1$*. Write* $\widetilde{M} = |M - M'|$ *and* $\widehat{M} = \max\{M, M'\}$*. We have:*

- *If* $D^5\widehat{M}^2(s^2/q)(k^2/n \vee 1) = o(1)$*, then no degree-*$D$ *test weakly separates* $\mathbb{P}$ *and* $\mathbb{Q}$*.*
- *If* $\widetilde{M}^{2/3}(s^2/q)k^2/n = \omega(1)$*,* $\widehat{M}^{-1/3}sk = \omega(1)$ *and* $\widetilde{M}^2k/\widehat{M}^2 = \omega(1)$ *then there exists a degree-3 test that strongly separates* $\mathbb{P}$ *and* $\mathbb{Q}$*.*

The upper and lower bounds match (up to log factors) provided $k^2 \geqslant n$, $\widehat{M} = O(1)$, $D \leqslant \mathrm{polylog}(n)$, and $q \geqslant 1/n$. The condition $q \geqslant 1/n$ is natural since without it there will be isolated vertices. The regime $k^2 < n$ is more complicated, and some open questions remain here even for simpler testing and recovery problems than those we study here; see Section 2.4.1 of [SW22] for discussion.

2.3 Formal relation to low-degree recovery

While there has been much work on studying the low degree polynomial framework for measuring the computational hardness of problems when one wishes to *detect* the presence of hidden structures, a recent line of work [SW22] has extended the framework to the setting where one wishes to *recover* a planted structure from noisy data. We use this recent work to show a connection between detection and recovery — namely, for any given recovery problem where one wishes to estimate some scalar quantity $g(X) \in \mathbb{R}$ of the planted distribution (e.g. $g(X) = X_1$) from some observation $Y \in \mathbb{R}^N$, there is a specific testing problem with the same computational hardness regimes and, in the other direction, for any testing problem $H_0 : Y \sim \mathbb{P}$ versus $H_1 : Y \sim \mathbb{Q}$, as long as $\mathbb{P}_X \ll \mathbb{Q}_X$ so that the likelihood ratio of the signal $d\mathbb{P}_X/d\mathbb{Q}_X$ exists, there is a corresponding recovery problem with the same computational hardness regimes.

To show these equivalences, we compare the object Corr from [SW22, Eq. (2)] to our notion of advantage Adv. In particular, we have

$$\mathsf{Corr}_{\leqslant D}(g(X), \widetilde{\mathbb{Q}}) := \sup_{f \text{ deg } D} \frac{\mathbb{E}_{\widetilde{\mathbb{Q}}}[g(X)f(Y)]}{\sqrt{\mathbb{E}_{\widetilde{\mathbb{Q}}}[f(Y)^2]}}, \tag{2.1}$$

where we have written $g(X)$ to indicate that the scalar quantity we are estimating depends on the underlying unknown signal X, and we treat $\widetilde{\mathbb{Q}}$ as the joint distribution of (X, Y). We also define Corr' that renormalizes Corr by dividing through by $\mathbb{E}_{\widetilde{\mathbb{Q}}}[g(X)]$:

$$\mathsf{Corr}'_{\leqslant D}(g(X), \widetilde{\mathbb{Q}}) := \frac{1}{\mathbb{E}_{\widetilde{\mathbb{Q}}}[g(X)]} \sup_{f \text{ deg } D} \frac{\mathbb{E}_{\widetilde{\mathbb{Q}}}[g(X)f(Y)]}{\sqrt{\mathbb{E}_{\widetilde{\mathbb{Q}}}[f(Y)^2]}}. \tag{2.2}$$

We compare Corr' to our advantage Adv defined in (2.4),

$$\mathsf{Adv}_{\leqslant D}(\mathbb{P}, \mathbb{Q}) := \sup_{f \text{ deg } D} \frac{\mathbb{E}_{\mathbb{P}}[f(Y)]}{\sqrt{\mathbb{E}_{\mathbb{Q}}[f(Y)^2]}}, \tag{2.3}$$

and, loosely, we notice that $\mathsf{Adv}_{\leqslant D}(\mathbb{P}, \mathbb{Q})$ equals $\mathsf{Corr}'_{\leqslant D}(g(X), \widetilde{\mathbb{Q}})$ when $\mathbb{Q}$ equals $\widetilde{\mathbb{Q}}$ and $\mathbb{P}$ is defined such that $\mathbb{E}_{\mathbb{P}}[f(Y)] = \mathbb{E}_{\widetilde{\mathbb{Q}}}[g(X)f(Y)]/\mathbb{E}_{\widetilde{\mathbb{Q}}}[g(X)]$. This gives rise to the following proposition.

In the testing and recovery problems respectively, the conjectured-hard region, i.e. where low-degree algorithms fail, is that where Adv and Corr' respectively are $1 + o(1)$ which for testing rules out f which weakly distinguish f in the testing problem (and $O(1)$ which rules out f which strongly distinguish).

For the case of recovery the notion of failure relates to the low-degree minimum mean squared error $\mathrm{MMSE}_{\leqslant D} := \inf_{f \text{ deg } D} \mathbb{E}_{\widetilde{\mathbb{Q}}}[(f(Y) - g(X))^2]$. Note the trivial estimator $f(Y) = \mathbb{E}_{\widetilde{\mathbb{Q}}}[g(X)]$, a constant polynomial, achieves MSE of $\mathbb{E}_{\widetilde{\mathbb{Q}}}[g(X)^2] - \mathbb{E}_{\widetilde{\mathbb{Q}}}[g(X)]^2$ and by Fact 1.1 of [SW22],

$$\mathrm{MMSE}_{\leqslant D} = \mathbb{E}_{\widetilde{\mathbb{Q}}}[g(X)^2] - \mathbb{E}_{\widetilde{\mathbb{Q}}}[g(X)]^2 \mathsf{Corr}'^2_{\leqslant D}(g(X), \widetilde{\mathbb{Q}}).$$

Thus $\mathsf{Corr}' = 1 + o(1)$ rules out f which do substantially better than the trivial estimator.

In the following proposition we assume the probability distributions are either discrete or have a density function.

Proposition 2.6.

1. *Given any recovery problem to estimate non-negative $g(X)$ given Y with joint distribution $(X, Y) \sim \mathbb{Q}$ and $0 < \mathbb{E}_{\mathbb{Q}}[g(X)] < \infty$, there exists a joint distribution $(X, Y) \sim \mathbb{P}$ such that for the planted testing problem $H_0: (X, Y) \sim \mathbb{Q}$ vs $H_1: (X, Y) \sim \mathbb{P}$,*

$$\mathsf{Adv}_{\leqslant D}(\mathbb{P}, \mathbb{Q}) = \mathsf{Corr}'_{\leqslant D}(g(X), \mathbb{Q}).$$

2. *Given a testing problem between two planted distributions $H_0 : (X, Y) \sim \mathbb{Q}$ and $H_1 : (X, Y) \sim \mathbb{P}$, if $\ell(X) := d\mathbb{P}_X / d\mathbb{Q}_X$ exists and $Y|X$ is the same under $\mathbb{P}$ and $\mathbb{Q}$, then for the estimation problem $g(X) = \ell(X)$ in $\mathbb{Q}$,*

$$\mathsf{Adv}_{\leqslant D}(\mathbb{P}, \mathbb{Q}) = \mathsf{Corr}'_{\leqslant D}(g(X), \mathbb{Q}).$$

The proof of Proposition 2.6 is included in Section 6, where we also show that our Propositions 2.8 and 2.9 generalize the cumulant upper bounds on Corr shown in [SW22].

Consider the important special case when we estimate the indicator of an event A on the signal, meaning that $g(X) = \mathbb{I}\{A\}$. In this case, Proposition 2.6 shows that the equivalent testing problem is one where we test between $H_1 : X \sim \mathbb{P}$ versus $H_0 : X \sim \mathbb{Q}$ where $\mathbb{Q}$ is the measure describing the randomness in the observed Y, and $\mathbb{P}$ is the measure of Y conditional on the occurrence of event A.

2.4 Proof overview

Main quantity to bound: advantage. In order to rule out weak separation between distributions $\mathbb{P} = \mathbb{P}_n$ and $\mathbb{Q} = \mathbb{Q}_n$ on $\mathbb{R}^{N_n}$, it will suffice to bound the degree-D "advantage," named as such to emphasize that it measures the ability of low-degree polynomials to outperform random guessing. The degree-D advantage, $\mathsf{Adv}_{\leqslant D}$, is defined as

$$\mathsf{Adv}_{\leqslant D}(\mathbb{P}, \mathbb{Q}) := \sup_{f \text{ deg } D} \frac{\mathbb{E}_{\mathbb{P}}[f]}{\sqrt{\mathbb{E}_{\mathbb{Q}}[f^2]}}, \tag{2.4}$$

where f ranges over polynomials $\mathbb{R}^N \to \mathbb{R}$ of degree at most D. The quantity $\mathsf{Adv}_{\leqslant D}$ is also the *norm of the degree-D likelihood ratio* (see [Hop18, KWB22]), but we will not use this interpretation here as the likelihood ratio is difficult to work with in our setting. We note that while the notion of separation is symmetric between $\mathbb{P}$ and $\mathbb{Q}$, the notion of advantage is not; for our purposes, we could just as easily work with $\mathsf{Adv}_{\leqslant D}(\mathbb{Q}, \mathbb{P})$ instead of $\mathsf{Adv}_{\leqslant D}(\mathbb{P}, \mathbb{Q})$. The following basic fact connects $\mathsf{Adv}_{\leqslant D}$ with strong/weak separation.

Lemma 2.7. *Fix a sequence $D = D_n$.*

- *If* $\mathsf{Adv}_{\leqslant D}(\mathbb{P},\mathbb{Q}) = O(1)$ *then no degree-*D *test strongly separates* $\mathbb{P}$ *and* $\mathbb{Q}$*.*
- *If* $\mathsf{Adv}_{\leqslant D}(\mathbb{P},\mathbb{Q}) = 1 + o(1)$ *then no degree-*D *test weakly separates* $\mathbb{P}$ *and* $\mathbb{Q}$*.*

The proof of Lemma 2.7, along with the proofs of all facts in this section, can be found in Section 5. In light of Lemma 2.7, it remains to bound $\mathsf{Adv}_{\leqslant D}$. We will provide a few general-purpose bounds, one for Gaussian problems and one for binary-valued problems. Both will involve recursively-defined quantities r_α, introduced in what follows.

Recursive definition for r_α. Suppose X is a random variable taking values in $\mathbb{R}^N$, which may have a different distribution under $\mathbb{P}$ and $\mathbb{Q}$. For $\alpha, \beta \in \mathbb{N}^N$ where $\mathbb{N} = \{0, 1, 2, \ldots\}$, define

$$|\alpha| := \sum_i \alpha_i, \qquad \alpha! := \prod_i \alpha_i!, \qquad \binom{\alpha}{\beta} := \prod_i \binom{\alpha_i}{\beta_i}, \qquad \text{and} \qquad X^\alpha := \prod_i X_i^{\alpha_i}.$$

Also define $\beta \leqslant \alpha$ to mean "$\beta_i \leqslant \alpha_i$ for all i" and define $\beta \lneq \alpha$ to mean "$\beta_i \leqslant \alpha_i$ for all i and for some i the inequality is strict: $\beta_i < \alpha_i$." With this notation in hand, define $r_\alpha = r_\alpha(X) \in \mathbb{R}$ for $\alpha \in \mathbb{N}^N$ recursively by

$$r_\alpha = \mathbb{E}_{\mathbb{P}}\left[X^\alpha\right] - \sum_{0 \leqslant \beta \lneq \alpha} r_\beta \binom{\alpha}{\beta} \mathbb{E}_{\mathbb{Q}}\left[X^{\alpha-\beta}\right]. \tag{2.5}$$

Bounds on advantage. We have the following general-purpose bounds on $\mathsf{Adv}_{\leqslant D}$ in terms of r_α defined in (2.5). The proofs are inspired by [SW22] and can be found in Section 5.

Proposition 2.8 (General additive Gaussian model)**.** *Suppose* $\mathbb{P}$ *and* $\mathbb{Q}$ *take the following form: to sample* $Y \sim \mathbb{P}$ *(or* $Y \sim \mathbb{Q}$*, respectively), first sample* $X \in \mathbb{R}^N$ *from an arbitrary prior* $\mathbb{P}_X$ *(or* $\mathbb{Q}_X$*, resp.), then sample* $Z \sim \mathcal{N}(0, I_N)$*, and set* $Y = X + Z$*. Define* r_α *as in* (2.5)*. Then*

$$\mathsf{Adv}_{\leqslant D}(\mathbb{P},\mathbb{Q}) \leqslant \sqrt{\sum_{\alpha \in \mathbb{N}^N,\, |\alpha| \leqslant D} \frac{r_\alpha^2}{\alpha!}}.$$

Proposition 2.9 (General binary observation model)**.** *Suppose* $\mathbb{P}$ *and* $\mathbb{Q}$ *each take the following form. To sample* $Y \sim \mathbb{P}$ *(or* $Y \sim \mathbb{Q}$*, respectively), first sample* $X \in \mathbb{R}^N$ *from an arbitrary prior* $\mathbb{P}_X$ *(or* $\mathbb{Q}_X$*, resp.) supported on* $X \in [\tau_0, \tau_1]^N$ *with* $0 < \tau_0 \leqslant \tau_1 < 1$*, then sample* $Y \in \{0,1\}^N$ *with entries conditionally independent given* X *and* $\mathbb{E}[Y_i|X] = X_i$*. Define* r_α *as in* (2.5)*. Then*

$$\mathsf{Adv}_{\leqslant D}(\mathbb{P},\mathbb{Q}) \leqslant \sqrt{\sum_{\alpha \in \{0,1\}^N,\, |\alpha| \leqslant D} \frac{r_\alpha^2}{(\tau_0(1-\tau_1))^{|\alpha|}}}.$$

Combinatorial properties of the r_α. The upshot of the two propositions above is that to show hardness of distinguishing $\mathbb{P}$ versus $\mathbb{Q}$, it suffices to bound the recursively defined r_α. This task is made easier by indentifying combinatorial properties that the r_α enjoy. In Section 3, we show general results for how properties of the probability spaces transfer to behaviour of the r_α. We may consider $\alpha \in \mathbb{N}^N$ as a multigraph (see Section 3.1), and we word the results in this language. Loosely speaking, the results we present are as follows:

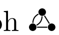
- If $\mathbb{P}$ and $\mathbb{Q}$ are multiplicative for disjoint graphs α and β then $r_{\alpha\cup\beta} = r_\alpha r_\beta$ (Lemma 3.1).
- If $\mathbb{E}_{\mathbb{P}}[X^\tau] = \mathbb{E}_{\mathbb{Q}}[X^\tau]$ for all trees τ, then the r value is zero on trees (Lemma 3.2).
- The r-values are indifferent to constant shifts to X (Lemma 3.3).
- If we scale X by a constant factor c to construct $\widetilde{r}$, then $\widetilde{r}_\alpha = c^{|\alpha|} r_\alpha$ (Lemma 3.4).

Putting it all together. In Section 4, we pivot back to considering our particular probability spaces $\mathbb{P}$ and $\mathbb{Q}$ and calculate the expected value of X^α as a function of properties of the graph α (see Lemma 4.1). This, together with the multiplicative and tree results for the r_α, allow us to bound Adv in the Gaussian case. The scaling and shifting properties of the r_α are used to show we can deduce the graph case from the Gaussian case.

3 The recursive algebra of planted vs planted

The recursively defined r_α play a central role in our proof. By Proposition 2.8, the 'advantage' Adv is bounded above by a sum of squares of the r_α; therefore, to show low-degree hardness for a distinguishing problem, it is enough to control the size of this sum of squares. In this section, we explore the combinatorial behaviour of these r_α and show that they exhibit very nice properties under only mild assumptions on the probability spaces P and Q. (We will write P and Q for the probability spaces in this section, both to emphasize that these results hold for general probability spaces and to ease the notational burden.) We assume throughout that both P and Q are symmetric, i.e. they are supported on X for which $X_{ij} = X_{ji}$.

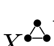
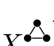
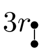
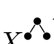
3.1 The graph interpretation

As in [SW22], it will be convenient to think of $\alpha \in \mathbb{N}^N$ as a multigraph, possibly with self-loops, on the vertex set $[n]$, where $N = n(n+1)/2$ and we take α_{ij}, for each $i \leqslant j$, to be the number of edges between vertices i and j. For example, for $n = 3$, $N = 6$ and if we fix the order to be $\alpha = (\alpha_{11}, \alpha_{12}, \alpha_{22}, \alpha_{13}, \alpha_{23}, \alpha_{33})$ then $(0, 2, 0, 1, 1, 0)$ is the graph  and, for $n = 2$, $N = 3$, $(1, 1, 0)$ is the graph  a single edge with a loop at one vertex. For graphs α and β, we consider β to be a subgraph of α, denoted $\beta \subseteq \alpha$, if the labelled edge set of β is a subset of the labelled edge set of α. For example, for the graph $\alpha =$ , the graphs $\beta =$  and $\beta' =$  are distinct subgraphs. For graph α and subgraph $\beta \subseteq \alpha$, define $\alpha \backslash \beta$ to be the graph obtained from α by first removing the labelled edges in β and then removing isolated vertices - e.g. $\backslash$ $=$ (not ). Similarly, let $\alpha \cap \beta$ denote the graph obtained by first taking the intersection of both graphs and then deleting any isolated vertices.

The usefulness of considering graphs with labelled edge sets is that it simplifies the expression for the recursion in the definition of r_α. To avoid confusion, write v_α for the vector that maps to the graph α. Note, first, that if $v_\beta \leqslant v_\alpha$, then $\beta \subseteq \alpha$ and vice-versa. However, the counts are different. For fixed α and β, with $v_\beta \leqslant v_\alpha$, there are $\binom{v_\alpha}{v_\beta}$ many distinct edge-labelled graphs β' such that $\beta' \subseteq \alpha$ and $v_{\beta'} = v_\beta$. Hence, for edge-labelled graphs, the equivalent recursive definition to (2.5) is as follows, where the sum is over edge-labelled subgraphs. For graph α, the term r_α is defined recursively by

$$r_\alpha = \mathbb{E}_P\left[X^\alpha\right] - \sum_{\varnothing \subseteq \beta \subsetneq \alpha} r_\beta \, \mathbb{E}_Q\left[X^{\alpha\backslash\beta}\right], \tag{3.1}$$

starting from the base case of the empty graph $r_\varnothing = 1$. For example, if the probability spaces are exchangeable (i.e. $\beta =$  and $\beta' =$  etc. have the same expectations under both P and Q) then

r $= \mathbb{E}_P\big[X$$\big] - \mathbb{E}_Q\big[X$$\big] - 3r$ $\mathbb{E}_Q\big[X$$\big] - 3r$ $\mathbb{E}_Q\big[X$$\big]$.

We may also define the r_α non-recursively as follows — see Remark 3.5 and the end of this section where we show the equivalence of the two definitions. For any set α let $\mathcal{P}(\alpha)$ be the set of partitions of α. For partition τ let $|\tau|$ denote the number of parts and $\gamma \in \tau$ indicates that the set γ is a part in the partition τ.

$$r_\alpha = \sum_{\varnothing \subseteq \delta \subseteq \alpha} \mathbb{E}_P[X^\delta] \sum_{\tau \in \mathcal{P}(\alpha \setminus \delta)} (-1)^{|\tau|}\, |\tau|! \prod_{\gamma \in \tau} \mathbb{E}_Q[X^\gamma] \tag{3.2}$$

We will use the notion of edge-labelled subgraphs, denoted $\subseteq$, to aid the proofs, but for the rest of the paper we consider $\alpha \in \mathbb{N}^N$, or equivalently α a graph without edge labels, and denote by $\leqslant$ the non-labelled subset or subgraph relation.

3.2 Combinatorial properties of the r-values

We will be interested in how properties of the probability spaces transfer to the behaviour of r_α. We will see that the r_α behave multiplicatively over taking disjoint unions if the following property holds for P and Q. Let $\alpha \cup \beta$ denote the disjoint union of α and β. We say the probability space A is multiplicative over disjoint unions if (3.3) holds:

$$\mathbb{E}_A\left[X^{\alpha\cup\beta}\right] = \mathbb{E}_A\left[X^\alpha\right]\mathbb{E}_A\left[X^\beta\right] \text{ for any graphs } \alpha \text{ and } \beta. \tag{3.3}$$

Lemma 3.1. *Suppose P and Q are symmetric and multiplicative over disjoint unions, i.e. they satisfy* (3.3)*, then for any α and β, we have $r_{\alpha\cup\beta} = r_\alpha r_\beta$.*

Proof. We proceed by induction on the number of edges. *Base case.* Suppose α consists of two disjoint edges, denoted by $\mid$ and $\vdots$. Then from (3.1),

$$r_{\mid\vdots} = \mathbb{E}_P\left[X^{\mid}\right]\mathbb{E}_P\left[X^{\vdots}\right] - r_\varnothing\,\mathbb{E}_Q\left[X^{\mid}\right]\mathbb{E}_Q\left[X^{\vdots}\right] - r_{\mid}\,\mathbb{E}_Q\left[X^{\vdots}\right] - r_{\vdots}\,\mathbb{E}_Q\left[X^{\mid}\right],$$

where we used the multiplicative property of (3.3) to deduce $\mathbb{E}_P\left[X^{\mid\vdots}\right] = \mathbb{E}_P\left[X^{\mid}\right]\mathbb{E}_P\left[X^{\vdots}\right]$ and, similarly, $\mathbb{E}_Q\left[X^{\mid\vdots}\right] = \mathbb{E}_Q\left[X^{\mid}\right]\mathbb{E}_Q\left[X^{\vdots}\right]$. Now substituting $r_\varnothing = 1$ along with $r_{\mid} = \mathbb{E}_P\left[X^{\mid}\right] - \mathbb{E}_Q\left[X^{\mid}\right]$ and the corresponding expression for $r_{\vdots}$, we get

$$r_{\mid\vdots} = \mathbb{E}_P\left[X^{\mid}\right]\mathbb{E}_P\left[X^{\vdots}\right] - \mathbb{E}_P\left[X^{\mid}\right]\mathbb{E}_Q\left[X^{\vdots}\right] - \mathbb{E}_P\left[X^{\vdots}\right]\mathbb{E}_Q\left[X^{\mid}\right] + \mathbb{E}_Q\left[X^{\mid}\right]\mathbb{E}_Q\left[X^{\vdots}\right] = r_{\mid} r_{\vdots}.$$

Inductive step. Fix $\tau = \alpha \cup \beta$ (where α, β aare disjoint) and assume the factorization of r holds for graphs with fewer than $|\tau| = |\alpha| + |\beta|$ edges. For any graph γ define z_γ by $z_\gamma := \mathbb{E}_P\left[X^\gamma\right] - r_\gamma$. Then, first note

$$z_\alpha\, z_\beta = \mathbb{E}_P\left[X^\alpha\right]\mathbb{E}_P\left[X^\beta\right] - r_\alpha\,\mathbb{E}_P\left[X^\beta\right] - r_\beta\,\mathbb{E}_P\left[X^\alpha\right] + r_\alpha\, r_\beta,$$

and that because α and β are disjoint and P satisfies (3.3), we have $\mathbb{E}_P\left[X^\alpha\right]\mathbb{E}_P\left[X^\beta\right] = \mathbb{E}_P\left[X^{\alpha\cup\beta}\right]$. Hence,

$$z_\alpha\, z_\beta = \mathbb{E}_P\left[X^{\alpha\cup\beta}\right] - r_\alpha\,\mathbb{E}_P\left[X^\beta\right] - r_\beta\,\mathbb{E}_P\left[X^\alpha\right] + r_\alpha\, r_\beta,$$

and now (back) substituting $\mathbb{E}_P\left[X^\alpha\right] = r_\alpha + z_\alpha$ and $\mathbb{E}_P\left[X^\beta\right] = r_\beta + z_\beta$ we find

$$r_\alpha\, r_\beta = \mathbb{E}_P\left[X^{\alpha\cup\beta}\right] - r_\alpha\, z_\beta - r_\beta\, z_\alpha - z_\alpha\, z_\beta.$$

By definition, $r_{\alpha\cup\beta} = \mathbb{E}_P\left[X^{\alpha\cup\beta}\right] - z_{\alpha\cup\beta}$; therefore, to complete the proof, it suffices to show the identity

$$z_{\alpha\cup\beta} = z_\beta\, r_\alpha + z_\alpha\, r_\beta + z_\alpha\, z_\beta. \tag{3.4}$$

Again, write $\tau = \alpha \cup \beta$ and note that by the definitions of r_τ and z_τ, we have

$$z_\tau = \mathbb{E}_P\left[X^\tau\right] - r_\tau = \sum_{\varnothing \subseteq \gamma \subsetneq \tau} r_\gamma \mathbb{E}_Q\left[X^{\tau \setminus \gamma}\right]. \tag{3.5}$$

Now observe that for any $\gamma \subsetneq \tau$, because $\tau = \alpha \cup \beta$, we have $\gamma = \gamma_\alpha \cup \gamma_\beta$ where $\gamma_\alpha = \gamma \cap \alpha$ and $\gamma_\beta = \gamma \cap \beta$. Thus, γ is a disjoint union of γ_α and γ_β with strictly fewer total edges than $\alpha \cup \beta$. Therefore, $(\alpha \cup \beta)\setminus\gamma = (\alpha\setminus\gamma_\alpha) \cup (\beta\setminus\gamma_\beta)$ where $\alpha\setminus\gamma_\alpha$ and $\beta\setminus\gamma_\beta$ are disjoint, so by assumption, $\mathbb{E}_Q\left[X^{(\alpha\cup\beta)\setminus\gamma}\right] = \mathbb{E}_Q\left[X^{\alpha\setminus\gamma_\alpha}\right]\mathbb{E}_Q\left[X^{\beta\setminus\gamma_\beta}\right]$ and by the inductive hypothesis $r_\gamma = r_{\gamma_\alpha} r_{\gamma_\beta}$. Hence, for any fixed $\gamma \subsetneq \alpha \cup \beta$,

$$r_\gamma\, \mathbb{E}_Q\left[X^{(\alpha\cup\beta)\setminus\gamma}\right] = r_{\gamma_\alpha}\, \mathbb{E}_Q\left[X^{\alpha\setminus\gamma_\alpha}\right] r_{\gamma_\beta}\, \mathbb{E}_Q\left[X^{\beta\setminus\gamma_\beta}\right]. \tag{3.6}$$

There are two special cases for (3.6). If $\gamma_\alpha = \alpha$, then

$$r_\gamma\, \mathbb{E}_Q\left[X^{(\alpha\cup\beta)\setminus\gamma}\right] = r_\alpha\, r_{\gamma_\beta}\, \mathbb{E}_Q\left[X^{\beta\setminus\gamma_\beta}\right], \tag{3.7}$$

and symmetrically for the case $\gamma_\beta = \beta$. Note that because γ is a strict subgraph of $\alpha \cup \beta$ either of $\gamma_\alpha = \alpha$ or $\gamma_\beta = \beta$ may hold but not both.

In the expression for $z_\tau = z_{\alpha\cup\beta}$ in (3.5) we take the sum over $\{\gamma : \varnothing \subseteq \gamma \subsetneq \alpha \cup \beta\}$ and partition it into sums over the sets $S_1 = \{\gamma : \gamma_\alpha = \alpha,\ \varnothing \subseteq \gamma_\beta \subsetneq \beta\}$, $S_2 = \{\gamma : \varnothing \subseteq \gamma_\alpha \subsetneq \alpha,\ \gamma_\beta = \beta\}$ and $S_3 = \{\gamma : \varnothing \subseteq \gamma_\alpha \subsetneq \alpha,\ \varnothing \subseteq \gamma_\beta \subsetneq \beta\}$. We begin with S_1. By (3.7),

$$\sum_{\gamma \in S_1} r_\gamma\, \mathbb{E}_Q\left[X^{(\alpha\cup\beta)\setminus\gamma}\right] = \sum_{\varnothing \subseteq \gamma_\beta \subsetneq \beta} r_\alpha\, r_{\gamma_\beta}\, \mathbb{E}_Q\left[X^{\beta\setminus\gamma_\beta}\right] = r_\alpha\, z_\beta. \tag{3.8}$$

The final step uses that $z_\beta = \sum_{\varnothing \subseteq \gamma_\beta \subsetneq \beta} r_{\gamma_\beta}\, \mathbb{E}_Q\left[X^{\beta\setminus\gamma_\beta}\right]$, similar to (3.5). Next, taking the sum over S_2 yields $r_\beta\, z_\alpha$ in the same way as in (3.8). Lastly, the sum over S_3 is given by

$$\sum_{\gamma \in S_3} r_\gamma\, \mathbb{E}_Q\left[X^{(\alpha\cup\beta)\setminus\gamma}\right] = \sum_{\varnothing \subseteq \gamma_\alpha \subsetneq \alpha,\, \varnothing \subseteq \gamma_\beta \subsetneq \beta} r_{\gamma_\alpha}\, \mathbb{E}_Q\left[X^{\alpha\setminus\gamma_\alpha}\right] r_{\gamma_\beta}\, \mathbb{E}_Q\left[X^{\beta\setminus\gamma_\beta}\right] = z_\alpha\, z_\beta. \tag{3.9}$$

By (3.5), we see that $z_\tau = z_{\alpha\cup\beta}$ can be obtained as the following sum:

$$\begin{aligned} z_{\alpha\cup\beta} = z_\tau = \sum_{\varnothing \subseteq \gamma \subsetneq \tau} r_\gamma \mathbb{E}_Q\left[X^{\tau\setminus\gamma}\right] &= \sum_{\gamma \in S_1} r_\gamma \mathbb{E}_Q\left[X^{\tau\setminus\gamma}\right] + \sum_{\gamma \in S_2} r_\gamma \mathbb{E}_Q\left[X^{\tau\setminus\gamma}\right] + \sum_{\gamma \in S_3} r_\gamma \mathbb{E}_Q\left[X^{\tau\setminus\gamma}\right] \\ &= r_\alpha\, z_\beta + r_\beta\, z_\alpha + z_\alpha\, z_\beta. \end{aligned}$$

where in the final step above we have used results (3.8)-(3.9). Thus, we find (3.4), confirming the identity as required. □

Lemma 3.2. *For all τ where τ is a forest, meaning a graph with no cycles, suppose that P and Q satisfy $\mathbb{E}_P\left[X^\tau\right] = \mathbb{E}_Q\left[X^\tau\right]$. Then, $r_\alpha = 0$ for any forest graph α.*

Proof. The proof is almost immediate by induction on the number of edges. For the base case we note $r_{\bullet\!-\!\bullet} = \mathbb{E}_P\left[X^{\bullet\!-\!\bullet}\right] - \mathbb{E}_Q\left[X^{\bullet\!-\!\bullet}\right] = 0$. For any fixed forest α and $\beta \subsetneq \alpha$, the graph β is a forest on strictly fewer edges and so by induction $r_\beta = 0$, but then $r_\alpha = \mathbb{E}_P\left[X^\alpha\right] - \mathbb{E}_Q\left[X^\alpha\right] = 0$. □

We also show that one can add a constant shift to the distribution without changing the values of the r_α. The proof is somewhat technical, so we relegate it to Section 5.

Lemma 3.3. *Let $\widetilde{X}$ be defined by $\widetilde{X}_{ij} = X_{ij} + y_{ij}$ where $y_{ij} \in \mathbb{R}$ is non-random for each pair i, j. For any probability spaces P and Q, let $r_\alpha = r_\alpha(P, Q, X)$ and $\widetilde{r}_\beta = \widetilde{r}_\beta(P, Q, \widetilde{X})$. Then, for all γ, we have that $r_\gamma = \widetilde{r}_\gamma$.*

The following lemma concerns the effect on r of scaling.

Lemma 3.4. *Fix $a \in \mathbb{R}$ and $a \neq 0$. Let $\widetilde{X}$ be defined by $\widetilde{X}_{ij} = aX_{ij}$. Then for any probability spaces P and Q, for $r_\alpha = r_\alpha(P, Q, X)$ and $\widetilde{r}_\beta = \widetilde{r}_\beta(P, Q, \widetilde{X})$, and for all γ, we have that $\widetilde{r}_\gamma = a^{|\gamma|} r_\gamma$, where $|\gamma|$ equals the number of edges in the graph γ, i.e. $|\gamma| = |E(\gamma)|$.*

Proof. This proof is a simple induction on $|\alpha|$. The base case is easy as $\widetilde{r}_\varnothing = r_\varnothing = 1$ as required. Now, fix α for some $|\alpha| > 1$, and assume we have proven the result for $|\beta| < |\alpha|$. Notice, that by definition, we have $\mathbb{E}_P\left[\widetilde{X}^\alpha\right] = a^{|\alpha|}\mathbb{E}_P\left[X^\alpha\right]$ and $\mathbb{E}_Q\left[\widetilde{X}^\beta\right] = a^{|\beta|}\mathbb{E}_Q\left[X^\beta\right]$. Therefore,

$$\widetilde{r}_\alpha = \mathbb{E}_P\left[\widetilde{X}^\alpha\right] - \sum_{\varnothing \subseteq \beta \subsetneq \alpha} \widetilde{r}_{\alpha\setminus\beta}\, \mathbb{E}_Q\left[\widetilde{X}^\beta\right] = a^{|\alpha|}\mathbb{E}_P\left[X^\alpha\right] - \sum_{\varnothing \subseteq \beta \subsetneq \alpha} \widetilde{r}_{\alpha\setminus\beta}\, a^{|\beta|}\mathbb{E}_Q\left[X^\beta\right].$$

By the inductive hypothesis, $\widetilde{r}_{\alpha\setminus\beta} = a^{|\alpha\setminus\beta|} r_{\alpha\setminus\beta}$ and so by the equation above we are done, as $a^{|\beta|}a^{|\alpha\setminus\beta|} = a^{|\alpha|}$. □

Remark 3.5. *We show that the non-recursive formula for the r_α given in* (3.2) *is equivalent to the recursive one in* (3.1)*. For convenience we restate* (3.2) *below:*

$$r_\alpha(P, Q) = \sum_{\varnothing \subseteq \delta \subseteq \alpha} \mathbb{E}_P[X^\delta] \sum_{\tau \in \mathcal{P}(\alpha\setminus\delta)} (-1)^{|\tau|}\, |\tau|! \prod_{\gamma \in \tau} \mathbb{E}_Q[X^\gamma]. \tag{3.10}$$

This follows by induction. Small cases may be checked. For the inductive step, assume the formula (3.10) *holds for all sets of size at most k and let α be any set of size $k+1$. Then by* (3.1)*, and noting the induction assumption holds for each β we have the following, writing a_δ for $\mathbb{E}_P[X^\delta]$ and b_γ for $\mathbb{E}_Q[X^\gamma]$,*

$$r_\alpha(P, Q) = a_\alpha - \sum_{\varnothing \subseteq \beta \subsetneq \alpha} r_\beta b_{\alpha\setminus\beta} = a_\alpha - \sum_{\varnothing \subseteq \beta \subsetneq \alpha} b_{\alpha\setminus\beta} \sum_{\varnothing \subseteq \delta \subseteq \beta} a_\delta \sum_{\tau \in \mathcal{P}(\beta\setminus\delta)} (-1)^{|\tau|}\, |\tau|! \prod_{\gamma \in \tau} b_\gamma.$$

We may now swap the order of summation to yield

$$r_\alpha(P, Q) = a_\alpha - \sum_{\varnothing \subseteq \delta \subsetneq \alpha} a_\delta \sum_{\varnothing \subseteq \beta \subseteq \alpha\setminus\delta} b_{\alpha\setminus\beta} \sum_{\tau \in \mathcal{P}(\beta\setminus\delta)} (-1)^{|\tau|}\, |\tau|! \prod_{\gamma \in \tau} b_\gamma.$$

For τ a partition of $\beta\setminus\delta$, let τ' be the partition constructed from τ by adding the part $\alpha\setminus\beta$ and note that τ' is a partition of $\alpha\setminus\delta$. Note each τ' appears $|\tau'| = |\tau| + 1$ times in the sum above, and hence

$$r_\alpha(P, Q) = a_\alpha - \sum_{\varnothing \subseteq \delta \subsetneq \alpha} a_\delta \sum_{\tau' \in \mathcal{P}(\alpha\setminus\delta)} (-1)^{|\tau'|-1}\, |\tau'|! \prod_{\gamma \in \tau'} b_\gamma$$

which matches the non-recursive expression for r as claimed.

4 Proof of Theorems 2.4 and 2.5

In this section we give the full proofs of the main results, Theorems 2.4 and 2.5.

Proof of Theorem 2.4. Hard regime. We start by proving the computational lower bound. By definition of the Additive Gaussian model, we can write the observed edge weights $Y = \{Y_{ij}\}_{i \leqslant j}$ as $Y = X + Z$, where Z consists of i.i.d. $\mathcal{N}(0,1)$ entries, and

$$X_{ij} = \begin{cases} \lambda/x_\ell & \sigma_i = \sigma_j = \ell \ \text{ for some } \ell \in [M], \\ 0 & \text{otherwise.} \end{cases} \tag{4.1}$$

Recall the sequence r_α, defined recursively via

$$r_\alpha = \mathbb{E}_{\mathbb{P}}\left[X^\alpha\right] - \sum_{0 \leqslant \beta \lneq \alpha} r_\beta \binom{\alpha}{\beta} \mathbb{E}_{\mathbb{Q}}\left[X^{\alpha-\beta}\right].$$

By Lemma 2.7 and Proposition 2.8, we have that if

$$\sum_{\alpha : |\alpha| \leqslant D} \frac{r_\alpha^2}{\alpha!} = 1 + o(1), \tag{4.2}$$

then no degree-D test can weakly separate $\mathbb{P}$ and $\mathbb{Q}$. Thus, to prove the computational lower bound, it suffices to show that (4.2) holds for $D^5 \lambda^2 M^2 (k^2/n \vee 1) = o(1)$.

We will demonstrate (4.2) by proving the following three facts. (We consider the sets α as graphs and write $V(\alpha)$ for the vertex set and $C(\alpha)$ for the set of connected components, see Section 3.1 for details.)

(i) For all α, the term r_α factorizes over the connected components of α. That is,

$$r_\alpha = \prod_{\beta \in \mathcal{C}(\alpha)} r_\beta.$$

(ii) If at least one connected component of α is a tree, then $r_\alpha = 0$.

(iii) For all α, where $|\alpha| = |E(\alpha)|$ counts the edges in the graph α,

$$|r_\alpha| \leqslant (|\alpha| + 1)^{|\alpha|} \left(\frac{\widehat{M}\lambda}{C}\right)^{|\alpha|} \left(\frac{k}{n}\right)^{|V(\alpha)|}.$$

Fact (i) follows directly from Lemma 3.1. To see Fact (ii), we note that by Lemma 3.2 it suffices to show that for any τ a tree, we have $\mathbb{E}_{\mathbb{P}}\left[X^\tau\right] = \mathbb{E}_{\mathbb{Q}}\left[X^\tau\right]$. But recall that for a tree, the number of edges is one less than the number of vertices, i.e. $|\tau| = |V(\tau)| - 1$ and τ consists of one connected component so that $|C(\tau)| = 1$. Thus, by Lemma 4.1 we are done. Fact (iii) follows from Lemma 4.2. We will state and prove Lemmas 4.1 and 4.2 at the end of this section.

Next, we argue that the three facts combined yield (4.2). From fact (iii), we have for each α,

$$\begin{aligned} r_\alpha^2 &\leqslant (|\alpha| + 1)^{2|\alpha|} \left(\frac{\widehat{M}\lambda}{C}\right)^{2|\alpha|} \left(\frac{k}{n}\right)^{2|V(\alpha)|} \\ &= (|\alpha| + 1)^{2|\alpha|} \left(\frac{\widehat{M}^2\lambda^2 k^2}{C^2 n}\right)^{|E(\alpha)|} \left(\frac{k^2}{n}\right)^{|V(\alpha)| - |E(\alpha)|} n^{-|V(\alpha)|}. \end{aligned}$$

From (ii), we know that r_α is nonzero only when all connected components of α contain at least one cycle. Denote

$$\mathcal{G}_{d,v} = \{\alpha : |E(\alpha)| = d;\ |V(\alpha)| = v;\ \text{for all } \beta \in \mathcal{C}(\alpha),\ \beta \text{ is not a tree}\}.$$

Note that for all d, v such that $v > d$, we have that $\mathcal{G}_{d,v} = \varnothing$ because if all connected components of α contains at least one cycle, we must have $|\alpha| \geqslant |V(\alpha)|$. Thus for $k^2 \geqslant n$, we have shown that for all d, v with $\alpha \in \mathcal{G}_{d,v}$,

$$r_\alpha^2 \leqslant (d+1)^{2d} \left(\frac{\widehat{M}^2 \lambda^2 k^2}{C^2 n} \right)^d n^{-v}.$$

On the other hand, for $k^2 < n$, we have

$$r_\alpha^2 \leqslant (d+1)^{2d} \left(\frac{\widehat{M}\lambda}{C} \right)^{2d} \left(\frac{k}{n} \right)^{2v} \leqslant (d+1)^{2d} \left(\frac{\widehat{M}\lambda}{C} \right)^{2d} n^{-v}.$$

Combined with the bound on r_α^2 for $k^2 \geqslant n$, we have shown that

$$r_\alpha^2 \leqslant (d+1)^{2d} \left(\left(\frac{\widehat{M}\lambda}{C} \right)^2 \left(\frac{k^2}{n} \vee 1 \right) \right)^d n^{-v}.$$

Next, we bound the size of $\mathcal{G}_{d,v}$ by counting the number of graphs with exactly d edges and v vertices:

$$|\mathcal{G}_{d,v}| \leqslant \binom{n}{v} \binom{v}{2}^d \leqslant n^v v^{2d}. \tag{4.3}$$

where the factor $\binom{n}{v}$ enumerates the possibilities for the vertex set in α; the $\binom{v}{2}^d$ factor counts the allocation of the d edges, allowing for edge multiplicity. Combining (4.6) (see Lemma 4.2 below) and (4.3) yields

$$\begin{aligned}
\sum_{\alpha\,:\,|\alpha|\leqslant D} \frac{r_\alpha^2}{\alpha!} &\leqslant r_0^2 + \sum_{d=1}^{D}\sum_{v=1}^{d}\sum_{\alpha\in\mathcal{G}_{d,v}} \frac{r_\alpha^2}{\alpha!} \\
&\leqslant 1 + \sum_{d=1}^{D}\sum_{v=1}^{d} |\mathcal{G}_{d,v}| \cdot (d+1)^{2d} \left(\left(\frac{\widehat{M}\lambda}{C}\right)^2 \left(\frac{k^2}{n} \vee 1\right)\right)^d n^{-v} \\
&\leqslant 1 + \sum_{d=1}^{D}\sum_{v=1}^{d} n^v v^{2d} (d+1)^{2d} \left(\left(\frac{\widehat{M}\lambda}{C}\right)^2 \left(\frac{k^2}{n} \vee 1\right)\right)^d n^{-v} \\
&= 1 + \sum_{d=1}^{D} \left((d+1)^2 \left(\frac{\widehat{M}\lambda}{C}\right)^2 \left(\frac{k^2}{n} \vee 1\right)\right)^d \sum_{v=1}^{d} v^{2d} \\
&\leqslant 1 + D\sum_{d=1}^{D} \left((D+1)^2 D^2 \left(\frac{\widehat{M}\lambda}{C}\right)^2 \left(\frac{k^2}{n} \vee 1\right)\right)^d \\
&\leqslant 1 + D\sum_{d=1}^{D} \left(\left(2D^2 \frac{\widehat{M}\lambda}{C}\right)^2 \left(\frac{k^2}{n} \vee 1\right)\right)^d \\
&= 1 + (1+o(1))\, 4D^5 \left(\frac{\widehat{M}\lambda}{C}\right)^2 \left(\frac{k^2}{n} \vee 1\right) = 1 + o(1),
\end{aligned}$$

where the last two equalities follow by the condition $D^5(\widehat{M}\lambda)^2(k^2/n \vee 1) = o(1)$, under which the summation over d is a geometrically decreasing sequence, dominated by the first term.

Easy regime. Next, we show that in the "easy" regime $\lambda^2\widetilde{M}^2(k^2/n \vee 1) = \omega(1)$ and $\widetilde{M}^2 k = \omega(\widehat{M}^2)$, there is a low-degree test that strongly separates $\mathbb{P}$ and $\mathbb{Q}$. When $\lambda^2\widehat{M}^2k^2/n = \omega(1)$, consider the algorithm that uses $\widehat{T} = \sum_i Y_{ii}$, the sum of the diagonal elements, as the test statistic. We can compute the first and second moments of $\widehat{T}$ under the two models using (4.1) to note that under $\mathbb{P}$, we have $Y_{ii} = \frac{\lambda}{x_\ell} + \mathcal{N}(0,1)$ if $\sigma_i = \sigma_j = \ell$ for some $\ell \in [M]$ and $Y_{ii} = \mathcal{N}(0,1)$ otherwise, where each community label ℓ is selected with probability $\frac{x_\ell k}{n}$ and no label is selected with probability $1 - \frac{k}{n}$. Under $\mathbb{Q}$, we replace x and M with x' and M'.

$$\begin{aligned}
\mathbb{E}_{\mathbb{P}}\left[\widehat{T}\right] &= n\,\mathbb{E}_{\mathbb{P}}\left[Y_{11}\right] = n\left[\sum_{\ell\in[M]} \frac{\lambda}{x_\ell}\cdot \mathbb{P}\{\sigma_1 = \ell\} + 0 \cdot \mathbb{P}\{\sigma_1 = \star\}\right] = n\sum_{\ell\in[M]}\frac{k\lambda}{n} = Mk\lambda, \\
\mathbb{E}_{\mathbb{Q}}\left[\widehat{T}\right] &= n\,\mathbb{E}_{\mathbb{Q}}\left[Y_{11}\right] = M'k\lambda, \\
\operatorname*{Var}_{\mathbb{P}}\left[\widehat{T}\right] &\leqslant n\,\mathbb{E}_{\mathbb{P}}\left[Y_{11}^2\right] = n\left[\sum_{\ell\in[M]}\left(\frac{\lambda^2}{x_\ell^2}+1\right)\cdot \mathbb{P}\{\sigma_1=\ell\} + 1\cdot\mathbb{P}\{\sigma_1=\star\}\right] = n + \sum_{\ell\in[M]}\frac{k\lambda^2}{x_\ell}, \\
\operatorname*{Var}_{\mathbb{Q}}\left[\widehat{T}\right] &\leqslant n\,\mathbb{E}_{\mathbb{Q}}\left[Y_{11}^2\right] = n + \sum_{\ell\in[M']}\frac{k\lambda^2}{x'_\ell}.
\end{aligned}$$

Note also that $M \min_\ell x_\ell \geqslant C$ implies $\max_\ell 1/x_\ell < M/C$; thus, $\sum_{\ell \leqslant M} \frac{1}{x_\ell} \leqslant M^2/C$. Hence, when $\widetilde{M}^2 k/\widehat{M}^2 = \omega(1)$ and $\widetilde{M}^2 \lambda^2 k^2/n = \omega(1)$,

$$\sqrt{\max\left\{\operatorname*{Var}_{\mathbb{Q}}\left[\widehat{T}\right], \operatorname*{Var}_{\mathbb{P}}\left[\widehat{T}\right]\right\}} = o\left(\left|\mathbb{E}_{\mathbb{P}}\left[\widehat{T}\right] - \mathbb{E}_{\mathbb{Q}}\left[\widehat{T}\right]\right|\right).$$

Thus, thresholding $\widehat{T}$ strongly separates $\mathbb{P}$ and $\mathbb{Q}$. □

Proof of Theorem 2.5. Hard regime. The proof proceeds by comparison to a corresponding Gaussian model, so that we can reuse the calculations in the proof of Theorem 2.4. Our starting point is Proposition 2.9. Define $X = X^{(q,s)}$ appropriately for our binary testing problem, i.e., $X^{(q,s)}_{ij} = q + s/x_\ell$ if $\sigma_i = \sigma_j = \ell$, and $X^{(q,s)}_{ij} = q$ otherwise. Let $\tau_0 = q$, and recall that we have a valid constant $\tau_1 < 1$ by assumption. Consider the additive Gaussian testing problem (as in Theorem 2.4) with the same parameters n, k, M, x as our binary model, and with $\lambda := s/\sqrt{q(1-\tau_1)}$. Let $X^{(\lambda)}$ denote the corresponding X as per Proposition 2.8, i.e., $X^{(\lambda)}_{ij} = \lambda/x_\ell$ if $\sigma_i = \sigma_j = \ell$, and $X^{(\lambda)}_{ij} = 0$ otherwise. Note $X^{(q,s)}_{ij} = (s/\lambda)X^{(\lambda)}_{ij} + q$ and so by Lemmas 3.3 and 3.4 we have, $r_\alpha(X^{(q,s)}) = (s/\lambda)^{|\alpha|} r_\alpha(X^{(\lambda)})$. By Proposition 2.9,

$$\begin{aligned}\mathsf{Adv}_{\leqslant D}(\mathbb{P}, \mathbb{Q}) &\leqslant \sqrt{\sum_{\alpha \in \{0,1\}^N,\, |\alpha| \leqslant D} \frac{r_\alpha\left(X^{(q,s)}\right)^2}{(q(1-\tau_1))^{|\alpha|}}} = \sqrt{\sum_{\alpha \in \{0,1\}^N,\, |\alpha| \leqslant D} r_\alpha\left(X^{(\lambda)}\right)^2} \\ &\leqslant \sqrt{\sum_{\alpha \in \mathbb{N}^N,\, |\alpha| \leqslant D} \frac{r_\alpha\left(X^{(\lambda)}\right)^2}{\alpha!}}.\end{aligned}$$

In other words, we have related the conclusion of Proposition 2.9 to the conclusion of Proposition 2.8 but with $s/\sqrt{q(1-\tau_1)}$ in place of λ. The result now follows by the proof of Theorem 2.4.

Easy regime. We now consider a signed triangle count $\widehat{R}$ as our test statistic. Let

$$\widehat{R} = \sum_{i<j<k} R_{ij} R_{ik} R_{jk} \qquad \text{where } R_{ij} = Y_{ij} - q. \tag{4.4}$$

Expectation and variance calculations for $\widehat{R}$ are computed in Lemma 5.1 of Section 5.5. Denote by $\widehat{M}$ the maximum of M and M'. Then,

$$\left|\mathbb{E}_{\mathbb{P}}\left[\widehat{R}\right] - \mathbb{E}_{\mathbb{Q}}\left[\widehat{R}\right]\right| = \tfrac{1}{3}\left|M - M'\right| s^3 k^3 \left(1 + O(n^{-1})\right),$$

and

$$\begin{aligned}&\max\left\{\operatorname*{Var}_{\mathbb{P}}\left[\widehat{R}\right], \operatorname*{Var}_{\mathbb{Q}}\left[\widehat{R}\right]\right\} \\ &\leqslant \frac{1}{C}\widehat{M}^2 k^5 s^6 + \widehat{M} k^4 s^4 q + \frac{1}{C}\widehat{M}^2 k^4 s^5 + \tfrac{1}{3} n^3 q^3 + n k^2 s q^2 + k^3 q^2 s + k^3 q s^2 + \tfrac{1}{3}\widehat{M} k^3 s^3\end{aligned} \tag{4.5}$$

where C is the constant from the assumption that $M \min_\ell x_\ell, M' \min_\ell x'_\ell > C$. Writing $\widetilde{M} = |M - M'|$, notice that to prove strong separation it suffices to show that each term in (4.5) is $o(\widetilde{M}^2 s^6 k^6)$. For the fourth term to be $o(\widetilde{M}^2 s^6 k^6)$ is equivalent to $\widetilde{M}^{2/3} s^2 k^2/(nq) = \omega(1)$, one of our assumptions. Similarly, for the first term to be $o(\widetilde{M}^2 s^6 k^6)$ is equivalent to $\widetilde{M}^2 k \widehat{M}^2 = \omega(1)$ another of our assumptions. For the last term to be $o(\widetilde{M}^2 s^6 k^6)$ is equivalent to $\widetilde{M}^{2/3} \widehat{M}^{-1/3} sk = \omega(1)$, which is implied by our assumption $\widehat{M}^{-1/3} sk = \omega(1)$. All other terms follow also due to these assumptions. □

Lemma 4.1. *For each $\alpha \in \mathbb{N}^N$, and for each $x = (x_1, \ldots, x_c)$ with $\sum_\ell x_\ell = 1$,*

$$\mathbb{E}_{\mathbb{P}}\left[X^\alpha\right] = \lambda^{|\alpha|}\left(\frac{k}{n}\right)^{|V(\alpha)|} \prod_{\beta \in C(\alpha)} \sum_{\ell=1}^{c} x_\ell^{|V(\beta)|-|\beta|}.$$

Proof. First consider β a connected graph. Note that for $(i,j) \in \beta$, if it is not the case that $i, j \in S_\ell$ for some ℓ then $X^{(i,j)} \sim \mathcal{N}(0,1)$ and so $\mathbb{E}_{\mathbb{P}}\left[X^{(i,j)}\right] = 0$ (here, we have used that our S_ℓ's do not overlap). Hence, for β connected,

$$\mathbb{E}_{\mathbb{P}}\left[X^\beta\right] = \sum_{\ell=1}^{c} \mathbb{P}\left(V(\beta) \in S_\ell\right)\left(\frac{\lambda}{x_\ell}\right)^{|\beta|} = \sum_{\ell=1}^{c}\left(\frac{x_\ell k}{n}\right)^{|V(\beta)|}\left(\frac{\lambda}{x_\ell}\right)^{|\beta|}.$$

Notice, it is now enough to show that the X^β's are independent for β's connected components of α, as this would imply that $\mathbb{E}_{\mathbb{P}}\left[X^\alpha\right] = \prod_{\beta \in C(\alpha)} \mathbb{E}_{\mathbb{P}}\left[X^\beta\right]$ and we have the result.

This independence follows because $X^{(i,j)}$ depends only on the events $[i \in S_\ell]$, $[j \in S_{\ell'}]$ for each ℓ, ℓ'; thus, X^β and $X^{\beta'}$ are independent as long as their vertex sets $V(\beta)$ and $V(\beta')$ do not overlap. As the vertex sets of connected components are mutually non-overlapping, we have finished the proof. □

Lemma 4.2. *Suppose $Mx_{(1)} \geqslant C$ and $M'x'_{(1)} \geqslant C$ where $x_{(1)} := \min_\ell x_\ell$ for some constant $C > 0$. Then there exists $n_0 \in \mathbb{N}$ such that for $n > n_0$, for all α, we have that r_α satisfies*

$$|r_\alpha| \leqslant (|\alpha + 1|)^{|\alpha|}\left(\frac{\widehat{M}\lambda}{C}\right)^{|\alpha|}\left(\frac{k}{n}\right)^{|V(\alpha)|}. \tag{4.6}$$

Proof. We will argue by induction on $|\alpha|$. A graph α with $|\alpha| = 1$ is either a tree with two vertices and one edge, or a self-loop with one vertex and one edge. If α is a tree, then $r_\alpha = 0$ and (4.6) trivially holds. Recall $\widehat{M} = \max\{M, M'\}$. If α is a self-loop, we have by Lemma 4.1,

$$r_\alpha = \mathbb{E}_{\mathbb{P}}\left[X^\alpha\right] - \mathbb{E}_{\mathbb{Q}}\left[X^\alpha\right] = M\lambda\left(\frac{k}{n}\right) - M'\lambda\left(\frac{k}{n}\right) \leqslant (|\alpha| + 1)^{|\alpha|}\left(\frac{\widehat{M}\lambda}{C}\right)^{|\alpha|}\left(\frac{k}{n}\right)^{|V(\alpha)|},$$

where the last inequality is because $C \leqslant Mx_{(1)} \leqslant 1$. We have shown (4.6) for $|\alpha| = 1$. Suppose (4.6) holds for all α with $|\alpha| \leqslant d - 1$; next, we show it also holds for $|\alpha| = d$.

If α is not connected, then each connected component $\beta \in \mathcal{C}(\alpha)$ has $|\beta| < d$. Thus, from the factorization lemma and the induction hypothesis, we have

$$|r_\alpha| = \prod_{\beta \in \mathcal{C}(\alpha)} |r_\beta| \leqslant \prod_{\beta \in \mathcal{C}(\alpha)} (|\beta| + 1)^{|\beta|}\left(\frac{\lambda\widehat{M}}{C}\right)^{|\beta|}\left(\frac{k}{n}\right)^{|V(\beta)|} \leqslant (|\alpha| + 1)^{|\alpha|}\left(\frac{\lambda\widehat{M}}{C}\right)^{|\alpha|}\left(\frac{k}{n}\right)^{|V(\alpha)|}.$$

Thus (4.6) holds. Next, we show (4.6) for α connected. If α is a tree, then by Fact (ii) we have $r_\alpha = 0$ and (4.6) holds. Therefore, it suffices to consider α that is not a tree. Recall that

$$r_\alpha = \mathbb{E}_{\mathbb{P}}\left[X^\alpha\right] - \mathbb{E}_{\mathbb{Q}}\left[X^\alpha\right] - \sum_{0 < \beta \lneq \alpha} r_\beta \binom{\alpha}{\beta} \mathbb{E}_{\mathbb{Q}}\left[X^{\alpha \setminus \beta}\right]. \tag{4.7}$$

For the first term in (4.7), we can apply Lemma 4.1 for connected α:

$$\begin{aligned}
\mathbb{E}_{\mathbb{P}}\left[X^{\alpha}\right] &= \lambda^{|\alpha|}\left(\frac{k}{n}\right)^{|V(\alpha)|}\sum_{\ell\in[M]} x_{\ell}^{|V(\alpha)|-|\alpha|} \\
&\leqslant \lambda^{|\alpha|}\left(\frac{k}{n}\right)^{|V(\alpha)|} M x_{(1)}^{|V(\alpha)|-|\alpha|} \\
&= \lambda^{|\alpha|}\left(\frac{k}{n}\right)^{|V(\alpha)|}\left(M x_{(1)}\right)^{|V(\alpha)|-|\alpha|} M^{1-|V(\alpha)|+|\alpha|} \\
&\leqslant \left(\frac{\lambda M}{C}\right)^{|\alpha|}\left(\frac{k}{n}\right)^{|V(\alpha)|} M^{1-|V(\alpha)|} \\
&\leqslant \left(\frac{\lambda M}{C}\right)^{|\alpha|}\left(\frac{k}{n}\right)^{|V(\alpha)|}
\end{aligned}$$

for large enough n. The first inequality is because assuming α is not a tree, $|V(\alpha)| \leqslant |\alpha|$ and $x_\ell \leqslant 1$; the second inequality is because $1 \geqslant M x_{(1)} \geqslant C$, thus $(M x_{(1)})^{|V(\alpha)|-|\alpha|} \leqslant C^{|V(\alpha)|-|\alpha|} \leqslant C^{-|\alpha|}$; the last inequality holds since $|V(\alpha)| \geqslant 1$.

Next, we bound the third term in (4.7). For each $\beta \subsetneq \alpha$ that is nonempty, $|\beta| < |\alpha| = d$. From the induction hypothesis, we have

$$|r_\beta| \leqslant (|\beta|+1)^{|\beta|}\left(\frac{\widehat{M}\lambda}{C}\right)^{|\beta|}\left(\frac{k}{n}\right)^{|V(\beta)|}.$$

Thus,

$$\begin{aligned}
&\left|r_\beta \mathbb{E}_{\mathbb{Q}} X^{\alpha\backslash\beta}\right| \leqslant (|\beta|+1)^{|\beta|}\left(\frac{\widehat{M}\lambda}{C}\right)^{|\beta|}\left(\frac{k}{n}\right)^{|V(\beta)|}\cdot\lambda^{|\alpha\backslash\beta|}\left(\frac{k}{n}\right)^{|V(\alpha\backslash\beta)|}\prod_{\gamma\in\mathcal{C}(\alpha\backslash\beta)}\sum_{\ell\in[M']}(x'_\ell)^{|V(\gamma)|-|\gamma|} \\
&= (|\beta|+1)^{|\beta|}\left(\frac{\widehat{M}\lambda}{C}\right)^{|\alpha|}\left(\frac{k}{n}\right)^{|V(\alpha)|}\left(\frac{\widehat{M}}{C}\right)^{-|\alpha\backslash\beta|}\left(\frac{k}{n}\right)^{-|V(\beta)\cap V(\alpha\backslash\beta)|}\prod_{\gamma\in\mathcal{C}(\alpha\backslash\beta)}\sum_{\ell\in[M']}(x'_\ell)^{|V(\gamma)|-|\gamma|} \\
&\leqslant (|\beta|+1)^{|\beta|}\left(\frac{\widehat{M}\lambda}{C}\right)^{|\alpha|}\left(\frac{k}{n}\right)^{|V(\alpha)|}\prod_{\gamma\in\mathcal{C}(\alpha\backslash\beta)}\left(\frac{\widehat{M}}{C}\right)^{-|\gamma|}\sum_{\ell\in[M']}(x'_\ell)^{|V(\gamma)|-|\gamma|}.
\end{aligned}$$

Next, we show that for all $\gamma \in \mathcal{C}(\alpha\backslash\beta)$, we have that $(\widehat{M}/C)^{-|\gamma|}\sum_{\ell\in[M']}(x'_\ell)^{|V(\gamma)|-|\gamma|} \leqslant 1$. Note that $|V(\gamma)| \leqslant |\gamma|+1$. We discuss the cases $|V(\gamma)| = |\gamma|+1$ and $|V(\gamma)| \leqslant |\gamma|$ separately. If $|V(\gamma)| = |\gamma|+1$, then

$$\left(\frac{\widehat{M}}{C}\right)^{-|\gamma|}\sum_{\ell\in[M']}(x'_\ell)^{|V(\gamma)|-|\gamma|} = \left(\frac{\widehat{M}}{C}\right)^{-|\gamma|} \leqslant 1.$$

If $|V(\gamma)| \leqslant |\gamma|$, then we have

$$\begin{aligned}
\left(\frac{\widehat{M}}{C}\right)^{-|\gamma|}\sum_{\ell\in[M']}(x'_\ell)^{|V(\gamma)|-|\gamma|} \leqslant \left(\frac{M}{C}\right)^{-|\gamma|} M'\,(x'_{(1)})^{|V(\gamma)|-|\gamma|} &\overset{(a)}{\leqslant} \left(\widehat{M}x'_{(1)}\right)^{|V(\gamma)|-|\gamma|}\widehat{M}^{1-|V(\gamma)|}C^{|\gamma|} \\
&\overset{(b)}{\leqslant} C^{|V(\gamma)|}\,\widehat{M}^{1-|V(|\gamma|)|} \overset{(c)}{\leqslant} 1,
\end{aligned}$$

where (a) is from $M' \leqslant M$; (b) is from $M'x'_{(1)} \geqslant C$ and $|V(\gamma)| \leqslant |\gamma|$; (c) is from $C \leqslant 1$, $\widehat{M} \geqslant 1$, and $|V(\gamma)| \geqslant 1$ for all $\gamma \in \mathcal{C}(\alpha\backslash\beta)$. We have shown that

$$\left| r_\beta \mathbb{E}_{\mathbb{Q}} X^{\alpha\backslash\beta} \right| \leqslant (|\beta|+1)^{|\beta|} \left(\frac{\widehat{M}\lambda}{C} \right)^{|\alpha|} \left(\frac{k}{n} \right)^{|V(\alpha)|}.$$

Plug in the values of $\mathbb{E}_{\mathbb{P}}[X^\alpha]$ and $\mathbb{E}_{\mathbb{Q}}[X^\alpha]$ to (4.7) to obtain

$$\begin{aligned} |r_\alpha| &\leqslant \left(\frac{\widehat{M}\lambda}{C} \right)^{|\alpha|} \left(\frac{k}{n} \right)^{|V(\alpha)|} + \lambda^{|\alpha|} \left(\frac{k}{n} \right)^{|V(\alpha)|} + \sum_{0<\beta\lneq\alpha} \binom{\alpha}{\beta} (|\beta|+1)^{|\beta|} \left(\frac{\widehat{M}\lambda}{C} \right)^{|\alpha|} \left(\frac{k}{n} \right)^{|V(\alpha)|} \\ &\leqslant \left[1 + 1 + \sum_{0<\beta\lneq\alpha} (|\beta|+1)^{|\beta|} \right] \left(\frac{\widehat{M}\lambda}{C} \right)^{|\alpha|} \left(\frac{k}{n} \right)^{|V(\alpha)|} \\ &\leqslant (|\alpha|+1)^{|\alpha|} \left(\frac{\widehat{M}\lambda}{C} \right)^{|\alpha|} \left(\frac{k}{n} \right)^{|V(\alpha)|}, \end{aligned}$$

where the last inequality is because

$$2 + \sum_{0<\beta\lneq\alpha} \binom{\alpha}{\beta} (|\beta|+1)^{|\beta|} = 2 + \sum_{0<\ell<|\alpha|} \binom{|\alpha|}{\ell} (\ell+1)^\ell \leqslant (|\alpha|+1)^{|\alpha|}.$$

We have shown that (4.6) holds for all α. □

5 Additional proofs

5.1 Proof of Lemma 2.7

First we prove the statement for weak separation. Assume, for the sake of contradiction, that some degree-D test $g : \mathbb{R}^N \to \mathbb{R}$ weakly separates $\mathbb{P}$ and $\mathbb{Q}$. Without loss of generality, we can shift and scale g so that $\mathbb{E}_{\mathbb{Q}}[g] = 0$ and $\mathbb{E}_{\mathbb{P}}[g] = 1$. Weak separation guarantees that for sufficiently large n, $\mathrm{Var}_{\mathbb{Q}}[g] = \mathbb{E}_{\mathbb{Q}}[g^2] \leqslant C$ for some positive constant $C > 0$. Defining $f = g + C$, we have

$$\mathsf{Adv}_{\leqslant D} \geqslant \frac{\mathbb{E}_{\mathbb{P}}[f]}{\sqrt{\mathbb{E}_{\mathbb{Q}}[f^2]}} = \frac{1+C}{\sqrt{\mathbb{E}_{\mathbb{Q}}[g^2] + C^2}} \geqslant \frac{1+C}{\sqrt{C+C^2}} = \sqrt{\frac{1+C}{C}},$$

which is a constant strictly greater than 1, contradicting $\mathsf{Adv}_{\leqslant D} = 1 + o(1)$. The proof for strong separation is identical, except now $C = o(1)$.

5.2 Proof of Proposition 2.8

The proof is similar to the proof of Theorem 2.2 in [SW22], so we only explain the differences. Our distribution $\mathbb{Q}$ plays the role of the single "planted" distribution in [SW22]. The only difference is that the quantity $\mathbb{E}[f(Y)x]$ from [SW22] needs to be replaced by our $\mathbb{E}_{\mathbb{P}}[f(Y)]$, which means (in the notation of [SW22]) the vector c needs to be redefined as $c_\alpha = \mathbb{E}_{\mathbb{P}}[h_\alpha(Y)] = \mathbb{E}_{\mathbb{P}}[X^\alpha]/\sqrt{\alpha!}$.

5.3 Proof of Proposition 2.9

Follow the proof of Theorem 2.7 in [SW22], but redefine $c = (c_\alpha)_{\alpha\in\{0,1\}^N}$ by $c_\alpha = \mathbb{E}_{\mathbb{P}}[\widetilde{X}^\alpha]$ where $\widetilde{X}_i = (\mu + 1/\mu)X_i - 1/\mu$ and $\mu = \sqrt{\frac{1-\tau_1}{\tau_0}}$. This gives the bound

$$\mathsf{Adv}_{\leqslant D}(\mathbb{P},\mathbb{Q}) \leqslant \sqrt{\sum_{\alpha\in\{0,1\}^N,\,|\alpha|\leqslant D} \frac{r_\alpha(\widetilde{X})^2}{(1+\tau_0-\tau_1)^{2|\alpha|}}}$$

where $r_\alpha(\widetilde{X})$ is defined in (2.5). Using Lemmas 3.3 and 3.4, we have $r_\alpha(\widetilde{X}) = (\mu + 1/\mu)^{|\alpha|} r_\alpha(X)$, so the above simplifies to give the result.

5.4 Proof of Lemma 3.3

Base case(s). Note that by definition $\widetilde{r}_\varnothing = r_\varnothing = 1$. Let $|\alpha| = 1$, i.e. $\alpha = \{ij\}$ for some $1 \leqslant i \leqslant j \leqslant n$. Then the base step follows directly from the definition

$$\widetilde{r}_\alpha = \mathbb{E}_P\left[\widetilde{X}^{ij}\right] - \mathbb{E}_Q\left[\widetilde{X}^{ij}\right] = \mathbb{E}_P\left[X^{ij}\right] + y_{ij} - \mathbb{E}_Q\left[X^{ij}\right] - y_{ij} = r_\alpha.$$

Inductive step. Fix α with $|\alpha| > 1$ and assume $\widetilde{r}_\beta = r_\beta$ for all $\beta \subsetneq \alpha$. Directly from the definition of r and the inductive hypothesis,

$$\begin{aligned}\widetilde{r}_\alpha &= \mathbb{E}_P\left[\widetilde{X}^\alpha\right] - \mathbb{E}_Q\left[\widetilde{X}^\alpha\right] - \sum_{\varnothing\subsetneq\beta\subsetneq\alpha} \widetilde{r}_{\alpha\setminus\beta}\, \mathbb{E}_Q\left[\widetilde{X}^\beta\right] \\ &= \mathbb{E}_P\left[\widetilde{X}^\alpha\right] - \mathbb{E}_Q\left[\widetilde{X}^\alpha\right] - \sum_{\varnothing\subsetneq\beta\subsetneq\alpha} r_{\alpha\setminus\beta}\, \mathbb{E}_Q\left[\widetilde{X}^\beta\right].\end{aligned} \tag{5.1}$$

We consider the third term, call it $\ast$. Writing y^η to indicate $\prod_{ij\in\eta} y_{ij}$, first notice that

$$\mathbb{E}_Q\left[\widetilde{X}^\beta\right] = \mathbb{E}_Q\left[\prod_{ij\in\beta}(X_{ij}+y_{ij})\right] = \mathbb{E}_Q\left[X^\beta\right] + \sum_{\varnothing\subsetneq\eta\subseteq\beta} y^\eta\, \mathbb{E}_Q\left[X^{\beta\setminus\eta}\right]; \tag{5.2}$$

hence,

$$\ast = \sum_{\varnothing\subsetneq\beta\subsetneq\alpha} r_{\alpha\setminus\beta}\, \mathbb{E}_Q\left[X^\beta\right] + \sum_{\varnothing\subsetneq\beta\subsetneq\alpha} r_{\alpha\setminus\beta} \sum_{\varnothing\subsetneq\eta\subseteq\beta} y^\eta\, \mathbb{E}_Q\left[X^{\beta\setminus\eta}\right].$$

If we let $\beta' = \beta\setminus\eta$, instead of summing over $\varnothing \subsetneq \beta \subsetneq \alpha$ and then $\varnothing \subsetneq \eta \subseteq \beta$, we may sum over $\varnothing \subsetneq \eta \subsetneq \alpha$ then $\varnothing \subseteq \beta' \subsetneq \alpha\setminus\eta$. Thus, noting also that $\alpha\setminus\beta = (\alpha\setminus\eta)\setminus\beta'$,

$$\ast = \sum_{\varnothing\subsetneq\beta\subsetneq\alpha} r_{\alpha\setminus\beta}\, \mathbb{E}_Q\left[X^\beta\right] + \sum_{\varnothing\subsetneq\eta\subsetneq\alpha} y^\eta \sum_{\varnothing\subseteq\beta'\subsetneq\alpha\setminus\eta} r_{(\alpha\setminus\eta)\setminus\beta'}\, \mathbb{E}_Q\left[X^{\beta'}\right].$$

But, by the definition of $r_{\alpha\setminus\eta}$,

$$\begin{aligned}\sum_{\varnothing\subseteq\beta'\subsetneq\alpha\setminus\eta} r_{(\alpha\setminus\eta)\setminus\beta'}\, \mathbb{E}_Q\left[X^{\beta'}\right] &= r_{\alpha\setminus\eta} + \sum_{\varnothing\subsetneq\beta'\subsetneq\alpha\setminus\eta} r_{(\alpha\setminus\eta)\setminus\beta'}\, \mathbb{E}_Q\left[X^{\beta'}\right] \\ &= \mathbb{E}_P\left[X^{\alpha\setminus\eta}\right] - \mathbb{E}_Q\left[X^{\alpha\setminus\eta}\right],\end{aligned}$$

which gives the following expression for $*$ where we no longer have the sum over β':

$$* = \sum_{\varnothing \subsetneq \beta \subsetneq \alpha} r_{\alpha \setminus \beta} \, \mathbb{E}_Q \left[X^{\beta} \right] + \sum_{\varnothing \subsetneq \eta \subsetneq \alpha} y^{\eta} \left(\mathbb{E}_P \left[X^{\alpha \setminus \eta} \right] - \mathbb{E}_Q \left[X^{\alpha \setminus \eta} \right] \right).$$

Substituting this expression for $*$ into our original expression for $\widetilde{r}_\alpha$ in (5.1), we have

$$\widetilde{r}_\alpha = \mathbb{E}_P \left[\widetilde{X}^{\alpha} \right] - \mathbb{E}_Q \left[\widetilde{X}^{\alpha} \right] - \sum_{\varnothing \subsetneq \beta \subsetneq \alpha} r_{\alpha \setminus \beta} \, \mathbb{E}_Q \left[X^{\beta} \right] - \sum_{\varnothing \subsetneq \eta \subsetneq \alpha} y^{\eta} \left(\mathbb{E}_P \left[X^{\alpha \setminus \eta} \right] - \mathbb{E}_Q \left[X^{\alpha \setminus \eta} \right] \right).$$

However, using (5.2) (and a similar result on $\mathbb{E}_P \left[\widetilde{X}^{\alpha} \right]$), we see that this last term is precisely what we need to cancel with the difference between $\mathbb{E}_P \left[\widetilde{X}^{\alpha} \right]$ and $\mathbb{E}_P \left[X^{\alpha} \right]$ and the difference between $\mathbb{E}_Q \left[\widetilde{X}^{\alpha} \right]$ and $\mathbb{E}_Q \left[X^{\alpha} \right]$, as

$$\sum_{\varnothing \subsetneq \eta \subsetneq \alpha} y^{\eta} \left(\mathbb{E}_P \left[X^{\alpha \setminus \eta} \right] - \mathbb{E}_Q \left[X^{\alpha \setminus \eta} \right] \right) = \sum_{\varnothing \subsetneq \eta \subseteq \alpha} y^{\eta} \left(\mathbb{E}_P \left[X^{\alpha \setminus \eta} \right] - \mathbb{E}_Q \left[X^{\alpha \setminus \eta} \right] \right).$$

Therefore,

$$\widetilde{r}_\alpha = \mathbb{E}_P \left[X^{\alpha} \right] - \mathbb{E}_Q \left[X^{\alpha} \right] - \sum_{\varnothing \subsetneq \beta \subsetneq \alpha} r_{\alpha \setminus \beta} \, \mathbb{E}_Q \left[X^{\beta} \right] = r_\alpha,$$

and we have proven the inductive step.

5.5 Calculations for signed triangle counts

In this section we analyse the degree 3 signed triangle count test statistic $\widehat{R}$, defined in (4.4), and show bounds on the expectation and variance of $\widehat{R}$, which will prove it strongly separates $\mathbb{P}$ and $\mathbb{Q}$ in the easy regime. Recall,

$$\widehat{R} = \sum_{i<j<k} R_{ij} \, R_{ik} \, R_{jk} \qquad \text{where } R_{ij} = Y_{ij} - q.$$

Lemma 5.1. *We let $\mathbb{P} = \mathbb{P}_{\text{Binary}}(n, k, q, s, M, x)$, given parameters n, k, q, s, M and $x \in \mathbb{R}^M$ with $\sum_{\ell \in [M]} x_\ell = 1$. Assume that $M \min_\ell x_\ell \geqslant C$. Then,*

$$\mathbb{E}_{\mathbb{P}} \left[\widehat{R} \right] = \frac{1}{3} M s^3 k^3 \left(1 + O(n^{-1}) \right),$$
$$\operatorname*{Var}_{\mathbb{P}} \left[\widehat{R} \right] \leqslant \frac{1}{C} M^2 k^5 s^6 + M k^4 s^4 q + \frac{1}{C} M^2 k^4 s^5 + \frac{1}{3} n^3 q^3 + n k^2 s q^2 + k^3 q^2 s + k^3 q s^2 + \frac{1}{3} M k^3 s^3.$$

Proof. Recall that in our model, the binary random variable Y_{ij} takes value 1 with probability $q + s/x_c$ if $\sigma_i = \sigma_j = c$ for some $c \in [M]$ and takes value 1 with probability q otherwise. Thus, we may calculate the expected values of R_{ij} conditioned on the community assignments of i and j:

$$\mathbb{E}_{\mathbb{P}} \left[R_{ij} \mid \sigma_i = c_i, \, \sigma_j = c_j \right] \quad = \quad \begin{cases} \frac{s}{x_c} & \text{if } c_i = c_j = c \text{ for some } c \in [M], \\ 0 & \text{otherwise.} \end{cases} \tag{5.3}$$

We now split the proof into expectation and variance calculations. All probabilities, expectations and variances will be with respect to $\mathbb{P}$, but we drop the subscript.

Expectation. Let

$$N^{\text{tri}} = \{\{ij,\, ik,\, jk\} \,:\, i,j,k \in [n],\; i < j < k\}\,,$$

and then we may express the signed triangle count $\widehat{R}$ by $\widehat{R} = \sum_{S \in N^{\text{tri}}} R_S$. Fix a set of edges in N^{tri}, w.l.o.g. $S = \{12, 13, 23\}$. Then, writing $[M]_\star$ for the set $\{\star, 1, \ldots, M\}$ (recall $\star$ denotes no community membership),

$$\begin{aligned}\mathbb{E}\left[R_S\right] &= \sum_{c_1,c_2,c_3 \in [M]_\star} \mathbb{E}\left[R_{12}\, R_{13}\, R_{23}\; |\sigma_1 = c_1,\, \ldots,\, \sigma_3 = c_3\right] P\left(\sigma_1 = c_1,\, \ldots,\, \sigma_3 = c_3\right) \\ &= \sum_{c_1,c_2,c_3 \in [M]_\star} \prod_{ij \in \{12,13,23\}} \mathbb{E}\left[R_{ij} \mid \sigma_1 = c_1,\, \ldots,\, \sigma_3 = c_3\right] \prod_{i=1}^{3} P\left(\sigma_i = c_i\right),\end{aligned}$$

as the expected values of R_{ij} and R_{ik} are independent conditional on the community assignments of i, j, k. Note by (5.3), $\mathbb{E}[R_{ij}|\sigma_i = c_i, \sigma_j = c_j]$ is equal to zero unless $c_i = c_j = c$ for some $c \in [M]$. Therefore the only non-zero terms in the sum above are those for which $c_1 = c_2 = c_3 = c$ for some $c \in [M]$. Let C_c be the event that $\sigma_1 = \sigma_2 = \sigma_3 = c$, then

$$\mathbb{E}\left[R_S\right] = \sum_{c=1}^{M} \prod_{ij \in S} \mathbb{E}\left[R_{ij} \mid C_c\right] \prod_{i=1}^{3} P\left(\sigma_i = c\right) = \sum_{c=1}^{M} \frac{s^3}{x_c^3} \left(\frac{k x_c}{n}\right)^3 = M k^3 s^3 n^{-3}.$$

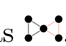
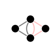
Because $|N^{\text{tri}}| = \binom{n}{3} = \frac{1}{3} n^3 (1 + O(\frac{1}{n}))$, the expectation of $\widehat{R}$ is as claimed.

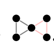
Variance. Recall $\widehat{R} = \sum_{S \in N^{\text{tri}}} R_S$ and so the variance is

$$\operatorname{Var}\left[\widehat{R}\right] = \sum_{S,T \in N^{\text{tri}}} \mathbb{E}\left[R_S R_T\right] - \mathbb{E}\left[R_S\right] \mathbb{E}[R_T]\,.$$

Note that if $V(S) \cap V(T) = \varnothing$, i.e. the sets of pairs have no vertices in common, then R_S and R_T are independent and these terms cancel in the expression above. Hence we need only sum over S, T with one overlapping vertex, with two overlapping vertices or equivalently one overlapping edge and lastly with all three vertices overlapping or equivalently $S = T$. Thus

$$\operatorname{Var}\left[\widehat{R}\right] \leqslant \sum_{\substack{S,T \in N^{\text{tri}} \\ |V(S) \cap V(T)| = 1}} \mathbb{E}\left[R_S R_T\right] + \sum_{\substack{S,T \in N^{\text{tri}} \\ |S \cap T| = 1}} \mathbb{E}\left[R_S R_T\right] + \sum_{S \in N^{\text{tri}}} \mathbb{E}\left[R_S^2\right]. \tag{5.4}$$

The terms above correspond to the sets of pairs overlapping as ,  and  respectively where the gray edges denote pairs in S and the pink edges denote pairs in T.

We begin by bounding the first term in (5.4), i.e. that corresponding to . Fix some pair of sets which overlap on one vertex, w.l.o.g. $S_1 = \{12, 13, 23\}$ and $T_1 = \{14, 15, 45\}$. Then, similarly to the expectation, again writing $[M]_\star$ for the set $\{\star, 1, \ldots, M\}$,

$$\mathbb{E}[R_{S_1} R_{T_1}] \quad = \quad \sum_{c_1,\ldots,c_5 \in [M]_\star} \prod_{ij \in S_1 \cup T_1} \mathbb{E}[R_{ij} \mid \sigma_1 = c_1, \ldots \sigma_5 = c_5] \prod_{i=1}^{5} P(\sigma_i = c_i)$$

as the expected values of R_{ij} and R_{ik} are independent conditional on the community assignments of i, j, k. Note that $\mathbb{E}[R_{ij}|\sigma_i = c_i, \sigma_j = c_j]$ is equal to zero unless $c_i = c_j = c$ for some $c \in [M]$.

Therefore the only non-zero terms in the sum above are those for which $c_1 = \ldots = c_5 = c$ for some $c \in [M]$. Let C_c be the event that $\sigma_1 = \ldots = \sigma_5 = c$, then

$$\mathbb{E}[R_{S_1} R_{T_1}] = \sum_{c=1}^{M} \prod_{ij \in S_1 \cup T_1} \mathbb{E}[R_{ij} \mid C_c] \prod_{i=1}^{5} P(\sigma_i = c) = \sum_{c=1}^{M} \frac{s^6}{x_c^6} \left(\frac{k x_c}{n} \right)^5 = k^5 s^6 n^{-5} \sum_{c=1}^{M} \frac{1}{x_c}.$$

Since there are at most n^5 ways we may pick $S, T \in N^{\mathrm{tri}}$ with $|V(S) \cap V(T)| = 1$, we may conclude that the first term of (5.4) is at most $k^5 s^6 \sum_{c=1}^{M} \frac{1}{x_c}$.

We next bound the second term in (5.4), i.e. that corresponding to . Similarly to above, fix some pair of sets which overlap on one edge, w.l.o.g. $S_2 = \{12, 13, 23\}$ and $T_2 = \{12, 14, 24\}$.

$$\begin{aligned}
\mathbb{E}[R_{S_2} R_{T_2}] &= \sum_{\underline{c} \in [M]_\star^4} \mathbb{E}[R_{12}^2 R_{13} R_{23} R_{14} R_{24} \mid \sigma_1 = c_1, \ldots, \sigma_4 = c_4] \prod_{i=1}^{4} P(\sigma_i = c_i) \\
&= \sum_{\underline{c} \in [M]_\star^4} \mathbb{E}[R_{12}^2 \mid \underline{\sigma} = \underline{c}] \prod_{ij \in \{13,23,14,24\}} \mathbb{E}[R_{ij} \mid \underline{\sigma} = \underline{c}] \prod_{i=1}^{4} P(\sigma_i = c_i) \qquad (5.5)
\end{aligned}$$

since, as before, R_{ij} and R_{ik} are independent when we have conditioned on the community assignments of i, j, k. Again, recall $\mathbb{E}[R_{ij} | \sigma_i = c_i, \sigma_j = c_j]$ is equal to zero unless $c_i = c_j = c$ for some $c \in [M]$. Thus for the product over $ij \in \{13, 23, 14, 24\}$ in (5.5) to be non-zero all vertices must have the same community assignment to some $c \in [M]$. Hence,

$$\mathbb{E}[Y_{S_2} Y_{T_2}] = \sum_{c=1}^{M} \mathbb{E}[R_{12}^2 \mid C_c] \prod_{ij \in \{13,23,14,24\}} \mathbb{E}[R_{ij} \mid C_c] \prod_{i=1}^{4} P(\sigma_i = c).$$

Calculate the conditional expectation of the square.

$$\mathbb{E}[R_{ij}^2 \mid \sigma_i = c_i, \sigma_j = c_j] = \begin{cases} q(1-q) + \frac{s}{x_c}(1-2q) & \text{if } c_i = c_j = c \text{ for some } c \in [M] \\ q(1-q) & \text{otherwise,} \end{cases} \qquad (5.6)$$

and thus,

$$\begin{aligned}
\mathbb{E}[R_{S_2} R_{T_2}] &= \sum_{c=1}^{M} \left[q(1-q) + \frac{s}{x_c}(1-2q) \right] \left(\frac{s}{x_c} \right)^4 \left(\frac{k x_c}{n} \right)^4 \\
&= M \left(\frac{ks}{n} \right)^4 q(1-q) + \left(\frac{ks}{n} \right)^4 s(1-2q) \sum_{c=1}^{M} \frac{1}{x_c}.
\end{aligned}$$

Since there are at most n^4 ways we may pick $S, T \in N^{\mathrm{tri}}$ with $|S \cap T| = 1$, we may conclude that the second term of (5.4) is at most $M k^4 s^4 q + k^4 s^5 \sum_{c=1}^{M} \frac{1}{x_c}$.

Lastly we bound the third (and last) term in (5.4), i.e. that corresponding to . Similarly to above, fix a set S (and T which entirely overlaps with it), w.l.o.g. $S_3 = \{12, 13, 23\}$. Calculate

$$\begin{aligned}
\mathbb{E}[R_{S_3}^2] &= P(D_0)\,(q(1-q))^3 + \sum_{i=1}^{3} \sum_{c=1}^{M} P[D_{i,c}] \left(q(1-q) + \frac{s}{x_c}(1-2q) \right)^i (q(1-q))^{3-i} \\
&\leqslant P(D_0) q^3 + \sum_{i=1}^{3} \sum_{c=1}^{M} P[D_{i,c}] \left(q + \frac{s}{x_c} \right)^i q^{3-i}
\end{aligned}$$

where $D_{i,c}$ denotes the set of community assignments such that i of R_{12}, R_{13}, R_{23} has distribution $\mathrm{Ber}(q + s/x_c)$ (while the others have distribution $\mathrm{Ber}(q)$), and D_0 denotes the set of community assignments where all three have distribution $\mathrm{Ber}(q)$. Observe $D_{1,c}$ is the set of assignments such that two vertices have label $c \in [M]$ and the other vertex has label in $\{\star, 1, \ldots, M\}\backslash\{c\}$ and thus $P(D_{1,c}) \leqslant 3(x_c k/n)^2$. Note $D_{2,c} = \varnothing$. Lastly $P(D_{3,c}) = \sum_{c\in M}(x_c k/n)^3$ as $D_{3,c}$ is the community assignment where each of the three vertices has label c. Then $P(D_0) = 1 - \sum_c P(D_{1,c}) - \sum_c P(D_{3,c})$. Substituting these bounds for D_0 and $D_{i,c}$ for $i = 1, 2, 3$ and writing $\rho_c = kx_c/n$ we get

$$\begin{aligned}\mathbb{E}[R_{S_3}^2] &\leqslant q^3(1 - 3\rho_c^2 - \rho_c^3) + 3\sum_{c=1}^{M} \rho_c^2 \left(q + \frac{s}{x_c}\right) q^2 + \sum_{c=1}^{M} \rho_c^3 \left(q + \frac{s}{x_c}\right)^3 \\ &= q^3 + 3n^{-2}k^2 sq^2 + n^{-3}k^3 \left(3q^2 s \sum_{c=1}^{M} x_c^2 + 3qs^2 + Ms^3\right).\end{aligned}$$

Since there are $\binom{n}{3}$ ways to pick $S \in N^{\mathrm{tri}}$ the third term of (5.4) is at most $\frac{1}{3}n^3 E[R_{S_3}^2]$,

$$\tfrac{1}{3}n^3 E[R_{S_3}^2] \leqslant \tfrac{1}{3}n^3 q^3 + nk^2 sq^2 + k^3 q^2 s + k^3 qs^2 + \tfrac{1}{3}Mk^3 s^3$$

where we substituted $\sum_{c\in M} x_c^2, \sum_{c\in M} x_c^3 \leqslant 1$. To finish, recall we assumed $M \min_c x_c > C$ for some constant C, and note this implies $\max_c 1/x_c < M/C$ and thus $\sum_c 1/x_c^2 < M^2/C$. Apply this to the bounds from the first and second terms of (5.4) and we are done. □

6 Relation to recovery: proofs and further discussion.

In this section we prove Proposition 2.6 showing the relation between the recovery problem studied in [SW22] and testing between two planted distributions. Loosely, Part 1 of the proposition states that for any recovery problem we may construct an equivalent testing problem, and Part 2 states that for testing problems where the likelihood ratio of the signals exists there is a corresponding recovery problem.

In [SW22] the authors show an upper bound on $\mathsf{Corr}(g(X), \mathbb{Q})$ in terms of cumulants κ_α where $\mathbb{Q}$ is additive Gaussian or Binomial. We show in Remark 6.1 below that our Propositions 2.8 and 2.9 which bound Adv in terms of cumulant-like quantities r_α recover these cumulant upper bounds proven in [SW22].

Proof of Prop 2.6, Part 1. We construct $\mathbb{P}_X$ (the planted part of the distribution $\mathbb{P}$) by size biasing $\mathbb{Q}_X$ by $g(X)$, then setting $Y|X$ in $\mathbb{P}$ to be the same as $Y|X$ in $\mathbb{Q}$. Suppose first that $\mathbb{Q}$ has density function $q(x, y)$. Let

$$p(x) = g(x)q(x)/\mathbb{E}_{\mathbb{Q}}[g(X)]$$

and let $p(y|x) = q(y|x)$. Note this does define a density function since

$$\int p(x, y)dxdy = \int p(y|x)p(x)dxdy = \frac{1}{\mathbb{E}_{\mathbb{Q}}[g(X)]}\int g(x)q(y|x)q(x)dxdy = 1.$$

Comparing the definitions of Adv in (2.3) and Corr' in (2.2) it suffices to show that $\mathbb{E}_{\mathbb{P}}[h(Y)] = \mathbb{E}_{\mathbb{Q}}[g(X)h(Y)]/E_{\mathbb{Q}}[g(X)]$ for any polynomial h. To see this holds note

$$\mathbb{E}_{\mathbb{P}}[h(Y)] = \int h(y)p(y|x)p(x)dxdy = \frac{1}{\mathbb{E}_{\mathbb{Q}}[g(X)]}\int g(x)h(y)q(y|x)q(x)dxdy = \frac{1}{\mathbb{E}_{\mathbb{Q}}[g(X)]}\mathbb{E}_{\mathbb{Q}}[g(X)\,h(Y)],$$

as required.

If we suppose that $\mathbb{Q}$ is a discrete distibution the proof is similar. Set

$$\mathbb{P}[X = x] = g(x)\mathbb{Q}[X = x]/\mathbb{E}_{\mathbb{Q}}[g(X)]$$

and let $\mathbb{P}[Y = y|X = x] = \mathbb{Q}[Y = y|X = x]$. Then, similarly one may check $\mathbb{P}$ is a probability measure and

$$\begin{aligned}\mathbb{E}_{\mathbb{P}}[h(Y)] &= \sum h(y)\mathbb{P}(Y = y|X = x)\mathbb{P}(X = x) \\ &= \frac{1}{\mathbb{E}_{\mathbb{Q}}[g(X)]}\sum g(x)h(y)\mathbb{Q}(Y = y|X = x)\mathbb{Q}(X = x) \\ &= \frac{1}{\mathbb{E}_{\mathbb{Q}}[g(X)]}\mathbb{E}_{\mathbb{Q}}[g(X)\, h(Y)],\end{aligned}$$

as required. □

Proof of Prop 2.6, Part 2. Suppose first that the joint distribution $(X, Y) \sim \mathbb{Q}$ (respectively $\sim \mathbb{P}$) has density function $q(x, y)$ (respectively $p(x, y)$).

Again, comparing the definitions of Adv in (2.3) and Corr' in (2.2) it suffices to show that for any polynomial h we have $\mathbb{E}_{\mathbb{P}}[h(Y)] = \mathbb{E}_{\mathbb{Q}}[g(X)h(Y)]/\mathbb{E}_{\mathbb{Q}}[g(X)] = \mathbb{E}_{\mathbb{Q}}[g(X)h(Y)]$, since $\mathbb{E}_{\mathbb{Q}}[g(X)] = 1$. We may calculate

$$\mathbb{E}_{\mathbb{Q}}[g(X)h(Y)] = \int g(x)h(y)q(y|x)q(x)dx'dy = \int h(y)q(y|x)p(x)dxdy$$

where the last line followed since $g(X) = \ell(X)$. Now recall that $p(y|x) = q(y|x)$ by assumption and thus by the above,

$$\mathbb{E}_{\mathbb{Q}}[g(X)h(Y)] = \int h(y)p(y|x)p(x)dxdy = \mathbb{E}_{\mathbb{P}}[h(Y)],$$

as required.

If we suppose that $\mathbb{Q}$ is discrete, then by definition of $g = g(X)$ and since $Y|X$ has the same distribution under $\mathbb{P}$ and $\mathbb{Q}$

$$\begin{aligned}\mathbb{P}_{\mathbb{Q}}[g(X)h(Y)] &= \sum_{x,y} g(x)h(y)\mathbb{P}_{\mathbb{Q}}[X = x]\mathbb{P}_{\mathbb{Q}}[Y = y|X = x] \\ &= \sum_{x,y} h(y)\mathbb{P}_{\mathbb{P}}[X = x]\mathbb{P}_{\mathbb{Q}}[Y = y|X = x] \;=\; \mathbb{E}_{\mathbb{P}}[h(Y)]\end{aligned}$$

as required. □

Remark 6.1. *In Proposition 2.6 we have seen that the quantities* $\mathsf{Adv}_{\leqslant D}(\mathbb{P}, \mathbb{Q})$ *and* $\mathsf{Corr}'_{\leqslant D}(g(X), \mathbb{Q})$ *are the same under some circumstances (i.e., for some triples* $g(X), \mathbb{P}, \mathbb{Q}$*). In this remark we observe that the upper bounds for* $\mathsf{Adv}_{\leqslant D}(\mathbb{P}, \mathbb{Q})$ *in terms of* r_α *in this paper imply the upper bounds on* Corr' *in terms of* k_α *in [SW22]. Note the proofs of the respective results follow the same sequence of steps. The* r_α *enjoy some nice properties, see Section 3, and it is instructive to see that they extend the cumulants* k_α *of [SW22].*

To be precise, for an additive Gaussian model, Theorem 2.2 of [SW22] shows $\mathsf{Corr}'^2_{\leqslant D}(g(X), \mathbb{Q}) \leqslant \sum_{|\alpha|\leqslant D} \kappa_\alpha^2/\alpha!$ *and we recover this result via Propositions 2.6 and 2.8. By Proposition 2.6, Part 1 and its proof, there exists a joint distribution* $(X, Y) \sim \mathbb{P}$ *such that* $g(X) = d\mathbb{P}_X/d\mathbb{Q}_X$ *and*

$$\mathsf{Adv}_{\leqslant D}(\mathbb{P}, \mathbb{Q}) = \mathsf{Corr}'_{\leqslant D}(g(X), \mathbb{Q}).$$

Now note that by Proposition 2.8, $\mathsf{Adv}_{\leqslant D}(\mathbb{P},\mathbb{Q}) \leqslant \sum_{|\alpha|\leqslant D} r_\alpha^2/\alpha!$ *and thus it suffices to show we have* $r_\alpha(\mathbb{P},\mathbb{Q}) = \kappa_\alpha(g(X),\mathbb{Q})$ *for the given* $g(X)$, $\mathbb{P}$ *and* $\mathbb{Q}$. *Recall the non-recursive expression for the* r_α *from* (3.2)*:*

$$r_\alpha(\mathbb{P},\mathbb{Q}) = \sum_{\varnothing\subseteq\delta\subseteq\alpha} \mathbb{E}_P[X^\delta] \sum_{\tau\in\mathcal{P}(\alpha\backslash\delta)} (-1)^{|\tau|}\, |\tau|! \prod_{\gamma\in\tau} \mathbb{E}_Q[X^\gamma].$$

The κ_α *in [SW22] are the joint cumulants of the random variables indexed by* α *and* $g(X)$*:*

$$\kappa_\alpha(g(X),\mathbb{Q}) = \sum_{\pi\in\mathcal{P}(\alpha\cup\star)} (-1)^{|\pi|-1}\, (|\pi|-1)! \prod_{\gamma\in\pi} \mathbb{E}_Q[X^\gamma]$$

where we denote $X^\star = g(X)$. *Note each partition* π *above contains* $\star$ *in some part, say* η, *and let* $\eta' = \eta\backslash\star$. *Let* π' *be the partition* $\pi\backslash\eta$ *and note each* $\pi \in \mathcal{P}(\alpha\cup\star)$ *gives rise to a unique* $\pi' \in \mathcal{P}((\alpha\cup\star)\backslash\eta) = \mathcal{P}(\alpha\backslash\eta')$ *with* $|\pi'| = |\pi|-1$. *Thus,*

$$\kappa_\alpha = \sum_{\varnothing\subseteq\eta\subseteq\alpha} \mathbb{E}_\mathbb{Q}[X^{\eta'} g(X)] \sum_{\pi'\in\mathcal{P}(\alpha\backslash\eta')} (-1)^{|\pi'|}\, |\pi'|! \prod_{\gamma\in\pi'} \mathbb{E}_Q[X^\gamma].$$

Now we simply note that $\mathbb{E}_\mathbb{Q}[X\eta' g(X)] = \mathbb{E}_\mathbb{P}[X\eta']$, *and thus* $r_\alpha = \kappa_\alpha$ *as required.*

7 Reduction from PDS recovery to testing 1 vs 2 communities

We recall Definition 2.2 of $\mathbb{P}_{\text{Gaussian}}(n,k,\lambda,M,x)$, which plants M communities of total expected size k, with signal strength λ, and expected proportions of nodes within each community $x = (x_1,\ldots,x_M)$. In the special case, as below, when each community has the same expected size, i.e. $x_i = 1/M$ for each i, we drop the last argument and write $\mathbb{P}_{\text{Gaussian}}(n,k,\lambda,M)$. Given a hypothesis testing problem to distinguish between $H_0 : Y \sim \mathbb{Q}$ and $H_1 : Y \sim \mathbb{P}$, we say an algorithm $B := B(Y)$ *strongly distinguishes* if $\mathbb{P}_0(B=1) = o(1)$ and $\mathbb{P}_1(B=0) = o(1)$, or in other words, if both the probability of type 1 and type 2 error approach zero with growing n.

Conjecture 7.1 (testing 1-vs-2 hypothesis)**.** *If* k *and* λ *scale with* n *such that* $\lambda^2(k^2/n \vee 1) = o(1)$, *then no sequence of randomized polynomial time algorithms* $B_n := B_n(Y)$ *can strongly distinguish between* $H_0 : Y \sim \mathbb{Q} = \mathbb{P}_{\text{Gaussian}}(n,k,\lambda,1)$ *and* $H_1 : Y \sim \mathbb{P} = \mathbb{P}_{\text{Gaussian}}(n,k,\lambda,2)$.

We will say that an algorithm, A_n, achieves *weak recovery* in $\mathbb{Q} = \mathbb{P}_{\text{Gaussian}}(n,k,\lambda,1)$ if there exists $\varepsilon > 0$ (not dependent on n) such that A_n returns a set I such that w.h.p. $|I\cap S| \geqslant \varepsilon k$ and $|I| \leqslant 1.1k$, where S is the planted set in $\mathbb{Q}$.

Theorem 7.2. *Assume the testing 1-vs-2 hypothesis (Conjecture 7.1). If* k *and* λ *scale with* n *such that* $k \geqslant \log\log n$, $\lambda^2(k^2/n \vee 1) = o(1)$ *and* $\lambda \geqslant k^{-1/2}\log n$, *then no sequence of randomized polynomial-time algorithms* A_n *achieves weak recovery in* $\mathbb{P}_{\text{Gaussian}}(n,k,\lambda,1)$.

High-level description. The intuition for the proof is as follows. Assume this were not true; namely, assume there exists a sequence of algorithms A_n that achieves weak recovery in $\mathbb{Q} = \mathbb{P}_{\text{Gaussian}}(n,k,\lambda,1)$. We will prove that if this is the case, one can construct a sequence of algorithms that can strongly distinguish between one and two communities, leading to a contradiction.

In more detail, consider that we have access to an algorithm $A := A_n$ that outputs a set of indices I, which weakly recovers the *single* planted community under $\mathbb{Q}$. Then we can boost this to output I' achieving exact recovery. The idea will be that we can then construct an algorithm $B := B_n$ that can run some (polynomial-time) checks on the output set I', given to us by A and boosting, that w.h.p. can distinguish between $\mathbb{P}$ or $\mathbb{Q}$. In designing the algorithm B, we have to consider both cases, $H_0 : Y \sim \mathbb{Q}$ and $H_1 : Y \sim \mathbb{P}$. Notice that under H_0, we know that A weakly recovers the single community; however, under the two-communities case, H_1, we have no guarantees on what set I will be returned by algorithm A, and this is the real challenge in constructing the algorithm B.

It turns out that the key to constructing the distinguishing algorithm B is to consider both the trace and sum of $Y_{I'}$, the data reduced to only the set I', which we have after boosting the set I returned by algorithm A. The high-level idea is that w.h.p., under H_0 we have a good idea of what values the sum and trace should take, as we are guaranteed that I' has large overlap with the single community. However, under H_1, either the sum will be too small or the trace too large, depending on the makeup of block I' relative to the two communities.

Let us think more about what can happen with the output of A and boosting, in the two-community case, H_1. If the returned block I' has large overlap with the pair of planted communities, the trace will likely be $> 3k\lambda/2$ (which is much larger than the expected trace in the single-community case) and so we construct B to return 1 if the trace is very large. On the other hand, if the returned block I has small overlap with the pair of planted communities, then the sum will likely be $< 99k\lambda/100$ (which is smaller than the expected sum in the single-community case) and therefore we construct B to also return 1 when the sum is too small. What the previous two sentences say is that constructing B to use a large trace or small sum threshold will correctly guess $\mathbb{P}$ under H_1 (when $\mathbb{P}$ is true), irrespective of the contents of set I' that the algorithm A and boosting return in this case.

One extra trick that helps control the noise error when we calculate the trace and sum of the returned block is that of *cloning* as in, for example, [BBH18]. Note that under $\mathbb{P}$ and $\mathbb{Q}$ the matrix Y can be written $Y = X + Z$ where Z is a Gaussian matrix independent from the signal X, which will take different forms depending on whether the data is generated under H_0 or H_1. Cloning means that given Y we may generate three matrices $Y^{(1a)}$, $Y^{(1b)}$ and $Y^{(2)}$: such that they have the distributions $Y^{(1a)} = X/2 + Z^{(1a)}$, $Y^{(1b)} = X/2 + Z^{(1b)}$ and $Y^{(2)} = X/\sqrt{2} + Z^{(2)}$. Thus, we have constructed three matrices which are sums of a scaled version of the original signal matrix with independent noise matrices $Z^{(1a)}$, $Z^{(1b)}$ and $Z^{(2)}$. (The signal strengths are now $\lambda' = \lambda/\sqrt{2}$ and $\lambda'' = \lambda/2$.) This allows us to run algorithm A on matrix $Y^{(1a)}$, returning output I, and then to boost I to I' using $Y^{(1b)}$, resulting in a set I' that is independent of $Z^{(2)}$. Finally, we use $Y^{(2)}$ for the trace and sum tests.

See Figure 7.2 for an illustration of the types of returned blocks I' that can be output by algorithm A and boosting in $H_0 : Y \sim \mathbb{Q}$, the single-community case, and $H_1 : Y \sim \mathbb{P}$, the two-community case. The figure also shows the approximate expected trace and sum of the submatrix restricted to I' in each of the cases.

Algorithm Definition. Mathematically, the algorithm $B = B_n$ is defined as follows — see also Fig. 7.1. Recall that we have been given Y, an $n \times n$ matrix and that A_n is an algorithm that achieves ε-weak recovery in $\mathbb{Q} = \mathbb{P}_{\text{Gaussian}}(n, k, \lambda, 1)$, the one-community model.

First, generate Z' and Z'', both $n \times n$ symmetric matrices, by independently sampling $Z'_{ij} \sim N(0, 1)$ for $i \leqslant j$ and setting $Z'_{ji} = Z'_{ij}$ (and independently generate Z'' in the same way). Then set

$$Y^{(1)} = \frac{Y + Z'}{\sqrt{2}}, \qquad Y^{(2)} = \frac{Y - Z'}{\sqrt{2}} \qquad \text{and} \qquad Y^{(1a)} = \frac{Y^{(1)} + Z''}{\sqrt{2}}, \qquad Y^{(1b)} = \frac{Y^{(1)} - Z''}{\sqrt{2}}.$$

Next, run $A = A_n$ on $Y^{(1a)}$ to produce output I. Denote by e_I the binary vector with support I. Now, let $v = Y^{(1b)} e_I$ and define I' by thresholding: set $i \in I'$ if $v_i > \lambda k/(\log \log k)$. Then, with

Algorithm Weak recovery to testing 1 vs 2

Inputs: Matrix $Y \in \mathbb{R}^{n\times n}$, algorithm $A = A_n$.

1. (**Pre-processing — cloning step**.)
   Generate independent matrices $Z', Z'' \sim N(0,1)^{\otimes n\times n}$ with independent Gaussian entries. Set $\lambda' = \lambda/\sqrt{2}$, $\lambda'' = \lambda/2$ and compute the matrices
   $$Y^{(1)} = \frac{1}{\sqrt{2}}(Y + Z') \quad \text{and} \quad Y^{(2)} = \frac{1}{\sqrt{2}}(Y - Z'),$$
   $$Y^{(1a)} = \frac{1}{\sqrt{2}}(Y^{(1)} + Z'') \quad \text{and} \quad Y^{(1b)} = \frac{1}{\sqrt{2}}(Y^{(1)} - Z'').$$
2. (**Weak recovery**.)
   Run algorithm A on $Y^{(1a)}$ to get output $I \subset [n]$. Let e_I be the vector with i-th entry 1 if $i \in I$ and 0 otherwise.
3. (**Boosting step**.)
   Let $v = Y^{(1b)} e_I$, then define $I' = \{\ell \,:\, v_\ell > k\lambda/(\log\log k)\}$ to get $I' \subset [n]$.
4. (**Check size, sum and trace of returned block**.)
   Consider submatrix $M = Y^{(2)}_{I'}$ and set $\lambda' = \lambda/\sqrt{2}$. If $|I'| - k \leqslant \sqrt{k}\log k$, $\mathrm{sum}(M) \geqslant \lambda' k^2 - k^{3/2}\log^2 k$ and $\mathrm{tr}(M) \leqslant 3\lambda' k/2$ then output 0. Otherwise, output 1.

Figure 7.1: Reduction from weak recovery of planted dense submatrix (PDS) to testing between one-block and two-block models. See also Fig. 7.2.

$$\lambda' = \lambda/\sqrt{2},$$
$$B = \begin{cases} 0 & \text{if } \big||I'| - k\big| \leqslant \sqrt{k}\log k, \;\; \mathrm{sum}(Y^{(2)}_{I'}) \geqslant \lambda' k^2 - k^{3/2}\log^2 k, \;\; \text{and } \mathrm{tr}(Y^{(2)}_{I'}) \leqslant 3\lambda' k/2, \\ 1 & \text{otherwise.} \end{cases} \tag{7.1}$$

To show that B distinguishes correctly w.h.p. requires some auxiliary lemmas. We give results on cloning from [BBH18]. We also record some well-known tail bounds for normal and binomially distributed random variables. We first prove Theorem 7.2 assuming these lemmas, then provide the lemmas in the next section.

Proof of Theorem 7.2. Let k and λ scale with n as prescribed. Also, let $\mathbb{Q} = \mathbb{P}_{\text{Gaussian}}(n,k,\lambda,1)$ and let $\mathbb{P} = \mathbb{P}_{\text{Gaussian}}(n,k,\lambda,2)$. Assume for the sake of contradiction that a randomized polynomial-time algorithm A_n achieves weak recovery in $\mathbb{Q}$. We will show that under this assumption, the algorithm B_n, defined in (7.1), achieves strong detection between $H_0 : Y \sim \mathbb{Q}$ and $H_1 : Y \sim \mathbb{P}$, leading to a contradiction.

Recall, to show that B achieves strong detection, we must prove that $\mathbb{P}_0(B = 1) = o(1)$ and $\mathbb{P}_1(B = 0) = o(1)$. In what follows, we call these proofs **Part 1** and **Part 2**.

Part 1: $\mathbb{P}_0(B = 1) = o(1)$. Throughout **Part 1** we assume the data is generated from $Y = X + Z \sim \mathbb{Q}$, the one-community model.

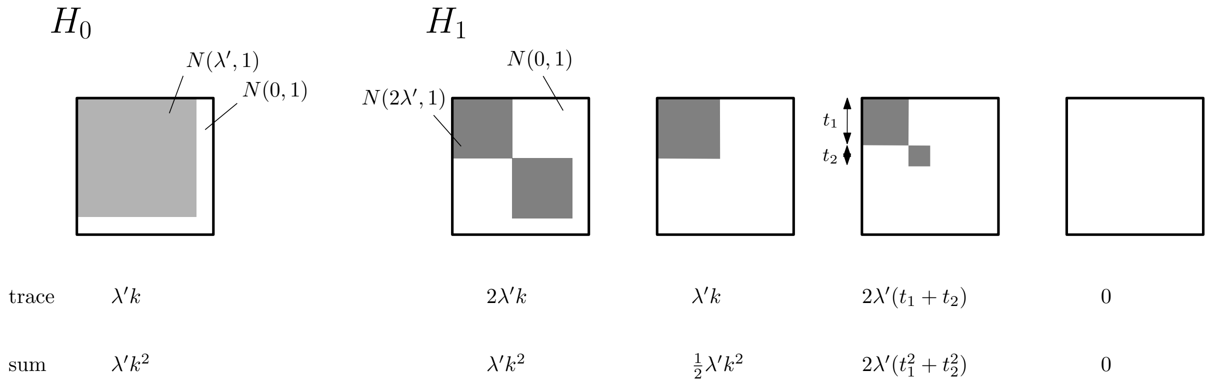


Figure 7.2: Diagram showing possible returned blocks of the matrix after running the recovery algorithm and boosting. In detail, we depict $Y^{(2)}_{I'}$, i.e. the submatrix of $Y^{(2)}$ restricted to the indices I' output by the recovery algorithm A_n run on $Y^{(1a)}$ and boosting on $Y^{(1b)}$ in the proof of Theorem 7.2. Here, the shaded regions represent entries $Y^{(2)}_{ij}$ where i, j are both part of the same planted community, and we have depicted just the submatrix Y_I for index sets I with size approximately k. Under H_0, we show the recovery algorithm and boosting returns a set I' with set-difference $o(k)$ from the planted structure S. Under H_1 there are no guarantees on what the algorithm will return, and we present a selection of possibilities: (i) the entire planted structure (both communities), (ii) one of the two planted communities, (iii) t_1 vertices from community 1 and t_2 in community 2 and (iv) none of the planted structure. For each possibility we give the expected values of the trace and sum of the submatrix $Y^{(2)}_{I'}$ (to leading order). In the proof we show that we can distinguish between H_0 and H_1 via the sum and trace of the returned block in $Y^{(2)}$.

By Lemma 7.3 it is equivalent to sample $Y^{(1a)}, Y^{(1b)}$ and $Y^{(2)}$ as follows. Construct X in the usual way: i.e. independently for each $i \in [n]$ sample the community label σ_i such that $\sigma_i = 1$ with probability k/n and $\star$ otherwise. For each $i, j \in [n]$ with $i \leqslant j$ $X_{ij} = \mathbf{1}[\sigma_i = \sigma_j = 1]$, and for $i > j$ set $X_{ij} = X_{ji}$.

Sample $n \times n$ matrix $Z^{(1a)}$. Independently for each $i, j \in [n]$ sample $Z^{(1a)}_{ij} \sim N(0, 1)$ and for $i > j$ set $Z^{(1)}_{ij} = Z^{(1)}_{ji}$. Independently sample $Z^{(1b)}$ and $Z^{(2)}$ in the same way. Then set

$$Y^{(1a)} = \tfrac{1}{2}X + Z^{(1a)}, \quad Y^{(1b)} = \tfrac{1}{2}X + Z^{(1b)} \quad \text{and} \quad Y^{(2)} = \tfrac{1}{\sqrt{2}}X + Z^{(2)}.$$

Let I be the returned set after running the algorithm A_n on $Y^{(1a)}$, and let I' be the returned set after the boosting step. The important observation is that I' is independent of $Z^{(2)}$.

We define a number of events which hold w.h.p. and show that if these events occur then deterministically the algorithm B returns 0.

Definition of events

Define E_0 to be the event that for some ε independent of n, $|S_1 \cap I| \geqslant \varepsilon k$ and $|I| \leqslant 1.1k$.

Define E_1 to be the event that $I' = S_1$ i.e. that after the boosting step, the set I' is exactly the planted community.

Let E_2 be the event that $||S_1| - k| \leqslant \sqrt{k} \log k$.

The next two events concern the sum and trace of the noise matrix $Z^{(2)}$ induced on the set I'. Recall since I' depends only on $Y^{(1a)}$ and $Y^{(1b)}$, it is independent of $Z^{(2)}$. Let E_3 be the event that

$$\mathrm{sum}(Z^{(2)}_I) \geqslant -k \log k \tag{7.2}$$

and let E_4 be the event that

$$\mathrm{tr}(Z_I^{(2)}) \leqslant \sqrt{k}\log k. \tag{7.3}$$

Claim 1a. Under H_0, if $E_0, \ldots, E_4$ hold then the algorithm B outputs 0.

Claim 1b. Under H_0, the events E_0, $\ldots$, E_4 hold w.h.p.

Proof of Claim 1a.

We will now calculate bounds on the sum and the trace. Notice that that $\mathrm{sum}(Y_{I'}^{(2)}) = \frac{1}{\sqrt{2}}\mathrm{sum}(X_{I'}) + \mathrm{sum}(Z_{I'}^{(2)})$. Let $t_1 = |S_1 \cap I'|$ and note that $\mathrm{sum}(X'_{I'}) = \frac{1}{\sqrt{2}}\lambda t_1^2 = \lambda' t_1^2$, where we recall $\lambda' = \lambda/\sqrt{2}$.

Recall, when E_1 holds then after boosting $I' = S_1$. Hence since E_0, E_1 and E_2 hold then,

$$|I'| = |S_1| \leqslant k + O(\sqrt{k}\log k) + o(k) = k + o(k).$$

Therefore

$$\mathrm{sum}(X_{I'}) = \lambda' k^2(1 + o(1)).$$

Hence, since E_2 holds,

$$\mathrm{sum}(Y_{I'}^{(2)}) \geqslant \lambda' k^2(1 + o(1)) - k\log k \geqslant \lambda' k^2(1 + o(1)) - \lambda' k^{3/2}\log^{-1} k,$$

where the second inequality follows since $\lambda' \geqslant k^{-1/2}\log^2 k$. Similarly, since E_0, E_1 and E_3 hold,

$$\mathrm{tr}(Y_{I'}^{(2)}) \leqslant \lambda' k(1 + o(1)) + \sqrt{k}\log k \leqslant \lambda' k(1 + o(1)) + \lambda' k\log^{-1} k.$$

Hence, w.h.p. the interval I' passes the three checks in (7.1) and B returns 0 as required, and we have proven the claim.

Proof of Claim 1b.

Event E_0 holds w.h.p.

By assumption on A_n, $\mathbb{P}_0(E_0) = 1 - o(1)$.

Event E_1 holds w.h.p.

Let $v = Y^{(1b)}e_I$. Fix $\ell \in [n]$ and calculate the ℓ-th entry of vector v as follows:

$$\begin{aligned} v_\ell &= [\tfrac{1}{2}Xe_I + Z^{(1b)}e_I]_\ell \\ &= \tfrac{1}{2}\sum_{i\in I} X_{\ell i} + \sum_{i\in I} Z_{\ell i}^{(1b)} \\ &= \begin{cases} \frac{1}{2}\lambda|I \cap S| + \sum_{v\in I} Z_{\ell i}^{(1b)} & \text{if } \ell \in S, \\ \sum_{i\in I} Z_{\ell i}^{(1b)} & \text{if } \ell \notin S. \end{cases} \end{aligned}$$

Recall that by the design of the algorithm B the set I is independent of the noise $Z^{(1b)}$. Thus $\sum_{i\in I} Z_{\ell i}^{(1b)}$ has distribution $N(0, |I|)$. Let $\mathcal{E}_\ell$ be the event that $|\sum_{i\in I} Z_{\ell i}^{(1b)}| \geqslant 2\sqrt{k\log n}$. Recall, that if E_0 holds then $|I| \leqslant 1.1k$ and $|I \cap S| \geqslant \varepsilon k$. Hence if E_0 holds we may apply Lemma 7.4 (with $m = 1$,

$r = \log n + \log\log n$ and $\sigma = (1.1k)^{1/2}$ note $\sigma\sqrt{2r} = (2.2k(\log n + \log\log n))^{1/2} \leqslant 2(k\log n)^{1/2})$ to see that

$$\mathbb{P}(\mathcal{E}_\ell) \leqslant \frac{2}{n\log n}.$$

Now note if E_0 and $\mathcal{E}_\ell$ hold and $\ell \in S$ then for some $\varepsilon > 0$ independent of n,

$$v_\ell \geqslant \tfrac{1}{2}\lambda\varepsilon k - 2k^{1/2}\log^{1/2} n \geqslant \tfrac{1}{2}\lambda\varepsilon k - 2k\lambda\log^{-1/2} n$$

where the second inequality followed since $\lambda \geqslant k^{-1/2}\log n$. But note that the RHS above is greater than the threshold $k\lambda/(\log\log k)$ for large n (since k grows with n). Hence if E_0 and $\mathcal{E}_\ell$ hold, and $\ell \in S$ then $\ell \in I'$. Similarly for $\ell \notin S$ if E_0 and $\mathcal{E}_\ell$ hold then

$$v_\ell \leqslant 2k^{1/2}\log^{1/2} n \leqslant 2k\lambda\log^{-1/2} n$$

and the RHS above is below the threshold $k\lambda/(\log\log k)$. Hence if E_0 and $\mathcal{E}_\ell$ hold, and $\ell \notin S$ then $\ell \notin I'$. We may now take a union bound over $\ell \in [n]$, and thus w.h.p. E_0 and all $\mathcal{E}_\ell$ hold and so w.h.p. $I' = S$.

Event E_2 holds w.h.p.
The event E_2 is an indicator that the single planted community has approximately expected size. Let $S_1 \subseteq [n]$ be the set of vertices labeled '1', i.e. $S_1 = \sigma^{-1}(1)$ for σ sampled as above. Note that $|S_1| \sim \mathrm{Ber}(n, k/n)$; thus, by Lemma 7.5, we have

$$\mathbb{P}_0(E_0) = \mathbb{P}_0\left(||S_1| - k| \leqslant \sqrt{k}\log k\right) \geqslant 1 - 2\exp\left\{-\frac{1}{3}(\log k)^2\right\} = 1 - o(1),$$

since k is growing with n. Note by construction the community assignment is the same for $Y^{(1a)}$, $Y^{(1b)}$ and $Y^{(2)}$.

Events E_3 and E_4 hold w.h.p.
We have already established E_0, E_1 and E_2 holds w.h.p. and now show that if E_0, E_1, E_2 hold then w.h.p. E_3 and E_4 hold.

Since E_0 and E_1 hold we have $|I'| = k(1 + o(1))$. Recall I' and $Z^{(2)}$ are independent hence if $|I'| = t$ then the sum of $Z^{(2)}_{I'}$ has distribution $N(0, t + 4\binom{t}{2}) = N(0, 2t^2 - t)$. (Since the trace/sum of diagonal is distributed as $N(0,t)$, and the sum above the diagonal is distributed as $N(0, \binom{t}{2})$.) Hence we may apply Lemma 7.4 (with $m = 1$, $r = \log k$ and $\sigma = \sqrt{2}t$), and w.h.p.

$$\mathrm{sum}(Z^{(2)}_{I'}) \geqslant -\sqrt{2}t\log^{1/2} k.$$

Since if E_1 and E_2 hold then $t = k(1 + o(1))$, and these hold w.h.p., then w.h.p.

$$\mathrm{sum}(Z^{(2)}_{I'}) \geqslant -\sqrt{2}t\log^{1/2} k \geqslant -k\log k,$$

i.e. E_3 holds.

Similarly if $|I'| = t$, and since I' is independent of $Z^{(2)}_{I'}$ then the trace of $Z^{(2)}_{I'}$ has distribution $N(0,t)$. Hence by Lemma 7.4 (with $m = 1$, $r = \log k$ and $\sigma = \sqrt{t}$), w.h.p.

$$\mathrm{tr}(Z^{(2)}_{I'}) \leqslant \sqrt{t}\log^{1/2} k.$$

Then, as earlier, w.h.p. $\mathrm{tr}(Z^{(2)}_{I'}) \leqslant \sqrt{k}\log k$ and hence w.h.p. E_4 holds.

This concludes the proof of Claim 1b, and thus we have established Part 1 of the proof.

Part 2: $\mathbb{P}_1(B = 0) = o(1)$. Throughout **Part 2** we assume the data is generated from $Y = X + Z \sim \mathbb{Q}$, the two community model.

Equivalently, we may sample $Y^{(1a)}$, $Y^{(1b)}$ and $Y^{(2)}$ as follows. Independently for each $i \in [n]$ sample the community label σ_i such that $\sigma_i = 1$ with probability $k/2n$ $\sigma_i = 2$ with probability $k/2n$ and $\star$ otherwise. For each $i, j \in [n]$ with $i \leqslant j$ $X_{ij} = 2\lambda \mathbf{1}[\sigma_i = \sigma_j = \ell]$ for $\ell \in \{1, 2\}$, and for $i > j$ set $X'_{ij} = X'_{ji}$.

Sample $n \times n$ matrix $Z^{(1a)}$ as in Part 1. Independently for each $i, j \in [n]$ sample $Z^{(1a)}_{ij} \sim N(0, 1)$ and for $i > j$ set $Z^{(1)}_{ij} = Z^{(1)}_{ji}$. Independently sample $Z^{(1b)}$ and $Z^{(2)}$ in the same way. Then set

$$Y^{(1a)} = \tfrac{1}{2}X + Z^{(1a)}, \quad Y^{(1b)} = \tfrac{1}{2}X + Z^{(1b)} \quad \text{and} \quad Y^{(2)} = \tfrac{1}{\sqrt{2}}X + Z^{(2)}.$$

This direction is a bit more subtle since we have no guarantees on what set I the algorithm A will return when given $Y^{(1)}$ with *two* planted communities — and so no knowledge of the set I' which will be returned after boosting see Figure 7.2.

Note if $||I'| - k| > \sqrt{k}\log k$ then the algorithm B returns 1 and we would be done, hence we may assume that $|I'| \in k \pm \sqrt{k}\log k$.

We define events E_1 that the planted communities have roughly the expected size. Additionally, E_2, E_3 concerning the maximum noise in $Z^{(2)}$ of the returned block (of size about k). In particular, E_2 will be the event that the total sum and E_3 that the trace of the noise in the returned block are small.

To define E_1, for $i = 1, 2$, let $S_i \subseteq [n]$ be the set of vertices labeled 'i', i.e. $S_i = \sigma^{-1}(i)$. Let E_1 be the event that $|S_1|, |S_2| \leqslant k/2 + \sqrt{k}\log k$.

Define E_2, as the event that the following bound holds:

$$\mathrm{sum}(Z^{(2)}_{I'}) < k \log k,$$

and define E_3 as the event that the following bound holds:

$$\mathrm{tr}(Z^{(2)}_{I'}) > -k^{1/2}\log k.$$

Claim 2a. *Under H_1, if E_1, E_2, E_3 hold then the algorithm B outputs 1.*

Claim 2b. *Under H_1, the events E_1, E_2, E_3 hold w.h.p.*

We first note that the proof follows by the claims before proving the claims. If B outputs 1 when $E_1, \ldots, E_4$ occur then $\mathbb{P}_1(B = 0) = 1 - \mathbb{P}_1(B = 1) \leqslant 1 - \mathbb{P}_1(E_1 \cap \ldots \cap E_4) \leqslant \mathbb{P}_1(E_1^c) + \ldots + \mathbb{P}_1(E_4^c) = o(1)$ and we would be done, as required.

Proof of Claim 2a. Assume E_1, E_2, E_3 hold and that the algorithm B outputs $B = 0$ (and seek a contradiction).

We will see that since E_2 and E_3 hold we get bounds on the sum and trace of $Y^{(2)}_{I'}$ in terms of the number of vertices of each planted community in I'. In particular, let $t_1 = t_1(I') = |S_1 \cap I'|$

and let $t_2 = t_2(I') = |S_2 \cap I'|$. See also Figure 7.2. Note that since E_1 holds and each t_i is a subset of S_i, then we know that $t_1, t_2 \leqslant k/2 + \sqrt{k}\log k$. Notice

$$\mathrm{sum}(Y_{I'}^{(2)}) = \mathrm{sum}(X'_{I'}) + \mathrm{sum}(Z_{I'}^{(2)}) = 2\lambda'(t_1^2 + t_2^2) + \mathrm{sum}(Z_{I'}^{(2)}).$$

Since E_2 holds,

$$\mathrm{sum}(Y_{I'}^{(2)}) \leqslant 2\lambda'(t_1^2 + t_2^2) + k\log k. \tag{7.4}$$

Similarly, since E_3 holds,

$$\mathrm{tr}(Y_{I'}^{(2)}) = 2\lambda'(t_1 + t_2) + \mathrm{tr}(Z_{I'}^{(2)}) \geqslant 2\lambda'(t_1 + t_2) - k^{1/2}\log k, \tag{7.5}$$

and hence since $\lambda' \geqslant k^{-1/2}\log^2 k$,

$$\mathrm{tr}(Y_{I'}^{(2)}) \geqslant 2\lambda'(t_1 + t_2) - \lambda' k \log^{-1} k. \tag{7.6}$$

In the next bit of the proof we show that since we assumed B outputs 0, this means the trace of $Y_{I'}^{(2)}$ is small, and thus $t_1 + t_2$ is small. Then we show $t_1 + t_2$ small implies that $Y_{I'}^{(2)}$ will have small sum, i.e. fail the sum test and the algorithm outputs 1, a contradiction.

Now, since B output 0, and by unpacking the definition in (7.1), the following must hold:

$$\mathrm{tr}(Y_{I'}) \leqslant \frac{3}{2}\lambda' k, \tag{7.7}$$

and thus by (7.6) and (7.7) we may deduce that

$$t_1 + t_2 \leqslant \tfrac{3}{4}k + \tfrac{1}{2}k\log^{-1} k. \tag{7.8}$$

Recall that for any y_1, y_2 with $y_1, y_2 \leqslant \eta$ that $y_1^2 + y_2^2 \leqslant \eta^2 + \min\{\eta^2, (y_1 + y_2 - \eta)^2\}$. Hence by (7.8) and since $t_1, t_2 \leqslant k/2 + \sqrt{k}\log k$ we have

$$t_1^2 + t_2^2 \leqslant (k/2 + \sqrt{k}\log k)^2 + (k/4)^2 \leqslant \tfrac{5}{16}k^2 + 2k^{3/2}\log k \leqslant \tfrac{1}{3}k^2.$$

Thus by (7.4),

$$\mathrm{sum}(Y_{I'}^{(2)}) \leqslant \tfrac{2}{3}\lambda' k^2 + k\log k.$$

Recalling that $\lambda' \geqslant k^{-1/2}\log^2 k$,

$$\mathrm{sum}(Y_{I'}^{(2)}) \leqslant \tfrac{2}{3}\lambda' k^2 + \lambda' k^{3/2}\log^{-1} k,$$

and so, inspecting the definition in (7.1), B returns '1', a contradiction, and we have proven Claim 2a.

Proof of Claim 2b.

Event E_1 holds w.h.p.
By Lemma 7.5 w.h.p. $|S_1| \leqslant k/2 + \sqrt{k}\log k$ for large k, and similarly for S_2. Hence, $\mathbb{P}_1(E_1) = 1 - o(1)$.

Events E_2 and E_3 hold w.h.p.
Let $t = |I'|$ then, as in part 1 of this proof, since $Z^{(2)}$ is independent of I', $\mathrm{sum}(Z_{I'}^{(2)}) \sim N(0, 2t^2 - t)$ and $\mathrm{tr}(Z_{I'}^{(2)}) \sim N(0, t)$. Also we may assume that $|I| \in k \pm \sqrt{k}\log k$. Thus the events E_2 and E_3 hold by applying Lemma 7.4, as in part 1 of this proof.

This completes the proof of Claim 2b, which concludes the proof of the theorem. □

7.1 Lemmas used to prove Theorem 7.2

For the Gaussian cloning trick, we note Lemma 10.2 of [BBH18], see also Figure 23 of that paper. Following notation of that paper, given a distribution $\mathbb{P}$, we denote by $\mathbb{P}^{\otimes n}$ (respectively $\mathbb{P}^{\otimes n \times n}$) the distribution $(X_1, \ldots, X_n)$ (respectively the distribution on $n \times n$ matrix) where the X_i are i.i.d and $X_i \sim \mathbb{P}$.

Lemma 7.3 (Gaussian cloning)**.** *Given $M \sim \mathcal{L}(A + N(0,1)^{\otimes n \times n})$ for any fixed matrix $A \in \mathbb{R}^{n \times n}$, sample $G = N(0,1)^{\otimes n \times n}$ and set $M^1 = (M+G)/\sqrt{2}$ and $M^2 = (M-G)/\sqrt{2}$. then $(M^1, M^2) \sim \mathcal{L}\Big(\frac{1}{\sqrt{2}}A + N(0,1)^{\otimes n \times n}\Big)^{\otimes 2}$.*

The following standard result will help us prove bounds on the noise in the returned block $Z_I^{(2)}$.

Lemma 7.4. *Let $X_1, \ldots, X_m$ be random variables, each of which is distributed as $N(0, \sigma_i^2)$ for some $\sigma_i \leqslant \sigma$. Then*

$$\mathbb{P}\big(\max_{i \in 1,\ldots,m} X_i \geqslant \sigma\sqrt{2(\log m + r)}\,\big) \leqslant e^{-r}.$$

We finally use a tail bound for binomial random variables that follows, for example, from Theorems 2.1 and 2.8 of [JŁR11].

Lemma 7.5. *Let random variables $X_1, \ldots, X_m$ be independent, with $0 \leqslant X_j \leqslant 1$ for each j. Let $S = \sum_{j=1}^m X_j$ and $\mu = \mathbb{E}(S)$. Then*

$$\mathbb{P}(|S - \mu| \geqslant x\sqrt{\mu}) \leqslant 2e^{-\frac{1}{3}x^2} \quad \text{for } 0 < x \leqslant \sqrt{\mu}\,.$$

Acknowledgements

This work began when the authors were visiting the Simons Institute for the Theory of Computing during the program on Computational Complexity of Statistical Inference in Fall 2021. We are grateful to Guy Bresler for helpful discussions.

References

[Abb17] Emmanuel Abbe. Community detection and stochastic block models: recent developments. *The Journal of Machine Learning Research*, 18(1):6446–6531, 2017.

[ACV13] Ery Arias-Castro and Nicolas Verzelen. Community detection in random networks. *arXiv preprint arXiv:1302.7099*, 2013.

[BAH$^+$22] Afonso S Bandeira, Ahmed El Alaoui, Samuel B Hopkins, Tselil Schramm, Alexander S Wein, and Ilias Zadik. The Franz-Parisi criterion and computational trade-offs in high dimensional statistics. *Advances in Neural Information Processing Systems*, 35:33831–33844, 2022.

[BBH18] Matthew Brennan, Guy Bresler, and Wasim Huleihel. Reducibility and computational lower bounds for problems with planted sparse structure. In *Conference On Learning Theory*, pages 48–166. PMLR, 2018.

[BHK$^+$19] Boaz Barak, Samuel B Hopkins, Jonathan Kelner, Pravesh K Kothari, Ankur Moitra, and Aaron Potechin. A nearly tight sum-of-squares lower bound for the planted clique problem. *SIAM Journal on Computing*, 48(2):687–735, 2019.

[BMR21] Jess Banks, Sidhanth Mohanty, and Prasad Raghavendra. Local statistics, semidefinite programming, and community detection. In *Proceedings of the 2021 ACM-SIAM Symposium on Discrete Algorithms (SODA)*, pages 1298–1316. SIAM, 2021.

[BR13] Quentin Berthet and Philippe Rigollet. Complexity theoretic lower bounds for sparse principal component detection. In *Conference on learning theory*, pages 1046–1066. PMLR, 2013.

[CX16] Yudong Chen and Jiaming Xu. Statistical-computational tradeoffs in planted problems and submatrix localization with a growing number of clusters and submatrices. *The Journal of Machine Learning Research*, 17(1):882–938, 2016.

[DKMZ11] Aurelien Decelle, Florent Krzakala, Cristopher Moore, and Lenka Zdeborová. Asymptotic analysis of the stochastic block model for modular networks and its algorithmic applications. *Physical Review E*, 84(6):066106, 2011.

[DKS17] Ilias Diakonikolas, Daniel M Kane, and Alistair Stewart. Statistical query lower bounds for robust estimation of high-dimensional gaussians and gaussian mixtures. In *58th Annual Symposium on Foundations of Computer Science (FOCS)*, pages 73–84. IEEE, 2017.

[FGR+17] Vitaly Feldman, Elena Grigorescu, Lev Reyzin, Santosh S Vempala, and Ying Xiao. Statistical algorithms and a lower bound for detecting planted cliques. *Journal of the ACM (JACM)*, 64(2):1–37, 2017.

[HKP+17] Samuel B Hopkins, Pravesh K Kothari, Aaron Potechin, Prasad Raghavendra, Tselil Schramm, and David Steurer. The power of sum-of-squares for detecting hidden structures. In *58th Annual Symposium on Foundations of Computer Science (FOCS)*, pages 720–731. IEEE, 2017.

[Hop18] Samuel B Hopkins. *Statistical Inference and the Sum of Squares Method.* PhD thesis, Cornell University, 2018.

[HS17] Samuel B Hopkins and David Steurer. Efficient bayesian estimation from few samples: community detection and related problems. In *58th Annual Symposium on Foundations of Computer Science (FOCS)*, pages 379–390. IEEE, 2017.

[HWX15] Bruce Hajek, Yihong Wu, and Jiaming Xu. Computational lower bounds for community detection on random graphs. In *Conference on Learning Theory*, pages 899–928. PMLR, 2015.

[JŁR11] Svante Janson, Tomasz Łuczak, and Andrzej Rucinski. *Random Graphs*, volume 45. John Wiley & Sons, 2011.

[KMOW17] Pravesh K Kothari, Ryuhei Mori, Ryan O'Donnell, and David Witmer. Sum of squares lower bounds for refuting any CSP. In *Proceedings of the 49th Annual ACM SIGACT Symposium on Theory of Computing*, pages 132–145, 2017.

[KWB22] Dmitriy Kunisky, Alexander S Wein, and Afonso S Bandeira. Notes on computational hardness of hypothesis testing: Predictions using the low-degree likelihood ratio. In *ISAAC Congress (International Society for Analysis, its Applications and Computation)*, pages 1–50. Springer, 2022.

[Moo17] Cristopher Moore. The computer science and physics of community detection: Landscapes, phase transitions, and hardness. *Bulletin of EATCS*, 1(121), 2017.

[SW22] Tselil Schramm and Alexander S Wein. Computational barriers to estimation from low-degree polynomials. *The Annals of Statistics*, 50(3):1833–1858, 2022.