The Annals of Statistics
2024, Vol. 52, No. 5, 2318–2348
https://doi.org/10.1214/24-AOS2437
© Institute of Mathematical Statistics, 2024

COMPUTATIONAL LOWER BOUNDS FOR GRAPHON ESTIMATION VIA LOW-DEGREE POLYNOMIALS

BY YUETIAN LUO[1,a] AND CHAO GAO[2,b]

[1]*Data Science Institute, University of Chicago,* [a]*yuetian@uchicago.edu*
[2]*Department of Statistics, University of Chicago,* [b]*chaogao@uchicago.edu*

Graphon estimation has been one of the most fundamental problems in network analysis and has received considerable attention in the past decade. From the statistical perspective, the minimax error rate of graphon estimation has been established by (*Ann. Statist.* **43** (2015) 2624–2652) for both stochastic block model (SBM) and nonparametric graphon estimation. The statistical optimal estimators are based on constrained least squares and have computational complexity exponential in the dimension. From the computational perspective, the best-known, polynomial-time estimator is based on universal singular value thresholding (USVT), but it can only achieve a much slower estimation error rate than the minimax one. It is natural to wonder if such a gap is essential. The computational optimality of the USVT or the existence of a computational barrier in graphon estimation has been a long-standing open problem. In this work, we take the first step toward it and provide rigorous evidence for the computational barrier in graphon estimation via low-degree polynomials. Specifically, in SBM graphon estimation, we show that for low-degree polynomial estimators, their estimation error rates cannot be significantly better than that of the USVT under a wide range of parameter regimes and in nonparametric graphon estimation, we show low-degree polynomial estimators achieve estimation error rates strictly slower than the minimax rate. Our results are proved based on the recent development of low-degree polynomials by (*Ann. Statist.* **50** (2022) 1833–1858), while we overcome a few key challenges in applying it to the general graphon estimation problem. By leveraging our main results, we also provide a computational lower bound on the clustering error for community detection in SBM with a growing number of communities and this yields a new piece of evidence for the conjectured Kesten–Stigum threshold for efficient community recovery. Finally, we extend our computational lower bounds to sparse graphon estimation and biclustering with additive Gaussian noise, and provide discussion on the optimality of our results.

1. Introduction. Network analysis has gained considerable research interest in the last couple of decades (Goldenberg et al. (2010), Bickel and Chen (2009), Girvan and Newman (2002), Wasserman and Faust (1994)). A key task in network analysis is to estimate the underlying network generating process. It is useful for many important applications such as studying network evolution (Pensky (2019)), predicting missing links (Miller, Jordan and Griffiths (2009), Airoldi, Costa and Chan (2013), Gao, Lu and Zhou (2015)), learning user preferences in recommender systems (Li et al. (2020)) and correcting errors in crowd-sourcing systems (Shah and Lee (2018)). In this paper, we are interested in the question: when could the underlying network generating process be estimated in a computationally efficient way?

A general representation for the generating process of unlabeled exchangeable networks was first introduced by Aldous (1981), Hoover (1979) and was further developed and named

Received September 2023; revised May 2024.
MSC2020 subject classifications. Primary 68Q17, 05C80, 62H30; secondary 62C20, 62G05.
Key words and phrases. Graphon estimation, computational lower bound, low-degree polynomials, community detection, Kesten–Stigum threshold, statistical–computational tradeoffs.

graphon in Lovász and Szegedy (2006), Diaconis and Janson (2008), Borgs et al. (2008). Specifically, in the graphon model, we observe an undirected graph of n nodes and the associated adjacency matrix $A \in \{0,1\}^{n\times n}$. The value of A_{ij} stands for the presence or the absence of an edge between the ith and the jth nodes. The sampling process of A is determined as follows: conditioning on $(\xi_1,\dots,\xi_n)$,

$$\text{for all } 1 \le i < j \le n, \quad A_{ij} = A_{ji} \sim \text{Bern}(M_{ij}) \quad \text{where } M_{ij} = f(\xi_i,\xi_j). \tag{1}$$

Here, the sequence $\{\xi_i\}$ are i.i.d. random variables sampled from an unknown distribution $\mathbb{P}_\xi$ supported on $[0,1]$. A common choice for $\mathbb{P}_\xi$ is the uniform distribution on $[0,1]$. In this paper, we allow $\mathbb{P}_\xi$ to be arbitrary so that the model (1) can be studied to its full generality. Conditioning on $(\xi_1,\dots,\xi_n)$, A_{ij}'s are mutually independent across all $1 \le i < j \le n$, and we adopt the convention that $A_{ii} = M_{ii} = 0$ for all $i \in [n]$. The function $f : [0,1]\times[0,1] \mapsto [0,1]$, which is assumed to be symmetric, is called graphon. In this work, we focus on this general graphon model and consider the problem of estimating f given A.

The concept of graphon plays a significant role in network analysis. It was originally developed as a limit of a sequence of graphs with growing sizes (Diaconis and Janson (2008), Lovász and Szegedy (2006), Lovász (2012)), and has been applied to various network analysis problems ranging from testing graph properties to characterizing distances between two graphs (Borgs et al. (2008, 2012), Lovász (2012)). The general graphon model in (1) captures many special models of interest. For example, when f is a constant function, it gives rise to the Erdős–Rényi random graph; when f is a blockwise constant function or $\mathbb{P}_\xi$ has a discrete support, it specializes to the stochastic block model (SBM) (Holland, Laskey and Leinhardt (1983)).

One challenge in graphon estimation is the nonidentifiability of f due to the fact that the latent random variables $\{\xi_i\}$ are unobservable. To overcome this, we follow the prior work Gao, Lu and Zhou (2015) and consider estimating f under the empirical loss:

$$\ell(\widehat{M}, M_f) := \frac{1}{\binom{n}{2}} \sum_{1\le i<j\le n} \big(\widehat{M}_{ij} - (M_f)_{ij}\big)^2, \tag{2}$$

where $\widehat{M} \in \mathbb{R}^{n\times n}$ and $(M_f)_{ij} := f(\xi_i,\xi_j)$.

There has been great interest in graphon estimation in the last decade (Wolfe and Olhede (2013), Airoldi, Costa and Chan (2013), Chan and Airoldi (2014), Gao, Lu and Zhou (2015), Klopp, Tsybakov and Verzelen (2017)) and we refer readers to Section 1.3 for detailed discussion. From the statistical perspective, Gao, Lu and Zhou (2015) provided the first characterization for the minimax error rate in graphon estimation. In particular, for the SBM with k blocks, the minimax estimation error rate is

$$\text{SBM class}: \quad \inf_{\widehat{M}} \sup_{M\in\mathcal{M}_k} \mathbb{E}\big(\ell(\widehat{M}, M)\big) \asymp \frac{k^2}{n^2} + \frac{\log k}{n}, \tag{3}$$

where $\mathcal{M}_k$ denotes the set of connectivity probability matrices in SBM with k communities and its exact definition is given Section 3. The minimax upper bound is achieved by a constrained least-squares estimator, which needs to search over all possible graphon matrices in $\mathcal{M}_k$ and is computationally inefficient, that is, with runtime exponential in n.

When f belongs to a Hölder space with smoothness index γ, the minimax estimation error rate is shown to be (Gao, Lu and Zhou (2015))

$$\text{Hölder class}: \quad \inf_{\widehat{M}} \sup_{f\in\mathcal{H}_\gamma(L)} \sup_{\mathbb{P}_\xi} \mathbb{E}\big(\ell(\widehat{M}, M_f)\big) \asymp \begin{cases} n^{-\frac{2\gamma}{\gamma+1}} & 0<\gamma<1, \\ \dfrac{\log n}{n} & \gamma \ge 1, \end{cases} \tag{4}$$

where $\mathcal{H}_\gamma(L)$ denotes the Hölder class to be introduced in Section 4. Again, computing the minimax optimal estimator is expensive, as it is based on first approximating a γ-smooth graphon with a blockwise constant matrix and then applying the constrained least-squares estimator.

From the computational perspective, the problem appears to be far less well understood. The best polynomial-time estimator so far for graphon estimation is the universal singular value thresholding (USVT) (Chatterjee (2015)), and its sharp error bound was obtained by Klopp and Verzelen (2019), Xu (2018),

$$\begin{aligned} &\text{SBM class}: \quad \sup_{M\in\mathcal{M}_k} \mathbb{E}\big(\ell(\widehat{M}_{\text{USVT}}, M)\big) \le C\frac{k}{n}, \\ &\text{Hölder class}: \quad \sup_{f\in\mathcal{H}_\gamma(L)} \sup_{\mathbb{P}_\xi} \mathbb{E}\big(\ell(\widehat{M}_{\text{USVT}}, M_f)\big) \le Cn^{-\frac{2\gamma}{2\gamma+1}}, \end{aligned} \tag{5}$$

for some constant $C > 0$ independent of n and k.

Comparing (3) and (4) with (5), we see that there is a big gap between the estimation error rate achieved by the USVT and the minimax rate. It has been conjectured in Xu (2018) that the error rates in (5) are optimal within the class of polynomial-time algorithms, but no rigorous evidence is provided there. The fundamental computational limits for graphon estimation have been a long-standing open problem in the community (Xu (2018), Gao and Ma (2021), Wu and Xu (2021)). In particular, in a recent survey about the statistical and computational limits for statistical problems with planted structures, Wu and Xu (2021) explicitly highlight "computational hardness of graphon estimation" in their Section 5 as one of the six prominent open problems in the field.

The gap on the performance of polynomial-time algorithms and unconstrained-time algorithms is quite common in high-dimensional statistical problems. There has been a flurry of progress in the statistics and theoretical computer science communities toward understanding the general "statistical–computational tradeoffs" phenomenon. This topic focuses on the gap between signal-to-noise ratio (SNR) requirements under which the problem is information-theoretically solvable vs. polynomial-time solvable. As the SNR increases, such problems often exhibit three phases of interest: (1) statistically unsolvable; (2) statistically solvable but computationally expensive, for example, with runtime exponential in the input dimension; (3) easily solvable in polynomial-time. Many frameworks such as average-case reduction, statistical query (SQ), sum-of-squares (SoS) hierarchy, optimization landscape and low-degree polynomials have been proposed to study this phenomenon, and we refer readers to Section 1.3 for a thorough discussion. Based on these frameworks, rigorous evidence for the computational barrier has been provided for a wide class of statistical problems, such as planted clique, sparse PCA, submatrix detection, tensor PCA, robust mean estimation and many others (Barak et al. (2019), Berthet and Rigollet (2013), Ma and Wu (2015), Zhang and Xia (2018), Brennan, Bresler and Huleihel (2018), Diakonikolas, Kane and Stewart (2017)).

Despite all these successes, the graphon estimation problem is a rare example where to our best knowledge essentially no progress has been made under any framework. We think there are two major challenges in establishing the computational lower bound for graphon estimation: (1) in this problem, we want to establish a computational lower bound for *estimation* error rate, while most existing frameworks are mainly designed for hypothesis testing. Two natural hypothesis testing problems associated with graphon estimation do not have computational barriers, as we will discuss in Appendix A in the Supplementary Material (Luo and Gao (2024)); (2) in contrast to the classical problems, such as planted clique or sparse PCA, there is no such canonical SNR quantity in graphon estimation, though it is often critical to understand this quantity in order to apply existing frameworks.

In this work, we overcome the above challenges and provide the first rigorous piece of evidence for the computational barrier in graphon estimation. The contributions of the paper are summarized below.

1.1. *Our contributions.* The main result of the paper is given by the following theorem.

THEOREM 1. *Suppose* $2 \leq k \leq \sqrt{n}$. *For any* $D \geq 1$, *there exists a universal constant* $c > 0$ *such that*

$$\inf_{\widehat{M} \in \mathbb{R}[A]^{n\times n}_{\leq D}} \sup_{M \in \mathcal{M}_k} \mathbb{E}\big(\ell(\widehat{M}, M)\big) \geq \frac{ck}{nD^4}. \tag{6}$$

Here, the notation $\widehat{M} \in \mathbb{R}[A]^{n\times n}_{\leq D}$ *means that for all* $(i, j) \in [n] \times [n]$, $\widehat{M}_{ij}$ *is a polynomial of* A *with degree no more than* D.

It has been widely conjectured in the literature that for a broad class of high-dimensional problems, degree-D polynomials are as powerful as the class of n^D (up to $\log n$ factors in the exponent) runtime algorithms (Hopkins (2018)). Therefore, by setting $D = \log^{1+\epsilon} n$ for any $\epsilon > 0$, Theorem 1 provides firm evidence that the best estimation error achieved by polynomial-time algorithms for graphon estimation under the SBM class cannot be faster than $\tilde{\Theta}(k/n)$. Up to logarithmic factors, this matches the upper bound achieved by USVT in (5).

We also establish a low-degree polynomial lower bound for graphon estimation under the Hölder class by approximating a smooth graphon via an SBM. See Theorem 4 in Section 4. Again, the statistical error rate in (4) is strictly faster than the one achieved by low-degree polynomial algorithms. Combining the two results, we make a step in resolving the open problem regarding the computational lower bounds for graphon estimation raised by Xu (2018), Gao and Ma (2021), Wu and Xu (2021).

1.2. *From community detection to graphon estimation.* Theorem 1 is proved by leveraging the recent advancement of low-degree polynomials developed by Schramm and Wein (2022). Compared with previous work (Hopkins and Steurer (2017), Hopkins (2018)) on hypothesis testing, the low-degree polynomial lower bound in Schramm and Wein (2022) is directly established for *estimation* problems under some prior distribution, and is thus particularly suitable for graphon estimation. Sharp computational lower bounds have been derived for several important examples in Schramm and Wein (2022) including the planted submatrix problem and the planted dense subgraph problem. However, unlike the examples in Schramm and Wein (2022), the graphon estimation problem does not have a natural prior distribution and SNR and, therefore, it is unclear how the general theorem of Schramm and Wein (2022) can be applied to such a setting.

To address this challenge, we consider another problem in network analysis called community detection. The goal of community detection is to recover the clustering structure of a network. For this purpose, a canonical model is the k-class SBM with within-class and between-class homogeneous connectivity probabilities, that is, for two nodes from the same community, the connectivity probability is set to be p and for two nodes from different communities, the connectivity probability is set to be q (Mossel, Neeman and Sly (2015b), Abbe, Bandeira and Hall (2016)). Unlike the general SBM that has $\frac{k(k-1)}{2}$ model parameters, the SBM used for community detection has only 2 parameters (p and q) and can be viewed as a subset. For this subset, not only all the joint cumulants required by the theorem of Schramm and Wein (2022) can be computed, but we also have a nature SNR that quantifies the statistical computational gap.

By applying Schramm and Wein (2022), we show that a nontrivial clustering error cannot be achieved by low-degree polynomial algorithms below the generalized Kesten–Stigum threshold (Kesten and Stigum (1966), Decelle et al. (2011), Chen and Xu (2016)). This result is of independent interest, and complements the recent progress by Hopkins and Steurer (2017), Bandeira et al. (2021), Banks, Mohanty and Raghavendra (2021), Brennan and Bresler (2020) on the computational limits of community detection. More importantly, the low-degree polynomial lower bound for community detection immediately implies the desired rate (6) for graphon estimation by carefully choosing a least favorable pair of p and q.

This connection between graphon estimation and community detection from the perspective of computational limit is quite surprising. Without any computational constraint, the statistical limits of the two problems are derived from very different arguments in the literature. While the minimax rate of graphon estimation is polynomial (Gao, Lu and Zhou (2015), Klopp, Tsybakov and Verzelen (2017), Gao et al. (2016)), the minimax rate of community detection is exponential (Zhang and Zhou (2016), Fei and Chen (2020)), and one cannot be derived from the other. In contrast, we show that the low-degree polynomial lower bounds for the two problems can be established through the same argument. Detailed discussion on the connection between the two problems will be given in Section 3 and Section 5.

1.3. *Related prior work*. *Graphon estimation* has received considerable attention in the past decade (Wolfe and Olhede (2013), Yang, Han and Airoldi (2014), Airoldi, Costa and Chan (2013), Olhede and Wolfe (2014), Chan and Airoldi (2014), Borgs, Chayes and Smith (2015), Chatterjee (2015), Gao, Lu and Zhou (2015), Klopp, Tsybakov and Verzelen (2017), Gao et al. (2016), Zhang, Levina and Zhu (2017), Klopp and Verzelen (2019)). The minimax error rates for a variety of graphon estimation problems, including (sparse) SBM graphon estimation, nonparametric graphon estimation, graphon estimation with missing entries have been established in Gao, Lu and Zhou (2015), Klopp, Tsybakov and Verzelen (2017), Gao et al. (2016), Klopp and Verzelen (2019). A number of efficient estimators for graphon estimation have been proposed (Airoldi, Costa and Chan (2013), Chatterjee (2015), Chan and Airoldi (2014), Zhang, Levina and Zhu (2017), Li et al. (2020), Gaucher and Klopp (2021)). In the SBM setting, Gaucher and Klopp (2021) showed that a tractable estimator based on variational inference can achieve the minimax rate under appropriate assumptions on the connectivity probability matrix and the clustering labels. Without these additional assumptions, the best polynomial-time estimators for SBM/nonparametric graphon estimation are provided and analyzed in Chatterjee (2015), Klopp and Verzelen (2019), Xu (2018), but they are far from optimal. Recently, graphon estimation in a bipartite graph, private graphon estimation and stochastic block smooth graphon model have also been considered in Choi (2017), Donier-Meroz et al. (2023), Borgs, Chayes and Smith (2015), Sischka and Kauermann (2022).

1.3.1. *Statistical–computational tradeoffs*. There has been a long line of work on studying the statistical–computational tradeoffs in high-dimensional statistical problems. One powerful approach to establish the computational lower bounds is based on the average-case reduction (Berthet and Rigollet (2013), Gao, Ma and Zhou (2017), Wang, Berthet and Samworth (2016), Ma and Wu (2015), Cai, Liang and Rakhlin (2017), Hajek, Wu and Xu (2015), Brennan, Bresler and Huleihel (2018), Brennan and Bresler (2020), Luo and Zhang (2022a), Pananjady and Samworth (2022)), and it requires a distribution over instances in a conjecturally hard problem to be mapped precisely to the target distribution. Once the reduction is done, all hardness results from the conjectured hard problem can be automatically inherited to the target problem. On the other hand, the conclusions rely on conjectures that have not been proved yet. For this reason, many recent literature aims to show computational hardness results under some restricted models of computation, such as sum-of-squares (Ma and Wigderson (2015), Hopkins et al. (2017), Barak et al. (2019)), statistical query (SQ) (Feldman et al. (2017), Diakonikolas, Kane and Stewart (2017), Diakonikolas, Kong and Stewart (2019), Feldman, Perkins and Vempala (2018)), class of circuit (Rossman (2008)), convex relaxation (Chandrasekaran and Jordan (2013)), local algorithms (Gamarnik and Sudan (2014)), low-degree polynomials (Hopkins and Steurer (2017), Kunisky, Wein and Bandeira (2022)) and others.

1.3.2. *Why the low-degree polynomial framework*. Among various ways to establish computational lower bounds, the low-degree polynomial framework is both clean and general. It has already been applied to many important high-dimensional problems and always leads to the same computational limits as conjectured in the literature. Compared with the low-degree polynomial method, the statistical query (SQ) framework is typically applied when the observed data consists of i.i.d. samples, but it is not clear how to cast graphon estimation into this form. The sum-of-squares (SoS) lower bounds provide strong evidence for the average-case hardness, but it is important to note that SoS lower bounds show hardness of certification problems. It does not necessarily imply hardness of estimation/recovery (Bandeira, Kunisky and Wein (2020), Banks, Mohanty and Raghavendra (2021)). Average-case reduction is often applied to hypothesis testing problems (Berthet and Rigollet (2013), Brennan, Bresler and Huleihel (2018)). To show the hardness of estimation from hypothesis testing, one often needs to further perform an extra reduction from estimation to testing. However, as we will see in Appendix A in the Supplementary Material (Luo and Gao (2024)), two natural hypothesis testing problems associated with graphon estimation do not have a statistical–computational gap.

1.3.3. *More literature on low-degree polynomials*. The idea of using low-degree polynomials to predict the statistical–computational gaps was recently developed in a line of work on studying the SoS hierarchy (Hopkins and Steurer (2017), Hopkins (2018), Barak et al. (2019)). Many state-of-art algorithms such as spectral methods and approximate messaging (AMP) (Donoho, Maleki and Montanari (2009)) can be represented as low-degree polynomials (Kunisky, Wein and Bandeira (2022), Gamarnik, Jagannath and Wein (2020), Montanari and Wein (2022)) and the "low" here typically means logarithmic in the dimension. In comparison to SoS computational lower bounds, the low-degree polynomial method is simpler to establish and appears to always yield the same results for natural average-case hardness problems. The majority of the existing low-degree polynomial hardness results are established for hypothesis testing problems based on the notion of *low-degree likelihood ratio*. Examples include unsupervised problems such as planted clique detection (Hopkins (2018), Barak et al. (2019)), community detection in SBM (Hopkins and Steurer (2017), Hopkins (2018), Jin et al. (2022)), spiked tensor model (Hopkins et al. (2017), Hopkins (2018), Kunisky, Wein and Bandeira (2022)), spiked Wishart model (Bandeira, Kunisky and Wein (2020)), sparse PCA (Ding et al. (2024)), spiked Wigner model (Kunisky, Wein and Bandeira (2022)), clustering in Gaussian mixture models (Löffler, Wein and Bandeira (2022), Lyu and Xia (2023)), planted vector recovery (Mao and Wein (2021)), independent component analysis (Auddy and Yuan (2023)) as well as supervised learning problems such as tensor regression (Luo and Zhang (2022b)) and mixed sparse linear regression (Arpino and Venkataramanan (2023)). Very recently, the low-degree polynomial method has also been extended to establish computational hardness in statistical estimation/recovery problems (Schramm and Wein (2022), Koehler and Mossel (2022), Wein (2023), Mao, Wein and Zhang (2023)) and random optimization (Gamarnik, Jagannath and Wein (2020), Wein (2022), Bresler and Huang (2022)). It is gradually believed that the low-degree polynomial method is able to capture the essence of what makes sum-of-squares algorithms, and more generally, polynomial-time algorithms succeed or fail (Hopkins (2018), Kunisky, Wein and Bandeira (2022)). However, there are a couple of important examples where the low-degree polynomials cannot predict the right computational threshold, such as the random 3-XOR-SAT problem (Kunisky, Wein and Bandeira (2022)). In those settings, low-degree polynomials can be outperformed by some "brittle" algebraic methods with almost no noise tolerance; we refer readers to Holmgren and Wein (2021), Zadik et al. (2022), Diakonikolas and Kane (2022) for more discussions. Finally, it is worth mentioning that although we focus on the low-degree polynomial framework, it has been demonstrated that this framework is closely related to many other frameworks, such as SoS, SQ, free-energy landscape and approximate message passing from various perspectives (Hopkins et al. (2017), Barak et al. (2019), Brennan et al. (2021), Bandeira et al. (2022), Montanari and Wein (2022)).

1.4. *Organization of the paper.* After the introduction of notation and preliminaries of low-degree polynomials in Section 2, we present our main results on the low-degree polynomial lower bounds for graphon estimation in SBM and nonparametric graphon estimation in Section 3 and Section 4, respectively. The low-degree polynomial lower bound for community detection in SBM is given in Section 5. Extensions of the main results to sparse graphon estimation and biclustering are given in Section 6. The proofs of the main results are presented in Section 7 and the rest of the proofs are deferred to the Appendices.

2. Notation and preliminaries. Define $\mathbb{N} = \{0, 1, 2, \dots\}$ and $[N] = \{1, \dots, N\}$ for an integer N. For $\boldsymbol{\alpha} \in \mathbb{N}^N$, define $|\boldsymbol{\alpha}| = \sum_{i=1}^N \boldsymbol{\alpha}_i$, $\boldsymbol{\alpha}! = \prod_{i=1}^N \boldsymbol{\alpha}_i!$, and for $X \in \mathbb{R}^N$, define $X^{\boldsymbol{\alpha}} = \prod_{i=1}^N X_i^{\boldsymbol{\alpha}_i}$. Given $\boldsymbol{\alpha}, \boldsymbol{\beta} \in \mathbb{N}^N$, we use $\boldsymbol{\alpha} \geq \boldsymbol{\beta}$ to mean $\boldsymbol{\alpha}_i \geq \boldsymbol{\beta}_i$ for all i. The operations $\boldsymbol{\alpha} + \boldsymbol{\beta}$ and $\boldsymbol{\alpha} - \boldsymbol{\beta}$ are performed entrywise. The notation $\boldsymbol{\beta} \lneq \boldsymbol{\alpha}$ means $\boldsymbol{\beta} \leq \boldsymbol{\alpha}$ and $\boldsymbol{\beta} \neq \boldsymbol{\alpha}$ (but not necessarily $\boldsymbol{\beta}_i < \boldsymbol{\alpha}_i$ for every i). Furthermore, for $\boldsymbol{\alpha} \geq \boldsymbol{\beta}$, we define $\binom{\boldsymbol{\alpha}}{\boldsymbol{\beta}} = \prod_{i=1}^N \binom{\boldsymbol{\alpha}_i}{\boldsymbol{\beta}_i}$. Sometimes, given $n \geq 1$ and $N = n(n-1)/2$, we will view $\boldsymbol{\alpha} \in \mathbb{N}^N$ as a multigraph (without self-loops) on vertex set $[n]$, that is, for each $i < j$, we let $\boldsymbol{\alpha}_{ij}$ represent the number of edges between vertices i and j. In this case, $V(\boldsymbol{\alpha}) \subseteq [n]$ denotes the set of vertices spanned by the edges of $\boldsymbol{\alpha}$. For any vector v, define its ℓ_2 norm as $\|v\|_2 = (\sum_i |v_i|^2)^{1/2}$. For any matrix $D \in \mathbb{R}^{p_1 \times p_2}$, the matrix Frobenius and spectral norms are defined as $\|D\|_{\mathrm{F}} = (\sum_{i,j} D_{ij}^2)^{1/2}$ and $\|D\| = \max_{u \in \mathbb{R}^{p_2}} \|Du\|_2 / \|u\|_2$, respectively. The notation I_r represents the r-by-r identity matrix and $\mathbf{1}_n$ is an all 1 vector in $\mathbb{R}^n$. For any two sequences of numbers, say $\{a_n\}$ and $\{b_n\}$, denote $a_n \asymp b_n$ or $a_n = \Theta(b_n)$ if there exists uniform constants $c, C > 0$ such that $ca_n \leq b_n \leq Ca_n$ for all n; $a_n \lesssim b_n$ means that $a_n \leq Cb_n$ holds for some constant $C > 0$ independent of n and $a_n = \tilde{\Theta}(b_n)$ if a_n/b_n and b_n/a_n are both bounded by $\mathrm{polylog}(n)$, that is, a_n and b_n are on the same order up to $\mathrm{polylog}(n)$ factors. Finally, throughout the paper, let c, c', c'', C be some constants independent of n and k, whose actual values may vary from line to line.

2.1. *Computational lower bounds for estimation via low-degree polynomials.* Consider the *general binary observation model* and suppose the signal $X \in [\tau_0, \tau_1]^N$ with $0 \leq \tau_0 < \tau_1 \leq 1$ is drawn from an arbitrary but known prior. We observe $Y \in \{0, 1\}^N$ where $\mathbb{E}[Y_i | X_i] = X_i$ and $\{Y_i\}_{i=1}^N$ are conditionally independent given X. Let $\mathbb{R}[Y]_{\leq D}$ denote the space of polynomials $g : \mathbb{R}^N \to \mathbb{R}$ of degree at most D of Y. Suppose the goal is to estimate a scalar quantity $x \in \mathbb{R}$, which is a function of X, then we have the following estimation lower bound for low-degree polynomial estimators.

PROPOSITION 1 (Schramm and Wein (2022)). *In the general binary model described above, denote $\mathbb{P}$ as the joint distribution of x and Y. Then for any $D \geq 1$, we have*

$$\inf_{g \in \mathbb{R}[Y]_{\leq D}} \mathbb{E}_{(x,Y) \sim \mathbb{P}} \big(g(Y) - x\big)^2 = \mathbb{E}\big(x^2\big) - \mathrm{Corr}_{\leq D}^2,$$

where the degree-D correlation $\mathrm{Corr}_{\le D}$ *is defined as*

$$\mathrm{Corr}_{\le D} := \sup_{\substack{g \in \mathbb{R}[Y]_{\le D} \\ \mathbb{E}_{\mathbb{P}}[g^2(Y)] \ne 0}} \frac{\mathbb{E}_{(x,Y)\sim\mathbb{P}}[g(Y)\cdot x]}{\sqrt{\mathbb{E}_{\mathbb{P}}[g^2(Y)]}}, \tag{7}$$

and satisfies the property

$$\mathrm{Corr}^2_{\le D} \le \sum_{\boldsymbol{\alpha} \in \{0,1\}^N, 0 \le |\boldsymbol{\alpha}| \le D} \frac{\kappa^2_{\boldsymbol{\alpha}}(x, X)}{(\tau_0(1-\tau_1))^{|\boldsymbol{\alpha}|}}.$$

Here, $\kappa_{\boldsymbol{\alpha}}(x, X)$ *is defined recursively by*

$$\begin{aligned} &\kappa_0(x, X) = \mathbb{E}(x) \quad \textit{and} \\ &\kappa_{\boldsymbol{\alpha}}(x, X) = \mathbb{E}(xX^{\boldsymbol{\alpha}}) - \sum_{0 \le \boldsymbol{\beta} \lneq \boldsymbol{\alpha}} \kappa_{\boldsymbol{\beta}}(x, X)\binom{\boldsymbol{\alpha}}{\boldsymbol{\beta}} \mathbb{E}[X^{\boldsymbol{\alpha}-\boldsymbol{\beta}}] \quad \textit{for } \boldsymbol{\alpha} \textit{ such that } |\boldsymbol{\alpha}| \ge 1. \end{aligned} \tag{8}$$

We note that Proposition 1 provides a general ℓ_2 estimation error lower bound for low-degree estimators of degree at most D. To show the low-degree polynomial lower bound in a specific problem, we then have to bound $\sum_{\boldsymbol{\alpha} \in \{0,1\}^N, 0 \le |\boldsymbol{\alpha}| \le D} \frac{\kappa^2_{\boldsymbol{\alpha}}(x,X)}{(\tau_0(1-\tau_1))^{|\boldsymbol{\alpha}|}}$, but to our knowledge there is no easy and unified way to do that. One important interpretation for $\kappa_{\boldsymbol{\alpha}}(x, X)$ is the following: if we view $\boldsymbol{\alpha}$ as a multiset $\{a_1, \dots, a_m\}$ with $m = \sum_{i=1}^N \boldsymbol{\alpha}_i$, which contains $\boldsymbol{\alpha}_i$ copies of i for all $i \in [N]$, then $\kappa_{\boldsymbol{\alpha}}(x, X)$ is the joint cumulant of a multiset of entries of the signal (Schramm and Wein (2022), Claim 2.14):

$$\kappa_{\boldsymbol{\alpha}}(x, X) = \kappa(x, X_{a_1}, \dots, X_{a_m}), \tag{9}$$

where $\kappa(\cdots)$ denotes the joint cumulant of a set of random variables and its formal definition and properties are provided in Appendix B.1 in the Supplementary Material (Luo and Gao (2024)). This fact about $\kappa_{\boldsymbol{\alpha}}(x, X)$ will be crucially used in our proofs of bounding $\mathrm{Corr}_{\le D}$ for graphon estimation.

A similar result as Proposition 1 holds under the general additive Gaussian noise model as well. We defer the result in that setting to Appendix B.2 in the Supplementary Material (Luo and Gao (2024)).

3. Computational limits for graphon estimation in the stochastic block model. We first define the parameter space of interest in SBM,

$$\begin{aligned} \mathcal{M}_k = \big\{ &M = (M_{ij}) \in [0,1]^{n\times n} : M_{ii} = 0 \text{ for } i \in [n], M_{ij} = M_{ji} = Q_{z_i z_j} \text{ for } i \ne j, \\ &\text{for some } Q = Q^\top \in [0,1]^{k\times k}, z \in [k]^n \big\}. \end{aligned} \tag{10}$$

In other words, the connectivity probability between the ith and the jth nodes, M_{ij}, only depends on Q through their clustering labels z_i and z_j. Given $M \in \mathcal{M}_k$, we observe a random graph with adjacency matrix $A \in \{0,1\}^{n\times n}$ and its generative progress is given in (1). The minimax rate of estimating $M \in \mathcal{M}_k$ is given in (3). It was shown in Gao, Lu and Zhou (2015) that the minimax rate can be achieved by the solution of the following constrained least-squares optimization,

$$\min_{M \in \mathcal{M}_k} \|A - M\|_{\mathrm{F}}^2, \tag{11}$$

which, by the definition of $\mathcal{M}_k$, is equivalent to

$$\begin{aligned}
&\min_{z\in[k]^n}\min_{Q\in\mathbb{R}^{k\times k}}\sum_{a,b\in[k]}\sum_{\substack{(i,j)\in z^{-1}(a)\times z^{-1}(b)\\ i\neq j}}(A_{ij}-Q_{ab})^2\\
&=\min_{z\in[k]^n}\sum_{a,b\in[k]}\sum_{\substack{(i,j)\in z^{-1}(a)\times z^{-1}(b)\\ i\neq j}}\big(A_{ij}-\overline{A}_{ab}(z)\big)^2,
\end{aligned}$$

where $z^{-1}(a) := \{i \in [n] : z_i = a\}$, and

$$\overline{A}_{ab}(z) := \begin{cases} \dfrac{1}{|z^{-1}(a)||z^{-1}(b)|}\displaystyle\sum_{i\in z^{-1}(a)}\sum_{j\in z^{-1}(b)} A_{ij} & a \neq b, \\ \dfrac{1}{|z^{-1}(a)|(|z^{-1}(a)|-1)}\displaystyle\sum_{\substack{i,j\in z^{-1}(a)\\ i\neq j}} A_{ij} & a = b. \end{cases}$$

Unfortunately, since the optimization problem involves searching over all clustering patterns, it is computationally expensive to solve and has runtime exponential in n.

This motivates a line of work on searching for polynomial-time algorithms. Among many of them, a prominent one is the universal singular value thresholding (USVT) proposed in Chatterjee (2015). It is a simple and versatile method for structured matrix estimation and has been applied to a variety of different problems such as low-rank matrix estimation, distance matrix completion, graphon estimation and ranking (Chatterjee (2015), Shah et al. (2016)). In particular, given the SVD of $A = U\Sigma V^\top = \sum_{i=1}^n \sigma_i(A)u_i v_j^\top$, where $\sigma_i(A)$ denotes the ith largest singular value of A, USVT estimates M by

$$\widehat{M}_{\mathrm{USVT}}(\tau) = \sum_{i:\sigma_i(A)>\tau}\sigma_i(A)u_i v_i^\top, \tag{12}$$

where τ is a carefully chosen tuning parameter. In the original paper by Chatterjee (2015), it was proved that USVT achieves the error rate $\sqrt{k/n}$ in estimating $M \in \mathcal{M}_k$. Later, the error rate of USVT was improved to k/n via a sharper analysis (Klopp and Verzelen (2019), Xu (2018)). Other polynomial-time algorithms in the literature (Airoldi, Costa and Chan (2013), Chan and Airoldi (2014), Zhang, Levina and Zhu (2017), Li et al. (2020), Borgs et al. (2021), Gaucher and Klopp (2021)) either achieve error rates no better than k/n or require additional assumptions on the matrix M. In this section, we will show that k/n is the best possible error rate that can be achieved by low-degree polynomial algorithms.

In order to apply the general tool given by Proposition 1, one needs to find a prior distribution supported on $\mathcal{M}_k$ and compute all the joint cumulants under this prior. It turns out that the analysis of the cumulants is intractable under the least favorable prior constructed by Gao, Lu and Zhou (2015) to prove the minimax lower bound. We need a simpler prior to apply Proposition 1. To this end, we introduce a special class of SBM models considered in the community detection literature, denoted by $\mathcal{M}_{k,p,q}$ $(0 \le q < p \le 1)$, whose definition is given by

$$\begin{aligned}
\mathcal{M}_{k,p,q} = \big\{&M = (M_{ij}) \in [0,1]^{n\times n} : M_{ii} = 0 \text{ for } i \in [n], \\
&M_{ij} = M_{ji} = p\mathbb{1}(z_i = z_j) + q\mathbb{1}(z_i \neq z_j) \text{ for } i \neq j \text{ for some } z \in [k]^n\big\}.
\end{aligned} \tag{13}$$

Since $\mathcal{M}_{k,p,q}$ is much simpler than $\mathcal{M}_k$, it is not clear that it would lead to a sharp computational lower bound. However, we will show that when the algorithms are restricted within the class of low-degree polynomials, graphon estimation under $\mathcal{M}_{k,p,q}$ can be as difficult as that under $\mathcal{M}_k$.

We consider the following natural prior distribution $\mathbb{P}_{\mathrm{SBM}(p,q)}$ supported on $\mathcal{M}_{k,p,q}$. In particular, $M \sim \mathbb{P}_{\mathrm{SBM}(p,q)}$ can be generated as follows: first, sample $z \in [k]^n$ according to $z_i \overset{i.i.d.}{\sim} \mathrm{Unif}\{1,\dots,k\}$ for all $i \in [n]$. Then let $M_{ij} = p\mathbb{1}(z_i = z_j) + q\mathbb{1}(z_i \neq z_j)$ for all $1 \le i < j \le n$ and $M_{ii} = 0$ for all $i \in [n]$. Our first main result shows that when the SNR of $\mathcal{M}_{k,p,q}$ is smaller than a certain threshold, then the estimation error of any low-degree polynomial estimator can be bounded from below.

THEOREM 2. *For any* $0 < r < 1$ *and* $D \ge 1$, *if*

$$\frac{(p-q)^2}{q(1-p)} \le \frac{r}{(D(D+1))^2}\left(\frac{k^2}{n} \wedge 1\right), \tag{14}$$

then we have

$$\inf_{\widehat{M} \in \mathbb{R}[A]^{n\times n}_{\le D}} \mathbb{E}_{A,M\sim\mathbb{P}_{\mathrm{SBM}(p,q)}}\big(\ell(\widehat{M}, M)\big) \ge \frac{(p-q)^2}{k} - (p-q)^2\left(\frac{1}{k^2} + \frac{r(2-r)}{(1-r)^2 n}\right). \tag{15}$$

Here, the notation $\widehat{M} \in \mathbb{R}[A]^{n\times n}_{\le D}$ *means that for all* $(i,j) \in [n]\times[n]$, *we have* $\widehat{M}_{ij} \in \mathbb{R}[A]_{\le D}$.

To understand the result of Theorem 2, let us consider the special case $k \le \sqrt{n}$ and ignore the second term on the right-hand side of (15). Then Theorem 2 indicates that whenever

$$\frac{n(p-q)^2}{k^2 q(1-p)} \ll 1, \tag{16}$$

the graphon estimation error cannot be better than $\frac{(p-q)^2}{k}$. We remark that $\frac{(p-q)^2}{k}$ is in fact a trivial error under the prior distribution $M \sim \mathbb{P}_{\mathrm{SBM}(p,q)}$, since it can be achieved by the constant estimator $\widehat{M}_{ij} = q$ for all $i \neq j$. One may recognize that the SNR condition (16) is related to the well-known *Kesten–Stigum* threshold (Kesten and Stigum (1966), Decelle et al. (2011)) in the literature of community detection (See Section 5 for more details). With arguments from statistical physics, it was conjectured that when the number of communities k is a constant, nontrivial community detection is possible in polynomial time whenever

$$\frac{n(p-q)^2}{k(p+(k-1)q)} > 1, \tag{17}$$

at least under the asymptotic regime $p = a/n$ and $q = b/n$ for some constants $a > b$. For general p and q such that $p \lesssim q < p < 0.99$, the two SNRs on the left-hand sides of (16) and (17) are of the same order. In fact, (16) could be regarded as an asymptotic extension or generalized version of (17) when k grows (Brennan and Bresler (2020)) and Chen and Xu (2016) conjectures (see their Conjecture 9) that it is the computational limits for community detection in SBM with a growing number of communities. Hence, Theorem 2 simply says nontrivial graphon estimation is not possible below the *generalized Kesten–Stigum* threshold under the parameter space $\mathcal{M}_{k,p,q}$.

To find a tight computational lower bound for graphon estimation under the original SBM class $\mathcal{M}_k$, we define

$$\mathcal{M}'_k = \bigcup_{0\le q\le p\le 1} \mathcal{M}_{k,p,q}. \tag{18}$$

Observe that $\mathcal{M}'_k \subset \mathcal{M}_k$, and we have

$$\begin{aligned} \inf_{\widehat{M} \in \mathbb{R}[A]^{n\times n}_{\le D}} \sup_{M\in\mathcal{M}_k} \mathbb{E}\big(\ell(\widehat{M}, M)\big) &\ge \inf_{\widehat{M} \in \mathbb{R}[A]^{n\times n}_{\le D}} \sup_{M\in\mathcal{M}'_k} \mathbb{E}\big(\ell(\widehat{M}, M)\big) \\ &\ge \inf_{\widehat{M} \in \mathbb{R}[A]^{n\times n}_{\le D}} \mathbb{E}_{A,M\sim\mathbb{P}_{\mathrm{SBM}(p,q)}}\big(\ell(\widehat{M}, M)\big). \end{aligned} \tag{19}$$

Since the above inequality holds for arbitrary $0 \le q \le p \le 1$, we can find a pair of p and q to maximize the right-hand side of (15) under the SNR constraint (14). This immediately leads to the following result.

COROLLARY 1. *Suppose $k \ge 2$. For any $D \ge 1$, there exists a universal constant $c > 0$ such that*

$$\inf_{\widehat{M} \in \mathbb{R}[A]_{\le D}^{n\times n}} \sup_{M \in \mathcal{M}_k} \mathbb{E}\big(\ell(M, \widehat{M})\big) \ge \frac{c}{D^4}\left(\frac{k}{n} \wedge \frac{1}{k}\right).$$

When $k \le \sqrt{n}$, the result in Corollary 1 reduces to Theorem 1. Under the low-degree polynomial conjecture (Hopkins (2018)) with $D = \log^{1+\epsilon} n$, the lower bound $\frac{k}{nD^4}$ matches the rate (5) achieved by USVT up to some logarithmic factors. This is a bit surprising since Corollary 1 is actually proved for a much smaller parameter space $\mathcal{M}_k'$ than the original one $\mathcal{M}_k$. This provides a valuable insight that in the regime $k \le \sqrt{n}$ the simple SBM prior $\mathbb{P}_{\mathrm{SBM}(p,q)}$ provides "computationally" a least favorable prior for graphon estimation.

When $k > \sqrt{n}$, however, the rate $\frac{1}{kD^4}$ does not match the performance of the USVT. This may result from the fact that the computational limits of the two spaces $\mathcal{M}_k'$ and $\mathcal{M}_k$ are different when k is large. We will verify in the following Section 3.1 that the rate $\frac{1}{kD^4}$ is actually sharp if we consider the smaller space $\mathcal{M}_k'$.

3.1. *Optimality of Theorem 2.* Our main result Theorem 2 leads to the lower bound rate $\frac{k}{n} \wedge \frac{1}{k}$ in Corollary 1 for graphon estimation under the SBM class $\mathcal{M}_k$. When $k > \sqrt{n}$, this rate becomes $\frac{1}{k}$, and does not match the upper bound achieved by USVT. In fact, since $\frac{1}{k}$ is even smaller than the minimax rate (3) when $k > n^{2/3}$, it cannot be the sharp. We will argue in this section that the suboptimal rate $\frac{1}{k}$ is due to the choice of the subset $\mathcal{M}_k'$ instead of an artifact of the proof of Theorem 2. The Bayes risk of Theorem 2 with respect to the prior $\mathbb{P}_{\mathrm{SBM}(p,q)}$ (supported on $\mathcal{M}_k'$) is optimal, and an improvement of the rate $\frac{1}{k}$ must involve a different subset.

Recall the definition $\mathcal{M}_k' = \bigcup_{0\le q\le p\le 1} \mathcal{M}_{k,p,q}$. Theorem 2 and the inequality (19) imply

$$\begin{aligned}
&\inf_{\widehat{M} \in \mathbb{R}[A]_{\le D}^{n\times n}} \sup_{M \in \mathcal{M}_k'} \mathbb{E}\big(\ell(\widehat{M}, M)\big) \\
&\quad \ge \inf_{\widehat{M} \in \mathbb{R}[A]_{\le D}^{n\times n}} \sup_{0\le q\le p\le 1} \mathbb{E}_{A,M\sim \mathbb{P}_{\mathrm{SBM}(p,q)}}\big(\ell(\widehat{M}, M)\big) \\
&\quad \ge \frac{c}{D^4}\left(\frac{k}{n} \wedge \frac{1}{k}\right).
\end{aligned}$$

The above lower bound cannot be improved. To see this, consider the following algorithm:

$$\widehat{M} = \begin{cases} \widehat{M}_{\mathrm{USVT}}(\tau) & k \le \sqrt{n}, \\ \widehat{M}_{\mathrm{mean}} & k \ge \sqrt{n}, \end{cases} \tag{20}$$

where $(\widehat{M}_{\mathrm{mean}})_{ij} = (\widehat{M}_{\mathrm{mean}})_{ji} = \sum_{1\le u<v\le n} A_{uv}/\binom{n}{2}$. When $k \le \sqrt{n}$, the USVT with $\tau \asymp \sqrt{n}$ achieves the rate $\frac{k}{n}$ (Xu (2018)). When $k > \sqrt{n}$, a straightforward calculation (see Appendix C.1 in the Supplementary Material (Luo and Gao (2024))) leads to

$$\sup_{0\le q\le p\le 1} \mathbb{E}_{A,M\sim \mathbb{P}_{\mathrm{SBM}(p,q)}}\big(\ell(\widehat{M}_{\mathrm{mean}}, M)\big) \le C\frac{1}{k}.$$

Algorithm 1 Low-degree Polynomial Algorithm for SBM Graphon Estimation

1: **Input:** A, p, q, k, r, t_1 and t_2.
2: (Fill the diagonal and transform the data) Let $\Lambda \in \mathbb{R}^{n\times n}$ be a diagonal matrix with i.i.d. $\mathrm{Bern}(p)$ entries on its diagonal and they are independent of A; let $\widetilde{A} = A + \Lambda - q\mathbf{1}_n\mathbf{1}_n^\top$.
3: (Power iteration) Generate an independent random matrix $B \in \mathbb{R}^{p\times r}$ with i.i.d. $N(0, 1)$ entries; compute $\widetilde{A}^{t_1}B$.
4: (Gradient descent) Run t_2 iterations of gradient descent (GD) with zero initialization on the objective $\min_{W\in\mathbb{R}^{r\times n}} \|\widetilde{A}^{t_1}BW - \widetilde{A}\|_{\mathrm{F}}^2$, that is, for $l = 0$ to $t_2 - 1$, compute

$$W_{l+1} = W_l - \eta B^\top \widetilde{A}^{t_1}\big(\widetilde{A}^{t_1}BW_l - \widetilde{A}\big) \quad \text{with } W_0 = \mathbf{0}.$$

5: **Output:** $\widehat{M} = \widetilde{A}^{t_1}BW_{t_2} + q\mathbf{1}_n\mathbf{1}_n^\top$.

In addition, for any $M \in \mathcal{M}_k'$, if we further assume $\frac{n}{\beta k} \le \sum_{i=1}^n \mathbb{1}((z_M)_i = a) \le \frac{n\beta}{k}$ for all $a \in [k]$ for some constant $\beta > 1$, we also have $\mathbb{E}_A(\ell(\widehat{M}_{\mathrm{mean}}, M)) \le C\frac{1}{k}$. In other words, the estimator (20) achieves the rate $\frac{k}{n} \wedge \frac{1}{k}$, and thus the lower bound cannot be improved.

As we have discussed in Section 3, our low-degree polynomial lower bounds for graphon estimation are derived by the connection to community detection. When $k > \sqrt{n}$, it is likely that the computational limits of the two problems are very different. A sharp lower bound for graphon estimation probably requires the construction of a very different subset than $\mathcal{M}_k'$. We leave this problem open.

3.2. *A matching low-degree polynomial upper bound for* $\mathbb{P}_{\mathrm{SBM}(p,q)}$. Though the error rate of USVT matches our low-degree polynomial lower bound when $k \le \sqrt{n}$, it is not strictly a low-degree polynomial algorithm, that is, its entry cannot be written as a polynomial of entries of A. In this section, we provide a rigorous low-degree polynomial algorithm with near-optimal guarantees. For technical convenience, we consider the setting $M \sim \mathbb{P}_{\mathrm{SBM}(p,q)}$ as defined in Section 3. [1] The algorithm is described below.

The main idea of Algorithm 1 is to simulate SVD via power iteration. However, power iteration does not lead to the right scaling without additional normalization, and this motivates us to run a further least-squares optimization to normalize the matrix. Least squares is not a low-degree algorithm since it involves matrix inverse, and this is simulated via gradient descent.

By simple counting, one can show that each entry of $\widehat{M}$ is a polynomial of entries of A, Λ, B with degree at most $2t_1t_2$. The guarantee of $\widehat{M}$ returned by Algorithm 1 is given as follows.

THEOREM 3. *Take $r = 2k$, $t_1 = t_2 = C'\log n$ and the stepsize of GD to be $\eta = \frac{1}{C''((\frac{n(p-q)}{k} + C''\sqrt{n})^{2t_1}k \vee (C''n)^{t_1+1})}$ for some large $C', C'' > 0$ in Algorithm 1. Then there exist $c, C, \bar{C} > 0$ depending only on C', C'' such that when $n \ge Ck\log^3 n$, we have with $\mathbb{P}_{\mathrm{SBM}(p,q)}$-probability at least $1 - n^{-\bar{C}}$, the $\widehat{M}$ in Algorithm 1 satisfies $\ell(\widehat{M}, M) \le \frac{c(k+\log n)\log^2 n}{n}$.*

To summarize, $\widehat{M}$ is a $O(\log^2 n)$-degree polynomial estimator that achieves the k/n error rate up to logarithmic factors. One important feature of Algorithm 1 is that it works for any $p, q \in [0, 1]$ and it automatically adapts between situations with a spectral gap or not. In addition, we note that in order to make the algorithm work, it is important to choose r to satisfy

[1] In fact, when $k \le \sqrt{n}$, one can show, via similar arguments in Gao, Lu and Zhou (2015), that the information theoretically optimal rate under $M \sim \mathbb{P}_{\mathrm{SBM}(p,q)}$ is still $\widetilde{\Theta}(1/n)$ for the least favorable pair of (p, q).

$r/k > 1$ for the sketching matrix B in Step 3. With this choice, the least-squares optimization is well conditioned and the gradient descent achieves a linear rate of convergence in the high SNR regime when there is a spectral gap. In the low SNR regime without a spectral gap, gradient descent after $t_2 = O(\log n)$ iterations stays close to the zero initialization, which still works for our purpose.

Compared with the low-degree upper bounds in Schramm and Wein (2022) where a single power iteration is needed in planted submatrix and dense subgraph problems, we have to run a logarithmic number of power iterations followed by a logarithmic number of iterations of gradient descent. The logarithmic number of power iterations seems to be necessary for us to extract the subspace information of A. In general, the proposed algorithm can understood as a principled way of simulating spectral algorithms via low-degree polynomials. On the other hand, we note that even though our algorithm is polynomial-time, it has degree $O(\log^2 n)$. It will be interesting to find a $O(\log n)$-degree algorithm to simulate spectral algorithms.

REMARK 1. Careful readers may notice that our estimator $\widehat{M}$ is a low-degree polynomial of entries of A as well as independently generated Λ and B, while our low-degree polynomial lower bounds in Section 3 are proved for the class of deterministic polynomials. However, this is not an issue since the low-degree polynomial lower bounds will continue to hold if we consider the class of polynomials of A, Λ and B. This is due to the fact that cumulants on two groups of independent random variables are zero (see Proposition 1 in Appendix B in the Supplementary Material (Luo and Gao (2024))). The same issue has also been dealt with in Claim A.1 by Schramm and Wein (2022).

4. Computational limits for nonparametric graphon estimation. Let us proceed to nonparametric graphon estimation. We first introduce a class of Hölder smooth graphon. Since graphons are symmetric functions, we only need to consider functions on $\mathcal{D} = \{(x, y) \in [0, 1] \times [0, 1] : x \geq y\}$. Define the derivative operator by

$$\nabla_{jk} f(x, y) = \frac{\partial^{j+k}}{(\partial x)^j (\partial y)^k} f(x, y),$$

and we adopt the convention $\nabla_{00} f(x, y) = f(x, y)$. Given a $\gamma > 0$, the Hölder norm of f is defined as

$$\|f\|_{\mathcal{H}_\gamma} = \max_{j+k \leq \lfloor \gamma \rfloor} \sup_{(x,y)\in\mathcal{D}} |\nabla_{jk} f(x, y)| + \max_{j+k=\lfloor \gamma \rfloor} \sup_{(x,y)\neq(x',y')\in\mathcal{D}} \frac{|\nabla_{jk} f(x, y) - \nabla_{jk} f(x', y')|}{(|x - x'| + |y - y'|)^{\gamma - \lfloor \gamma \rfloor}},$$

and the Hölder class with smoothness parameter $\gamma > 0$ and radius $L > 0$ is defined as

$$\mathcal{H}_\gamma(L) = \{\|f\|_{\mathcal{H}_\gamma} \leq L : f(x, y) = f(y, x) \text{ for } x \geq y\}.$$

Finally, the class of smooth graphon of interest is

$$\mathcal{F}_\gamma(L) = \{0 \leq f \leq 1 : f \in \mathcal{H}_\gamma(L)\}.$$

The minimax rate of estimating $f \in \mathcal{F}_\gamma(L)$ is given by (4). Note that this rate can also be written as

$$\min_k \left(\frac{k^2}{n^2} + \frac{\log k}{n} + k^{-2(\gamma\wedge 1)}\right) \asymp \begin{cases} n^{-\frac{2\gamma}{\gamma+1}} & 0 < \gamma < 1, \\ \dfrac{\log n}{n} & \gamma \geq 1, \end{cases} \tag{21}$$

where the first term $\frac{k^2}{n^2} + \frac{\log k}{n}$ is the minimax rate of graphon estimation under the SBM class $\mathcal{M}_k$, and the second term $k^{-2(\gamma\wedge 1)}$ is the error of approximating a nonparametric graphon

$f \in \mathcal{F}_\gamma(L)$ by an SBM with k blocks (Lemma 2.1 of Gao, Lu and Zhou (2015)). A rate-optimal estimator can be constructed by the same constrained least-squares optimization (11) with k chosen to be $\lceil n^{\frac{1}{1+\gamma\wedge 1}} \rceil$, that is, the solution to the bias-variance tradeoff (21). Despite its statistical optimality, solving (11) is computationally intractable.

In terms of polynomial-time algorithms, it was proved by Xu (2018) that the USVT estimator (12) with tuning parameter $\tau \asymp \sqrt{n}$ achieves the rate (5). Just as (21), the suboptimal rate (5) can also be written in the form of bias-variance tradeoff,

$$\min_k \left(\frac{k}{n} + k^{-2\gamma} \right) \asymp n^{-2\gamma/(2\gamma+1)}, \tag{22}$$

where $\frac{k}{n}$ is the error rate of estimating a rank-k matrix, and $k^{-2\gamma}$ is the error of approximating a nonparametric graphon $f \in \mathcal{F}_\gamma(L)$ by a rank-k matrix (Proposition 1 of Xu (2018)). The optimal choice of k is given by $\lceil n^{\frac{1}{1+2\gamma}} \rceil$. Other polynomial-time algorithms in the literature (Airoldi, Costa and Chan (2013), Chan and Airoldi (2014), Zhang, Levina and Zhu (2017), Li et al. (2020), Borgs et al. (2021)) either achieve error rates no better than $n^{-2\gamma/(2\gamma+1)}$ or require additional assumptions on f. In the following result, we provide a lower bound for nonparametric graphon estimation within the class of low-degree polynomials.

THEOREM 4. *Suppose $\gamma > 0.5$. For any $D \geq 1$, there exists $c > 0$ only depending on L and γ such that*

$$\inf_{\widehat{M} \in \mathbb{R}[A]^{n\times n}_{\leq D}} \sup_{f \in \mathcal{F}_\gamma(L)} \sup_{\mathbb{P}_\xi} \mathbb{E}\big(\ell(\widehat{M}, M_f)\big) \geq c n^{-\frac{2\gamma+1}{2\gamma+2}} / D^4.$$

Theorem 4 is proved by similar arguments that lead to Theorem 1. A simple calculation shows that the low-degree polynomial lower bound $n^{-\frac{2\gamma+1}{2\gamma+2}}$ is strictly slower than the statistical rate (21) by a factor scales polynomially in n whenever $\gamma > 0.5$. It confirms that the minimax rate in nonparametric graphon estimation cannot be achieved by the class of low-degree polynomials when $\gamma > 0.5$, providing rigorous evidence for the statistical–computational gap.

Careful readers may notice the gap between the low-degree polynomial lower bound and the upper bound achieved by USVT. We believe this is due to the fact that Theorem 4 is proved based on Theorem 2, where we use the SBM model class $\mathcal{M}'_k$, that is, SBM class with two parameters (p, q), to approximate a Hölder smooth graphon. To be specific, the optimal choice of p, q in Theorem 2 would satisfy $p - q \asymp \frac{k}{\sqrt{n}}$. At the same time, to guarantee that SBM(p, q) is a γ-Hölder smooth graphon, we need the condition $p - q \lesssim 1/k^\gamma$ (see Proposition 4 in Appendix D in the Supplementary Material (Luo and Gao (2024))), that is, $\frac{k^2}{n} \lesssim k^{-2\gamma}$. So, our choice of k is from the tradeoff between $\frac{k}{n}$ and $k^{-2\gamma-1}$, which is different from the tradeoff (22) for USVT. To close this gap, we believe that a more sophisticated SBM class is needed to approximate Hölder smooth graphons, which is beyond the scope of the paper and we leave it as an interesting future direction.

5. Computational limits for community detection in SBM.

The key to the derivation of the computational lower bound for graphon estimation is the understanding of community detection under the distribution $\mathbb{P}_{\mathrm{SBM}(p,q)}$ supported on $\mathcal{M}_{k,p,q}$. For any $M \in \mathcal{M}_{k,p,q}$ with $p > q$, there exists a unique $z \in [k]^n$ such that

$$M_{ij} = p\mathbb{1}(z_i = z_j) + q\mathbb{1}(z_i \neq z_j).$$

We write such z as z_M to emphasize its dependence on M. The membership matrix Z_M is defined by: for $i \in [n]$, $(Z_M)_{ii} = 0$, for all $i \neq j$,

$$(Z_M)_{ij} = \mathbb{1}\big((z_M)_i = (z_M)_j\big) = \frac{M_{ij} - q}{p - q}. \tag{23}$$

The goal of the community detection is to recover the clustering labels z_M or the membership matrix Z_M.

The problem of community detection has been widely studied in the literature (Bickel and Chen (2009), Rohe, Chatterjee and Yu (2011), Lei and Rinaldo (2015), Jin (2015)). When $k = 2$, groundbreaking work by Mossel, Neeman and Sly (2015b, 2018), Massoulié (2014) shows that nontrivial community detection (better than random guess) is possible if and only if $\frac{n(p-q)^2}{2(p+q)} > 1$. Sharp SNR thresholds have also been derived for partial recovery and exact recovery (Mossel, Neeman and Sly (2015a), Abbe, Bandeira and Hall (2016)). We refer the readers to Abbe (2017), Moore (2017) for extensive reviews on the topic.

It turns out that the problem starts to exhibit a statistical–computational gap as k gets larger. When k is a large *constant*, with arguments from statistical physics, it was conjectured in the literature that nontrivial community detection is possible in polynomial time whenever the SNR exceeds the *Kesten–Stigum* threshold (Kesten and Stigum (1966), Decelle et al. (2011)), which is sharply characterized by $\frac{n(p-q)^2}{k(p+(k-1)q)} > 1$ at least under the asymptotic regime $p = a/n$ and $q = b/n$ for some constants $a > b > 0$. In contrast, the information-theoretic limit only requires $\frac{n(p-q)^2}{pk\log k}$ to be large for nontrivial community detection (Banks et al. (2016), Zhang and Zhou (2016)), so there is a (constant level) statistical–computational gap. The algorithmic side of this conjecture has been resolved in Abbe and Sandon (2018), while rigorous evidence of the computational lower bound has been much more elusive and was partially provided by Hopkins and Steurer (2017), Bandeira et al. (2021), Banks, Mohanty and Raghavendra (2021). There is also a statistical–computational gap for the detection version of the problem and statistical/computational thresholds for detection and recovery problems are the same when k is a constant (Bandeira et al. (2021), Banks, Mohanty and Raghavendra (2021)).

In this section, we focus on the problem of community detection with a potentially growing k as n grows. Different from the constant k regime, in Appendix A in the Supplementary Material (Luo and Gao (2024)), we illustrate that two natural hypothesis testing problems associated with SBM do not have a statistical–computational gap when k grows (at least there is not a statistical–computational gap scaling polynomially in n). However, it was conjectured in Chen and Xu (2016), Brennan and Bresler (2020) that there is still a statistical–computational gap for the recovery problem in SBM with a growing number of communities and the computational limit is given by the generalized Kesten–Stigum threshold (16).

Our goal of this section is to present a low-degree polynomial lower bound for recovery in SBM with growing k under the following loss function:

$$\ell(\widehat{Z}, Z) = \frac{1}{\binom{n}{2}} \sum_{1 \le i < j \le n} (\widehat{Z}_{ij} - Z_{ij})^2.$$

Compared with Hamming loss of estimating the clustering labels, the above loss for estimating the membership matrix avoids the identifiability issue due to label switching. Under the distribution $M \sim \mathbb{P}_{\mathrm{SBM}(p,q)}$, it is easy to show that a trivial error of community detection is

$$\mathbb{E}_{A,M\sim\mathbb{P}_{\mathrm{SBM}(p,q)}}\big(\ell(\widehat{Z}, Z_M)\big) = \frac{1}{k} - \frac{1}{k^2},$$

achieved by $\widehat{Z} = \frac{1}{k}\mathbf{1}_{n\times n}$ where $\mathbf{1}_{n\times n}$ denotes a $n \times n$ matrix with all 1 in its entries. Random guess would achieve a slightly worse error $\frac{2}{k}(1-\frac{1}{k})$ under the same setting. Therefore, we say that an algorithm $\widehat{Z}$ can achieve nontrivial community detection if its error is much smaller than $\frac{1}{k} - \frac{1}{k^2}$. When $\frac{n(p-q)^2}{pk^2} > C$ for some sufficiently large constant $C > 0$, nontrivial community detection is possible, and polynomial-time algorithms including spectral clustering (Chin, Rao and Vu (2015), Abbe and Sandon (2018)) and semidefinite programming (SDP) (Guédon and Vershynin (2016), Li, Chen and Xu (2021)) would work.[2] Next, we provide a result in the other direction.

THEOREM 5. *For any $D \geq 1$, suppose*

$$\frac{(p-q)^2}{q(1-p)} \leq \frac{1}{2(D(D+1))^2}\left(\frac{k^2}{n} \wedge 1\right),$$

then

$$\inf_{\widehat{Z}\in\mathbb{R}[A]^{n\times n}_{\leq D}} \mathbb{E}_{A,M\sim\mathbb{P}_{\mathrm{SBM}(p,q)}}\big(\ell(\widehat{Z}, Z_M)\big) \geq \frac{1}{k} - \frac{1}{k^2} - \frac{3}{n}. \tag{24}$$

In particular, when $k \leq \sqrt{n}$ and

$$\frac{n(p-q)^2}{k^2 q(1-p)} \leq \frac{1}{2(D(D+1))^2}, \tag{25}$$

the lower bound (24) holds.

Theorem 5 shows that when the SNR $\frac{n(p-q)^2}{k^2q(1-p)}$ is small, no low-degree polynomial algorithm can achieve nontrivial community detection, which provides firm evidence of the conjecture of the *generalized Kesten–Stigum* threshold for community detection in SBM with a growing number of communities. In fact, Theorem 5 can be viewed as a rearrangement of Theorem 2. Given the relation (23), the loss functions of graphon estimation and community detection can be linked through $\ell(\widehat{M}, M) = (p-q)^2\ell(\widehat{Z}, Z_M)$.

As we have mentioned above, there are a couple of existing pieces of evidence for the computational limits of community detection in SBM when k is a constant (Hopkins and Steurer (2017), Bandeira et al. (2021), Banks, Mohanty and Raghavendra (2021)). While when the number of communities grows, to our knowledge, there is only one piece of evidence for the hardness of recovery in SBM via average-case reduction from secrete-leakage planted clique (Brennan and Bresler, 2020, Section 14.1). They considered establishing the computational lower bound for a testing problem where the null is the Erdős–Rényi random graph and the alternative is a variant of imbalanced SBM (ISBM) with two features: first, the averaged number of degrees under the null and alternative are matched; second, the ISBM under the alternative is a mean-field analogy of the original SBM so that the testing problem becomes harder and it matches the hardness of the recovery problem. The reduction result is significant as all existing computational hardness evidence for secrete-leakage planted clique can be inherited to the testing problem they consider. The limitation is that they do not directly handle the estimation problem under the original SBM model; moreover, their reduction only works when $k = o(n^{1/3})$, while our computational lower bound is valid as long as $k \leq \sqrt{n}$.

[2]For completeness, the performance of SDP under the model $M \sim \mathbb{P}_{\mathrm{SBM}(p,q)}$ is given in Appendix E.2 in the Supplementary Material (Luo and Gao (2024)).

6. Extensions and discussion. Our main results can also be extended to the following settings. In Section 6.1, we consider sparse graphon estimation, and present a corresponding low-degree polynomial lower bound. Section 6.2 considers the estimation problem under a biclustering structure with additive Gaussian noise, which can be regarded as an extension of the SBM to rectangular matrices.

6.1. *Computational lower bound for sparse graphon estimation.* Network observed in practice is often sparse in the sense that the total number of edges is of order $o(n^2)$. The problem of sparse graphon estimation is typically more complex than the dense one and has also been widely considered in the literature (Bickel and Chen (2009), Bickel, Chen and Levina (2011), Borgs et al. (2018, 2019), Klopp, Tsybakov and Verzelen (2017), Gao et al. (2016), Borgs et al. (2021)). This section will focus on the sparse SBM model. Given any $0 < \rho < 1$, the class of probability matrices is defined as

$$\begin{aligned}\mathcal{M}_{k,\rho} = \{ & M = (M_{ij}) \in [0,\rho]^{n\times n} : M_{ii} = 0 \text{ for } i \in [n], M_{ij} = M_{ji} = Q_{z_i z_j} \text{ for } i \neq j, \\ & \text{for some } z \in [k]^n, Q = Q^\top \in [0,\rho]^{k\times k}\}.\end{aligned} \tag{26}$$

The minimax rate for sparse graphon estimation has been derived by Klopp, Tsybakov and Verzelen (2017), Gao et al. (2016),

$$\inf_{\widehat{M}} \sup_{M \in \mathcal{M}_{k,\rho}} \mathbb{E}(\ell(\widehat{M}, M)) \asymp \rho\left(\frac{k^2}{n^2} + \frac{\log k}{n}\right) \wedge \rho^2.$$

By solving a constrained least-squares optimization problem $\min_{M\in\mathcal{M}_{k,\rho}} \|A - M\|_F^2$ similar to (11), one achieves the rate $\rho(\frac{k^2}{n^2} + \frac{\log k}{n})$. The other part of the minimax rate ρ^2 can be trivially achieved by $\widehat{M} = \mathbf{0}$. In terms of polynomial time algorithms, Klopp and Verzelen (2019) considered a USVT estimator with tuning parameter $\tau \asymp \sqrt{n\rho}$, and showed that as long as $\rho \geq \frac{\log n}{n}$,

$$\ell(\widehat{M}_{\mathrm{USVT}}(\tau), M) \leq C\frac{\rho k}{n},$$

with high probability.[3]

The goal of this section is to show that the above rate cannot be improved by a polynomial-time algorithm. This is given by the following theorem.

THEOREM 6. *Suppose $2 \leq k \leq \sqrt{n}$ and $\rho \geq \frac{ck^2}{n}$ for some small $0 < c < 1$. Then for any $D \geq 1$, there exists a universal constant $c' > 0$ such that*

$$\inf_{\widehat{M}\in\mathbb{R}[A]^{n\times n}_{\leq D}} \sup_{M\in\mathcal{M}_{k,\rho}} \mathbb{E}(\ell(M, \widehat{M})) \geq \frac{c'\rho k}{nD^4}.$$

6.2. *Computational lower bound for biclustering.* Biclustering is another popular model of interest and has found a lot of applications in the literature (Hartigan (1972), Choi and Wolfe (2014), Rohe, Qin and Yu (2016), Chi, Allen and Baraniuk (2017), Mankad and Michailidis (2014)). Similar to SBM, many different problems have been considered for biclustering, such as recovery of the clustering structure, signal estimation and signal detection (detecting whether the signal matrix is zero or not). A line of early work has studied

[3]For completeness, an in-expectation bound is established in Appendix F.1 in the Supplementary Material (Luo and Gao (2024)) for a spectral algorithm.

the statistical and computational limits for detection or recovery in biclustering with one planted cluster (Balakrishnan et al. (2011), Kolar et al. (2011), Butucea and Ingster (2013), Butucea, Ingster and Suslina (2015), Ma and Wu (2015), Cai, Liang and Rakhlin (2017), Brennan, Bresler and Huleihel (2018), Schramm and Wein (2022)) and their extensions to a growing number of clusters have been considered in Chen and Xu (2016), Cai, Liang and Rakhlin (2017), Banks et al. (2018), Brennan and Bresler (2020), Dadon, Huleihel and Bendory (2024). In this section, we are more interested in the latter case.

Define the following parameter space of rectangular matrices with biclustering structure,

$$\mathcal{M}_{k_1,k_2} = \{M \in \mathbb{R}^{n_1 \times n_2} : M_{ij} = Q_{z_i z_j} \text{ for some } Q \in \mathbb{R}^{k_1 \times k_2}, z_1 \in [k_1]^{n_1}, z_2 \in [k_2]^{n_2}\}.$$

We observe $Y = M + E$, where $M \in \mathcal{M}_{k_1,k_2}$ and E has i.i.d. $N(0,1)$ entries. In this section, we are primarily interested in estimating M given Y and the loss of interest is $\ell(\widehat{M}, M) = \frac{1}{n_1 n_2} \sum_{i\in[n_1], j\in[n_2]} (\widehat{M}_{ij} - M_{ij})^2$. The minimax rate has been derived by Gao et al. (2016),

$$\inf_{\widehat{M}} \sup_{M \in \mathcal{M}_{k_1,k_2}} \mathbb{E}\big(\ell(\widehat{M}, M)\big) \asymp \frac{k_1 k_2}{n_1 n_2} + \frac{\log k_1}{n_2} + \frac{\log k_2}{n_1}, \tag{27}$$

and it is achieved by a constrained least-squares estimator that is computationally intractable. In terms of polynomial-time algorithms, a heuristic two-way extension of the Lloyd's algorithm has been proposed in Gao et al. (2016), but there is no theoretical guarantee. Let us instead consider a simple spectral algorithm,

$$\widehat{M} = \mathop{\arg\min}_{M:\mathrm{rank}(M) \le k_1 \wedge k_2} \|Y - M\|_{\mathrm{F}}^2.$$

Its theoretical guarantee is given by the following result.

PROPOSITION 2. *There exists* $C > 0$ *such that* $\sup_{M\in\mathcal{M}_{k_1,k_2}} \mathbb{E}(\ell(\widehat{M}, M)) \le C \frac{k_1 \wedge k_2}{n_1 \wedge n_2}$.

Compared with the minimax rate (27), the rate achieved by the spectral algorithm is not optimal. We will show that this rate is indeed the best one that can be achieved by a polynomial-time algorithm, at least in certain regimes of the problem. To this end, consider a subset of $\mathcal{M}_{k_1,k_2}$, denoted by $\mathcal{M}_{k_1,k_2,\lambda}$, whose definition is given by

$$\mathcal{M}_{k_1,k_2,\lambda} = \{M \in \mathbb{R}^{n_1 \times n_2} : M_{ij} = Q_{z_i z_j} \text{ for some } z_1 \in [k_1]^{n_1}, z_2 \in [k_2]^{n_2},$$
$$Q \in \mathbb{R}^{k_1 \times k_2} \text{ such that } Q_{ii} = \lambda \text{ for all } i \in [k_1 \wedge k_2] \text{ and } Q_{ij} = 0 \text{ otherwise}\}.$$

We also consider a prior distribution $\mathbb{P}_{\mathrm{BC}(\lambda)}$ supported on $\mathcal{M}_{k_1,k_2,\lambda}$. The sampling process $M \sim \mathbb{P}_{\mathrm{BC}(\lambda)}$ is given as follows: first, generate $z_1 \in [k_1]^n$, $z_2 \in [k_2]^n$ such that $(z_1)_i, (z_2)_j \overset{i.i.d.}{\sim}$ Unif$\{1, \dots, k_1 \wedge k_2\}$ independently for all $i \in [n_1]$ and $j \in [n_2]$; then let $M_{ij} = \lambda \mathbb{1}((z_1)_i = (z_2)_j)$. The following result gives a lower bound for the class of low-degree polynomial algorithms.

THEOREM 7. *For any* $0 < r < 1$ *and* $D \ge 1$, *if*

$$\lambda^2 \le \frac{r}{(D(D+1))^2} \min\left(1, \frac{k_1^2 \wedge k_2^2}{n_1 \vee n_2}\right) \tag{28}$$

holds, *then*

$$\inf_{\widehat{M} \in \mathbb{R}[Y]_{\le D}^{n_1 \times n_2}} \mathbb{E}_{Y, M\sim\mathbb{P}_{BC(\lambda)}}\big(\ell(\widehat{M}, M)\big) \ge \frac{\lambda^2}{k_1 \wedge k_2} - \left(\frac{\lambda^2}{k_1^2 \wedge k_2^2} + \frac{r(2-r)\lambda^2}{(1-r)^2(n_1 \vee n_2)}\right). \tag{29}$$

Since $\mathcal{M}_{k_1,k_2,\lambda} \subset \mathcal{M}_{k_1,k_2}$, the lower bound (29) is also valid for $\mathcal{M}_{k_1,k_2}$ under the SNR condition (28). To obtain a tight lower bound for $\mathcal{M}_{k_1,k_2}$, we can maximize the right-hand side of (29) under the constraint (28). This leads to the following biclustering lower bound.

COROLLARY 2. *Suppose $k_1 \wedge k_2 \geq 2$. For any $D \geq 1$, there exists a universal constant $c > 0$ such that*

$$\inf_{\widehat{M}\in\mathbb{R}[Y]_{\leq D}^{n_1\times n_2}} \sup_{M\in\mathcal{M}_{k_1,k_2}} \mathbb{E}(\ell(\widehat{M}, M)) \geq \frac{c}{D^4}\left(\frac{k_1\wedge k_2}{n_1\vee n_2} \wedge \frac{1}{k_1\wedge k_2}\right).$$

When $k_1 \wedge k_2 \leq \sqrt{n_1 \vee n_2}$ and $n_1 \asymp n_2$, the lower bound matches the rate of the spectral algorithm.

7. Proofs of the main results in Section 3.

In this section, we provide proofs for the low-degree polynomial lower bound of graphon estimation in SBM. We will use Proposition 1 to prove our results and one of the key parts there is to understand $\kappa_{\boldsymbol{\alpha}}(x, X)$. We will first introduce a few preliminary results regarding $\kappa_{\boldsymbol{\alpha}}(x, X)$, then prove Theorem 2 in Section 7.1 and finally prove Corollary 1 in Section 7.2.

Note that $\kappa_{\boldsymbol{\alpha}}(x, X)$ depends on the prior of X. Now, let us introduce the following *uniform SBM prior with fixed first vertex*.

DEFINITION 1. Consider a k-class SBM with n vertices. We say $X \in \mathbb{R}^{n(n-1)/2}$ is drawn from the uniform SBM prior with fixed first vertex and parameter $\lambda > 0$ if it is generated as follows: (1) generate a membership vector $z \in [k]^n$ such that $z_1 = 1$, $z_j \overset{i.i.d.}{\sim} \text{Unif}\{1, \dots, k\}$ for $j = 2, \dots, n$; (2) let $X = \text{vec}(\lambda Z_{ij} : i < j)$, where the symmetric matrix $Z \in \{0, 1\}^{n\times n}$ is the corresponding membership matrix of z. Here, the notation $\text{vec}(Z_{ij} : i < j)$ means the vectorization of the upper triangular matrix of Z by column.

Then we have the following bounds on $|\kappa_{\boldsymbol{\alpha}}(x, X)|$ when X is drawn from the prior in Definition 1.

PROPOSITION 3. *Suppose X is generated from the uniform SBM prior with fixed first vertex and parameter λ. Denote the membership vector of X as z and the first entry of X as x, that is, $x = \mathbb{1}(z_1 = z_2)$. Then for any multigraph $\boldsymbol{\alpha}$ on X with $|\boldsymbol{\alpha}| \geq 1$, we have*:

(i) *if $\boldsymbol{\alpha}$ is a disconnected or $\boldsymbol{\alpha}$ is connected but $2 \notin V(\boldsymbol{\alpha})$, then $\kappa_{\boldsymbol{\alpha}}(x, X) = 0$;*
(ii) *if $\boldsymbol{\alpha}$ is connected, $2 \in V(\boldsymbol{\alpha})$ and $1 \notin V(\boldsymbol{\alpha})$, then $\kappa_{\boldsymbol{\alpha}}(x, X) = 0$;*
(iii) *if $\boldsymbol{\alpha}$ is connected, $2 \in V(\boldsymbol{\alpha})$ and $1 \in V(\boldsymbol{\alpha})$, then $|\kappa_{\boldsymbol{\alpha}}(x, X)| \leq \lambda^{|\boldsymbol{\alpha}|+1}(1/k)^{|V(\boldsymbol{\alpha})|-1} \times (|\boldsymbol{\alpha}| + 1)^{|\boldsymbol{\alpha}|}$.*

PROOF. Throughout the proofs, we will view $\boldsymbol{\alpha}$ as a multigraph of n vertices.

Proof of (i). By (9), we know that $\kappa_{\boldsymbol{\alpha}}(x, X)$ is the joint cumulant of a group of random variables, say $\mathcal{G}$. For the case $\boldsymbol{\alpha}$ is disconnected, $\mathcal{G}$ could be divided into $\mathcal{G}_1$ and $\mathcal{G}_2$ and $\mathcal{G}_1$ and $\mathcal{G}_2$ are independent of each other. Thus, by Proposition 1 in Appendix B.1 in the Supplementary Material (Luo and Gao (2024)), $\kappa_{\boldsymbol{\alpha}}(x, X)$ is zero. Similarly, for the case $\boldsymbol{\alpha}$ is connected but $2 \notin V(\boldsymbol{\alpha})$, we know in the prior for X, z_1 is known and fixed, so if $2 \notin V(\boldsymbol{\alpha})$, x will be independent of X. By the same argument, $\kappa_{\boldsymbol{\alpha}}(x, X)$ will be zero.

Proof of (ii). First, for any connected $\boldsymbol{\alpha}$, we have

$$\begin{aligned}\mathbb{E}[X^{\boldsymbol{\alpha}}] &= \lambda^{|\boldsymbol{\alpha}|}\mathbb{P}(\text{all vertices in } V(\boldsymbol{\alpha}) \text{ are in the same community}) \\ &= \lambda^{|\boldsymbol{\alpha}|} \cdot \left(\frac{1}{k}\right)^{|V(\boldsymbol{\alpha})|-1}, \\ \mathbb{E}[xX^{\boldsymbol{\alpha}}] &= \lambda^{|\boldsymbol{\alpha}|+1}\mathbb{P}(\text{all vertices in } V(\boldsymbol{\alpha}) \cup \{1,2\} \text{ are in the same community}) \\ &= \lambda^{|\boldsymbol{\alpha}|+1}\left(\frac{1}{k}\right)^{|V(\boldsymbol{\alpha})\cup\{1,2\}|-1}.\end{aligned} \tag{30}$$

Next, we prove the claim by induction. When $|\boldsymbol{\alpha}| = 0$, $\kappa_0(x, X) = \mathbb{E}(x) = \frac{\lambda}{k}$. Then, for $\boldsymbol{\alpha}$ such that $|\boldsymbol{\alpha}| = 1$, $2 \in V(\boldsymbol{\alpha})$ and $1 \notin V(\boldsymbol{\alpha})$, we have

$$\kappa_{\boldsymbol{\alpha}}(x, X) \overset{(8)}{=} \mathbb{E}[xX^{\boldsymbol{\alpha}}] - \kappa_0(x, X)\mathbb{E}[X^{\boldsymbol{\alpha}}] \overset{(30)}{=} \lambda^{|\boldsymbol{\alpha}|+1}\left(\frac{1}{k}\right)^{|V(\boldsymbol{\alpha})|} - \frac{\lambda}{k}\lambda^{|\boldsymbol{\alpha}|}\left(\frac{1}{k}\right)^{|V(\boldsymbol{\alpha})|-1} = 0.$$

Now, assume that given any $t \geq 2$ and any $\boldsymbol{\alpha}$ with $2 \in V(\boldsymbol{\alpha})$, $1 \notin V(\boldsymbol{\alpha})$ and $|\boldsymbol{\alpha}| < t$, $\kappa_{\boldsymbol{\alpha}}(x, X) = 0$. Then for any such $\boldsymbol{\alpha}$ with $|\boldsymbol{\alpha}| = t$, we have

$$\begin{aligned}\kappa_{\boldsymbol{\alpha}}(x, X) &\overset{(8)}{=} \mathbb{E}(xX^{\boldsymbol{\alpha}}) - \sum_{0 \leq \boldsymbol{\beta} \lneq \boldsymbol{\alpha}} \kappa_{\boldsymbol{\beta}}(x, X)\binom{\boldsymbol{\alpha}}{\boldsymbol{\beta}}\mathbb{E}[X^{\boldsymbol{\alpha}-\boldsymbol{\beta}}] \\ &\overset{(a)}{=} \mathbb{E}(xX^{\boldsymbol{\alpha}}) - \kappa_0(x, X)\mathbb{E}[X^{\boldsymbol{\alpha}}] = \lambda^{|\boldsymbol{\alpha}|+1}\left(\frac{1}{k}\right)^{|V(\boldsymbol{\alpha})|} - \frac{\lambda}{k}\lambda^{|\boldsymbol{\alpha}|}\left(\frac{1}{k}\right)^{|V(\boldsymbol{\alpha})|-1} = 0,\end{aligned}$$

where (a) is because for any $\boldsymbol{\beta}$ such that $|\boldsymbol{\beta}| \geq 1$, $1 \notin V(\boldsymbol{\beta})$ since $\boldsymbol{\beta}$ is a subgraph of $\boldsymbol{\alpha}$, and thus $\kappa_{\boldsymbol{\beta}}(x, X) = 0$ for either the case $2 \notin V(\boldsymbol{\beta})$ by the result we have proved in part (i) and the case $2 \in V(\boldsymbol{\beta})$ by the induction assumption. This finishes the induction, and we have that for any $\boldsymbol{\alpha}$ such that $|\boldsymbol{\alpha}| \geq 1$, $2 \in V(\boldsymbol{\alpha})$ and $1 \notin V(\boldsymbol{\alpha})$, $\kappa_{\boldsymbol{\alpha}}(x, X) = 0$.

Proof of (iii). First, for any connected subgraph $\boldsymbol{\beta}$ of $\boldsymbol{\alpha}$,

$$\begin{aligned}\mathbb{E}[X^{\boldsymbol{\alpha}-\boldsymbol{\beta}}] &= \lambda^{|\boldsymbol{\alpha}-\boldsymbol{\beta}|}\mathbb{P}(\text{each connected component in } \boldsymbol{\alpha}-\boldsymbol{\beta} \text{ belongs to the same community}) \\ &= \lambda^{|\boldsymbol{\alpha}-\boldsymbol{\beta}|}\left(\frac{1}{k}\right)^{|V(\boldsymbol{\alpha}-\boldsymbol{\beta})|-\mathcal{C}(\boldsymbol{\alpha}-\boldsymbol{\beta})},\end{aligned} \tag{31}$$

where $\mathcal{C}(\boldsymbol{\alpha}-\boldsymbol{\beta})$ denotes the number of connected components in $\boldsymbol{\alpha}-\boldsymbol{\beta}$.

Next, we prove the claim by induction. Recall that when $|\boldsymbol{\alpha}| = 0$, $\kappa_0(x, X) = \frac{\lambda}{k}$. Then, for $\boldsymbol{\alpha}$ such that $|\boldsymbol{\alpha}| = 1$, $2 \in V(\boldsymbol{\alpha})$ and $1 \in V(\boldsymbol{\alpha})$, we have

$$\begin{aligned}\kappa_{\boldsymbol{\alpha}}(x, X) &\overset{(8)}{=} \mathbb{E}[xX^{\boldsymbol{\alpha}}] - \kappa_0(x, X)\mathbb{E}[X^{\boldsymbol{\alpha}}] \overset{(30)}{=} \lambda^{|\boldsymbol{\alpha}|+1}\left(\frac{1}{k}\right)^{|V(\boldsymbol{\alpha})|-1} - \frac{\lambda}{k}\lambda^{|\boldsymbol{\alpha}|}\left(\frac{1}{k}\right)^{|V(\boldsymbol{\alpha})|-1} \\ &= \lambda^{|\boldsymbol{\alpha}|+1}\left(\frac{1}{k}\right)^{|V(\boldsymbol{\alpha})|-1} - \frac{\lambda}{k}\lambda^{|\boldsymbol{\alpha}|}\left(\frac{1}{k}\right)^{|V(\boldsymbol{\alpha})|-1} = \lambda^{|\boldsymbol{\alpha}|+1}\left(\frac{1}{k}\right)^{|V(\boldsymbol{\alpha})|-1}(1 - 1/k),\end{aligned}$$

thus $|\kappa_{\boldsymbol{\alpha}}(x, X)| \leq \lambda^{|\boldsymbol{\alpha}|+1}(1/k)^{|V(\boldsymbol{\alpha})|-1}(|\boldsymbol{\alpha}| + 1)^{|\boldsymbol{\alpha}|}$ holds for $|\boldsymbol{\alpha}| = 1$.

Now, assume that given any $t \geq 2$ and any $\boldsymbol{\alpha}$ with $2 \in V(\boldsymbol{\alpha})$, $1 \in V(\boldsymbol{\alpha})$ and $|\boldsymbol{\alpha}| < t$, $|\kappa_{\boldsymbol{\alpha}}(x, X)| \leq \lambda^{|\boldsymbol{\alpha}|+1}(1/k)^{|V(\boldsymbol{\alpha})|-1}(|\boldsymbol{\alpha}| + 1)^{|\boldsymbol{\alpha}|}$. Then for any such $\boldsymbol{\alpha}$ with $|\boldsymbol{\alpha}| = t$, we have

$$\begin{aligned}|\kappa_{\boldsymbol{\alpha}}(x, X)| &\overset{(8)}{=} \left|\mathbb{E}(xX^{\boldsymbol{\alpha}}) - \sum_{0 \leq \boldsymbol{\beta} \lneq \boldsymbol{\alpha}} \kappa_{\boldsymbol{\beta}}(x, X)\binom{\boldsymbol{\alpha}}{\boldsymbol{\beta}}\mathbb{E}[X^{\boldsymbol{\alpha}-\boldsymbol{\beta}}]\right| \\ &\leq |\mathbb{E}(xX^{\boldsymbol{\alpha}})| + \sum_{0 \leq \boldsymbol{\beta} \lneq \boldsymbol{\alpha}} |\kappa_{\boldsymbol{\beta}}(x, X)|\binom{\boldsymbol{\alpha}}{\boldsymbol{\beta}}\mathbb{E}[X^{\boldsymbol{\alpha}-\boldsymbol{\beta}}]\end{aligned} \tag{32}$$

$$
\begin{aligned}
&\overset{\text{Part }(i)}{=} \left|\mathbb{E}(xX^{\boldsymbol{\alpha}})\right| + \sum_{0\le\boldsymbol{\beta}\lneq\boldsymbol{\alpha}:\boldsymbol{\beta}\text{ is connected}} \left|\kappa_{\boldsymbol{\beta}}(x,X)\right|\binom{\boldsymbol{\alpha}}{\boldsymbol{\beta}}\mathbb{E}\left[X^{\boldsymbol{\alpha}-\boldsymbol{\beta}}\right] \\
&\overset{(30),(31)}{=} \lambda^{|\boldsymbol{\alpha}|+1}\left(\frac{1}{k}\right)^{|V(\boldsymbol{\alpha})|-1} \\
&\quad + \sum_{0\le\boldsymbol{\beta}\lneq\boldsymbol{\alpha}:\boldsymbol{\beta}\text{ is connected}} \left|\kappa_{\boldsymbol{\beta}}(x,X)\right|\binom{\boldsymbol{\alpha}}{\boldsymbol{\beta}}\lambda^{|\boldsymbol{\alpha}-\boldsymbol{\beta}|}\left(\frac{1}{k}\right)^{|V(\boldsymbol{\alpha}-\boldsymbol{\beta})|-\mathcal{C}(\boldsymbol{\alpha}-\boldsymbol{\beta})} \\
&\overset{(a)}{=} \lambda^{|\boldsymbol{\alpha}|+1}\left(\frac{1}{k}\right)^{|V(\boldsymbol{\alpha})|-1} + \left|\kappa_0(x,X)\right|\lambda^{|\boldsymbol{\alpha}|}\left(\frac{1}{k}\right)^{|V(\boldsymbol{\alpha})|-1} \\
&\quad + \sum_{\substack{0\lneq\boldsymbol{\beta}\lneq\boldsymbol{\alpha},\\ \boldsymbol{\beta}\text{ is connected },\\ 1\in V(\boldsymbol{\beta}),2\in V(\boldsymbol{\beta})}} \left|\kappa_{\boldsymbol{\beta}}(x,X)\right|\binom{\boldsymbol{\alpha}}{\boldsymbol{\beta}}\lambda^{|\boldsymbol{\alpha}-\boldsymbol{\beta}|}\left(\frac{1}{k}\right)^{|V(\boldsymbol{\alpha}-\boldsymbol{\beta})|-\mathcal{C}(\boldsymbol{\alpha}-\boldsymbol{\beta})} \\
&\overset{(b)}{=} \lambda^{|\boldsymbol{\alpha}|+1}\left(\frac{1}{k}\right)^{|V(\boldsymbol{\alpha})|-1} + \lambda^{|\boldsymbol{\alpha}|+1}\left(\frac{1}{k}\right)^{|V(\boldsymbol{\alpha})|} \\
&\quad + \sum_{\substack{0\lneq\boldsymbol{\beta}\lneq\boldsymbol{\alpha},\\ \boldsymbol{\beta}\text{ is connected },\\ 1\in V(\boldsymbol{\beta}),2\in V(\boldsymbol{\beta})}} \lambda^{|\boldsymbol{\beta}|+1}(1/k)^{|V(\boldsymbol{\beta})|-1}\left(|\boldsymbol{\beta}|+1\right)^{|\boldsymbol{\beta}|}\binom{\boldsymbol{\alpha}}{\boldsymbol{\beta}}\lambda^{|\boldsymbol{\alpha}-\boldsymbol{\beta}|} \\
&\quad \times\left(\frac{1}{k}\right)^{|V(\boldsymbol{\alpha}-\boldsymbol{\beta})|-\mathcal{C}(\boldsymbol{\alpha}-\boldsymbol{\beta})},
\end{aligned}
$$

where in (a), we separate the term $\boldsymbol{\beta}=0$ in the summation and then use the results proved in (i)–(ii) of this proposition; (b) is because $\kappa_0(x,X)=\frac{\lambda}{k}$ and by the induction assumption.

Next,

$$
\begin{aligned}
\left|\kappa_{\boldsymbol{\alpha}}(x,X)\right| &\le \lambda^{|\boldsymbol{\alpha}|+1}\left(\frac{1}{k}\right)^{|V(\boldsymbol{\alpha})|-1} + \lambda^{|\boldsymbol{\alpha}|+1}\left(\frac{1}{k}\right)^{|V(\boldsymbol{\alpha})|} \\
&\quad + \sum_{\substack{0\lneq\boldsymbol{\beta}\lneq\boldsymbol{\alpha},\\ \boldsymbol{\beta}\text{ is connected },\\ 1\in V(\boldsymbol{\beta}),2\in V(\boldsymbol{\beta})}} \lambda^{|\boldsymbol{\beta}|+1}(1/k)^{|V(\boldsymbol{\beta})|-1}\left(|\boldsymbol{\beta}|+1\right)^{|\boldsymbol{\beta}|}\binom{\boldsymbol{\alpha}}{\boldsymbol{\beta}}\lambda^{|\boldsymbol{\alpha}-\boldsymbol{\beta}|} \\
&\quad \times\left(\frac{1}{k}\right)^{|V(\boldsymbol{\alpha}-\boldsymbol{\beta})|-\mathcal{C}(\boldsymbol{\alpha}-\boldsymbol{\beta})} \\
&\overset{(a)}{\le} 2\lambda^{|\boldsymbol{\alpha}|+1}\left(\frac{1}{k}\right)^{|V(\boldsymbol{\alpha})|-1} + \lambda^{|\boldsymbol{\alpha}|+1}\left(\frac{1}{k}\right)^{|V(\boldsymbol{\alpha})|-1} \sum_{\substack{0\lneq\boldsymbol{\beta}\lneq\boldsymbol{\alpha},\\ \boldsymbol{\beta}\text{ is connected },\\ 1\in V(\boldsymbol{\beta}),2\in V(\boldsymbol{\beta})}} \left(|\boldsymbol{\beta}|+1\right)^{|\boldsymbol{\beta}|}\binom{\boldsymbol{\alpha}}{\boldsymbol{\beta}} \\
&\le \lambda^{|\boldsymbol{\alpha}|+1}\left(\frac{1}{k}\right)^{|V(\boldsymbol{\alpha})|-1}\left(2+\sum_{0\lneq\boldsymbol{\beta}\lneq\boldsymbol{\alpha}}\left(|\boldsymbol{\beta}|+1\right)^{|\boldsymbol{\beta}|}\binom{\boldsymbol{\alpha}}{\boldsymbol{\beta}}\right) \\
&\overset{(b)}{=} \lambda^{|\boldsymbol{\alpha}|+1}\left(\frac{1}{k}\right)^{|V(\boldsymbol{\alpha})|-1}\left(2+\sum_{\ell=1}^{|\boldsymbol{\alpha}|-1}(\ell+1)^{\ell}\binom{|\boldsymbol{\alpha}|}{\ell}\right) \qquad (33)
\end{aligned}
$$

$$\leq \lambda^{|\boldsymbol{\alpha}|+1}\left(\frac{1}{k}\right)^{|V(\boldsymbol{\alpha})|-1}\left(2+\sum_{\ell=1}^{|\boldsymbol{\alpha}|-1}|\boldsymbol{\alpha}|^{\ell}\binom{|\boldsymbol{\alpha}|}{\ell}\right)$$

$$\leq \lambda^{|\boldsymbol{\alpha}|+1}\left(\frac{1}{k}\right)^{|V(\boldsymbol{\alpha})|-1}\left(\sum_{\ell=0}^{|\boldsymbol{\alpha}|}|\boldsymbol{\alpha}|^{\ell}\binom{|\boldsymbol{\alpha}|}{\ell}\right)$$

$$= \lambda^{|\boldsymbol{\alpha}|+1}\left(\frac{1}{k}\right)^{|V(\boldsymbol{\alpha})|-1}\left(|\boldsymbol{\alpha}|+1\right)^{|\boldsymbol{\alpha}|},$$

where (a) is due to the following Lemma 1 and in (b) we use the fact $\sum_{\boldsymbol{\beta}:|\boldsymbol{\beta}|=\ell}\binom{\boldsymbol{\alpha}}{\boldsymbol{\beta}}=\binom{|\boldsymbol{\alpha}|}{\ell}$ for any $\ell \leq |\boldsymbol{\alpha}|$.

LEMMA 1. *Given any connected multigraph $\boldsymbol{\alpha}$, suppose $\boldsymbol{\beta}$ is a connected subgraph of $\boldsymbol{\alpha}$, then*

$$\left|V(\boldsymbol{\alpha}-\boldsymbol{\beta})\right| + \left|V(\boldsymbol{\beta})\right| - \mathcal{C}(\boldsymbol{\alpha}-\boldsymbol{\beta}) \geq \left|V(\boldsymbol{\alpha})\right|,$$

where $\mathcal{C}(\boldsymbol{\alpha}-\boldsymbol{\beta})$ denotes the number of connected component in the graph $\boldsymbol{\alpha}-\boldsymbol{\beta}$.

PROOF. First, it is clear that $V(\boldsymbol{\alpha}-\boldsymbol{\beta}) \cup V(\boldsymbol{\beta}) \supseteq V(\boldsymbol{\alpha})$. Since both $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$ are connected multigraphs, for each connected component in $\boldsymbol{\alpha}-\boldsymbol{\beta}$, it must have at least a common vertex with $\boldsymbol{\beta}$. Moreover, that common vertex is counted twice in computing $|V(\boldsymbol{\alpha}-\boldsymbol{\beta})| + |V(\boldsymbol{\beta})|$. This completes the proof. □

This finishes the induction and the proof of this proposition. □

The next lemma counts the number of connected $\boldsymbol{\alpha}$ in (iii) of Proposition 3 such that $\kappa_{\boldsymbol{\alpha}}(x, X)$ is nonzero.

LEMMA 2. *Given any $d \geq 1$, $0 \leq h \leq d-1$, the number of connected $\boldsymbol{\alpha}$ such that $1 \in V(\boldsymbol{\alpha})$, $2 \in V(\boldsymbol{\alpha})$, $|\boldsymbol{\alpha}| = d$ and $|V(\boldsymbol{\alpha})| = d+1-h$ is at most $n^{d-h-1}d^{d+h}$.*

PROOF. We view $\boldsymbol{\alpha}$ as a multigraph on $[n]$ and count the number of ways to construct such $\boldsymbol{\alpha}$. The counting strategy is the following: we start with adding Vertex 2 to $\boldsymbol{\alpha}$ and then add $(d-h)$ edges such that for each edge there, it will introduce a new vertex; then we add the rest of h edges on these existing vertices. In the first stage above, we can also count different cases by considering when will Vertex 1 be introduced in adding new vertices.

- If Vertex 1 is the first vertex to be added after Vertex 2, then the number of such choices of $\boldsymbol{\alpha}$ is at most $(nd)^{d-h-1}(d^2)^h$. Here, $(nd)^{d-h-1}$ is because for each of the rest of $d-h-1$ edges, the number of choices for the starting vertex is at most d since there are at most $(d+1)$ vertices in $\boldsymbol{\alpha}$ and the number of choices for a newly introduced vertex is at most n. $(d^2)^h$ comes from that in the second stage, the choice of each extra edge is at most $\binom{(d+1)}{2} = (d+1)d/2 \leq d^2$.
- By the same counting strategy, if Vertex 1 is the second vertex to be added in the first stage, then the number of such choices of $\boldsymbol{\alpha}$ is at most $(nd)^{d-h-1}(d^2)^h$.
- $\cdots$
- If Vertex 1 is the $(d-h)$th vertex to be added in the first stage, then the number of such choices of $\boldsymbol{\alpha}$ is at most $(nd)^{d-h-1}(d^2)^h$.

By adding them together, the number of choices of connected $\boldsymbol{\alpha}$ such that $1 \in V(\boldsymbol{\alpha})$, $2 \in V(\boldsymbol{\alpha})$, $|\boldsymbol{\alpha}| = d$ and $|V(\boldsymbol{\alpha})| = d + 1 - h$ is at most

$$(d-h)(nd)^{d-h-1}(d^2)^h \le d(nd)^{d-h-1}(d^2)^h = n^{d-h-1}d^{d+h}. \qquad \square$$

In the following Proposition 4, we bound $\sum_{\boldsymbol{\alpha}\in\mathbb{N}^N, 0\le|\boldsymbol{\alpha}|\le D} \frac{\kappa_{\boldsymbol{\alpha}}^2(x,X)}{\boldsymbol{\alpha}!}$ when X is generated from the uniform SBM prior with fixed first vertex.

PROPOSITION 4. *Under the same setting as in Proposition 3, for any $D \ge 1$, we have*

$$\sum_{\boldsymbol{\alpha}\in\mathbb{N}^N, 0\le|\boldsymbol{\alpha}|\le D} \frac{\kappa_{\boldsymbol{\alpha}}^2(x,X)}{\boldsymbol{\alpha}!} \le \frac{\lambda^2}{k^2} - \frac{\lambda^2}{n} + \frac{\lambda^2}{n}\sum_{h=0}^{d}\left(D^2(D+1)^2\lambda^2\right)^h \sum_{d=h}^{D}\left(D(D+1)^2\frac{n\lambda^2}{k^2}\right)^{d-h}.$$

In particular, for any $0 < r < 1$, if $\lambda^2 \le \frac{r}{(D(D+1))^2}\min(1, \frac{k^2}{n})$, then we have

$$\sum_{\boldsymbol{\alpha}\in\mathbb{N}^N, 0\le|\boldsymbol{\alpha}|\le D} \frac{\kappa_{\boldsymbol{\alpha}}^2(x,X)}{\boldsymbol{\alpha}!} \le \frac{\lambda^2}{k^2} + \frac{r(2-r)\lambda^2}{(1-r)^2 n}.$$

PROOF. First, we have $\kappa_0(x,X) = \mathbb{E}(x) = \lambda/k$. Then

$$\begin{aligned}
\sum_{\boldsymbol{\alpha}\in\mathbb{N}^N, 0\le|\boldsymbol{\alpha}|\le D} \frac{\kappa_{\boldsymbol{\alpha}}^2(x,X)}{\boldsymbol{\alpha}!} &\le \sum_{\boldsymbol{\alpha}\in\mathbb{N}^N, 0\le|\boldsymbol{\alpha}|\le D} \kappa_{\boldsymbol{\alpha}}^2(x,X) \\
&\overset{\text{Proposition 3}(i)(ii)}{=} \kappa_0^2(x,X) + \sum_{\substack{\boldsymbol{\alpha}\in\mathbb{N}^N, 1\le|\boldsymbol{\alpha}|\le D,\\ \boldsymbol{\alpha}\text{ connected }, 1\in V(\boldsymbol{\alpha}), 2\in V(\boldsymbol{\alpha})}} \kappa_{\boldsymbol{\alpha}}^2(x,X) \\
&\overset{\text{Proposition 3}(iii),\text{Lemma 2}}{\le} \frac{\lambda^2}{k^2} + \sum_{d=1}^{D}\sum_{h=0}^{d-1} n^{d-h-1}d^{d+h} \\
&\quad \times \left(\lambda^{d+1}(1/k)^{d-h}(d+1)^d\right)^2 \\
&= \frac{\lambda^2}{k^2} + \frac{\lambda^2}{n}\sum_{d=1}^{D}\sum_{h=0}^{d-1}\left(\frac{nd(d+1)^2\lambda^2}{k^2}\right)^d\left(\frac{dk^2}{n}\right)^h \\
&\le \frac{\lambda^2}{k^2} + \frac{\lambda^2}{n}\sum_{d=1}^{D}\sum_{h=0}^{d-1}\left(\frac{nD(D+1)^2\lambda^2}{k^2}\right)^d\left(\frac{Dk^2}{n}\right)^h \\
&\le \frac{\lambda^2}{k^2} - \frac{\lambda^2}{n} + \frac{\lambda^2}{n}\sum_{d=0}^{D}\sum_{h=0}^{d}\left(\frac{nD(D+1)^2\lambda^2}{k^2}\right)^d\left(\frac{Dk^2}{n}\right)^h \\
&= \frac{\lambda^2}{k^2} - \frac{\lambda^2}{n} + \frac{\lambda^2}{n}\sum_{h=0}^{d}\left(D^2(D+1)^2\lambda^2\right)^h\sum_{d=h}^{D}\left(D(D+1)^2\frac{n\lambda^2}{k^2}\right)^{d-h}.
\end{aligned}$$

This shows the first conclusion.

For any $0 < r < 1$ and $D \geq 1$, if $\lambda^2 \leq \frac{r}{(D(D+1))^2} \min(1, \frac{k^2}{n})$, by the above result, we have

$$\begin{aligned}\sum_{\boldsymbol{\alpha}\in\mathbb{N}^N, 0\leq|\boldsymbol{\alpha}|\leq D} \frac{\kappa_{\boldsymbol{\alpha}}^2(x, X)}{\boldsymbol{\alpha}!} &\leq \frac{\lambda^2}{k^2} - \frac{\lambda^2}{n} + \frac{\lambda^2}{n} \sum_{h=0}^{d} \big(D^2(D+1)^2\lambda^2\big)^h \sum_{d=h}^{D} \bigg(D(D+1)^2 \frac{n\lambda^2}{k^2}\bigg)^{d-h} \\ &\leq \frac{\lambda^2}{k^2} - \frac{\lambda^2}{n} + \frac{\lambda^2}{n} \sum_{h=0}^{d} r^h \sum_{d=h}^{D} r^{d-h} \leq \frac{\lambda^2}{k^2} - \frac{\lambda^2}{n} + \frac{\lambda^2}{n} \frac{1}{(1-r)^2} \\ &= \frac{\lambda^2}{k^2} + \frac{r(2-r)\lambda^2}{(1-r)^2 n}.\end{aligned}$$

This completes the proof of this proposition. □

7.1. *Proof of Theorem* 2. PROOF. First, since M is drawn uniformly at random from $\mathcal{M}_{k,p,q}$, by symmetry, we have

$$\begin{aligned}&\inf_{\widehat{M}\in\mathbb{R}[A]^{n\times n}_{\leq D}} \mathbb{E}_{A,M\sim\mathbb{P}_{\mathrm{SBM}(p,q)}}\big[\ell(\widehat{M}, M)\big] \\ &\quad = \inf_{g\in\mathbb{R}[A]_{\leq D}} \mathbb{E}_{A,M\sim\mathbb{P}_{\mathrm{SBM}(p,q)}}\big[\big(g(A) - M_{12}\big)^2\big] \\ &\quad = \inf_{g\in\mathbb{R}[A]_{\leq D}} \sum_{j=1}^{k} \mathbb{E}_{A,M\sim\mathbb{P}_{\mathrm{SBM}(p,q)}}\big[\big(g(A) - M_{12}\big)^2 \big| (z_M)_1 = j\big]\mathbb{P}\big((z_M)_1 = j\big) \qquad (34) \\ &\quad = \inf_{g\in\mathbb{R}[A]_{\leq D}} \mathbb{E}_{A,M\sim\mathbb{P}_{\mathrm{SBM}(p,q)}}\big[\big(g(A) - M_{12}\big)^2 \big| (z_M)_1 = 1\big] \\ &\quad = \inf_{g\in\mathbb{R}[A]_{\leq D}} \mathbb{E}_{A,M\sim\mathbb{P}'_{\mathrm{SBM}(p,q)}}\big[\big(g(A) - M_{12}\big)^2\big],\end{aligned}$$

where $\mathbb{P}'_{\mathrm{SBM}(p,q)}$ is the restriction of $\mathbb{P}_{\mathrm{SBM}(p,q)}$ on $\mathcal{M}_{k,p,q}$ such that $(z_M)_1 = 1$.

The graphon estimation problem in SBM fits in the general binary observation model described in Section 2.1. Thus, we can apply the results in Proposition 1. Following the notation in Section 2.1, in our context, we have $x = M_{12}$, $X = \mathrm{vec}(M_{ij} : i < j)$ encodes the upper triangular entries of M and $Y = \mathrm{vec}(A_{ij} : i < j)$ encodes the upper triangular entries of A. Thus, $N = n(n-1)/2$ and the law of X is supported on $[q, p]$. By Proposition 1, we have

$$\begin{aligned}&\inf_{g\in\mathbb{R}[A]_{\leq D}} \mathbb{E}_{A,M\sim\mathbb{P}'_{\mathrm{SBM}(p,q)}}\big[\big(g(A) - M_{12}\big)^2\big] \\ &\quad = \mathbb{E}_{A,M\sim\mathbb{P}'_{\mathrm{SBM}(p,q)}}\big(M_{12}^2\big) - \mathrm{Corr}_{\leq D}^2 \qquad (35) \\ &\quad \geq \mathbb{E}_{A,M\sim\mathbb{P}'_{\mathrm{SBM}(p,q)}}\big(M_{12}^2\big) - \sum_{\boldsymbol{\alpha}\in\{0,1\}^N, 0\leq|\boldsymbol{\alpha}|\leq D} \frac{\kappa_{\boldsymbol{\alpha}}^2(M_{12}, X)}{(q(1-p))^{|\boldsymbol{\alpha}|}},\end{aligned}$$

where $\kappa_{\boldsymbol{\alpha}}(M_{12}, X)$ is recursively defined as

$$\kappa_0(M_{12}, X) = \mathbb{E}_{A,M\sim\mathbb{P}'_{\mathrm{SBM}(p,q)}}[M_{12}] = p/k + (1-1/k)q = q + (p-q)/k;$$

$$\kappa_{\boldsymbol{\alpha}}(M_{12}, X) = \mathbb{E}_{A,M\sim\mathbb{P}'_{\mathrm{SBM}(p,q)}}\big[M_{12}X^{\boldsymbol{\alpha}}\big] - \sum_{0\leq\boldsymbol{\beta}\lneq\boldsymbol{\alpha}} \kappa_{\boldsymbol{\beta}}(x, X)\binom{\boldsymbol{\alpha}}{\boldsymbol{\beta}}\mathbb{E}\big[X^{\boldsymbol{\alpha}-\boldsymbol{\beta}}\big] \qquad (36)$$

$$\overset{(a)}{=} \mathbb{E}_{A,M\sim\mathbb{P}'_{\mathrm{SBM}(p,q)}}\left[M_{12}X^{\boldsymbol{\alpha}}\right] - \sum_{0\leq\boldsymbol{\beta}\lneq\boldsymbol{\alpha}} \kappa_{\boldsymbol{\beta}}(x, X)\mathbb{E}\left[X^{\boldsymbol{\alpha}-\boldsymbol{\beta}}\right] \quad \text{for } \boldsymbol{\alpha} \text{ such that } |\boldsymbol{\alpha}| \geq 1,$$

where in (a), we use the fact $\boldsymbol{\alpha} \in \{0,1\}^N$ so that $\binom{\boldsymbol{\alpha}}{\boldsymbol{\beta}} = 1$.

Directly computing $\kappa_{\boldsymbol{\alpha}}(M_{12}, X)$ is complicated. Here, we do a transformation and let $\overline{X} = (X-q)/\sqrt{q(1-p)}$, $\overline{M}_{12} = (M_{12}-q)/\sqrt{q(1-p)}$. By the interpretation of $\kappa_{\boldsymbol{\alpha}}(x, X)$ in (9), we have

$$\begin{aligned}\kappa_0(\overline{M}_{12}, \overline{X}) &\overset{(a)}{=} \kappa_0\left(\frac{M_{12}}{\sqrt{q(1-p)}}, X\right) - \frac{q}{\sqrt{q(1-p)}} \\ &= \frac{1}{\sqrt{q(1-p)}}\kappa_0(M_{12}, X) - \frac{q}{\sqrt{q(1-p)}} \\ &\overset{(36)}{=} \frac{p-q}{k\sqrt{q(1-p)}},\end{aligned} \tag{37}$$

where (a) is by Proposition 2 in Appendix B.1 in the Supplementary Material (Luo and Gao (2024)). For any $\boldsymbol{\alpha}$ such that $|\boldsymbol{\alpha}| \geq 1$, by Proposition 2, we have

$$\kappa_{\boldsymbol{\alpha}}(\overline{M}_{12}, \overline{X}) = \frac{1}{(q(1-p))^{(|\boldsymbol{\alpha}|+1)/2}} \cdot \kappa_{\boldsymbol{\alpha}}(M_{12}, X). \tag{38}$$

By plugging (37) and (38) into (34) and (35), we have

$$\begin{aligned}&\inf_{g\in\mathbb{R}[A]_{\leq D}} \mathbb{E}_{A,M\sim\mathbb{P}'_{\mathrm{SBM}(p,q)}}\left[\left(g(A) - M_{12}\right)^2\right] \\ &\geq \mathbb{E}_{A,M\sim\mathbb{P}'_{\mathrm{SBM}(p,q)}}\left(M_{12}^2\right) - \sum_{\boldsymbol{\alpha}\in\{0,1\}^N, 0\leq|\boldsymbol{\alpha}|\leq D} \frac{\kappa_{\boldsymbol{\alpha}}^2(M_{12}, X)}{(q(1-p))^{|\boldsymbol{\alpha}|}} \\ &= q^2 + \frac{p^2-q^2}{k} - \kappa_0^2(M_{12}, X) - \sum_{\boldsymbol{\alpha}\in\{0,1\}^N, 0\lneq|\boldsymbol{\alpha}|\leq D} \frac{\kappa_{\boldsymbol{\alpha}}^2(M_{12}, X)}{(q(1-p))^{|\boldsymbol{\alpha}|}} \\ &\overset{(36),(38)}{=} q^2 + \frac{p^2-q^2}{k} - \left(q + (p-q)/k\right)^2 + q(1-p)\kappa_0^2(\overline{M}_{12}, \overline{X}) \\ &\quad - \sum_{\boldsymbol{\alpha}\in\{0,1\}^N, 0\leq|\boldsymbol{\alpha}|\leq D} q(1-p)\kappa_{\boldsymbol{\alpha}}^2(\overline{M}_{12}, \overline{X}) \\ &\overset{(37)}{=} \frac{(p-q)^2}{k} - q(1-p) \sum_{\boldsymbol{\alpha}\in\{0,1\}^N, 0\leq|\boldsymbol{\alpha}|\leq D} \kappa_{\boldsymbol{\alpha}}^2(\overline{M}_{12}, \overline{X}).\end{aligned} \tag{39}$$

Recall that Z_M is the membership matrix associated with M. Since X represents the upper triangular entries of $q\mathbf{1}_n\mathbf{1}_n^\top + (p-q)Z_M$ where $\mathbf{1}_n$ represents an all 1 vector, after the transformation, $\overline{X}$ encodes upper triangular entries of $\frac{(p-q)Z_M}{\sqrt{q(1-p)}}$ and $\overline{M}_{12}$ is the first entry of $\overline{X}$.

Notice that $\mathbb{P}'_{\mathrm{SBM}(p,q)}$ is exactly the uniform SBM prior with fixed first vertex defined in Definition 1 with $\lambda = \frac{(p-q)}{\sqrt{q(1-p)}}$. By Proposition 4, we have

$$\begin{aligned}&\inf_{\widehat{M}\in\mathbb{R}[A]_{\leq D}^{n\times n}} \mathbb{E}_{A,M\sim\mathbb{P}_{\mathrm{SBM}(p,q)}}\left[\ell(\widehat{M}, M)\right] \\ &\overset{(34)}{=} \inf_{g\in\mathbb{R}[A]_{\leq D}} \mathbb{E}_{A,M\sim\mathbb{P}'_{\mathrm{SBM}(p,q)}}\left[\left(g(A) - M_{12}\right)^2\right]\end{aligned} \tag{40}$$

$$\overset{(39)}{\geq} \frac{(p-q)^2}{k} - q(1-p) \sum_{\boldsymbol{\alpha}\in\{0,1\}^N, 0\leq|\boldsymbol{\alpha}|\leq D} \kappa_{\boldsymbol{\alpha}}^2(\overline{M}_{12}, \overline{X})$$

$$\overset{(a)}{=} \frac{(p-q)^2}{k} - q(1-p) \sum_{\boldsymbol{\alpha}\in\{0,1\}^N, 0\leq|\boldsymbol{\alpha}|\leq D} \kappa_{\boldsymbol{\alpha}}^2(\overline{M}_{12}, \overline{X})/\boldsymbol{\alpha}!$$

$$\overset{\text{Proposition 4}}{\geq} \frac{(p-q)^2}{k} - q(1-p) \cdot \left(\frac{(p-q)^2}{k^2 q(1-p)} + \frac{r(2-r)(p-q)^2}{(1-r)^2 n q(1-p)}\right)$$

$$= \frac{(p-q)^2}{k} - (p-q)^2\left(\frac{1}{k^2} + \frac{r(2-r)}{(1-r)^2 n}\right),$$

where (a) is because $\boldsymbol{\alpha} \in \{0, 1\}^N$. This completes the proof of this theorem. □

7.2. *Proof of Corollary* 1. PROOF. Since $k \geq 2$ and $n \geq k^2 \geq 2k$, by Theorem 2 we have there exists a small enough $r > 0$ such that when $\frac{(p-q)^2}{q(1-p)} \leq \frac{r}{(D(D+1))^2} \min(1, \frac{k^2}{n})$, we have

$$\inf_{\widehat{M}\in\mathbb{R}[A]_{\leq D}^{n\times n}} \mathbb{E}_{A,M\sim\mathbb{P}_{\mathrm{SBM}(p,q)}} \ell(\widehat{M}, M) \geq c_r \frac{(p-q)^2}{k} \tag{41}$$

for some constant $c_r > 0$ depends on r only.

Then we take $\epsilon \leq q \leq p \leq 1-\epsilon$ for some $\epsilon > 0$ such that $\frac{(p-q)^2}{q(1-p)} \geq c\frac{r}{(D(D+1))^2} \min(1, \frac{k^2}{n})$ for some $1 > c > 0$, and we have

$$\begin{aligned}\inf_{\widehat{M}\in\mathbb{R}[A]_{\leq D}^{n\times n}} \sup_{M\in\mathcal{M}_k} \mathbb{E}\big(\ell(\widehat{M}, M)\big) &\geq \inf_{\widehat{M}\in\mathbb{R}[A]_{\leq D}^{n\times n}} \mathbb{E}_{A,M\sim\mathbb{P}_{\mathrm{SBM}(p,q)}} \ell(\widehat{M}, M)\\ &\overset{(41)}{\geq} c_r \frac{(p-q)^2}{k} = c_r' \frac{r}{(D(D+1))^2} q(1-p)\left(\frac{k}{n} \wedge \frac{1}{k}\right)\\ &\geq \frac{c}{D^4}\left(\frac{k}{n} \wedge \frac{1}{k}\right),\end{aligned}$$

where c depends on ϵ and r only. □

Acknowledgments. The authors would like to thank an Associated Editor, and two anonymous referees for their helpful suggestions, which helped improve the presentation and quality of this paper.

Funding. The research of C. Gao is supported in part by NSF Grants ECCS-2216912 and DMS-2310769, NSF Career Award DMS-1847590 and an Alfred Sloan fellowship.

SUPPLEMENTARY MATERIAL

Supplement to "Computational lower bounds for graphon estimation via low-degree polynomials" (DOI: 10.1214/24-AOS2437SUPP; .pdf). This supplement contains further discussion and proofs.

REFERENCES

ABBE, E. (2017). Community detection and stochastic block models: Recent developments. *J. Mach. Learn. Res.* **18** 6446–6531. MR3827065

ABBE, E., BANDEIRA, A. S. and HALL, G. (2016). Exact recovery in the stochastic block model. *IEEE Trans. Inf. Theory* **62** 471–487. MR3447993 https://doi.org/10.1109/TIT.2015.2490670

ABBE, E. and SANDON, C. (2018). Proof of the achievability conjectures for the general stochastic block model. *Comm. Pure Appl. Math.* **71** 1334–1406. MR3812075 https://doi.org/10.1002/cpa.21719

AIROLDI, E. M., COSTA, T. B. and CHAN, S. H. (2013). Stochastic blockmodel approximation of a graphon: Theory and consistent estimation. *Adv. Neural Inf. Process. Syst.* **26**.

ALDOUS, D. J. (1981). Representations for partially exchangeable arrays of random variables. *J. Multivariate Anal.* **11** 581–598. MR0637937 https://doi.org/10.1016/0047-259X(81)90099-3

ARPINO, G. and VENKATARAMANAN, R. (2023). Statistical–computational tradeoffs in mixed sparse linear regression. In *Proceedings of Thirty Sixth Conference on Learning Theory* **195** 921–986.

AUDDY, A. and YUAN, M. (2023). Large Dimensional Independent Component Analysis: Statistical Optimality and Computational Tractability. ArXiv preprint. Available at arXiv:2303.18156. MR4639578

BALAKRISHNAN, S., KOLAR, M., RINALDO, A., SINGH, A. and WASSERMAN, L. (2011). Statistical and computational tradeoffs in biclustering. In *NIPS* 2011 *Workshop on Computational Trade-Offs in Statistical Learning*, *Vol.* 4.

BANDEIRA, A. S., BANKS, J., KUNISKY, D., MOORE, C. and WEIN, A. (2021). Spectral planting and the hardness of refuting cuts, colorability, and communities in random graphs. In *Conference on Learning Theory* 410–473. PMLR.

BANDEIRA, A. S., EL ALAOUI, A., HOPKINS, S., SCHRAMM, T., WEIN, A. S. and ZADIK, I. (2022). The Franz-Parisi criterion and computational trade-offs in high dimensional statistics. *Adv. Neural Inf. Process. Syst.* **35** 33831–33844.

BANDEIRA, A. S., KUNISKY, D. and WEIN, A. S. (2020). Computational hardness of certifying bounds on constrained PCA problems. In 11*th Innovations in Theoretical Computer Science Conference*. *LIPIcs. Leibniz Int. Proc. Inform.* **151** Art. No. 78, 29. Schloss Dagstuhl. Leibniz-Zent. Inform., Wadern. MR4048181

BANKS, J., MOHANTY, S. and RAGHAVENDRA, P. (2021). Local statistics, semidefinite programming, and community detection. In *Proceedings of the* 2021 *ACM-SIAM Symposium on Discrete Algorithms* (*SODA*) 1298–1316. SIAM, Philadelphia, PA. MR4262512 https://doi.org/10.1137/1.9781611976465.79

BANKS, J., MOORE, C., NEEMAN, J. and NETRAPALLI, P. (2016). Information-theoretic thresholds for community detection in sparse networks. In *Conference on Learning Theory* 383–416. PMLR.

BANKS, J., MOORE, C., VERSHYNIN, R., VERZELEN, N. and XU, J. (2018). Information-theoretic bounds and phase transitions in clustering, sparse PCA, and submatrix localization. *IEEE Trans. Inf. Theory* **64** 4872–4994. MR3819345 https://doi.org/10.1109/tit.2018.2810020

BARAK, B., HOPKINS, S., KELNER, J., KOTHARI, P. K., MOITRA, A. and POTECHIN, A. (2019). A nearly tight sum-of-squares lower bound for the planted clique problem. *SIAM J. Comput.* **48** 687–735. MR3945259 https://doi.org/10.1137/17M1138236

BERTHET, Q. and RIGOLLET, P. (2013). Complexity theoretic lower bounds for sparse principal component detection. In *Conference on Learning Theory* 1046–1066.

BICKEL, P. J. and CHEN, A. (2009). A nonparametric view of network models and Newman–Girvan and other modularities. *Proc. Natl. Acad. Sci. USA* **106** 21068–21073.

BICKEL, P. J., CHEN, A. and LEVINA, E. (2011). The method of moments and degree distributions for network models. *Ann. Statist.* **39** 2280–2301. MR2906868 https://doi.org/10.1214/11-AOS904

BORGS, C., CHAYES, J. and SMITH, A. (2015). Private graphon estimation for sparse graphs. *Adv. Neural Inf. Process. Syst.* **28**.

BORGS, C., CHAYES, J. T., COHN, H. and GANGULY, S. (2021). Consistent nonparametric estimation for heavy-tailed sparse graphs. *Ann. Statist.* **49** 1904–1930. MR4319235 https://doi.org/10.1214/20-aos1985

BORGS, C., CHAYES, J. T., COHN, H. and ZHAO, Y. (2018). An L^p theory of sparse graph convergence II: LD convergence, quotients and right convergence. *Ann. Probab.* **46** 337–396. MR3758733 https://doi.org/10.1214/17-AOP1187

BORGS, C., CHAYES, J. T., COHN, H. and ZHAO, Y. (2019). An L^p theory of sparse graph convergence I: Limits, sparse random graph models, and power law distributions. *Trans. Amer. Math. Soc.* **372** 3019–3062. MR3988601 https://doi.org/10.1090/tran/7543

BORGS, C., CHAYES, J. T., LOVÁSZ, L., SÓS, V. T. and VESZTERGOMBI, K. (2008). Convergent sequences of dense graphs. I. Subgraph frequencies, metric properties and testing. *Adv. Math.* **219** 1801–1851. MR2455626 https://doi.org/10.1016/j.aim.2008.07.008

BORGS, C., CHAYES, J. T., LOVÁSZ, L., SÓS, V. T. and VESZTERGOMBI, K. (2012). Convergent sequences of dense graphs II. Multiway cuts and statistical physics. *Ann. of Math.* (2) **176** 151–219. MR2925382 https://doi.org/10.4007/annals.2012.176.1.2

BRENNAN, M. and BRESLER, G. (2020). Reducibility and statistical–computational gaps from secret leakage. In *Conference on Learning Theory* 648–847. PMLR.

BRENNAN, M., BRESLER, G. and HULEIHEL, W. (2018). Reducibility and computational lower bounds for problems with planted sparse structure. In *Conference on Learning Theory* 48–166. PMLR.

BRENNAN, M. S., BRESLER, G., HOPKINS, S., LI, J. and SCHRAMM, T. (2021). Statistical query algorithms and low degree tests are almost equivalent. In *Conference on Learning Theory* 774–774. PMLR.

BRESLER, G. and HUANG, B. (2022). The algorithmic phase transition of random k-SAT for low degree polynomials. In 2021 *IEEE* 62*nd Annual Symposium on Foundations of Computer Science—FOCS* 2021 298–309. IEEE Comput. Soc., Los Alamitos, CA. MR4399691

BUTUCEA, C. and INGSTER, Y. I. (2013). Detection of a sparse submatrix of a high-dimensional noisy matrix. *Bernoulli* **19** 2652–2688. MR3160567 https://doi.org/10.3150/12-BEJ470

BUTUCEA, C., INGSTER, Y. I. and SUSLINA, I. A. (2015). Sharp variable selection of a sparse submatrix in a high-dimensional noisy matrix. *ESAIM Probab. Stat.* **19** 115–134. MR3374872 https://doi.org/10.1051/ps/2014017

CAI, T. T., LIANG, T. and RAKHLIN, A. (2017). Computational and statistical boundaries for submatrix localization in a large noisy matrix. *Ann. Statist.* **45** 1403–1430. MR3670183 https://doi.org/10.1214/16-AOS1488

CHAN, S. and AIROLDI, E. (2014). A consistent histogram estimator for exchangeable graph models. In *International Conference on Machine Learning* 208–216. PMLR.

CHANDRASEKARAN, V. and JORDAN, M. I. (2013). Computational and statistical tradeoffs via convex relaxation. *Proc. Natl. Acad. Sci. USA* **110** E1181–E1190. MR3047651 https://doi.org/10.1073/pnas.1302293110

CHATTERJEE, S. (2015). Matrix estimation by universal singular value thresholding. *Ann. Statist.* **43** 177–214. MR3285604 https://doi.org/10.1214/14-AOS1272

CHEN, Y. and XU, J. (2016). Statistical–computational tradeoffs in planted problems and submatrix localization with a growing number of clusters and submatrices. *J. Mach. Learn. Res.* **17** 882–938. MR3491121

CHI, E. C., ALLEN, G. I. and BARANIUK, R. G. (2017). Convex biclustering. *Biometrics* **73** 10–19. MR3632347 https://doi.org/10.1111/biom.12540

CHIN, P., RAO, A. and VU, V. (2015). Stochastic block model and community detection in sparse graphs: A spectral algorithm with optimal rate of recovery. In *Conference on Learning Theory* 391–423. PMLR.

CHOI, D. (2017). Co-clustering of nonsmooth graphons. *Ann. Statist.* **45** 1488–1515. MR3670186 https://doi.org/10.1214/16-AOS1497

CHOI, D. and WOLFE, P. J. (2014). Co-clustering separately exchangeable network data. *Ann. Statist.* **42** 29–63. MR3161460 https://doi.org/10.1214/13-AOS1173

DADON, M., HULEIHEL, W. and BENDORY, T. (2024). Detection and recovery of hidden submatrices. *IEEE Trans. Signal Inf. Process. Netw.* **10** 69–82. MR4692776 https://doi.org/10.1109/tsipn.2024.3352264

DECELLE, A., KRZAKALA, F., MOORE, C. and ZDEBOROVÁ, L. (2011). Asymptotic analysis of the stochastic block model for modular networks and its algorithmic applications. *Phys. Rev. E* **84** 066106.

DIACONIS, P. and JANSON, S. (2008). Graph limits and exchangeable random graphs. *Rend. Mat. Appl.* (7) **28** 33–61. MR2463439

DIAKONIKOLAS, I. and KANE, D. (2022). Non-Gaussian component analysis via lattice basis reduction. In *Conference on Learning Theory* 4535–4547. PMLR.

DIAKONIKOLAS, I., KANE, D. M. and STEWART, A. (2017). Statistical query lower bounds for robust estimation of high-dimensional Gaussians and Gaussian mixtures (extended abstract). In 58*th Annual IEEE Symposium on Foundations of Computer Science—FOCS* 2017 73–84. IEEE Comput. Soc., Los Alamitos, CA. MR3734219 https://doi.org/10.1109/FOCS.2017.16

DIAKONIKOLAS, I., KONG, W. and STEWART, A. (2019). Efficient algorithms and lower bounds for robust linear regression. In *Proceedings of the Thirtieth Annual ACM-SIAM Symposium on Discrete Algorithms* 2745–2754. SIAM, Philadelphia, PA. MR3909639 https://doi.org/10.1137/1.9781611975482.170

DING, Y., KUNISKY, D., WEIN, A. S. and BANDEIRA, A. S. (2024). Subexponential-time algorithms for sparse PCA. *Found. Comput. Math.* **24** 865–914. MR4760356 https://doi.org/10.1007/s10208-023-09603-0

DONIER-MEROZ, E., DALALYAN, A. S., KRAMARZ, F., CHONÉ, P. and D'HAULTFOEUILLE, X. (2023). Graphon Estimation in bipartite graphs with observable edge labels and unobservable node labels. ArXiv preprint. Available at arXiv:2304.03590.

DONOHO, D. L., MALEKI, A. and MONTANARI, A. (2009). Message-passing algorithms for compressed sensing. *Proc. Natl. Acad. Sci.* **106** 18914–18919.

FEI, Y. and CHEN, Y. (2020). Achieving the Bayes error rate in synchronization and block models by SDP, robustly. *IEEE Trans. Inf. Theory* **66** 3929–3953. MR4115142 https://doi.org/10.1109/TIT.2020.2966438

FELDMAN, V., GRIGORESCU, E., REYZIN, L., VEMPALA, S. S. and XIAO, Y. (2017). Statistical algorithms and a lower bound for detecting planted cliques. *J. ACM* **64** Art. 8, 37. MR3664576 https://doi.org/10.1145/3046674

FELDMAN, V., PERKINS, W. and VEMPALA, S. (2018). On the complexity of random satisfiability problems with planted solutions. *SIAM J. Comput.* **47** 1294–1338. MR3827195 https://doi.org/10.1137/16M1078471

GAMARNIK, D., JAGANNATH, A. and WEIN, A. S. (2020). Low-degree hardness of random optimization problems. In 2020 *IEEE* 61*st Annual Symposium on Foundations of Computer Science* 131–140. IEEE Comput. Soc., Los Alamitos, CA. MR4232029 https://doi.org/10.1109/FOCS46700.2020.00021

GAMARNIK, D. and SUDAN, M. (2014). Limits of local algorithms over sparse random graphs [extended abstract]. In *ITCS'*14*—Proceedings of the* 2014 *Conference on Innovations in Theoretical Computer Science* 369–375. ACM, New York. MR3359490

GAO, C., LU, Y., MA, Z. and ZHOU, H. H. (2016). Optimal estimation and completion of matrices with biclustering structures. *J. Mach. Learn. Res.* **17** 5602–5630. MR3569248

GAO, C., LU, Y. and ZHOU, H. H. (2015). Rate-optimal graphon estimation. *Ann. Statist.* **43** 2624–2652. MR3405606 https://doi.org/10.1214/15-AOS1354

GAO, C. and MA, Z. (2021). Minimax rates in network analysis: Graphon estimation, community detection and hypothesis testing. *Statist. Sci.* **36** 16–33. MR4194201 https://doi.org/10.1214/19-STS736

GAO, C., MA, Z. and ZHOU, H. H. (2017). Sparse CCA: Adaptive estimation and computational barriers. *Ann. Statist.* **45** 2074–2101. MR3718162 https://doi.org/10.1214/16-AOS1519

GAUCHER, S. and KLOPP, O. (2021). Optimality of variational inference for stochasticblock model with missing links. *Adv. Neural Inf. Process. Syst.* **34** 19947–19959.

GIRVAN, M. and NEWMAN, M. E. J. (2002). Community structure in social and biological networks. *Proc. Natl. Acad. Sci. USA* **99** 7821–7826. MR1908073 https://doi.org/10.1073/pnas.122653799

GOLDENBERG, A., ZHENG, A. X., FIENBERG, S. E., AIROLDI, E. M. et al. (2010). A survey of statistical network models. *Found. Trends Mach. Learn.* **2** 129–233.

GUÉDON, O. and VERSHYNIN, R. (2016). Community detection in sparse networks via Grothendieck's inequality. *Probab. Theory Related Fields* **165** 1025–1049. MR3520025 https://doi.org/10.1007/s00440-015-0659-z

HAJEK, B., WU, Y. and XU, J. (2015). Computational lower bounds for community detection on random graphs. In *Conference on Learning Theory* 899–928.

HARTIGAN, J. A. (1972). Direct clustering of a data matrix. *J. Amer. Statist. Assoc.* **67** 123–129.

HOLLAND, P. W., LASKEY, K. B. and LEINHARDT, S. (1983). Stochastic blockmodels: First steps. *Soc. Netw.* **5** 109–137. MR0718088 https://doi.org/10.1016/0378-8733(83)90021-7

HOLMGREN, J. and WEIN, A. S. (2021). Counterexamples to the low-degree conjecture. In 12*th Innovations in Theoretical Computer Science Conference*. *LIPIcs. Leibniz Int. Proc. Inform.* **185** Art. No. 75, 9. Schloss Dagstuhl. Leibniz-Zent. Inform., Wadern. MR4214319

HOOVER, D. N. (1979). Relations on probability spaces and arrays of random variables. Preprint, Institute for Advanced Study, Princeton, NJ 2 275.

HOPKINS, S. B., KOTHARI, P. K., POTECHIN, A., RAGHAVENDRA, P., SCHRAMM, T. and STEURER, D. (2017). The power of sum-of-squares for detecting hidden structures. In 58*th Annual IEEE Symposium on Foundations of Computer Science—FOCS* 2017 720–731. IEEE Comput. Soc., Los Alamitos, CA. MR3734275 https://doi.org/10.1109/FOCS.2017.72

HOPKINS, S. B. and STEURER, D. (2017). Efficient Bayesian estimation from few samples: Community detection and related problems. In 58*th Annual IEEE Symposium on Foundations of Computer Science—FOCS* 2017 379–390. IEEE Comput. Soc., Los Alamitos, CA. MR3734245 https://doi.org/10.1109/FOCS.2017.42

HOPKINS, S. B. K. (2018). Statistical Inference and the Sum of Squares Method. Ph.D. Thesis. MR3864930

JIN, J. (2015). Fast community detection by SCORE. *Ann. Statist.* **43** 57–89. MR3285600 https://doi.org/10.1214/14-AOS1265

JIN, J., KE, T., TURNER, P. and ZHANG, A. (2022). Phase transition for detecting a small community in a large network. In *The* 11*th International Conference on Learning Representations*.

KESTEN, H. and STIGUM, B. P. (1966). Additional limit theorems for indecomposable multidimensional Galton–Watson processes. *Ann. Math. Stat.* **37** 1463–1481. MR0200979 https://doi.org/10.1214/aoms/1177699139

KLOPP, O., TSYBAKOV, A. B. and VERZELEN, N. (2017). Oracle inequalities for network models and sparse graphon estimation. *Ann. Statist.* **45** 316–354. MR3611494 https://doi.org/10.1214/16-AOS1454

KLOPP, O. and VERZELEN, N. (2019). Optimal graphon estimation in cut distance. *Probab. Theory Related Fields* **174** 1033–1090. MR3980311 https://doi.org/10.1007/s00440-018-0878-1

KOEHLER, F. and MOSSEL, E. (2022). Reconstruction on trees and low-degree polynomials. *Adv. Neural Inf. Process. Syst.* **35** 18942–18954.

KOLAR, M., BALAKRISHNAN, S., RINALDO, A. and SINGH, A. (2011). Minimax localization of structural information in large noisy matrices. In *Advances in Neural Information Processing Systems* 909–917.

KUNISKY, D., WEIN, A. S. and BANDEIRA, A. S. (2022). Notes on computational hardness of hypothesis testing: Predictions using the low-degree likelihood ratio. In *Mathematical Analysis*, *Its Applications and Computation*. *Springer Proc. Math. Stat.* **385** 1–50. Springer, Cham. MR4461037 https://doi.org/10.1007/978-3-030-97127-4_1

LEI, J. and RINALDO, A. (2015). Consistency of spectral clustering in stochastic block models. *Ann. Statist.* **43** 215–237. MR3285605 https://doi.org/10.1214/14-AOS1274

LI, X., CHEN, Y. and XU, J. (2021). Convex relaxation methods for community detection. *Statist. Sci.* **36** 2–15. MR4194200 https://doi.org/10.1214/19-STS715

LI, Y., SHAH, D., SONG, D. and YU, C. L. (2020). Nearest neighbors for matrix estimation interpreted as blind regression for latent variable model. *IEEE Trans. Inf. Theory* **66** 1760–1784. MR4077516 https://doi.org/10.1109/tit.2019.2950299

LÖFFLER, M., WEIN, A. S. and BANDEIRA, A. S. (2022). Computationally efficient sparse clustering. *Inf. Inference* **11** 1255–1286. MR4526323 https://doi.org/10.1093/imaiai/iaac019

LOVÁSZ, L. (2012). *Large Networks and Graph Limits*. *American Mathematical Society Colloquium Publications* **60**. Amer. Math. Soc., Providence, RI. MR3012035 https://doi.org/10.1090/coll/060

LOVÁSZ, L. and SZEGEDY, B. (2006). Limits of dense graph sequences. *J. Combin. Theory Ser. B* **96** 933–957. MR2274085 https://doi.org/10.1016/j.jctb.2006.05.002

LUO, Y. and GAO, C. (2024). Supplement to "Computational Lower Bounds for Graphon Estimation via Low-degree Polynomials." https://doi.org/10.1214/24-AOS2437SUPP

LUO, Y. and ZHANG, A. R. (2022a). Tensor clustering with planted structures: Statistical optimality and computational limits. *Ann. Statist.* **50** 584–613. MR4382029 https://doi.org/10.1214/21-aos2123

LUO, Y. and ZHANG, A. R. (2022b). Tensor-on-tensor regression: Riemannian optimization, over-parameterization, statistical–computational gap, and their interplay. arXiv preprint. Available at arXiv:2206.08756.

LYU, Z. and XIA, D. (2023). Optimal estimation and computational limit of low-rank Gaussian mixtures. *Ann. Statist.* **51** 646–667. MR4600996 https://doi.org/10.1214/23-aos2264

MA, T. and WIGDERSON, A. (2015). Sum-of-squares lower bounds for sparse PCA. In *Advances in Neural Information Processing Systems* 1612–1620.

MA, Z. and WU, Y. (2015). Computational barriers in minimax submatrix detection. *Ann. Statist.* **43** 1089–1116. MR3346698 https://doi.org/10.1214/14-AOS1300

MANKAD, S. and MICHAILIDIS, G. (2014). Biclustering three-dimensional data arrays with plaid models. *J. Comput. Graph. Statist.* **23** 943–965. MR3270705 https://doi.org/10.1080/10618600.2013.851608

MAO, C. and WEIN, A. S. (2021). Optimal spectral recovery of a planted vector in a subspace. ArXiv preprint. Available at arXiv:2105.15081.

MAO, C., WEIN, A. S. and ZHANG, S. (2023). Detection-recovery gap for planted dense cycles. In *The Thirty Sixth Annual Conference on Learning Theory* 2440–2481. PMLR.

MASSOULIÉ, L. (2014). Community detection thresholds and the weak Ramanujan property. In *STOC'14—Proceedings of the* 2014 *ACM Symposium on Theory of Computing* 694–703. ACM, New York. MR3238997

MILLER, K. T., JORDAN, M. and GRIFFITHS, T. (2009). Nonparametric latent feature models for link prediction. In *Advances in Neural Information Processing Systems*, *Vol*. 22.

MONTANARI, A. and WEIN, A. S. (2022). Equivalence of approximate message passing and low-degree polynomials in rank-one matrix estimation. ArXiv preprint. Available at arXiv:2212.06996.

MOORE, C. (2017). The computer science and physics of community detection: Landscapes, phase transitions, and hardness. *Bull. Eur. Assoc. Theor. Comput. Sci. EATCS* **121** 26–61. MR3699594

MOSSEL, E., NEEMAN, J. and SLY, A. (2015a). Consistency thresholds for the planted bisection model [extended abstract]. In *STOC'15—Proceedings of the* 2015 *ACM Symposium on Theory of Computing* 69–75. ACM, New York. MR3388184

MOSSEL, E., NEEMAN, J. and SLY, A. (2015b). Reconstruction and estimation in the planted partition model. *Probab. Theory Related Fields* **162** 431–461. MR3383334 https://doi.org/10.1007/s00440-014-0576-6

MOSSEL, E., NEEMAN, J. and SLY, A. (2018). A proof of the block model threshold conjecture. *Combinatorica* **38** 665–708. MR3876880 https://doi.org/10.1007/s00493-016-3238-8

OLHEDE, S. C. and WOLFE, P. J. (2014). Network histograms and universality of blockmodel approximation. *Proc. Natl. Acad. Sci. USA* **111** 14722–14727.

PANANJADY, A. and SAMWORTH, R. J. (2022). Isotonic regression with unknown permutations: Statistics, computation and adaptation. *Ann. Statist.* **50** 324–350.

PENSKY, M. (2019). Dynamic network models and graphon estimation. *Ann. Statist.* **47**.

ROHE, K., CHATTERJEE, S. and YU, B. (2011). Spectral clustering and the high-dimensional stochastic blockmodel. *Ann. Statist.* **39** 1878–1915.

ROHE, K., QIN, T. and YU, B. (2016). Co-clustering directed graphs to discover asymmetries and directional communities. *Proc. Natl. Acad. Sci. USA* **113** 12679–12684.

ROSSMAN, B. (2008). On the constant-depth complexity of k-clique. In *Proceedings of the Fortieth Annual ACM Symposium on Theory of Computing* 721–730. ACM, New York.

SCHRAMM, T. and WEIN, A. S. (2022). Computational barriers to estimation from low-degree polynomials. *Ann. Statist.* **50** 1833–1858.

SHAH, D. and LEE, C. (2018). Reducing crowdsourcing to graphon estimation, statistically. In *International Conference on Artificial Intelligence and Statistics* 1741–1750. PMLR.

SHAH, N., BALAKRISHNAN, S., GUNTUBOYINA, A. and WAINWRIGHT, M. (2016). Stochastically transitive models for pairwise comparisons: Statistical and computational issues. In *International Conference on Machine Learning* 11–20. PMLR.

SISCHKA, B. and KAUERMANN, G. (2022). Stochastic block smooth graphon model. arXiv preprint. Available at arXiv:2203.13304.

WANG, T., BERTHET, Q. and SAMWORTH, R. J. (2016). Statistical and computational trade-offs in estimation of sparse principal components. *Ann. Statist.* **44** 1896–1930.

WASSERMAN, S. and FAUST, K. (1994). Social Network Analysis: Methods and Applications.

WEIN, A. S. (2022). Optimal low-degree hardness of maximum independent set. *Math. Stat. Learn.* **4** 221–251.

WEIN, A. S. (2023). Average-case complexity of tensor decomposition for low-degree polynomials. In *Proceedings of the* 55*th Annual ACM Symposium on Theory of Computing* 1685–1698.

WOLFE, P. J. and OLHEDE, S. C. (2013). Nonparametric graphon estimation. ArXiv preprint. Available at arXiv:1309.5936.

WU, Y. and XU, J. (2021). Statistical problems with planted structures: Information-theoretical and computational limits. In *Information-Theoretic Methods in Data Science* 383. 13.

XU, J. (2018). Rates of convergence of spectral methods for graphon estimation. In *International Conference on Machine Learning* 5433–5442. PMLR.

YANG, J., HAN, C. and AIROLDI, E. (2014). Nonparametric estimation and testing of exchangeable graph models. In *Artificial Intelligence and Statistics* 1060–1067. PMLR.

ZADIK, I., SONG, M. J., WEIN, A. S. and BRUNA, J. (2022). Lattice-based methods surpass sum-of-squares in clustering. In *Conference on Learning Theory* 1247–1248. PMLR.

ZHANG, A. and XIA, D. (2018). Tensor svd: Statistical and computational limits. *IEEE Trans. Inf. Theory* **64** 7311–7338.

ZHANG, A. Y. and ZHOU, H. H. (2016). Minimax rates of community detection in stochastic block models. *Ann. Statist.* 2252–2280.

ZHANG, Y., LEVINA, E. and ZHU, J. (2017). Estimating network edge probabilities by neighbourhood smoothing. *Biometrika* **104** 771–783.