RESEARCH-ARTICLE

Sharp Phase Transitions in Estimation with Low-Degree Polynomials

YOUNGTAK SOHN, Brown University, Providence, RI, United States

ALEXANDER S WEIN, University of California, Davis, Davis, CA, United States

Open Access Support provided by:

University of California, Davis

Brown University

Published: 15 June 2025

Citation in BibTeX format

STOC '25: 57th Annual ACM Symposium on Theory of Computing
June 23 - 27, 2025
Prague, Czechia

Conference Sponsors:
SIGACT

https://doi.org/10.1145/3717823.3718294
ISBN: 9798400715105

Sharp Phase Transitions in Estimation with Low-Degree Polynomials

Youngtak Sohn
Brown University
Providence, USA
youngtak_sohn@brown.edu

Alexander S. Wein
University of California at Davis
Davis, USA
aswein@ucdavis.edu

Abstract

High-dimensional planted problems, such as finding a hidden dense subgraph within a random graph, often exhibit a gap between statistical and computational feasibility. While recovering the hidden structure may be statistically possible, it is conjectured to be computationally intractable in certain parameter regimes. A powerful approach to understanding this hardness involves proving lower bounds on the efficacy of low-degree polynomial algorithms. We introduce new techniques for establishing such lower bounds, leading to novel results across diverse settings: planted submatrix, planted dense subgraph, the spiked Wigner model, and the stochastic block model. Notably, our results address the estimation task — whereas most prior work is limited to hypothesis testing — and capture sharp phase transitions such as the "BBP" transition in the spiked Wigner model (named for Baik, Ben Arous, and Péché) and the Kesten–Stigum threshold in the stochastic block model. Existing work on estimation either falls short of achieving these sharp thresholds or is limited to polynomials of very low (constant or logarithmic) degree. In contrast, our results rule out estimation with polynomials of degree n^{δ} where n is the dimension and $\delta > 0$ is a constant, and in some cases we pin down the optimal constant δ. Our work resolves open problems posed by Hopkins & Steurer (2017) and Schramm & Wein (2022), and provides rigorous support within the low-degree framework for conjectures by Abbe & Sandon (2018) and Lelarge & Miolane (2019).

CCS Concepts

• **Theory of computation → Random network models**; **Computational complexity and cryptography**.

Keywords

Statistical-Computational Gaps, Community Detection, Low-Degree Polynomials

ACM Reference Format:
Youngtak Sohn and Alexander S. Wein. 2025. Sharp Phase Transitions in Estimation with Low-Degree Polynomials. In *Proceedings of the 57th Annual ACM Symposium on Theory of Computing (STOC '25), June 23–27, 2025, Prague, Czechia.* ACM, New York, NY, USA, 12 pages. https://doi.org/10.1145/3717823.3718294

STOC '25, Prague, Czechia

ACM ISBN 979-8-4007-1510-5/25/06
https://doi.org/10.1145/3717823.3718294

1 Introduction

The task of discovering a hidden "signal" of interest buried in a large noisy dataset, is central to modern statistics and data science. In addition to the statistical question of discerning the weakest possible signal, the high-dimensionality of these problems also poses the computational challenge of finding an algorithm of practical runtime. As a testbed for studying the fundamental limitations of what is achievable in these settings, we focus on the following four canonical models for planted signal in a random matrix or random graph.

- **Planted Submatrix**: For a sparsity parameter $\rho \in [0,1]$ and a signal-to-noise parameter $\lambda \geq 0$, we observe the $n \times n$ matrix $Y = \lambda\theta\theta^\top + Z$ where $\theta \in \mathbb{R}^n$ has i.i.d. Bernoulli(ρ) entries and $Z \in \mathbb{R}^{n\times n}$ is a symmetric matrix with $\{Z_{ij}\}_{i\leq j}$ i.i.d. $\mathcal{N}(0,1)$. Given Y, the goal is to estimate θ.
- **Planted Dense Subgraph**: For a sparsity parameter $\rho \in [0,1]$ and edge probabilities $0 \leq p_0 \leq p_1 \leq 1$, we observe a random graph on n vertices whose adjacency matrix $Y = (Y_{ij})_{1\leq i<j\leq n} \in \{0,1\}^{\binom{n}{2}}$ is generated as follows. First draw $\theta \in \mathbb{R}^n$ with i.i.d. Bernoulli(ρ) entries. Conditioned on θ, draw $Y_{ij} \sim \text{Bernoulli}(p_0 + (p_1 - p_0)\theta_i\theta_j)$ independently for each $i < j$. Thus, edges within the planted subgraph have probability p_1 while others have probability p_0. Given Y, the goal is to estimate θ.
- **Spiked Wigner Model**: This is a canonical model for a low-rank matrix corrupted by additive Gaussian noise. Let $U \in \mathbb{R}^{n\times m}$ have entries i.i.d. from some prior π with mean 0 and variance 1. We observe the $n \times n$ matrix $Y = X + Z$ where $Z \in \mathbb{R}^{n\times n}$ is symmetric with $\{Z_{ij}\}_{i\leq j}$ i.i.d. $\mathcal{N}(0,1)$, and $X = \sqrt{\lambda/n}\,UU^\top$ for a signal-to-noise parameter $\lambda \geq 0$. Given Y, the goal is to estimate X. Unlike most prior work, we allow m to potentially grow with n.
- **Stochastic Block Model (SBM)**: This is a canonical model for community detection in random graphs. To generate an n-vertex graph, first each vertex $i \in [n]$ is independently assigned a community label $\sigma_i^\star$, drawn uniformly from $[q]$ where q is the number of communities. For edge probabilities $0 \leq p_0 \leq p_1 \leq 1$, we observe the graph with adjacency matrix $Y = (Y_{ij})_{1\leq i<j\leq n} \in \{0,1\}^{\binom{n}{2}}$ generated as follows. Independently for each $i < j$, draw $Y_{ij} \sim \text{Bernoulli}(p_0 + (p_1 - p_0)\mathbb{1}_{\sigma_i^\star=\sigma_j^\star})$. Thus, within-community edges have probability p_1 while cross-community edges have probability p_0. Given Y, the goal is to estimate whether two given vertices (say, vertices 1, 2) are in the same community.

> We will also consider the more general SBM where the communities can have different relative sizes and each pair of communities can have a different connection probability.

In each of these models we will be assuming an asymptotic regime where $n \to \infty$, and the other parameters (such as ρ, p_0, p_1, m) may scale with n in some prescribed way or may be designated as fixed "constants" (such as π, q) that do not depend on n. Asymptotic notation such as $O(\cdot)$, $o(\cdot)$, $\Omega(\cdot)$, $\omega(\cdot)$, $\Theta(\cdot)$ will always pertain to this limit. The objective will be to *estimate* (a.k.a. *recover*) the planted signal (such as θ) to some desired accuracy, which may be measured in terms of mean squared error, or in terms of achieving some success metric *with high probability*, i.e., success probability $1 - o(1)$ as $n \to \infty$. We assume the parameters of the model (such as λ, ρ, π) are known to the statistician, but not the latent variables such as θ, U.

The models defined above are all well studied, and we defer a thorough literature review to Section 2. Some prior work focuses on determining the *statistical limits*, that is, for what values of the parameters is it possible versus impossible to succeed, with no restrictions on the estimator. Other work focuses on finding estimators that can be computed *efficiently* (say, in polynomial time). Notably, all the above models appear to exhibit *statistical-computational gaps*, meaning there is a "possible-but-hard" regime of parameters where some estimator is known to succeed via "brute-force" search, yet no polynomial-time algorithm is known to succeed. In such cases, it is desirable to understand whether this hardness is inherent: does there really not exist a poly-time algorithm, or do we just need to work harder to find one? Our focus in this work will be on identifying this transition between "easy" (poly-time solvable) and "hard." Notably, the best known algorithms for the above models have "sharp" phase transitions in their behavior, where the problem abruptly becomes easy once a signal-to-noise parameter passes a certain threshold. Our aim will be to prove matching lower bounds, showing hardness below this precise threshold.

A key challenge in this endeavor is that classical notions of complexity such as NP-hardness are not applicable here, since we are dealing with *average-case* problems where the goal is to succeed for (not all but) "most" random inputs from a particular distribution. Instead, a popular and versatile approach for rigorously vindicating the type of phase transitions we are interested in, is to study the behavior of *low-degree polynomial estimators* [101]. To this end, we consider the task of estimating a particular scalar quantity, denoted x. For instance, in the planted submatrix problem, a natural choice is $x := \theta_1$, the first entry of the signal vector. The class of algorithms we will consider are multivariate polynomials $f \in \mathbb{R}[Y]$ in the input variables Y_{ij}, of degree (at most) some parameter D (which may scale with n). The best performance over all such algorithms is measured by the *degree-D minimum mean squared error*,

$$\mathsf{MMSE}_{\le D} := \inf_{\substack{f \in \mathbb{R}[Y] \\ \deg(f) \le D}} \mathbb{E}[(f(Y) - x)^2], \tag{1}$$

where the expectation is over the joint distribution of (x, Y) as specified by the model. This scalar MMSE can generally be directly related to its vector analogue; see Section 2.2. If the quantity $\mathsf{MMSE}_{\le D}$ is "small" (appropriately defined), we will say degree-D polynomials succeed at the estimation task, and if $\mathsf{MMSE}_{\le D}$ is "large" then we will say degree-D polynomials fail. The degree D serves as a measure of an algorithm's complexity, and the above framework allows us to quantify the difficulty of an estimation task by the degree that is required to solve it. Our main results will establish that in the conjectured "hard" regimes for each of the four models, $\mathsf{MMSE}_{\le D}$ does not even beat the "trivial" estimator $f(Y) \equiv \mathbb{E}[x]$, that is, $\mathsf{MMSE}_{\le D} \ge (1 - o(1))\mathsf{MMSE}_{\le 0}$, for some D scaling as n^δ for a constant $\delta > 0$.

The above notion of "degree complexity" has conjectural connections to the more traditional notion of time complexity (runtime). For the style of problems we are considering, polynomials of degree $O(\log n)$ tend to be powerful enough to capture the best known poly-time algorithms, so if we can prove failure of super-logarithmic polynomials, this is considered an indication that there is no poly-time algorithm for the task (see [52, Conjecture 2.2.4]). For this reason, "low" degree typically means $O(\log n)$, unless stated otherwise. More generally, polynomials of degree D are expected to have the same power as algorithms of runtime $n^{\tilde{O}(D)}$ where $\tilde{O}(\cdot)$ hides a polylog(n) factor (see [52, Hypothesis 2.1.5]), so for example, degree n^δ corresponds to time $\exp(n^{\delta \pm o(1)})$ for fixed $\delta > 0$. We refer the reader to [101] and references therein for further discussion of this framework. For the four models studied in this work, we will give low-degree *upper bounds*, confirming that degree-$O(\log n)$ polynomials succeed at estimation in the "easy" regime where poly-time algorithms are known. This shows that low-degree estimators are a meaningful class of algorithms to consider here.

It is worth noting that we will depart from most prior work on low-degree complexity: the original work on this topic considered the setting of *hypothesis testing* (a.k.a. *detection*) rather than estimation [53, 54] (see [52, 68] for exposition), and there have been numerous follow-up works in this setting, some of which are discussed in Section 2. The case of estimation that we consider here has also received attention [55, 56, 62, 67, 73, 74, 78, 86, 101, 107] but it is more difficult to analyze and our mathematical toolbox is less complete. While results on testing can sometimes shed light on hardness of estimation, this is not the case for many problems — including some that we study here — due to gaps between the testing and estimation thresholds. For this reason it is important to have tools that directly address the estimation task. Existing lower bounds for low-degree estimation either do not reach the type of sharp thresholds that we aim for in this work, or are limited to very low degree (falling short of the super-logarithmic standard discussed above). Some examples of sharp thresholds at very low degree appear in [55, 86, 101]. Lower bounds at very low degree are already interesting but do not necessarily give a reliable prediction for the true computational threshold, as discussed in, e.g., [109]. Our work is the first to simultaneously capture sharp thresholds and rule out super-logarithmic degree in estimation. Specifically, our results will rule out polynomials of degree n^δ for a constant $\delta > 0$, sometimes for the *optimal* constant δ. Being the first results of this type, our work gives further credibility to the low-degree polynomial framework by demonstrating its ability to capture sharp estimation thresholds that coincide with the suspected computational limits. We note that concurrent and independent work of [56] also proves hardness at degree n^δ up to a sharp threshold, namely

the Kesten–Stigum threshold in the broadcast tree model, using techniques that are rather different from ours.

We remark that there are other frameworks for explaining statistical-computational gaps, including average-case reductions (see [20]), the statistical query model (see [22]), the sum-of-squares hierarchy (see [97]), and the overlap gap property (see [46]). However, none of these appear able to capture the sharp estimation thresholds that we study here, at least with current techniques. One exception is the methods based on statistical physics (see [66, 110]), which tend to be non-rigorous but do have an extraordinary track record of predicting sharp thresholds in estimation problems, such as the Kesten–Stigum (KS) threshold in the stochastic block model (SBM) [34] which we will also study here. These methods essentially postulate the optimality of specific algorithms such as *belief propagation (BP)* or *approximate message passing (AMP)*, which tend to be captured by the low-degree class (see [27, 58, 60, 86]). Since physics-style methods are quite different from low-degree methods, it is valuable to corroborate those predictions with rigorous evidence from low-degree approaches. This is especially true because low-degree methods appear to make more reliable predictions for a wider range of problems and scaling regimes, with *tensor PCA* [98] being a prime example of a setting where the physics methods struggle (see [108]). For some models — namely planted dense subgraph and spiked Wigner — we will uncover a new phase transition in a scaling regime where no physics-style prediction of hardness exists (and even if physics methods were applied here, it's not so well-established whether they should be considered reliable in such regimes).

Finally, we note that concurrent and independent work [39] gives a different form of low-degree evidence for one of the threshold phenomena that we study, namely the KS threshold in the SBM. Rather than our approach of giving unconditional bounds on $\mathsf{MMSE}_{\le D}$, their work assumes a conjecture on a certain optimality of low-degree polynomials for hypothesis testing, and uses this to deduce hardness of non-trivial recovery by $\exp(n^{0.99})$-time algorithms. We refer to this type of argument as a *detection-to-recovery reduction*. We discuss the comparison to our work further in Section 2.5. Another recent independent work is [72], which gives evidence for sharp recovery thresholds in certain graph matching models, also using a detection-to-recovery reduction.

1.1 Our Contributions

1.1.1 Example: Planted Submatrix. We now illustrate our results in more detail, focusing on the planted submatrix problem. To recap the setting, a principal submatrix of size roughly $\rho n \times \rho n$ with elevated mean λ is planted in an $n \times n$ symmetric Gaussian matrix. The goal is to estimate θ_1, the indicator for whether vertex 1 belongs to the planted submatrix. We state (a simplification of) our main result for this setting.

Theorem 1.1 (See Theorem 2.2). *Consider the planted submatrix model with $\rho = o(1)$. For any constant $\epsilon > 0$ there exists a constant $C \equiv C(\epsilon) > 0$ such that the following holds for all sufficiently large n. If*

$$\lambda \le (1-\epsilon)(\rho\sqrt{en})^{-1} \qquad \text{and} \qquad D \le \lambda^{-2}/C$$

then

$$\mathsf{MMSE}_{\le D} := \inf_{\substack{f \in \mathbb{R}[Y] \\ \deg(f) \le D}} \mathbb{E}[(f(Y)-\theta_1)^2] \ge \rho - C\rho^2.$$

To interpret this, we restrict the following discussion to the regime $\Omega(1/\sqrt{n}) \le \rho \le o(1)$ for ease of exposition, with the general case discussed in Section 2.2. First note that the trivial MSE achieved by $f(Y) \equiv \mathbb{E}[\theta_1] = \rho$ is $\mathrm{Var}(\theta_1) = \rho - \rho^2$, so we have met our stated goal of showing $\mathsf{MMSE}_{\le D} \ge (1-o(1))\mathsf{MMSE}_{\le 0}$. Next, a poly-time algorithm based on *approximate message passing (AMP)* is known to achieve near-perfect estimation of θ — meaning $o(\rho n)$ misclassification errors, with high probability — when $\lambda \ge (1+\epsilon)(\rho\sqrt{en})^{-1}$ for an arbitrary constant $\epsilon > 0$ [49]. Our condition on λ holds below this sharp threshold. Finally, below the AMP threshold $\lambda \le (1-\epsilon)(\rho\sqrt{en})^{-1}$, the best known algorithms have runtime $\exp(\tilde{O}(\lambda^{-2}))$ [40, 51] (so long as $\lambda \gg (\rho n)^{-1/2}$, which is information-theoretically necessary [25, 63]; we use $\gg$ to hide polylog(n) factors). Recalling the heuristic correspondence between degree and runtime discussed above, our condition on D coincides with this runtime $\exp(\tilde{O}(\lambda^{-2}))$. Thus, our lower bound suggests that below the sharp AMP threshold, runtime $\exp(\tilde{\Omega}(\lambda^{-2}))$ is required. This shows optimality of the existing algorithms in a strong sense, pinning down both the sharp AMP threshold and also the precise degree $\Omega(\lambda^{-2})$. This resolves two different open questions of [101], where coarser low-degree lower bounds were shown.

We emphasize that the hardness of estimation established above cannot be deduced from the statistical limits of estimation [25, 63] nor the computational limits of hypothesis testing [21, 75], since these have different thresholds; see [101].

1.1.2 Other Models. We now summarize our contributions for the other three models presented above, with more details in Section 2.

For planted dense subgraph, we give the same lower bound as for planted submatrix but with the substitution $\lambda^2 = (p_1-p_0)^2/[p_0(1-p_0)]$; see Theorem 2.4. In particular, our lower bound reaches a sharp threshold, and we also show that degree-$O(\log n)$ polynomials succeed above this threshold. However, the matching poly-time algorithm does not seem to be known. Thus, our result suggests a new phase transition phenomenon that may be achievable with an AMP-style algorithm.

The spiked Wigner model and stochastic block model both have well-established conjectured computational thresholds for the onset of weak recovery, i.e., non-trivial estimation. These thresholds are known as the *BBP transition* (after Baik, Ben Arous, and Péché [8]) and *Kesten–Stigum (KS) bound* [34, 61], respectively. Specifically, it was conjectured in [69, Conjecture 10] (based on [65, 71]) and [3, Conjecture 2] (based on [34]) that below the BBP transition and the KS bound, respectively, non-trivial estimation is impossible in polynomial time for these models. Existing low-degree lower bounds show hardness of the associated hypothesis testing problem below these thresholds [9, 54, 68], but no such results were known for estimation and this was stated as an open problem by Hopkins and Steurer in one of the first papers on the low-degree framework [54]. We resolve this, giving estimation lower bounds at degree n^{δ} for a constant $\delta > 0$, below the sharp BBP and KS thresholds. We refer to Theorem 2.6 for the spiked Wigner model, and to Theorem 2.8 for the SBM. Altogether, our results provide rigorous

evidence towards the conjectures of [69] and [3] by establishing the corresponding low-degree hardness. (We do not realistically expect to resolve these conjectures outright, as this would imply a solution to the P versus NP problem.) In the spiked Wigner model, we additionally go beyond the setting of the conjecture by allowing the rank m to grow with n.

1.1.3 Low-Degree Upper Bounds. In the discussion above, we have focused on low-degree *lower bounds*, that is, proving failure of degree-D polynomials. Such results are most meaningful when complemented by the corresponding low-degree *upper bound*, that is, we would like to show $\mathsf{MMSE}_{\le D}$ is "small" for some $D = O(\log n)$ in the "easy" regime where efficient algorithms are known, in order to rigorously establish a phase transition in the behavior of polynomials.

Indeed, we will show for our four models of interest that degree-$O(\log n)$ polynomials succeed above the sharp thresholds appearing in our lower bounds. More specifically, for planted submatrix and planted dense subgraph we will show that degree-$O(\log n)$ polynomials achieve *strong recovery* above the threshold, meaning $\mathsf{MMSE}_{\le D} = o(\mathsf{MMSE}_{\le 0})$. For spiked Wigner and SBM we will give a similar result but for *weak recovery*, meaning $\mathsf{MMSE}_{\le D} = (1-\Omega(1))\mathsf{MMSE}_{\le 0}$. The full details, along with further discussion, are presented in Section 2.

For planted submatrix and planted dense subgraph, our upper bound is proved by constructing a polynomial estimator that is the average of certain "tree-structured" monomials with a particular structure. For spiked Wigner, a sharp low-degree upper bound above the BBP transition was previously known in the rank-1 case ($m = 1$), based on self-avoiding walks [54]. We will extend this to larger m. Similarly, for the SBM, a sharp low-degree upper bound above the KS threshold is known, also based on self-avoiding walks [54]. We will include the proof for completeness.

1.2 Proof Techniques

Here we present the main ideas in the lower bound proofs. The full proofs can be found in the full version [104].

1.2.1 Overview of the Proof Strategy. Consider for now the generic setting where the goal is to estimate some scalar x given Y. Instead of $\mathsf{MMSE}_{\le D}$, it will be more convenient to work with the *degree-D maximum correlation*

$$\mathsf{Corr}_{\le D} := \sup_{\substack{f\in\mathbb{R}[Y]\\ \deg(f)\le D}} \frac{\mathbb{E}[f(Y)\cdot x]}{\sqrt{\mathbb{E}[f(Y)^2]\cdot\mathbb{E}[x^2]}}.$$

We have taken a normalization such that $\mathsf{Corr}_{\le D}\in[0,1]$, which deviates from the one used by [101]. The following fact shows that $\mathsf{Corr}_{\le D}$ and $\mathsf{MMSE}_{\le D}$ are directly related.

Fact 1.2 ([101], Fact 1.1). $\mathsf{MMSE}_{\le D} = (1-\mathsf{Corr}^2_{\le D})\,\mathbb{E}[x^2]$.

We aim to give an upper bound on $\mathsf{Corr}_{\le D}$, and the main difficulty is the factor $\mathbb{E}[f(Y)^2]$ in the denominator, since Y does not have independent entries. We will make use of the underlying independent random variables from which Y is generated, for instance Z and θ in the planted submatrix problem.

Choose a basis $\{\phi_\alpha\}_{\alpha\in\mathcal{I}}$ for $\mathbb{R}[Y]_{\le D}$ (polynomials in the entries of Y of degree $\le D$) so that any candidate polynomial estimator can be expanded as

$$f(Y) = \sum_\alpha \hat{f}_\alpha \phi_\alpha(Y)$$

for some vector of (real) coefficients $\hat{f} = (\hat{f}_\alpha)_{\alpha\in\mathcal{I}}$. In the settings we consider there will always be a vector W of independent random variables from which x and Y are generated. We will need to construct an orthonormal collection $\{\psi_\beta\}_{\beta\in\mathcal{J}}$ of polynomials in (the entries of) W. The meaning of "orthonormal" here is $\langle\psi_\beta,\psi_{\beta'}\rangle := \mathbb{E}[\psi_\beta(W)\cdot\psi'_\beta(W)] = \mathbb{1}_{\beta=\beta'}$. There is no requirement on the degrees of the polynomials ψ_β nor on their span (they need not form a full basis), but it will be advantageous to use a "rich" collection of polynomials. We emphasize that, unlike $\{\psi_\beta\}$, the basis $\{\phi_\alpha\}$ need not be orthogonal. Working in the Hilbert space of random variables, we can bound the norm of $f(Y)$ using its projections onto the orthonormal directions ψ_β:

$$\mathbb{E}[f(Y)^2] \ge \sum_\beta \mathbb{E}[f(Y)\cdot\psi_\beta(W)]^2.$$

Equivalently, $\mathbb{E}[f(Y)^2] \ge \|M\hat{f}\|^2$ where $M = (M_{\beta\alpha})_{\beta\in\mathcal{J},\alpha\in\mathcal{I}}$ is defined by

$$M_{\beta\alpha} = \mathbb{E}[\phi_\alpha(Y)\cdot\psi_\beta(W)].$$

Also define the vector $c = (c_\alpha)_{\alpha\in\mathcal{I}}$ by

$$c_\alpha = \mathbb{E}[\phi_\alpha(Y)\cdot x].$$

The core of our approach will be to construct a certain vector u, which then implies a bound on $\mathsf{Corr}_{\le D}$.

Proposition 1.3. *Suppose there is a vector $u = (u_\beta)_{\beta\in\mathcal{J}}$ satisfying the linear equations $u^\top M = c^\top$, or equivalently,*

$$\sum_{\beta\in\mathcal{J}} u_\beta M_{\beta\alpha} = c_\alpha \tag{2}$$

for all $\alpha\in\mathcal{I}$. Then

$$\mathsf{Corr}_{\le D} \le \frac{\|u\|}{\sqrt{\mathbb{E}[x^2]}} := \sqrt{\frac{\sum_{\beta\in\mathcal{J}} u_\beta^2}{\mathbb{E}[x^2]}}.$$

Proof. Recalling the fact $\mathbb{E}[f(Y)^2] \ge \|M\hat{f}\|^2$,

$$\sqrt{\mathbb{E}[x^2]}\cdot\mathsf{Corr}_{\le D} = \sup_f \frac{\mathbb{E}[f(Y)\cdot x]}{\sqrt{\mathbb{E}[f(Y)^2]}} \le \sup_{\hat{f}} \frac{c^\top\hat{f}}{\|M\hat{f}\|}$$
$$= \sup_{\hat{f}} \frac{u^\top M\hat{f}}{\|M\hat{f}\|} \le \sup_{\hat{f}} \frac{\|u\|\cdot\|M\hat{f}\|}{\|M\hat{f}\|} = \|u\|.$$

□

Provided x lies in the span of $\{\psi_\beta\}$, it is not difficult to construct *some* vector u satisfying the constraints $u^\top M = c^\top$, as one solution is $u_\beta = \mathbb{E}[\psi_\beta\cdot x]$ which can be seen from the expansion $x = \sum_\beta \mathbb{E}[x\cdot\psi_\beta]\,\psi_\beta$. However, this gives only the trivial bound $\mathsf{Corr}_{\le D}\le 1$. Ideally, we would take u to be the minimum-norm solution to the constraints, which can be found using the Moore–Penrose pseudoinverse: $u^\top = c^\top M^+$. The issue here is that this value of u seems difficult to work with explicitly since we do not have a closed-form expression for M^+. Instead, our approach will be to manually construct a more tractable solution u with a small enough

norm. As we will see in the subsequent sections, our proof leverages a key geometric insight that "tree-structured" polynomials are most informative for estimation in the planted submatrix/subgraph problems whereas "path-shaped" polynomials are most effective for spiked Wigner and SBM. In Section 1.2.4, we illustrate our construction of u for the planted submatrix model. Although this solution is explicit in that model, we emphasize that for the stochastic block model, our construction of u is defined recursively over certain set of graphs, necessitating a delicate analysis to control its ℓ_2 norm.

We are hopeful that the above strategy may be useful to prove new low-degree lower bounds in settings beyond those considered in this work. This new approach has not appeared before in its current form, although it takes some inspiration from the prior works [101, 107][1] which also crucially use orthogonal polynomials in the underlying independent random variables. Compared to [107], our approach is a generalization (the ψ_β may not form a basis) and simplification (we only need to construct u solving $u^\top M = c^\top$ rather than a left-inverse for M). While [107] studies a different problem from us (tensor decomposition, which does not appear to have sharp threshold behavior), [101] studies some of the same problems that we do, and we sharpen the bounds in a qualitative way, allowing us to capture sharp thresholds. We will give more details on the comparison to [101] in Section 1.2.4 below.

1.2.2 Removing "Bad" Terms. We will designate some values of α and β as "good," denoted $\alpha \in \hat{\mathcal{I}}$ or $\beta \in \hat{\mathcal{J}}$, and all others as "bad" (denoted, e.g., $\alpha \notin \hat{\mathcal{I}}$). The choice of $\hat{\mathcal{I}} \subseteq \mathcal{I}$ and $\hat{\mathcal{J}} \subseteq \mathcal{J}$ will be problem-specific and will depend on the estimand x, but intuitively the good basis elements represent those that are potentially useful, or "informative," for estimation of x whereas the bad ones represent those that "uninformative." Under certain conditions we will be able to remove the bad terms from consideration. We refer to Sections 1.2.4 and 1.2.5 for specific examples.

Lemma 1.4. *Suppose the following conditions are met:*

- $u_\beta = 0$ *for all* $\beta \notin \hat{\mathcal{J}}$.
- *For each* $\alpha \notin \hat{\mathcal{I}}$ *there exist* $\hat{\alpha} \in \hat{\mathcal{I}}$ *and* $\mu \in \mathbb{R}$ *such that* $c_\alpha = \mu c_{\hat{\alpha}}$ *and* $M_{\beta\alpha} = \mu M_{\beta\hat{\alpha}} \ \ \forall \beta \in \hat{\mathcal{J}}$.

If (2) *holds for all* $\alpha \in \hat{\mathcal{I}}$ *then* (2) *holds for all* $\alpha \in \mathcal{I}$.

Proof. Fix $\alpha \notin \hat{\mathcal{I}}$ and the corresponding $\hat{\alpha}, \mu$. We verify (2):

$$c_\alpha = \mu c_{\hat{\alpha}} = \mu \sum_\beta u_\beta M_{\beta\hat{\alpha}} = \mu \sum_{\beta \in \hat{\mathcal{J}}} u_\beta M_{\beta\hat{\alpha}} = \sum_{\beta \in \hat{\mathcal{J}}} u_\beta M_{\beta\alpha} = \sum_\beta u_\beta M_{\beta\alpha}.$$

□

Remark 1.5. *In our applications of this framework,* $\mathcal{J}$ *will be a set consisting of pairs* (β, γ) *where* $\beta \in \mathcal{I}$ *and* γ *is a certain "coloring" of* β *that encodes the signal structure. The analogous construction applies to the sets* $\hat{\mathcal{I}}$ *and* $\hat{\mathcal{J}}$. *Thus, the role of* β *in Lemma 1.4 will be played by a pair* (β, γ), *abbreviated as* $\beta\gamma$.

[1] Additionally, we thank Jonathan Niles-Weed for discussions that helped inspire this approach.

1.2.3 Notation. For indexing polynomials we will often use elements $\alpha \in \{0,1\}^S$ or $\alpha \in \mathbb{N}^S$ for some finite set S, where $\mathbb{N} := \{0, 1, 2, \ldots\}$. For such elements α, β we define $|\alpha| := \sum_{k \in S} \alpha_k$, $\alpha! := \prod_{k \in S} \alpha_k!$, and $\binom{\alpha}{\beta} := \prod_{k \in S} \binom{\alpha_k}{\beta_k}$. Also, $\beta \le \alpha$ means $\beta_k \le \alpha_k$ for all $k \in S$; and $\beta \lneq \alpha$ means $\beta \le \alpha$ and there exists $k \in S$ where $\beta_k < \alpha_k$. For a vector $X = (X_k)_{k \in S}$, we write $X^\alpha := \prod_{k \in S} (X_k)^{\alpha_k}$.

An element $\alpha \in \{0,1\}^S$ can equivalently be viewed as a subset of S, specifically the subset $\{k \in S : \alpha_k = 1\}$. We may abuse notation and identify α with this subset. For $\alpha, \beta \in \{0,1\}^S$, the meaning of $\beta \le \alpha$ is $\beta \subseteq \alpha$, and the meaning of $\beta \lneq \alpha$ is $\beta \subsetneq \alpha$. An element $\alpha \in \mathbb{N}^S$ can be viewed as a multiset, containing α_k copies of each $k \in S$.

One common choice of S will be $[n] := \{1, 2, \ldots, n\}$. Another will be either $\{(i,j) : 1 \le i < j \le n\}$ or $\{(i,j) : 1 \le i \le j \le n\}$, which we will abbreviate as $\binom{[n]}{2}$ or $\binom{[n]}{2}$ ("multichoose" notation) respectively. We will also abbreviate $\alpha_{(i,j)}$ as α_{ij} in this case. Note that $\alpha \in \{0,1\}^{\binom{[n]}{2}}$ can be viewed as a (simple) graph on vertex set $[n]$, namely the graph that includes edge (i,j) whenever $\alpha_{ij} = 1$. Similarly $\alpha \in \mathbb{N}^{\binom{[n]}{2}}$ can be viewed as a multigraph on vertex set $[n]$ (now with self-loops and parallel edges allowed), namely the multigraph with α_{ij} copies of edge (i,j) for all $i \le j$. With this view, we will often refer to graph-theoretic properties of α. We use $V(\alpha) \subseteq [n]$ to denote the vertex set of α, by which we mean the set of non-isolated vertices. In other words, $i \in V(\alpha)$ if there exists some edge (i,j), possibly a self-loop (i,i). We call α *connected* if every pair of vertices in $V(\alpha)$ has a path connecting them. The empty graph $\alpha = 0$ is considered to be connected. We write $\alpha = \mathbb{1}_{(i,j)}$ for the graph consisting of a single edge between vertices i and j. As above, we may abuse notation and identify α with the associated edge set, e.g., writing $\alpha \cap \beta$ for the common edges between two graphs. We may also explicitly write $E(\alpha)$ for the edge set of α, that is, the (multi)set of pairs (i,j) with $i \le j$ corresponding to the edges of α (where parallel edges are included multiple times based on their multiplicity).

1.2.4 Example: Planted Submatrix. We now sketch how to apply the above framework to planted submatrix, and how this differs from the prior work [101]. To recap the setup, we have parameters $\lambda \ge 0$ and $\rho \in [0,1]$, and the observed data is $(Y_{ij})_{1 \le i \le j \le n}$ generated as follows: $Y = X + Z$ where $X_{ij} = \lambda \theta_i \theta_j$, $\theta \in \{0,1\}^n$ is i.i.d. Bernoulli(ρ), and $\{Z_{ij}\}_{i \le j}$ are i.i.d. $\mathcal{N}(0,1)$. The goal is to estimate $x := \theta_1$.

To construct orthogonal polynomials, we will make use of the multivariate Hermite polynomials $\{H_\alpha\}$ for $\alpha \in \mathbb{N}^{\binom{[n]}{2}}$ (see [105]). These are well known to be orthogonal with respect to Gaussian measure, and we use the normalization for which they are orthonormal: $\mathbb{E}[H_\alpha(Z) \cdot H_\beta(Z)] = \mathbb{1}_{\alpha=\beta}$. As our basis for $\mathbb{R}[Y]_{\le D}$, choose $\phi_\alpha(Y) := H_\alpha(Y)$ for $\alpha \in \mathcal{I} := \{\alpha \in \mathbb{N}^{\binom{[n]}{2}} : |\alpha| \le D\}$. For our orthonormal polynomials in the underlying independent random variables $W = (Z, \theta)$, choose

$$\psi_{\beta\gamma}(Z, \theta) = H_\beta(Z) \left(\frac{\theta - \rho}{\sqrt{\rho(1-\rho)}} \right)^\gamma$$

for $\beta\gamma \in \mathcal{J} := \{(\beta, \gamma) : \beta \in \mathbb{N}^{\binom{[n]}{2}}, |\beta| \le D, \gamma \in \{0,1\}^n\}$. We will often view γ as a subset $\gamma \subseteq [n]$ as described above. Note that

$\{\psi_{\beta\gamma}\}$ is orthonormal as required. With this setup, we will compute (with details deferred to the full version [104])

$$c_\alpha = \frac{\lambda^{|\alpha|}\rho^{|V(\alpha)\cup\{1\}|}}{\sqrt{\alpha!}},$$

$$M_{\beta\gamma,\alpha} = \mathbb{1}_{\beta\le\alpha}\cdot\mathbb{1}_{\gamma\subseteq V(\alpha-\beta)}\cdot\sqrt{\frac{\beta!}{\alpha!}}\binom{\alpha}{\beta}\lambda^{|\alpha-\beta|}\rho^{|V(\alpha-\beta)|}\left(\frac{1-\rho}{\rho}\right)^{|\gamma|/2}.$$

Towards applying Lemma 1.4, we define $\hat{\mathcal{I}}$ and $\hat{\mathcal{J}}$ as follows.

- By convention, $0 \in \hat{\mathcal{I}}$. For $|\alpha| \ge 1$ we include α in $\hat{\mathcal{I}}$ if and only if α (when viewed as a multigraph) is connected with $1 \in V(\alpha)$.
- We include $\beta\gamma$ in $\hat{\mathcal{J}}$ if and only if $\beta \in \hat{\mathcal{I}}$ and $\gamma \subseteq V(\beta)\cup\{1\}$. (The union with $\{1\}$ only matters in the case $\beta = 0$.)

Later, we will check the conditions in Lemma 1.4, allowing the "bad" terms $\alpha \notin \hat{\mathcal{I}}$ and $\beta \notin \hat{\mathcal{J}}$ to be disregarded. This step is analogous to [101], where disconnected graphs were similarly removed from consideration. Now, combining Proposition 1.3 and Lemma 1.4, our goal is to choose values $(u_{\beta\gamma})_{\beta\gamma\in\hat{\mathcal{J}}}$ and verify

$$\sum_{\beta\gamma\in\hat{\mathcal{J}}} u_{\beta\gamma}M_{\beta\gamma,\alpha} = c_\alpha \qquad \forall\alpha\in\hat{\mathcal{I}}.$$

Then our final bound will be

$$\mathrm{Corr}_{\le D} \le \frac{\|u\|}{\sqrt{\mathbb{E}[x^2]}} = \rho^{-1/2}\sqrt{\sum_{\beta\gamma\in\hat{\mathcal{J}}} u_{\beta\gamma}^2}$$

since we are choosing $u_{\beta\gamma} = 0$ for $\beta\gamma \notin \hat{\mathcal{J}}$.

Using the support structure of M and the fact $M_{\alpha 0,\alpha} = 1$, our constraints can be written as

$$u_{\alpha 0} = c_\alpha - \sum_{\beta\le\alpha}\left(\sum_{\gamma\subseteq V(\alpha-\beta)} u_{\beta\gamma}M_{\beta\gamma,\alpha}\right) \qquad \forall\alpha\in\hat{\mathcal{I}}. \tag{3}$$

For all $\alpha\gamma \in \hat{\mathcal{J}}$ with $\gamma \ne 0$, we refer to the values $u_{\alpha\gamma}$ as *free variables*. We are free to choose values for these variables and then the remaining values $u_{\alpha 0}$ are determined by the recurrence (3). Setting all free variables to zero recovers the existing method of [101]. This amounts to using only the orthonormal polynomials $\psi_\beta(Z) = H_\beta(Z)$, which means only the "noise" is being used to lower-bound $\mathbb{E}[f(Y)^2]$. We will improve on this by using both the signal and noise. Specifically, we will set the free variables in such a way to zero out the term in parentheses in (3). As a result, we will have simply $u_{\alpha 0} = c_\alpha$ for all $\alpha \in \hat{\mathcal{I}}$. This avoids some difficult-to-control recursive blowup of the u values that is present in [101]. Our specific construction for u is

$$u_{\alpha\gamma} = \left(-\sqrt{\frac{\rho}{1-\rho}}\right)^{|\gamma|}\cdot c_\alpha \qquad \forall\alpha\gamma\in\hat{\mathcal{J}}.$$

The remaining details, including the computation of $\|u\|$, are deferred to the full version [104]. The dominant contributions to $\|u\|$ come from α that (viewed as a multigraph) are trees containing vertex 1.

1.2.5 Other Models. We now summarize how the arguments change for the other models we consider. For the planted dense subgraph problem, the choice for the orthonormal set $\{\psi_{\beta\gamma}\}$ is more subtle. However, we ultimately manage to use a similar construction for u as above. In particular, the set of "good" terms and "bad" terms are the same as those used for the planted submatrix problem, besides the consideration of simple graphs rather than multigraphs.

The proofs for the spiked Wigner model and SBM are more involved. A key difference lies in the definition of the "good" terms. This difference stems, in part, from the dependence of the estimand x on both vertices 1 and 2. Specifically, "good" α now satisfy $1, 2 \in V(\alpha)$ and every $v \in V(\alpha)\setminus\{1,2\}$ has degree at least 2. The intuition behind the last condition is that α that contains a leaf node which is not 1, 2 is "uninformative" for the estimation of x.

Furthermore, we cannot obtain a closed-form solution for u as in the prior cases. Consequently, the values of $u = (u_{\beta\gamma})$ are defined implicitly through a recursive formula: $(u_{\alpha\gamma})_\gamma$ is constructed based on the values of $(u_{\beta\gamma})_\gamma$ for $|\beta| < |\alpha|$. The construction is designed to minimize $\sum_\gamma u_{\alpha\gamma}^2$ subject to a linear constraint at each step. The key aspect of the analysis involves controlling their growth. Crucially, we show that $\sum_\gamma u_{\alpha\gamma}^2$ does not grow exponentially in $|\alpha|$ (as in the bounds from [101]) but exponentially in $|\alpha| - |V(\alpha)| + 1$ (which represents edges in excess of a tree). This turns out to be crucial to capture the sharp thresholds that we aim for at degree n^δ.

For planted submatrix and planted dense subgraph, the dominant contributions to $\|u\|$ come from trees containing vertex 1, where vertex 1 is distinguished because we are estimating θ_1. The importance of trees here appears sensible, given that the optimal algorithm is AMP, which can essentially be expressed as a sum of such "tree-structured" polynomials [13, 27, 58, 60, 86]. For spiked Wigner and SBM, there are two distinguished vertices 1, 2 since we will be estimating pairwise quantities, and the dominant contributions to $\|u\|$ come from simple paths from vertex 1 to 2. This reflects the fact that the optimal algorithms are based on (self-avoiding or non-backtracking) paths [19, 54, 64, 81].

2 Main Results

We now present our main results, with the proofs deferred to the full version [104].

2.1 Correlation and Weak Recovery

In all the models we consider, the observation will be called Y and the scalar quantity to be estimated will be called x. We will state our results in terms of $\mathrm{Corr}_{\le D}$, as defined in Section 1.2.1:

$$\mathrm{Corr}_{\le D} := \sup_{f\in\mathbb{R}[Y]_{\le D}} \frac{\mathbb{E}[f(Y)\cdot x]}{\sqrt{\mathbb{E}[f(Y)^2]\cdot\mathbb{E}[x^2]}} \in [0,1].$$

To show low-degree hardness of estimation, our goal will be to rule out *weak recovery*, defined as $\mathrm{Corr}_{\le D} = \Omega(1)$. That is, we aim to prove $\mathrm{Corr}_{\le D} = o(1)$ for $D = n^{\Omega(1)}$, which by Fact 1.2 implies

$$\mathsf{MMSE}_{\le D} \ge (1-o(1))\mathbb{E}[x^2] \ge (1-o(1))\mathsf{MMSE}_{\le 0},$$

meaning degree-D polynomials have no significant advantage over the trivial estimator $f(Y) \equiv \mathbb{E}[x]$.

For our low-degree upper bounds, we will assume a regime where $\mathbb{E}[x]^2 = o(\mathbb{E}[x^2])$, implying $\mathsf{MMSE}_{\le 0} = (1-o(1))\mathbb{E}[x^2]$.

We will aim to establish either *strong recovery*, meaning $\mathrm{Corr}_{\le D} = 1 - o(1)$ or equivalently $\mathrm{MMSE}_{\le D} = o(\mathrm{MMSE}_{\le 0})$, or *weak recovery*, meaning $\mathrm{Corr}_{\le D} = \Omega(1)$ or equivalently $\mathrm{MMSE}_{\le D} = (1 - \Omega(1))\mathrm{MMSE}_{\le 0}$.

2.2 Planted Submatrix

Definition 2.1 (Planted Submatrix Model). *For parameters $\lambda \ge 0$ and $\rho \in [0,1]$, observe the $n \times n$ matrix $Y = \lambda\theta\theta^\top + Z$ where $\theta \in \{0,1\}^n$ is i.i.d.* Bernoulli(ρ)*, and Z has entries $Z_{ij} = Z_{ji} \sim \mathcal{N}(0,1)$ where $\{Z_{ij}\}$ are independent. The goal is to estimate $x := \theta_1$.*

Theorem 2.2. *Consider the planted submatrix model. For any constant $\epsilon > 0$ there exists a constant $C \equiv C(\epsilon) > 0$ for which the following holds. If*

$$\lambda \le (1-\epsilon)(\rho\sqrt{en})^{-1}\sqrt{1-\rho} \qquad \text{and} \qquad D \le \lambda^{-2}/C$$

then

$$\mathrm{Corr}_{\le D} \le C\sqrt{\frac{\rho}{1-\rho}}.$$

The full version [104] also includes a matching upper bound (added after the STOC submission): $\mathrm{Corr}_{\le O(\log n)}$ approaches 1 for λ above a sharp threshold. Provided $\rho = o(1)$, the thresholds for λ in the lower and upper bounds match, and the lower bound shows $\mathrm{Corr}_{\le D} = o(1)$ as desired.

As discussed in Section 2.1, the lower bound implies $\mathrm{MMSE}_{\le D} \ge (1 - o(1))\mathrm{MMSE}_{\le 0}$, so that the degree-$D$ MMSE is essentially no better than the trivial MMSE, which in this case is $\mathrm{MMSE}_{\le 0} = \rho - \rho^2$. As pointed out in [101], the (scalar) $\mathrm{MMSE}_{\le D}$ that we are working with can be directly related to its vector analogue (see the full version [104]).

Discussion. This model has been thoroughly studied. The statistical limits are well understood [24, 25, 63], as are the computational limits of hypothesis testing between the planted submatrix model and the "null" model $\lambda = 0$ [21, 75]. Our focus is instead on the computational limits of estimation, which has a different threshold (see [101]).

We will focus on the regime $\rho = n^{\gamma-1}$ for a constant $\gamma \in (0,1)$. When $\gamma > 1/2$, an algorithm based on *approximate message passing (AMP)* achieves near-perfect estimation of θ — meaning $o(\rho n)$ misclassification errors, with high probability — provided $\lambda \ge (1+\epsilon)(\rho\sqrt{en})^{-1}$ [49]. This sharp threshold represents the best known performance of any poly-time algorithm. Aside from a general belief that AMP algorithms are powerful, there were no convincing lower bounds reaching this sharp threshold prior to our work. Low-degree lower bounds were given in [101, Appendix E] showing that polynomials of very low degree (a specific constant times $\log n$) fail to reach the sharp threshold, but the bound falls away from the threshold as the degree increases beyond that.

In the other regime, $\gamma < 1/2$, the AMP threshold can be beaten by a very simple algorithm: thresholding the diagonal entries of the observed matrix gives exact recovery of θ (with high probability) provided $\lambda \gg 1$ (where throughout we use $\gg$ to hide a polylog(n) factor). This regime does not have a sharp threshold but rather a smooth tradeoff between signal strength and runtime, with the best known algorithms achieving exact recovery in runtime $\exp(\tilde{O}(\lambda^{-2}))$ for all λ in the range $(\rho n)^{-1/2} \ll \lambda \le$ polylog(n) [40, 51]; estimation becomes statistically impossible when $\lambda \ll (\rho n)^{-1/2}$ [25, 63]. These algorithms also work for $\gamma \ge 1/2$. The low-degree lower bounds of [101] rule out degree-D polynomials when $\lambda \le \Omega(\min\{1, (\rho\sqrt{n})^{-1}\}/D^2)$, which captures the poly-time threshold at $\lambda \approx 1$ but not the specific runtime needed when $\lambda \ll 1$. Prior to our work, the only known lower bounds that capture this specific runtime $\exp(\tilde{O}(\lambda^{-2}))$ were of a rather different nature, ruling out certain Markov chains [14].

Theorem 2.2 gives a comprehensive low-degree lower bound showing optimality of the algorithms above in a strong sense: if λ lies below the AMP threshold, namely $\lambda \le (1-\epsilon)(\rho\sqrt{en})^{-1}$, then the polynomial degree required for estimation is $\Omega(\lambda^{-2})$, suggesting that runtime $\exp(\tilde{\Omega}(\lambda^{-2}))$ is necessary. This resolves two different open questions of [101]: pinning down the sharp AMP threshold and also the precise degree $\Omega(\lambda^{-2})$.

The lower and upper bounds together establish an "all-or-nothing" phase transition for low-degree polynomials in the regime $\gamma > 1/2$: the low-degree MMSE jumps sharply from near-trivial to near-perfect at the AMP threshold. This can be viewed as a computational analogue of, e.g., [93]. While the upper bound confirms that degree-$O(\log n)$ polynomials succeed above the AMP threshold, it remains open to show that degree-$\tilde{O}(\lambda^{-2})$ polynomials succeed below the AMP threshold for $(\rho n)^{-1/2} \ll \lambda \ll 1$. Success of degree-$O(\log n)$ polynomials for $\lambda \gg 1$ is established by [101].

The case $\rho = \Theta(1)$ has also been studied [35, 47, 71, 86]. In this case the relevant scaling for λ is $\lambda = c/\sqrt{n}$ for a constant c. The best known algorithm is again based on AMP [35, 71], but here its MSE converges to some nontrivial constant depending on c, ρ in contrast to the all-or-nothing behavior above. It has been shown that constant-degree polynomials cannot surpass the precise MSE achieved by AMP [86], and extending this result to higher degree remains an interesting open question. The bound that we prove here applies to higher degree polynomials but does not reach the optimal MSE value. In fact, our bound $\mathrm{Corr}_{\le D} \le C\sqrt{\rho/(1-\rho)}$ becomes vacuous in the regime $\rho = \Theta(1)$ unless ρ is a very small constant.

2.3 Planted Dense Subgraph

Definition 2.3 (Planted Dense Subgraph Model). *For parameters $\rho \in [0,1]$ and $p_0, p_1 \in [0,1]$, we observe $Y = (Y_{ij}) \in \{0,1\}^{\binom{[n]}{2}}$ generated as follows.*

- *A planted signal $\theta = (\theta_i)_{1 \le i \le n} \in \{0,1\}^n$ is drawn with i.i.d.* Bernoulli(ρ) *entries.*
- *Conditioned on θ, $Y_{ij} \sim$ Bernoulli$(p_0 + (p_1 - p_0)\theta_i\theta_j)$ is sampled independently for each $i < j$.*

The goal is to estimate $x := \theta_1$.

Without loss of generality we will assume $p_0 \le p_1$, since otherwise one can consider the complement graph instead. The main result is as follows.

Theorem 2.4. *Consider the planted dense subgraph model with $0 < p_0 \le p_1 \le 1$, and define*

$$\lambda := \frac{p_1 - p_0}{\sqrt{p_0(1-p_0)}}. \tag{4}$$

For any constant $\epsilon > 0$, there exists a constant $C \equiv C(\epsilon) > 0$ for which the following holds. If

$$\lambda \le (1-\epsilon)(\rho\sqrt{en})^{-1}\sqrt{1-\rho} \qquad \text{and} \qquad D \le \lambda^{-2}/C$$

then

$$\mathrm{Corr}_{\le D} \le C\sqrt{\frac{\rho}{1-\rho}}.$$

Note that this matches our result for planted submatrix (Theorem 2.2) with the substitution (4). The full version [104] also includes a matching upper bound (added after the STOC submission).

Discussion. Compared to planted submatrix, the planted dense subgraph problem has a wider variety of different regimes with different behaviors, depending on the scaling of p_0, p_1. Many statistical results exist for this model [7, 28, 106], as well as computational lower bounds for testing [48], and positive algorithmic results [6, 16, 28, 84]. Computational limits for estimation were considered by [101], and our result provides a sharper refinement.

To simplify the discussion, we focus for now on the regime considered by [48] where $1/n^2 \ll p_0 \ll 1$ and $p_1 = cp_0$ for a constant $c > 1$, and furthermore we restrict our attention to the case $1/\sqrt{n} \ll \rho \ll 1$ which is most relevant for us due to the detection-recovery gap (see [21, Conjecture 2.2]). Under these assumptions, the results of [101] already give matching upper and lower bounds for low-degree polynomials on a coarse scale, namely, the threshold occurs at $n\rho^2 p_0 = \tilde{\Theta}(1)$. This is achieved by a very simple algorithm that selects vertices of large degree. Our result refines this, showing a low-degree phase transition at the sharp threshold $n\rho^2 p_0 = [e(c-1)^2]^{-1}$. To our knowledge, this threshold has not appeared before in the literature, except in one specific scaling regime considered by [84] where $p_0, p_1 = \Theta(1/n)$, $\rho = \Theta(1)$ and $\rho \to 0$ *after* $n \to \infty$. Notably, a matching poly-time algorithm that achieves this threshold has not appeared in the literature (outside the specific regime of [84]). Our upper bound implies an algorithm of quasi-polynomial runtime $n^{O(\log n)}$, by directly evaluating the polynomial term-by-term. We expect it should be possible to give a poly-time algorithm by approximately evaluating this "tree-structured" polynomial via the "color coding" trick [5, 54], which has been used by [79, 80] to approximate tree-structured polynomials. We also expect that a more practical algorithm should be possible using an AMP approach similar to [36, 49], and we leave this for future work.

Another aspect in which our result improves on [101] is in the bound on D, namely, in the above scaling regime of [48] we show that degree $D = \Omega(1/p_0)$ is necessary below the sharp threshold. We expect this is essentially optimal: while it has not appeared in the literature, an algorithm of runtime $\exp(\tilde{O}(1/p_0))$ can be obtained for $(\rho n)^{-1} \ll p_0 \ll 1$ by a simple adaptation of the spiked Wigner results in [40]. The algorithm is based on a brute-force search over size-ℓ subsets of vertices, with $\ell = \tilde{\Theta}(1/p_0)$.

We will discuss a few other popular scaling regimes. While not a focus of our work, a notable special case of the planted dense subgraph model is the well-known *planted clique* problem, where $p_1 = 1$ and $p_0 = 1/2$. Here, the testing and estimation problems coincide in difficulty, both transitioning from hard to easy at $\rho \approx 1/\sqrt{n}$. More precisely, there is an AMP algorithm for exact recovery above $\rho = 1/\sqrt{en}$ [36], but there are in fact better poly-time algorithms that reach $\rho = \epsilon/\sqrt{n}$ for an arbitrary constant $\epsilon > 0$ [4]; so there is not actually a sharp threshold here but rather a smooth tradeoff between runtime and clique size. Our result, as it should, does not indicate a sharp threshold here, as it becomes limited to constant degree in the planted clique regime. Stronger low-degree lower bounds for planted clique estimation were known previously, reaching the "correct" degree $D \approx \log^2 n$ [101], and the analogous result for testing was known even earlier (see [52]). However, we note that planting a clique with $1/\sqrt{n} \ll \rho \ll 1$ in a dense graph with $p_0 = 1 - c/(n\rho^2)$, or equivalently planting an independent set of size $n\rho$ with $p_0 = c/(n\rho^2)$, exhibits a sharp threshold at $c = 1/e$ for estimation by low-degree polynomials, as we establish in Theorem 2.4. To our knowledge, this sharp phase transition has not appeared in the previous literature.

Another interesting regime of planted dense subgraph that we do not focus on is the *log-density* regime where $\rho \ll 1/\sqrt{n}$ and $p_1 = n^{-\alpha}$, $p_0 = n^{-\beta}$ for constants $0 < \alpha < \beta$ [16]. We do not expect a sharp threshold here, and the low-degree limits for testing have been characterized using a somewhat delicate conditioning argument [37]. The known algorithms for estimation match the lower bounds for testing (which is an easier problem), so we expect the limits of estimation and testing to coincide. Still, proving optimal bounds on $\mathsf{MMSE}_{\le D}$ remains a difficult technical challenge, which we leave for future work.

2.4 Spiked Wigner Model

Definition 2.5 (Spiked Wigner Model). *For a parameter $\lambda \ge 0$, let*

$$X = \sqrt{\frac{\lambda}{n}} UU^\top$$

where $U \in \mathbb{R}^{n\times m}$ with entries i.i.d. from some prior π. Observe

$$Y_{ij} = X_{ij} + Z_{ij} \qquad \text{for} \qquad 1 \le i \le j \le n,$$

where Z_{ij} are i.i.d. $\mathcal{N}(0,1)$. The goal is to estimate $x := X_{1,2}$.

The estimand $X_{1,2}$ is representative of any off-diagonal entry of X, by symmetry, while the diagonal entries have negligible contribution to the matrix mean squared error $\|\hat{X} - X\|_F^2$. It is common to put $\mathcal{N}(0,2)$ instead of $\mathcal{N}(0,1)$ on the diagonal of Z, which would only make the problem harder, so our lower bound still holds (see [101, Claim A.2]). The upper bound below also remains valid because it does not use the diagonal entries.

Theorem 2.6. *Consider the spiked Wigner model and assume the prior π satisfies $\mathbb{E}[\pi] = 0$ and $\mathbb{E}[\pi^2] = 1$.*

(a) *(Lower Bound) Suppose that for some $c, \nu > 0$, $\mathbb{E}|\pi|^k \le (ck)^{\nu k}$ for any integer $k \ge 1$. There exist constants $\delta, C > 0$ depending only on c, ν such that if $D \le n^\delta$ then*

$$\mathrm{Corr}_{\le D} \le C\sqrt{\frac{m}{n}\sum_{d=1}^{D}\lambda^d}.$$

Consequently,

- *if $\lambda < 1$ is fixed and $D \le n^\delta$, then $\mathrm{Corr}_{\le D} \le C\sqrt{\frac{m\lambda}{n(1-\lambda)}}$,*
- *if $\lambda = 1$ and $D \le n^\delta$, then $\mathrm{Corr}_{\le D} \le C\sqrt{Dm/n}$, and*
- *if $\lambda = O(1)$, $m \le n^{1-\Omega(1)}$, and $D = o(\log n)$, then $\mathrm{Corr}_{\le D} = o(1)$.*

(b) *(Upper Bound) Suppose that* $K \equiv \mathbb{E}[\pi^4] < \infty$ *and* $\lambda \geq 1 + \eta$, *where* K *and* $\eta \in (0, 1)$ *are constants (not depending on* n*). Then there exist a constant* $c \equiv c(K) > 0$ *and a universal constant* $C > 0$ *such that if* $m = m_n$ *satisfies* $m = o(n)$, *then for* $D \geq \frac{C}{\eta} \log(n/m)$ *and large enough* n,

$$\mathrm{Corr}_{\leq D} \geq c\eta.$$

Discussion. Most prior work pertains to the case $m = O(1)$ (and often $m = 1$), and we focus on this case for now. We also assume that λ and the spike prior π are fixed, i.e., not depending on n. The spiked Wigner model was first studied in random matrix theory, where a sharp phase transition at $\lambda = 1$ was discovered in the eigenvalues and eigenvectors of Y [15, 26, 44, 76, 94]. This is known as the "BBP" transition, by analogy to the similar phase transition discovered by Baik, Ben Arous, and Péché in the related *spiked Wishart* model [8]. These results immediately imply an algorithm, based on the m leading eigenvectors, for weak recovery when $\lambda > 1$. Here, weak recovery means $\langle \hat{X}, X\rangle/(\|\hat{X}\|_F \|X\|_F)$ converges to some positive constant depending on λ, where the estimator $\hat{X}$ is obtained from Y by truncating the eigendecomposition to the m leading terms. When $\lambda \leq 1$, the m leading eigenvectors fail to achieve weak recovery.

Later, a number of works asked the statistical question of whether *any* method can estimate the signal below the BBP threshold [38, 42, 45, 69, 70] (see [83] for a survey), or even detect its presence [11, 43, 85, 95]. Statistically speaking, the answer turns out to be "yes" for some spike priors π, including sufficiently sparse priors but not $\mathcal{N}(0, 1)$ or $\mathrm{Unif}(\{\pm 1\})$. However, regardless of the prior (as long as it does not depend on n), no *poly-time* algorithm is known to achieve weak recovery below $\lambda = 1$. It has been conjectured that this hardness is inherent, on the basis of statistical physics methods, namely AMP and the associated free energy barriers [70] (see Conjecture 10 in [69]). Low-degree lower bounds show hardness for hypothesis testing against the "null" model ($\lambda = 0$) whenever $\lambda < 1$ [9, 68], providing indirect evidence that the seemingly more difficult estimation task should also be hard. We note that our focus is on weak recovery because strong recovery, meaning $\langle \hat{X}, X\rangle/(\|\hat{X}\|_F \|X\|_F) \to 1$, is information-theoretically impossible; see [69, 96].

Our lower bound shows low-degree hardness of weak recovery when $\lambda \leq 1$, notably covering the critical case $\lambda = 1$. This holds for all priors π satisfying a mild moment condition (which, for instance, covers all priors with bounded support). We rule out polynomials of degree n^{δ} for a particular constant $\delta > 0$ depending only on c, v. This resolves the conjecture mentioned above, within the low-degree framework. The testing results [9] suggest that this hardness can be extended to any $\delta < 1$, and this remains an open problem.

Another consequence of our lower bound is that logarithmic degree is required, even in the easy regime. Combined with the upper bound, we know that when $\lambda > 1$ is fixed and $m \leq n^{1-\Omega(1)}$, the degree complexity of weak recovery is exactly on the order $D = \Theta(\log n)$.

Our results also extend to the case of growing m (a.k.a. symmetric matrix factorization), which has received recent attention [12, 23, 57, 77, 96, 102]. We show that the above phenomenon persists, that is, the low-degree threshold for weak recovery remains at $\lambda = 1$ as long as $m \ll n$. To our knowledge, no prior work has explored computational hardness in the growing-m regime, so our results uncover a new phase transition that was not previously substantiated. As m grows, the testing and estimation thresholds separate, so this new phenomenon could not have been uncovered by studying the testing problem.

Our low-degree upper bound is an extension of [54], which handles the $m = 1$ case using self-avoiding walks. We expect that our estimator can be made into a poly-time algorithm using the "color coding" trick [5], as in [54]. Alternatively, poly-time weak recovery based on eigenvectors might be deduced from the spectral analysis by [57]. More sophisticated poly-time algorithms that aim to optimize the precise mean squared error are discussed in [23, 96, 102].

Our lower bound becomes vacuous once $m = \Omega(n)$, and for good reason: in this regime, the degree-1 estimator $f(Y) = Y_{1,2}$ achieves correlation $\sqrt{\frac{m\lambda/n}{m\lambda/n+1}}$, which gives weak recovery as long as λ is of constant order.

2.5 Stochastic Block Model

Definition 2.7 (Stochastic Block Model). *Let* $q \geq 2$ *be the number of communities. Let* $\pi = (\pi_k)_{k\in[q]} \in \mathbb{R}^q_{>0}$ *be a vector whose entries sum to 1, representing a probability distribution over* $[q]$. *Let* $Q \in \mathbb{R}^{q\times q}_{>0}$ *be a symmetric matrix with positive entries. Observe* $Y \in \{0,1\}^{\binom{[n]}{2}}$, *the adjacency matrix for a graph, generated as follows. The community labels* $\sigma^\star = (\sigma_i^\star)_{1\leq i\leq n} \in [q]^n$ *for the* n *vertices are drawn as* $\sigma_i^\star \overset{iid}{\sim} \pi$. *Conditional on* $\sigma^\star$, $Y_{ij} \sim \mathrm{Bernoulli}(Q_{\sigma_i^\star,\sigma_j^\star})$ *is sampled independently for each* $i < j$. *The goal is to estimate* $x := Q_{\sigma_1^\star,\sigma_2^\star} - \mathbb{E}[Q_{\sigma_1^\star,\sigma_2^\star}]$.

Equivalently, the probability of having an edge between i and j, conditioned on $\sigma^\star$, is $Q_{\sigma_i^\star,\sigma_j^\star}/n$. We will mostly focus on the sparse regime with a fixed number of communities, i.e. regard q, π, Q as fixed and consider the limit $n \to \infty$, although our proof for the low-degree lower bound applies more generally, e.g. when $q \ll n^{1/8}$ in the *symmetric* SBM where $\pi_k \equiv 1/q$ and the diagonal (resp. off-diagonal) entries of Q are the same (see Remark 2.9). In the sparse regime, we cannot hope to achieve strong recovery, even information-theoretically, due to the presence of isolated vertices, and thus we focus on weak recovery.

We are interested in recovering $(Q_{\sigma_i^\star,\sigma_j^\star})_{1\leq i<j\leq n}$, the membership matrix. We have chosen x so that $\mathrm{Corr}_{\leq D}$ is directly related to the MMSE

$$\mathrm{MMSE}_{\leq D} = \inf_{f\in\mathbb{R}[Y]_{\leq D}} \mathbb{E}[(f(Y) - Q_{\sigma_1^\star,\sigma_2^\star})^2], \tag{5}$$

as described in Section 2.1. Here, note that the centering term $\mathbb{E}[Q_{\sigma_1^\star,\sigma_2^\star}]$ that appears in x has been omitted from (5) because it does not affect $\mathrm{MMSE}_{\leq D}$. Note that in the *symmetric* SBM, estimating $Q_{\sigma_1^\star,\sigma_2^\star}$ is equivalent to estimating $\mathbb{1}_{\sigma_1^\star=\sigma_2^\star}$. In fact, our lower bound also rules out estimation of $(\mathbb{1}_{\sigma_1^\star=k} - \pi_k)(\mathbb{1}_{\sigma_2^\star=\ell} - \pi_\ell)$ for *any* $k, \ell \in [q]$; see the full version [104].

Define $d := \mathbb{E}[Q_{\sigma_1^\star,\sigma_2^\star}] > 0$, which is (asymptotically) the average degree of the observed graph. Assume that the average degree of each vertex is the same regardless of its community label:

$$\sum_{\ell=1}^{q} Q_{k,\ell}\pi_\ell = d \qquad \forall k \in [q]. \tag{6}$$

Violation of condition (6) allows for weak recovery, regardless of q, π, Q as long as these are fixed as $n \to \infty$, via simple degree counting (see e.g. [92, Proposition 4.1]). Thus, we may assume the condition (6) without loss of generality (see also [3, 10, 54]). Define the stochastic matrix

$$T := \frac{1}{d}\mathrm{diag}(\pi)Q \tag{7}$$

and let $1 = \lambda_1(T) \geq |\lambda_2(T)| \geq \ldots \geq |\lambda_q(T)|$ denote the eigenvalues of T in decreasing order of magnitude. A central role will be played by the parameter

$$\lambda := |\lambda_2(T)|.$$

Theorem 2.8. *Consider the stochastic block model with parameters q, π, Q such that Eq. (6) holds.*

(a) *(Lower Bound) There exist constants $\delta, C > 0$ depending only on q, π, Q such that if $D \leq n^\delta$ then*

$$\mathsf{Corr}_{\leq D} \leq \sqrt{\frac{C}{n}\sum_{t=1}^{D}(d\lambda^2)^t}.$$

Consequently, if $d\lambda^2 \leq 1$ and $D \leq n^\delta$, then $\mathsf{Corr}_{\leq D} = o(1)$.

(b) *(Upper Bound [3, 54]) If q, π, Q are fixed with $d\lambda^2 > 1$, then for large enough n we have $\mathsf{Corr}_{\leq C\log n} \geq \eta$ for some constants $C \equiv C(q, \pi, Q) > 0$ and $\eta \equiv \eta(q, \pi, Q) > 0$.*

Remark 2.9. *Although Theorem 2.8 is stated only for the sparse SBM with a constant number of communities, our proof reveals that if Condition (6) is satisfied, the low-degree lower bound remains valid when $\|Q\|_\infty \leq n^{1-\epsilon}$ and $\left(\frac{q}{\pi_{\min}}\right)^4 \frac{d}{Q_{\min}} \leq n^{1-\epsilon}$ for a constant $\epsilon > 0$ where $\pi_{\min} := \min_{k\in[q]} \pi_k$ and $Q_{\min} := \min_{k,\ell\in[q]} Q_{k,\ell}$. Namely, if $d\lambda^2 \leq 1 - \eta$ for a constant $\eta > 0$ then $\mathsf{Corr}_{\leq n^\delta} = o(1)$ holds for some $\delta = \delta(\epsilon, \eta) > 0$.*

Note that in Definition 2.7, we assumed $\pi \in \mathbb{R}^q_{>0}$ and $Q \in \mathbb{R}^{q\times q}_{>0}$. We may assume the former without loss of generality since otherwise we can remove the empty communities. On the other hand, the latter condition that the connectivity matrix Q has positive entries is likely a proof artifact, and we leave the question of resolving the low-degree hardness when some entries of Q are 0 as an open problem.

The upper bound (b) for $d\lambda^2 > 1$ follows from the upper bounds derived in the works [3, 54], which we include for completeness. The main contribution of this work is the lower bound (a) for $d\lambda^2 \leq 1$.

Discussion. The stochastic block model (SBM) is a special case of inhomogeneous random graphs [18] that has been extensively studied as a model for communities in statistics and social sciences, see e.g. [17, 50, 100, 103], and for analyzing clustering algorithms in computer science, see e.g. [30, 32, 41, 59, 82]. See [1, 87] for survey articles.

Our focus is on the sparse regime where the edge probabilities are proportional to $1/n$, the number of communities q is held constant, and the objective is weak recovery. In the sparse regime, the landmark work [34] first predicted a sharp computational phase transition at the so-called *Kesten–Stigum (KS) threshold* $d\lambda^2 = 1$ based on a heuristic analysis of the *belief propagation (BP)* algorithm. First identified by Kesten and Stigum [61] in the context of multi-type branching processes, the KS threshold $d\lambda^2 = 1$ has since played an important role in other areas, including phylogenetic reconstructions [33, 90, 99].

To simplify the discussion, we focus on the symmetric SBM where the diagonal (resp. off-diagonal) entries of Q are the same. A sequence of works has established that poly-time algorithms achieve weak recovery above the KS threshold $d\lambda^2 > 1$ for $q = 2$ [19, 81, 89] and for general $q \geq 3$ [3]. Below the KS threshold $d\lambda^2 \leq 1$, weak recovery is information-theoretically impossible for $q = 2$ [88] or $q = 3, 4$ and d large enough [91]. For $q \geq 5$, a statistical-computational gap appears: no known poly-time algorithm succeeds below the KS threshold, yet it is information-theoretically possible to do so [2, 10, 31]. However, regardless of the prior π or the probability matrix Q, it has been conjectured by [3] based on the prediction by [34] that no poly-time algorithm can achieve weak recovery below $d\lambda^2 < 1$. Low-degree lower bounds for hypothesis testing support the presumed hardness below the KS threshold [9, 54], and the concurrent work [39] makes this precise with a detection-to-recovery reduction. However, proving bounds on $\mathsf{MMSE}_{\leq D}$ remained a difficult technical challenge that was posed as an open question by Hopkins and Steurer in one of the first papers on the low-degree framework [54]. We resolve this, proving that degree-n^δ polynomials fail to achieve weak recovery below the sharp KS threshold $d\lambda^2 \leq 1$. We show this for a particular constant $\delta > 0$ but we expect the result to hold for any constant $\delta < 1$, as suggested by [39], and we leave this as an open problem.

Prior to our work, coarser low-degree lower bounds for estimation were obtained by [73], who additionally studied the case of a growing number of communities and the related task of graphon estimation. Our proof of Theorem 2.8(a) reveals that, in the symmetric SBM where the average degree d is of order constant, community detection is low-degree hard below the KS threshold as long as $q \ll n^{1/8}$ (see Remark 2.9). The work [29] extends our techniques to show that the KS bound remains the threshold for low-degree recovery as long as $q \ll n^{1/2}$, and this is tight in the sense that the KS threshold can be surpassed in polynomial time when $q \gg n^{1/2}$. The concurrent work [39] also argues hardness of weak recovery below the KS bound for slowly growing $q = n^{o(1)}$. Since their approach exploits hardness of testing, we expect it may not work for $q = n^{\Omega(1)}$, due to the large detection-recovery gap that appears here [29].

Acknowledgments

We are grateful to Jonathan Niles-Weed for discussions that helped inspire the framework presented in Section 1.2.1. We are grateful to Elchanan Mossel for numerous discussions and valuable insights. We thank Lenka Zdeborová for introducing us to the spiked Wigner model with growing rank (also known as symmetric matrix factorization) and providing the reference [77].

Funding: ASW was partially supported by an Alfred P. Sloan Research Fellowship and NSF CAREER Award CCF-2338091.

References

[1] Emmanuel Abbe. 2018. Community detection and stochastic block models: recent developments. *Journal of Machine Learning Research* 18, 177 (2018), 1–86.

[2] Emmanuel Abbe and Colin Sandon. 2016. Crossing the KS threshold in the stochastic block model with information theory. In *2016 IEEE International Symposium on Information Theory (ISIT)*. IEEE, 840–844.

[3] Emmanuel Abbe and Colin Sandon. 2018. Proof of the achievability conjectures for the general stochastic block model. *Communications on Pure and Applied Mathematics* 71, 7 (2018), 1334–1406.

[4] Noga Alon, Michael Krivelevich, and Benny Sudakov. 1998. Finding a large hidden clique in a random graph. *Random Structures & Algorithms* 13, 3-4 (1998), 457–466.

[5] Noga Alon, Raphael Yuster, and Uri Zwick. 1995. Color-coding. *Journal of the ACM (JACM)* 42, 4 (1995), 844–856.

[6] Brendan PW Ames. 2013. Robust convex relaxation for the planted clique and densest k-subgraph problems. *arXiv preprint arXiv:1305.4891* (2013).

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

[8] Jinho Baik, Gérard Ben Arous, and Sandrine Péché. 2005. Phase Transition of the Largest Eigenvalue for Nonnull Complex Sample Covariance Matrices. *Annals of Probability* (2005), 1643–1697.

[9] Afonso S Bandeira, Jess Banks, Dmitriy Kunisky, Christopher Moore, and Alexander S Wein. 2021. Spectral planting and the hardness of refuting cuts, colorability, and communities in random graphs. In *Conference on Learning Theory*. PMLR, 410–473.

[10] Jess Banks, Cristopher Moore, Joe Neeman, and Praneeth Netrapalli. 2016. Information-theoretic thresholds for community detection in sparse networks. In *Conference on Learning Theory*. PMLR, 383–416.

[11] Jess Banks, Cristopher Moore, Roman Vershynin, Nicolas Verzelen, and Jiaming Xu. 2018. Information-theoretic bounds and phase transitions in clustering, sparse PCA, and submatrix localization. *IEEE Transactions on Information Theory* 64, 7 (2018), 4872–4894.

[12] Jean Barbier, Justin Ko, and Anas A Rahman. 2024. A multiscale cavity method for sublinear-rank symmetric matrix factorization. *arXiv preprint arXiv:2403.07189* (2024).

[13] Mohsen Bayati, Marc Lelarge, and Andrea Montanari. 2015. Universality in polytope phase transitions and message passing algorithms. *Annals of Applied Probability: an official journal of the Institute of Mathematical Statistics* 25, 2 (2015), 753–822.

[14] Gérard Ben Arous, Alexander S Wein, and Ilias Zadik. 2023. Free Energy Wells and Overlap Gap Property in Sparse PCA. *Communications on Pure and Applied Mathematics* 76, 10 (2023), 2410–2473.

[15] Florent Benaych-Georges and Raj Rao Nadakuditi. 2011. The eigenvalues and eigenvectors of finite, low rank perturbations of large random matrices. *Advances in Mathematics* 227, 1 (2011), 494–521.

[16] Aditya Bhaskara, Moses Charikar, Eden Chlamtac, Uriel Feige, and Aravindan Vijayaraghavan. 2010. Detecting high log-densities: an $O(n^{1/4})$ approximation for densest k-subgraph. In *Proceedings of the forty-second ACM symposium on Theory of computing*. 201–210.

[17] P.J. Bickel and A. Chen. 2009. A nonparametric view of network models and Newman-Girvan and other modularities. *Proceedings of the National Academy of Sciences* 106, 50 (2009), 21068–21073.

[18] Béla Bollobás, Svante Janson, and Oliver Riordan. 2007. The phase transition in inhomogeneous random graphs. *Random Structures & Algorithms* 31, 1 (2007), 3–122.

[19] Charles Bordenave, Marc Lelarge, and Laurent Massoulié. 2015. Non-backtracking spectrum of random graphs: community detection and non-regular ramanujan graphs. In *2015 IEEE 56th Annual Symposium on Foundations of Computer Science*. IEEE, 1347–1357.

[20] Matthew Brennan and Guy Bresler. 2020. Reducibility and statistical-computational gaps from secret leakage. In *Conference on Learning Theory*. PMLR, 648–847.

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

[22] Matthew S Brennan, Guy Bresler, Sam Hopkins, Jerry Li, and Tselil Schramm. 2021. Statistical Query Algorithms and Low Degree Tests Are Almost Equivalent. In *Conference on Learning Theory*. PMLR, 774–774.

[23] Joël Bun, Romain Allez, Jean-Philippe Bouchaud, and Marc Potters. 2016. Rotational invariant estimator for general noisy matrices. *IEEE Transactions on Information Theory* 62, 12 (2016), 7475–7490.

[24] C Butucea and YI Ingster. 2013. Detection of a sparse submatrix of a high-dimensional noisy matrix. *Bernoulli: a Journal of Mathematical Statistics and Probability* 19, 5 B (2013), 2652–2688.

[25] Cristina Butucea, Yuri I Ingster, and Irina A Suslina. 2015. Sharp variable selection of a sparse submatrix in a high-dimensional noisy matrix. *ESAIM: Probability and Statistics* 19 (2015), 115–134.

[26] Mireille Capitaine, Catherine Donati-Martin, and Delphine Féral. 2009. The largest eigenvalues of finite rank deformation of large Wigner matrices: convergence and nonuniversality of the fluctuations. *The Annals of Probability* 37, 1 (2009), 1–47.

[27] Houssam El Cheairi and David Gamarnik. 2024. Algorithmic Universality, Low-Degree Polynomials, and Max-Cut in Sparse Random Graphs. *arXiv preprint arXiv:2412.18014* (2024).

[28] Yudong Chen and Jiaming Xu. 2016. Statistical-computational tradeoffs in planted problems and submatrix localization with a growing number of clusters and submatrices. *Journal of Machine Learning Research* 17, 27 (2016), 1–57.

[29] Byron Chin, Elchanan Mossel, Youngtak Sohn, and Alexander S Wein. 2025. Stochastic block models with many communities and the Kesten–Stigum bound. *arXiv preprint arXiv:2503.03047* (2025).

[30] A. Coja-Oghlan. 2010. Graph partitioning via adaptive spectral techniques. *Combinatorics, Probability and Computing* 19, 02 (2010), 227–284.

[31] Amin Coja-Oghlan, Florent Krzakala, Will Perkins, and Lenka Zdeborová. 2018. Information-theoretic thresholds from the cavity method. *Advances in Mathematics* 333 (2018), 694–795.

[32] A. Condon and R.M. Karp. 2001. Algorithms for graph partitioning on the planted partition model. *Random Structures and Algorithms* 18, 2 (2001), 116–140.

[33] C. Daskalakis, E. Mossel, and S. Roch. 2011. Evolutionary trees and the Ising model on the Bethe lattice: a proof of Steel's conjecture. *Probability Theory and Related Fields* 149, 1-2 (2011), 149—189. http://www.springerlink.com/content/k76630287587433x/

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

[35] Yash Deshpande and Andrea Montanari. 2014. Information-theoretically optimal sparse PCA. In *2014 IEEE International Symposium on Information Theory*. IEEE, 2197–2201.

[36] Yash Deshpande and Andrea Montanari. 2015. Finding hidden cliques of size $\sqrt{N/e}$ in nearly linear time. *Foundations of Computational Mathematics* 15 (2015), 1069–1128.

[37] Abhishek Dhawan, Cheng Mao, and Alexander S Wein. 2023. Detection of dense subhypergraphs by low-degree polynomials. *arXiv preprint arXiv:2304.08135* (2023).

[38] Mohamad Dia, Nicolas Macris, Florent Krzakala, Thibault Lesieur, Lenka Zdeborová, et al. 2016. Mutual information for symmetric rank-one matrix estimation: A proof of the replica formula. *Advances in Neural Information Processing Systems* 29 (2016).

[39] Jingqiu Ding, Yiding Hua, Lucas Slot, and David Steurer. 2025. Low degree conjecture implies sharp computational thresholds in stochastic block model. *arXiv preprint arXiv:2502.15024* (2025).

[40] Yunzi Ding, Dmitriy Kunisky, Alexander S Wein, and Afonso S Bandeira. 2023. Subexponential-time algorithms for sparse PCA. *Foundations of Computational Mathematics* (2023), 1–50.

[41] M.E. Dyer and A.M. Frieze. 1989. The solution of some random NP-hard problems in polynomial expected time. *Journal of Algorithms* 10, 4 (1989), 451–489.

[42] Ahmed El Alaoui and Florent Krzakala. 2018. Estimation in the spiked Wigner model: a short proof of the replica formula. In *International Symposium on Information Theory (ISIT)*. IEEE, 1874–1878.

[43] Ahmed El Alaoui, Florent Krzakala, and Michael I Jordan. 2020. Fundamental limits of detection in the spiked Wigner model. *The Annals of Statistics* 48, 2 (2020), 863–885.

[44] Delphine Féral and Sandrine Péché. 2007. The largest eigenvalue of rank one deformation of large Wigner matrices. *Communications in mathematical physics* 272 (2007), 185–228.

[45] Alyson K Fletcher and Sundeep Rangan. 2018. Iterative reconstruction of rank-one matrices in noise. *Information and Inference: A Journal of the IMA* 7, 3 (2018), 531–562.

[46] David Gamarnik. 2021. The overlap gap property: A topological barrier to optimizing over random structures. *Proceedings of the National Academy of Sciences* 118, 41 (2021), e2108492118.

[47] David Gamarnik, Aukosh Jagannath, and Subhabrata Sen. 2021. The overlap gap property in principal submatrix recovery. *Probability Theory and Related Fields* 181, 4 (2021), 757–814.

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

[49] Bruce Hajek, Yihong Wu, and Jiaming Xu. 2018. Submatrix localization via message passing. *Journal of Machine Learning Research* 18, 186 (2018), 1–52.

[50] P.W. Holland, K.B. Laskey, and S. Leinhardt. 1983. Stochastic blockmodels: First steps. *Social Networks* 5, 2 (1983), 109 – 137. https://doi.org/10.1016/0378-8733(83)90021-7

[51] Guy Holtzman, Adam Soffer, and Dan Vilenchik. 2020. A greedy anytime algorithm for sparse PCA. In *Conference on Learning Theory*. PMLR, 1939–1956.

[52] Samuel Hopkins. 2018. *Statistical Inference and the Sum of Squares Method*. Ph. D. Dissertation. Cornell University.

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

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

[55] Han Huang and Elchanan Mossel. 2024. Low Degree Hardness for Broadcasting on Trees. *arXiv preprint arXiv:2402.13359* (2024).

[56] Han Huang and Elchanan Mossel. 2025. Optimal Low degree hardness for Broadcasting on Trees. *arXiv preprint arXiv:2502.04861* (2025).

[57] Jiaoyang Huang. 2018. Mesoscopic perturbations of large random matrices. *Random Matrices: Theory and Applications* 7, 02 (2018), 1850004.

[58] Misha Ivkov and Tselil Schramm. 2024. Semidefinite programs simulate approximate message passing robustly. In *Proceedings of the 56th Annual ACM Symposium on Theory of Computing*. 348–357.

[59] M. Jerrum and G.B. Sorkin. 1998. The Metropolis algorithm for graph bisection. *Discrete Applied Mathematics* 82, 1-3 (1998), 155–175.

[60] Chris Jones and Lucas Pesenti. 2024. Fourier Analysis of Iterative Algorithms. *arXiv preprint arXiv:2404.07881* (2024).

[61] H. Kesten and B. P. Stigum. 1966. Additional limit theorems for indecomposable multidimensional Galton-Watson processes. *Ann. Math. Statist.* 37 (1966), 1463–1481.

[62] Frederic Koehler and Elchanan Mossel. 2021. Reconstruction on Trees and Low-Degree Polynomials. *arXiv preprint arXiv:2109.06915* (2021).

[63] Mladen Kolar, Sivaraman Balakrishnan, Alessandro Rinaldo, and Aarti Singh. 2011. Minimax localization of structural information in large noisy matrices. *Advances in Neural Information Processing Systems* 24 (2011).

[64] Florent Krzakala, Cristopher Moore, Elchanan Mossel, Joe Neeman, Allan Sly, Lenka Zdeborová, and Pan Zhang. 2013. Spectral redemption in clustering sparse networks. *Proceedings of the National Academy of Sciences* 110, 52 (2013), 20935–20940.

[65] Florent Krzakala, Jiaming Xu, and Lenka Zdeborová. 2016. Mutual information in rank-one matrix estimation. In *2016 IEEE Information Theory Workshop (ITW)*. IEEE, 71–75.

[66] Florent Krzakala and Lenka Zdeborová. 2024. Statistical physics methods in optimization and machine learning. *Lecture Notes* (2024).

[67] Dmitriy Kunisky, Cristopher Moore, and Alexander S Wein. 2024. Tensor cumulants for statistical inference on invariant distributions. *arXiv preprint arXiv:2404.18735* (2024).

[68] Dmitriy Kunisky, Alexander S Wein, and Afonso S Bandeira. 2019. Notes on computational hardness of hypothesis testing: Predictions using the low-degree likelihood ratio. In *ISAAC Congress (International Society for Analysis, its Applications and Computation)*. Springer, 1–50.

[69] Marc Lelarge and Léo Miolane. 2019. Fundamental limits of symmetric low-rank matrix estimation. *Probability Theory and Related Fields* 173 (2019), 859–929.

[70] Thibault Lesieur, Florent Krzakala, and Lenka Zdeborová. 2015. MMSE of probabilistic low-rank matrix estimation: Universality with respect to the output channel. In *2015 53rd Annual Allerton Conference on Communication, Control, and Computing (Allerton)*. IEEE, 680–687.

[71] Thibault Lesieur, Florent Krzakala, and Lenka Zdeborová. 2015. Phase transitions in sparse PCA. In *International Symposium on Information Theory (ISIT)*. IEEE, 1635–1639.

[72] Zhangsong Li. 2025. Algorithmic contiguity from low-degree conjecture and applications in correlated random graphs. *arXiv preprint arXiv:2502.09832* (2025).

[73] Yuetian Luo and Chao Gao. 2023. Computational Lower Bounds for Graphon Estimation via Low-degree Polynomials. *arXiv preprint arXiv:2308.15728* (2023).

[74] Yuetian Luo and Anru R Zhang. 2022. Tensor clustering with planted structures: Statistical optimality and computational limits. *The Annals of Statistics* 50, 1 (2022), 584–613.

[75] Zongming Ma and Yihong Wu. 2015. COMPUTATIONAL BARRIERS IN MINIMAX SUBMATRIX DETECTION. *The Annals of Statistics* (2015), 1089–1116.

[76] Mylène Maïda. 2007. Large deviations for the largest eigenvalue of rank one deformations of Gaussian ensembles. *Electronic Journal of Probability* 12 (2007), 1131–1150.

[77] Antoine Maillard, Florent Krzakala, Marc Mézard, and Lenka Zdeborová. 2022. Perturbative construction of mean-field equations in extensive-rank matrix factorization and denoising. *Journal of Statistical Mechanics: Theory and Experiment* 2022, 8 (2022), 083301.

[78] Cheng Mao, Alexander S Wein, and Shenduo Zhang. 2023. Detection-recovery gap for planted dense cycles. In *The Thirty Sixth Annual Conference on Learning Theory*. PMLR, 2440–2481.

[79] Cheng Mao, Yihong Wu, Jiaming Xu, and Sophie H Yu. 2021. Testing network correlation efficiently via counting trees. *arXiv preprint arXiv:2110.11816* (2021).

[80] Cheng Mao, Yihong Wu, Jiaming Xu, and Sophie H Yu. 2023. Random graph matching at Otter's threshold via counting chandeliers. In *Proceedings of the 55th Annual ACM Symposium on Theory of Computing*. 1345–1356.

[81] Laurent Massoulié. 2014. Community detection thresholds and the weak Ramanujan property. In *Proceedings of the forty-sixth annual ACM symposium on Theory of computing*. 694–703.

[82] Frank McSherry. 2001. Spectral partitioning of random graphs. In *Foundations of Computer Science, 2001. Proceedings. 42nd IEEE Symposium on*. IEEE, 529–537.

[83] Léo Miolane. 2018. Phase transitions in spiked matrix estimation: information-theoretic analysis. *arXiv preprint arXiv:1806.04343* (2018).

[84] Andrea Montanari. 2015. Finding one community in a sparse graph. *Journal of Statistical Physics* 161 (2015), 273–299.

[85] Andrea Montanari, Daniel Reichman, and Ofer Zeitouni. 2015. On the limitation of spectral methods: From the gaussian hidden clique problem to rank-one perturbations of gaussian tensors. *Advances in Neural Information Processing Systems* 28 (2015).

[86] Andrea Montanari and Alexander S Wein. 2024. Equivalence of approximate message passing and low-degree polynomials in rank-one matrix estimation. *Probability Theory and Related Fields* (2024), 1–53.

[87] Cristopher Moore. 2017. The computer science and physics of community detection: Landscapes, phase transitions, and hardness. *arXiv preprint arXiv:1702.00467* (2017).

[88] Elchanan Mossel, Joe Neeman, and Allan Sly. 2012. Stochastic block models and reconstruction. *arXiv preprint arXiv:1202.1499* (2012).

[89] Elchanan Mossel, Joe Neeman, and Allan Sly. 2018. A proof of the block model threshold conjecture. *Combinatorica* 38, 3 (2018), 665–708.

[90] E. Mossel, S. Roch, and A. Sly. 2011. On the inference of large phylogenies with long branches: How long is too long? *Bull. Math. Bio.* 73, 7 (2011), 1627–1644.

[91] Elchanan Mossel, Allan Sly, and Youngtak Sohn. 2023. Exact phase transitions for stochastic block models and reconstruction on trees. In *Proceedings of the 55th Annual ACM Symposium on Theory of Computing*. 96–102.

[92] Elchanan Mossel, Allan Sly, and Youngtak Sohn. 2024. Weak recovery, hypothesis testing, and mutual information in stochastic block models and planted factor graphs. *arXiv preprint arXiv:2406.15957* (2024).

[93] Jonathan Niles-Weed and Ilias Zadik. 2023. It was "all" for "nothing": sharp phase transitions for noiseless discrete channels. *IEEE Transactions on Information Theory* 69, 8 (2023), 5188–5202.

[94] Sandrine Péché. 2006. The largest eigenvalue of small rank perturbations of Hermitian random matrices. *Probability Theory and Related Fields* 134 (2006), 127–173.

[95] Amelia Perry, Alexander S Wein, Afonso S Bandeira, and Ankur Moitra. 2018. Optimality and sub-optimality of PCA I: Spiked random matrix models. *The Annals of Statistics* 46, 5 (2018), 2416–2451.

[96] Farzad Pourkamali, Jean Barbier, and Nicolas Macris. 2024. Matrix Inference in Growing Rank Regimes. *IEEE Transactions on Information Theory* 70, 11 (2024), 8133–8163. https://doi.org/10.1109/TIT.2024.3422263

[97] Prasad Raghavendra, Tselil Schramm, and David Steurer. 2018. High dimensional estimation via sum-of-squares proofs. In *Proceedings of the International Congress of Mathematicians: Rio de Janeiro 2018*. World Scientific, 3389–3423.

[98] Emile Richard and Andrea Montanari. 2014. A statistical model for tensor PCA. *Advances in neural information processing systems* 27 (2014).

[99] Sebastien Roch and Allan Sly. 2017. Phase transition in the sample complexity of likelihood-based phylogeny inference. *Probability Theory and Related Fields* 169, 1 (2017), 3–62.

[100] Karl Rohe, Sourav Chatterjee, and Bin Yu. 2011. Spectral clustering and the high-dimensional stochastic blockmodel. *The Annals of Statistics* 39, 4 (2011), 1878–1915.

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

[102] Guilhem Semerjian. 2024. Matrix denoising: Bayes-optimal estimators via low-degree polynomials. *Journal of Statistical Physics* 191, 10 (2024), 139.

[103] T.A.B. Snijders and K. Nowicki. 1997. Estimation and prediction for stochastic blockmodels for graphs with latent block structure. *Journal of Classification* 14, 1 (1997), 75–100.

[104] Youngtak Sohn and Alexander S Wein. 2025. Sharp phase transitions in estimation with low-degree polynomials. *arXiv preprint arXiv:2502.14407* (2025).

[105] Gábor Szegö. 1975. *Orthogonal polynomials*. Vol. 23. American Mathematical Soc.

[106] N Verzelen and E Arias-Castro. 2015. Community detection in sparse random networks. *Annals of Applied Probability* 25, 6 (2015), 3465–3510.

[107] Alexander S Wein. 2023. Average-case complexity of tensor decomposition for low-degree polynomials. In *Proceedings of the 55th Annual ACM Symposium on Theory of Computing*. 1685–1698.

[108] Alexander S Wein, Ahmed El Alaoui, and Cristopher Moore. 2019. The Kikuchi hierarchy and tensor PCA. In *2019 IEEE 60th Annual Symposium on Foundations of Computer Science (FOCS)*. IEEE, 1446–1468.

[109] Xifan Yu, Ilias Zadik, and Peiyuan Zhang. 2024. Counting Stars is Constant-Degree Optimal For Detecting Any Planted Subgraph. *arXiv preprint arXiv:2403.17766* (2024).

[110] Lenka Zdeborová and Florent Krzakala. 2016. Statistical physics of inference: Thresholds and algorithms. *Advances in Physics* 65, 5 (2016), 453–552.

Received 2024-11-04; accepted 2025-02-01