COMPUTATIONAL LOWER BOUNDS VIA ALMOST ORTHONORMAL POLYNOMIALS

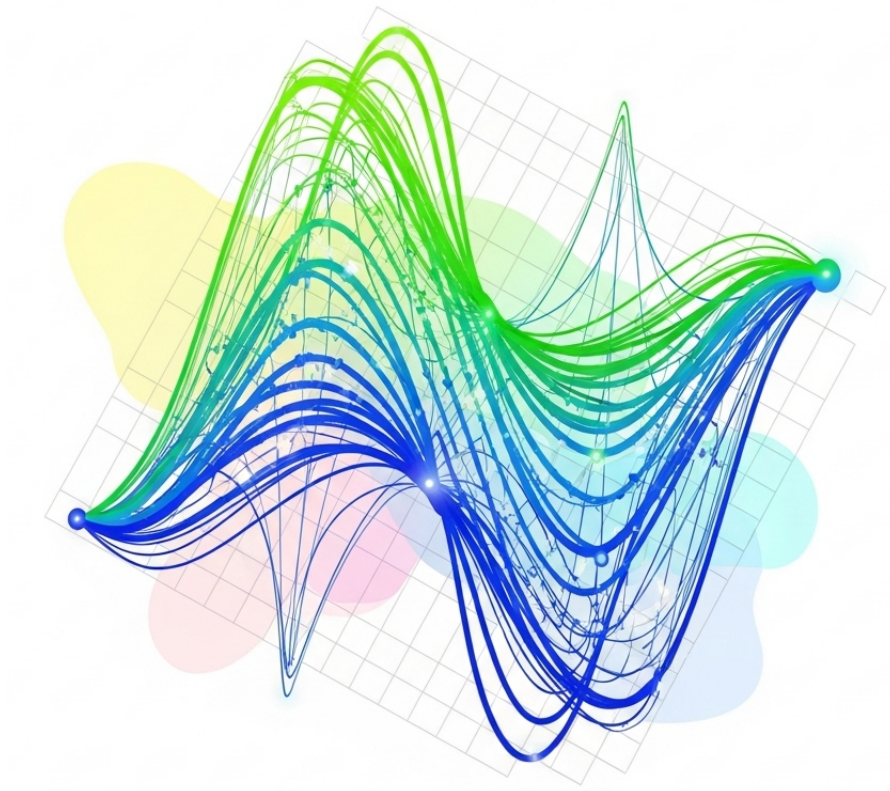
SIMONE MARIA GIANCOLA*†

Supervision: Alexandra Carpentier, Christophe Giraud and Nicolas Verzelen

Last Modified on September 23, 2025

Abstract

Low-degree lower bounds are an established method to claim average-case hardness for algorithms. When the null hypothesis is simple, an explicit orthonormal basis suffices to start the derivation. When it is not, advanced proof methods rely on a complex matrix inversion via recursive relations. In this work, we present a new proof technique to find low-degree lower bounds for any type of null distribution. It relies on finding an "almost orthonormal" basis. Specifically, we exploit symmetries in the distributions and the graphical structure to adjust orthogonal polynomials for simple null hypotheses. With the right fixes, correlations decay fast enough to reproduce the classical proof. We present the steps for a motivating example: the planted sub-matrix model. The observation is a binary matrix such that the entries are more likely to be positive if they belong to a latent clique of vertices of random size. After establishing a low-degree lower bound for this model, we argue that our method works for many others. The progression is adaptive, with sections for readers not familiar with low-degree polynomials and average-case hardness in general.



*Laboratoire de Mathématiques d'Orsay - Université Paris-Saclay, France

†simonegiancola09@gmail.com

Image credit: Gemini; prompt: *Generate an image with the following prompt "Computational lower bounds via almost orthonormal polynomials". Do not put text, and make the background white.*

CONTENTS

List of symbols 4
1 Introduction 5
1.I Contribution 7
1.II Related work 9
2 Low-degree method (for unfamiliar readers) 10
3 Main result 13
3.I Discussion 14
3.II A working representation of the advantage 16
3.III First idea: grouping by invariants, skeleton graphs 17
3.IV Second idea: almost orthonormality 19
4 More details on the skeletons formalism 20
4.I Invariant graph-theoretic objects at the level of skeletons 23
4.II Proof ingredients 25
4.III Key graph properties 30
5 Construction and properties of the almost orthonormal basis 34
5.I An informal discussion about what we want 38
5.II Formalization 39
6 Bounding the advantage: proof of the main theorem 51
6.I Digression: element-wise bound 54
6.II Finalization 55
6.III Adaptation for perturbation on signal size 56
6.IV Final remarks and extensions 56

References 59
A Proofs of the basis and invariance lemmas 62
B Matching bounds 63
B.I Global sum statistic 64
B.II Line-sum 64
B.III Comments on optimality of global sum and line-sum statistics 65
B.IV Formal analysis of global sum statistic under signal strength perturbations 66
C Average-case hardness and statistical-to-computational gaps (for unfamiliar readers) 70

ACKNOWLEDGEMENTS

I am very lucky to have met Christophe Giraud during the "Statistics and probability in high dimension" course at Orsay. The interesting interaction we had and his curiosity led to this work. Thank you for believing in me and including so much expertise in the process. I cannot thank enough also Alexandra Carpentier and Nicolas Verzelen for teaching me with patience and very experienced mathematical eyes. Your skills and culture amaze me. I am certain I will learn a great amount of mathematics in the three years of PhD to come.

The quality of teaching here in France has been surprisingly rewarding due to the exceptional fleet of professors I have met. Thank you for being so obsessed with quality and passing a good message to aspiring mathematicians.

The ten months at Orsay were amazing also thanks to my classmates. It is warming to see that each and every one has a great passion for this field. Thank you for always teaching me something.

To friends outside of school in Paris, I cannot help but mention Anna, Carmine, and Gabriele: I really had a great time with you. We crossed paths in three different ways, but I hope we will stick together in the years that come.

Grazie alla mia famiglia per credere in me, sempre. Grazie per spingermi a fare quello che mi rende calmo e felice, anche se non lo capite. Ai miei genitori, a mio fratello, il nonno, gli zii e cugini vicini o lontani che siano. Grazie a mia nonna che ha lasciato un segno indelebile nella mia vita; dalla cultura del sacrificio, ai principi, all'amore per la musica.

Per finire, grazie agli amici di sempre da Milano. Anche se sono spesso via, siete con me. Ogni volta che torno, mi fate sentire a casa. In particolare, grazie a Gianluca e Marta che in molti frangenti mi hanno suppo(soppo)rtato.

$[n]$	integers from 1 to n
$\lesssim, \gtrsim, \approx$	inequality/equality up to constants
$\lesssim_{\log}, \gtrsim_{\log}, \approx_{\log}$	inequality/equality up to poly-logarithmic factors
$\mathbf{Y}$	observation matrix
$\mathbf{X}$	latent signal
n	observation size
k	signal size
λ	signal strength/magnitude
$\mathrm{H}_0^{\text{noise}}$	pure noise null hypothesis
H_0	null hypothesis
H_1	alternative hypothesis
η	signal strength perturbation
ζ	signal size perturbation
$\boldsymbol{\theta}$	parameter vector
$\mathcal{P}_{\boldsymbol{\theta}}$	parameterized probability distribution
D	degree of a polynomial
$\mathrm{Adv}_{(\leqslant \mathrm{D})}(\mathrm{H}_0, \mathrm{H}_1)$	advantage
π	labelling of vertices
Π_ℓ	injective functions from $[\ell]$ to $[n]$
σ	permutation
$G = (V, E)$	skeleton graph in abstract space
$\pi(G) = (\pi(V), \pi(E))$	labelled graph in observation space
$\psi(\cdot; j), \psi(\cdot; G)$	basis element
$\boldsymbol{\psi}$	basis vector
$\alpha_G, \alpha_i, \boldsymbol{\alpha}$	coefficient of basis element, coefficients vector
$\mathcal{G}_{(\leqslant D)}$	skeletons set
$\mathcal{M}$	set of matchings
$\mathcal{M}_{\mathrm{PM}}$	set of perfect matchings
$\mathcal{M}^\star$	set of star matchings
$G_\triangle = (V_\triangle, E_\triangle)$	symmetric difference graph
#CC	number of connected components in symmetric difference graph
c_{si}	constant in assumptions 3.1 - 3.7
$\#\mathrm{CC}_{\mathsf{pure}}$	number of pure connected components in symmetric difference graph
U	unmatched vertices in symmetric difference graph
$\boldsymbol{M}$	matched vertices in symmetric difference graph
$\boldsymbol{M}_{\mathrm{PM}}$	perfectly matched vertices in symmetric difference graph
$\boldsymbol{M}_{\mathrm{SM}}$	semi-matched vertices in symmetric difference graph
$\mathrm{d}(\cdot, \cdot)$	graph edit distance
$\mathrm{Aut}(G)$	automorphism group of a graph
$\mathcal{M}_{\mathrm{shadow}}(\cdot, \cdot, \cdot)$	set of shadow matchings
$\Pi(\boldsymbol{M})$	pairs of injections that give $\boldsymbol{M}$ as matching
P_G	canonical basis element of definition 2.18
$\widehat{P}_G$	centered basis element
$\overline{P}_G$	corrected basis element
$\nu(G)$	dominant rescaling factor
$\widetilde{P}_G$	corrected rescaled basis
$\boldsymbol{G}$	Gram matrix of the basis

The question of proving negative results for problems in statistics is long-standing, and dual to positive results. Older literature focused mainly on finding impossibility/feasibility theorems for important problems without constraining the space of functions considered. This line of work of information-theoretic lower and upper bounds flourished (Arias-Castro and Verzelen 2014; B. Clarke, J. Clarke, and Yu 2014; Commenges 2015; Duchi 2024; Ebrahimi, Soofi, and Soyer 2010; Kullback 1978; Verzelen and Arias-Castro 2015). Often, the claims argue that the optimal function to solve a given problem is already well-known in statistics, e.g. the maximum-likelihood estimator or the likelihood ratio (Kunisky, Wein, and Afonso S. Bandeira 2019).

This work pertains attacking the same problems but from the point of view of what algorithms cannot achieve. While there are known cases in which the best possible function from information theory coincides with an algorithm (Mondelli and Montanari 2018), there are others where we have a so-called **statistical-to-computational gap** (Afonso S. Bandeira, Perry, and Wein 2018; Kunisky, Wein, and Afonso S. Bandeira 2019; Zdeborová and Krzakala 2016). The phenomenon is widely present and hints at the existence of scenarios in which even the best algorithm cannot cover all instances where the signal (information) rises above the noise (randomness). Gaps arise in many fields and subfields. Notably, problems spanning hypothesis testing (Kunisky 2020), estimation (Even, Giraud, and Verzelen 2024, 2025a,b; Schramm and Wein 2022; Sohn and Wein 2025), refutation (Kothari et al. 2023), and optimization (Huang and Sellke 2025; Wein 2020) share this phenomenology.

The low-degree method is a technique to tackle algorithmic lower bounds and find such gaps. It arose as the "key step" in another procedure: the method of sum-of-squares (Barak, Hopkins, et al. 2016; Hopkins et al. 2017). Later, it matured into a separate field, with independent motivations (Kunisky, Wein, and Afonso S. Bandeira 2019). Let us mention briefly some of its features.
On the positive side, its flexibility made it widely applicable, with plenty of recent works using it in different flavours. Without being exhaustive we mention (Arpino and Venkataramanan 2023; Ding et al. 2023; Even, Giraud, and Verzelen 2024, 2025a; Kothari et al. 2023; Rush et al. 2022) and reroute the reader to the review of Wein (2025a) for a comprehensive summary. Notably, these span hypothesis testing, estimation, refutation, and optimization as above (Wein 2025a,b).

Apart from using it as a workhorse for proofs, other objectives of current research on the method are:

- to refine the construction (Buhai et al. 2025; Kothari et al. 2023; Kunisky 2020);
- to find ways to use it when a certain class of orthogonal polynomials is not explicit (Kunisky 2020; Schramm and Wein 2022; Sohn and Wein 2025).

Concerning the former, the main negative aspect is that the low-degree method relies on conjectured inclusions of class of functions and algorithms, so it is subject to adjustments. See in particular (Hopkins et al. 2017, hyp. 2.1.5) (Wein 2025a, hyp. 3.1), the counterexample of Buhai et al. (2025) and the new formalism of Kunisky (2024a,b). The positive argument is that it appears to work well under proper assumptions.
We are interested in the latter. Let us clarify it further. For a bottom-up justification of the following definitions we reroute non-familiar readers to appendix C. It is a friendly summary of the transition from information-theoretic lower bounds to computational bounds aimed at explaining what is a statistical-to-computational gap and why we would be interested in it.

Suppose we observe a random matrix $\mathbf{Y} \in \{-1,1\}^{n\times n}$ which depends on a latent (i.e. unobserved) random structure $\mathbf{X}$, and we want to extract from $\mathbf{Y}$ knowledge about $\mathbf{X}$. To model it, we say $\mathbf{Y}$ is sampled from an unknown distribution $\mathcal{P}_{\boldsymbol{\theta}}$ depending on a latent parameter $\boldsymbol{\theta} \in \Theta \subseteq \mathbb{R}^K$ which influences the strength/size of the latent random variables (see the discussion in appendix C). From now onwards, we term $\mathbf{Y}$ the **observation**, and $\mathbf{X}$ the **signal** with **structure** $\boldsymbol{\theta}$. In principle, Θ is a huge space, so we can say very little about the latent structure of $\mathbf{Y}$ if we only know $\mathbf{Y} \sim \mathcal{P}_{\boldsymbol{\theta}}$ for some $\boldsymbol{\theta} \in \Theta$. A common situation in statistics is that expert knowledge makes our life a lot easier. In practice we restrict the possible values $\boldsymbol{\theta}$ can take.
Let us then take a very favourable case where we know that $\mathbf{Y}$ was sampled from either of two distributions. From $\{\boldsymbol{\theta} \in \Theta\}$ we move to a pair $(\boldsymbol{\theta}, \boldsymbol{\theta}') \in \Theta \times \Theta$. We define two major types of problem when this simplification happens:

Problem 1.1 (Detection)**.** *Let $\Theta^{\text{noise}} \subset \Theta$ be the subspace of parameters such that $\mathbf{Y} \sim \mathcal{P}_{\boldsymbol{\theta}^{\text{noise}}}$ has no structure for all $\boldsymbol{\theta}^{\text{noise}} \in \Theta^{\text{noise}}$, informally termed "pure noise". Suppose $\boldsymbol{\theta}^{\text{noise}} \in \Theta^{\text{noise}}, \boldsymbol{\theta} \in \Theta$. Let $\mathrm{H}_0^{\text{noise}}$ be the hypothesis that*

$\mathbf{Y} \sim \mathcal{P}_{\boldsymbol{\theta}^{\text{noise}}}$. *Let* H_1 *be the hypothesis that* $\mathbf{Y} \sim \mathcal{P}_{\boldsymbol{\theta}}$ *is sampled from any other distribution. Study the behavior for given* $\boldsymbol{\theta} \in \Theta$ *of the hypothesis test* ***between distributions****:*

$$\mathrm{H}_0^{\text{noise}} : \mathbf{Y} \sim \mathcal{P}_{\boldsymbol{\theta}^{\text{noise}}}, \qquad \mathrm{H}_1 : \mathbf{Y} \sim \mathcal{P}_{\boldsymbol{\theta}}. \tag{1.2}$$

Namely, understand when we can distinguish between a pure noise matrix and a matrix with structure. We implicitly assume that all pure noise distributions in Θ^{noise} *are indistinguishable in the sense that they induce the same distribution on* $\mathbf{Y}$*. The particular* $\boldsymbol{\theta}^{\text{noise}}$ *does not influence this problem.*

Problem 1.3 (Complex detection)**.** *Let* $\mathcal{P}_{\boldsymbol{\theta}}, \mathcal{P}_{\boldsymbol{\theta}'}$ *be distributions of* $\mathbf{Y}$*. Study the behavior for given* $(\boldsymbol{\theta}, \boldsymbol{\theta}') \in \Theta^{\times 2}$ *with* $\theta_j \neq \theta'_j$ *for only one index* $j \in [K]$ *of the hypothesis test* ***between distributions****:*

$$\mathrm{H}_0 : \mathbf{Y} \sim \mathcal{P}_{\boldsymbol{\theta}}, \qquad \mathrm{H}_1 : \mathbf{Y} \sim \mathcal{P}_{\boldsymbol{\theta}'}. \tag{1.4}$$

Namely, understand when we can distinguish two structured distributions, since the other cases collapse to problem 1.1.

In the even more favourable scenario, we know $\boldsymbol{\theta}$, and just need to infer information about the latent structure. This corresponds to an additional simplification to $\mathbf{Y} \sim \mathcal{P}_{\boldsymbol{\theta}}$, which leaves us with only one question:

Problem 1.5 (Estimation, or recovery)**.** *Assume* $\mathbf{Y} \sim \mathcal{P}_{\boldsymbol{\theta}}$ *for some* $\boldsymbol{\theta} \in \Theta$*. Study for given* $\boldsymbol{\theta}$ *the behavior of an estimator of the latent signal* $\mathbf{X}$ *in* $\mathbf{Y}$*.*

Problems 1.1 - 1.3 are hypotheses tests. In particular, problem 1.3 is an example of "complex" testing. The common approach is to build a test function $\mathrm{t}(\cdot)$ and threshold it properly. For problem 1.5 we want an estimator. There are many ways to write uniform bounds for success/failure of classes of functions for these problems (see appendix C).

In this document, we **focus on impossibility results for complex testing between distributions with algorithms taking a maximum time to compute**, i.e. problem 1.3, with the constraint that we allow test functions to take a maximum computation time. Namely, we seek regions of Θ were no procedure up to some computational time can solve with small error probability problem 1.3. We focus on a toy model (eqn. 1.6) to present clearly the proof technique. Later in subsection 6.IV we discuss a large class of models with respective assumptions that enjoy an analogous result.

We **apply the low-degree method**, a technique to show, at least conjecturally, that computational time-constrained algorithms cannot solve problems 1.1 - 1.3 - 1.5 in prescribed regions of Θ. In section 2, we explain why it boils down to upper bounding a quantity termed advantage (definition 1.10), which depends on the hypotheses $\mathrm{H}_0, \mathrm{H}_1$, thus implicitly on the parameters $(\boldsymbol{\theta}, \boldsymbol{\theta}')$. The dependency on the distributions considered is natural, and led earlier works (see the review of Wein (2025a)) to decompose quantities inside the expectations according in the Hilbert space of the null distribution (eqn. 1.13). Usually, this is easy for problem 1.1, but **way harder** for complex testing and estimation, i.e. problem 1.3, or problem 1.5 when we take the "null" to be the only distribution $\mathcal{P}_{\boldsymbol{\theta}}$. In practice, the main problem in these last two cases is the presence of a latent structure.

We overcome this issue with a **new proof technique**.

DOCUMENT STRUCTURE In the remaining part of this section we present our contribution (subsec. 1.I) and discuss related work in subsection 1.II. We then present the low-degree method for non-experts (sec. 2), and define the objects pertinent to it. In particular, this section is intended for readers that are not familiar with the subject. Our main results and assumptions are in section 3. We also discuss a specific representation of the object we bound (subsec. 3.II), and our two main ideas in subsections 3.III - 3.IV. One crucial notion we will introduce is that of skeleton graphs: in section 4 we clarify the construction and speak at large of it. From the skeletons, we reach a formulation of an "almost orthonormal" basis in section 5. In simple words, it is a basis satisfying a key property we present in definition 1.14. Section 6 concludes with the proof of the main theorem: an upper bound on the performance of polynomials at varying degree. As a bonus, subsection 6.IV briefly sketches in which sense the proof technique extends to many more models almost directly.

Appendix A reports the proofs of the lemmas of sections 2 - 3. Appendix B shows a matching lower bound of our negative result. It clarifies an argument we make in section 3, in particular remarks 3.5 - 3.10. Appendix C is an introduction to information-theoretic and algorithmic lower bounds in average-case hardness. In particular, we formalize the notion of statistical-to-computational gap.

NOTATION Most of the symbols are standard. We use the shorthand $[n] := \{1, \dots, n\}$. To denote inequalities/equalities up to constants we write $\gtrsim, \lesssim, \approx$ and $\lesssim_{\log}, \gtrsim_{\log}, \approx_{\log}$ for relations that hold up to

poly-logarithmic factors in the sample size n. The letters $c, C, c_\cdot, c^\cdot, C_\cdot, C^\cdot$ always mean constants. Since $\mathrm{H}_0^{\text{noise}}, \mathrm{H}_0, \mathrm{H}_1, \mathrm{H}$ are hypotheses of distributions depending on $\boldsymbol{\theta}$, when we write $\mathbb{E}_{\mathrm{H}}[\cdot], \mathbb{P}_{\mathrm{H}}$ we mean the expectation (resp. probability) with respect to the distribution considered.
The only difference we make is between what is random and what is *not*, what is scalar, what is vectorial and what is matricial. For example, $a, b, c, x, y, z, \alpha, \beta, \gamma$ is a scalar, while $\mathrm{a}, \mathrm{b}, \mathrm{c}, \mathrm{x}, \mathrm{y}, \mathrm{z}, \upalpha, \upbeta, \upgamma$ is a random scalar. Similarly, $\boldsymbol{a}, \boldsymbol{b}, \boldsymbol{c}, \boldsymbol{x}, \boldsymbol{y}, \boldsymbol{z}, \boldsymbol{\alpha}, \boldsymbol{\beta}, \boldsymbol{\gamma}$ is a vector; $\mathbf{a}, \mathbf{b}, \mathbf{c}, \mathbf{x}, \mathbf{y}, \mathbf{z}, \boldsymbol{\upalpha}, \boldsymbol{\upbeta}, \boldsymbol{\upgamma}$ is a random vector. Again, $\boldsymbol{A}, \boldsymbol{B}, \boldsymbol{C}, \boldsymbol{X}, \boldsymbol{Y}, \boldsymbol{Z}, \boldsymbol{\Lambda}, \boldsymbol{\Psi}, \boldsymbol{\Theta}$ is a matrix; $\mathbf{A}, \mathbf{B}, \mathbf{C}, \mathbf{X}, \mathbf{Y}, \mathbf{Z}, \boldsymbol{\Lambda}, \boldsymbol{\Psi}, \boldsymbol{\Theta}$ is a random matrix. An expectation such as $\mathbb{E}_{\mathrm{x}}[\mathrm{x}yz] = \int xyz \, \mathrm{d}\mathbb{P}[x]$ is such that y is deterministic, and we integrate out against x which is deterministic once it is expressed inside an integral, keeping z random throughout.
We denote graphs $G = (V, E)$ via edges and vertices in a certain abstract space. Graphs in the space of the observation are always labelled through a given injection π from the abstract space to the observation space. The graph notation is independent of the rest, and we use $\boldsymbol{A}$ for sets of pairs related to graph theory objects. All the other objects are explicit, and graphs are never random in this work.

1.1 *Contribution*

In this document we propose a new method to prove low-degree lower bounds. We consider as **toy example** an instance of the planted sub-matrix model:

$$\mathbf{Y} \in \{-1, 1\}^{n \times n}, \text{ such that for all } i, j \in [n] \qquad \mathrm{Y}_{ij} = \begin{cases} 1 & \text{with probability } \frac{1+\mathbf{x}_{ij}}{2} \\ -1 & \text{with probability } \frac{1-\mathbf{x}_{ij}}{2} \end{cases}, \qquad \mathrm{X}_{ij} = \mathrm{x}_i \mathrm{x}_j, \tag{1.6}$$

where $\mathbf{x} \in \{0, \sqrt{\lambda}\}^n$ is a binary vector such that:

$$\mathrm{x}_i \overset{\text{i.i.d.}}{\sim} \sqrt{\lambda}\mathrm{Ber}\left(\frac{k}{n}\right), \qquad \lambda \in [0, 1], \ k \in \mathbb{N}. \tag{1.7}$$

In words: there is a signal $\mathbf{x}$ that acts on the observation as $\mathbf{X} = \mathbf{x}\mathbf{x}^\top$ and increases the probability of $\mathrm{Y}_{ij} = 1$ when $\mathrm{x}_i = \mathrm{x}_j = \sqrt{\lambda}$. The matrix $\mathbf{X} = \mathbf{x}\mathbf{x}^\top$ is the latent signal with structure $\boldsymbol{\theta}$. It is a shift of the classical Erdös-Rényi random graph with a planted clique of random size. There are many ways to study versions of this model with the low-degree method and related approaches (Alon, Krivelevich, and Sudakov 1998; Brennan and Bresler 2020; Gamarnik, Moore, and Zdeborová 2022; Hopkins et al. 2017; Kothari et al. 2023).

For given $(k, \lambda) =: \boldsymbol{\theta} \in \Theta := \mathbb{N} \times [0, 1]$ the planted sub-matrix model is a distribution $\mathcal{P}_{\boldsymbol{\theta}}$. We can study problems 1.1 - 1.3 - 1.5 in the language of statistics.

Definition 1.8 (Pure noise parameters of planted sub-matrix)**.** *For the model of equation 1.6 we define the subspace of pure noise parameters* $\Theta^{\text{noise}} := \{(k, \lambda) \mid k\lambda = 0\}$. *Notice that for all* $\boldsymbol{\theta}^{\text{noise}} \in \Theta^{\text{noise}}$ *it holds that* $\mathbf{X}$ *is a null matrix almost surely and* $\mathrm{Y}_{ij} \overset{\text{i.i.d.}}{\sim} \mathrm{Rad}(1/2)$ *for all* $i, j \in [n]$. *All pure noise parameters induce the same distribution on* $\mathbf{Y}$ *and the definition is in accordance with problem 1.1.*

Remark 1.9 (Problems 1.1 - 1.3 - 1.5 for the planted sub-matrix model)**.** *The "pure noise" scenario of definition 1.8 is such that* $\mathrm{Y}_{ij} \overset{\text{i.i.d.}}{\sim} \mathrm{Rad}(1/2)$ *for all* $i, j \in [n]$ *for all* $\boldsymbol{\theta}^{\text{noise}} \in \Theta^{\text{noise}}$. *In accordance with definition 1.8 and problem 1.1 all observations* $\mathbf{Y} \sim \mathrm{H}(\boldsymbol{\theta}^{\text{noise}})$ *are indistinguishable, in the sense that they induce the same distribution and there is no difference in doing hypothesis test with one or another.*
The complex test of problem 1.3 is of two major types:

- *we perturb the signal strength* $\boldsymbol{\theta} = (k, \lambda)$ *to* $\boldsymbol{\theta}' = (k, \lambda + \eta)$ *for some* $\lambda > 0, \eta > 0$;
- *we perturb the signal size* $\boldsymbol{\theta} = (k, \lambda)$ *to* $\boldsymbol{\theta}' = (k + \zeta, \lambda)$ *for some* $k > 0, \zeta > 0$.

The case where n changes between the two distributions is easy to detect as we observe an n-dimensional matrix, so it is trivial.
Estimation in problem 1.5 boils down to estimating for a given error measure the support of the clique, i.e. the sites j such that $\mathrm{x}_j = \sqrt{\lambda}$ *where* $\mathbf{Y} \sim \mathrm{H}(\boldsymbol{\theta})$ *for some* $\boldsymbol{\theta} \in \Theta \backslash \Theta^{\text{noise}}$.

As we said, we focus on complex hypothesis testing, the immediate example being testing between two non-zero values of λ or k above. We call this **complex** testing because we want to distinguish distributions both having signal (see problem 1.3). We defined problems 1.1 - 1.5 because non-triviality of complex testing

depends on them (see remark 3.6) and our technique extends to these two, albeit being superfluous for detection.

While the model is quite standard and we recover known results, the importance of the work is the **explanation in a controlled setting of the proof technique**, which applies in far more instances (see subsec. 6.IV). Due to its direct nature, we hope that this method will spark finer results and more explicit derivations of low-degree lower bounds.

The main quantity in low-degree lower bounds is the advantage. We define it below.

Definition 1.10 (Advantage)**.** *Let* $\mathrm{H}_0, \mathrm{H}_1$ *be hypotheses of two probability distributions. We define the advantage as:*

$$\mathrm{Adv}_{(\leqslant D)}(\mathrm{H}_0, \mathrm{H}_1) := \sup_{\substack{f:\{-1,1\}^{n\times n}\mapsto\mathbb{R} \\ f:\deg(f)\leqslant D}} \frac{\mathbb{E}_{\mathrm{H}_1}[f(\mathbf{Y})]}{\sqrt{\mathbb{E}_{\mathrm{H}_0}\left[f^2(\mathbf{Y})\right]}}, \tag{1.11}$$

where the supremum is taken over polynomials.

Remark 1.12. *The adaptation for our three problems is immediate; the distributions are parameterized, and the advantage depends on the parameters implicitly.*
For detection (prob. 1.1) we take $\mathcal{P}_{\boldsymbol{\theta}^{\mathrm{noise}}}$ *the distribution of* $\mathrm{H}_0^{\mathrm{noise}}$*, for some* $\boldsymbol{\theta}^{\mathrm{noise}}$ *and* $\mathcal{P}_{\boldsymbol{\theta}}$ *the distributions of* H_1 *for some* $\boldsymbol{\theta} \in \Theta$*.*
For complex testing (prob. 1.3), we consider $\mathcal{P}_{\boldsymbol{\theta}}$ *the distribution of* H_0 *and* $\mathcal{P}_{\boldsymbol{\theta}'}$ *the distribution of* H_1 *for a pair* $(\boldsymbol{\theta}, \boldsymbol{\theta}') \in \Theta^{\times 2}$*.*
In our paper, all hypotheses correspond to a distribution so it is unambiguous.
In the case of estimation (prob. 1.5), at a given $\boldsymbol{\theta} \in \Theta$*, there is an analog notion of advantage where we take integrals with respect to* $\mathcal{P}_{\boldsymbol{\theta}}$ *at the numerator and denominator (see (Schramm and Wein 2022; Sohn and Wein 2025)).*

We seek an upper bound on this quantity because it quantifies how polynomials of degree up to D are able to make the mean under one distribution large with respect to the fluctuations under another. For the special case of estimation it is a matter of making the mean large with respect to fluctuations under the same distribution, like in the second moment method. For a full derivation in complex testing and detection see section 2.
A function that makes the objective of the advantage large is said to *separate* (in the weak or strong sense, see definitions 2.10 - 2.12). Therefore, if we show that no polynomial can separate, we claim the problem is hard to study up to polynomials of some degree, and extrapolate conjecturally that it is the same for algorithms taking a time that depends on D (see sec. 2). In particular, polynomial-time algorithms correspond in the low-degree formalism to $D \approx_{\log} \log n$.

The simplest way to upper bound the advantage is to *upper bound* the numerator and *lower bound* the denominator. The former is easy (it is linear), so let us focus on the latter. Interpreting integrals under H_0 as a Hilbert space with associated inner product $\langle f, g\rangle_{\mathrm{H}_0} := \mathbb{E}_{\mathrm{H}_0}[fg]$ the advantage turns into an optimization problem over the coefficients of the expansion:

$$\mathrm{Adv}_{(\leqslant D)}(\mathrm{H}_0, \mathrm{H}_1) = \sup_{\boldsymbol{\alpha}} \frac{\mathbb{E}_{\mathrm{H}_1}\left[\sum_{j\in\mathrm{basis}} \alpha_j \psi(\mathbf{Y}; j)\right]}{\sqrt{\sum_{i,j\in\mathrm{basis}} \mathbb{E}_{\mathrm{H}_0}\left[\alpha_i\alpha_j \psi(\mathbf{Y}; i)\psi(\mathbf{Y}; j)\right]}}, \tag{1.13}$$

for $(\psi(\mathbf{Y}; j))_{j\in\mathrm{basis}}$ a basis of polynomials of degree less than D. For simple null hypotheses such as in $\mathrm{H}_0^{\mathrm{noise}}$ from problem 1.1 there are known nice orthonormal bases. When this happens, the denominator is just $\|\boldsymbol{\alpha}\|_2$ and the expression greatly simplifies. Much of the recent work on low-degree polynomials aims to overcome the issues of generic distributions in H_0. The simplest example of null hypothesis that is not trivial is when $\mathrm{H}_0 \equiv \mathrm{H}_0(\boldsymbol{\theta})$ is a distribution with $\lambda \neq 0$ and $k \neq 0$ as in problem 1.3: it has itself a signal, and we want to distinguish it from H_1 which has larger signal. We call it a **complex** testing problem (see prob. 1.3).

THIS WORK We will build an *almost* orthonormal basis of polynomials for complex testing. We define this property as:

Definition 1.14 (Almost orthonormal basis)**.** *A basis* $(\psi(\cdot; j))_{j\in\mathrm{basis}}$ *for a Hilbert space induced by a distribution* $\mathbb{Q}$ *is almost orthonormal when it satisfies the following inequalities up to constants:*

$$\|\boldsymbol{\alpha}\|_2 \lesssim \sqrt{\sum_{i,j} \mathbb{E}_{\mathbb{Q}}\left[\alpha_i\alpha_j \psi(\mathbf{Y}; i)\psi(\mathbf{Y}; j)\right]} =: \|\boldsymbol{\alpha}\|_{\mathbb{E}_{\mathbb{Q}}[\boldsymbol{\psi}\boldsymbol{\psi}^\top]} \lesssim \|\boldsymbol{\alpha}\|_2. \tag{1.15}$$

In particular, for complex testing (problem 1.3) we have $\mathbb{Q} = \mathcal{P}_{\boldsymbol{\theta}}$ the distribution of $\mathrm{H}_0 \equiv \mathrm{H}_0(\boldsymbol{\theta})$, and are interested in cases $\boldsymbol{\theta} \notin \Theta^{\text{noise}}$.
For estimation (problem 1.5) we seek an almost orthonormal basis with respect to the distribution $\mathbb{Q} = \mathcal{P}_{\boldsymbol{\theta}}$ which is at the numerator and denominator of the advantage (def. 1.10).

Once we have an almost orthonormal basis the technique to bound the advantage degrades to models where the distribution of the null hypothesis H_0 admits an **explicit** orthonormal basis. The **key steps** in building it rely on the symmetry and conditional independence of the problem. If it exists, the main theorem (thm. 3.8) follows smoothly as the simplification is very deep.

1.II *Related work*

There are two axes of similarity with literature: (i) finding computational lower bounds for non-trivial null hypotheses, and (ii) exploiting the invariance of the distribution we consider in problems 1.1 - 1.3.
Concerning the former, to the best of our knowledge, the oldest work deriving algorithmic lower bounds with the low-degree method under non-simple null hypotheses is (Kunisky 2020). In particular, Kunisky (2020) is able to go beyond the assumption of a Gaussian null hypothesis by leveraging results on exponential families of random variables. These enjoy explicit orthonormal bases. Later, the work of Schramm and Wein (2022) unveiled an implicit way to upper bound the advantage (def. 1.10) with a recursive construction. The positive aspect of the technique is that it is very general. The price to pay is that the method is rather complex. Many works followed their paradigm (see the discussion in (Wein 2025a)). A sharpening of (Schramm and Wein 2022) also recently appeared as a preprint (Sohn and Wein 2025). However, the improved precision is again traded-off with a complex derivation. A peculiar aspect of this technique is the appearance of *cumulants* in the inequalities (see (Schramm and Wein 2022, thm. 2.2, rem. 2.3)), which have no clear justification. Properties of such cumulants are crucial for obtaining the bounds (Even, Giraud, and Verzelen 2024, 2025a,b).
Invariance in statistics has instead a long history. The seminal books of Lehmann (1970) and Lehmann and Casella (1998) study situations in which invariant distributions enjoy nice properties, especially for hypothesis testing. The conjugation of these ideas with the low-degree method already started, with works largely overlapping with physics ideas (Kunisky, Moore, and Wein 2024; Montanari and Wein 2022; Semerjian 2024). The fact that physics comes up simultaneously with symmetry is not a surprise, and motivates us to discuss alternatives to the low-degree method.

OTHER CONNECTIONS While the low-degree method originated from the computer science/statistics community, many formalisms arose in other fields. The idea is always the same; one seeks to restrict the class of functions to a class of *computable* functions believed to include algorithms, or suggest their uselessness in proper regimes. Among these, we find (non-exhaustively): the sum-of-squares hierarchy of relaxations itself (Barak and Steurer 2014), proofs of average-case reductions (Brennan and Bresler 2020), techniques to show that Markov chain Monte Carlo algorithms fail (Arous, Wein, and Zadik 2020; Jerrum 1992), the statistical query approach (Feldman 2017; Reyzin 2020; Steinhardt 2016; Szörényi 2009), and plenty of methods from statistical physics (Afonso S Bandeira and Alaoui 2022; Barbier 2024; Barbier et al. 2024; Gamarnik, Moore, and Zdeborová 2022; Gamarnik and Zadik 2019; Zdeborová and Krzakala 2016). We comment some below.

In nice cases, predictions match across methods. In general, the low-degree pathway is the widest (Wein 2025a,b). Motivated by these facts, an exciting task that arose is classifying or showing equivalence between any of them.
This ambitious direction has reached strong but narrow results. In certain classes of models, the low-degree method may be "equivalent" in some proper sense to: the best known algorithm from statistical physics (termed AMP, for approximate message passing) (Montanari and Wein 2022), an established method (termed Franz-Parisi, after the authors) to prove geometric impediments in the complexity landscape (Afonso S Bandeira and Alaoui 2022; Chen et al. 2025), or other types of so-called "free energy barrier" results such as the overlap-gap property (Barbier et al. 2024; Gamarnik, Moore, and Zdeborová 2022; Gamarnik and Zadik 2019, 2022; R and Kızıldağ 2025). Notably, there are also plenty of connections within fields that do not include the low-degree method under proper assumptions. For a very comprehensive overview, we reroute the reader to the presentation of Wein (2025b) and the review of Wein (2025a).
For further arguments, we suggest consulting (Kunisky, Wein, and Afonso S. Bandeira 2019; Wein 2025a) and the references therein. It is also useful to compare with the complementary physics side of the literature, well exposed in (Afonso S. Bandeira, Perry, and Wein 2018; Zdeborová and Krzakala 2016).

In this section we review the low-degree method for hypothesis testing (e.g. probs. 1.1 - 1.3) with an eye towards readers that are not experts of the literature. Coming back to the beginning of section 1, it is one of the most used computational models to fix the issue of older unconstrained bounds. We postpone the broader context of these unconstrained bounds and how these are linked to algorithmic bounds to appendix C.

The low-degree method is a restriction to the computational class of polynomials of low degree. Following Wein (2025a, hyp. 3.1), it means that we roughly believe/conjecture the following inclusions:

$$\text{“Degree } O(1) \text{ polynomials} \subseteq \text{polynomial-time algorithms} \subseteq \text{degree } O(\log n) \text{ polynomials”}.$$

In full generality, if we wish to extend the analogy we believe in the inclusions of Hopkins et al. (2017, hyp. 2.1.5) and Wein (2025a, hyp. 3.2):

$$\text{“exp}\left\{D/\log^C n\right\}\text{- time algorithms} \subseteq \text{degree } D \text{ polynomials} \subseteq n^{O(D)}\text{-time algorithms”},$$

for some large C and $D \equiv D(n)$ increasing with n. While these have to be taken with care, we can in principle study the behavior of the best performance over type I and type II errors for functions of a given degree, and then come back later to the subtleties. In general, it makes sense to say that if polynomials of a very large degree cannot solve a problem, then no algorithm up to some runtime can. Similarly, if a polynomial of low degree can solve a problem, we expect to write down an algorithm that computes the polynomial and solves it. For larger justifications and issues, we reroute the reader to subsection 1.II where we provide robust references.

Remark 2.1. *We state our definitions in the context of our planted sub-matrix model for problems 1.1 - 1.3, but they are fairly more general and adapt to any sequence of hypotheses $((\mathrm{H}_0^{(n)}, \mathrm{H}_1^{(n)}))_{n\in\mathbb{N}}$ with associated distributions $((\mathcal{P}_{\boldsymbol{\theta}(n)}, \mathcal{P}_{\boldsymbol{\theta}'(n)}))_{n\in\mathbb{N}}$ of random variables taking values in some N-dimensional space where $N \equiv N_n$. For our little model, we have $N = \binom{n}{2}$ if we observe only half of the matrix $\mathbf{Y}$ and ignore the diagonal, or $N = n^2$ if we observe all of it, and $\mathbf{Y}$ lives in a hypercube. For tensors, we would have $N = n^p$ and $\mathbf{Y}$ tensorial in some field for example.*

Remark 2.2. *We state our definitions asymptotically for simplicity. One could adapt the whole explanation to a non-asymptotic setting for the restricted computational class. The ways in which this is feasible or not are problem-dependent.*

We formulate two notions of success for a test as in other works (Kunisky, Wein, and Afonso S. Bandeira 2019; Wein 2025a). The analogues in information theory and the algorithmic counterpart are in the appendix (equations C.3 - C.4 - C.6 - C.7).

Definition 2.3 (Strong detection)**.** *For a hypothesis test $(\mathrm{H}_0, \mathrm{H}_1)$ between distributions with given parameters sequences $\boldsymbol{\theta} \equiv \boldsymbol{\theta}(n), \boldsymbol{\theta}' \equiv \boldsymbol{\theta}'(n)$, we say a test $\mathsf{t} : \{-1,1\}^{n\times n} \mapsto \{0,1\}$ performs strong detection if:*

$$\mathbb{P}_{\mathrm{H}_0}\left[\mathsf{t}(\mathbf{Y}) = 1\right] + \mathbb{P}_{\mathrm{H}_1}\left[\mathsf{t}(\mathbf{Y}) = 0\right] = o\left(1\right), \qquad \text{as } n \to \infty. \tag{2.4}$$

In words: the function performs asymptotically no errors.

Definition 2.5 (Weak detection)**.** *For a hypothesis test $(\mathrm{H}_0, \mathrm{H}_1)$ between distributions with given parameters sequences $\boldsymbol{\theta} \equiv \boldsymbol{\theta}(n), \boldsymbol{\theta}' \equiv \boldsymbol{\theta}'(n)$, we say a test $\mathsf{t} : \{-1,1\}^{n\times n} \mapsto \{0,1\}$ performs weak detection if:*

$$\mathbb{P}_{\mathrm{H}_0}\left[\mathsf{t}(\mathbf{Y}) = 1\right] + \mathbb{P}_{\mathrm{H}_1}\left[\mathsf{t}(\mathbf{Y}) = 0\right] \leqslant 1 - \Omega\left(1\right), \qquad \text{as } n \to \infty. \tag{2.6}$$

In words: the function beats random guessing by a non-negligible margin.

In particular, these two are different notions of a positive result, and we can adapt them to degree D polynomials by adding proper constraints.

Example 2.7. *Consider the following criterions that are a study of type I and type II errors restricted to thresholding polynomials:*

$$\inf_{\substack{f:\{-1,1\}^{n\times n}\mapsto\{0,1\}\\ f:\deg(f)\leqslant D}} \inf_{\xi} \left\{\mathbb{P}_{\mathrm{H}_0}\left[\mathbb{1}_{f(\mathbf{Y})\geqslant\xi} = 1\right] + \mathbb{P}_{\mathrm{H}_1}\left[\mathbb{1}_{f(\mathbf{Y})\geqslant\xi} = 0\right]\right\} \geqslant 1 - o\left(1\right), \qquad \forall \boldsymbol{\theta} \in \Theta_{(\leqslant D)\,\mathsf{imp}}, \tag{2.8}$$

$$\inf_{\substack{f:\{-1,1\}^{n\times n}\mapsto\{0,1\}\\ f:\deg(f)\leqslant D}} \inf_{\xi} \left\{\mathbb{P}_{\mathrm{H}_0}\left[\mathbb{1}_{f(\mathbf{Y})\geqslant\xi} = 1\right] + \mathbb{P}_{\mathrm{H}_1}\left[\mathbb{1}_{f(\mathbf{Y})\geqslant\xi} = 0\right]\right\} = o\left(1\right), \qquad \forall \boldsymbol{\theta} \notin \Theta_{(\leqslant D)\,\mathsf{imp}}, \tag{2.9}$$

where in words we mean that we take the best ξ thresholding each f attaining minimal type I and type II errors. We say that:

- *if we are in the first scenario, then there is no weakly or strongly detecting degree D test, but there might be higher degree computable tests;*
- *if we are in the second scenario, we know there is a degree D test that performs strong detection, hence also weak detection. Moreover, since there are degree-D tests, the same happens if we remove the constraint.*

In particular, we hope that the region of impossibility of polynomials roughly coincides with the region of impossibility of algorithms with some runtime D', i.e. that $\Theta_{(\leqslant D)\,\mathsf{imp}} \approx \Theta_{(\leqslant D')\mathsf{alg}\,\mathsf{imp}}$. For $D \approx_{\log} \log n$ we expect that we capture polynomial time algorithms. In any case, we are interested in both the weak and strong sense of having low type I plus type II error.

It turns out that showing this directly is not nice. The quickest simplification is using a looser criterion, that of separability. There are two versions of it. The former is the strong version:

Definition 2.10 (Strong separation)**.** *A function $f : \{-1,1\}^{n\times n} \mapsto \mathbb{R}$ strongly separates a hypothesis test $(\mathrm{H}_0, \mathrm{H}_1)$ between parametric distributions for given parameter sequences $\boldsymbol{\theta} \equiv \boldsymbol{\theta}(n), \boldsymbol{\theta}' \equiv \boldsymbol{\theta}'(n)$ if:*

$$\max\left\{\mathrm{Var}_{\mathrm{H}_0}[f], \mathrm{Var}_{\mathrm{H}_1}[f]\right\} = o\left(\left|\mathbb{E}_{\mathrm{H}_0}[f] - \mathbb{E}_{\mathrm{H}_1}[f]\right|\right). \tag{2.11}$$

The latter relaxes the vanishing requirement to just boundedness.

Definition 2.12 (Weak separation)**.** *A function $f : \{-1,1\}^{n\times n} \mapsto \mathbb{R}$ weakly separates a hypothesis test $(\mathrm{H}_0, \mathrm{H}_1)$ between parametric distributions for given parameter sequences $\boldsymbol{\theta} \equiv \boldsymbol{\theta}(n), \boldsymbol{\theta}' \equiv \boldsymbol{\theta}'(n)$ if:*

$$\max\left\{\mathrm{Var}_{\mathrm{H}_0}[f], \mathrm{Var}_{\mathrm{H}_1}[f]\right\} = O\left(\left|\mathbb{E}_{\mathrm{H}_0}[f] - \mathbb{E}_{\mathrm{H}_1}[f]\right|\right). \tag{2.13}$$

To clear matters our, we argue that separability is a sufficient condition for detection.

Lemma 2.14. *If there exists a weakly (strongly) separating function then there exists a weakly (strongly) detecting test. Such test thresholds the separating function.*

Proof. Follows by an application of Chebyshev's inequality. □

Having found a sufficient condition, we just work towards a nicer formulation for establishing bounds. We want to compare quantities, so it is natural to consider their ratio:

$$\frac{\left|\mathbb{E}_{\mathrm{H}_0}[f] - \mathbb{E}_{\mathrm{H}_1}[f]\right|}{\max\left\{\mathrm{Var}_{\mathrm{H}_0}[f], \mathrm{Var}_{\mathrm{H}_1}[f]\right\}}. \tag{2.15}$$

To simplify, we recenter the null to have $\mathbb{E}_{\mathrm{H}_0}[f] = 0$ at no loss, and lower bound the maximum in the denominator.[1] In doing these, we get to a larger quantity: $\mathbb{E}_{\mathrm{H}_1}[f]/\sqrt{\mathbb{E}_{\mathrm{H}_0}[f^2]}$. It is some L^2-looking criterion to compare the variance in the null and the first moment in the alternative. For a given test, if it is large in some proper sense we know we are either weakly or strongly separating distributions in the sense of definitions 2.12 - 2.10. Moreover, by lemma 2.14, we have the detection analogues and conclude that the problem is solvable in the asymptotic regimes. With this machinery, we can show the positive result that in a given region there is a function (possibly a polynomial of some degree) that attains low type I and type II error jointly.

EXTRAPOLATION To prove negative results we just "flip the sock". If we show that no polynomial up to some degree performs strong (weak) detection, then we are **drawn to believe** that there are no strongly (weakly) detecting tests. We get to the workhorse of papers in the low-degree method: the advantage of definition 1.10. We rewrite it below in the form of problem 1.3 for convenience.

$$\mathrm{Adv}_{(\leqslant D)}(\mathrm{H}_0, \mathrm{H}_1) = \sup_{\substack{f:\{-1,1\}^{n\times n}\mapsto\mathbb{R} \\ \deg(f)\leqslant D}} \frac{\mathbb{E}_{\mathrm{H}_1}[f(\mathbf{Y})]}{\sqrt{\mathbb{E}_{\mathrm{H}_0}\left[f^2(\mathbf{Y})\right]}}. \tag{2.16}$$

We are interested in three behaviors of the advantage, depending on $\boldsymbol{\theta} \equiv \boldsymbol{\theta}(n)$ and $D \equiv D(n)$:

1. when it is unbounded, and so there exists a strongly separating function;

[1] By lower bounding the maximum we lose something, and might recover suboptimal results. In these cases, a certain "conditional" low-degree method solves the issue (see the discussion at the end of (Kunisky, Wein, and Afonso S. Bandeira 2019, sec. 1.2)).

2. when it is bounded but not vanishing, $O(1)$, in which case there is no strongly separating function, but there is a weakly separating function;

3. when it is $1 + o(1)$, case in which there is not even a weakly separating function.

Using definitions 2.3 - 2.5, we will take case #1 as evidence of easiness of the problem in the strong and weak sense, case #2 as evidence of hardness in the strong sense, and easiness in the weak sense, case #3 as evidence of hardness in both senses.

Eventually, the best-case scenario of the low-degree method is a combination of the following steps, largely dependent on the conjectured inclusions at the beginning of this section:

1. showing that for all given $\boldsymbol{\theta} \in \Theta_{(\leqslant D)\,\mathsf{imp}}$ the advantage is vanishing (or bounded);

2. showing that for all $\boldsymbol{\theta} \notin \Theta \backslash \Theta_{(\leqslant D)\,\mathsf{imp}}$ the advantage is bounded but not vanishing (or diverging);

3. in a finer way, showing that for all $\boldsymbol{\theta} \notin \Theta \backslash \Theta_{(\leqslant D)\,\mathsf{imp}}$ there exists an algorithm attaining strong (or weak) detection when $D \approx_{\log} \log n$;

4. even better, showing that for all $\boldsymbol{\theta} \notin \Theta \backslash \Theta_{(\leqslant D)\,\mathsf{imp}}$ there exists an algorithm that takes $\exp\{D/\log^C n\}$ time to perform strong (or weak) detection at varying D.

INTUITION To tackle steps #2, #3, #4 we take inspiration from step #1, where we prove an upper bound on the advantage. Informally, the hardest object to bound at given D should hint at the best-performing degree D polynomial. When $D \approx_{\log} \log n$, this gives an indication of the best-performing poly-time algorithm, which is expected to be robust to a slight perturbation from $\Theta_{(\leqslant D)\,\mathsf{imp}}$ to its complement, and hence to "start working" just outside the impossible phase. Therefore, we need to attack step #1.

USUAL TECHNIQUE To upper bound the advantage the trick is to rely on the statistical structure. We already mentioned it is a L^2-looking criterion, and while this analogy has a clear formalization (see e.g. (Kunisky, Wein, and Afonso S. Bandeira 2019)), we do not need it in its entirety. If we have an orthonormal basis, the representation of equation 1.13 greatly simplifies:

$$\mathrm{Adv}_{(\leqslant D)}(\mathrm{H}_0, \mathrm{H}_1) = \sup_{\substack{f:\{-1,1\}^{n\times n}\mapsto\mathbb{R} \\ \deg(f)\leqslant D}} \frac{\mathbb{E}_{\mathrm{H}_1}\left[\sum_{j\in\text{basis}} \alpha_j \psi(\mathbf{Y}; j)\right]}{\|\boldsymbol{\alpha}\|_2} = \sup_{\boldsymbol{\alpha}:\|\boldsymbol{\alpha}\|_2=1} \sum_{j\in\text{basis}} \alpha_j \mathbb{E}_{\mathrm{H}_1}\left[\psi(\mathbf{Y}; j)\right], \tag{2.17}$$

and the complicated fraction became a linear sum. Upper bounding terms to get a bounded sum is then a problem-dependent task.

While the existence of an orthonormal basis is not an issue in separable spaces, its tractability or it being explicit is not a guarantee. For pure noise null hypotheses as in problem 1.1, it is (see lem. 2.23):

Definition 2.18 (Canonical monomials)**.** *Let $G = (V, E)$ be a graph over $V = \{v_1, \dots, v_\ell\}$ vertices. Denote Π_ℓ the set of injective mappings that label the graph. Define:*

$$\begin{aligned} P : \mathcal{G} \times \Pi_\cdot \times \mathbb{R}^{n\times n} &\to \mathbb{R} \\ (G = (V, E), \Pi_{|V|}, \boldsymbol{Y}) &\mapsto \prod_{(i,j)\in E} Y_{\pi(i),\pi(j)} \end{aligned} \tag{2.19}$$

We term "canonical basis" of polynomials of degree less than D the set $\left(1, \left(P_{G,\pi}\right)_{(\pi,G),(\leqslant D)}\right)$.
Here, the (π, G) pairs cover all non-empty labelled graphs with less than D edges and no isolated vertices.
We show it is an orthonormal basis for distributions allowed in $\mathrm{H}_0^{\text{noise}}$ of problem 1.1 in lemma 2.23.

Remark 2.20. *We add the constant function $f(\boldsymbol{Y}) \equiv 1$ because it is needed and not present in the monomials. Morally, it corresponds to the empty graph $G = \varnothing$.*

Remark 2.21. *We use the notion of labelled graph $\pi(G)$ for later purposes. One can just think of an element of this basis as a product over edges (i, j). When we sum over G, π, we morally sum over labelled graphs in the entries of the matrix.*

Remark 2.22. *In our specific model the underlying graph is simple (no double edges), and has no self-loops, i.e. no Y_{ii} terms in the basis.*

Lemma 2.23. *Under the distributions allowed in* $\mathrm{H}_0^{\text{noise}}$ *from problem 1.1 the canonical basis* $\left(1, \left(P_{G,\pi}\right)_{\pi\in\Pi_{|V|}}\right)_{G\in\mathcal{G}_{(\leqslant D)}}$ *is an orthonormal basis of polynomials of degree less than* D *from* $\{-1,1\}^{n\times n}$ *to* $\mathbb{R}$.

Proof. See appendix A. □

Contrarily, for structured null hypotheses, such as those in problem 1.3, or for estimation,[2] which is problem 1.5, finding an orthonormal basis is a hard task. As we mentioned in subsection 1.II, this impediment was the main motivation behind the work of Schramm and Wein (2022). There, the authors and the works that followed rely on a careful lower bound on the denominator to get to an upper bound on the advantage as in equation 2.17. The resulting terms in the sum are recursively defined, and the technology is very involved. Our proposal summarized in subsection 1.I is more explicit and direct. In the next section, we present its consequence.

3 MAIN RESULT

The main result is a bound on the advantage (def. 1.10) in problem 1.3 for the planted sub-matrix model of equation 1.6. As we said, it supports the claim that polynomials up to some degree fail at solving the hypothesis test. It also works for problem 1.1, but it is superfluous (see remark 3.6). The main ideas behind an adaptation to a larger class of models are in subsection 6.IV.
In particular, to obtain results about the denominator of the advantage (def. 1.10), which is independent of H_1, we only need to work on the first three parameters (n,k,λ) and the degree D.

Assumption 3.1. *The degree is such that* $D \geqslant 2$. *Moreover:*

$$\max\left\{\frac{k}{n}, \frac{\lambda k}{\sqrt{n}}, \lambda\right\} \leqslant D^{-8c_{\text{si}}}, \tag{3.2}$$

for some large universal constant $c_{\text{si}} > 0$. *For the sake of this document, we do not optimize it.*

Thanks to assumption 3.1, we will simplify greatly the denominator of the advantage (def. 1.10).

Proposition 3.3 (Preliminary version of proposition 5.82)**.** *Suppose we want to study problem 1.3 for the planted sub-matrix model of equation 1.6. If the* $\boldsymbol{\theta}$ *of* H_0 *satisfies assumption 3.1 there exists a basis for the distribution in* H_0 *that is almost orthonormal in the sense of definition 1.14.*

Remark 3.4 (Important comment on almost orthonormality)**.** *We want an upper bound on the advantage (def. 1.10). Therefore, almost orthonormality may look redundant, as we only need the upper bound in definition 1.14, i.e. that for some basis* $\|\boldsymbol{\alpha}\|_{\mathbb{E}_{\mathrm{H}_0}[\boldsymbol{\psi}\boldsymbol{\psi}^\top]} \gtrsim \|\boldsymbol{\alpha}\|_2$ *for all* $\boldsymbol{\alpha}$ *vectors of coefficients of the decomposition. While this is true to obtain an upper bound, there is no guarantee that it will be tight. Definition 1.14 ensures that the representation through the almost orthonormal basis and that through the existing orthonormal basis differ by a multiplicative constant.*

Remark 3.5 (Interpretation)**.** *There are three conditions, each with a precise meaning. Recall that we want to prove a negative result for algorithms.*

- *We require* k/n *to be small because otherwise we saturate the information-theoretic bound in the sense of appendix C; in regimes where the signal is as large as the observation the problem has no statistical-to-computational gap.*
- *We need* $\lambda k/\sqrt{n}$ *to be small because if it is not then a line-sum statistic can solve problem 1.3. Actually, it can estimate (i.e. problem 1.5) the latent locations* $\{j \in [n] \mid \mathrm{x}_j = 1\}$ *which is a harder task in the sense of definition C.8. We justify this in the next remark and show it in appendix B.*
- *We postulate the condition on* λ *to make the first two results at the right scale.*

Remark 3.6 (Interesting regimes)**.** *Complex testing (prob. 1.3) is a generalization of detection (prob. 1.1) where the null distribution can have a signal. Therefore, to prove a negative result about complex testing we need detection to be in an easy regime. In addition to this, we can claim that complex testing is easier (in the sense of definition C.8) than estimation of the position of the signal (prob. 1.5), as we can trivially use an estimator for the signal to solve the hypothesis testing question. This suggests that complex testing is in between detection and estimation. The practical*

[2] One can prove hardness of estimation by bounding an analog of the advantage on only one distribution: it has signal in the denominator and numerator. See (Schramm and Wein 2022; Sohn and Wein 2025).

implication is that if we want complex testing to be an "interesting" problem of its own then we need detection to strictly dominate estimation (again, in the sense of def. C.8). In terms of the signal conditions of the assumption, this results in an added condition that $k \gtrsim_{\log} \sqrt{n}$. This region of parameters is said to have a ***detection-recovery*** *gap.*[3]
If we do not consider the case when $k \gtrsim_{\log} \sqrt{n}$, then we still have a hardness result for complex testing, but we already know that detection, which is an easier problem (in the sense of definition C.8) is hard, so we are proving a superfluous result.

To upper bound the full expression, we add an assumption on H_1.

Assumption 3.7 (Perturbation)**.** *The perturbations for complex testing (problem 1.3) satisfy the following conditions, where c_{si} is the same constant of assumption 3.1:*

- *if we test a perturbation on the strength of the signal λ against $\lambda + \eta$, then $\eta \leqslant \lambda/D$ and $\eta k^2/n \leqslant D^{-8c_{\mathrm{si}}}$;*
- *if we test a perturbation on the size of the signal k against $k + \zeta$ then $\zeta \leqslant k/2D$ and $\zeta\sqrt{\lambda/n} \leqslant D^{-8c_{\mathrm{si}}}$.*

The conditions on the c_{si} constant may vary across the types of perturbations here. Morally, c_{si} is always a constant but it might be different when we test on λ or on k.

The main result is a fine control on the advantage.

Theorem 3.8. *Suppose we want to study problem 1.3 for the planted sub-matrix model of equation 1.6. Recall the definition of advantage (def. 1.10) and the fact that $\mathrm{H}_0, \mathrm{H}_1$ are hypotheses of distributions, so that the notation $\mathrm{Adv}(\mathrm{H}_0, \mathrm{H}_1)$ is unambiguous.*
Let assumptions 3.1 and 3.7 hold for $(\boldsymbol{\theta}, \boldsymbol{\theta}') \in \Theta^{\times 2}$ parameterizing the distribution in H_0 and H_1 respectively. Then for all $D \geqslant 2$:

$$\mathrm{Adv}_{(\leqslant D)}(\mathrm{H}_0, \mathrm{H}_1) \leqslant 1 + \frac{1}{D}. \tag{3.9}$$

Remark 3.10 (Interpretation)**.** *In both cases of complex testing (prob. 1.3) we take the η or ζ perturbation to be smaller by at least a D factor than the actual signal, i.e. $\eta \leqslant \lambda/D$ and $\zeta \leqslant k/2D$. Had we not assumed this, the intuition is that we would have degraded to a detection task (prob. 1.1). For sufficiently large perturbations the null hypothesis "looks like pure noise" even if it has signal from the perspective of the alternative.*
If the other assumption does not hold, then there is a matching algorithm, as we said in remark 3.5 for the other conditions of assumption 3.1. We sketch the argument in appendix B.

3.1 *Discussion*

COMPARISON WITH LITERATURE Our model in equation 1.6 is closely related to planted clique, which is one of the pillars of average-case hardness. The detection problem (prob. 1.1) was amply studied in literature (Brennan and Bresler 2020; Hopkins et al. 2017; Kunisky, Wein, and Afonso S. Bandeira 2019; Wein 2025a), as well as the estimation problem (prob. 1.5) (Alon, Krivelevich, and Sudakov 1998; Schramm and Wein 2022; Sohn and Wein 2025). Instead, the complex testing scenario of problem 1.3 was somehow overlooked. While it is true that the technique of Schramm and Wein (2022) is well-suited for testing between distributions with signal, the novelty of our result is the proof technique. The almost orthonormal basis (def. 1.14) greatly simplifies the steps to bound the advantage and is not limited to the planted sub-matrix model (see extensions in subsection 6.IV).

ALGORITHMIC IMPLICATIONS As we mentioned in section 2, the low-degree conjecture states that polynomial-time algorithms correspond to polynomials of degree $\log n$, and an *impossibility result* for polynomial-time algorithms corresponds to showing that the advantage is bounded in some proper sense for polynomials of degree slightly larger than $\log n$, e.g. $(\log n)^{1+\epsilon}$ for all $\epsilon > 0$. We can see from our statement in theorem 3.8 that for $D = \omega(\log n)$ we have $\mathrm{Adv}_{(\leqslant D)}(\mathrm{H}_0, \mathrm{H}_1) = 1 + o(1)$, which according to definition 2.5 - 2.12 means weak separation (and thus weak detection conjecturally) is impossible in the signal regimes of assumptions 3.1 - 3.7. Logically, since weak detection is impossible, so is strong detection.[4]

[3] A way to find this condition is to compare the detection threshold $\lambda \approx_{\log} n/k^2$ and the estimation threshold, which is $\lambda \approx_{\log} \sqrt{n}/k$, imposing that they are distinct. The reason why these two are expected to be the detection and recovery threshold is in the analysis of appendix B.
[4] Notice how this is in accordance with the fact that $\mathrm{Adv}_{(\leqslant D)}(\mathrm{H}_0, \mathrm{H}_1) = O(1)$ and definition C.8.

NON-CONJECTURAL SIDE We can find poly-time algorithms or information-theoretic barriers (in the sense of appendix C) as soon as we violate either of the conditions in assumptions 3.1 - 3.7. In appendix B we propose all the informal concentration arguments for perturbations of k, λ and all possible relaxations of the assumptions, plus one fully formalized.[5] This is the best-case scenario where we have a matching positive result. The simple statistics:

$$\mathsf{s}_{\text{global}}(\boldsymbol{Y}) := \sum_{i,j} Y_{ij}, \qquad \text{and} \qquad \mathsf{s}_{\text{line}}(\boldsymbol{Y}) := \#\left\{ j \mid \sum_{i \neq j} Y_{ij} \geqslant \omega \right\}, \tag{3.11}$$

for some well-chosen ω, distinguish $(\mathrm{H}_0, \mathrm{H}_1)$ as soon as we get out of the parameters allowed by assumptions 3.1 - 3.7. As we expect not to find a better algorithm, we extrapolate that the low-degree method captures the right behavior of polynomial time algorithms. The following propositions summarize the discussion in appendix B:

Proposition 3.12 (Informal)**.** *Consider assumption 3.1 for the planted sub-matrix model of equation 1.6 and problem 1.3, where we perturb either λ or k in the signal.*

- *If we relax the condition on k/n and consider precision up to poly-logarithmic factors then there exists an efficiently computable function able to solve the detection problem 1.1 optimally among all functions: we hit an information-theoretic bound.*
- *If we relax the condition on $\lambda k/\sqrt{n}$ up to poly-logarithmic factors, then there is a polynomial-time algorithm solving complex testing (prob. 1.3).*
- *Similarly, if we relax assumption 3.7 up to poly-logarithmic factors there exists a polynomial-time algorithm solving complex testing.*

In words, if we break either of the interesting inequalities in assumptions 3.1 - 3.7 there is a statistic.

The only one we fully formalize is as follows.

Proposition 3.13. *Consider the instance of problem 1.3 where H_0 is a distribution $\mathcal{P}_{\boldsymbol{\theta}}$ with $\boldsymbol{\theta} = (k, \lambda)$ and H_1 is a distribution $\mathcal{P}_{\boldsymbol{\theta}'}$ where $\boldsymbol{\theta}' = (k, \lambda + \eta)$ for some $\eta > 0$. Let assumption 3.1 hold, and suppose assumption 3.7 does not hold. Quantitatively, suppose $k^2\eta/n > 4\sqrt{2\ln 8}$. Then the statistic $\mathsf{s}_{\text{global}}(\boldsymbol{Y})$ from equation 3.11 is able to perform weak detection (def. 2.5) by weakly separating (def. 2.12) the distributions in the null and in the alternative. In equations, the following holds:*

$$\mathbb{P}_{\mathrm{H}_0}\left[\mathsf{s}_{\text{global}}(\mathbf{Y}) - \mu_0 \geqslant \xi\right] + \mathbb{P}_{\mathrm{H}_1}\left[\mathsf{s}_{\text{global}}(\mathbf{Y}) - \mu_0 < \xi\right] \leqslant 8\sqrt{\mathrm{e}^{-\phi^2/8n^2}} + \epsilon < 1, \qquad \forall \epsilon > 0, \tag{3.14}$$

where $\mu_0 = n(n-1)/n^2k^2\lambda, \phi = n(n-1)/n^2k^2\eta$ and ξ is any value in the interval $(n\sqrt{-2\ln p_{\mathrm{c}}/8}, \phi - n\sqrt{-2\ln p_{\mathrm{c}}/8})$, with $p_{\mathrm{c}} = 8\sqrt{\mathrm{e}^{-\phi^2/8n^2}} + \epsilon$.

Remark 3.15. *If we believe in the low-degree conjecture of section 6, this result suggests that the low-degree method captures up to poly-logarithmic factors the behavior of some "simple" statistics expected to be optimal. See subsection B.III for alternative arguments in favour of this conclusion.*

BEYOND POLYNOMIAL TIME An extension of the low-degree conjecture we also mentioned at the beginning of section 2 states that degree $D = n^\delta$ polynomials correspond to algorithms with runtime $n^{\widetilde{\Theta}(D)} = \exp\left\{n^{\delta \pm o(1)}\right\}$, where $\widetilde{\Theta}(\cdot)$ hides poly-logarithmic factors. The generic correspondence for negative results is that if the advantage is bounded (or asymptotically one) for some $D \equiv D(n) \leqslant t(n)\mathsf{polylog}(n)$ then algorithms with runtime $n^{t(n)}$ cannot distinguish the hypotheses. In other words, since theorem 3.8 is a negative result, for any $t(n)$ choice we will only be able to say when $n^{t(n)}$ algorithms are *conjecturally expected not to work*. Below, we comment on this formalism with two interesting aspects.

The first natural question is how sub-exponential algorithms perform under our assumptions. We say an algorithm is sub-exponential if for some $\delta \in (0,1)$ it has runtime $n^{n^\delta} = \exp\left\{\widetilde{O}\left(n^\delta\right)\right\}$, where $\widetilde{O}(\cdot)$ hides poly-logarithmic factors (Kunisky, Wein, and Afonso S. Bandeira 2019). The motivation is simple: we define polynomial-time algorithms to have runtime n^{poly}, where poly is any constant polynomial, and the immediate

[5] The others are analogous.

super-class of exponential type allow for n^δ to be a vanishing monomial in n. In the sub-exponential case, $t(n) = n^\delta$ for any $\delta > 0$. We find that their advantage is bounded as $1 + 1/D$ going faster to zero, but in a *much narrower* region of parameter space, since assumption 3.1 has inequalities that shrink exponentially fast with D. The second question is if we can rule out somehow weak separation and not strong separation. Ideally, we need the advantage to be bounded but not vanishing, which in theorem 3.8 corresponds to $D = \Theta(1)$. From the identification with $n^{t(n)}$ algorithms and $D(n) \leqslant t(n)\mathsf{polylog(n)}$ we already see that we cannot have this level of accuracy in our bound. Moreover, we would be in a regime where the signal quantities in assumption 3.1 are potentially not vanishing.

In the next subsection, we start to work with bases to rewrite the advantage from equation 1.11 into a specific version of equation 1.13, which will be equation 3.27.

3.II *A working representation of the advantage*

ADJUST THE BASIS While we cannot hope for an orthonormal basis in problem 1.3 as we said in section 2 there is an orthonormal basis for problem 1.1 to start from, which is that of definition 2.18. It is orthonormal for the pure noise scenario by lemma 2.23: the null hypothesis of problem 1.1 of distinguishing a Rademacher matrix from a matrix with signal. Let us call this hypothesis $\mathrm{H}_0^{\text{noise}}$ to be explicit. Under $\mathrm{H}_0^{\text{noise}}$ we have $\mathrm{Y}_{ij} \overset{\text{i.i.d.}}{\sim} \mathrm{Rad}(1/2)$ for all $i \leqslant j$, under H_0 of problem 1.3 the observation is not i.i.d., but the canonical basis is still a basis.

Lemma 3.16. *In all the probability distributions we consider, the canonical basis* $\left(1, \left(P_{G,\pi}\right)_{\pi \in \Pi_{|V|}}\right)_{G \in \mathcal{G}_{(\leqslant D)} \cup \{\varnothing\}}$ *is a basis of polynomials of degree less than D from* $\{-1,1\}^{n \times n}$ *to* $\mathbb{R}$.

Proof. See appendix A. □

In particular, for a generic hypothesis H_ℓ of a probability distribution with $\boldsymbol{\theta} = (k, \lambda)$, we have:

$$\begin{aligned}
\left\langle P_{G^{(1)},\pi^{(1)}}, P_{G^{(2)},\pi^{(2)}} \right\rangle_{\mathrm{H}_\ell} &:= \mathbb{E}_{\mathrm{H}_\ell}\left[P_{G^{(1)},\pi^{(1)}} P_{G^{(2)},\pi^{(2)}}\right] \\
&= \mathbb{E}_{\mathrm{H}_\ell}\left[\prod_{\substack{(i,j)\in\pi^{(1)}(G^{(1)}) \\ (i,j)\in\pi^{(2)}(G^{(2)})}} \mathrm{Y}_{ij}^2 \prod_{\substack{(i,j)\in\pi^{(1)}(G^{(1)}) \\ (i,j)\notin\pi^{(2)}(G^{(2)})}} \mathrm{Y}_{ij} \prod_{\substack{(i,j)\notin\pi^{(1)}(G^{(1)}) \\ (i,j)\in\pi^{(2)}(G^{(2)})}} \mathrm{Y}_{ij}\right] \\
&= \mathbb{E}_{\mathbf{X}}\left[\mathbb{E}_{\mathbf{Y}|\mathbf{X}}\left[\prod_{\substack{(i,j)\in\pi^{(1)}(G^{(1)}) \\ (i,j)\notin\pi^{(2)}(G^{(2)})}} \mathrm{Y}_{ij} \prod_{\substack{(i,j)\notin\pi^{(1)}(G^{(1)}) \\ (i,j)\in\pi^{(2)}(G^{(2)})}} \mathrm{Y}_{ij}\right]\right] \\
&= \mathbb{E}_{\mathbf{X}}\left[\prod_{\substack{(i,j)\in\pi^{(1)}(G^{(1)}) \\ (i,j)\notin\pi^{(2)}(G^{(2)})}} \mathrm{X}_{ij} \prod_{\substack{(i,j)\notin\pi^{(1)}(G^{(1)}) \\ (i,j)\in\pi^{(2)}(G^{(2)})}} \mathrm{X}_{ij}\right] \\
&= \lambda^{|E_\triangle|}\left(\frac{k}{n}\right)^{|V_\triangle|},
\end{aligned} \tag{3.17}$$

where $G_\triangle = (V_\triangle, E_\triangle)$ is the symmetric difference graph induced by the edges of $\pi^{(1)}(G^{(1)})$ and $\pi^{(2)}(G^{(2)})$. Crucially, the final formula is due to the conditional independence of $\mathrm{Y}_{ij} \mid \mathrm{X}_{ij}$ for all edges and by the fact that the $\mathrm{X}_{ij} = \mathrm{x}_i\mathrm{x}_j$ are products of i.i.d. Bernoulli random variables dilated by a λ factor. By convention, when the graphs are the same, the symmetric difference is the empty graph, and we return one.[6]
From this observation, orthonormality at $\lambda k = 0$ is immediate, and in the general case we lose it since $\lambda k \neq 0$.

Thanks to the last lemma, we *always* have a decomposition of the advantage in the canonical basis like in equation 1.13, but only for $\mathrm{H}_0^{\text{noise}}$ we write at the denominator $\|\boldsymbol{\alpha}\|_2$. In the generic case where H_0 has $\lambda \neq 0$ and $k \neq 0$ the basis is not orthonormal.

[6] This is in accordance with the fact that $\mathrm{Y}_{ij}^2 \overset{\text{a.s.}}{=} 1$ for all (i,j) and all distributions considered.

TO SUMMARIZE We see that the advantage from definition 1.10 is the usual object to prove low-degree lower bounds. However, its main weakness is that it relies crucially on an understanding of how orthonormal polynomials enter into the picture. This is due to the fact that the denominator does not have a nice form if the decomposition is not orthonormal. At the same time, the spurious correlations have a precise form, i.e. the symmetric difference of the underlying graphs considered (eqn. 3.17). This structure is in good terms with the symmetries of the problem. Namely:

- any two pairs of labelled graphs that have the same symmetric difference have the same correlation;
- the probability distributions H_ℓ enjoy a permutation symmetry: there are no preferred locations for the signal (see lem. 3.23).

Combining these we attempt to collect terms by invariance in two nested ways: by labellings that have the same correlation and by permutations. Going up at this level of symmetry allows us to not count twice objects that are merely the same with regard to randomness. To formalize this, in subsection 3.III we discuss the invariance by permutations, thanks to which we decompose the advantage differently. We argue that this regrouping greatly simplifies how we handle correlations between labelled graphs.
In subsection 3.IV we then clarify what it means for a basis to be almost orthonormal, as we need for our purposes. In particular, we present a sufficient condition for almost orthonormality (definition 1.14) to hold.

3.III *First idea: grouping by invariants, skeleton graphs*

As hinted in the previous subsection, working over labelled graphs is superfluous: the randomness of the problem and in particular the way monomials correlate (i.e. equation 3.17) do not crucially depend on the way we labelled the two graphs. In this subsection, we present the formalism of skeletons which seeks to go up the ladder of generality and reach the highest level of collection by invariant quantities.
As we did in definition 2.18, we consider a graph $G = (V, E)$ where $V = \{v_1, \dots v_\ell\}$ are its nodes/vertices for some $\ell \geqslant 2$ and where E is the set of edges. We write $|V|$ the number of nodes, and $|E|$ the number of edges.

Assumption 3.18. *Throughout the text all graphs have no isolated vertices. In other words, they are induced by their edge set.*

A skeleton is a collection of vertices and edges without regard to the labellings. In other words, it is the equivalence class of the given (V, E) pair. However, there are few ways to parameterize it, so we will adopt choices that are useful for deriving results. Since it is relevant only for combinatorial purposes, we postpone the details to section 4 and just present the mere definition.
In what follows Π_ℓ is the set of injective mappings $\pi : V \to [n]$ with $|V| = \ell$.

Definition 3.19 (Skeleton)**.** *Let $G = (V, E)$ be a graph, where $V = \{v_1, \dots, v_\ell\}$, $\ell \leqslant n$, and there are no isolated nodes. We say a graph $\pi(G)$ on ℓ labelled vertices $\{i_1, \dots, i_\ell\} \subset [n]$ is in the skeleton G when there is a labelling $\pi : V \to [n]$, with $\pi \in \Pi_\ell$, such that G and $\pi(G)$ are isomorphic through π (def. 4.2). Namely, we have $\pi(G) \cong G$. A skeleton is an equivalence class of graphs. We consider skeletons up to isomorphism, meaning that two isomorphic graphs in the abstract space are the same skeleton. Within a skeleton, we include all distinct objects arising from injections $\pi \in \Pi_{|V|}$. In particular, we also count automorphic labelled graphs. For extensive clarifications, see remark 4.7 and all of the section 4.*

Remark 3.20 (Alternative view)**.** *The skeleton formalism is useful if we need to refer to the original vertices with labels unambiguously. Otherwise, we could have taken an unlabelled graph, or the skeleton with removed labels. The issue with the unlabelled formulation is that it is somehow non-standard to refer to edges of an unlabelled graph. In practice, the set $V = \{v_1, \dots, v_\ell\}$ is an abstract set of vertices that we label through π.*

Definition 3.21 (Set of skeletons)**.** *For a given $D \equiv D(n) \leqslant n$, we define $\mathcal{G}_{(\leqslant D)}$ as the set of skeletons with less than D edges quotiented by the isomorphism relation over graphs in the abstract space.*

Remark 3.22. *Coming back to the basis in definition 2.18, we have that $(P_{G,\pi})_{(\pi,G),(\leqslant D)}$ has the same elements as $\left((P_{G,\pi})_{\pi\in\Pi_{|V|}}\right)_{G\in\mathcal{G}_{(\leqslant D)}}$.*

SYMMETRY The form of the correlation in equation 3.17 is emblematic. While we work over labelled graphs, two *different* pairs of labelled graphs that have the same symmetric difference have the same correlation. Working with skeletons allows us to fix the "shape" of the two graphs. Once we fix two skeletons $G^{(1)}, G^{(2)}$, the many pairs $\pi^{(1)} \in \Pi_{|V^{(1)}|}$ and $\pi^{(2)} \in \Pi_{|V^{(2)}|}$ can be regrouped into sets of pairs that have the same symmetric difference. Had we worked at the level of labelled graphs, we would have not done this as nicely, especially with regard to the second invariance, that of permutations. We discuss it below.

Working over skeletons also allows us to exploit the permutation symmetry of the problem. The next lemma formalizes what we meant earlier by "the signal has no preferred location".

Lemma 3.23. *Fix any degree $D > 0$. Then, the value of the advantage $\mathrm{Adv}_{(\leqslant D)}(\mathrm{H}_0, \mathrm{H}_1)$ in problems 1.1 - 1.3 is achieved by a function $f^\star$ such that $f^\star(\boldsymbol{Y})$ is invariant by permutation. In other words, for any bijection $\sigma : [n] \mapsto [n]$, we have $f^\star(\boldsymbol{Y}) = f^\star(\boldsymbol{Y}_\sigma)$ where $\boldsymbol{Y}_\sigma := (Y_{\sigma(i),\sigma(j)})_{i,j}$.*

Proof. See appendix A. □

Since there is an invariant optimal polynomial, it makes sense to seek a basis over invariant polynomials. The way to build a naïve proposal is to take the canonical basis of definition 2.18 and symmetrize it. In words, this means collecting all basis elements that are in the same skeleton.[7] We write for a given skeleton G:

$$P_G(\boldsymbol{Y}) = \sum_{\pi \in \Pi_{|V|}} P_{G,\pi}, \tag{3.24}$$

and wonder if it is enough. Intuitively, the symmetrized object P_G should be invariant to permutations because we construct it by using all of them, and should still keep the explanatory power that the $(P_{G,\pi})_{\pi \in \Pi_{|V|}}$ had. We check this in the next lemma.

Lemma 3.25. *Consider problems 1.1 - 1.3. Let f be a polynomial invariant to permutations, of degree less than D, with domain in $\{-1,1\}^{n \times n}$. There exist numerical values $(\alpha_G)_{G \in \mathcal{G}_{(\leqslant D)} \cup \{\varnothing\}}$ such that $f(Y) = \alpha_\varnothing + \sum_{G \in \mathcal{G}_{(\leqslant D)}} \alpha_G P_G(Y)$. In other words, the collection $\left(1, (P_G)_{G \in \mathcal{G}_{(\leqslant D)} \cup \{\varnothing\}}\right)$ with $P_G = \sum_{\pi \in \Pi_{|V|}}$ is a basis of invariant polynomials of degree less than D.*

Proof. See appendix A. □

Combining lemmas 3.23 - 3.25, we have the following lemma.

Lemma 3.26. *Let $D > 0$. For any pairs of hypotheses from problems 1.1 and 1.3 the advantage decomposes along $\left(1, (P_G)_{G \in \mathcal{G}_{(\leqslant D)} \cup \{\varnothing\}}\right)$ in the sense of equation 1.13. In equations, we have:*

$$\mathrm{Adv}_{(\leqslant D)}(\mathrm{H}_0, \mathrm{H}_1) = \sup_{(\alpha_G)_{G \in \mathcal{G}_{(\leqslant D)} \cup \{\varnothing\}}} \frac{\mathbb{E}_{\mathrm{H}_1}\left[\sum_{G \in \mathcal{G}_{(\leqslant D)}} \alpha_G P_G\right]}{\sqrt{\mathbb{E}_{\mathrm{H}_0}\left[\left(\sum_{G \in \mathcal{G}_{(\leqslant D)}} \alpha_G P_G\right)^2\right]}}. \tag{3.27}$$

Moreover, the same holds for any alternative basis over invariant polynomials of degree less than D.

Proof. By lemma 3.23 the advantage is attained by an invariant function. Using lemma 3.25, we decompose any invariant function as $f = \alpha_\varnothing + \sum_{G \in \mathcal{G}_{(\leqslant D)}} \alpha_G P_G$. We have the chain of equalities:

$$\begin{aligned}
\mathrm{Adv}_{(\leqslant D)}(\mathrm{H}_0, \mathrm{H}_1) &= \sup_{\substack{f:\{-1,1\}^{n \times n} \mapsto \mathbb{R} \\ f:\deg(f) \leqslant D}} \frac{\mathbb{E}_{\mathrm{H}_1}\left[f(\mathbf{Y})\right]}{\sqrt{\mathbb{E}_{\mathrm{H}_0}\left[f^2(\mathbf{Y})\right]}} \\
&= \sup_{\substack{f:\{-1,1\}^{n \times n} \mapsto \mathbb{R} \\ f:\deg(f) \leqslant D \\ f \text{ invariant to permutations}}} \frac{\mathbb{E}_{\mathrm{H}_1}\left[f(\mathbf{Y})\right]}{\sqrt{\mathbb{E}_{\mathrm{H}_0}\left[f^2(\mathbf{Y})\right]}} \\
&= \sup_{(\alpha_G)_{G \in \mathcal{G}_{(\leqslant D)} \cup \{\varnothing\}}} \frac{\mathbb{E}_{\mathrm{H}_1}\left[\sum_{G \in \mathcal{G}_{(\leqslant D)}} \alpha_G P_G\right]}{\sqrt{\mathbb{E}_{\mathrm{H}_0}\left[\left(\sum_{G \in \mathcal{G}_{(\leqslant D)}} \alpha_G P_G\right)^2\right]}}.
\end{aligned} \tag{3.28}$$

□

[7] Here we start noticing that the redundant formalism of labelled graphs in the definition is useful.

Once we know that the advantage depends only on skeletons and that the correlations of the canonical basis group in terms of how they intersect, we seek to make the most of these regroupings. The essence of our new proof technique is that there exists an almost orthonormal basis (def. 1.14). The path to present it clearly is precisely to go at this level of invariance, where counting becomes easier because we make the most of the symmetries. In the next section, we discuss the idea behind the claim of proposition 3.3 and further intuition on how to establish the actual result, which is proposition 5.82.

3.iv *Second idea: almost orthonormality*

The issue with *any* basis is that the denominator of equation 3.27 is a highly coupled quadratic form where each basis element interacts with the other.[8] Characterizing the interaction of the basis with *any* set of coefficients $(\alpha_G)_{G\in\mathcal{G}_{(\leqslant D)}}$ is far from trivial in general. In this section, we will sketch how we deal with this aspect to prove an inequality as in definition 1.14, which we formalize in proposition 5.82. Then, establishing theorem 3.8 is a routine consequence.

As in section 1 the simplest thing we can hope for is that for a given a basis decomposition:

$$f(\cdot) = \sum_{G\in\mathcal{G}_{(\leqslant D)}} \alpha_G \psi(\cdot, G), \tag{3.29}$$

definition 1.14 holds. In another perspective, we hope that the decomposition along the basis becomes as good as the Euclidean norm of the coefficients if we let $n \to \infty$. Mathematically, we want to find a good $(\psi(\cdot, G))_{G\in\mathcal{G}_{(\leqslant D)}}$ set such that:

$$\|\boldsymbol{\alpha}\|_2^2 = \|f\|_{\mathrm{H}_0}^2 = \sum_{G\in\mathcal{G}_{(\leqslant D)}} \alpha_G^2 \|\psi(\cdot, G)\|_{\mathrm{H}_0}^2 + \sum_{\substack{G^{(1)}\neq G^{(2)} \\ G^{(1)},G^{(2)}\in\mathcal{G}_{(\leqslant D)}}} \alpha_{G^{(1)}}\alpha_{G^{(2)}} \left\langle \psi(\cdot, G^{(1)}), \psi(\cdot, G^{(2)}) \right\rangle_{\mathrm{H}_0} \approx \|\boldsymbol{\alpha}\|_2^2. \tag{3.30}$$

We propose next three different and complementary views on our objective.

COVARIANCE VIEW We want to establish if under the interesting scaling the basis is almost orthonormal (def. 1.14). As a reminder, it needs to hold for any invariant polynomial of degree less than D. Indeed, while it cannot work all the way for any function, we can hope that if we sum over skeletons in $\mathcal{G}_{(\leqslant D)}$ over less than D edges decomposing any invariant polynomial f of degree less than D then it will be enough.
To start simplifying, let us take a normalized basis: we make each $\psi(\cdot, G)$ have $\|\psi(\cdot; G)\|_{\mathrm{H}_0} = 1$. We can also regard the expression for the norm of f decomposed along the basis $\{\psi(\cdot, G)\}_{G\in\mathcal{G}_{(\leqslant D)}\cup\{\varnothing\}}$ as a quadratic form as in definition 1.14:

$$\|f\|_{\mathrm{H}_0}^2 = \boldsymbol{\alpha}^\top \mathbb{E}_{\mathrm{H}_0}\left[\boldsymbol{\psi}\boldsymbol{\psi}^\top\right]\boldsymbol{\alpha} = \|\boldsymbol{\alpha}\|^2_{\mathbb{E}_{\mathrm{H}_0}[\boldsymbol{\psi}\boldsymbol{\psi}^\top]}, \qquad \boldsymbol{\psi} \coloneqq (\psi(\cdot, G))_{G\in\mathcal{G}_{(\leqslant D)}\cup\{\varnothing\}}. \tag{3.31}$$

In particular, the matrix in the quadratic form is the expectation of a rank-one matrix $\mathbb{E}_{\mathrm{H}_0}\left[\boldsymbol{\psi}\boldsymbol{\psi}^\top\right]$ that collects the basis. It is a Gram matrix, with unit diagonal once we normalize and take the expectation. What we morally wish is that:

$$\mathbb{E}_{\mathrm{H}_0}\left[\boldsymbol{\psi}\boldsymbol{\psi}^\top\right] \approx \boldsymbol{I}_{|\mathcal{G}_{(\leqslant D)}|+1}. \tag{3.32}$$

Since the matrix is Gram, it is positive semi-definite.

GERSHGORIN VIEW Another perspective is to use Gershgorin's circle theorem (see (Potters and Bouchaud 2020, chap. 1.2.1) and (Horn and Johnson 2012, chap. 6)). The matrix $\mathbb{E}_{\mathrm{H}_0}\left[\boldsymbol{\psi}\boldsymbol{\psi}^\top\right]$ has unit diagonal entries. Then, each eigenvalue is at least into one of the "Gershgorin circles", which are circles centered at the diagonal terms, i.e. 1 and with radius being the sum over columns or rows that discard the diagonal. Since the matrix is symmetric, we need not consider the distinction of rows and columns. Then, we can say that all eigenvalues are within:

$$1 \pm \sup_{G^{(1)}\in\mathcal{G}_{(\leqslant D)}} \sum_{\substack{G^{(2)}\in\mathcal{G}_{(\leqslant D)}\cup\{\varnothing\} \\ G^{(2)}\neq G^{(1)}}} \left|\mathrm{CoV}_{\mathrm{H}_0}\left[\psi(\cdot; G^{(1)}), \psi(\cdot; G^{(2)})\right]\right|. \tag{3.33}$$

We then want the sum of covariances to be small.

[8] Alternatively, as we will do later, we see it as an eigenvalue of a matrix of integrals over products of polynomials.

APPROXIMATE ISOMETRIES We want the $\mathbb{E}_{H_0}\left[\boldsymbol{\psi}\boldsymbol{\psi}^\top\right]$ matrix to be an approximate isometry, where ϵ the approximation factor is small.

It turns out that for restricted isometry conditions all the three are equivalent, so we can either:

1. show that $\mathbb{E}_{H_0}\left[\boldsymbol{\psi}\boldsymbol{\psi}^\top\right]$ is such that $\left\|\mathbb{E}_{H_0}\left[\boldsymbol{\psi}\boldsymbol{\psi}^\top\right] - \boldsymbol{I}\right\|_{\text{op}} \leqslant \epsilon$;
2. show that $\mathbb{E}_{H_0}\left[\boldsymbol{\psi}\boldsymbol{\psi}^\top\right]$ is an ϵ-approximate isometry;
3. show that $\sup_{G^{(1)}\in\mathcal{G}_{(\leqslant D)}} \sum_{\substack{G^{(2)}\in\mathcal{G}_{(\leqslant D)}\cup\{\varnothing\}\\ G^{(2)}\neq G^{(1)}}} \left|\text{CoV}_{H_0}\left[\psi(\cdot;G^{(1)}),\psi(\cdot;G^{(2)})\right]\right| \leqslant \epsilon$.

Remark 3.34. *The control by an arbitrary ϵ is even too strong. If we are able to show that the outer diagonal entries are jointly small, then the non-infinitesimal version of the statements above bounds the eigenvalues inside an interval of* 1*. For example, if:*

$$\sup_{G^{(1)}\in\mathcal{G}_{(\leqslant D)}} \sum_{\substack{G^{(2)}\in\mathcal{G}_{(\leqslant D)}\cup\{\varnothing\}\\ G^{(2)}\neq G^{(1)}}} \left|\text{CoV}_{H_0}\left[\psi(\cdot;G^{(1)}),\psi(\cdot;G^{(2)})\right]\right| \leqslant \frac{1}{2}, \tag{3.35}$$

then the eigenvalues of the Gram matrix $\mathbb{E}_{H_0}\left[\boldsymbol{\psi}\boldsymbol{\psi}^\top\right]$ are all in the strip $[1/2, 3/2]$.

In particular, it will basically always be a sum of non-zero entries, as for each $G^{(1)}, G^{(2)}$ there always exist a pair $(\pi^{(1)}, \pi^{(2)})$ such that the correlation is non-zero. However, the magnitude of each will be rather small. In section 5 we will show that an adjustment of the canonical basis of definition 2.18 is an approximate isometry in the sense above, which means that the eigenvalues are controlled, which means that it is an almost orthonormal basis (def.1.14).

Remark 3.36. *The interesting and key step in the technique of proving theorem 3.8 is this construction of an almost orthonormal basis $\boldsymbol{\psi}$ (def. 1.14). Once we have the property of almost orthonormality with proposition 3.3:*

$$\|\boldsymbol{\alpha}\|_2 \gtrsim \|\boldsymbol{\alpha}\|_{\mathbb{E}_{H_0}[\boldsymbol{\psi}\boldsymbol{\psi}^\top]} \gtrsim \|\boldsymbol{\alpha}\|_2 \qquad \textit{for all } \boldsymbol{\alpha} = (\alpha)_{G\in\mathcal{G}_{(\leqslant D)}}, \tag{3.37}$$

the bound on the advantage follows by canonical arguments since there is a direct upper bound with the form of equation 2.17 up to constants:

$$\text{Adv}_{(\leqslant D)}(H_0, H_1) \lesssim \sup_{\boldsymbol{\alpha}:\|\boldsymbol{\alpha}\|_2=1} \mathbb{E}_{H_1}\left[\sum_{G\in\mathcal{G}_{(\leqslant D)}\cup\{\varnothing\}} \alpha_G \psi(\mathbf{Y};G)\right]. \tag{3.38}$$

In the next section we properly define many objects that emerge from the invariance of labellings. These are crucial for the construction of an almost orthonormal basis in section 5.

4 MORE DETAILS ON THE SKELETONS FORMALISM

In the planted sub-matrix model (eqn. 1.6), the observation is:

$$\{-1,1\}^{n\times n} \ni \mathbf{Y} = \begin{cases} 1 & \text{with probability } \frac{1+\mathbf{X}_{ij}}{2} \\ -1 & \text{with probability } \frac{1-\mathbf{X}_{ij}}{2} \end{cases}, \qquad \mathbf{X}_{ij} = \mathbf{x}_i\mathbf{x}_j, \qquad \mathbf{x}_i \overset{\text{i.i.d.}}{\sim} \sqrt{\lambda}\text{Ber}\left(\frac{k}{n}\right). \tag{4.1}$$

We want to use the low-degree method to derive a negative result for algorithms, as we argued in section 2. To apply it we need to consider polynomials of degree less than D, where in particular for the interesting case $D \approx_{\log} \log n$. Since we work on a nice binary space over $\{-1,1\}^{n\times n}$, these end up being represented through an "adjustment" of the canonical basis for when the random variables are Rademacher distributed, i.e. when $\mathbf{Y} \sim H_0^{\text{noise}}$ as in the null of problem 1.1.[9] As we saw in definition 2.18, this canonical basis is a collection of monomials $P_{G,\pi}(\boldsymbol{Y}) = \prod_{(i,j)\in E} Y_{\pi(i),\pi(j)}$ for $\pi(G)$ a labelled graph, plus the unit function, i.e. the empty graph. For a given set of edges $\mathbb{T}$ in a labelled graph $\pi(G)$, this perspective naturally gives rise to a graph with random variables at each (i,j), which is in turn a product of random variables at each vertex. Since we will fix the degree to be less than D, we will then consider all graphs with vertices in $[n]$ that have less than D edges, and automatically less than $2D$ nodes since there are no isolated vertices. Summing over all $\mathbb{T}$, or equivalently all edge labellings, is not trivial, but we can make some simplifications. In particular, we group graphs up to isomorphisms.

[9] An i.i.d. Rademacher distribution is the analogue of pure noise in this model: there is no signal.

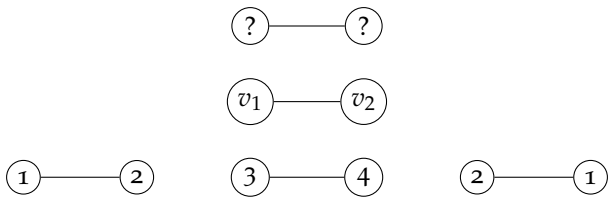


Figure 1: Three isomorphic graphs inside the same skeleton, two of which are automorphic in the observation space
In the formalism of def. 3.19 the first and the third count as two distinct graphs and arise from three injections of the form $\pi : \{v_1, v_2\} \to [n]$, the skeleton, a formalization of the "shape" graph with question marks just above.

CLARIFICATIONS AND CONTEXT In section 3 we introduced skeletons, with definition 3.19 and the justification of seeking the highest level of abstraction to keep the interesting invariant quantities of the problem. We clarify which combinatorial choices we adopt. Then, in the next subsection we define the invariant objects at the level of skeletons that we need.

Definition 4.2 (Graph isomorphism)**.** *Two labelled graphs $G = (V, E), G' = (V', E')$ are isomorphic when there is a bijection $\varphi : V \mapsto V'$ between their vertex sets such that vertices in G are adjacent if and only if they are after applying φ. When two labelled graphs are isomorphic, we write $G \simeq G'$.*

Definition 4.3 (Graph automorphism)**.** *An automorphism of a labelled graph $G = (V, E)$ is a permutation $\sigma : V \mapsto V$ that preserves edge connections, i.e. such that two vertices $(v, v') \in E$ if and only if $(\sigma(v), \sigma(v')) \in \sigma(G)$ the graph obtained after applying the permutation.*
In other words, it is an isomorphism of G to itself.

Definition 4.4 (Automorphism group)**.** *The automorphism group of a labelled graph $G = (V, E)$ is the set of permutations that preserve edge connections. We denote it as $\mathrm{Aut}(G)$. Its size is the number of such permutations, written as $|\mathrm{Aut}(G)|$.*

Lemma 4.5. *Two graphs induced by edges are isomorphic if and only if there exists a permutation matrix $\boldsymbol{P}$ such that $\boldsymbol{A}_{E'} = \boldsymbol{P}\boldsymbol{A}_E\boldsymbol{P}^{-1}$, where $\boldsymbol{A}_E, \boldsymbol{A}_{E'}$ are the respective adjacency matrices.*

Proof. The E, E' graphs are in bijection with their adjacency matrix representation. The operation of permuting E and isolated nodes in $[n]\backslash V$ is represented through $\boldsymbol{P}$, which permutes the columns of the adjacency matrix. □

Remark 4.6. *In our computation, we need to consider the number of permutations that fixes each $\boldsymbol{A}_E$ adjacency matrix for E an edge set. This is the size of the automorphism group of the graph which we denote by $\mathrm{Aut}(G)$.*

Remark 4.7 (What we are counting, what we are not counting)**.** *When referring to a skeleton, we take it in $G \in \mathcal{G}_{(\leqslant D)}$ the space of graphs quotiented by isomorphism. We consider isomorphic skeleton graphs as the same graph. Therefore, if $G^{(1)} \simeq G^{(2)}$ then $P_{G^{(1)},\pi} = P_{G^{(2)},\pi}$ for all labellings, but we consider the polynomial only once. This means that the polynomial represented through $P_{G,\pi}$, ignoring the indexes of the variables x_i is uniquely counted across various $G \in \mathcal{G}_{(\leqslant D)}$.*
*Instead, we choose to consider skeletons "up to automorphisms", in the sense that when building a given skeleton G we place inside **all** the labellings $\pi \in \Pi_{|V|}$ without quotienting the set. Doing so, we will slightly over-count, but it is easier to deal with. The polynomials $P_{G,\pi^{(1)}}, P_{G,\pi^{(2)}}$ for fixed G and different $\pi^{(1)} \neq \pi^{(2)}$ are then the same polynomial in the labelled variables when the underlying graphs are automorphic.*
Since we include automorphisms, the graphs in fig. 1 are all counted inside the same skeleton. The skeleton would be the graph with v_1, v_2; the shape the graph with question marks. In terms of polynomials, we are saying we consider inside P_G (eqn. 3.24) all the three: $P_{G,\pi^{(1)}} = Y_{12}$, $P_{G,\pi^{(2)}} = Y_{21}$ and $P_{G,\pi^{(3)}} = Y_{34}$. The first and third graph are counted twice: they are both present in the skeleton, despite being each related by a permutation (automorphic in the sense of definition 4.3). In figures 2 - 3 we discuss a non-trivial example.

POTENTIAL CORRECTIONS Alternatively, we could consider all possible labellings of vertices neglecting automorphisms. The sums for the symmetrized monomial P_G of equation 3.24 in this case would be:

$$\sum_{\pi \in \Pi_{|V|} \backslash \simeq_{\mathrm{Aut}}} , \tag{4.8}$$

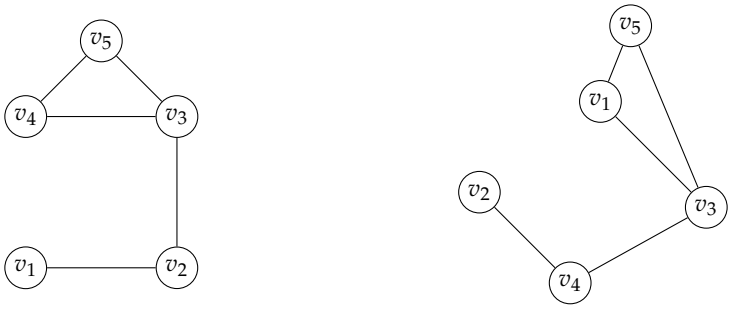


Figure 2: Two isomorphic skeletons in the abstract space
Since skeletons in $\mathcal{G}_{(\leqslant D)}$ are taken from the space of graphs up to isomorphism (def. 4.2) the two graphs above are the same skeleton, i.e. they are counted only once in the abstract space. They correspond to the same $G \in \mathcal{G}_{(\leqslant D)}$. Notice how the "arrangement" of the nodes in space is irrelevant, i.e. the way in which vertices and edges are printed on paper.

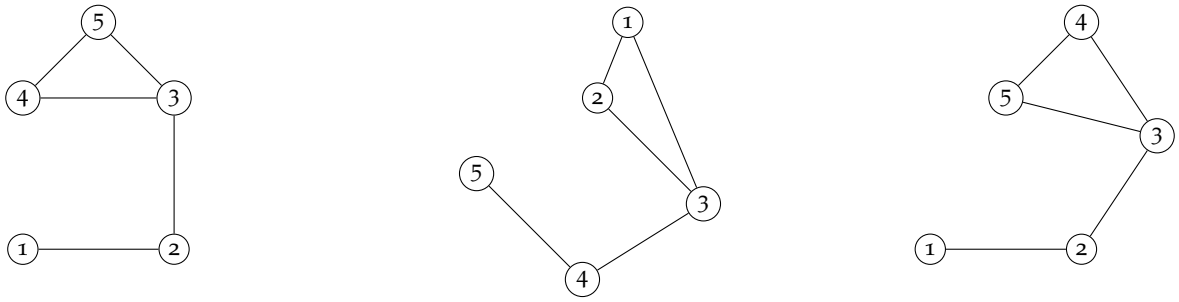


Figure 3: Three labellings of the skeleton in figure 2, two of which are automorphic
Consider the skeleton of figure 2, for simplicity represented through the graph on the left of such figure. The three graphs above correspond to three different labellings $\pi^{(1)}, \pi^{(2)}, \pi^{(3)} \in \Pi_5$. The leftmost graph is such that $\pi^{(1)}(v_i) = i$ for all $i \in [5]$. The center graph is such that $\pi^{(2)}(i) = 5 - i$ for all $i \in [5]$ and the rightmost graph corresponds to the labelling $\pi^{(3)}(i) = i$ for all $i \in \{1, 2, 3\}$, $\pi^{(3)}(4) = 5, \pi^{(3)}(5) = 4$. In particular, the labellings $\pi^{(1)}, \pi^{(3)}$ return two automorphic graphs (def. 4.3) in the observation space. In the skeleton formalism, they are **both** counted as distinct. The center graph is of course also present in the enumeration.

where $\simeq_{\text{Aut}}$ quotients the labellings by automorphisms. It turns out that it is less helpful to group as such. We stick to our choice, and potentially correct the sums for the various P_G polynomials over skeletons as:

$$\frac{1}{|\text{Aut}(G)|}\sum_{\pi\in\Pi_{|V|}}. \tag{4.9}$$

In short: in our formalism the first and third graph in figures 1 - 3 are distinct, and both included when enumerating the labellings of their skeleton.

Thanks to definition 3.19, the way we see E makes it so that skeletons are induced by E only,[10] which fixes:

- the number of vertices $|V_E|$;
- the number of edges $|E|$;
- the "type" of connections.

In particular, for a given graph induced by E, there is a representation in terms of a sequence of tuples $\{(i,\nu_i)\}_{i\geqslant 1}$, where to each $i\in[n]$ we pair its neighbors set $\nu_i\subseteq V_E\backslash\{i\}$, where V_E is the set of vertices induced by the edges E. Notice we also specify that for all $i\notin V_E$ we have $\nu_i=\varnothing$.

To give yet another perspective, we could say that two sub-graphs induced by edge sets E,E' are such that $E\simeq E'$ when there is a permutation $\sigma:[n]\to[n]$ such that the respective representations are equivalent. In particular, it acts on both i and ν_i as $\sigma\circ\nu_i=\nu_{\sigma(i)}$. For this reason a sum over a skeleton is permutation invariant. A skeleton can then be seen as a set of edges that are isomorphic to each other in the sense above. Since it is an equivalence relation, it partitions the set of graphs. We write $\{E:E\cong E'\}$ when we want to fix a skeleton and consider its various edge profiles. The sub-graphs of the complete graph $\mathscr{K}_n$ over $[n]$ vertices are then partitioned into various equivalence classes containing isomorphic graphs.

In the next subsection we define relevant objects at the level of skeletons. These are fundamental for deriving the results of section 3. Working at the such symmetry level of skeletons is crucial here: the objects that arise from the integrals are invariant to permutations, so counting at a lower level (e.g. if we consider labelled graphs) is superfluous and more complicated.

4.1 *Invariant graph-theoretic objects at the level of skeletons*

For a given pair of skeleton graphs, irrespectively of how they are "decorated" with labels, we can identify different useful notions of how the labelled graphs intersect, and group the two equivalence classes in subsets of pairs accordingly.

MATCHING OF NODES Consider two skeletons $G^{(1)}=(V^{(1)},E^{(1)}),G^{(2)}=(V^{(2)},E^{(2)})$. We write $\boldsymbol{M}\equiv\boldsymbol{M}(G^{(1)},G^{(2)})$ for a set of pairs of nodes $(v^{(1)},v^{(2)})\in V^{(1)}\times V^{(2)}$ where no node in $V^{(1)}$ or $V^{(2)}$ appears twice. Write $\mathcal{M}$ for the set of all possible matchings of nodes.

Remark 4.10 (Size)**.** *Naturally, the size of a matching is the number of vertices it fixes, i.e. the size of the subset of $V^{(1)}$ taken, which is equal to the size of the subset of $V^{(2)}$ taken.*

PAIRINGS SET For a generic $\boldsymbol{M}\in\mathcal{M}$:

$$\Pi(\boldsymbol{M}):=\left\{\pi^{(1)}\in\Pi_{V^{(1)}},\pi^{(2)}\in\Pi_{V^{(2)}}:\forall(v^{(1)},v^{(2)})\in V^{(1)}\times V^{(1)},\{\pi^{(1)}(v^{(1)})=\pi^{(2)}(v^{(2)})\}\iff\{(v^{(1)},v^{(2)})\in\boldsymbol{M}\}\right\}, \tag{4.11}$$

which is in words the set of pairs of labellings of two skeletons that make the labelled graphs match through $\boldsymbol{M}$.

SYMMETRIC DIFFERENCE GRAPH For any $\left(\pi^{(1)},\pi^{(2)}\right)\in\Pi(\boldsymbol{M})$ the "overlap" of labelled graphs is constant and equal to $\boldsymbol{M}$. For this reason, we define $G_\triangle\equiv G_\triangle(\boldsymbol{M},G^{(1)},G^{(2)})=(V_\triangle,E_\triangle)$. The graph $G_\triangle$ is the symmetric difference graph associated to $(G^{(1)},G^{(2)},\boldsymbol{M})$. To construct it, we take the symmetric difference of edges in both graphs and the vertices in such set. Alternatively, we join $G^{(1)}$ and $G^{(2)}$ according to the matching $\boldsymbol{M}$, in the sense that we merge the vertices present in the pairs in $\boldsymbol{M}$. Then, we remove edges that are present in both graphs, and isolated nodes. Notice how this definition generalizes equation 3.17.

[10] In other words: since we do not consider isolated vertices the edge set induces the vertices in its connections.

We write $\#\mathrm{CC} \equiv \#\mathrm{CC}(\boldsymbol{M})$ for the number of connected components in $G_\triangle$, and $\#\mathrm{CC}_{\mathrm{pure}} \equiv \#\mathrm{CC}_{\mathrm{pure}}(\boldsymbol{M})$ for the number of connected components in $G_\triangle$ that are composed of nodes exclusively in either $G^{(1)}$, or $G^{(2)}$. This is the number of connected components that are "untouched" from the matching process.

TYPES OF VERTICES IN A MATCHING For a matching $\boldsymbol{M}$, there are vertices that are left out. We denote these as $U^{(1)}(\boldsymbol{M})$, resp. $U^{(2)}(\boldsymbol{M})$. We term them **unmatched** nodes. In words, they are the sets of nodes in $G^{(1)}$, resp. $G^{(2)}$ that are not matched. We have that:

$$\begin{aligned} U^{(1)}(\boldsymbol{M}) &:= V^{(1)} \backslash \{w \in V^{(1)} : \exists (v^{(1)}, v^{(2)}) \in \boldsymbol{M} \text{ s.t. } w = v^{(1)}\}, \\ U^{(2)}(\boldsymbol{M}) &:= V^{(2)} \backslash \{w \in V^{(2)} : \exists (v^{(1)}, v^{(2)}) \in \boldsymbol{M} \text{ s.t. } w = v^{(2)}\}. \end{aligned} \tag{4.12}$$

The connection between matched and unmatched nodes is, for $i \in \{1, 2\}$:

$$\left|V^{(i)}\right| - |\boldsymbol{M}| = \left|U^{(i)}\right|. \tag{4.13}$$

There are two main types of matched nodes. For the set of pairs nodes in $G^{(1)}$, resp. $G^{(2)}$ that are matched but that are adjacent to a node in $U^{(1)}$, resp. $U^{(2)}$ we define:

$$\boldsymbol{M}_{\mathrm{SM}}(\boldsymbol{M}) := \{(v^{(1)}, v^{(2)}) \in \boldsymbol{M} : v^{(1)} \text{adjacent to a node in } U^{(1)} \text{ or } v^{(2)} \text{adjacent to a node in } U^{(2)}\}. \tag{4.14}$$

This is the set of **semi-matched nodes (SM)**. The remaining pairs of nodes $\boldsymbol{M} \backslash \boldsymbol{M}_{\mathrm{SM}}(\boldsymbol{M})$ are said to be **perfectly matched (PM)** as they are only incident to matched nodes.
Using this distinction we construct an object that is crucial for our proof technique.

SHADOW MATCHINGS Consider two sets of unmatched nodes $\overline{U}^{(1)} \subset V^{(1)}, \overline{U}^{(2)} \subset V^{(2)}$ and a set of node matches $\underline{\boldsymbol{M}} \subset \mathcal{M}$. We define the set of shadow matchings of this triplet as:

$$\mathcal{M}_{\mathrm{shadow}}(\overline{U}_1, \overline{U}_2, \underline{\boldsymbol{M}}) := \left\{\boldsymbol{M}' \in \mathcal{M} : \;\; U^{(1)}(\boldsymbol{M}') = \overline{U}^{(1)},\; U^{(2)}(\boldsymbol{M}') = \overline{U}^{(2)},\; \boldsymbol{M}_{\mathrm{SM}}(\boldsymbol{M}') = \underline{\boldsymbol{M}}\right\}. \tag{4.15}$$

In words, it is the set of all matchings that lead to the set $\underline{\boldsymbol{M}}$ of semi-matched nodes and to the sets $\overline{U}_1, \overline{U}_2$ of unmatched nodes in resp. $G^{(1)}, G^{(2)}$. We say that these matchings satisfy a given **shadow** $(\overline{U}_1, \overline{U}_2, \underline{\boldsymbol{M}})$. The only thing that can vary between two elements of $\mathcal{M}_{\mathrm{shadow}}(\overline{U}_1, \overline{U}_2, \underline{\boldsymbol{M}})$ is the matching of the nodes that are not in $\overline{U}_1, \overline{U}_2$, or part of a pair of nodes in $\underline{\boldsymbol{M}}$. This matching must ensure that all of these nodes are perfectly matched.

TYPES OF MATCHINGS It is also useful to classify matchings further. We will sometimes mention the set $\mathcal{M}_{\mathrm{PM}}$ of matchings in $\mathcal{M}$ such that the nodes in V are **perfectly matched** in the sense that it leads to $G_\triangle$ being the empty graph (with $E_\triangle = \varnothing$). We can have a perfect match if and only if the graphs belong to the same skeleton. Then, $\mathcal{M}_{\mathrm{PM}} \neq \varnothing$ if and only $G^{(1)}$ and $G^{(2)}$ are equal up to a labelling of the nodes, i.e. $G^{(1)} \simeq G^{(2)}$ if $G^{(1)}, G^{(2)} \in \mathcal{G}_{(\leqslant D)}$, where $\simeq$ is the equivalence relation of def. 4.2.
Moreover, we define $\mathcal{M}^\star$, the set of matchings of nodes such that each $\boldsymbol{M} \in \mathcal{M}^\star$ satisfies that all connected components of $G^{(1)}, G^{(2)}$ have at least one node present in $\boldsymbol{M}$. Note that for any $\boldsymbol{M} \in \mathcal{M}^\star$, we have

$$\#\mathrm{CC}_{\mathrm{pure}}(\boldsymbol{M}) = 0. \tag{4.16}$$

Example 4.19 shows three graphs that help visualize this crucial construction.

DISTANCE BETWEEN GRAPHS Another crucial ingredient of our proof is establishing a control of the correlation of skeletons in terms of a proper distance notion. We introduce for $G^{(1)}, G^{(2)} \in \mathcal{G}_{(\leqslant D)}$ the following distance:

$$\mathrm{d}\left(G^{(1)}, G^{(2)}\right) := \min_{\boldsymbol{M} \in \mathcal{M}} |E_\triangle|. \tag{4.17}$$

Note that if $G^{(1)} \neq G^{(2)}$, it is strictly larger than 0. Otherwise, it is 0.

Remark 4.18. *This distance is known in literature as the graph edit distance (Serratosa 2021).*

Example 4.19. *In figures 4 - 5 - 6 we discuss graphically some relevant constructions.*

Thanks to this formalism, we can present a four-pager outline of the proof idea.

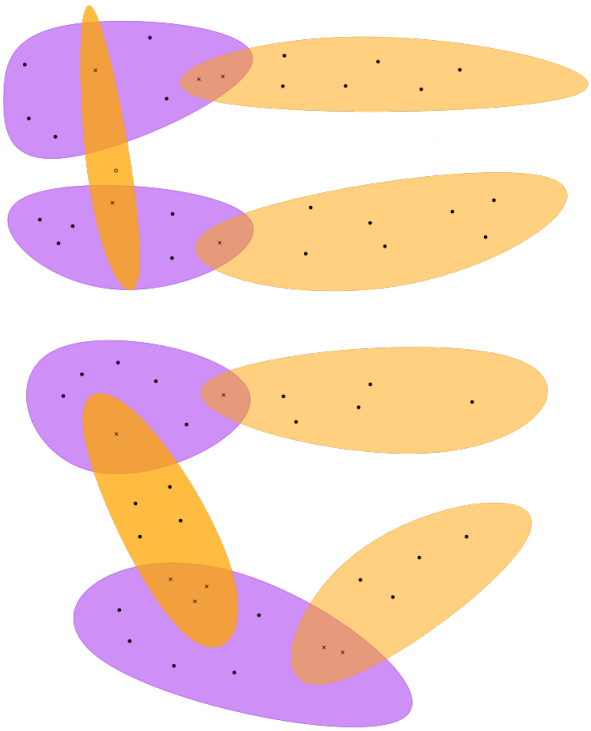


Figure 4: Matching exemplified
In this depiction, the vertex sets of two graphs $G^{(1)}, G^{(2)}$ once labelled by $\pi^{(1)}, \pi^{(2)}$ are colored respectively in purple and orange. Vertices are black dots, or black crosses if they are present in both graphs. They match according to some $\boldsymbol{M} \in \mathcal{M}^\star$, since each connected component in both graphs is impacted, and $\#\mathrm{CC}_{\mathrm{pure}} = 0$. It is important to notice that from this schematic view we cannot identify all matched vertices since they depend on the edges that our "blobs" have. We can only say that the crosses are in $\boldsymbol{M}$. For the full image, refer to fig. 6.

4.11 *Proof ingredients*

(a) FIND A SPARSE AND NICE BASIS The monomial basis from definition 2.18 is nice: the inner product of its terms is explicit. Combining equation 3.17 and assumption 3.1, we have:

$$\mathbb{E}_{\mathrm{H}_0}\left[P_{G^{(1)},\pi^{(1)}} P_{G^{(2)},\pi^{(2)}}\right] = \lambda^{|E_\triangle|}\left(\frac{k}{n}\right)^{|V_\triangle|} = o\left(1\right), \tag{4.20}$$

where $G_\triangle$ is the **symmetric difference** of the two labelled graphs. A symmetric difference graph has:

- untouched nodes $U^{(1)}, U^{(2)}$ that are not common and neither are their neighbors;
- boundary nodes, the semi-matched vertices in $\boldsymbol{M}_{\mathrm{SM}}$;
- perfectly matched pairs of nodes in $\boldsymbol{M}_{\mathrm{PM}}$.

The Gram matrix of this basis is *very dense*: all inner products are non-zero. It is difficult to study. We seek cancellations in the inner products. A good starting point is to center each term. Considering:

$$P_{G,\pi} - \mathbb{E}_{\mathrm{H}_0}\left[P_{G,\pi}\right], \tag{4.21}$$

we would notice that when the graphs of two basis elements are such that $(\pi^{(1)}, \pi^{(2)}) \in \Pi(\varnothing)$ the inner product is zero by independence. However, this is not enough, and we can exploit more independence: each connected component of the graph is a conditionally independent random variable. Then for a skeleton G with m connected components:

$$\prod_{s=1}^{m} P_{G_s,\pi} - \mathbb{E}_{\mathrm{H}_0}\left[P_{G_s,\pi}\right], \tag{4.22}$$

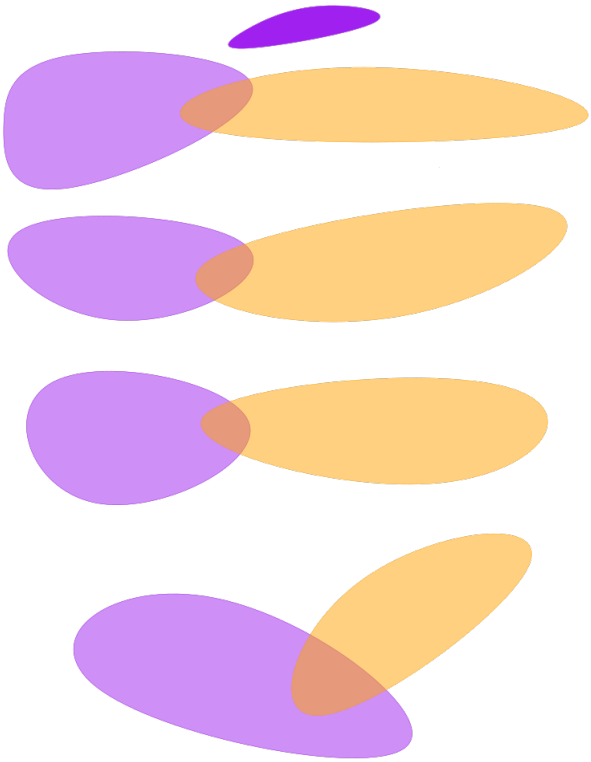


Figure 5: A matching not in $\mathcal{M}^\star$
The vertex sets of $G^{(1)}, G^{(2)}$ labelled by $\pi^{(1)}, \pi^{(2)}$ do not intersect over all connected components: there is a purple connected component that is left alone, and $\#\mathrm{CC}_{\mathrm{pure}} = 1$.

centers the canonical polynomial over a labelled graph (G, π) component by component. When we compare two labelled graphs $(G^{(1)}, \pi^{(1)})$ and $(G^{(2)}, \pi^{(2)})$, their inner product in this basis is null unless each connected component has at least a shared vertex with at least a component from the other graph. Such a property is a consequence of the fact that components are centered and independent random variables. We say graphs of this kind are *interconnected*. There are a lot fewer graphs of this type, so the Gram matrix under this basis is *sparser*. The entries are non-zero if and only if $(\pi^{(1)}, \pi^{(2)}) \in \Pi(\boldsymbol{M})$ for some $\boldsymbol{M} \in \mathcal{M}^\star$, which we defined in subsection 4.I.

(B) GROUP BY SYMMETRIES As we discussed in subsection 4.I, the issue with the label-by-label view is that the inner product of two basis elements depends on quantities that are invariant with respect to the labels $(\pi^{(1)}, \pi^{(2)})$. Namely, the correlation of the canonical monomial basis of definition 2.18 is constant across labellings $(\pi^{(1)}, \pi^{(2)}) \in \Pi(\boldsymbol{M})$. The symmetric difference of equation 3.17 is merely a result about how the *skeleton graphs* intersect, and not about where in the graph they do so. This motivates us to consider a basis of the form (see def. 5.11):

$$\overline{P}_G = \sum_{\pi \in \Pi_{|V|}} \prod_{s=1}^{m} P_{G_s,\pi} - \mathbb{E}_{\mathrm{H}_0}\left[P_{G_s,\pi}\right], \tag{4.23}$$

where $\Pi_{|V|}$ is the set of injections (labellings) of the vertices of the graph. We justified the idea in subsection 3.III. Effectively, the basis we build is invariant to permutations like the function attaining the advantage of definition 1.10:

Lemma (Lemmas 3.23 - 3.25 - 3.26 in the main text)**.** *In the planted sub-matrix model of equation 1.6 (and all models invariant by permutations), the advantage (def. 1.10) is equal to the advantage restricted to invariant functions. Therefore, we can work on a decomposition of the advantage as in equation 1.13 where the basis is a basis for invariant polynomials of degree less than D.*

The basis of equation 4.23 is *dense but smaller*: each term is non-zero, as there is always a choice of labellings $(\pi^{(1)}, \pi^{(2)}) \in \Pi(\boldsymbol{M})$ by which two skeletons $(G^{(1)}, G^{(2)})$ can match with $\boldsymbol{M} \neq \varnothing$, but the size of the basis is only the number of such skeletons over less than D edges, which is the size of the $\mathcal{G}_{(\leqslant D)}$ set from definition 3.21 (plus one including the constant function).
We want to show that this grouping gives a dense matrix where the off-diagonal terms are globally small.

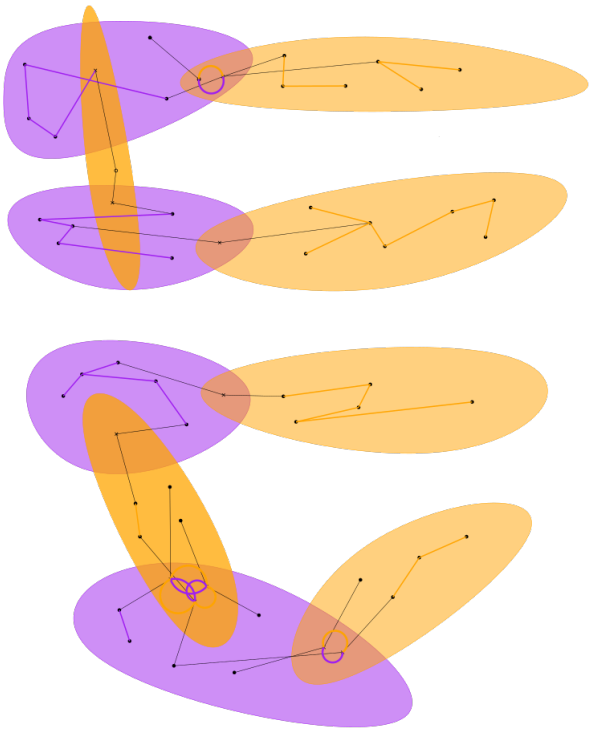


Figure 6: Full view of fig. 4
To highlight which edges are in the symmetric difference, we add in thick colored lines all connections within shared vertices in both edge sets (so that they do not appear in the symmetric difference). Then, we connect the perfectly matched vertices with the semi-matched vertices with black lines. We use purple (resp. orange) lines to form connections on the purple (resp. orange) connected components. Semi-matched nodes "bridge" perfectly matched nodes and unmatched nodes.

(c) BACK TO LABEL VS LABEL Our first key observation is that our candidate basis is non-zero a lot less frequently than the first basis; only for labelled graphs that interconnect each connected component, i.e. for which the matching is such that $\boldsymbol{M} \in \mathcal{M}^\star$. Our second key observation is that the candidate basis, when it is non-zero, has inner products up to constants being still a symmetric difference:

Proposition (Proposition 5.30 in the main text)**.** *Suppose $G^{(1)}, G^{(2)}$ are skeletons with labellings $\left(\pi^{(1)}, \pi^{(2)}\right) \in \Pi(\boldsymbol{M})$. Let assumption 3.1 hold. Then:*

- *if $\boldsymbol{M} \notin \mathcal{M}^\star$ we have $\mathbb{E}_{\mathrm{H}_0}\left[\overline{P}_{G^{(1)},\pi^{(1)}}\overline{P}_{G^{(2)},\pi^{(2)}}\right] = 0$;*
- *if $\boldsymbol{M} \in \mathcal{M}_{\mathrm{PM}}$ we have:*

$$\left|\frac{\mathbb{E}_{\mathrm{H}_0}\left[\overline{P}_{G^{(1)},\pi^{(1)}}\overline{P}_{G^{(2)},\pi^{(2)}}\right] - \mathbb{E}_{\mathrm{H}_0}\left[P_{G^{(1)},\pi^{(1)}}P_{G^{(2)},\pi^{(2)}}\right]}{\mathbb{E}_{\mathrm{H}_0}\left[P_{G^{(1)},\pi^{(1)}}P_{G^{(2)},\pi^{(2)}}\right]}\right| = o\left(1\right), \tag{4.24}$$

- *if $\boldsymbol{M} \in \mathcal{M}^\star$ it holds that:*

$$\left|\mathbb{E}_{\mathrm{H}_0}\left[\overline{P}_{G^{(1)},\pi^{(1)}}\overline{P}_{G^{(2)},\pi^{(2)}}\right]\right| \leqslant (1 + o\left(1\right))\mathbb{E}_{\mathrm{H}_0}\left[P_{G^{(1)},\pi^{(1)}}P_{G^{(2)},\pi^{(2)}}\right], \tag{4.25}$$

Proof. Use the binomial theorem and the key property of lemma 5.27:

$$\mathbb{E}_{\mathrm{H}_0}\left[P_{G^{(1)},\pi^{(1)}}\right]\mathbb{E}_{\mathrm{H}_0}\left[P_{G^{(2)},\pi^{(2)}}\right] \leqslant \left(\frac{k}{n}\right)^{|\boldsymbol{M}|}\mathbb{E}_{\mathrm{H}_0}\left[P_{G^{(1)},\pi^{(1)}}P_{G^{(2)},\pi^{(2)}}\right], \tag{4.26}$$

where $\boldsymbol{M}$ is the matching of $(\pi^{(1)}, \pi^{(2)})$. □

Then, we morally have a lot less symmetric differences contributions to sum over labellings.

(C)-BIS GRAPH VS GRAPH This leads to two key results. First, we can normalize the basis. There is a dominating term in the variance, i.e. such that $\mathbb{E}_{H_0}\left[\overline{P}_G^2\right] \asymp \nu(G)$, which is definition 5.40

$$\nu(G) = \frac{n!}{(n-|V|)!}|\mathrm{Aut}(G)|. \tag{4.27}$$

Now that we can normalize the basis, we look at the rescaled covariances. Our second key result is that the basis of equation 4.23 normalized by $1/\sqrt{\nu(G)}$ has a quantitative bound in the off-diagonal terms:

Proposition (Proposition 5.30 in the main text)**.** *Consider the $\overline{P}$ basis of equation 4.23 rescaled by $1/\sqrt{\nu(G)}$. Let assumption 3.1 hold. Then if D diverges with n:*

$$\mathrm{Var}_{H_0}\left[\frac{1}{\sqrt{\nu(G)}}\overline{P}_G\right] \asymp 1, \tag{4.28}$$

Moreover, if $G^{(1)} \neq G^{(2)}$ are two skeletons:

$$\left|\mathrm{CoV}_{H_0}\left[\frac{1}{\sqrt{\nu(G^{(1)})\nu(G^{(2)})}}\overline{P}_{G^{(1)}}\overline{P}_{G^{(2)}}\right]\right| \lesssim D^{-4c_{\mathrm{si}}\,\mathrm{d}(G^{(1)},G^{(2)})}, \tag{4.29}$$

Here d *is the edit distance of the two labelled graphs: the more graphs are far away, the less they correlate.*
We need a quantitative control since we will show later that the sum over graphs is negligible, so we want to establish later that all terms are smaller than the cardinality of the sum.

Proof. The dominating factor $\nu(G)$ is the correlation over perfect matchings: these are never negligible, but they appear if and only if $G^{(1)} \simeq G^{(2)}$ are isomorphic skeletons (def. 4.2) and they match perfectly when labelled. Therefore, it is present only when we compute the variance. Whether we compute the variance or the covariance, we have an additional correction term we want to bound:

$$\begin{aligned}
\mathsf{correct} :=& \sum_{\boldsymbol{M}\in\mathcal{M}^\star\setminus\mathcal{M}_{\mathrm{PM}}}\sum_{(\pi^{(1)},\pi^{(2)})\in\Pi(\boldsymbol{M})}\frac{1}{\sqrt{\nu(G^{(1)})\nu(G^{(2)})}}\mathbb{E}_{H_0}\left[\overline{P}_{G^{(1)},\pi^{(1)}}\overline{P}_{G^{(2)},\pi^{(2)}}\right],\\
\lesssim& \frac{1}{\sqrt{\left|\mathrm{Aut}(G^{(1)})\right|\left|\mathrm{Aut}(G^{(2)})\right|}}\sum_{\boldsymbol{M}\in\mathcal{M}^\star\setminus\mathcal{M}_{\mathrm{PM}}} n^{|U^{(1)}|+|U^{(2)}|/2}\lambda^{|E_\triangle|}\left(\frac{k}{n}\right)^{|V_\triangle|}\\
=& \frac{1}{\sqrt{\left|\mathrm{Aut}(G^{(1)})\right|\left|\mathrm{Aut}(G^{(2)})\right|}}\sum_{\boldsymbol{M}\in\mathcal{M}^\star\setminus\mathcal{M}_{\mathrm{PM}}}\left(\frac{\lambda k}{\sqrt{n}}\right)^{|U^{(1)}|+|U^{(2)}|}\lambda^{|E_\triangle|-|U^{(1)}|-|U^{(2)}|}\left(\frac{k}{n}\right)^{|V_\triangle|-|U^{(1)}|-|U^{(2)}|}.
\end{aligned} \tag{4.30}$$

The final expression comes from an application of the proposition in step (C), a bound on counting labellings, and using the fact that the inner product in the old basis (eqn. 4.20) is invariant for fixed $\boldsymbol{M}$. In the second step, we grouped terms nicely to apply our assumption 3.1. Working on the fact that graphs have no isolated nodes, we showed in lemma 4.59 that:

$$|E_\triangle| \geqslant |V_\triangle| - \#\mathrm{CC}, \tag{4.31}$$

namely that the symmetric difference is at least a forest of connected components. Combining some inequalities & identities from lemma 4.59 we also establish:

$$|E_\triangle| - |U^{(1)}| - |U^{(2)}| \geqslant \max\left\{0,\, \mathrm{d}(G^{(1)},G^{(2)}) - |U^{(1)}| - |U^{(2)}|, 1 - |U^{(1)}| - |U^{(2)}|\right\} \quad \text{since } |E_\triangle| \geqslant 1; \tag{4.32}$$

$$|\boldsymbol{M}_{\mathrm{SM}}| - \#\mathrm{CC} \geqslant 0 \tag{4.33}$$

$$|V_\triangle| - |U^{(1)}| - |U^{(2)}| = |\boldsymbol{M}_{\mathrm{SM}}|. \tag{4.34}$$

With these in mind, we are ready to show that equation 4.30 is vanishing. Using the signal condition from assumption 3.1:

$$\max\left\{\frac{\lambda k}{\sqrt{n}}, \frac{k}{n}, \lambda\right\} \leqslant D^{-8c_{\mathrm{si}}}, \tag{4.35}$$

for some large constant $c_{\mathrm{si}} > 0$, we can collect all powers inside a single term:

$$\mathsf{correct} \lesssim \frac{1}{\sqrt{|\mathrm{Aut}(G^{(1)})||\mathrm{Aut}(G^{(2)})|}}\sum_{\boldsymbol{M}\in\mathcal{M}^\star\setminus\mathcal{M}_{\mathrm{PM}}}\left(D^{-8c_{\mathrm{si}}}\right)^{\mathsf{pow}} \tag{4.36}$$

$$\mathsf{pow} := |U^{(1)}| + |U^{(2)}| + \max\left\{0,\, \mathrm{d}(G^{(1)},G^{(2)}) - |U^{(1)}| - |U^{(2)}|, 1 - |U^{(1)}| - |U^{(2)}|\right\} + |\boldsymbol{M}_{\mathrm{SM}}|. \tag{4.37}$$

Discussing the cases in the maximum some algebra shows that:

$$-8_{c_{\mathrm{si}}}\mathsf{pow} \leqslant -8c_{\mathrm{si}} \max\left\{ \mathrm{d}(G^{(1)}, G^{(2)}), 1, |U^{(1)}| + |U^{(2)}| + |\boldsymbol{M}_{\mathrm{SM}}| \right\}. \tag{4.38}$$

In practice the only dependence we need to control is through the number of untouched vertices $|U^{(1)}|, |U^{(2)}|$ and the number of semi-matched pairs of vertices $|\boldsymbol{M}_{\mathrm{SM}}|$. Such a triplet gives rise to a "shadow matching" from subsection 4.I. Therefore, grouping by sizes and using the bound on shadow matchings from lemma 4.64:

$$\begin{aligned}
\mathsf{correct} &\lesssim {}^{1}\!/\!\sqrt{|\mathrm{Aut}(G^{(1)})||\mathrm{Aut}(G^{(2)})|} \sum_{\boldsymbol{M}\in\mathcal{M}^{\star}\setminus\mathcal{M}_{\mathrm{PM}}} D^{-8c_{\mathrm{si}}\max\{\mathrm{d}(G^{(1)},G^{(2)}),1,|U^{(1)}|+|U^{(2)}|+|\boldsymbol{M}_{\mathrm{SM}}|\}} \\
&\leqslant \sum_{(U^{(1)},U^{(2)},\underline{\boldsymbol{M}}) \text{ triplets}} D^{-8c_{\mathrm{si}}\max\{\mathrm{d}(G^{(1)},G^{(2)}),1,|U^{(1)}|+|U^{(2)}|+|\boldsymbol{M}_{\mathrm{SM}}|\}} \\
&\leqslant \sum_{\substack{1\leqslant u_1\leqslant 2D\\ 1\leqslant u_2\leqslant 2D\\ 1\leqslant s\leqslant 2D}} N_{u_1,u_2,s}\cdot D^{-8c_{\mathrm{si}}\max\{\mathrm{d}(G^{(1)},G^{(2)}),1,u_1+u_2+s\}}.
\end{aligned} \tag{4.39}$$

It remains to count the number of triplets at the level of skeletons, i.e. $N_{u_1,u_2,s}$. For graphs over less than D edges, and so less than $2D$ vertices, it is certainly less than $(2D)^{u_1+u_2+2s}$. Plugging it inside:

$$\mathsf{correct} \lesssim \sum_{\substack{1\leqslant u_1\leqslant 2D\\ 1\leqslant u_2\leqslant 2D\\ 1\leqslant s\leqslant 2D}} (2D)^{u_1+u_2+2s} D^{-8c_{\mathrm{si}}\max\{\mathrm{d}(G^{(1)},G^{(2)}),1,u_1+u_2+s\}}, \tag{4.40}$$

For a fairly large constant we can counter the exploding $(2D)^{u_1+u_2+2s} \leqslant D^{2(u_1+u_2+2s)}$ factor and get that $\mathsf{correct} \leqslant D^{-4c_{\mathrm{si}}\max\{\mathrm{d}(G^{(1)},G^{(2)}),1\}}$. □

(D) ACROSS GRAPHS We constructed a Gram matrix where the diagonal is unity and the off-diagonals are small in a controlled way. It remains to show that this control is just enough. Noticing that there are less than $(d+D)^{2d} \leqslant (D)^{4d}$ graphs over less than D edges at distance d from a given one (this is lemma 4.74) for c large enough the co-variances are globally smaller than unity, and the eigenvalues of the Gram matrix are finely controlled around 1 when D diverges. Mathematically for all $G^{(1)}$ skeletons:

Proposition (Proposition 5.82 in the main text)**.** *This is a specialized version of proposition 3.3 for the basis of definition 5.42.*
Suppose assumption 3.1 holds and $D \equiv D(n) \to \infty$ as $n \to \infty$. The $\overline{P}$ basis rescaled by ${}^{1}\!/\!\sqrt{\nu(G)}$ decorated with the unit function (which is the $\widetilde{P}_G$ basis of def. 5.42) is almost orthonormal in the sense of definition 1.14.

Proof. By the discussion of subsection 3.IV, we just need to show that the Gram matrix of correlations of the basis is an approximate isometry. With the proposition of step (C)-BIS, and the Gershgorin view of bounding the off diagonals (from subsec. 3.IV), we just need to count skeletons (denoting the empty graph as the unit function):

$$\sum_{G^{(2)}\neq G^{(1)}, G^{(2)}\in\mathcal{G}_{(\leqslant D)}\cup\{\varnothing\}} \left|\mathrm{CoV}_{\mathrm{H}_0}\left[{}^{1}\!/\!\sqrt{\nu(G^{(1)})}\overline{P}^{(1)}_{G}, {}^{1}\!/\!\sqrt{\nu(G^{(2)})}\overline{P}_{G^{(2)}}\right]\right| \leqslant \sum_{d=1}^{D} D^{4d} D^{-4c_{\mathrm{si}}\max\{d,1\}} \leqslant D^{-c} = o(1). \tag{4.41}$$

Letting $\boldsymbol{\psi} := ({}^{1}\!/\!\sqrt{\nu(G)}\overline{P}_G)_{G\in\mathcal{G}_{(\leqslant D)}\cup\{\varnothing\}}$ where $\varnothing$ denotes the empty graph/constant function this implies that the eigenvalues of $\mathbb{E}_{\mathrm{H}_0}\left[\boldsymbol{\psi}\boldsymbol{\psi}^{\top}\right]$ are all within $1 \pm o(1)$ as $n \to \infty$, and the claim follows. □

(E) ON THE ADVANTAGE Since the eigenvalues are in a vanishing window around unity, we have an approximate isometry by decomposing along the basis of equation 4.23 rescaled. Such a basis is an almost orthornormal basis in the sense of definition 1.14. In equations, it holds that all coefficient decompositions of polynomials of degree less than D satisfy $\|\boldsymbol{\alpha}\|_{\mathbb{E}_{\mathrm{H}_0}[\boldsymbol{\psi}\boldsymbol{\psi}^{\top}]} \asymp \|\boldsymbol{\alpha}\|_2$. Thanks to step (D), the proof of theorem 3.8 is a matter of applying the technique for orthonormal bases of a detection problem (prob. 1.1), since the advantage simplifies to a linear representation of the form of equation 3.38. We get to the final result:

Theorem (Theorem 3.8 in the main text)**.** *Suppose assumptions 3.1 - 3.7 hold for the parameters of the complex testing problem (prob. 1.3) of the planted sub-matrix model (eqn. 1.6). Then if $D \equiv D(n) \to \infty$ as $n \to \infty$:*

$$\mathrm{Adv}_{(\leqslant D)}(\mathrm{H}_0, \mathrm{H}_1) \leqslant 1 + o(1). \tag{4.42}$$

In the next subsection we establish properties of the graph-theoretic objects from subsection 4.I. We use them throughout the main arguments.

4.III *Key graph properties*

Lemma 4.43. *Suppose $G^{(1)}, G^{(2)}$ are two skeleton graphs. Let $G^{(1)}$ have ℓ vertices and $G^{(2)}$ have s vertices. The number of matchings, of size $|\boldsymbol{M}| = q$ is:*

$$\binom{\ell}{q}\binom{s}{q}q!. \tag{4.44}$$

Therefore, the number of matchings is:

$$\sum_{q=1}^{\min\{\ell,s\}}\binom{\ell}{q}\binom{s}{q}q!. \tag{4.45}$$

Proof. A matching of size q is nothing but a pairing of subsets of $V^{(1)}, V^{(2)}$. For this, we choose q vertices out of the ℓ possible in $V^{(1)}$, then q vertices out of the s possible in $V^{(2)}$. Ultimately, we pair them in $q!$ ways, since a pairing is just an injective function from the sampled subset of $V^{(1)}$ to the sampled subset of $V^{(2)}$. □

When considering the symmetrized monomials P_G (see eqn. 3.24) we sum over all injections $\pi \in \Pi_{|V|}$, so we work at the level of all labellings of a skeleton. It is important to know the number of graphs that a skeleton groups.

Lemma 4.46. *Let G be a skeleton over ℓ vertices. Then in the current choice clarified in remark 4.7 it represents $n!/(n-\ell)!$ labelled graphs if we sum over it. If $n \gg \ell$ then $\#\{\pi : \pi(G) \simeq G\} \sim n^{\ell}$.*

Proof. As per remark 4.7 we group inside the possible ways to label G all injections $\pi \in \Pi_{\ell}$. Then number of injections from ℓ vertices to n destinations is $n!/(n-\ell)!$. The asymptotic is standard. □

Remark 4.47. *The size of the automorphism group (def. 4.4) of a graph depends strongly on its edges. Different graphs have different symmetries. It is highly non-trivial. By grouping **all** injections in the symmetrization P_G of equation 3.24 for the basis of monomials $P_{G,\pi}$ (def. 2.18) we avoid having to consider the automorphism group at the level of counting how many graphs are inside a skeleton.*

While for each skeleton we include the contribution of the automorphism group by considering all injections, it is important to know its magnitude. This is especially because the way in which we overlook the automorphism group will make it pop up later.

Lemma 4.48. *For any non-empty graph $G = (V, E)$ over ℓ vertices, we have $|\mathrm{Aut}(G)| \leqslant \ell!$.*

Proof. The most symmetric scenario is when the graph is a clique $G = \mathscr{K}_{\ell}$, so that the connectivity of each vertex is maximized. By definition 4.4, the automorphism group of a graph is the set of permutations that preserve the connections. In a clique, any permutation returns the same graph, so $|\mathrm{Aut}(\mathscr{K}_{\ell})| = \ell!$ is maximal. □

Lemma 4.49. *If $\pi^{(1)}(G), \pi^{(2)}(G)$ are labellings of a skeleton G then they have the same number of connected components.*

Proof. Both graphs are isomorphic to G, so there are labellings $\pi^{(1)}, \pi^{(2)}$ of G such that $\pi^{(i)}(G) \simeq G$ for $i \in \{1, 2\}$. The graphs $\pi^{(1)}(G), \pi^{(2)}(G)$ have the same number of isolated vertices contributing in the same way to the total number of connected components. We ignore them. Thus, there is a bijection φ between $V^{(1)} \backslash V^{(1)}_{\mathrm{iso}}$ and $V^{(2)} \backslash V^{(2)}_{\mathrm{iso}}$. If $i \overset{\gamma}{\leftrightarrow} j$ for γ a path, then $\varphi(i) \overset{\gamma}{\leftrightarrow} \varphi(j)$ and vice versa for φ^{-1} the inverse of the bijection. Therefore, the relation $i \in \nu_j \iff j \in \nu_i$ in $G^{(1)}$ translates into the relation $\gamma(j) \in \nu_{\gamma(i)} \iff \gamma(i) \in \nu_{\gamma(j)}$ and the number of connected components is preserved.
Alternatively, we could show that $G^{(1)}$ has the same number of connected components of G and the same for $G^{(2)}$, working on the labellings. □

Lemma 4.50. *Let $G^{(1)}, G^{(2)}$ be given skeletons in $\mathcal{G}_{(\leqslant D)}$. It holds that:*

$$|\mathcal{M}_{\mathrm{PM}}| = \left|\mathrm{Aut}\left(G^{(1)}\right)\right| \mathbb{1}\left\{G^{(1)} = G^{(2)}\right\}, \tag{4.51}$$

and if $G^{(1)} \simeq G^{(2)} \simeq G = (V, E)$ for all $\boldsymbol{M} \in \mathcal{M}_{PM}$ we have also:

$$|\Pi(\boldsymbol{M})| = \frac{n!}{(n - |V|)!}. \tag{4.52}$$

Proof. **(perfect matchings)** Different skeletons do not admit a perfect matching. For $G^{(1)} \simeq G^{(2)}$ the number of perfect matchings is the size of the automorphism group by definition. **(labellings of a perfect matching)** For $G^{(1)} \simeq G^{(2)} \simeq G$ the number of pairs of labellings that gives a perfect matching is the number of labellings of the graph G. For $|V|$ vertices, this is the number of injections from $|V|$ to n, i.e. $|\Pi_{|V|}| = \frac{n!}{(n-|V|)!}$. □

Lemma 4.53. *The number of pairs of $(\pi^{(1)}, \pi^{(2)}) \in \Pi_{V^{(1)}} \times \Pi_{V^{(2)}}$ giving rise to the matching $\boldsymbol{M} \subset V^{(1)} \times V^{(2)}$ satisfies*

$$|\Pi(\boldsymbol{M})| = \frac{n!}{(n - (|V^{(1)}| + |V^{(2)}| - |\boldsymbol{M}|))!}. \tag{4.54}$$

Proof. If the graphs and their matching are fixed, it remains to choose the vertices that are not labeled. There are $\binom{n}{|V^{(1)}|+|V^{(2)}|-|\boldsymbol{M}|}$ ways to do so. Then, we correct by the permutations of vertices, which are $|V^{(1)}| + |V^{(2)}| - |\boldsymbol{M}|$. The claim follows.
Alternatively, it is the number of injections from the full space of vertices of dimension n to the size of the vertices to take, of dimension $|V^{(1)}| + |V^{(2)}| - |\boldsymbol{M}|$. □

Lemma 4.55. *We have:*

$$\frac{\sqrt{(n - |V^{(1)}|)!(n - |V^{(2)}|)!}}{(n - (|V^{(1)}| + |V^{(2)}| - |\boldsymbol{M}|))!} \leqslant n^{(|V^{(1)}|+|V^{(2)}|)/2-|\boldsymbol{M}|}. \tag{4.56}$$

Proof. Stirling's formula on the LHS returns the RHS. For a non-asymptotic bound, we use $|\boldsymbol{M}| \leqslant \min\left\{|V^{(1)}|, |V^{(2)}|\right\}$ and $|V^{(1)}| + |V^{(2)}| = \min\left\{|V^{(1)}|, |V^{(2)}|\right\} + \max\left\{|V^{(1)}|, |V^{(2)}|\right\}$. Let us lighten notation using a, b for the sizes of the vertex sets and m for the size of the matching. We have:

$$\frac{\sqrt{(n-a)!(n-b)!}}{(n-(a+b-m))!} = \sqrt{\frac{(n-a)!}{(n-(a+b-m))!}}\sqrt{\frac{(n-b)!}{(n-(a+b-m))!}} = \sqrt{\prod_{j=a}^{a+b-m}(n-j)}\sqrt{\prod_{j=b}^{a+b-m}(n-j)!}. \tag{4.57}$$

The condition $m \leqslant a \wedge b$ ensures that the numerator is larger than the denominator in both fractions. As $n - j \leqslant n$ for all j in the products:

$$\sqrt{\prod_{j=a}^{a+b-m}(n-j)}\sqrt{\prod_{j=b}^{a+b-m}(n-j)!} \leqslant \sqrt{n^{a-m}}\sqrt{n^{b-m}} = n^{\frac{a+b}{2}-m}. \tag{4.58}$$

□

Lemma 4.59. *Recall the construction of invariant objects of subsection 4.I. In particular, the symmetric difference graph $G_\triangle$, with its connected components $\#\mathrm{CC}, \#\mathrm{CC}_{\mathrm{pure}}$, the unmatched vertex sets $U^{(1)}, U^{(2)}$, semi and perfectly matched vertices $\boldsymbol{M}_{\mathrm{SM}}, \boldsymbol{M}_{\mathrm{PM}}$ and the edit distance between graphs $\mathrm{d}(\cdot,\cdot)$.*
We have for any $\boldsymbol{M} \in \mathcal{M}$:

$$\left|V_\triangle\right| = \left|U^{(1)}\right| + \left|U^{(2)}\right| + |\boldsymbol{M}_{\mathrm{SM}}|, \tag{4.60}$$

$$\left|E_\triangle\right| \geqslant \max\left\{\left|V_\triangle\right| - \#\mathrm{CC},\, \mathrm{d}\left(G^{(1)}, G^{(2)}\right)\right\}, \tag{4.61}$$

$$|\boldsymbol{M}_{\mathrm{SM}}| \geqslant \#\mathrm{CC} - \#\mathrm{CC}_{\mathrm{pure}}, \tag{4.62}$$

and

$$\left|V_\triangle\right| \geqslant 2\#\mathrm{CC}. \tag{4.63}$$

Proof. Consider two skeletons $G^{(1)}, G^{(2)}$ with labellings $(\pi^{(1)}, \pi^{(2)})$. Suppose $(\pi^{(1)}, \pi^{(2)}) \in \Pi(\boldsymbol{M})$.
(claim #1) The vertices in the symmetric difference $V_\triangle$ are of two types:

- those that are unmatched and their neighbors are unmatched;

- those that are in the symmetric difference because they "bridge" unmatched nodes with matched nodes.

In other words, the former are vertices for which the incident edges come only form either $\pi^{(1)}$ or $\pi^{(2)}$, not both. Equivalently, they are those forming the $U^{(1)}, U^{(2)}$ sets. The latter are vertices that have, say, incident edges only from $\pi^{(1)}$ and some incident edges that are shared. Through the edges only in $\pi^{(1)}$ the vertex is in the symmetric difference induced by edges. Equivalently, there exists a tuple in $\boldsymbol{M}_{\text{SM}}(\boldsymbol{M})$ where the vertex appears. The claim follows.
(claim #2) We prove that $|E_{\triangle}|$ is larger than both quantities, hence being larger than their maximum.
By definition $\mathrm{d}(G^{(1)}, G^{(2)}) = \min_{\boldsymbol{M}\in\mathcal{M}} |E'_{\triangle}|$, so all $|E_{\triangle}| \geqslant \mathrm{d}(G^{(1)}, G^{(2)})$, which verifies the first.
The edges in the symmetric difference induce the vertices in the symmetric difference. Namely, each vertex appears at least in one edge, and is not isolated. The minimal number of edges given a number of vertices corresponds to a tree, which has #vertices -1 edges. The symmetric difference graph is induced by the symmetric difference of edges. Its connected components are at most the sum of connected components of the two original graphs, since it is made of the edges that appear uniquely in either. For each of these connected components in $G_{\triangle}$, by the fact that we have no isolated nodes, there is at least a tree. Therefore, there are $|E_{\triangle}| \geqslant |V_{\triangle}| - \#CC$ edges at least. This proves the second lower bound.
(claim #3) As previously said, the symmetric difference graph has connected components arising from edges of the two labellings that appear uniquely in either. Each connected component can be pure or non-pure. For each non-pure connected component there is at least a vertex $v^{\star}$ that is shared by both labellings. Such vertex is necessarily in a tuple in $\boldsymbol{M}_{\text{SM}}$ by its definition. Therefore, to each connected component that is non-pure there is at least one semi-matched vertex. The claim follows.
(claim #4) The vertices in the symmetric difference $G_{\triangle}$ are induced by the edges in the symmetric difference $E_{\triangle}$. Such edges will form #CC connected components, each induced by the relative edge sets that partition $E_{\triangle}$. Since to each connected component there corresponds at least an edge, there correspond at least two vertices. The claim follows.

□

Lemma 4.64. *We have*

$$|\mathcal{M}_{\text{shadow}}\left(\overline{U}_1, \overline{U}_2, \underline{\boldsymbol{M}}\right)| \leqslant \min\left\{\left|\mathrm{Aut}\left(G^{(1)}\right)\right|, \left|\mathrm{Aut}\left(G^{(2)}\right)\right|\right\}. \tag{4.65}$$

Proof. Throughout the proof, we use the graph-theoretic objects defined in subsection 4.I.
(#0 conventions) We introduce some proof-specific notation. Let $\underline{\boldsymbol{M}}^{(1)}$ and $\underline{\boldsymbol{M}}^{(2)}$ be the semi-matched nodes of $G^{(1)}$ and $G^{(2)}$ respectively. The sets $\boldsymbol{M}^{(1)}_{\text{PM}}, \boldsymbol{M}^{(2)}_{\text{PM}}$ are analogous. Notice that they are sets of vertices, not sets of pairs of vertices. We also define $\partial(v) \coloneqq \{v' \in V \mid ((v, v') \in E)\}$, the operator that takes a vertex and returns its neighborhood in a generic graph. We use the shorthands:

$$\{v^{(1)}, \partial(v^{(1)})\} \coloneqq \left\{(v^{(1)}, v) \mid v \in \partial(v^{(1)})\right\} \tag{4.66}$$

to denote the set of edges incident to $v^{(1)} \in V^{(1)}$, and the special subset of $\partial(v^{(1)})$

$$\partial(v^{(1)}; \text{PM}) \coloneqq \left\{w^{(1)} \in V^{(1)} \mid (v^{(1)}, w^{(1)}) \in E^{(1)}, \text{and } w^{(1)} \in \boldsymbol{M}^{(1)}_{\text{PM}}\right\}, \tag{4.67}$$

which is the neighborhood of $v^{(1)}$ of perfectly matched vertices. Lastly, a permutation $\sigma : V \to V$ applied to a subset of its domain V acts element-wise, and if it is applied to a set of edges it is applied element-wise to each entry. In equations, when we write $\sigma(E)$ we mean that $(i, j) \mapsto (\sigma(i), \sigma(j))$ for all $(i, j) \in E$. Accordingly, a permutation acts on a graph $G = (V, E)$ as $\sigma(G) = (\sigma(V), \sigma(E))$.
(#1 reduction) For the statement to be non-trivial, we need that the perfectly matched vertices in each graph are of the same number, and that there is at least one shadow matching. Otherwise, $\boldsymbol{M}_{\text{PM}} = \varnothing$. Therefore, without loss of generality let us assume there is at least one shadow matching $\boldsymbol{M} = \underline{\boldsymbol{M}} \cup \boldsymbol{M}_{\text{PM}} \in \mathcal{M}_{\text{shadow}}(\overline{U}^{(1)}, \overline{U}^{(2)}, \underline{\boldsymbol{M}})$, and that $|\boldsymbol{M}^{(1)}_{\text{PM}}| = |\boldsymbol{M}^{(2)}_{\text{PM}}|$. The set $\boldsymbol{M}_{\text{PM}}$ contains pairs of vertices from $G^{(1)}$ and $G^{(2)}$.
(#2 extended permutation) To find another shadow matching, we can only change the vertices in $\boldsymbol{M}_{\text{PM}} = \{(v_i^{(1)}, v_i^{(2)})\}$; those that are perfectly matched. The others are fixed. Since we need to return a matching, there are two possible ways in which we can explore all candidates:

- either we permute the vertices $\boldsymbol{M}^{(1)}_{\text{PM}}$ of the first graph with a permutation $\widetilde{\sigma}^{(1)} : \boldsymbol{M}^{(1)}_{\text{PM}} \mapsto \boldsymbol{M}^{(1)}_{\text{PM}}$;
- or we permute the vertices $\boldsymbol{M}_{\text{PM};2}$ of the second graph with a permutation $\widetilde{\sigma}^{(2)} : \boldsymbol{M}^{(2)}_{\text{PM}} \mapsto \boldsymbol{M}^{(2)}_{\text{PM}}$.

The permutation $\widetilde{\sigma}^{(1)}$ corresponds uniquely to a permutation $\sigma^{(1)}$ over all $V^{(1)}$ vertices which has as fixed points the unmatched and semi-matched vertices. Namely, it corresponds to $\sigma^{(1)} : V^{(1)} \mapsto V^{(1)}$ such that $\sigma^{(1)}(v_i^{(1)}) = v_i^{(1)}$ for all $v_i^{(1)}$ in $\overline{U}^{(1)}, \underline{\boldsymbol{M}}^{(1)}$. The same holds for $\widetilde{\sigma}^{(2)}$. From now on, we work on the $\sigma^{(1)}, \sigma^{(2)}$ extensions.

(#3 symmetry) We argue that not all permutations of this kind return shadow matchings. It is then sufficient to show that by permuting the nodes of the first graph in all the allowed ways we have an upper bound by $|\mathrm{Aut}(G^{(1)})|$. More explicitly, we want to show that each permutation $\sigma^{(1)}$ that is of the type described in step #2 and returns a shadow matching is an automorphism of $G^{(1)}$ (def. 4.3). The result for $G^{(2)}$ is symmetric, we omit it.

(#4 relevant partitions) By construction, the vertex and edge sets of $G^{(1)}, G^{(2)}$ admit useful partitions. For $r = 1, 2$:

$$V^{(r)} = \overline{U}^{(r)} \cup \underline{\boldsymbol{M}}^{(r)} \cup \boldsymbol{M}_{\mathrm{PM}}^{(r)} \tag{4.68}$$

$$E^{(r)} = \left[E^{(r)}\right]_{\mathrm{U,U}} \cup \left[E^{(r)}\right]_{\mathrm{U,SM}} \cup \left[E^{(r)}\right]_{\mathrm{SM,SM}} \cup \left[E^{(r)}\right]_{\mathrm{SM,PM}} \cup \left[E^{(r)}\right]_{\mathrm{PM,PM}}, \tag{4.69}$$

which are respectively edges within unmatched nodes, within an unmatched node and a semi-matched node, within semi-matched and semi-matched, within semi-matched and perfectly matched, within perfectly matched nodes. For example, there are by construction no edges $(v^{(r)}, w^{(r)})$ such that $v^{(r)} \in \overline{U}^{(r)}$ and $w^{(r)} \in \underline{\boldsymbol{M}}^{(r)}$. If we apply $\sigma^{(1)}$ go $G^{(1)}$, the decomposition is analogous but tedious to write since it depends implicitly on $\sigma^{(1)}$. Next, we show that the peculiar type of permutation required greatly simplifies the expression.

(#5 easy cases) Since $\sigma^{(1)}$ has fixed points at unmatched and semi-matched vertices by step #2, we know that $\sigma^{(1)}(\overline{U}^{(1)}) = \overline{U}^{(1)}$ and $\sigma^{(2)}(\underline{\boldsymbol{M}}^{(2)}) = \underline{\boldsymbol{M}}^{(2)}$. Therefore:

$$\left[\sigma^{(1)}(E^{(1)})\right]_{\mathrm{U,U}} = \left[E^{(1)}\right]_{U,U}, \qquad \left[\sigma^{(1)}(E^{(1)})\right]_{\mathrm{U,SM}} = \left[E^{(1)}\right]_{\mathrm{U,SM}}, \qquad \text{and} \qquad \left[\sigma^{(1)}(E^{(1)})\right]_{\mathrm{SM,SM}} = \left[E^{(r)}\right]_{\mathrm{SM,SM}}. \tag{4.70}$$

(#6 hard cases) The remaining sets from the decomposition of equation 4.69 are more complex. We treat them separately. We use the notation of step #0.

(#6.1 first term) Observing that $\sigma^{(1)}$ acts on vertices $v \in \boldsymbol{M}_{\mathrm{SM}}^{(1)}$ as an identity we make the following changes of indexing:

$$\begin{aligned} E_{\mathrm{SM,PM}}^{(1)} &= \bigcup_{v \in \boldsymbol{M}_{\mathrm{SM}}^{(1)}} \{v, \partial(v; \mathrm{PM})\} \\ &= \bigcup_{\sigma^{(1)}(v) \in \boldsymbol{M}_{\mathrm{SM}}^{(1)}} \left\{\sigma^{(1)}(v), \partial(\sigma^{(1)}(v); \mathrm{PM})\right\} \\ &= \left[\sigma^{(1)}(E^{(1)})\right]_{\mathrm{SM,PM}}, \end{aligned} \tag{4.71}$$

which holds since we know that semi-matched vertices are a fixed point of $\sigma^{(1)}$, so the first term is again the set of edges between semi-matched vertices and their respective perfectly matched neighbors.

(#6.2 second term) The second term requires to use the shadow constraint, since there are no fixed-point tricks. The set $E_{\mathrm{PM,PM}}^{(1)}$ is made of edges from the sub-graph of $G^{(1)}$ that contains only perfectly matched vertices with $G^{(2)}$. Then, we have $E_{\mathrm{PM,PM}}^{(1)} = E_{\mathrm{PM,PM}}^{(2)}$. Since $E_{\mathrm{PM,PM}}^{(2)}$ is fixed (it is unaffected by $\sigma^{(1)}$), the conclusion is that

$$\left[\sigma^{(1)}(E^{(1)})\right]_{\mathrm{PM,PM}} = \left[E^{(2)}\right]_{\mathrm{PM,PM}} = \left[E^{(1)}\right]_{\mathrm{PM,PM}}. \tag{4.72}$$

(#7 finalization) Combining equations 4.70 - 4.71 - 4.72 with the edge decomposition of equation 4.69 for the permuted graph $\sigma^{(1)}(G^{(1)})$ we find that:

$$\begin{aligned} \sigma^{(1)}(E^{(1)}) &= \left[\sigma^{(1)}(E^{(1)})\right]_{\mathrm{U,U}} \cup \left[\sigma^{(1)}(E^{(1)})\right]_{\mathrm{U,SM}} \cup \left[\sigma^{(1)}(E^{(r)})\right]_{\mathrm{SM,SM}} \cup \left[\sigma^{(1)}(E^{(1)})\right]_{\mathrm{SM,PM}} \cup \left[\sigma^{(1)}(E^{(1)})\right]_{\mathrm{PM,PM}} \\ &= \left[E^{(1)}\right]_{\mathrm{U,U}} \cup \left[E^{(1)}\right]_{\mathrm{U,SM}} \cup \left[E^{(r)}\right]_{\mathrm{SM,SM}} \cup \left[E^{(1)}\right]_{\mathrm{SM,PM}} \cup \left[E^{(1)}\right]_{\mathrm{PM,PM}}. \end{aligned} \tag{4.73}$$

This means that an edge is in $E^{(1)}$ if and only if it is in $\sigma^{(1)}(E^{(1)})$, so $\sigma^{(1)}$ is an automorphism in the sense of definition 4.3. The claim follows by the simplifications of steps #1 - #2 - #3.

□

Lemma 4.74. *Consider $G^{(1)} \in \mathcal{G}_{(\leqslant D)} \cup \{\varnothing\}$. It holds that:*

$$\left|\left\{G^{(2)} \mid \mathrm{d}(G^{(1)}, G^{(2)}) = d\right\}\right| \leqslant (d+D)^{2d}. \tag{4.75}$$

Proof. Recall that $\mathrm{d}(G^{(1)}, G^{(2)}) = \min_{\boldsymbol{M}\in\mathcal{M}} |E_{\triangle}|$. If such distance is fixed at d, we need to choose how to make the edges of $G^{(1)}, G^{(2)}$ pair into a set of size d. In $G^{(1)}$ we have $|E^{(1)}| \leqslant D$ edges, and in the symmetric difference we have d edges. In the worst case, we have $(|E^{(1)}| + d)^2 \leqslant (D+d)^2$ options for pairs of edges in the symmetric difference. Since there are d such pairs, the claim follows. □

Lemma 4.76. *Two graphs, each with one connected component, are connected if and only if they share a vertex.*

Proof. **($\Longrightarrow$)** If two connected graphs are connected, for each vertex in G there is a path γ using edges in $G^{(1)} \cup G^{(2)}$ that joins it with every vertex in $G^{(2)}$. This means there must be at least a vertex that is common to both. Indeed, having a shared edge is weaker as it means having two shared vertices, having no shared vertices means there is no path.
($\Longleftarrow$) If two graphs share a vertex and satisfy the assumptions, then they are trivially connected. □

Now that we presented all the objects we need for the basis construction, we present a guide through the basis that will be almost orthonormal. In particular, in the next section we argue that the canonical basis and a first modification do not exploit all the properties of the problem, and reach a final formulation in definition 5.11.

5 CONSTRUCTION AND PROPERTIES OF THE ALMOST ORTHONORMAL BASIS

We start with a soft motivation for our final basis proposal, inspired by the arguments in the proof sketch of subsection 4.II where we sought a sparse enough basis over labelled graphs. Throughout, we consider problem 1.3 for the planted sub-matrix model of equation 1.6, so the distribution under H_0 is such that $\lambda k \neq 0$, i.e. there is a signal in the null distribution. We will also largely use the graph-theoretic objects defined in subsection 4.I.

The $P_{G,\pi}$ basis of def. 2.18 is a basis (lem 3.16) but is **largely** not orthonormal. Let us discuss informally why. First, we provide an example to show how easy it is to make two basis elements correlate.

Example 5.1. *Consider $G^{(1)}, G^{(2)}$ skeletons and labellings $\pi^{(1)}, \pi^{(2)}$, which may also not belong to any matching $\boldsymbol{M}$. Under the planted sub-matrix model of equation 1.6, a simple computation gives that $\mathbb{E}_{\mathrm{H}_0}\left[P_{G^{(1)},\pi^{(1)}} P_{G^{(2)},\pi^{(2)}}\right] = \lambda^{|E_{\pi^{(1)}\triangle\pi^{(2)}}|} \left(k/n\right)^{|V_{\pi^{(1)}\triangle\pi^{(2)}}|}$. Such quantity is non-null for all graphs $G^{(1)} \neq G^{(2)}$ across all labellings of the two, and non-null across all non-perfect matchings o the same $G^{(1)} = G^{(2)}$ graph.*

Intuitively, the basis $\left(1, ((P_{G,\pi})_{\pi\in\Pi_{|V|}})_{G\in\mathcal{G}_{(\leqslant D)}}\right)$ is very dense, and regrouping by symmetries to construct P_G as in the idea of subsection 3.III will sum over a very dense matrix. However, there is an immediate adjustment that hints at the right adjustment. We present them next.

With the intention of sparsifying the basis at the level of labelled graphs, we can just recenter each canonical monomial from definition 2.18, and then group by symmetries. The result is a "centered basis".

Definition 5.2 (Invariant centered monomials)**.** *Let $G = (V, E)$ be a skeleton over $V = \{v_1, \dots, v_\ell\}$ vertices. Denote Π_ℓ the set of injective mappings. Define:*

$$\begin{aligned}
\widehat{P} : \mathcal{G} \times \mathbb{R}^{n\times n} &\to \mathbb{R} \\
(G = (V, E), \boldsymbol{Y}) &\mapsto \sum_{\pi\in\Pi_{|V|}} P_{G,\pi} - \mathbb{E}_{\mathrm{H}_0}\left[P_{G,\pi}\right] \\
(G, \boldsymbol{Y}) &\mapsto \sum_{\pi\in\Pi_{|V|}} \prod_{(i,j)\in E} Y_{\pi(i)\pi(j)} - \mathbb{E}_{\mathrm{H}_0}\left[\prod_{(i,j)\in E} Y_{\pi(i)\pi(j)}\right].
\end{aligned} \tag{5.3}$$

In words: we sum over the possible realizations of the skeletons a monomial arising from the product of the entries of $\mathbf{Y}$ along its edges, centered.
When there is no ambiguity, we write $\widehat{P}_G(\mathbf{Y}) = \widehat{P}_G$ for simplicity. When writing $\widehat{P}_{G,\pi}$ we consider only one term in the sum.

Lemma 5.4. *The set of invariant centered monomials decorated with the unit function* $\left(1, \left(\widehat{P}_G\right)_{G\in\mathcal{G}_{(\leqslant D)}}\right)$ *is a basis of invariant polynomials of degree less than D.*

Proof. Since we just shift the invariant basis, we reapply lemma 3.25 to conclude. □

Let us establish some algebraic properties of this object. The first moment is zero, as it is the sum of centered random variables. In equations:

$$\mathbb{E}_{\mathrm{H}_0}\left[\widehat{P}_G(\mathbf{Y})\right] = \sum_{\pi\in\Pi_{|V|}} \mathbb{E}_{\mathrm{H}_0}\left[P_{G,\pi}\right] - \mathbb{E}_{\mathrm{H}_0}\left[P_{G,\pi}\right] = 0. \tag{5.5}$$

We now move to computing the variance. The key adjustment is that when the labelled graphs will be non-overlapping then $\mathbb{E}_{\mathrm{H}_0}\left[\widehat{P}_{G,\pi^{(1)}}\widehat{P}_{G,\pi^{(2)}}\right] = 0$.

Lemma 5.6 (Relation to canonical basis)**.** *Suppose assumption 3.1 holds, or even just the weaker conditions in the proof. Consider two skeletons* $G^{(1)}, G^{(2)}$*:*

1. *if* $\boldsymbol{M} = \varnothing$ *then* $\mathbb{E}_{\mathrm{H}_0}\left[\widehat{P}_{G^{(1)},\pi^{(1)}}\widehat{P}_{G^{(2)},\pi^{(2)}}\right] = 0$*;*
2. *otherwise if D diverges with n:*

$$\mathbb{E}_{\mathrm{H}_0}\left[\widehat{P}_{G^{(1)},\pi^{(1)}}\widehat{P}_{G^{(2)},\pi^{(2)}}\right] \asymp \mathbb{E}_{\mathrm{H}_0}\left[P_{G^{(1)},\pi^{(1)}}P_{G^{(2)},\pi^{(2)}}\right], \tag{5.7}$$

and if D does not diverge we have equality up to constants.

Remark 5.8. *From here onwards, the language of matchings from section 4 is relevant.*

Proof. **(claim #1)** If $\boldsymbol{M} = \varnothing$ then the latent variables in $G^{(1)}$ and $G^{(2)}$ are independent. By this fact:

$$\mathbb{E}_{\mathrm{H}_0}\left[\widehat{P}_{G^{(1)},\pi^{(1)}}\widehat{P}_{G^{(2)},\pi^{(2)}}\right] = \mathbb{E}_{\mathrm{H}_0}\left[P_{G^{(1)},\pi^{(1)}}P_{G^{(2)},\pi^{(2)}}\right] - \mathbb{E}_{\mathrm{H}_0}\left[P_{G^{(1)},\pi^{(1)}}\right]\mathbb{E}_{\mathrm{H}_0}\left[P_{G^{(2)},\pi^{(2)}}\right] = 0, \tag{5.9}$$

since the first integral decouples.
(claim #2) When the matching is not empty, it induces correlations. In particular:

$$\begin{aligned}\mathbb{E}_{\mathrm{H}_0}\left[\widehat{P}_{G^{(1)},\pi^{(1)}}\widehat{P}_{G^{(2)},\pi^{(2)}}\right] &= \mathbb{E}_{\mathrm{H}_0}\left[P_{G,\pi^{(1)}}P_{G,\pi^{(2)}}\right] - \mathbb{E}_{\mathrm{H}_0}\left[P_{G,\pi^{(1)}}\right]\mathbb{E}_{\mathrm{H}_0}\left[P_{G,\pi^{(2)}}\right] \\ &= \lambda^{|\pi^{(1)}\triangle\pi^{(2)}|}\left(\frac{k}{n}\right)^{|V_{\pi^{(1)}\triangle\pi^{(2)}}|}\left(1-\lambda^{2|E|-|\pi^{(1)}\triangle\pi^{(2)}|}\left(\frac{k}{n}\right)^{2|V|-|V_{\pi^{(1)}\triangle\pi^{(2)}}|}\right) \\ &\lesssim \lambda^{|\pi^{(1)}\triangle\pi^{(2)}|}\left(\frac{k}{n}\right)^{|V_{\pi^{(1)}\triangle\pi^{(2)}}|},\end{aligned} \tag{5.10}$$

which is constant for all $\left(\pi^{(1)}, \pi^{(2)}\right) \in \Pi(\boldsymbol{M})$. If D diverges, as $\lambda = o(1), k/n = o(1)$ by assumption 3.1 the upper bound is also a lower bound asymptotically since the minus term in the parenthesis is a $o(1)$. By a similar reasoning, if D does not necessarily diverge, we bound up to constants since $D \geqslant 2$. □

We wonder if this is enough.

The issue for the basis[11] of definition 5.2 is two-fold. On one side, we are not using all the conditional independence of the model. On the other, it has still non-negligible cross terms in the variance once we allow for graphs that have more than one connected component. Since we want tight bounds, we need to find a proper generalization of these basis functions that:

- are null in expectation;
- have null covariance if the induced graphs are disconnected;
- have lower covariance if the induced graphs are connected.

[11] Minor aspect: we always complete with the unit function.

A solution is to center the connected components (which are independent), each by each.

In practice, we know by lem. 4.49 that the number of connected components is an invariant for graphs in a given skeleton. Then, let G be a skeleton with m connected components and ℓ vertices. In particular, for $\pi \in \Pi_\ell$ which induces the edges $\mathbb{T}$ there is a decomposition $\mathbb{T} = \mathbb{T}_1 \cup \cdots \cup \mathbb{T}_m$ into disjoint sets where each $\mathbb{T}_s, \mathbb{T}_r$ pair not only has empty intersection, but also the vertices of the graph induced are disjoint, as a consequence of lemma 4.76. Since disconnected graphs are independent upon conditioning, we have a decomposition of G into independent random realizations over sub-graphs induced by $\{G_s\}_{s=1}^m$. We use these for our corrected monomial basis.

Definition 5.11 (Corrected invariant monomials and the corrected invariant monomials basis)**.** *Let $G = (V, E)$ be a skeleton (def. 3.21) over $V = \{v_1, \ldots, v_\ell\}$ vertices. Denote Π_ℓ the set of injective mappings. Define*

$$\begin{aligned}
\overline{P} : \mathcal{G} \times \mathbb{R}^{n\times n} &\to \mathbb{R} \\
(G = (V, E), \mathbf{Y}) &\mapsto \sum_{\pi\in\Pi_{|V|}} \overline{P}_{G,\pi}(\mathbf{Y}) \\
(G, \mathbf{Y}) &\mapsto \sum_{\pi\in\Pi_{|V|}} \prod_{s=1}^{m} \widehat{P}_{G_s,\pi}(\mathbf{Y}), \\
(G, \mathbf{Y}) &\mapsto \sum_{\pi\in\Pi_{|V|}} \prod_{s=1}^{m} \prod_{(i,j)\in E_s} Y_{\pi(i)\pi(j)} - \mathbb{E}_{\mathrm{H}_0}\left[\prod_{(i,j)\in E_s} Y_{\pi(i)\pi(j)}\right]
\end{aligned} \tag{5.12}$$

where $G = (G_1, \ldots, G_m)$ is the decomposition into connected components of the graph G.
The basis of corrected invariant monomials is $\left(1, \left(\overline{P}_G\right)_{G\in\mathcal{G}_{(\leqslant D)}}\right)$.
In words: we sum over the possible realizations of the skeletons a product of monomials arising from the centered polynomial P over the connected components. Each centered polynomial is the product over the edges of said connected component.
When there is no ambiguity, we write $\overline{P}_{G,\pi}(\mathbf{Y}) = \overline{P}_{G,\pi}$.

Lemma 5.13. *The collection $\left(1, \left(\overline{P}_G\right)_{G\in\mathcal{G}_{(\leqslant D)}}\right)$ is a basis of polynomials of degree less than D with domain in $\{-1, 1\}^{n\times n}$.*

Proof. If a polynomial basis is such that each term shifts by lower degree polynomials it remains a basis. More explicitly, we reason as follows.
We use an induction argument combined with lemma 3.16. Each $\overline{P}_G$ for $G \in \mathcal{G}_{(\leqslant D)}$ is a polynomial with degree equal to the degree of P_G, since each other term removes connected components and integrates them separately. Let us express each $\overline{P}_G = P_G + \text{rest}$. By induction, graphs with one connected component span the space spanned by elements in P_G that correspond to it, since $\overline{P}_G = P_G$ in this case. Graphs with up to two connected components span the orthogonal complement of the space spanned by graphs with one connected component since each $\overline{P}_G = P_G + \text{rest}$ is such that P_G is not spanned by graphs with one connected component, and so are its linear combinations. Graphs with up to $m + 1$ connected components span the orthogonal complement of graphs with up to m connected components by the same reasoning. For each level of number of connected components, we have $\text{span}\{\overline{P}_G\} = \text{span}\{P_G\}$, and the claim follows by lemma 3.16, which states that the P_G polynomials form a basis. In particular, the "uniqueness of the representation" is more involved in notation but just analogous.[12]

□

The expectation over a single skeleton is again zero. Indeed, separate connected components are independent (the latent Bernoulli random variables are):

$$\mathbb{E}_{\mathrm{H}_0}\left[\overline{P}_G(\mathbf{Y})\right] = \sum_{\pi\in\Pi_{|V|}} \mathbb{E}_{\mathrm{H}_0}\left[\overline{P}_{G,\pi}\right] = \sum_{\pi\in\Pi_{|V|}} \prod_{s=1}^{m} \mathbb{E}_{\mathrm{H}_0}\left[\widehat{P}_{G_s,\pi}\right] = 0, \tag{5.14}$$

where in the penultimate step we used independence of sub-graphs. Since each random variable is centered, we can also say the following for the inner product of two $(G^{(1)}, \pi^{(1)}), (G^{(2)}, \pi^{(2)})$ not necessarily in the same equivalence class. Without loss of generality, let $G^{(1)} = \bigcup_{s=1}^m G_s^{(1)}$ and $G^{(2)} = \bigcup_{t=1}^r G_t^{(2)}$ decompose into m, r

[12]There can be terms where the added P_G is the same but lower degree polynomials are different, but these do not break uniqueness, since the lower degree terms by induction induce a unique representation.

connected components forming disconnected graphs, i.e. $G^{(1)}$ has m connected components and $G^{(2)}$ has r connected components. Then, suppose that there is a connected component in $G^{(1)}$, say $G^{(1)}_{s^\star}$ that induces a graph that is not connected to any of the others. In particular, it is by default not connected to any other $G^{(1)}_s$ where $s \neq s^\star$, but also not connected to any $G^{(2)}_t$ for $t \in [r]$ once labelled. As a graph, its latent Bernoulli random variables are independent of the others, and we can take out its expectation:

$$\begin{aligned}\left\langle \overline{P}_{G^{(1)},\pi^{(1)}}, \overline{P}_{G^{(2)},\pi^{(2)}} \right\rangle_{\mathrm{H}_0} &= \mathbb{E}_{\mathrm{H}_0}\left[\overline{P}_{G^{(1)},\pi^{(1)}}\overline{P}_{G^{(2)},\pi^{(2)}}\right] \\ &= \mathbb{E}_{\mathrm{H}_0}\left[\prod_{s=1}^{m}\prod_{t=1}^{r} P_{G^{(1)}_s,\pi^{(1)}} P_{G^{(2)}_t,\pi^{(2)}}\right] \\ &= \mathbb{E}_{\mathrm{H}_0}\left[\prod_{s=1}^{m}\prod_{t=1}^{r} \mathbb{1}_{\{s\neq s^\star\}} P_{G^{(1)}_s,\pi^{(1)}} P_{G^{(2)}_t,\pi^{(2)}}\right] \mathbb{E}_{\mathrm{H}_0}\left[P_{G^{(1)}_{s^\star},\pi^{(1)}}\right] \\ &= 0.\end{aligned} \tag{5.15}$$

In the language of matchings, we would say that the pair $(\pi^{(1)}, \pi^{(2)}) \in \Pi(\boldsymbol{M})$ for $\boldsymbol{M} \notin \mathcal{M}^\star$. An example of such graphs tuple is fig. 10. Thus, the alignment is non-trivial if and only if each connected component of $G^{(1)}$ is connected (shares a vertex in the induced graph) with at least one connected component of $G^{(2)}$ and vice versa. For example, see figs. 11-12. For the sake of clarity, we place this condition into a definition.

Definition 5.16 (interconnectivity)**.** *We say two graphs $\pi^{(1)}(G^{(1)}), \pi^{(2)}(G^{(2)})$ are interconnected if each connected component of $\pi^{(1)}(G^{(1)})$ shares a vertex with at least one connected component of $\pi^{(2)}(G^{(2)})$ and vice versa.*

Remark 5.17. *Consider two skeletons $G^{(1)}, G^{(2)}$. A necessary but not sufficient condition for interconnectivity of two labellings $\pi^{(1)}, \pi^{(2)}$ is that $(\pi^{(1)}, \pi^{(2)}) \in \Pi(\boldsymbol{M})$ for some $\boldsymbol{M} \in \mathcal{M}$. A necessary and sufficient condition for interconnectivity is that $(\pi^{(1)}, \pi^{(2)}) \in \Pi(\boldsymbol{M})$ for $\boldsymbol{M} \in \mathcal{M}^\star$. Indeed, the $\mathcal{M}^\star$ set was the set of matchings where each connected component is matched to at least one connected component of the other graph.*
When computing moments of the new basis, we will work on matchings $\boldsymbol{M} \in \mathcal{M}^\star$ which are the only non-trivial ones.

This observation induces a specific factoring in the integrals. Let $G^{(1)}, G^{(2)}$ be again generic with m and r connected components respectively. Suppose further that each connected component of $\pi^{(1)}(G^{(1)})$ is connected to at least one connected component of $\pi^{(2)}(G^{(2)})$ and vice versa. In other words, let they be interconnected, i.e. $(\pi^{(1)}, \pi^{(2)}) \in \Pi(\boldsymbol{M})$ for $\boldsymbol{M} \in \mathcal{M}^\star$. If we take the union of graphs $\pi^{(1)}(G^{(1)}) \cup \pi^{(2)}(G^{(2)})$, it will induce a graph $G_{\pi^{(1)}\cup\pi^{(2)}}$ with a peculiar structure: each connected component of $G_{\pi^{(1)}\cup\pi^{(2)}}$ contains vertices that belong to at least one connected component from each original graph $\pi^{(1)}(G^{(1)}), \pi^{(2)}(G^{(2)})$. By construction, these vertex sets intersect only inter-edge sets, i.e. there is no vertex that is shared by both $G^{(1)}_s, G^{(1)}_{s'}$ for $s \neq s'$ but there has to be at least one vertex shared between some $G^{(1)}_s$ and some $G^{(2)}_t$ for each tuple $(\pi^{(1)}, \pi^{(2)}) \in \Pi(\boldsymbol{M})$ with $\boldsymbol{M} \in \mathcal{M}^\star$ taken from any pair of equivalence classes. The random variables $P_{G^{(1)}_s,\pi^{(1)}}(\mathbf{Y}), P_{G^{(2)}_t,\pi}(\mathbf{Y})$, once centered, induce a "centered" version of $G_{\pi^{(1)}\cup\pi^{(2)}}$, that can be factored in its connected components.

Lemma 5.18. *Consider $G^{(1)}, G^{(2)}$ two skeletons. If $(\pi^{(1)}, \pi^{(2)}) \in \Pi(\boldsymbol{M})$ for $\boldsymbol{M} \in \mathcal{M}^\star$, the connected components of the graph $G_{\pi^{(1)}\cup\pi^{(2)}}$, which is the union of $\pi^{(1)}(G^{(1)}), \pi^{(2)}(G^{(2)})$, are at most* $\min(m, r)$.

Proof. Since each connected component of $\pi^{(1)}(G^{(1)})$ has to share a vertex with at least another one from $\pi^{(2)}(G^{(2)})$ the minimal type of graph is as in fig. 11, when $m = r$. If $m > r$, or vice versa, we can only make the new vertex set in $\pi^{(1)}(G^{(1)})$ overlap with an existing one in $\pi^{(2)}(G^{(2)})$, without overlapping with any of the old ones from $\pi^{(1)}(G^{(1)})$. Thus, the number of connected components does not increase: it remains $r = \min\{m, r\}$. □

With the previous basis candidate of definition 5.2, we could upper bound the inner product of two labelled graphs $\pi^{(1)}(G^{(1)}), \pi^{(2)}(G^{(2)})$ by:

$$\lambda^{|\pi^{(1)}\triangle\pi^{(2)}|}\left(\frac{k}{n}\right)^{\left|V_{\pi^{(1)}\triangle\pi^{(2)}}\right|} \mathbb{1}\left\{\pi^{(1)}(G^{(1)}) \text{ conn. } \pi^{(2)}(G^{(2)})\right\}. \tag{5.19}$$

In the new basis, it is not as immediate to say the same: since we center each connected component, the inner product has a non-trivial expression. For $G^{(1)} = \bigcup_{s=1}^{m} G^{(1)}_s$ and $G^{(2)} = \bigcup_{t=1}^{r} G^{(2)}_t$ two skeletons and $\pi^{(1)} \in \Pi_{|V^{(1)}|}, \pi^{(2)} \in \Pi_{|V^{(2)}|}$ labellings it holds that:

$$\left\langle \overline{P}_{G^{(1)},\pi^{(1)}}, \overline{P}_{G^{(2)},\pi^{(2)}} \right\rangle_{\mathrm{H}_0} = \mathbb{E}_{\mathrm{H}_0}\left[\widehat{P}_{G^{(1)},\pi^{(1)}}(\mathbf{Y})\widehat{P}_{G^{(2)},\pi^{(2)}}(\mathbf{Y})\right] = \mathbb{E}_{\mathrm{H}_0}\left[\prod_{s=1}^{m}\prod_{r=1}^{t} \widehat{P}_{G^{(1)}_s,\pi^{(1)}}\widehat{P}_{G^{(2)}_t,\pi^{(2)}}\right]. \tag{5.20}$$

In the next subsection, we propose the intuition behind the quantitative bound we establish in proposition 5.30.

5.1 *An informal discussion about what we want*

By construction, if $\pi^{(1)}(G^{(1)}), \pi^{(2)}(G^{(2)})$ have an isolated connected component in the graph $G_{\pi^{(1)}\cup\pi^{(2)}}$ then the covariance is null, i.e. the case of fig. 10. Therefore, let us suppose they are interconnected (def. 5.16). In other words, $(\pi^{(1)}, \pi^{(2)}) \in \Pi(\boldsymbol{M})$ for some $\boldsymbol{M} \in \mathcal{M}^\star$. Without loss of generality, let $m \leqslant r$. Then $G_{\pi^{(1)}\cup\pi^{(2)}}$, the union graph of $\pi^{(1)}(G^{(1)}), \pi^{(2)}(G^{(2)})$, factors into $G^{(1)}_{\pi^{(1)}\cup\pi^{(2)}} \cup \cdots G^{(h)}_{\pi^{(1)}\cup\pi^{(2)}}$ connected components where $h \leqslant m$ by lemma 5.18. Each of these is independent once conditioning on the latent variables, and the expectation of their product decouples. For each $G^{(w)}_{\pi^{(1)}\cup\pi^{(2)}}$ assign a set of "incident" edge sets $\partial_w(\pi^{(1)}), \partial_w(\pi^{(2)})$ that are the ones inducing the connected component. Then:

$$\mathbb{E}_{\mathrm{H}_0}\left[\overline{P}_{G^{(1)},\pi^{(1)}}\overline{P}_{G^{(2)},\pi^{(2)}}\right] = \prod_{w=1}^{h}\mathbb{E}_{\mathrm{H}_0}\left[\overline{P}_{G^{(1)},\partial_w(\pi^{(1)})}\overline{P}_{G^{(2)},\partial_w(\pi^{(2)})}\right]. \tag{5.21}$$

In other words: the labellings $\pi^{(1)}, \pi^{(2)}$ are partitioned into the different connected components and for each $\partial_w(\pi^{(1)})$ we take only the labelling pertinent to the w^{th} connected component in the union graph.
By construction, the sets in $\partial_w(\pi^{(1)}), \partial_w(\pi^{(2)})$ only overlap in between and not within, meaning that an edge set in $\partial_w(\pi^{(1)})$ can only have shared vertices/edges with an edge set in $\partial_w(\pi^{(2)})$ and vice versa. Moreover, the pair $(\partial_w(\pi^{(1)}), \partial_w(\pi^{(2)}))$ induces the graph corresponding to the connected component $G^{(w)}_{\pi^{(1)}\cup\pi^{(2)}}$. When we take the expectation of this connected component we have cancellations: each edge in $G^{(w)}_{\pi^{(1)}\cup\pi^{(2)}}$ can appear:

1. just in $\pi^{(1)}(G^{(1)}_s)$ for only one $\pi^{(1)}(G^{(1)}_s) \in \partial_w(\pi^{(1)})$;
2. just in $G^{(2)}_t$ for only one $\pi^{(2)}(G^{(2)}_t) \in \partial_w(\pi^{(2)})$;
3. in a unique pair $(\pi^{(1)}(G^{(1)}_s), \pi^{(2)}(G^{(2)}_t))$.

When case #3 happens, we have $\mathrm{Y}^2_{ij} \overset{a.s.}{=} 1$ and do not integrate over it when we take the outer expectation. At the same time, it appears also inside the inner expectation $\mathbb{E}_{\mathrm{H}_0}\left[\prod_{(i,j)\in E^{(1)}_s}\mathrm{Y}_{\pi^{(1)}(i)\pi^{(1)}(j)}\right]$ and $\mathbb{E}_{\mathrm{H}_0}\left[\prod_{(i,j)\in E^{(2)}_t}\mathrm{Y}_{\pi^{(2)}(i)\pi^{(2)}(j)}\right]$ of the minus terms. We can see that the first term in eqn. 5.20, the one with no inner expectations, will be just a big product in explicit form:

$$\mathsf{firstterm} = \lambda^{\#\{\text{edges appearing uniquely in either } \partial_w(\pi^{(1)}),\partial_w(\pi^{(2)})\}}\left(\frac{k}{n}\right)^{\#\{\text{vertices of uniquely appearing edges}\}}, \tag{5.22}$$

while for the others, we can see them as successive removals of connected component from either of $\partial_w(\pi^{(1)})$ or $\partial_w(\pi^{(2)})$. When we remove an edge set, we take its expectation as a single term,[13] and we claim its contribution is smaller than if it were included in the full integral. Indeed, when it interacts with the other connected components, it has to cancel out some vertices/edges that do not cancel out when it integrates alone. At the same time, the remaining integral also has more vertices/edges. Letting m_w, r_w be the number of connected components respectively in $\partial_w(\pi^{(1)}), \partial_w(\pi^{(2)})$, with $\sum_{w=1}^h m_w = m$ and $\sum_{w=1}^h r_w = r$, we have a chain of inequalities for all $S \subset \partial_w(\pi^{(1)}), T \subset \partial_w(\pi^{(2)})$:

$$\begin{aligned}
\mathbb{E}_{\mathrm{H}_0}\left[\prod_{s\in\partial_w(\pi^{(1)})}P_{G^{(1)}_s,\pi^{(1)}}\prod_{t\in\partial_w(\pi^{(2)})}P_{G^{(2)}_t,\pi^{(2)}}\right] &= \mathbb{E}_{\mathrm{H}_0}\left[\prod_{s\in\partial_w(\pi^{(1)})}\mathbf{Y}^{\square\pi^{(1)}(G^{(1)}_s)}\prod_{t\in\partial_w(\pi^{(2)})}\mathbf{Y}^{\square\pi^{(2)}(G^{(2)}_t)}\right] \\
&\geqslant \mathbb{E}_{\mathrm{H}_0}\left[\prod_{s\in S}\mathbf{Y}^{\square\pi^{(1)}(G^{(1)}_s)}\prod_{t\in T}\mathbf{Y}^{\square\pi^{(2)}(G^{(2)}_t)}\right] \\
&\times\prod_{s\notin S}\mathbb{E}_{\mathrm{H}_0}\left[\mathbf{Y}^{\square\pi^{(1)}(G^{(1)}_s)}\right]\prod_{t\notin T}\mathbb{E}_{\mathrm{H}_0}\left[\mathbf{Y}^{\square\pi^{(2)}(G^{(2)}_t)}\right] \\
&\geqslant \prod_{s\in\partial_w(\pi^{(1)})}\mathbb{E}_{\mathrm{H}_0}\left[\mathbf{Y}^{\square\pi^{(1)}(G^{(1)}_s)}\right]\prod_{t\in\partial_w(\pi^{(2)})}\mathbb{E}_{\mathrm{H}_0}\left[\mathbf{Y}^{\square\pi^{(2)}(G^{(2)}_t)}\right].
\end{aligned} \tag{5.23}$$

[13] e.g. it appears as $\mathbb{E}_{\mathrm{H}_0}\left[\prod_{(i,j)\in E^{(1)}_s}\mathrm{Y}_{\pi^{(1)}(i)\pi^{(1)}(j)}\right]$ in the multiplication.

Remark 5.24. *We have to be careful: while the inequality is true, there are $2^{r_h+m_h}$ terms in each h connected component to take into account. Then, while each is smaller, it might be that when summing they are not at all. The product structure of eqn. 5.20 is such that for each $j \in [m_h], i \in [r_h]$ there are $\binom{m_h}{j}\binom{r_h}{i}$ terms that have:*

- *an integral over j connected components from $\partial_w(\pi^{(1)})$ and i connected components from $\partial_w(\pi^{(2)})$;*
- *multiplied by a product of $m_h - j$ split integrals of connected components from $\partial_w(\pi^{(1)})$;*
- *multiplied by a product of $r_h - i$ split integrals of connected components from $\partial_w(\pi^{(2)})$.*

We need to consider the gained factor at each removal of an edge set from either $\partial_w(\pi^{(1)}), \partial_w(\pi^{(2)})$ and show that the sum over $(i,j) \in [m_h] \times [r_h]$ is globally contributing as a controlled multiple of the symmetric difference. In other words, to prove for all $w \in [h]$:

$$|\mathbb{E}_{H_0}\left[\overline{P}_{G^{(1)},\partial_w(\pi^{(1)})}\overline{P}_{G^{(2)},\partial_w(\pi^{(2)})}\right]| \lesssim f(n)\lambda^{\#\{\text{edges appearing uniquely in either } \partial_w(\pi^{(1)}),\partial_w(\pi^{(2)})\}}\left(\frac{k}{n}\right)^{\#\{\text{vertices of uniquely appearing edges}\}}, \tag{5.25}$$

we will need to quantify how much we gain by removing a connected component, and show that jointly all terms including the first one are $f(n)$ times the first one, where $f(n)$ grows slow enough to control it with vanishing terms in later results. If it holds for all $w \in [h]$, taking the product, it holds jointly for the full union graph and we find:

$$\begin{aligned}\mathbb{E}_{H_0}\left[\overline{P}_{G^{(1)},\pi^{(1)}}\overline{P}_{G^{(2)},\pi^{(2)}}\right] &\lesssim \prod_{w=1}^{h} f(n)\lambda^{\#\{\text{edges appearing uniquely in either } \partial_w(\pi^{(1)}),\partial_w(\pi^{(2)})\}}\left(\frac{k}{n}\right)^{\#\{\text{vertices of uniquely appearing edges}\}} \\ &= \lambda^{|\pi^{(1)}\triangle\pi^{(2)}|}\left(\frac{k}{n}\right)^{|V_{\pi^{(1)}\triangle\pi^{(2)}}|}\cdot f(n),\end{aligned} \tag{5.26}$$

since the sum of uniquely appearing edges and respective uniquely appearing vertices across connected components is the symmetric difference of $\pi^{(1)}, \pi^{(2)}$.

In the next subsection, we formalize the idea.

5.11 *Formalization*

We first establish a control on the decoupling of graphs arising from canonical monomials.

Lemma 5.27 (One step improvement)**.** *Recall the definition of canonical monomials (def. 2.18). Suppose $G^{(1)}, G^{(2)}$ are skeletons with labellings $\left(\pi^{(1)}, \pi^{(2)}\right) \in \Pi(\boldsymbol{M})$, where $\boldsymbol{M}$ is the set of matched nodes. Then:*

$$\mathbb{E}_{H_0}\left[P_{G^{(1)},\pi^{(1)}}P_{G^{(2)},\pi^{(2)}}\right] \geqslant \left(\frac{n}{k}\right)^{|\boldsymbol{M}|}\mathbb{E}_{H_0}\left[P_{G^{(1)},\pi^{(1)}}\right]\mathbb{E}_{H_0}\left[P_{G^{(2)},\pi^{(2)}}\right]. \tag{5.28}$$

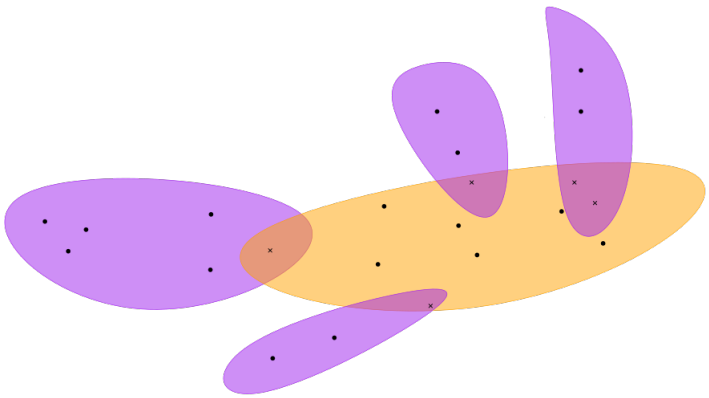
Proof. On the RHS we have the symmetric difference of the graphs, which is an integral over $G_{\triangle}$. By hypothesis, it has $|\boldsymbol{M}|$ matched nodes. On the RHS we have the product of the integrals along the two graphs, decoupled. All the matched nodes will be free, and counted twice instead of once, i.e. once for each graph. This gives a discount factor of $(n/k)^{|\boldsymbol{M}|}$ between the two at least without considering edges. □

Example 5.29. *We visualize this reasoning in figures 7b - 7b - 8a - 8b - 9.*

It remains to formalize how the $\overline{P}, \widehat{P}$ of defs. 5.11 and def. 5.2 are related. In particular, we will use that the $\widehat{P}$ basis is up to constants equal to the canonical monomial basis P of def. 2.18 when it is non-zero, and that the $\overline{P}$ basis is null in even more instances.

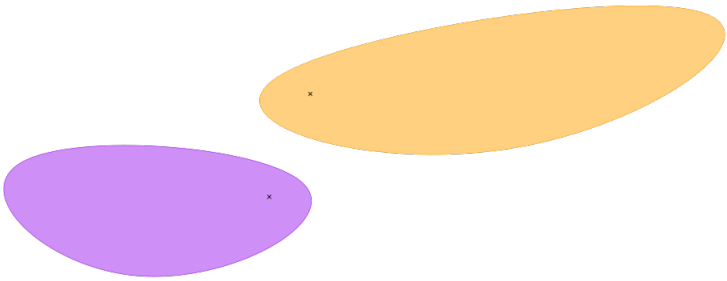
Proposition 5.30. *Suppose $G^{(1)}, G^{(2)}$ are skeletons with labellings $\left(\pi^{(1)}, \pi^{(2)}\right) \in \Pi(\boldsymbol{M})$. Let assumption 3.1 hold. Then:*

- *if $\boldsymbol{M} \notin \mathcal{M}^{\star}$ we have $\mathbb{E}_{H_0}\left[\overline{P}_{G^{(1)},\pi^{(1)}}\overline{P}_{G^{(2)},\pi^{(2)}}\right] = 0$;*



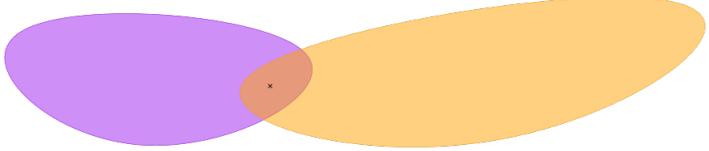
(a) Generic scenario before we remove one orange edge set

The orange edge set induces an orange vertex set that "overlaps" with some purple vertex sets. Among these, there are necessarily shared vertices, and there may be shared edges.



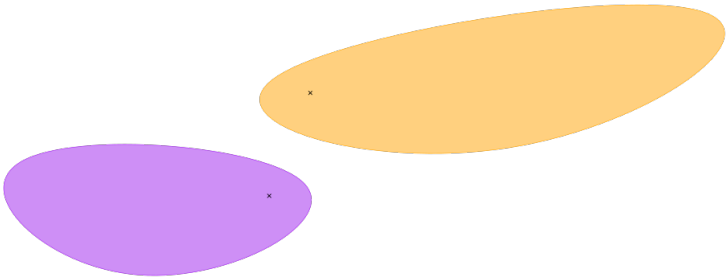
(b) Generic scenario when we remove one orange edge set

Upon removal from the joint integration, the shared parts become "free" and they contribute to the integral.



(a) Minimal scenario before we remove one edge set

In the worst case, we have two edge sets inducing vertex sets that overlap only over one vertex.



(b) Minimal scenario when we remove one edge set

Upon removal, the vertex that was shared is free in both vertex sets, so from one, joint vertex, we have two "free" vertices in two separate integrals.

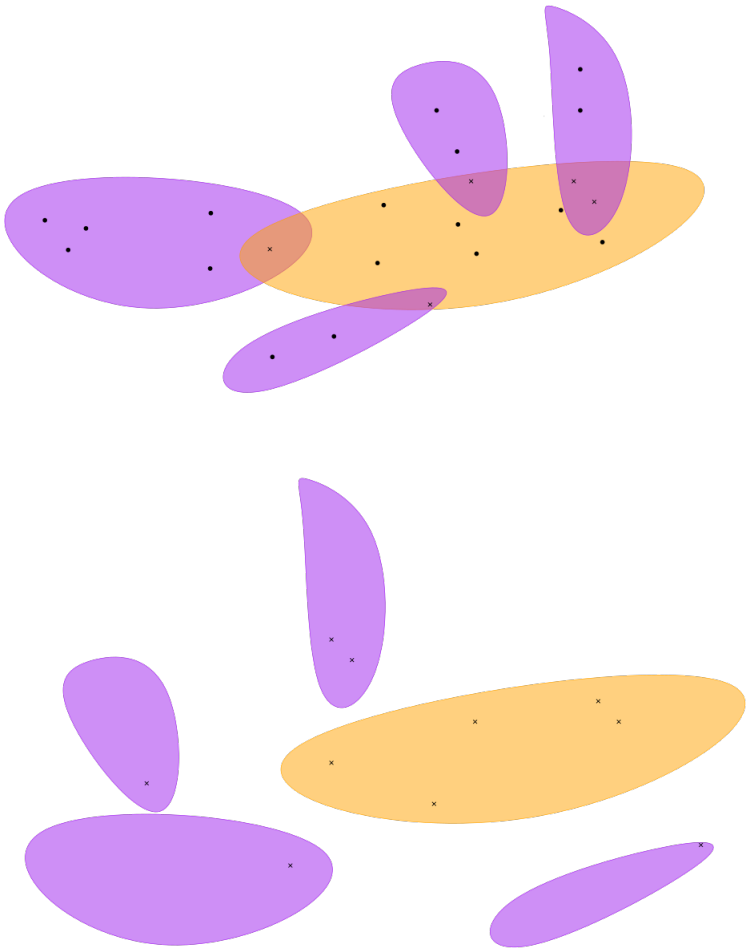


Figure 9: Full image
A graphical instantiation for the formalization of the action of removing connected components from the integral, i.e. lemma 5.27.

- *if* $\boldsymbol{M} \in \mathcal{M}_{\mathrm{PM}}$ *we have:*

$$\left|\frac{\mathbb{E}_{\mathrm{H}_0}\left[\overline{P}_{G^{(1)},\pi^{(1)}}\overline{P}_{G^{(2)},\pi^{(2)}}\right]-\mathbb{E}_{\mathrm{H}_0}\left[P_{G^{(1)},\pi^{(1)}}P_{G^{(2)},\pi^{(2)}}\right]}{\mathbb{E}_{\mathrm{H}_0}\left[P_{G^{(1)},\pi^{(1)}}P_{G^{(2)},\pi^{(2)}}\right]}\right| = o\left(1\right), \tag{5.31}$$

or without taking the limit the bound is by a factor $D^{\widetilde{c}k}/n = o\left(1\right)$*;*

- *if* $\boldsymbol{M} \in \mathcal{M}^{\star}$ *it holds that:*

$$\left|\mathbb{E}_{\mathrm{H}_0}\left[\overline{P}_{G^{(1)},\pi^{(1)}}\overline{P}_{G^{(2)},\pi^{(2)}}\right]\right| \leqslant (1+o\left(1\right))\mathbb{E}_{\mathrm{H}_0}\left[P_{G^{(1)},\pi^{(1)}}P_{G^{(2)},\pi^{(2)}}\right], \tag{5.32}$$

or without taking the limit the coefficient is a constant.

Proof. **(claim #1)** Follows by equation 5.15.
(claims #2-#3) We apply lem. 5.27 recursively on the unfolding of the $\overline{P}$ basis. The difference of the two bases is a sum over all ways of removing connected components from the full integrals. Without loss of generality, let m, r be the number of connected components of $G^{(1)}, G^{(2)}$ respectively. Then:

$$\begin{aligned}
&\left|\mathbb{E}_{\mathrm{H}_0}\left[\overline{P}_{G^{(1)},\pi^{(1)}}\overline{P}_{G^{(2)},\pi^{(2)}}\right]-\mathbb{E}_{\mathrm{H}_0}\left[P_{G^{(1)},\pi^{(1)}}P_{G^{(2)},\pi^{(2)}}\right]\right| \\
&= \sum_{S\subseteq[m],T\subseteq[r]} \mathbb{1}\{S\neq\varnothing \vee T\neq\varnothing\}\mathbb{E}_{\mathrm{H}_0}\left[P_{(G_s^{(1)})_{s\notin S},\pi^{(1)}}P_{(G_t^{(2)})_{t\notin T},\pi^{(2)}}\right]\prod_{s\in S}\mathbb{E}_{\mathrm{H}_0}\left[P_{G_s^{(1)},\pi^{(1)}}\right]\prod_{t\in T}\mathbb{E}_{\mathrm{H}_0}\left[P_{G_t^{(2)},\pi^{(2)}}\right] \\
&=: R
\end{aligned} \tag{5.33}$$

In particular, the indicator makes sure that we do not take the full product, and we disregard the (-1) power appearing from the subtractions. We can use lem. 5.27 recursively on each term in the sum. Element-wise, we have:

$$\mathbb{E}_{\mathrm{H}_0}\left[P_{(G_s^{(1)})_{s\notin S},\pi^{(1)}}P_{(G_t^{(2)})_{t\notin T},\pi^{(2)}}\right]\prod_{s\in S}\mathbb{E}_{\mathrm{H}_0}\left[P_{G_s^{(1)},\pi^{(1)}}\right]\prod_{t\in T}\mathbb{E}_{\mathrm{H}_0}\left[P_{G_t^{(2)},\pi^{(2)}}\right] \leqslant \left(\frac{k}{n}\right)^{|\boldsymbol{M}(S,T,[m],[r])|}\mathbb{E}_{\mathrm{H}_0}\left[P_{G^{(1)},\pi^{(1)}}P_{G^{(2)},\pi^{(2)}}\right], \tag{5.34}$$

where $|\boldsymbol{M}(S,T,[m],[r])|$ is the set of matched vertices arising from the recursion:

(Ro) take the large symmetric difference graph between $[m]\backslash S, [r]\backslash T$;

(R1) for each term in S recursively apply lem. 5.27;

(R2) for each term in T recursively apply lem. 5.27.

- form $\boldsymbol{M}_{\mathrm{PM}}(S,T,[m],[r])$ as the set of perfect matches between the symmetric difference of $[m]\backslash S, [r]\backslash T$ and the connected components S, T taken alone.

In other words, they are matches (v, v') of the following three mutually exclusive types (see lem. 5.39#1):

(T1) v is a vertex in a c.c. in S and v' is a vertex in a c.c. in $[r]\backslash T$;

(T2) v is a vertex in a c.c. in T and v' is a vertex in a c.c. in $[m]\backslash S$;

(T3) v is a vertex in a c.c. in S and v' is a vertex in a c.c. in T.

There are $2^{m_S+m_T}$ ways to take connected components forming a big integral with $|S| = m_S, |T| = r_T$ terms. Moreover, $m_S \in \{0,\dots,m\}, m_T \in \{0,\dots,r\}$, so we can group the big sum by sizes to find:

$$\begin{aligned}
R &\leqslant \sum_{S\subseteq[m],T\subseteq[r]} \mathbb{1}\{S\neq\varnothing, T\neq\varnothing\}\left(\frac{k}{n}\right)^{|\boldsymbol{M}(S,T,[m],[r])|}\mathbb{E}_{\mathrm{H}_0}\left[P_{G^{(1)},\pi^{(1)}}P_{G^{(2)},\pi^{(2)}}\right] \\
&= \mathbb{E}_{\mathrm{H}_0}\left[P_{G^{(1)},\pi^{(1)}}P_{G^{(2)},\pi^{(2)}}\right]\sum_{S\subseteq[m],T\subseteq[r]} \mathbb{1}\{S\neq\varnothing \vee T\neq\varnothing\}\left(\frac{k}{n}\right)^{|\boldsymbol{M}(S,T,[m],[r])|} \\
&\leqslant \mathbb{E}_{\mathrm{H}_0}\left[P_{G^{(1)},\pi^{(1)}}P_{G^{(2)},\pi^{(2)}}\right]\sum_{\mu=0}^{m}\sum_{\rho=0}^{r}\mathbb{1}\{\mu+\rho>0\}2^{\mu+\rho}\left(\frac{k}{n}\right)^{\min_{S:|S|=\mu,T:|T|=\rho}|\boldsymbol{M}(S,T,[m],[r])|}.
\end{aligned} \tag{5.35}$$

In words, we grouped by subsets of size μ, ρ and took the minimal number of matchings associated to the operation we explained just above in such subsets. We use lemma 5.39#2 which we prove just after:

$$\mu + \rho - |\boldsymbol{M}(S, T, [m], [r])| \leqslant 0, \qquad \forall S : |S| = \mu, T : |T| = \rho, \tag{5.36}$$

thanks to which (concentrating on the sum) taking $S^\star, T^\star$ the optimizers of the minimum:

$$\begin{aligned}
&\sum_{\mu=0}^{m}\sum_{\rho=0}^{r} \mathbb{1}\{\mu+\rho>0\} 2^{\mu+\rho} \left(\frac{k}{n}\right)^{\min_{S:|S|=\mu,T:|T|=\rho}|\boldsymbol{M}(S,T,[m],[r])|} \\
\leqslant &\sum_{\mu=0}^{m}\sum_{\rho=0}^{r} \mathbb{1}\{\mu+\rho>0\} \left(2\frac{k}{n}\right)^{|\boldsymbol{M}(S^\star,T^\star,[m],[r])|} \\
\leqslant &(D^2+1)\left(2\frac{k}{n}\right) \\
= &\, o\left(1\right),
\end{aligned} \tag{5.37}$$

where we used in the last step assumption 3.1, i.e. that $k/n \leqslant D^{-8c_{\text{si}}}$ for some large $c_{\text{si}} > 0$ constant, and in particular $8c_{\text{si}} \geqslant 3$. Putting it all together, we proved that:[14]

$$\left|\frac{\mathbb{E}_{\text{H}_0}\left[\overline{P}_{G^{(1)},\pi^{(1)}}\overline{P}_{G^{(2)},\pi^{(2)}}\right] - \mathbb{E}_{\text{H}_0}\left[P_{G^{(1)},\pi^{(1)}}P_{G^{(2)},\pi^{(2)}}\right]}{\mathbb{E}_{\text{H}_0}\left[P_{G^{(1)},\pi^{(1)}}P_{G^{(2)},\pi^{(2)}}\right]}\right| \leqslant 2(D^2+1)\frac{k}{n}. \tag{5.38}$$

Claim #3 follows as a special case since we assumed $\boldsymbol{M} \in \mathcal{M}^\star$ and $\mathcal{M}_{\text{PM}} \subset \mathcal{M}^\star$. Claim #2 follows by rearranging terms and using the reverse triangular inequality on the numerator.

□

Lemma 5.39. *The following are true in the above proof:*

1. *the construction of $\boldsymbol{M}(S, T, [m], [r])$ in equation 5.34 via (R0) - (R1) - (R2) is unambiguous and in bijection with the three types (T1) - (T2) - (T3) mentioned just after;*
2. *$\mu + \rho \leqslant |\boldsymbol{M}(S, T, [m], [r])|$ for all feasible combinations;*

In particular, they hold if we take the minimum over S, T such that $|S| = \mu, |T| = \rho$.

Proof. **(claim #1)** For given $G^{(1)}, G^{(2)}$ connected components interact only across and not within, in the sense that $G^{(1)}_s, G^{(1)}_{s'}$ do not have common vertices when $s \neq s'$. The polynomials over $s \notin S, t \notin T$ are all inside the same integral. By simple algebra using def. 2.18 a product of different polynomials over the canonical basis returns the symmetric difference graph of the underlying graphs. Therefore, in the (R0) step we are considering the large integral. Using lemma 5.27, we can bound the product of two integrals by the integral of the product times a power of k/n which is the number of perfect matches between the two graphs. In steps (R1) - (R2) of the recursion we then couple all the integrals together bringing inside the same large expectation all the terms in S, T that were integrated alone. As a result, we will have the integral over the whole product of the two graphs $G^{(1)}, G^{(2)}$ discounted by k/n times the sum of successive perfect matches between the graphs explored in $G^{(1)}, G^{(2)}$. When a node is perfectly matched, it belongs unambiguously to two separate connected components, one from $G^{(1)}$, one from $G^{(2)}$. Suppose such vertex is then in $G^{(1)}_s, G^{(2)}_t$ for some $s \in [m], t \in [r]$. It can pop out in the recursion steps (R1) - (R2) if and only if it satisfies either of the conditions #1-#2-#3 in the proof, as otherwise it is a perfectly matched node already in the symmetric difference graph of step (R0), which does not influence the applications of lemma 5.27. This proves that there is a bijection between the construction in the recursion (R0) - (R1) - (R2) and the criterions #1-#2-#3 for the vertices appearing in the power. Moreover, since the vertex belongs to a unique pair, the order in which we add the connected components does not matter in the recursion, and the expression is unambiguous.
(claim #2) The sizes μ, ρ correspond to the number of connected components in the integrals taken alone. Since $\boldsymbol{M} \in \mathcal{M}^\star$ each connected component has at least a matched node. It follows that $\mu + \rho \leqslant \boldsymbol{M}(S, T, [m], [r])$, since:

- for at least one vertex in each c.c. in S there is a vertex in either a c.c. in T or in the symmetric difference graph of c.c.s in $[m]\backslash S, [r]\backslash T$;

[14] Notice that in the P basis of def. 2.18 the inner product is always positive.

- for at least one vertex in each c.c. in T there is a vertex in either a c.c. in S or in the symmetric difference graph of c.c.s in $[m]\backslash S, [r]\backslash T$.

Thus, for each of the μ and for each of the ρ connected components there is a matched vertex in $\boldsymbol{M}(S, T, [m], [r])$ according to the three exhaustive cases in the proof. □

Having this similarity of basis products we can compute the variance and the covariance. Then, we clarify in which sense the new basis is almost orthonormal.

Definition 5.40 (Dominating term)**.** *Let G be a skeleton. Define:*

$$\nu(G) := \frac{n!}{(n-|V|)!}|\mathrm{Aut}(G)|. \tag{5.41}$$

Let us then define our final basis.

Definition 5.42 (Corrected rescaled monomial basis)**.** *Recall the form of $\overline{P}_G$ for G a skeleton from def. 5.11. Define:*

$$\widetilde{P}_G := \frac{1}{\sqrt{\nu(G)}}\overline{P}_G. \tag{5.43}$$

We term "rescaled basis" the collection $\left(1, \left(\widetilde{P}_G\right)_{G\in\mathcal{G}_{(\leqslant D)}}\right)$. Since we just rescale elements, it is still a basis for invariant polynomials of degree less than D (lemma 5.13).

Proposition 5.44 (Variance and covariance over skeleton)**.** *Consider the $\overline{P}$ basis of def. 5.11. Let assumption 3.1 hold. Then if D diverges with n:*

$$\mathrm{Var}_{\mathrm{H}_0}\left[\widetilde{P}_G\right] = \mathrm{Var}_{\mathrm{H}_0}\left[\frac{1}{\sqrt{\nu(G)}}\overline{P}_G\right] \asymp 1, \tag{5.45}$$

and if D does not diverge with n we have a quantitative correction, meaning equal up to constants. Moreover, if $G^{(1)} \neq G^{(2)}$ are two skeletons:

$$\left|\mathrm{CoV}_{\mathrm{H}_0}\left[\frac{1}{\sqrt{\nu(G^{(1)})\nu(G^{(2)})}}\overline{P}_{G^{(1)}}\overline{P}_{G^{(2)}}\right]\right| \lesssim D^{-c\,\mathrm{d}(G^{(1)},G^{(2)})}, \tag{5.46}$$

where $c > 0$ is a positive factor that we can choose. Later we will take $c = 4c_{\mathrm{si}}$.

Proof. Let us work on a generic tuple $G^{(1)}, G^{(2)}$. We have:

$$\mathrm{CoV}_{\mathrm{H}_0}\left[\widetilde{P}_{G^{(1)}}\widetilde{P}_{G^{(2)}}\right] = \sum_{\boldsymbol{M}\in\mathcal{M}^\star}\sum_{(\pi^{(1)},\pi^{(2)})\in\Pi(\boldsymbol{M})}\mathbb{E}_{\mathrm{H}_0}\left[\widetilde{P}_{G^{(1)},\pi^{(1)}}\widetilde{P}_{G^{(2)},\pi^{(2)}}\right], \tag{5.47}$$

where we applied proposition 5.30#1 to remove matchings that are not in $\mathcal{M}^\star$. If $G^{(1)} \simeq G^{(2)}$ then we can have perfect matchings, otherwise not. We distinguish the two cases.
(PM) When there are perfect matchings:

$$\mathrm{CoV}_{\mathrm{H}_0}\left[\widetilde{P}_{G^{(1)}}\widetilde{P}_{G^{(2)}}\right] = \mathrm{CoV}_{\mathrm{H}_0}\left[\widetilde{P}^2_{G^{(1)}}\right] = \frac{1}{\nu(G^{(1)})}\mathrm{Var}_{\mathrm{H}_0}\left[\overline{P}_{G^{(1)}}\right]. \tag{5.48}$$

The variance has form:

$$\begin{aligned}
\mathrm{Var}_{\mathrm{H}_0}\left[\overline{P}^{(1)}_G\right] &= \sum_{\boldsymbol{M}\in\mathcal{M}_{\mathrm{PM}}}\sum_{(\pi^{(1)},\pi^{(2)})\in\Pi(\boldsymbol{M})}\mathbb{E}_{\mathrm{H}_0}\left[\overline{P}_{G^{(1)},\pi^{(1)}}\overline{P}_{G^{(1)},\pi^{(2)}}\right] + \sum_{\boldsymbol{M}\in\mathcal{M}^\star\backslash\mathcal{M}_{\mathrm{PM}}}\sum_{(\pi^{(1)},\pi^{(2)})\in\Pi(\boldsymbol{M})}\mathbb{E}_{\mathrm{H}_0}\left[\overline{P}_{G^{(1)},\pi^{(1)}}\overline{P}_{G^{(1)},\pi^{(2)}}\right] \\
&\geqslant \sum_{\boldsymbol{M}\in\mathcal{M}_{\mathrm{PM}}}\sum_{(\pi^{(1)},\pi^{(2)})\in\Pi(\boldsymbol{M})}\left(1 - D^{\tilde{c}}\frac{k}{n}\right)\underbrace{\mathbb{E}_{\mathrm{H}_0}\left[P_{G^{(1)},\pi^{(1)}}P_{G^{(1)},\pi^{(2)}}\right]}_{=1} \\
&\quad - \sum_{\boldsymbol{M}\in\mathcal{M}^\star\backslash\mathcal{M}_{\mathrm{PM}}}\sum_{(\pi^{(1)},\pi^{(2)})\in\Pi(\boldsymbol{M})}\left|\mathbb{E}_{\mathrm{H}_0}\left[\overline{P}_{G^{(1)},\pi^{(1)}}\overline{P}_{G^{(1)},\pi^{(2)}}\right]\right| \\
&= \left(1 - D^{\tilde{c}}\frac{k}{n}\right)\sum_{\boldsymbol{M}\in\mathcal{M}_{\mathrm{PM}}}|\Pi(\boldsymbol{M})| - \sum_{\boldsymbol{M}\in\mathcal{M}^\star\backslash\mathcal{M}_{\mathrm{PM}}}\sum_{(\pi^{(1)},\pi^{(2)})\in\Pi(\boldsymbol{M})}\left|\mathbb{E}_{\mathrm{H}_0}\left[\overline{P}_{G^{(1)},\pi^{(1)}}\overline{P}_{G^{(1)},\pi^{(2)}}\right]\right| \\
&= \left(1 - D^{\tilde{c}}\frac{k}{n}\right)\frac{n!}{(n-|V^{(1)}|)!}|\mathrm{Aut}(G^{(1)})| - \sum_{\boldsymbol{M}\in\mathcal{M}^\star\backslash\mathcal{M}_{\mathrm{PM}}}\sum_{(\pi^{(1)},\pi^{(2)})\in\Pi(\boldsymbol{M})}\left|\mathbb{E}_{\mathrm{H}_0}\left[\overline{P}_{G^{(1)},\pi^{(1)}}\overline{P}_{G^{(1)},\pi^{(2)}}\right]\right|
\end{aligned} \tag{5.49}$$

where we used lemmas 4.53 - 4.50 and the bound of proposition 5.30#3. The same steps but the other inequality of proposition 5.30#3 establish an upper bound on the variance of the form (notice we flip two signs):

$$\mathrm{Var}_{\mathrm{H}_0}\left[\overline{P}_{G^{(1)}}\right] \leqslant \left(1 + D^{\tilde{c}}\frac{k}{n}\right)\frac{n!}{(n-|V^{(1)}|)!}|\mathrm{Aut}(G^{(1)})| + \sum_{\boldsymbol{M}\in\mathcal{M}^\star\setminus\mathcal{M}_{\mathrm{PM}}}\sum_{(\pi^{(1)},\pi^{(2)})\in\Pi(\boldsymbol{M})}\left|\mathbb{E}_{\mathrm{H}_0}\left[\overline{P}_{G^{(1)},\pi^{(1)}}\overline{P}_{G^{(1)},\pi^{(2)}}\right]\right|. \tag{5.50}$$

In particular, $D^{\tilde{c}}\frac{k}{n} \leqslant D^{\tilde{c}-8c_{\mathrm{si}}} \leqslant 1/2$ once c_{si} is large enough in assumption 3.1. If we normalize by $\nu(G^{(1)})$, the first term is asymptotically $\asymp 1$, or upper and lower bound by $1 \pm c$ for a positive constant. What we wish to show is that the second term is negligible.

(NO-PM) When there are no perfect matchings, i.e. when $G^{(1)} \not\simeq G^{(2)}$, we only have the second term in equation 5.50, and we wish to show it is small in a proper sense.

(work on the correction term) We are then hinted to normalize the $\overline{P}$ polynomials of definition 5.11 by the candidate dominating factor of definition 5.40, obtaining the rescaled basis of definition 5.42. For this, we want to show that the correction term

$$\mathsf{correct} \coloneqq \sum_{\boldsymbol{M}\in\mathcal{M}^\star\setminus\mathcal{M}_{\mathrm{PM}}}\sum_{(\pi^{(1)},\pi^{(2)})\in\Pi(\boldsymbol{M})}\frac{1}{\sqrt{\nu(G^{(1)})\nu(G^{(2)})}}\mathbb{E}_{\mathrm{H}_0}\left[\overline{P}_{G^{(1)},\pi^{(1)}}\overline{P}_{G^{(2)},\pi^{(2)}}\right], \tag{5.51}$$

is less than $D^{-c\,\mathrm{d}(G^{(1)},G^{(2)})}$ for some constant $c > 0$ as in the claim. By proposition 5.30#2, we have that:[15]

$$\begin{aligned}\mathsf{correct} &\leqslant \sum_{\boldsymbol{M}\in\mathcal{M}^\star\setminus\mathcal{M}_{\mathrm{PM}}}\sum_{(\pi^{(1)},\pi^{(2)})\in\Pi(\boldsymbol{M})}\frac{1}{\sqrt{\nu(G^{(1)})\nu(G^{(2)})}}\left|\mathbb{E}_{\mathrm{H}_0}\left[\overline{P}_{G^{(1)},\pi^{(1)}}\overline{P}_{G^{(2)},\pi^{(2)}}\right]\right| \\ &\leqslant \sum_{\boldsymbol{M}\in\mathcal{M}^\star\setminus\mathcal{M}_{\mathrm{PM}}}\frac{2}{\sqrt{\nu(G^{(1)})\nu(G^{(2)})}}\lambda^{|E_\triangle|}\left(\frac{k}{n}\right)^{|V_\triangle|}|\Pi(\boldsymbol{M})|,\end{aligned} \tag{5.52}$$

where we used that for fixed $\boldsymbol{M}$ the sets of vertices are fixed (i.e. we fix the skeletons). Using lemmas 4.50 - 4.53 - 4.55, we can bound:

$$\frac{1}{\sqrt{\nu(G^{(1)})\nu(G^{(2)})}}|\Pi(\boldsymbol{M})| \leqslant \frac{1}{\sqrt{\left|\mathrm{Aut}\left(G^{(1)}\right)\right|\left|\mathrm{Aut}\left(G^{(2)}\right)\right|}}n^{|V^{(1)}|+|V^{(2)}|/2-|\boldsymbol{M}|} = \frac{1}{\sqrt{\left|\mathrm{Aut}\left(G^{(1)}\right)\right|\left|\mathrm{Aut}\left(G^{(2)}\right)\right|}}n^{|U^{(1)}|+|U^{(2)}|/2}. \tag{5.53}$$

Recall the invariant objects defined in subsection 4.I. From lem. 4.59, we recall that the following bounds hold on the graph structure:

$$|E_\triangle| \geqslant \max\left\{|V_\triangle| - \#\mathrm{CC},\, \mathrm{d}(G^{(1)}, G^{(2)})\right\} \tag{5.54}$$

$$|V_\triangle| = |U^{(1)}| + |U^{(2)}| + |\boldsymbol{M}_{\mathrm{SM}}| \tag{5.55}$$

$$|\boldsymbol{M}_{\mathrm{SM}}| \geqslant \#\mathrm{CC} \tag{5.56}$$

$$|V_\triangle| \geqslant 2\#\mathrm{CC}. \tag{5.57}$$

Let us rearrange terms to make some clever adjustments in the big sum so that the next steps are more explicit. Reordering:

$$\begin{aligned}\mathsf{correct} &\leqslant \frac{2}{\sqrt{\left|\mathrm{Aut}(G^{(1)})\right|\left|\mathrm{Aut}(G^{(2)})\right|}}\sum_{\boldsymbol{M}\in\mathcal{M}^\star\setminus\mathcal{M}_{\mathrm{PM}}}n^{|U^{(1)}|+|U^{(2)}|/2}\lambda^{|E_\triangle|}\left(\frac{k}{n}\right)^{|V_\triangle|} \\ &= \frac{2}{\sqrt{\left|\mathrm{Aut}(G^{(1)})\right|\left|\mathrm{Aut}(G^{(2)})\right|}}\sum_{\boldsymbol{M}\in\mathcal{M}^\star\setminus\mathcal{M}_{\mathrm{PM}}}\left(\frac{\lambda k}{\sqrt{n}}\right)^{|U^{(1)}|+|U^{(2)}|}\lambda^{|E_\triangle|-|U^{(1)}|-|U^{(2)}|}\left(\frac{k}{n}\right)^{|V_\triangle|-|U^{(1)}|-|U^{(2)}|}\end{aligned} \tag{5.58}$$

We can see that each term in the sum is vanishing. We now aim to show they are *globally* vanishing. In particular, we seek something explicit in the exploding factor $n^{|U^{(1)}|+|U^{(2)}|}$, since we want to counter it, and attempt to reorder terms to not lose any contribution that is vanishing.

[15] We have a control by a $1 + D^{\tilde{c}}k/n \leqslant 2$ factor.

To continue, we exploit the graph structure. Combining the inequalities & identities above in eqns. 5.54 to 5.57 we find:

$$|E_\triangle| - |U^{(1)}| - |U^{(2)}| \geqslant \max\left\{0,\, \mathrm{d}(G^{(1)}, G^{(2)}) - |U^{(1)}| - |U^{(2)}|, 1 - |U^{(1)}| - |U^{(2)}|\right\} \quad \text{since } |E_\triangle| \geqslant 1; \tag{5.59}$$

$$|\boldsymbol{M}_{\mathrm{SM}}| - \#\mathrm{CC} \geqslant 0 \tag{5.60}$$

$$|V_\triangle| - |U^{(1)}| - |U^{(2)}| = |\boldsymbol{M}_{\mathrm{SM}}|. \tag{5.61}$$

Using the signal condition of assumption 3.1:

$$\max\left\{\frac{\lambda k}{\sqrt{n}}, \frac{k}{n}, \lambda\right\} \leqslant D^{-8c_{\mathrm{si}}}, \tag{5.62}$$

for some large constant $c_{\mathrm{si}} > 0$, we can collect all powers inside a single term:

$$\mathsf{correct} \leqslant \frac{2}{\sqrt{|\mathrm{Aut}(G^{(1)})||\mathrm{Aut}(G^{(2)})|}} \sum_{\boldsymbol{M}\in\mathcal{M}^\star\setminus\mathcal{M}_{\mathrm{PM}}} \left(D^{-8c_{\mathrm{si}}}\right)^{\mathsf{pow}} \tag{5.63}$$

$$\mathsf{pow} := |U^{(1)}| + |U^{(2)}| + \max\left\{0,\, \mathrm{d}(G^{(1)}, G^{(2)}) - |U^{(1)}| - |U^{(2)}|, 1 - |U^{(1)}| - |U^{(2)}|\right\} + |\boldsymbol{M}_{\mathrm{SM}}|. \tag{5.64}$$

Discussing the cases in the maximum some algebra shows that:

$$-8_{c_{\mathrm{si}}}\mathsf{pow} \leqslant -8c_{\mathrm{si}} \max\left\{\mathrm{d}(G^{(1)}, G^{(2)}), 1, |U^{(1)}| + |U^{(2)}| + |\boldsymbol{M}_{\mathrm{SM}}|\right\}. \tag{5.65}$$

Recall the construction of shadow matchings from section 4. Starting from:

$$\begin{aligned}
\mathsf{correct} &\leqslant 2/\sqrt{|\mathrm{Aut}(G^{(1)})||\mathrm{Aut}(G^{(2)})|} \sum_{\boldsymbol{M}\in\mathcal{M}^\star\setminus\mathcal{M}_{\mathrm{PM}}} D^{-8c_{\mathrm{si}}\max\{\mathrm{d}(G^{(1)},G^{(2)}),1,|U^{(1)}|+|U^{(2)}|+|\boldsymbol{M}_{\mathrm{SM}}|\}} \\
&= 2/\sqrt{|\mathrm{Aut}(G^{(1)})||\mathrm{Aut}(G^{(2)})|} \sum_{(U^{(1)},U^{(2)},\underline{\boldsymbol{M}})\text{ triplets}} D^{-8c_{\mathrm{si}}\max\{\mathrm{d}(G^{(1)},G^{(2)}),1,|U^{(1)}|+|U^{(2)}|+|\boldsymbol{M}_{\mathrm{SM}}|\}} \left|\mathcal{M}_{\mathsf{shadow}}\left(U^{(1)}, U^{(2)}, \underline{\boldsymbol{M}}\right)\right| \\
&\leqslant 2 \sum_{(U^{(1)},U^{(2)},\underline{\boldsymbol{M}})\text{ triplets}} D^{-8c_{\mathrm{si}}\max\{\mathrm{d}(G^{(1)},G^{(2)}),1,|U^{(1)}|+|U^{(2)}|+|\boldsymbol{M}_{\mathrm{SM}}|\}} \\
&\leqslant 2 \sum_{\substack{1\leqslant u_1\leqslant 2D\\ 1\leqslant u_2\leqslant 2D\\ 1\leqslant s\leqslant 2D}} N_{u_1,u_2,s} \cdot D^{-8c_{\mathrm{si}}\max\{\mathrm{d}(G^{(1)},G^{(2)}),1,u_1+u_2+s\}}
\end{aligned} \tag{5.66}$$

where we used lemma 4.64, which we remind bounds the size of the set of shadow matchings by the minimal size of the automorphism group, and grouped by sizes, with the observation that we consider graphs with less than $2D$ vertices. It remains to count the number of triplets at the level of skeletons, i.e. $N_{u_1,u_2,s}$. For graphs over less than D edges, and so less than $2D$ vertices, it is certainly less than $(2D)^{u_1+u_2+2s}$, where the upper bound corresponds to choosing from $2D$ vertices at each vertex. Plugging it inside:

$$\mathsf{correct} \leqslant 2 \sum_{\substack{1\leqslant u_1\leqslant 2D\\ 1\leqslant u_2\leqslant 2D\\ 1\leqslant s\leqslant 2D}} (2D)^{u_1+u_2+2s} D^{-8c_{\mathrm{si}}\max\{\mathrm{d}(G^{(1)},G^{(2)}),1,u_1+u_2+s\}}, \tag{5.67}$$

where we discuss the two cases of the maximum, using $D \geqslant 2$ and $-8c_{\mathrm{si}} + 4 \leqslant 0$. By lemma 5.70 we find:

$$\mathsf{correct} \leqslant 2 \sum_{\substack{1\leqslant u_1\leqslant 2D\\ 1\leqslant u_2\leqslant 2D\\ 1\leqslant s\leqslant 2D}} D^{(4-8c_{\mathrm{si}})\max\{\mathrm{d}(G^{(1)},G^{(2)}),1\}} \leqslant D^{6+(4-8c_{\mathrm{si}})\max\{\mathrm{d}(G^{(1)},G^{(2)}),1\}}. \tag{5.68}$$

To conclude, we just need to impose that (using $D \geqslant 2$):

$$7 + (4 - 8c_{\mathrm{si}}) \max\left\{\mathrm{d}(G^{(1)}, G^{(2)}), 1\right\} \leqslant -4c_{\mathrm{si}} \max\left\{\mathrm{d}(G^{(1)}, G^{(2)}), 1\right\}, \tag{5.69}$$

which holds for $c_{\mathrm{si}} \geqslant 11/4$. □

Lemma 5.70. *If $4 - 8c_{\text{si}} \leqslant 0$, we have:*

$$2(u_1 + u_2 + 2s) - 8c_{\text{si}} \max\Big\{ \mathrm{d}(G^{(1)}, G^{(2)}), 1, u_1 + u_2 + s \Big\} \leqslant (4 - 8c_{\text{si}}) \max\Big\{ \mathrm{d}(G^{(1)}, G^{(2)}), 1 \Big\} \leqslant 0. \tag{5.71}$$

Proof. For simplicity, let $u_1 + u_2 =: u$ and $\mathrm{d}(G^{(1)}, G^{(2)}) =: \mathrm{d}$. We discuss two cases separately.
(case #1) If $\max\{\mathrm{d}, 1\} \geqslant u + s$ then in particular $\max\{\mathrm{d}, 1\} \geqslant s$ and:

$$2u + 4s - 8c_{\text{si}} \max\{\mathrm{d}, 1\} \leqslant 2u + 2s + 2s - 8c_{\text{si}} \max\{\mathrm{d}, 1\} \leqslant (4 - 8c_{\text{si}}) \max\{\mathrm{d}, 1\}. \tag{5.72}$$

(case #2) If $\max\{\mathrm{d}, 1\} < u + s$ then:

$$2u + 4s - 8c_{\text{si}} \max\{\mathrm{d}, 1\} \leqslant 2u + 4s - 8c_{\text{si}}(u + s) \leqslant 4u + 4s - 8c_{\text{si}}(u + s) = (4 - 8c_{\text{si}})(u + s), \tag{5.73}$$

and the condition $4 - 8c_{\text{si}} \leqslant 0$ ensures that it is less than $(4 - 8c_{\text{si}}) \max\{\mathrm{d}, 1\}$. □

Combining the two statements of proposition 5.44 we conclude that the $\widetilde{P}$ basis of def. 5.42 is almost orthonormal (def. 1.14) with the degree of "almost" controlled by an inverse of D factor involving the distance. We will prove it is sufficient to control the covariance globally over the basis and make it look like an orthonormal basis.

Remark 5.74. *There is a hierarchy across the bases we consider. Let us ignore for a moment the sum over the equivalence class. For a single edge set $\mathbb{T}$ induced by a labelling $\pi^{(1)}(G^{(1)})$, we have the following trivial equivalences:*

- *if $\mathbb{E}_{\mathrm{H}_0}\left[P_{G^{(1)},\pi^{(1)}}\right] = 0$ then $P_{G^{(1)},\pi^{(1)}} \overset{\text{a.s.}}{=} \widehat{P}_{G^{(1)},\pi^{(1)}}(\mathbf{Y})$;*
- *if $\pi^{(1)}(G^{(1)})$ has a single connected component then $\widehat{P}_{G^{(1)},\pi^{(1)}}(\mathbf{Y}) \overset{\text{a.s.}}{=} \overline{P}_{G^{(1)},\pi^{(1)}}(\mathbf{Y})$;*
- *if $\pi^{(1)}(G^{(1)})$ has a single connected component and $\mathbb{E}_{\mathrm{H}_0}\left[P_{G^{(1)},\pi^{(1)}}\right] = 0$ then $P_{G^{(1)},\pi^{(1)}} = \mathbb{E}_{\mathrm{H}_0}\left[P_{G^{(1)},\pi^{(1)}}(\mathbf{Y})\right] \overset{\text{a.s.}}{=} \overline{P}_{G^{(1)},\pi^{(1)}}(\mathbf{Y})$.*

Alternatively for $\pi^{(1)}, \pi^{(2)}$ two labellings of skeletons $G^{(1)}, G^{(2)}$:

- *under the classic basis:*

$$\Big\langle P_{G^{(1)},\pi^{(1)}}, P_{G^{(2)},\pi^{(2)}} \Big\rangle_{\mathrm{H}_0} = \lambda^{|\pi^{(1)} \triangle \pi^{(2)}|} \left(\frac{k}{n}\right)^{|V_{\pi^{(1)} \triangle \pi^{(2)}}|}; \tag{5.75}$$

- *under the $\widehat{P}$ basis:*

$$\begin{aligned} \Big\langle \widehat{P}_{G^{(1)},\pi^{(1)}}, \widehat{P}_{G^{(2)},\pi^{(2)}} \Big\rangle_{\mathrm{H}_0} &= \lambda^{|\pi^{(1)} \triangle \pi^{(2)}|} \left(\frac{k}{n}\right)^{|V_{\pi^{(1)} \triangle \pi^{(2)}}|} - \lambda^{|\pi^{(1)}| + |\pi^{(2)}|} \left(\frac{k}{n}\right)^{|V_{\pi^{(1)}}| + |V_{\pi^{(2)}}|} \\ &\asymp \lambda^{|\pi^{(1)} \triangle \pi^{(2)}|} \left(\frac{k}{n}\right)^{|V_{\pi^{(1)} \triangle \pi^{(2)}}|} \mathbb{1}\Big\{ \pi^{(1)}, \pi^{(2)} \text{ induce conn. graphs} \Big\}; \end{aligned} \tag{5.76}$$

- *under the $\overline{P}$ basis:*

$$\begin{aligned} \left| \Big\langle \overline{P}_{G^{(1)},\pi^{(1)}}, \overline{P}_{G^{(2)},\pi^{(2)}} \Big\rangle_{\mathrm{H}_0} \right| &= \text{non-trivial expression} \\ &\lesssim \lambda^{|\pi^{(1)} \triangle \pi^{(2)}|} \left(\frac{k}{n}\right)^{|V_{\pi^{(1)} \triangle \pi^{(2)}}|} \mathbb{1}\Big\{ \Big(\pi^{(1)}, \pi^{(2)}\Big) \in \pi(\boldsymbol{M}), \text{ for } \boldsymbol{M} \in \mathcal{M}^{\star} \Big\}; \end{aligned} \tag{5.77}$$

  *which is larger if the indicator is non-zero, but it is **a lot less frequently so**.*

In this perspective, we first remove correlations arising from disconnected graphs, then correlations arising from graphs with some disconnected component.

Example 5.78. *In figures 10-15 we place some examples of how different graphs correlate in the basis decomposition we propose. We imagine there are two edge sets induced by labellings $\pi^{(1)}, \pi^{(2)}$ on skeletons $G^{(1)}, G^{(2)}$. The vertices they induce form sets of connected components respectively in purple and orange. They might or might not have shared vertices, and their connections form the "overlap graph".*

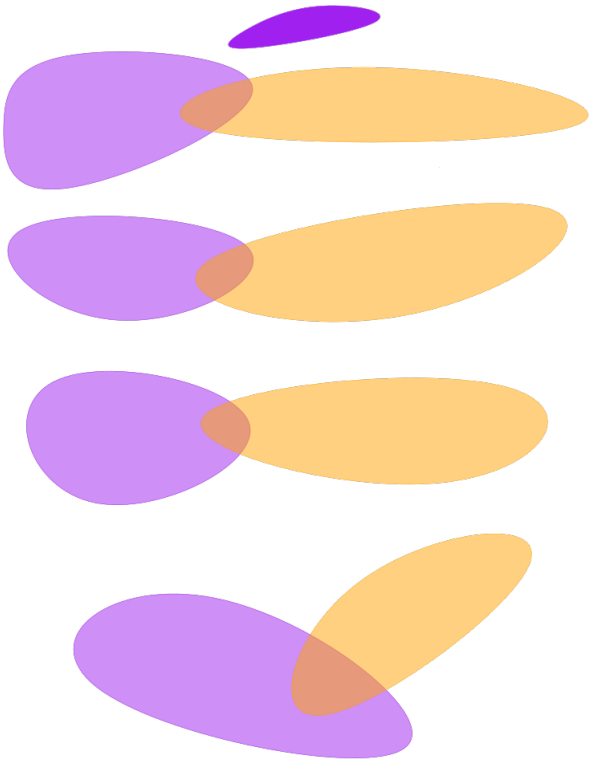


Figure 10: Null alignment
This example has $\left\langle \overline{P}_{G^{(1)},\pi^{(1)}}, \overline{P}_{G^{(2)},\pi^{(2)}} \right\rangle_{\mathrm{H}_0} = 0$ as there is an isolated connected component in dark purple. We are not interested in these.

To finalize the computation we show that the basis is almost orthonormal in the sense of definition 1.14. Having the right control of each entry (prop. 5.44#2), it is a matter of upper bounding the way to count skeletons. The best way to see if a normalized basis is almost orthonormal in the sense of subsection 3.IV is to use the Gershgorin interpretation, which we remind required to find a bound of the form:

$$\sup_{G^{(1)} \in \mathcal{G}_{(\leqslant D)} \cup \{\varnothing\}} \sum_{\substack{G^{(2)} \in \mathcal{G}_{(\leqslant D)} \\ G^{(1)} \neq G^{(2)}}} \left| \mathrm{CoV}_{\mathrm{H}_0} \left[\psi(\cdot; G^{(1)}) \psi(\cdot; G^{(2)}) \right] \right| \leqslant \frac{1}{2}, \tag{5.79}$$

or even better by some $D^{-\mathsf{pow}}$ where pow is some possibly explicit control.

Remark 5.80 (Important)**.** *The corrected rescaled monomial basis of definition 5.42 is a collection of $\widetilde{P}_G$ polynomials enriched with the unit function. We can write the basis into a vector $\boldsymbol{\psi}$, where by convention the first term is $\psi_1 = 1$ and the others are the $\widetilde{P}_G$ for $G \in \mathcal{G}_{(\leqslant D)}$. In particular, the order in which we write them does not matter (the outer product $\boldsymbol{G} = \boldsymbol{\psi}\boldsymbol{\psi}^\top$ is rotation invariant). A generic entry is $G_{G^{(1)},G^{(2)}} = \mathbb{E}_{\mathrm{H}_0}\left[\psi_{G^{(1)}}\psi_{G^{(2)}}\right] = \mathbb{E}_{\mathrm{H}_0}\left[\widetilde{P}_{G^{(1)}}\widetilde{P}_{G^{(2)}}\right]$ or $G_{1,G^{(1)}} = \mathbb{E}_{\mathrm{H}_0}\left[\psi_0\psi_{G^{(1)}}\right] = \mathbb{E}_{\mathrm{H}_0}\left[1 \cdot \widetilde{P}_{G^{(1)}}\right]$. In the former case, we use proposition 5.44, in the latter, we simply notice that $\mathbb{E}_{\mathrm{H}_0}\left[\psi_0\psi_{G^{(1)}}\right] = 0$, since the basis is centered. If $N = |\mathcal{G}_{(\leqslant D)}| + 1$, and we choose a reference order $G_1, \ldots, G_{|\mathcal{G}_{(\leqslant D)}|}$ for the labelled graphs to form $\boldsymbol{\psi}$ the Gram matrix is:*

$$\boldsymbol{G} = \begin{array}{c} \\ \psi_1 \\ \psi_2 \\ \psi_3 \\ \vdots \\ \psi_N \end{array} \begin{array}{c} \begin{array}{ccccc} \psi_1 & \psi_2 & \psi_3 & \cdots & \psi_N \end{array} \\ \begin{pmatrix} G_{11} & 0 & 0 & \cdots & 0 \\ 0 & G_{22} & G_{23} & \cdots & G_{2N} \\ 0 & G_{32} & G_{33} & \cdots & G_{3N} \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & G_{N2} & G_{N3} & \cdots & G_{NN} \end{pmatrix} \end{array}. \tag{5.81}$$

In particular:

- *the off-diagonal sums of $\boldsymbol{G}$ are independent of the first row/column;*

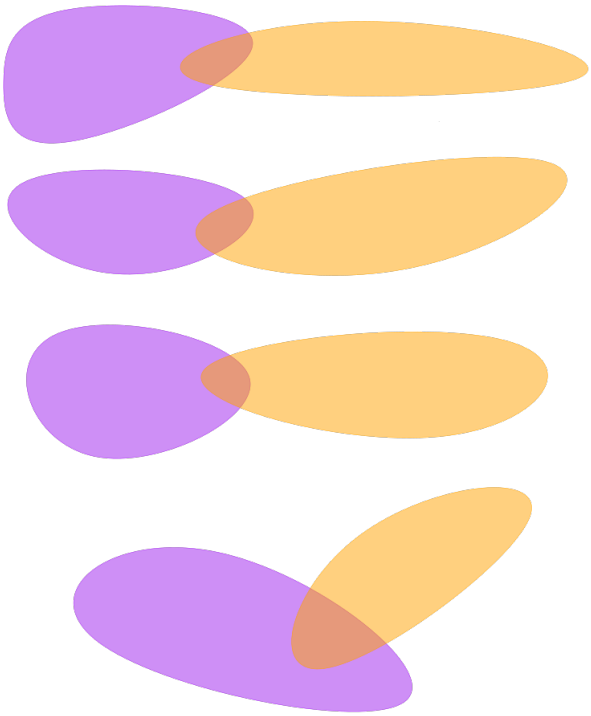


Figure 11: Non-null *minimal* alignment vertex sets
Each vertex set is connected to one vertex set of the other graph. The alignment is non-trivial.

- *the diagonal terms are asymptotically unity if D diverges by proposition 5.44#1, or the quantitative bound gives us that they are bounded below by $1/2$ strictly;*
- *proposition 5.44#2 bounds each green term in absolute value by $D^{-4c_{\mathrm{si}}\,\mathrm{d}(G^{(1)},G^{(2)})}$;*
- *the Gram matrix is symmetric so the bound on the green terms is the same on the other side of the matrix.*

For this reason in the following proposition we ignore the first row/column: it suffices to work out how the off-diagonal sum over the skeletons compares with the diagonal terms.

Proposition 5.82. *Suppose assumption 3.1 holds. Recall the construction of the $\widetilde{P}$ basis from def. 5.42. Define the Gram matrix:*

$$\boldsymbol{G} \coloneqq \left(\mathbb{E}_{\mathrm{H}_0}\left[\widetilde{P}_{G^{(1)}}\widetilde{P}_{G^{(2)}}\right]\right)_{G^{(1)},G^{(2)}\in\mathcal{G}_{(\leqslant D)}\cup\{\varnothing\}}, \tag{5.83}$$

where by $\varnothing$ we mean that we add the constant function to the underlying basis (refer to the remark above, remark 2.20 and definition 5.42). For all $G^{(1)} \neq G^{(2)}$ when D diverges:

$$\boldsymbol{G}_{G^{(1)},G^{(1)}} \asymp 1 \qquad \textit{and} \qquad \sum_{\substack{G^{(2)}\in\mathcal{G}_{(\leqslant D)}\cup\{\varnothing\}\\ G^{(2)}\neq G^{(1)}}} \left|\boldsymbol{G}_{G^{(1)},G^{(2)}}\right| \lesssim D^{-c_{\mathrm{si}}\,\mathrm{d}(G^{(1)},G^{(2)})}. \tag{5.84}$$

As a consequence the matrix is an approximate isometry. If D does not diverge, it is nevertheless diagonal dominant, meaning that the eigenvalues are a constant multiple of unity.

Proof. The consequence is immediate for D diverging with n and for the weaker result it suffices to notice that the quantitative bounds chain the eigenvalues into the $[1/2, 3/2]$ strip in the worst case when $D = 2$ since c_{si} is a large constant. We make this explicit by proving the claims in equation 5.84.
The first is true since we rescale by the dominating term $\nu(G^{(1)})$: it is a mere restatement of proposition 5.44.
The second claim follows by counting the number of skeletons that are at fixed distance from a given one,

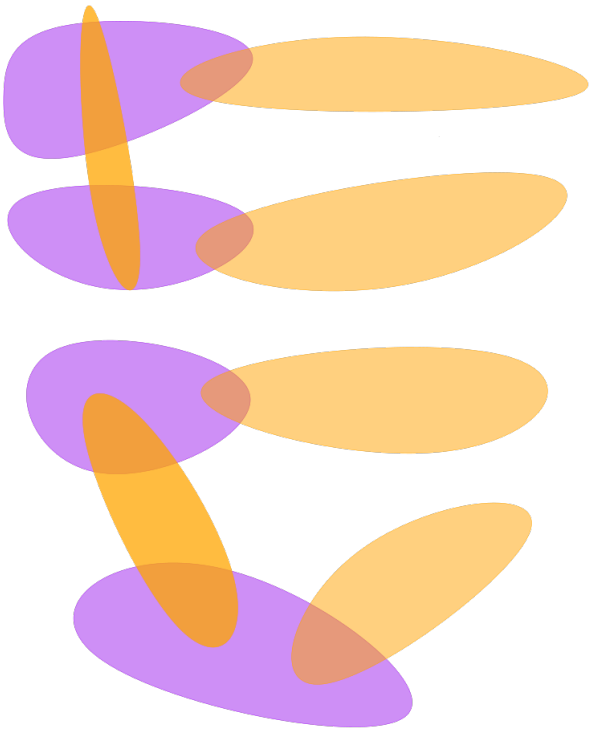


Figure 12: Non-null alignment vertex sets
Each vertex set is connected to at least one vertex set of the other graph. The alignment is non-trivial.

which is lem. 4.74. The covariance of a graph and the unit function is null (i.e. the discussion of remark 5.80). We find:

$$\begin{aligned}
\sum_{G^{(2)}\in\mathcal{G}_{(\leqslant D)}} \left|\mathbb{E}_{H_0}\left[\widetilde{P}_{G^{(1)}}\widetilde{P}_{G^{(2)}}\right]\right| &\lesssim \sum_{G^{(2)}\in\mathcal{G}_{(\leqslant D)}} D^{-4c_{\text{si}}\,\mathrm{d}(G^{(1)},G^{(2)})} \\
&= \sum_{d=1}^{D} \#\left\{G^{(2)} \mid \mathrm{d}(G^{(1)},G^{(2)}) = d\right\} D^{-4c_{\text{si}}d} \\
&\leqslant \sum_{d=1}^{D} (d+D)^{2d} D^{-4c_{\text{si}}d} \\
&\leqslant \sum_{d=1}^{D} D^{d(4-c_{\text{si}})} \\
&\leqslant D^{(4-c_{\text{si}})+1} \\
&\leqslant D^{-c_{\text{si}}/2},
\end{aligned} \tag{5.85}$$

where we used $c_{\text{si}} \geqslant 10$. Therefore, we have:

$$G_{G^{(1)},G^{(1)}} - \sum_{G^{(2)}\neq G^{(1)}\in\mathcal{G}_{(\leqslant D)}} \left|G_{G^{(1)},G^{(2)}}\right| \geqslant 1 - D^{-c_{\text{si}}/2} \geqslant \frac{1}{2}, \tag{5.86}$$

since $D \geqslant 2$ and c_{si} is large enough (say, larger than 2). The eigenvalues of the Gram matrix are all in the strip $[1/2, 3/2]$ and get closer exponentially fast to being in $[1-\epsilon, 1+\epsilon]$. In any way, we have:

$$\|\boldsymbol{\alpha}\|_2 \lesssim \|\boldsymbol{\alpha}\|_{\boldsymbol{G}} \lesssim \|\boldsymbol{\alpha}\|_2, \qquad \boldsymbol{\alpha} = (\alpha_G)_{G\in\mathcal{G}_{(\leqslant D)}}, \qquad \|\boldsymbol{\alpha}\|_{\boldsymbol{G}} = \boldsymbol{\alpha}^\top \boldsymbol{G}\boldsymbol{\alpha}, \tag{5.87}$$

since the eigenvalues are approximately a constant multiple of unity. The last equation is the requirement in definition 1.14, so the basis in $\boldsymbol{G}$ is almost orthonormal. □

Comparing with section 3 we have all the formal results announced. It remains to prove the main theorem. We do so in the next section.

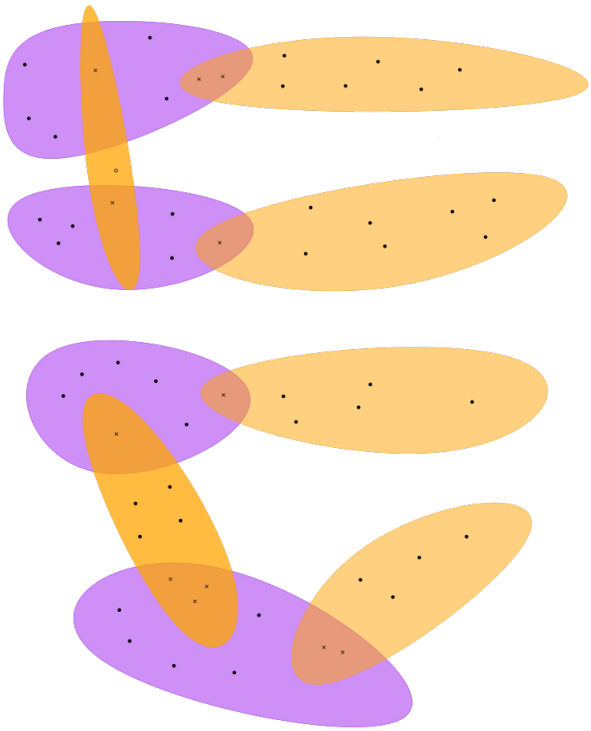


Figure 13: Shared vertices depiction
We highlight how the $q = 9$ number of shared vertices could be. They are in the region where the colored vertex sets overlap. This is not sufficient to understand the number of edges in the overlap graph as discussed in figure 6.

6 BOUNDING THE ADVANTAGE: PROOF OF THE MAIN THEOREM

In this final section we prove theorem 3.8.

The seemingly involved expression of the advantage in equation 3.27 simplifies greatly thanks to our almost orthonormal basis (def. 1.14), which is the one from definition 5.42. Using proposition 5.82, we can renormalize the coefficients:

$$\begin{aligned}
\mathrm{Adv}_{(\leqslant D)}(\mathrm{H}_0, \mathrm{H}_1) &= \sup_{(\alpha_G)_{G\in\mathcal{G}_{(\leqslant D)}\cup\{\varnothing\}}} \frac{\mathbb{E}_{\mathrm{H}_1}\left[\sum_{G\in\mathcal{G}_{(\leqslant D)}\cup\{\varnothing\}} \alpha_G \widetilde{P}_G\right]}{\sqrt{\mathbb{E}_{\mathrm{H}_0}\left[\left(\sum_{G\in\mathcal{G}_{(\leqslant D)}\cup\{\varnothing\}} \alpha_G \widetilde{P}_G\right)^2\right]}} \\
&\lesssim \sup_{(\alpha_G)_{G\in\mathcal{G}_{(\leqslant D)}\cup\{\varnothing\}}} \frac{\mathbb{E}_{\mathrm{H}_1}\left[\sum_{G\in\mathcal{G}_{(\leqslant D)}\cup\{\varnothing\}} \alpha_G \widetilde{P}_G\right]}{\|(\alpha_G)_{G\in\mathcal{G}_{(\leqslant D)}\cup\{\varnothing\}}\|_2} \qquad (6.1) \\
&= \sup_{(\alpha_G)_{G\in\mathcal{G}_{(\leqslant D)}\cup\{\varnothing\}}:\|(\alpha_G)_{G\in\mathcal{G}_{(\leqslant D)}\cup\{\varnothing\}}\|_2=1} \mathbb{E}_{\mathrm{H}_1}\left[\sum_{G\in\mathcal{G}_{(\leqslant D)}\cup\{\varnothing\}} \alpha_G \widetilde{P}_G\right],
\end{aligned}$$

which is equation 3.38. To begin, we take out the constant function corresponding to the empty graph, which contributes as $\alpha_\varnothing \mathbb{E}_{\mathrm{H}_1}\left[P_\varnothing\right] = \alpha_\varnothing \leqslant 1$. Since we want to show that the advantage is $1 + o(1)$, i.e. that weak separation is impossible (def. 2.12), notice that the way we choose the constant function is irrelevant: in the $\lesssim$ step above we lost a constant, but we can just absorb it in the empty graph element of the basis to get back 1. In other words, what matters is showing that the part of the advantage where we sum over $\mathcal{G}_{(\leqslant D)}$ is vanishing. Therefore, we move to studying the expectation under H_1 of the non-trivial basis elements. This reduces to computing the expectation under H_1 of each connected component. Indeed, each labelled graph inside a

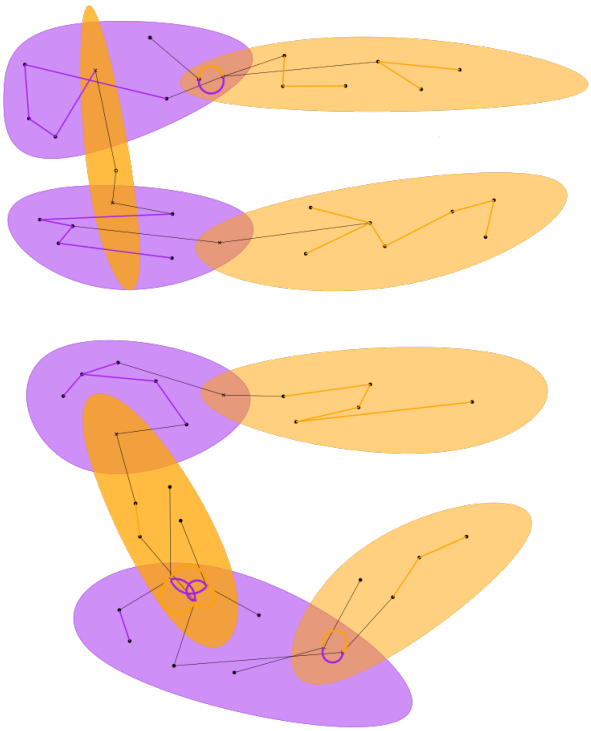


Figure 14: Minimal tree for symmetric difference
To obtain the minimal number of edges in the symmetric difference, we add in thick colored lines all connections within shared vertices in both edge sets (so that they do not appear in the symmetric difference). Then, we connect the shared vertices with the non-shared vertices with black lines, then we use purple (resp. orange) lines to form a tree on the purple (resp. orange) connected components. These types of matchings attain the lower bound $|E_{\triangle}| = |V_{\triangle}| - \#\text{CC}$ of lemma 4.59.

$G \in \mathcal{G}_{(\leqslant D)}$ (see defs. 3.19 - 3.21) has the same expectation. We find, using the language of section 4 and the definition of the $\nu(G)$ dominating term (def. 5.40):

$$\begin{aligned}\mathbb{E}_{\mathrm{H}_1}\left[\widetilde{P}_G\right] &= |\Pi(|V|)|\mathbb{E}_{\mathrm{H}_1}\left[\widetilde{P}_{G,\pi}\right] \\ &= |\Pi(|V|)|\frac{1}{\sqrt{\nu(G)}}\mathbb{E}_{\mathrm{H}_1}\left[\prod_{\ell=1}^{m} P_{G_\ell,\pi} - \mathbb{E}_{\mathrm{H}_0}\left[P_{G_{\ell,\pi}}\right]\right] \\ &= \frac{1}{\sqrt{\nu(G)}}|\Pi(|V|)|\prod_{\ell=1}^{m}\mathbb{E}_{\mathrm{H}_1}\left[P_{G_\ell,\pi} - \mathbb{E}_{\mathrm{H}_0}\left[P_{G_{\ell,\pi}}\right]\right]\end{aligned} \tag{6.2}$$

for any $G \in \mathcal{G}_{(\leqslant D)}, \pi \in \Pi_{|V|}$, where in the last step we used independence. A skeleton becomes a labelled graph by taking any injection, so $|\Pi(|V|)| = \frac{n!}{(n-|V|)!}$ the number of injections from $|V|$ to n.[16] Concerning the expectation, different structures of $(\mathrm{H}_0, \mathrm{H}_1)$ are analogous but return distinct quantities. In what follows, we focus on testing $\mathrm{H}_0 : \mathcal{P}_{\boldsymbol{\theta}}$ for $\boldsymbol{\theta} = (k, \lambda)$ against $\mathrm{H}_1 : \mathcal{P}_{\boldsymbol{\theta}'}$ for $\boldsymbol{\theta}' = (\lambda + \eta, k)$, which is problem 1.3 in the version of perturbed signal strength. The analysis when we perturb k follows by the same principles (see subsec. 6.III). Morally, the expression of the expectation under the planted sub-matrix model is generic, and we just need to adjust the signal strength. In the spirit of equation 3.17 for all $\ell \in [m]$:

$$\mathbb{E}_{\mathrm{H}_1}\left[P_{G_{\ell,\pi}}\right] = (\lambda + \eta)^{|E_\ell|}\left(\frac{k}{n}\right)^{|V_\ell|} \tag{6.3}$$

$$\mathbb{E}_{\mathrm{H}_0}\left[P_{G_{\ell,\pi}}\right] = \lambda^{|E_\ell|}\left(\frac{k}{n}\right)^{|V_\ell|}. \tag{6.4}$$

[16] This is lemma 4.50#2.

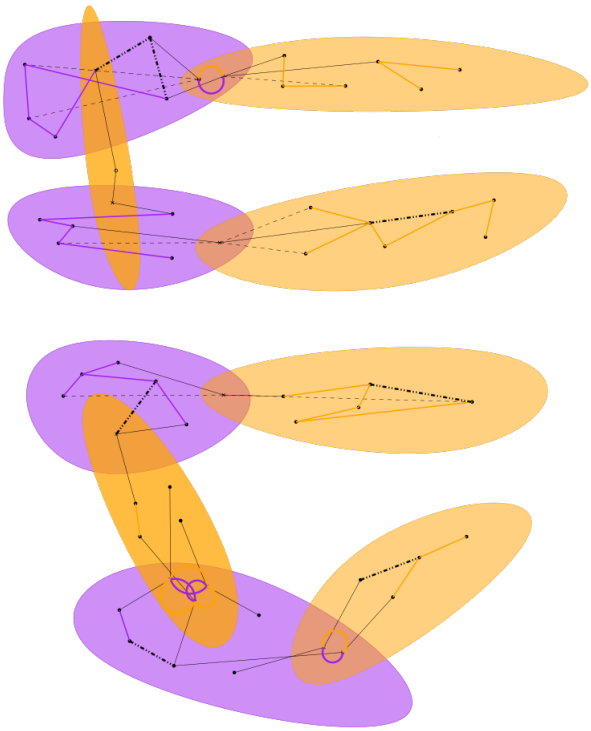


Figure 15: A *redundant* graph
To the graph of fig. 14 we add some redundant connections in the symmetric difference. In dashed lines we include more connections within shared and non-shared vertices. In dotted dashed lines we add connections that remove the tree structure in the non-shared vertices.

Combining with the explicit form of $\nu(G)$ in definition 5.40 and after some algebraic manipulations we find:

$$\mathbb{E}_{H_1}\left[\widetilde{P}_G\right] = \sqrt{\frac{n!}{(n-|V|)!}\frac{}{|\mathrm{Aut}(G)|}}\left(\frac{k}{n}\right)^{|V|}\lambda^{|E|}\prod_{\ell=1}^{m}\left(1+\frac{\eta}{\lambda}\right)^{|E_\ell|}-1. \tag{6.5}$$

If we want something simpler, we can open the product with the binomial theorem and apply a rough bound on the geometric sum:

$$\left(1+\frac{\eta}{\lambda}\right)^{|E_\ell|}-1 = \sum_{t=1}^{|E_\ell|}\binom{|E_\ell|}{t}\left(\frac{\eta}{\lambda}\right)^t \tag{6.6}$$

$$\leqslant \sum_{t=1}^{|E_\ell|}\left(|E_\ell|\frac{\eta}{\lambda}\right)^t \tag{6.7}$$

$$\leqslant |E_\ell|\left(|E_\ell|\frac{\eta}{\lambda}\right) \qquad \eta \leqslant \frac{\lambda}{D}, 1 \leqslant |E_\ell| \leqslant D, \tag{6.8}$$

where the condition on η, λ is justified in remark 3.10 and in assumption 3.7. Putting it all together:

$$\begin{aligned}\mathrm{Adv}_{(\leqslant D)}(H_0, H_1) &\leqslant 1 + \sup_{(\alpha_G)_{G\in\mathcal{G}_{(\leqslant D)}}=\boldsymbol{\alpha}:\|\boldsymbol{\alpha}\|_2^2=1}\mathbb{E}_{H_1}\left[\sum_{G\in\mathcal{G}_{(\leqslant D)}}\alpha_G\widetilde{P}_G\right] \\ &\leqslant 1 + \sup_{(\alpha_G)_{G\in\mathcal{G}_{(\leqslant D)}}=\boldsymbol{\alpha}:\|\boldsymbol{\alpha}\|_2^2=1}\sum_{G\in\mathcal{G}_{(\leqslant D)}}\alpha_G\sqrt{\frac{n!}{(n-|V|)!}\frac{}{|\mathrm{Aut}(G)|}}\left(\frac{k}{n}\right)^{|V|}\lambda^{|E|}\prod_{\ell=1}^{m}|E_\ell|^2\frac{\eta}{\lambda} \\ &\leqslant 1 + \sum_{G\in\mathcal{G}_{(\leqslant D)}}\left(\frac{k}{\sqrt{n}}\right)^{|V|}\lambda^{|E|}D^{2m}\left(\frac{\eta}{\lambda}\right)^m,\end{aligned} \tag{6.9}$$

where we used $\alpha_G \leqslant 1$ and $\sqrt{n!/(n-|V|)!/|\mathrm{Aut}(G)|} \leqslant n^{|V|/2}$. In particular, with these two steps we completely neglect the connectivity of the graph. We remove the dependency on the number of connections in each connected

component and the potential discount by the size of the automorphism group, and more in general the dependence on G inside the sum that is not possible to group by quantities related to graphs $G \in \mathcal{G}_{(\leqslant D)}$. With these simplifications, we rewrite the upper bound as

$$\mathrm{Adv}_{(\leqslant D)}(\mathrm{H}_0, \mathrm{H}_1) \leqslant 1 + \sum_{\substack{2 \leqslant a \leqslant 2D \\ 1 \leqslant b \leqslant D \\ 1 \leqslant m \leqslant D}} N_{a,b,m} \left(\frac{k}{\sqrt{n}}\right)^a \lambda^b D^{2m} \left(\frac{\eta}{\lambda}\right)^m, \tag{6.10}$$

where:

$$N_{a,b,m} := \# \left\{ G \in \mathcal{G}_{(\leqslant D)} \mid |V| = a, |E| = b, m \text{ connected components} \right\}. \tag{6.11}$$

To upper bound further, we go step-by-step.

6.1 *Digression: element-wise bound*

Since we want to show that the sum over $G \in \mathcal{G}_{(\leqslant D)}$ is small, we will first check that each term in the worst case is small. As is evident from the expression, it depends on $|V|, |E|, s$, respectively the number of vertices, edges and connected components. To understand which come into play, we discuss them separately.
To simplify matters, we will take D of the order polylog(n) and write $\lesssim_{\log}$ to discard a polynomial in D with constant degree.

Remark 6.12. *Throughout, the intuition is three-fold:*

- *in appendix B we show that the optimal algorithm is a sum-type statistic like $\sum_{(i,j)\in E} Y_{ij}$, which is the symmetrization of Y_{ij}, a "very small tree" over two vertices and many connected components;*
- *our new basis is not impacted by many connected components;*
- *the nice bounds $|E| \geqslant |V| - m$ and $|V| \geqslant 2m$ from lemma 4.59 are worst-case when $\lambda \leqslant 1, k/n < 1$ and coincide with the optimal algorithm.*

OPTIMAL ALGORITHM The small tree corresponds to $|E| = 1, |V| = 2, m = 1$, which gives:

$$\lambda \left(\frac{k}{\sqrt{n}}\right)^2 \frac{\eta}{\lambda} D^2 \lesssim_{\log} 1 \iff \eta \frac{k^2}{n} \lesssim_{\log} 1. \tag{6.13}$$

In what follows, we will show that when there is only one connected component and when there are many the optimal algorithm is still the upper bound.

SINGLE CONNECTED COMPONENT When the graph has $m = 1$ connected component(s) we have:

$$\begin{aligned} \lambda^{|E|} \left(\frac{k}{\sqrt{n}}\right)^{|V|} \frac{\eta}{\lambda} D^2 &\leqslant \lambda^{|V|-1} \left(\frac{k}{\sqrt{n}}\right)^{|V|} \frac{\eta}{\lambda} D^2 \\ &= \left(\lambda \frac{k}{\sqrt{n}}\right)^{|V|-2} \left(\eta \frac{k^2}{n} D\right)^2. \end{aligned} \tag{6.14}$$

Imposing that it is $\lesssim_{\log} 1$, the first term is small by assumption 3.1 since $|V| \geqslant 2$, the second coincides with the condition we got from the algorithm. Again we get $\eta k^2/n \lesssim_{\log} 1$.

MANY CONNECTED COMPONENTS When $m \geqslant 1$ we have:

$$\begin{aligned} \lambda^{|E|} \left(\frac{k}{\sqrt{n}}\right)^{|V|} \left(\frac{\eta}{\lambda}\right)^m D^{2m} &\leqslant \lambda^{|V|-m} \left(\frac{k}{\sqrt{n}}\right)^{|V|} \left(\frac{\eta}{\lambda}\right)^m D^{2m} \\ &= \left(\lambda \frac{k}{\sqrt{n}}\right)^{|V|-2m} \left(\eta \frac{k^2}{n} D^2\right)^m. \end{aligned} \tag{6.15}$$

Since there are no isolated vertices, $|V| \geqslant 2m$, and we can use again assumption 3.1 to reduce to the optimal algorithm condition that $\eta k^2/n \lesssim_{\log} 1$ and $m = 1$.
To conclude, we move to the full sum.

Let us group the sum in terms of $|V|, |E|, s$ which are the influencing factors. In other words, we consider equation 6.10. The first observation is that the cardinality of the sum in equation 6.10 is bounded by $(2D) \cdot D \cdot D \leqslant D^4$, which is polynomial in D. The size of the set $N_{a,b,m}$ is also bounded, but we have to be careful and counterbalance it with our signal terms.

Lemma 6.16. *It holds that* $N_{a,b,m} \leqslant D^{4b}$.

Proof. We have $N_{a,b,m} \leqslant N_{a,b} \leqslant a^{2b} \leqslant D^{4b}$ where we used $D \geqslant 2$ and the constraints that a skeleton over a vertices with b edges is an abstract graph where we need to choose where to place edges and each edge has less than a^2 choices to make. Moreover, inside $\mathcal{G}_{(\leqslant D)}$ we have $a \leqslant 2D$. □

Combining these two estimates with equation 6.10 we find:

$$\mathrm{Adv}_{(\leqslant D)}(\mathrm{H}_0, \mathrm{H}_1) \leqslant 1 + D^4 \sup_{(a,b,m) \text{ feasible}} D^{4b+2m} \lambda^b \left(\frac{k}{\sqrt{n}}\right)^a \left(\frac{\eta}{\lambda}\right)^m, \tag{6.17}$$

where by "feasible" we mean that we take into account all triplets (a, b, m) of numbers of vertices, edges and connected components of skeletons in $\mathcal{G}_{(\leqslant D)}$. The three key feasibility conditions we use come from lemma 4.59:

- $b \geqslant a - m$ (i.e. $|E| \geqslant |V| - m$);
- $a \geqslant 2m$ (i.e. $|V| \geqslant 2m$, no isolated edges).

Moreover, we use the algorithm intuition of the previous subsection. Bounding the term inside the supremum, for all feasible triplets:

$$D^{4b+2m} \lambda^b \left(\frac{k}{\sqrt{n}}\right)^a \left(\frac{\eta}{\lambda}\right)^m = \left(\lambda \frac{k}{\sqrt{n}}\right)^{a-2m} \left(\eta \frac{k^2}{n}\right)^m \cdot \lambda^{b+m-a} D^{4b+2m}. \tag{6.18}$$

An application of assumptions 3.1 - 3.7 concludes the proof. Indeed, including the D^4 term from equation 6.17:

$$\mathrm{Adv}_{(\leqslant D)}(\mathrm{H}_0, \mathrm{H}_1) \leqslant 1 + \sup_{(a,b,m) \text{ feasible}} D^{4+4b+2m-8c_{\mathrm{si}}(a-2m+m+b+m-a)}. \tag{6.19}$$

Rearranging, we can use $4 \leqslant 4b$, $b \geqslant 1$, and $m \leqslant b$ to have all terms depending on the number of edges. Then, a simple upper bound is:

$$\mathrm{Adv}_{(\leqslant D)}(\mathrm{H}_0, \mathrm{H}_1) \leqslant 1 + \sup_{1 \leqslant b \leqslant D} D^{10b - 8c_{\mathrm{si}}(b)} \leqslant D^{10-8c_{\mathrm{si}}} \leqslant 1 + \frac{1}{D}, \text{ if } 10 - 8c_{\mathrm{si}} \leqslant -1 \iff c_{\mathrm{si}} \geqslant \frac{11}{8}, \tag{6.20}$$

which is the case, since c_{si} is a large constant.

Remark 6.21. *With another path, we found that the worst-case bound is at the number of edges $b = 1$, which corresponds to the optimal algorithm.*

EXPLICIT ROUTE We attempt to see finely how the optimal algorithm and the assumptions come into play. As we know, the two assumptions ensure that all terms inside the parenthesis are $\leqslant D^{-f_i}$ for some $f_i > 0$ positive factor. To be clear, in the statement we require that they are all less than a common $D^{-8c_{\mathrm{si}}}$. We see that we have three signal terms, so three factors, and two annoying positive powers, namely $4b$ and $2m$. This is an over-determined system. We can choose two particular f_1, f_2 to cancel the D^{4b} and D^{2m} terms and then freely choose f_3 to make the quantity as small as we wish. In particular, we take into account the D^4 factor outside (from the cardinality of the sum, namely eqn. 6.17), and want to obtain that overall $\mathrm{Adv}_{(\leqslant D)}(\mathrm{H}_0, \mathrm{H}_1) \leqslant 1 + 1/D$. Using the factor perspective:

$$\mathrm{Adv}_{(\leqslant D)}(\mathrm{H}_0, \mathrm{H}_1) \leqslant 1 + D^{\mathsf{pow}}, \qquad \mathsf{pow} := 4 + 4b + 2m - f_1(a - 2m) - f_2 m - f_3(b + m - a). \tag{6.22}$$

We wish to find $(f_1, f_2, f_3) \subset \mathbb{R}_+$ such that $\mathsf{pow} \leqslant -1$ for all feasible (a, b, m).
One direct solution is $f_3 = 4$, which cancels the b, then $f_1 = 8$ and any $f_2 \geqslant 6$. The generic form is recovered perhaps by regrouping pow and using some trivial bounds:

$$\begin{aligned} \mathsf{pow} &= 4 + b(4 - f_3) + a(f_3 - f_1) + m(2 + 2f_1 - f_2 - f_3) \\ &\leqslant b(4 - f_3) + a(f_3 - f_1) + m(6 + 2f_1 - f_2 - f_3), \end{aligned} \tag{6.23}$$

since $4 \leqslant 4m$ as $m \geqslant 1$. Then, using that $a \geqslant 2$, $b \geqslant 1$, $m \geqslant 1$ we just need to enforce that all coefficients are negative. To see that the optimal algorithm still pops up, under the condition $f_3 \geqslant 4$, we can use $b \geqslant a - m$ to find:

$$\begin{aligned} \mathsf{pow} &\leqslant a(-f_1 + f_3 + 4 - f_3) + m(2 + 2f_1 - f_2 - f_3 + f_3 - 4) \\ &= a(-f_1 + 4) + m(2 + 2f_1 - f_2 - 4), \end{aligned} \tag{6.24}$$

and for $f_1 \geqslant 4$ we can discard the first parenthesis to find that we just need $f_2 \geqslant 2f_1 - 1$ in the worst case when $m = 1$. Needless to say, assumption 3.1 on the null hypotheses and assumption 3.7 on the perturbation satisfy these requirements.

6.III *Adaptation for perturbation on signal size*

In this section, we computed everything explicitly for $(\mathrm{H}_0, \mathrm{H}_1)$ as in problem 1.3 where we perturb the signal *strength* λ to $\lambda + \eta$. The reasoning for when we perturb the signal *size* from k to $k + \zeta$ is analogous. We find:

$$\mathbb{E}_{\mathrm{H}_1}\left[P_{G_{\ell,\pi}}\right] = \lambda^{|E_\ell|}\left(\frac{k+\zeta}{n}\right)^{|V_\ell|}, \tag{6.25}$$

for each connected component, which is equation 6.3. Then, being careful with $|V_\ell| \geqslant 2$ for all ℓ connected components, the inequality in the chain starting from equation 6.6 becomes:

$$\left(1 + \frac{\zeta}{k}\right)^{|V_\ell|} - 1 \leqslant |V_\ell|\left(|V_\ell|\frac{\zeta}{k}\right)^2 \leqslant D^6\frac{\zeta^2}{k^2}, \qquad \text{under } \zeta \leqslant \frac{k}{2D}. \tag{6.26}$$

Again, we assumed $\zeta \lesssim_{\log} k$, or precisely that $\zeta \leqslant k/2D$, i.e. that the perturbation is small enough as to not get into the pure noise regime (see remark 3.10). Under the rescaled basis, we find that the advantage takes form:

$$\mathrm{Adv}_{(\leqslant D)}(\mathrm{H}_0, \mathrm{H}_1) \leqslant 1 + \sum_{\substack{2\leqslant a\leqslant 2D \\ 1\leqslant b\leqslant D \\ 1\leqslant m\leqslant D}} N_{a,b,m}\left(\frac{k}{\sqrt{n}}\right)^a \lambda^b D^{6m}\left(\frac{\zeta}{k}\right)^{2m}, \tag{6.27}$$

which is the analog of equation 6.10, with slightly changed coefficients. If we use the same intuition on the optimal algorithm of subsection 6.I, we find that the condition $\zeta \lesssim_{\log} \sqrt{n/\lambda}$ extends nicely throughout. Replicating subsection 6.II, the quantitative condition is that $\zeta\sqrt{\lambda}/\sqrt{n} \leqslant D^{-8c_{\mathrm{si}}}$, which is in accordance with assumption 3.7 when c_{si} is large enough. For example, we can ask that $c_{\mathrm{si}} \geqslant 13/8$ for this last inequality to hold.

6.IV *Final remarks and extensions*

OUR CONTRIBUTION is a very explicit technique to derive low-degree lower bounds for a typical problem: the planted sub-matrix model (eqn. 1.6) in the setting of complex testing (prob 1.3). In particular, the advantage from definition 1.10 is vanishing in the regimes of assumption 3.1 - 3.7, and no degree D polynomial can perform weak detection (cf. definition 2.5) between H_0 and H_1 observations $\mathbf{Y}$. Using the well-grounded suggestion at the heart of low-degree polynomials (see sec. 2) we expect that for $D \approx_{\log} \log n$ we recover known results up poly logarithmic factors. The novelty in the technique is the explicit way of recovering these bounds via the construction of an almost orthonormal basis (def. 1.14). Using conditional independence and symmetries in the distributions, the canonical basis of definition 2.18 turns into a basis where the correlations are vanishing, i.e. that of definition 5.42. We hope that thanks to this new strategy finer proofs will follow for other models.

As a matter of fact the technology of skeletons is not tailored to models as that of equation 1.6 only. Let us briefly outline why this is the case **for perturbations only on the size of the signal** k into $k + \zeta$.
Firstly, any hypothesis test with $(\mathrm{H}_0, \mathrm{H}_1)$ invariant to permutations admits an expression of the advantage as in equation 3.27; we never used the model explicitly in the discussion of subsection 3.III. It is also quite natural to translate the binary matrix case in which $\mathbf{Y} \in \{-1, +1\}^{n\times n}$ to a generic observation where:

$$\mathrm{Y}_{ij} = \begin{cases} 1 - q & \text{with probability } q + \mathrm{X}_{ij} \\ -q & \text{with probability } 1 - q - \mathrm{X}_{ij} \end{cases}, \qquad p := q + \lambda. \tag{6.28}$$

In looking for an invariant basis, the formalism of skeletons is still helpful, and in the proofs of section 5 we have far more room in the inequalities. If we maximize the use of conditional independence, a basis like that of definition 5.42 is almost orthonormal as long as we have an analog of lemma 5.27, and a control on the correlation like in equation 3.17. To be precise, we can keep similar results to the proofs once we establish that for a given model the following assumptions hold.

Assumption 6.29 (Generalization)**.** *For an observation as in equation 6.28 it holds that:*

(A1) The distribution of $\mathbf{Y}$ *is invariant to permutations* $\sigma : [n] \mapsto [n]$ *acting as* $\sigma(\boldsymbol{Y}) = (Y_{\sigma(i)\sigma(j)})_{i,j\in[n]}$.

(A2) For any skeletons $G^{(1)}, G^{(2)} \in \mathcal{G}_{(\leqslant D)}$ *and labellings* $(\pi^{(1)}, \pi^{(2)}) \in \Pi(\boldsymbol{M})$ *it holds that:*

$$\left|\mathbb{E}_{\mathrm{H}_0}\left[P_{G^{(1)},\pi^{(1)}}\right]\mathbb{E}_{\mathrm{H}_0}\left[P_{G^{(2)},\pi^{(2)}}\right]\right| \leqslant c_{\mathrm{m}2}\left(D^{c_{\mathrm{m}1}}\frac{k}{n}\right)^{|\boldsymbol{M}_{\mathrm{PM}}|}\left|\mathbb{E}_{\mathrm{H}_0}\left[P_{G^{(1)},\pi^{(1)}}P_{G^{(2)},\pi^{(2)}}\right]\right|, \tag{6.30}$$

for some constants $c_{\mathrm{m},i}$*, where* P *is the canonical basis (def. 2.18) and* $\boldsymbol{M}_{\mathrm{PM}}$ *is the set of perfect matchings from section 4.*

(A3) For some constants $c_{\mathrm{v},i}$ *where* $i = 1, 2$ *it holds that:*

$$\mathbb{E}_{\mathrm{H}_0}\left[P_{G^{(1)},\pi^{(1)}}P_{G^{(2)},\pi^{(2)}}\right] \leqslant c_{\mathrm{v}2}\left(D^{c_{\mathrm{v}1}}\lambda\right)^{|E_\triangle|}(\overline{qr})^{|E^{(1)}|+|E^{(2)}|-|E_\triangle|}\left(D^{c_{\mathrm{v}1}}\frac{k}{n}\right)^{|V_\triangle|-\#\mathrm{CC}}, \tag{6.31}$$

where again $E_\triangle, V_\triangle, \#\mathrm{CC}$ *are respectively edges, vertices and connected components of the symmetric difference (see sec. 4) and:*

$$\overline{q} := q(1-q) = \mathbb{E}_{\mathrm{H}_0}\left[\mathrm{Y}_{ij}^2 \mid \mathrm{X}_{ij} = 0\right], \qquad \overline{p} := p(1-q)^2 + (1-p)q^2 = \mathbb{E}_{\mathrm{H}_0}\left[\mathrm{Y}_{ij}^2 \mid \mathrm{X}_{ij} = \lambda\right], \qquad \overline{r} := \frac{\overline{p}}{\overline{q}}. \tag{6.32}$$

(A4) The parameters are in a region:

$$\max\left\{\frac{k}{n}, \frac{\lambda k}{\sqrt{np}}, \frac{\lambda}{q}\right\} \leqslant D^{-8c_{\mathrm{si}}}, \tag{6.33}$$

for some large enough $c_{\mathrm{si}} > 0$.

Coming back to previous explanations, we can make an analogy for each of them. Assumption (A1) is a requirement on the permutation symmetry of the distribution (i.e. lemma 3.23). Moving to (A2), we are asking that the model satisfies a bound like that of lemma 5.27, for when we wanted to make the $\overline{P}$ basis of definition 5.11 close enough to the canonical basis of definition 2.18. In particular, we are adding some D factors and degrading from $\boldsymbol{M}$ in lemma 5.27 to $\boldsymbol{M}_{\mathrm{SM}} \subset \boldsymbol{M}$ only the perfect matches in (A2). Assumption (A3) morally says that the correlation between graphs under H_0 is like the symmetric difference, modulo the fact that we rescale the matrix to take values $\mathrm{Y}_{ij} \in \{-q, 1-q\}$ and adding some D factors in the powers involving the symmetric difference. The regime assumption in (A4) is also analogous to our assumption 3.1 once we take into account the rescaling.
By working out the proofs of section 5 it only takes mechanical time to realize that under assumption 6.29 a result similar to proposition 5.82 holds for the Gram matrix of correlations. Namely, we can use the same almost orthonormal basis (def. 1.14): that of definition 5.42 modulo an adjusted rescaling by $1/\overline{q}$ such that the condition of definition 1.14 holds.
The question is what kinds of models satisfy assumption 6.29#(A1) - (A2) - (A3). We provide two examples below.

GENERALIZED PLANTED SUB-MATRIX Under H_0 we sample a clique at random of size $\mathrm{Bin}(n, k/n)$ and choose to add a spike of magnitude λ in the sampled entries. Notice that by equation 6.28 we are however not in a centered case, which is $q = 1/2$, and we allow for different connection probabilities: within the clique it is p and outside it is q. For this model, the alternative hypothesis is that we sampled from a distribution where a fraction ϵ of the entries in the clique is removed, which morally means that $\zeta = -\epsilon k$. It is more natural to talk of a negative alteration rather than a positive alteration but the interpretation is analogous, as we want the size of ζ, i.e. $|\zeta|$ which is just expressed as a portion of the size of the signal with ϵ to be small for hardness and large for the existence of an algorithm.

STOCHASTIC BLOCK MODEL Under H_0 we sample a random partition of $[n]$ into k groups (a multinomial) and set $X_{ij} = \lambda$ if (i, j) are in the same group and zero otherwise. There is a block structure in the matrix, and we think of the perturbation in the alternative hypothesis as observing a matrix where we picked a group at random and removed it from such group with probability ϵ. Again we have a negative perturbation that is proportional to k.

We choose not to check explicitly assumption 6.29 for the generalized planted sub-matrix model and the stochastic block model. Rather, we mention that for these observations the basis of definition 5.42 is again almost orthonormal (def. 1.14) just like in the result of prop. 5.82. It then takes another computation similar to section 6 to show the analogue of theorem 3.8. In this case, the perturbation is ϵ, and we will assume we are in its hardness regime.

Assumption 6.34 (Generalized perturbation on the size of the signal)**.** *It holds that* $\epsilon\lambda k^2/n\sqrt{q} \leqslant D^{-8c_{\mathrm{si}}}$ *where* c_{si} *is the same large enough constant of assumption 6.29#(A4).*

As for the easier case, it is a matter of working out the final inequalities to obtain a weak detection hardness result in the spirit of definition 2.5 like for our main inequality: theorem 3.8.

Theorem 6.35 (Generalization of main theorem)**.** *For models and alterations as the generalized planted sub-matrix model or the stochastic block model, suppose that assumptions 6.29#(A4) and 6.34 hold. Then:*

$$\mathrm{Adv}_{(\leqslant D)}(H_0, H_1) \leqslant 1 + \frac{1}{D}, \qquad \textit{for all } D \geqslant 2. \tag{6.36}$$

FURTHER There are other permutation invariant models in the sense of assumption 6.29#(A1) that allow for these bounds. It is always a matter of double-checking that the interesting regimes match the regimes allowed by the proof computations. It is also interesting to allow for a fixed size of the signal. For example a slight modification on the model of equation 1.6 is that $|\mathrm{supp}(x)| \sim \mathrm{Geom}(k)$ instead of $x_i \overset{\text{i.i.d.}}{\sim} \mathrm{Ber}(k/n)$ which adds some correlations between entries: these can be dealt with using coupling techniques, which we do not outline here.
Believing enough in the power of this method, one could further explore what kinds of invariances and chains of assumptions allow going beyond permutation invariance (i.e. ass. 6.29#(A1)).

Alon, Noga, Michael Krivelevich, and Benny Sudakov (1998). "Finding a Large Hidden Clique in a Random Graph". In: *Random Structures & Algorithms* 13.3-4, pp. 457–466. ISSN: 1098-2418. DOI: 10.1002/(SICI)1098-2418(199810/12)13:3/4<457::AID-RSA14>3.0.CO;2-W (cit. on pp. 7, 14).

Arias-Castro, Ery and Nicolas Verzelen (2014). "Community Detection in Dense Random Networks". In: *The Annals of Statistics* 42.3, pp. 940–969. ISSN: 0090-5364. JSTOR: 43556311 (cit. on p. 5).

Arous, Gérard Ben, Alexander S. Wein, and Ilias Zadik (June 2020). *Free Energy Wells and Overlap Gap Property in Sparse PCA*. DOI: 10.48550/arXiv.2006.10689. arXiv: 2006.10689 [cs, math, stat] (cit. on p. 9).

Arpino, Gabriel and Ramji Venkataramanan (Mar. 2023). *Statistical-Computational Tradeoffs in Mixed Sparse Linear Regression*. arXiv: 2303.02118 [cs, math, stat] (cit. on p. 5).

Bandeira, Afonso S and Ahmed El Alaoui (2022). "The Franz–Parisi Criterion and Computational Trade-offs in High Dimensional Statistics". In: (cit. on p. 9).

Bandeira, Afonso S., Amelia Perry, and Alexander S. Wein (Apr. 2018). *Notes on Computational-to-Statistical Gaps: Predictions Using Statistical Physics*. arXiv: 1803.11132 [cs, stat] (cit. on pp. 5, 9).

Barak, Boaz, Samuel B. Hopkins, et al. (Apr. 2016). *A Nearly Tight Sum-of-Squares Lower Bound for the Planted Clique Problem*. DOI: 10.48550/arXiv.1604.03084. arXiv: 1604.03084 [cs] (cit. on p. 5).

Barak, Boaz and David Steurer (May 2014). *Sum-of-Squares Proofs and the Quest toward Optimal Algorithms*. DOI: 10.48550/arXiv.1404.5236. arXiv: 1404.5236 [cs] (cit. on p. 9).

Barbier, Damien (Aug. 2024). *How to Escape Atypical Regions in the Symmetric Binary Perceptron: A Journey through Connected-Solutions States*. DOI: 10.48550/arXiv.2408.04479. arXiv: 2408.04479 [cond-mat] (cit. on p. 9).

Barbier, Damien et al. (Apr. 2024). "On the Atypical Solutions of the Symmetric Binary Perceptron". In: *Journal of Physics A: Mathematical and Theoretical* 57.19, p. 195202. ISSN: 1751-8121. DOI: 10.1088/1751-8121/ad3a4a (cit. on p. 9).

Brennan, Matthew and Guy Bresler (2020). "Reducibility and Statistical-Computational Gaps from Secret Leakage". In: *Proceedings of Thirty Third Conference on Learning Theory*. PMLR, pp. 648–847 (cit. on pp. 7, 9, 14).

Buhai, Rares-Darius et al. (May 2025). *The Quasi-Polynomial Low-Degree Conjecture Is False*. DOI: 10.48550/arXiv.2505.17360. arXiv: 2505.17360 [cs] (cit. on p. 5).

Chen, Siyu et al. (June 2025). *An Optimized Franz-Parisi Criterion and Its Equivalence with SQ Lower Bounds*. DOI: 10.48550/arXiv.2506.06259. arXiv: 2506.06259 [math] (cit. on p. 9).

Clarke, Bertrand, Jennifer Clarke, and Chi Wai Yu (Feb. 2014). "Statistical Problem Classes and Their Links to Information Theory". In: *Econometric Reviews* 33.1-4, pp. 337–371. ISSN: 0747-4938. DOI: 10.1080/07474938.2013.807190 (cit. on p. 5).

Commenges, Daniel (Nov. 2015). *Information Theory and Statistics: An Overview*. DOI: 10.48550/arXiv.1511.00860. arXiv: 1511.00860 [math] (cit. on p. 5).

Ding, Yunzi et al. (Jan. 2023). "Subexponential-Time Algorithms for Sparse PCA". In: *Foundations of Computational Mathematics*. ISSN: 1615-3383. DOI: 10.1007/s10208-023-09603-0 (cit. on p. 5).

Duchi, John (2024). "Statistics and Information Theory - Lecture Notes" (cit. on p. 5).

Ebrahimi, Nader, Ehsan S. Soofi, and Refik Soyer (2010). "Information Measures in Perspective". In: *International Statistical Review* 78.3, pp. 383–412. ISSN: 1751-5823. DOI: 10.1111/j.1751-5823.2010.00105.x (cit. on p. 5).

Even, Bertrand, Christophe Giraud, and Nicolas Verzelen (June 2024). "Computation-Information Gap in High-Dimensional Clustering". In: *Proceedings of Thirty Seventh Conference on Learning Theory*. PMLR, pp. 1646–1712 (cit. on pp. 5, 9).

— (July 2025a). *Computational Barriers for Permutation-Based Problems, and Cumulants of Weakly Dependent Random Variables*. DOI: 10.48550/arXiv.2507.07946. arXiv: 2507.07946 [math] (cit. on pp. 5, 9).

— (June 2025b). *Computational Lower Bounds in Latent Models: Clustering, Sparse-Clustering, Biclustering*. DOI: 10.48550/arXiv.2506.13647. arXiv: 2506.13647 [math] (cit. on pp. 5, 9).

Feldman, Vitaly (Apr. 2017). *A General Characterization of the Statistical Query Complexity*. DOI: 10.48550/arXiv.1608.02198. arXiv: 1608.02198 [cs] (cit. on p. 9).

Gamarnik, David, Cristopher Moore, and Lenka Zdeborová (Nov. 2022). "Disordered Systems Insights on Computational Hardness". In: *Journal of Statistical Mechanics: Theory and Experiment* 2022.11, p. 114015. ISSN: 1742-5468. DOI: 10.1088/1742-5468/ac9cc8. arXiv: 2210.08312 [cond-mat, stat] (cit. on pp. 7, 9).

Gamarnik, David and Ilias Zadik (Dec. 2019). *The Landscape of the Planted Clique Problem: Dense Subgraphs and the Overlap Gap Property*. DOI: 10.48550/arXiv.1904.07174. arXiv: 1904.07174 [math] (cit. on p. 9).

Gamarnik, David and Ilias Zadik (Apr. 2022). "Sparse High-Dimensional Linear Regression. Estimating Squared Error and a Phase Transition". In: *The Annals of Statistics* 50.2, pp. 880–903. ISSN: 0090-5364, 2168-8966. DOI: 10.1214/21-AOS2130 (cit. on p. 9).

Hopkins, Samuel B. et al. (Oct. 2017). *The Power of Sum-of-Squares for Detecting Hidden Structures*. DOI: 10.48550/arXiv.1710.05017. arXiv: 1710.05017 [cs] (cit. on pp. 5, 7, 10, 14).

Horn, Roger A. and Charles R. Johnson (Oct. 2012). *Matrix Analysis*. 2nd ed. Cambridge University Press. ISBN: 978-0-521-83940-2 978-1-139-02041-1 978-0-521-54823-6. DOI: 10.1017/cbo9781139020411 (cit. on p. 19).

Huang, Brice and Mark Sellke (Jan. 2025). *Strong Low Degree Hardness for Stable Local Optima in Spin Glasses*. DOI: 10.48550/arXiv.2501.06427. arXiv: 2501.06427 [cond-mat] (cit. on p. 5).

Jerrum, Mark (1992). "Large Cliques Elude the Metropolis Process". In: *Random Structures and Algorithms* 3.4, pp. 347–359. ISSN: 10429832, 10982418. DOI: 10.1002/rsa.3240030402 (cit. on p. 9).

Kothari, Pravesh K. et al. (Mar. 2023). *Is Planted Coloring Easier than Planted Clique?* DOI: 10.48550/arXiv.2303.00252. arXiv: 2303.00252 [cs] (cit. on pp. 5, 7).

Kullback, Solomon (1978). *Information Theory and Statistics*. Reprint. Gloucester, Mass: Smith. ISBN: 978-0-8446-5625-0 (cit. on p. 5).

Kunisky, Dmitriy (2020). "Hypothesis Testing with Low-Degree Polynomials in the Morris Class of Exponential Families". In: (cit. on pp. 5, 9).

— (Mar. 2024a). *Low Coordinate Degree Algorithms I: Universality of Computational Thresholds for Hypothesis Testing*. DOI: 10.48550/arXiv.2403.07862. arXiv: 2403.07862 [math] (cit. on p. 5).

— (Dec. 2024b). *Low Coordinate Degree Algorithms II: Categorical Signals and Generalized Stochastic Block Models*. DOI: 10.48550/arXiv.2412.21155. arXiv: 2412.21155 [math] (cit. on p. 5).

Kunisky, Dmitriy, Cristopher Moore, and Alexander S. Wein (Apr. 2024). *Tensor Cumulants for Statistical Inference on Invariant Distributions*. DOI: 10.48550/arXiv.2404.18735. arXiv: 2404.18735 [math] (cit. on pp. 9, 62).

Kunisky, Dmitriy, Alexander S. Wein, and Afonso S. Bandeira (July 2019). *Notes on Computational Hardness of Hypothesis Testing: Predictions Using the Low-Degree Likelihood Ratio*. arXiv: 1907.11636 [cs, math, stat] (cit. on pp. 5, 9–12, 14, 15, 62, 66).

Lehmann, Erich L. (1970). *Testing Statistical Hypotheses*. 5. print. A Wiley Publication in Mathematical Statistics. New York u.a: Wiley. ISBN: 978-0-471-52470-0 (cit. on pp. 9, 62).

Lehmann, Erich L. and G. Casella (1998). *Theory of Point Estimation*. Springer Texts in Statistics. New York: Springer-Verlag. ISBN: 978-0-387-98502-2. DOI: 10.1007/b98854 (cit. on p. 9).

Mondelli, Marco and Andrea Montanari (July 2018). *Fundamental Limits of Weak Recovery with Applications to Phase Retrieval*. arXiv: 1708.05932 [cs, math, stat] (cit. on p. 5).

Montanari, Andrea and Alexander S. Wein (Dec. 2022). *Equivalence of Approximate Message Passing and Low-Degree Polynomials in Rank-One Matrix Estimation*. arXiv: 2212.06996 [math, stat] (cit. on pp. 9, 62).

O'Donnell, Ryan (June 2014). *Analysis of Boolean Functions*. 1st ed. Cambridge University Press. ISBN: 978-1-107-03832-5 978-1-139-81478-2 978-1-107-47154-2. DOI: 10.1017/cbo9781139814782 (cit. on p. 62).

Potters, Marc and Jean-Philippe Bouchaud (2020). *A First Course in Random Matrix Theory: For Physicists, Engineers and Data Scientists*. Cambridge: Cambridge University Press. ISBN: 978-1-108-48808-2. DOI: 10.1017/9781108768900 (cit. on p. 19).

R, Abhishek Hegade K. and Eren C. Kızıldağ (June 2025). *Large Average Subtensor Problem: Ground-State, Algorithms, and Algorithmic Barriers*. DOI: 10.48550/arXiv.2506.17118. arXiv: 2506.17118 [math] (cit. on p. 9).

Reyzin, Lev (May 2020). *Statistical Queries and Statistical Algorithms: Foundations and Applications*. DOI: 10.48550/arXiv.2004.00557. arXiv: 2004.00557 [cs] (cit. on p. 9).

Rush, Cynthia et al. (Dec. 2022). *Is It Easier to Count Communities than Find Them?* arXiv: 2212.10872 [cs, math, stat] (cit. on p. 5).

Schramm, Tselil and Alexander S. Wein (June 2022). "Computational Barriers to Estimation from Low-Degree Polynomials". In: *The Annals of Statistics* 50.3. ISSN: 0090-5364. DOI: 10.1214/22-AOS2179. arXiv: 2008.02269 [cs, math, stat] (cit. on pp. 5, 8, 9, 13, 14).

Semerjian, Guilhem (Oct. 2024). "Matrix Denoising: Bayes-Optimal Estimators Via Low-Degree Polynomials". In: *Journal of Statistical Physics* 191.10, p. 139. ISSN: 1572-9613. DOI: 10.1007/s10955-024-03359-9 (cit. on pp. 9, 62).

Serratosa, Francesc (Aug. 2021). "Redefining the Graph Edit Distance". In: *SN Computer Science* 2.6, p. 438. ISSN: 2661-8907. DOI: 10.1007/s42979-021-00792-5 (cit. on p. 24).

Sohn, Youngtak and Alexander S. Wein (Feb. 2025). *Sharp Phase Transitions in Estimation with Low-Degree Polynomials*. DOI: 10.48550/arXiv.2502.14407. arXiv: 2502.14407 [math] (cit. on pp. 5, 8, 9, 13, 14).

Steinhardt, Jacob (2016). "Memory, Communication, and Statistical Queries". In: (cit. on p. 9).

Szörényi, Balázs (2009). "Characterizing Statistical Query Learning: Simplified Notions and Proofs". In: *Lecture Notes in Computer Science*, pp. 186–200. ISSN: 0302-9743, 1611-3349. DOI: 10.1007/978-3-642-04414-4_18 (cit. on p. 9).

Verzelen, Nicolas and Ery Arias-Castro (Dec. 2015). "Community Detection in Sparse Random Networks". In: *The Annals of Applied Probability* 25.6, pp. 3465–3510. ISSN: 1050-5164, 2168-8737. DOI: 10.1214/14-AAP1080 (cit. on p. 5).

Wein, Alexander S. (Nov. 2020). *Optimal Low-Degree Hardness of Maximum Independent Set*. DOI: 10.48550/arXiv.2010.06563. arXiv: 2010.06563 [cs] (cit. on p. 5).

— (June 2025a). *Computational Complexity of Statistics: New Insights from Low-Degree Polynomials*. DOI: 10.48550/arXiv.2506.10748. arXiv: 2506.10748 [math] (cit. on pp. 5, 6, 9, 10, 14).

— (2025b). *Unifying Statistical Physics and Low-Degree Polynomials?* (Cit. on pp. 5, 9).

Zdeborová, Lenka and Florent Krzakala (Sept. 2016). "Statistical Physics of Inference: Thresholds and Algorithms". In: *Advances in Physics* 65.5, pp. 453–552. ISSN: 0001-8732, 1460-6976. DOI: 10.1080/00018732.2016.1211393. arXiv: 1511.02476 [cond-mat, stat] (cit. on pp. 5, 9).

In this paragraph we prove some lemmas from sections 2 - 3. The idea that the projection of an optimal separator to the class of low-degree polynomials preserves the same original properties comes from the fact that the space of polynomials is a linear subspace, and the orthogonal projection theorem applies (see the discussion in (Kunisky, Wein, and Afonso S. Bandeira 2019)). Some earlier works looked at distributions that enjoy invariance properties (Kunisky, Moore, and Wein 2024; Montanari and Wein 2022; Semerjian 2024). In particular, we could derive similar statements with more work by applying a formal version of the Hunt-Stein theorem from (Lehmann 1970) as in Montanari and Wein (2022, lem. 4.4).

Proof of lemma 2.23. Orthonormality is immediate from the fact that the inner product is the symmetric difference of graphs, i.e. equation 3.17 and $\lambda k = 0$. We are left to prove that it is a basis.
For a generic f, we have that there exist coefficients and powers such that:

$$f = a + \sum_r b_r \prod_{(i,j)\in E_r} Y_{ij}^{\beta_{ij}^{(r)}}, \tag{A.1}$$

where in particular $|E_r| \leqslant D$. Since $Y_{ij} \in \pm 1$, we can simply set $\beta_{ij}^{(r)} \equiv 1$. The representation:

$$f = a + \sum_r b_r \prod_{(i,j)\in E_r} Y_{ij}, \tag{A.2}$$

is written in terms of the set of monomials $P_{G,\pi}$ decorated with the unit function as follows. We set $\alpha_\varnothing = a$. Then, we take graphs $G = (V, E) \in \mathcal{G}_{\leqslant D}$ together with labellings $\pi \in \Pi_V$ such that for all r there exist a pair (G, π) satisfying $\prod_{(i,j)\in E_r} Y_{ij} = \prod_{i,j\in E} Y_{\pi(i),\pi(j)}$. Ultimately, we set $b_r = \alpha_{G,\pi}$. To ensure existence, it suffices to remark that there are no double edges or self-loops. For uniqueness, we reason by contradiction. Suppose there are two representations of a non-trivial polynomial f. Namely, there exist two distinct pairs of coefficients $(\alpha_\varnothing, (\alpha_{G,\pi})_{G,\pi})$ and $(\beta_\varnothing, (\beta_{G,\pi})_{G,\pi})$ with which we can express $f \not\equiv 0$. Then:

$$0 \equiv f - f = \alpha_\varnothing - \beta_\varnothing - \sum_{G,\pi} (\alpha_{G,\pi} - \beta_{G,\pi}) P_{G,\pi}. \tag{A.3}$$

The right-hand side is a polynomial and it is zero everywhere on the hypercube. We want to show it is then zero everywhere. The simplest way to see it is to recenter the hypercube to take values $\{0, 1\}^{n\times n}$, and observe that we can proceed by induction. Alternatively, there are standard references such as the book of O'Donnell (2014).
In passing from equation A.1 to equation A we gain that the underlying graph is simple (not a multi-graph), or equivalently that the edges appear in the polynomial only once in each summand. From this, we have a precise ordering of graphs if the number of edges that appear. Starting from the smallest ones, we cancel all coefficients by evaluating the null function $f - f$ at the smallest one-edge graphs. Going up, we can proceed to show that all two edge graphs have null coefficients by the same reasoning. By induction, the statement holds until we have considered all possible graphs.
Consequently, all the coefficients of $f - f$ are zero, reaching a contradiction as the representations of f had to be distinct, and the decomposition is unique. □

Proof of lemma 3.16. Orthogonality is lost since the symmetric difference is non-zero. The collection remains a basis since the reasoning of the proof of lemma 2.23 does not change. □

Proof of lemma 3.23. Let $f^\star$ be a polynomial of degree less than D attaining the maximal advantage. By invariance of the probability distributions $\mathrm{H}_0, \mathrm{H}_1$ any permutation of $f^\star$ defined as the function $f^\star(\boldsymbol{PY})$ for $\boldsymbol{P}$ a permutation matrix is optimal. The "symmetrization" of the optimal function $f^\star_{\mathrm{inv}}(\boldsymbol{Y}) = 1/n! \sum_{\boldsymbol{P} \text{ perm. mat}} f^\star(\boldsymbol{PY})$ is still a polynomial of degree less than D, which is invariant to permutations. Consider:

$$\frac{\mathbb{E}_{\mathrm{H}_1}\left[\frac{1}{n!}\sum_{\boldsymbol{P} \text{ perm. mat}} f^\star(\boldsymbol{PY})\right]}{\sqrt{\mathbb{E}_{\mathrm{H}_0}\left[\left(\frac{1}{n!}\sum_{\boldsymbol{P} \text{ perm. mat}} f^\star(\boldsymbol{PY})\right)^2\right]}}. \tag{A.4}$$

By invariance of H_1 with respect to permutations, the numerator is equal to the numerator in the optimal value of $\mathrm{Adv}_{(\leqslant D)}(H_0, H_1)$ evaluated at $f^\star$. For the denominator, by convexity of the square function and invariance with respect to permutations (in this case of the second moment of $f^\star$):

$$\mathbb{E}_{H_0}\left[\left(\frac{1}{n!}\sum_{\boldsymbol{P}\text{ perm. mat}} f^\star(\boldsymbol{P}\mathbf{Y})\right)^2\right] \leqslant \mathbb{E}_{H_0}\left[(f^\star)^2\right]. \tag{A.5}$$

The argument of the advantage is larger for an invariant function. Therefore, the optimizer in the advantage is attained by an invariant function. □

Proof of lemma 3.25. Consider the space of polynomials taking values in $\{-1,1\}^{n\times n}$ of degree less than D. The monomials $P_{G,\pi}$ and the unit function are a basis of this space by lemma 3.16. Grouping the monomials by permutations, i.e. summing over $\sum_{\pi\in\Pi_{|V|}}$ they form a set of polynomials invariant by permutations. We construct the candidate basis set with these functions decorated with the unit function, and want to show the set spans invariant polynomials of degree less than D. For f a generic permutation invariant polynomial of degree less than D, we have in particular that it is decomposable in the $P_{G,\pi}$ basis, since it is a polynomial. Mathematically, there exist coefficients such that:

$$f = \alpha_\varnothing + \sum_{G\in\mathcal{G}_{\leqslant D},\pi} \alpha_{G,\pi} P_{G,\pi}, \tag{A.6}$$

where $\alpha_\varnothing$ corresponds to the constant function. In particular, for any $G \in \mathcal{G}_{\leqslant D}$ labelled in two different ways with $\pi^{(1)} \neq \pi^{(2)} \in \Pi_{|V|}$ there exists a permutation matrix $\boldsymbol{P}$ such that:

$$\alpha_{G,\pi^{(1)}} = \mathbb{E}_{H_0}\left[f(\mathbf{Y})P_{G,\pi^{(1)}}(\mathbf{Y})\right] = \mathbb{E}_{H_0}\left[f(\boldsymbol{P}\mathbf{Y})P_{G,\pi^{(1)}}\right] = \mathbb{E}_{H_0}\left[f(\mathbf{Y})P_{G,\pi^{(2)}}\right] = \alpha_{G,\pi^{(2)}}. \tag{A.7}$$

Since the coefficients are the same across $\pi \in \Pi_{|V|}$, we collect them and write f as a linear combination of elements of the candidate basis:

$$f = \alpha_\varnothing + \sum_{G\in\mathcal{G}_{\leqslant D}}\sum_{\pi\in\Pi_{|V|}} \underbrace{\alpha_{G,\pi}}_{\alpha_G} P_{G,\pi} = \alpha_\varnothing + \sum_{G\in\mathcal{G}_{\leqslant D}} \alpha_G \sum_{\pi\in\Pi_{|V|}} P_{G,\pi} = \alpha_\varnothing + \sum_{G\in\mathcal{G}_{\leqslant D}} \alpha_G P_G. \tag{A.8}$$

Having a decomposition, we can show uniqueness in the same way as in the proof of lemma 3.16. Therefore, $\left(1, (P_G)_{G\in\mathcal{G}_{\leqslant D}}\right)$ is a basis for invariant polynomials of degree less than D. □

B MATCHING BOUNDS

In this section we put elements of the proof of propositions 3.12 and a complete argument for proposition 3.13. Namely, we discuss what happens when we relax either of the conditions from assumption 3.1 or assumption 3.7. The idea is that:

- if $k \neq o(n)$ then the signal is approximately of the size of full observation matrix; we reach an information-theoretic threshold easily and there is no gap;
- if $\lambda k/\sqrt{n} \neq o(1)$ then a line-sum statistic performs detection (with a caveat for perturbations of λ);
- if $\eta \neq o(n/k^2)$ or $\zeta \neq o\left(\sqrt{\frac{n}{\lambda}}\right)$ then a global sum statistic performs detection.

To make matters clear, we repeat that we do not comment on the condition on λ because it is just to scale everything correctly.

With results of this kind, we are certain that our main theorem 3.8 is tight up to poly-logarithmic factors. Let us explain thoroughly the two sides of this argument. On the negative, under assumptions 3.1 - 3.7 specialized at $D \approx_{\log} \log n$ our result says that no polynomial of degree D can perform weak detection in the sense of definition 2.5. On the positive side, again up to poly-logarithmic factors, as soon as we violate either of the conditions we show there are two options. Either there is no gap at all when we hit the information-theoretic bound, see appendix C; or there is a working algorithm. Therefore, the low-degree conjecture of section 2 captures the expected behavior of known algorithms (see subsec. B.III for further comments).

In the two next subsections, we analyze the algorithms with concentration arguments using $O_{\mathbb{P}}(\cdot)$ to denote a term that is bounded with high probability. In the third subsection, we present the formal derivation for the global sum statistic, i.e. the proof of proposition 3.13.

Remark B.1. *As $n \to \infty$ poly-logarithmic factors are negligible. The claim "up to poly-logarithmic factors" is thus rather rough for sharp, low-dimensional claims. It is somewhat standard in the low-degree method.*

B.1 *Global sum statistic*

The global sum statistic from equation 3.11:

$$s_{\text{global}}(\mathbf{Y}) = \sum_{i \neq j} Y_{ij}, \tag{B.2}$$

is able to distinguish (H_0, H_1) under mild conditions on the parameters. The intuition is as follows. With high probability, $\mathbf{X} = (x_i x_j)_{i \neq j}$ concentrates since it is a matrix of Bernoulli random variables. Indeed, we have $X_{ij} \sim \lambda \text{Ber}(k^2/n^2)$ and in particular $x_i \overset{\text{i.i.d.}}{\sim} \sqrt{\lambda}\text{Ber}(k/n)$. With high probability, the support of the vector x is of size:

$$\sum_{i=1}^{n} \frac{1}{\sqrt{\lambda}} x_i = |\text{supp}(x)| = k \pm O_{\mathbb{P}}\left(\sqrt{k}\right), \tag{B.3}$$

so that the size of the portion of the matrix with signal is with high probability within $k^2 \pm O_{\mathbb{P}}\left(k^{3/2}\right)$. For large enough k, the correction should not matter. Then, the random variable $s_{\text{global}}(\mathbf{Y})$ has the following distribution for the first perturbation problem:

$$s_{\text{global}}(\mathbf{Y}) = \begin{cases} \sum_{\substack{i \neq j \\ X_{ij}=1}} \text{Rad}\left(1+\lambda/2\right) + \sum_{\substack{i \neq j \\ X_{ij}=0}} \text{Rad}(1/2) & \text{under } H_0 \\ \sum_{\substack{i \neq j \\ X_{ij}=1}} \text{Rad}\left(1+\lambda+\eta/2\right) + \sum_{\substack{i \neq j \\ X_{ij}=0}} \text{Rad}(1/2) & \text{under } H_1. \end{cases} \tag{B.4}$$

Again, by concentration there are $k^2 \pm O_{\mathbb{P}}\left(k^{3/2}\right)$ entries in the first sum, and $n^2 - k^2 \mp O_{\mathbb{P}}\left(k^{3/2}\right)$ entries in the second. Using concentration:

$$\sum_{\substack{i \neq j \\ X_{ij}=0}} \text{Rad}(1/2) = O_{\mathbb{P}}(n) \text{ under both } H_0, H_1, \tag{B.5}$$

since we assume that $k = o(n)$ (otherwise there is an information-theoretic bound). The difference is in the two sums with the signal. Under the null and the alternative the sum is of roughly $k^2 \pm O_{\mathbb{P}}\left(k^{3/2}\right)$ order and the Y_{ij} are independent once the entries X_{ij} are fixed. Using these facts:

$$\sum_{\substack{i \neq j \\ X_{ij}=1}} \text{Rad}\left(1+\lambda/2\right) = \begin{cases} k^2\lambda + O_{\mathbb{P}}(k) & \text{under } H_0 \\ k^2(\lambda+\eta) + O_{\mathbb{P}}(k) & \text{under } H_1 \end{cases}. \tag{B.6}$$

If we roughly impose that the distributions are separated we wish that:

$$k^2(\lambda+\eta) - k - n \gtrsim_{\log} k^2\lambda + k + n \tag{B.7}$$

which after simplifications up to log factors means that the s_{global} statistic works when $k^2\eta \gtrsim_{\log} n$ which coincides with the opposite condition of assumption 3.7.

Remark B.8. *The same reasoning for a perturbation ζ on the size of the signal k as in problem 1.3 gives the feasibility condition $\zeta\sqrt{\lambda}/\sqrt{n} \gtrsim_{\log} 1$ which matches assumption 3.7.*

B.2 *Line-sum*

Alternatively, we could check if there are "many" lines that are large. This corresponds to thresholding the line-sum statistic from equation 3.11

$$s_{\text{line}}(\boldsymbol{Y}) = \#\left\{ j : \sum_{i \neq j} \boldsymbol{Y}_{ij} \geqslant \omega \right\}, \tag{B.9}$$

for some well-chosen threshold ω. The same reasoning holds on concentration of $\sum_{i=1}^{n} x_i/\sqrt{\lambda} = k \pm O_{\mathbb{P}}\left(\sqrt{k}\right)$, so defining the clique set $\mathscr{C} = \{j : x_j = 1\}$ it holds that:

$$\begin{aligned}\sum_{i\neq j} Y_{ij} &= \begin{cases} \sum_{i\neq j} \mathrm{Rad}(1/2) & \text{i.i.d. if } j \notin \mathscr{C} \text{ under both} \\ \sum_{\substack{i\neq j \\ i\in\mathscr{C}}} \mathrm{Rad}(1+\lambda/2) + \sum_{\substack{i\neq j \\ i\notin\mathscr{C}}} \mathrm{Rad}(1/2) & \text{i.i.d. if } j \in \mathscr{C} \text{ under } H_0 \\ \sum_{\substack{i\neq j \\ i\in\mathscr{C}}} \mathrm{Rad}(1+\lambda+\eta/2) + \sum_{\substack{i\neq j \\ i\notin\mathscr{C}}} \mathrm{Rad}(1/2) & \text{i.i.d. if } j \in \mathscr{C} \text{ under } H_1 \end{cases} \\ &= \begin{cases} O_{\mathbb{P}}\left(\sqrt{n}\right) & \text{if } j \notin \mathscr{C} \text{ under both} \\ \lambda k + O_{\mathbb{P}}\left(\sqrt{k}\right) + O_{\mathbb{P}}\left(\sqrt{n}\right) & \text{if } j \in \mathscr{C} \text{ under } H_0 \\ (\lambda+\eta)k + O_{\mathbb{P}}\left(\sqrt{k}\right) + O_{\mathbb{P}}\left(\sqrt{n}\right) & \text{if } j \in \mathscr{C} \text{ under } H_1 \end{cases}. \end{aligned} \tag{B.10}$$

First, for the entries to be at all visible we need up to log factors $\lambda k \gtrsim_{\log} \sqrt{n}$. Secondly, we impose that there is a clear distinction between the sums over the cliques under H_0 and H_1. Mathematically, we require:

$$(\lambda+\eta)k - \sqrt{k} - \sqrt{n} \gtrsim_{\log} \lambda k + \sqrt{k} + \sqrt{n}. \tag{B.11}$$

Reordering and using again leading orders we obtain the condition $k\eta \gtrsim_{\log} \sqrt{n}$. Notice that the condition is $\eta \gtrsim_{\log} \sqrt{n}/k$ and by $k \gtrsim_{\log} \sqrt{n}$ (recall remark 3.6) we are not in the regime of assumption 3.7 as $\sqrt{n}/k \gtrsim_{\log} n/k^2$. This holds true unless we have no detection-recovery gap: in the case of a line-sum statistic we get close to a matching algorithm for the hardness result of theorem 3.8 when λ is perturbed.
To fix this little inconsistency, we just need to realize that once the condition $\lambda k \gtrsim_{\log} \sqrt{n}$ holds then we can *estimate* the clique with high probability. Consequently, since we expect complex testing (prob. 1.3) to be an easier problem than estimation (prob. 1.5), complex testing should be easy in this regime in the sense of definition C.8. In practical terms, we can identify k columns in $\mathbf{Y}$ where the line sum is well above the value of a line sum without signal. These are columns $j \in \mathscr{C}$ above. Taking the sub-matrix $(Y_{ij})_{i,j\in\mathscr{C}} \in \{-1,1\}^{k\times k}$ we aim to estimate the signal strength and see if it is closer to λ or to $\lambda+\eta$. Since for given $j \in \mathscr{C}$ the $O_{\mathbb{P}}\left(\sqrt{n}\right)$ oscillations came from entries $i \notin \mathscr{C}$ that did not have signal we know that:

$$\sum_{\substack{i\neq j \\ i,j\in\mathscr{C}}} Y_{ij} = \begin{cases} \lambda k + O_{\mathbb{P}}\left(\sqrt{k}\right) & \text{under } H_0 \\ (\lambda+\eta)k + O_{\mathbb{P}}\left(\sqrt{k}\right) & \text{under } H_1 \end{cases}, \qquad \text{i.i.d. for all } j \in \mathscr{C}. \tag{B.12}$$

So if we use the signal estimator

$$\widehat{\lambda}(\mathbf{Y}) := \frac{1}{k^2} \sum_{j\in\mathscr{C}} \sum_{\substack{i\neq j \\ i\in\mathscr{C}}} Y_{ij}, \tag{B.13}$$

then:

$$\widehat{\lambda}(\mathbf{Y}) = \begin{cases} \lambda + O_{\mathbb{P}}\left(\frac{1}{\sqrt{k}}\right) & \text{under } H_0 \\ \lambda + \eta + O_{\mathbb{P}}\left(\frac{1}{\sqrt{k}}\right) & \text{under } H_1 \end{cases}. \tag{B.14}$$

To separate the distributions we need $\eta \gtrsim_{\log} 1/\sqrt{k}$. The intersection of the perturbation assumption 3.7 and $\eta \gtrsim_{\log} 1/\sqrt{k}$ is a non-empty interval when $n \gtrsim_{\log} k^{3/2}$. Since we keep the condition that $k/n = o(1)$ from assumption 3.1, combining these two observations complex testing (prob. 1.3) is solved with high probability by a line sum algorithm as soon as we break the inequality $\lambda k/\sqrt{n} = o(1)$ for k of small enough order with respect to n.

Remark B.15. *Contrarily, if we have $\lambda k/\sqrt{n} \gtrsim_{\log} 1$ for a perturbation on k with magnitude ζ the heuristic suggests that line-sum distinguishes when $k + \sqrt{k} \lesssim_{\log} k + \zeta - \sqrt{k+\zeta}$ which holds up to log factors when $\zeta \gtrsim_{\log} \sqrt{k}$. Noticing that $\sqrt{k} \lesssim_{\log} \sqrt{n/\lambda}$ the region is included in assumption 3.7. We directly have a working algorithm.*

In the next subsection we discuss the weaknesses and arguments in favour of these conclusions.

B.III *Comments on optimality of global sum and line-sum statistics*

Our main theorem (thm. 3.8) is a negative result and as soon as we violate any of its assumptions we know there exists an algorithm.
The weakness of this argument is that while there exist an algorithm, nothing tells us this should be optimal. Formally, the Neyman-Pearson lemma ensures that if we allow for unconstrained computation time, regressing to an information-theoretic setting as in appendix C, then we know the optimal test is a well-chosen threshold on the likelihood ratio (Kunisky, Wein, and Afonso S. Bandeira 2019). However, once we restrict to polynomial time algorithms there is no guarantee that the optimal algorithm is the sum or the line-sum statistic of equations B.2 - B.9. There might as well be an algorithm that under both assumptions 3.1 - 3.7 is able to perform weak detection (def. 2.5) for complex testing, which is problem 1.3. While this is a possibility, we present three key arguments to argue that it is not expected.

(A1) The sum and line-sum statistics (eqns. B.2 - B.9) are optimal among invariant statistics, where by invariance we mean invariance in the sense of lemma 3.23. Since we know the objective function in the advantage (def. 1.10) is attained by an invariant function, we know that no function can do better than our candidates in terms of weak separation (def. 2.12).

(A2) While one could further argue that weak separation (def. 2.12) is a *sufficient* condition for weak detection (def. 2.5), and not a characterization, the wide belief is that once we restrict to degree D polynomials these criterions become rather equivalent (see the discussion in (Kunisky, Wein, and Afonso S. Bandeira 2019), especially regarding hypercontractivity).

(A3) Lastly, the low-degree method is a heuristic, but it appears to catch the expected algorithmic thresholds for a wide collection of problems (we mention some in subsection 1.II). Modulo some adjustments it is well-established as a method for those that believe in it. At the very worst, it rules out thresholding polynomials up to some degree. This is already a large class of test functions that includes spectral methods (see (Kunisky, Wein, and Afonso S. Bandeira 2019)).

We now move to an exemplified proof via concentration inequalities for the global sum statistic of equation B.2. This is the proof of proposition 3.13.

B.IV *Formal analysis of global sum statistic under signal strength perturbations*

Concentration arguments as in the previous two subsections are clean but avoid the many details. Here we add these details for the $\mathsf{s}_{\text{global}}$ statistic/algorithm (eqn. B.2) when we test $\boldsymbol{\theta} = (k, \lambda)$ against $\boldsymbol{\theta}' = (k, \lambda + \eta)$ in the sense of problem 1.3. At the cost of similar arguments all the other combinations of perturbations and thresholding statistics are analogous.

Proof of proposition 3.13. We want to bound both the probability of a type I error and of a type II error. This means that we want to upper bound:

$$\mathbb{P}_{\mathrm{H}_0}\left[\mathsf{s}_{\text{global}}(\mathbf{Y}) - \mu_0 \geqslant \xi\right], \qquad \text{and} \qquad \mathbb{P}_{\mathrm{H}_1}\left[\mathsf{s}_{\text{global}}(\mathbf{Y}) - \mu_0 < \xi\right], \tag{B.16}$$

for some well-chosen threshold $\xi \in \left(0, {}^{n(n-1)}\!/_{n^2}k^2\eta\right)$, where $\mu_0 := {}^{n(n-1)}\!/_{n^2}k^2\lambda$ is the expectation of $\mathsf{s}_{\text{global}}$ under H_0. Similarly, we define $\mu_1 := {}^{n(n-1)}\!/_{n^2}k^2(\lambda + \eta)$, and $\phi := \mu_1 - \mu_0 > 0$. By the equation above we have two events to consider respectively under H_0 and H_1:

$$A_0 \equiv A_0^{(\xi)} := \left\{\mathsf{s}_{\text{global}}(\mathbf{Y}) - \mu_0 \geqslant \xi\right\}, \qquad A_1 \equiv A_1^{(\xi)} := A_0^{\mathrm{c}} = \left\{\mathsf{s}_{\text{global}}(\mathbf{Y}) - \mu_0 < \xi\right\}. \tag{B.17}$$

While these are largely complex to analyze, we can condition on high probability events that simplify their evaluation. Inspired by the discussion of the previous subsection we define:

$$B \equiv B^{(\delta)} := \{||\mathrm{supp}(\mathbf{x})| - k| \leqslant \delta k\}, \qquad C \equiv C^{(\chi)} := \left\{\left|\sum_{\substack{i\neq j\\ \mathrm{x}_i\mathrm{x}_j=0}} \mathrm{Y}_{ij}\right| \leqslant \chi\right\}, \tag{B.18}$$

where $\delta \in (0,1), \chi \in (0, {}^{n(n-1)}\!/_2)$ are parameters we will choose together with ξ. The following fact allows us to decompose the probabilities.

Fact B.19. *Let $\mathbb{P}$ be a probability measure and A, B, C be three measurable events. Then:*

$$\mathbb{P}[A] \leqslant \mathbb{P}[A \mid B, C] + \mathbb{P}[B^{\mathrm{c}}] + \mathbb{P}[C^{\mathrm{c}} \mid B]. \tag{B.20}$$

Proof. We just decompose the probabilities over and over. Mathematically, we have:

$$\mathbb{P}[A] = \mathbb{P}[A \mid B]\,\mathbb{P}[B] + \mathbb{P}[A \mid B^{\mathrm{c}}]\,\mathbb{P}[B^{\mathrm{c}}] \leqslant \mathbb{P}[A \mid B] + \mathbb{P}[B^{\mathrm{c}}], \tag{B.21}$$

and we bound further in this style to make C pop-up. We have:

$$\mathbb{P}[A \mid B] = \mathbb{P}[A \mid B, C]\,\mathbb{P}[C \mid B] + \mathbb{P}[A \mid B, C^{\mathrm{c}}]\,\mathbb{P}[C^{\mathrm{c}} \mid B] \leqslant \mathbb{P}[A \mid B, C] + \mathbb{P}[C^{\mathrm{c}} \mid B]. \tag{B.22}$$

The claim follows. □

It remains to show that this whole sum is small using concentration of measure results.
The easiest term is the unconditional probability. We have that $|\mathrm{supp}(\mathbf{x})| \sim \mathrm{Bin}(n, k/n)$ so that:

$$\mathbb{P}_{\mathrm{H}_0}[B^{\mathrm{c}}] = \mathbb{P}[|\mathrm{Bin}(n, k/n) - k| \geqslant \delta k] \leqslant 2\exp\left\{-\frac{\delta^2 k}{3}\right\}, \tag{B.23}$$

by concentration of the binomial distribution. We can make it small upon choosing well δ.

For the conditional probabilities, we need to be more careful. The key observation is that both events can lose the dependence structure if we take the worst-case. Let us be clearer with an explicit example. Consider $\mathbb{P}_{\mathrm{H}_0}[C^{\mathrm{c}} \mid B]$. For each fixed realization of the support $\mathbf{x}$ the random variables Y_{ij} such that $x_i x_j = 0$ are independent Rademacher distributed random variables. The number of these pairs depends on the size of the support, but we always have a Gaussian concentration. Assuming the size of the support is s we have that:

$$\mathbb{P}[C^{\mathrm{c}} \mid |\mathrm{supp}(\mathbf{x})| = s] = \mathbb{P}\left[\left|\sum_{r=1}^{\bar{s}} \mathrm{Rad}(1/2)\right| \geqslant \frac{\chi}{2}\right] \leqslant 2\exp\left\{-\frac{\chi^2}{8\bar{s}}\right\}, \qquad \bar{s} := \frac{n(n-1)}{2} - \frac{s(s-1)}{2}, \tag{B.24}$$

where we notice that we divided by 2 since we were summing over $i \neq j$ but the matrix is symmetric. Therefore, conditional on the event B, we have an overall bound by the worst size possible, which is in this case the maximal size since:

$$\sup_{s \in [k(1-\delta), k(1+\delta)]} \exp\left\{-\frac{\chi^2}{8\bar{s}^2}\right\} = \exp\left\{-\frac{\chi^2}{2(n(n-1) - k(1-\delta)(k(1-\delta)-1))}\right\} \leqslant \exp\left\{-\frac{\chi^2}{2n^2}\right\}, \tag{B.25}$$

since $k/n \leqslant D^{-8c_{\mathrm{si}}}$ and $D \geqslant 2$ under assumption 3.1. Notice that here we only apply the information-theoretic part of the conditions (see discussion at the beginning of this section and remark 3.5).
For the other two conditional probabilities, the reasoning is analogous, but we need to make two different bounds.

BOUND ON NULL HYPOTHESIS TERM The following event inclusion is useful to simplify the expression:

$$\begin{aligned}\{A_0 \mid B, C\} &= \left\{ \sum_{\substack{i \neq j \\ x_i x_j = 1}} Y_{ij} - \mu_0 \geqslant \xi - \sum_{\substack{i \neq j \\ x_i x_j = 0}} Y_{ij} \mid B, C \right\} \\ &\subseteq \left\{ \sum_{\substack{i \neq j \\ x_i x_j = 1}} Y_{ij} - \mu_0 \geqslant \xi - \chi \mid B, C \right\},\end{aligned} \tag{B.26}$$

where we used in particular that event C holds. For fixed $\mathbf{x}$, the sum on the last RHS is a sum of independent Rademacher distributed random variables with mean $1+\lambda/2$ and unit variance. If the size of the support is s we find again:

$$\mathbb{P}\left[\sum_{r=1}^{s(s-1)/2} \mathrm{Rad}(1+\lambda/2) - \frac{\mu_0}{2} \geqslant \frac{\xi - \chi}{2}\right] \leqslant \exp\left\{-\frac{(\xi-\chi)^2}{2(s(s-1))}\right\}, \qquad \forall \chi < \xi, \tag{B.27}$$

where we stress that we will have to take into account the condition $\chi < \xi$. Taking the worst-case value among supports allowed by event B we bound as before under assumption 3.1:

$$\sup_{s\in[k(1-\delta),k(1+\delta)]} \exp\left\{-\frac{(\xi-\chi)^2}{2(s(s-1))}\right\} \leqslant \exp\left\{-\frac{(\xi-\chi)^2}{2n^2}\right\}. \tag{B.28}$$

Therefore, in the worst-case we have a bound of this type on $\mathbb{P}_{H_0}[A_0 \mid B, C]$.

BOUND ON ALTERNATIVE HYPOTHESIS TERM We seek something similar, now using that under the H_1 distribution the mean is not μ_0 but rather $\mu_1 = \mu_0 + \phi$, so we need to recenter concentration terms. Let us establish the useful inclusion:

$$\begin{aligned}
\{A_1 \mid B, C\} &= \left\{ \sum_{\substack{i\neq j \\ x_i x_j = 1}} Y_{ij} - \mu_0 - \phi \leqslant \xi - \sum_{\substack{i\neq j \\ x_i x_j = 0}} Y_{ij} - \phi \mid B, C \right\} \\
&\subseteq \left\{ \sum_{\substack{i\neq j \\ x_i x_j = 1}} Y_{ij} - \mu_0 - \phi \leqslant \xi + \chi - \phi \mid B, C \right\} \\
&= \left\{ \sum_{\substack{i\neq j \\ x_i x_j = 1}} Y_{ij} - \mu_1 \leqslant \xi + \chi - \phi \mid B, C \right\}.
\end{aligned} \tag{B.29}$$

If we fix $\mathbf{x}$ in the last event, and take it to have a support allowed by the event B of size s, we need to evaluate the tail probability of a sum of i.i.d. centered Rademacher random variables. By Gaussian concentration, under the necessary condition that $\xi + \chi - \phi < 0$ we find:

$$\mathbb{P}\left[\sum_{r=1}^{s(s-1)/2} \mathrm{Rad}(1+\lambda+\eta/2) - \frac{\mu_1}{2} \leqslant \frac{\xi+\chi-\phi}{2}\right] \leqslant \exp\left\{-\frac{(\xi+\chi-\phi)^2}{2(s(s-1))}\right\}. \tag{B.30}$$

One last time, the worst-case upper bound inside the event B is by the conditions that we keep in assumption 3.1:

$$\sup_{s\in[k(1-\delta),k(1+\delta)]} \exp\left\{-\frac{(\xi+\chi-\phi)^2}{2(s(s-1))}\right\} \leqslant \exp\left\{-\frac{(\xi+\chi-\phi)^2}{2n^2}\right\}. \tag{B.31}$$

PUTTING IT ALL TOGETHER We aim to threshold properly s_{global} and show it attains a small type I and type II error. The variables in the following bound are the threshold ξ and the conditioning events thresholds (δ, χ). We start by combining equations B.23 - B.25 - B.28 - B.31 and fact B.19 for a valid triplet (ξ, χ, δ) such that $\chi < \xi, \xi + \chi - \phi < 0$. In equations, the sum of type I and type II error admits the bound:

$$\begin{aligned}
\mathbb{P}_{H_0}\left[s_{\text{global}}(\mathbf{Y}) - \mu_0 \geqslant \xi\right] + \mathbb{P}_{H_1}\left[s_{\text{global}}(\mathbf{Y}) - \mu_0 < \xi\right] &= \mathbb{P}_{H_0}[A_0] + \mathbb{P}_{H_1}[A_1] \\
&\leqslant \mathbb{P}_{H_0}[A_0 \mid B, C] + \mathbb{P}_{H_1}[A_1 \mid B, C] + 2\mathbb{P}_{H_0}[B^c \mid C] + 2\mathbb{P}_{H_0}[C^c] \\
&\leqslant \exp\left\{-\frac{(\xi-\chi)^2}{2n^2}\right\} + \exp\left\{-\frac{(\xi+\chi-\phi)^2}{2n^2}\right\} \\
&\quad + 2\exp\left\{-\frac{\chi^2}{2n^2}\right\} + 4\exp\left\{-\frac{\delta^2 k}{3}\right\},
\end{aligned} \tag{B.32}$$

where we used that the events B^c and $C^c \mid B$ do not change whether we integrate under the distribution of H_0 or of H_1. The remaining part of the proof is just an analysis argument to make the RHS smaller than $1 - \Omega(1)$ when the condition on the perturbation of assumption 3.7 is violated, so when $\eta^{k^2}/n \gtrsim_{\log} 1$ approximately.

Let us bound each term separately and then combine them. We seek to make them all less than $p/4$ for some fixed $p \in (0, 1)$ that depends on (n, k, η). The choices we make reflect the heuristic of subsection B.I, and

happen to be the minimum required for weak separation (def. 2.12). Taking $\delta = c_\delta/\sqrt{k}$ for some $c_\delta > 0$ we have that:

$$4\exp\left\{-\frac{\delta^2 k}{3}\right\} = 4\exp\left\{-\frac{c_\delta^2}{3}\right\} \leqslant \frac{p}{4} \iff c_\delta \geqslant \sqrt{-3\ln\left(\frac{p}{16}\right)}. \tag{B.33}$$

Letting $\chi = c_\chi n$ for some $c_\chi > 0$ we have that:

$$2\exp\left\{-\frac{\chi^2}{n^2}\right\} = 2\exp\left\{-\frac{c_\chi^2}{2}\right\} \leqslant \frac{p}{4} \iff c_\chi \geqslant \sqrt{-2\ln\left(\frac{p}{8}\right)}. \tag{B.34}$$

These are the decoupled terms. Concerning the coupled terms, we see that there needs to be a non-trivial interval for ξ to exist. Combining the conditions to have our concentration inequalities such interval is:

$$\chi < \xi < \phi - \chi = \frac{n(n-1)}{n^2}k^2\eta - \chi, \tag{B.35}$$

which is a non-empty interval if $\phi > 2\chi = 2c_\chi n$. Once we impose that the interval exists, we have that $\xi - \chi > (c_\xi - c_\chi)n > 0$ for some $c_\xi > c_\chi$ thanks to which:

$$\exp\left\{-\frac{(\xi-\chi)^2}{2n^2}\right\} \leqslant \exp\left\{-\frac{(c_\xi - c_\chi)^2}{2}\right\} \leqslant \frac{p}{4} \iff c_\xi - c_\chi \geqslant \sqrt{-2\ln\left(\frac{p}{4}\right)}. \tag{B.36}$$

Using the same reasoning, we can find a $\phi/n > c_\phi > c_\xi$ such that $\phi - \xi - \chi > (c_\phi - c_\xi - c_\chi)n$ implying:

$$\exp\left\{-\frac{(\phi-\xi-\chi)^2}{2n^2}\right\} \leqslant \exp\left\{-\frac{(c_\phi - c_\xi - c_\chi)^2}{2}\right\} \leqslant \frac{p}{4} \iff c_\phi - c_\xi - c_\chi \geqslant \sqrt{-2\ln\left(\frac{p}{4}\right)}. \tag{B.37}$$

Reordering equations B.33 - B.34 - B.36 - B.37, together with the fact that $\phi > 2\chi$ and $\phi > nc_\phi$ we see that we want to find a $p \in (0,1)$ for given (n,k,η) such that the following intervals are non-empty:

$$\begin{aligned}
\sqrt{-3\ln\left(\frac{p}{16}\right)} &\leqslant c_\delta < \infty \\
\sqrt{-2\ln\left(\frac{p}{8}\right)} &\leqslant c_\chi \\
c_\chi + \sqrt{-2\ln\left(\frac{p}{4}\right)} &\leqslant c_\xi \leqslant \frac{\phi}{n} + c_\chi \\
c_\chi + c_\xi + \sqrt{-2\ln\left(\frac{p}{4}\right)} &\leqslant c_\phi \leqslant \frac{\phi}{n} + c_\chi + c_\xi.
\end{aligned} \tag{B.38}$$

The first is trivially satisfied for c_δ large enough. The second includes c_χ which appears in the condition $\phi > 2\chi = 2c_\chi n$. The last two intervals exist as soon as $\phi > n\sqrt{-2\ln\left(\frac{p}{4}\right)}$. Therefore, we need to impose two conditions:

$$\frac{\phi}{n} > 2c_\chi \geqslant 2\sqrt{-2\ln\left(\frac{p}{8}\right)}, \qquad \frac{\phi}{n} > \sqrt{-2\ln\left(\frac{p}{4}\right)}. \tag{B.39}$$

The inequality $\frac{\phi}{n} > 2c_\chi$ is the opposite of the perturbation in assumption 3.7 as $\phi/n = n(n-1)/n^3 k^2\eta > c_\chi > 0$ is satisfied for all $n \geqslant 2$ when $k^2/2n\eta \geqslant c_\chi$. The other two, for fixed (n,k,η) identify a region $p \in (p_c, 1)$ of allowed probabilities. The expression of p_c is explicit:

$$p_c > 8\sqrt{e^{-\phi^2/8n^2}}, \tag{B.40}$$

where the RHS is less than unity, making the interval of allowed values of p non-empty, when $k^2\eta/n > 4\sqrt{2\ln 8}$, which is again a violation of assumption 3.7. Retracing our steps back, under the last condition there exists an interval for c_χ, and automatically equations B.39 are satisfied, which implies that all constants c_χ, c_ξ, c_ϕ exist. The constant c_δ always exist for any given p, and the final probability is:

$$\mathbb{P}_{H_0}\left[s_{\text{global}}(\mathbf{Y}) - \mu_0 \geqslant \xi\right] + \mathbb{P}_{H_1}\left[s_{\text{global}}(\mathbf{Y}) - \mu_0 < \xi\right] \leqslant p_c = 8\sqrt{e^{-\phi^2/8n^2}} + \epsilon < 1, \qquad \forall \epsilon > 0. \tag{B.41}$$

In particular, the threshold ξ is in the non-empty interval $nc_\chi < \xi < \phi - nc_\chi$, and we take the constant:

$$c_\chi = \sqrt{-2\ln\left(\frac{p_c}{8}\right)}. \tag{B.42}$$

□

C AVERAGE-CASE HARDNESS AND STATISTICAL-TO-COMPUTATIONAL GAPS (FOR UNFAMILIAR READERS)

A common problem in statistics is to understand "emergence" of structure, where by emergence we mean the statistical threshold at which the latent random variable is visible. We suppose our observation $\mathbf{Y}$ is for example a random matrix, sampled from an unknown distribution, and aim to recover information about the distribution from the data. In an ideal setting, we have all the deterministic information possible, and wish to infer only the latent random structure.

Let us propose a motivating example. For given $(k, \lambda) \in \mathbb{R}_+^2$, a form of the planted sub-matrix model as in equation 1.6 is such that there is a latent set $\mathcal{S}$ of around k indexes $i \in [n]$. Therefore, there is a $|\mathcal{S}| \times |\mathcal{S}|$ sub-matrix of $\mathbf{Y}$ that has morally more $+1$ entries. The increased number of positive entries compared to the rest is precisely a notion of structure.

Remark C.1. *Since λ enters a probability, the parameter space is without loss of generality $(k, \lambda) \in \mathbb{R}_+ \times [0, 1]$. The case in which $\lambda < 0$ is symmetric.*

WHAT IS A STATISTICAL THRESHOLD? In this statistician-friendly formulation, we know all deterministic features of the problem: the model, the model size n, the signal size k, and the signal strength λ. What we want to understand is the nature of $\mathcal{S}$. In particular, any other formulation where we do not know more is naturally harder to solve, in any sense possible. For the purpose of this document, we restrict to the optimistic scenario where only the latent information is unknown. The only issue is that we do not know $\mathcal{S}$ *a priori*: it is random and there are $\binom{n}{k}$ possible realizations. In practice, if we plot the matrix it does not have a block structure. Intuitively, if λ is large, i.e. we plant a strong signal, or k is "large", i.e. we plant a "big" sub-matrix, it should be "visible", but we are interested in formalizing what *large*, *big*, and *visible* mean. For this purpose, we control randomness by taking a setting where (k, λ) vary jointly, possibly with other auxiliary parameters. Let us store them abstractly in a vector $\boldsymbol{\theta} \in \Theta \subset \mathbb{R}^K$. The interpretation is that we see the planted sub-matrix problem as a model that depends on its parameters $\boldsymbol{\theta}$ and want to answer questions such as problems 1.1 - 1.3 - 1.5, reported here informally for convenience:

(Q1) In which regions of Θ can we understand if we observed a matrix with structure or not at all?

(Q2) In which regions of Θ can we understand if we observed a matrix with large/strong structure?

(Q3) In which regions of Θ can we find the latent structure?

Remark C.2. *A common trait is that the regions exhibit dependence within parameters. For example, we would have $\lambda \equiv \lambda(n), k \equiv k(n)$ and n, so that as n varies also (k, λ) do. The simplifying approach is to take $n \to \infty$, and then potentially refine with non-asymptotic results.*

A statistical threshold is the boundary at which the behavior of a problem with randomness changes. When it is unveiled, we find well-behaved and identifiable regions of different statistical behaviors; from one to the other, a new characteristic of the problem emerges. As an example, suppose $\boldsymbol{\theta} \neq \boldsymbol{\theta}'$ where $\boldsymbol{\theta}'$ is a slight perturbation of $\boldsymbol{\theta}$ that goes just outside the region where we answer (Q1) positively. Then, we are *sure* that there is not enough information in $\boldsymbol{\theta}'$. In other words, in moving from $\boldsymbol{\theta}'$ to $\boldsymbol{\theta}$ in parameter space the signal emerges from noise, and the boundary we crossed is sharp in some quantifiable sense. While the questions above are informal, there are canonical formulations in statistics that formalize them.

We want to answer (Q1) - (Q2) - (Q3). Problems 1.1 - 1.3 - 1.5 formalize them in the language of classical statistics, but what does it mean to "solve them well"? According to which measure of goodness?

SUCCESS CRITERION In problems 1.1 - 1.3 - 1.5 we glissed over what "solving well" means. For estimation, it is common practice to consider a loss with respect to the ground truth. In some cases, this is an ℓ^p loss. Since there is randomness involved, we will take the expectation of it and require obtaining guarantees on the loss over the randomness. For example:

(C1) if the loss is on average larger than a certain value, we have a negative result;

(C2) if it is smaller than a certain value, we have a positive result;

where positive and negative merely mean that we are sure to be better/worse than such loss on average. The trick is then to understand what is the best we can do, or the least considering the trivial method, e.g. random

guessing, and comparing. For hypothesis testing, the analog is to study the probability of making a mistake, which is the sum of type I and type II errors.
The key aspect of the comparison step is understand what we want to study. In the next two paragraphs, we present a seasoned and a modern view on hypothesis tests such as problems 1.1 - 1.3.

CLASSICAL Statistics has long focused on universal guarantees over any possible function. For example, studying (C1) above boils down to proving a result such as:

"For any $\boldsymbol{\theta} \in \Theta_{\mathsf{imp}}$ the type I and type II error are larger than $1/2 - o(1)$ as $n \to \infty$".

The formalization is:

$$\inf_{f:\{-1,1\}^{n\times n}\mapsto\{0,1\}} \left\{\mathbb{P}_{\mathrm{H}_0}\left[f(\mathbf{Y})=1\right] + \mathbb{P}_{\mathrm{H}_1}\left[f(\mathbf{Y})=0\right]\right\} \geqslant 1 - o(1), \qquad \forall \boldsymbol{\theta} \in \Theta_{\mathsf{imp}}, \tag{C.3}$$

and tells us that for any function there is no construction that performs significantly better than random guessing for problem configurations in Θ_{imp}. In other words, Θ_{imp} is a region of impossible problem instances, where we cannot answer (Q1) or (Q2). An example of companion result of the (C2) type is for example:

"For any $\boldsymbol{\theta} \notin \Theta_{\mathsf{imp}}$ there exists an explicit $f^\star$ (or more than one) that attains small type I and type II error as $n \to \infty$".

In full analogy, we write this as:

$$\inf_{f:\{-1,1\}^{n\times n}\mapsto\{0,1\}} \left\{\mathbb{P}_{\mathrm{H}_0}\left[f(\mathbf{Y})=1\right] + \mathbb{P}_{\mathrm{H}_1}\left[f(\mathbf{Y})=0\right]\right\} \leqslant 1 - \Omega(1), \qquad \forall \boldsymbol{\theta} \notin \Theta_{\mathsf{imp}}, \tag{C.4}$$

and $f^\star$ attaining it known, e.g. some complicated integral.

Remark C.5. *While we only presented an example, there are many ways to write down these statements, according to the notion of solving "well" the problem. In the positive result just above, we could have asked to be $o(1)$. We could have written these down non-asymptotically with precise quantitative versions of $o(\cdot), O(\cdot)$. The theory is rather flexible in this sense.*

The main bottleneck of these principles is that in many cases the optimal function $f^\star$ is known, explicit, but takes **exponential time to compute**. Given an observation $\mathbf{Y}$ with $n \gg 1$, we are hopeless to have the answer before the age of the universe. This non-practicality observation has drawn research to the scenario in which we replace functions with algorithms.

COMPUTATIONAL BOUNDS To capture the behavior of efficient functions, we rephrase equation C.3 as:

$$\inf_{\substack{f:\{-1,1\}^{n\times n}\mapsto\{0,1\}\\ f \text{ computable in poly-time}}} \left\{\mathbb{P}_{\mathrm{H}_0}\left[f(\mathbf{Y})=1\right] + \mathbb{P}_{\mathrm{H}_1}\left[f(\mathbf{Y})=0\right]\right\} \geqslant 1 - o(1), \qquad \forall \boldsymbol{\theta} \in \Theta_{\mathsf{algo\,imp}}, \tag{C.6}$$

and equation C.4 as:

$$\inf_{\substack{f:\{-1,1\}^{n\times n}\mapsto\{0,1\}\\ f \text{ computable in poly-time}}} \left\{\mathbb{P}_{\mathrm{H}_0}\left[f(\mathbf{Y})=1\right] + \mathbb{P}_{\mathrm{H}_1}\left[f(\mathbf{Y})=0\right]\right\} \leqslant 1 - \Omega(1), \qquad \forall \boldsymbol{\theta} \notin \Theta_{\mathsf{algo\,imp}}. \tag{C.7}$$

In particular, by the fact that we restrict the optimization, we have $\Theta_{\mathsf{algo\,imp}} \supseteq \Theta_{\mathsf{imp}}$. With these types of inclusions we can create a hierarchy of hardness of problems which is useful to define.

Definition C.8 (Domination)**.** *Since we look at problems over a phase diagram of parameters, two problems $\mathsf{P}_1, \mathsf{P}_2$ are put in relation depending on how large an area they cover with a positive answer. Suppose two given problems $\mathsf{P}_1, \mathsf{P}_2$ depend on a set of parameters in $\mathbb{R}^K$. Given a criterion for returning a solution, e.g. a loss, they are in order of domination $\mathsf{P}_1 \succ \mathsf{P}_2$ if P_1 is solved in a strictly larger subset of $\mathbb{R}^K$. In particular with this definition the unconstrained method (eqn C.4) always dominates poly-time algorithms (eqn C.7).*

When the inclusion $\Theta_{\mathsf{algo\,imp}} \supset \Theta_{\mathsf{imp}}$ is strict we say we have a **statistical-to-computational gap**. There are regions of the parameter space where problem instances are solvable by a generic function but not by algorithms. The importance of studying existence of gaps is merely practical: for a given problem, it is not sufficient to know when it is solvable, but rather important to know when it is *efficiently* solvable.
Since the notion of gap is formal, in principle one can hope for an $f^\star_{\mathsf{algo}}$ optimizing equation C.7 and a negative

result uniformly over the class of algorithms like equation C.6. However, there is no well-known formalism that captures algorithms with functions and vice versa. In recent years, the approaches proposed consisted in restricting to a certain "computational class". While none of them has a precise answer,[17] modulo some details they tend to make the same predictions on the Θ_{imp} regions across interesting problems.
For the purpose of this document, we focus on the low-degree method, which we discuss at length in section 2 of the main text. It is one of the most flexible and mathematically established. Some alternatives and comments on connections are in subsection 1.II.

[17] Rather, they all work with the conjecture that they are capturing algorithms under nice assumptions.