

MIT Open Access Articles

A Nearly Tight Sum-of-Squares Lower Bound for the Planted Clique Problem

The MIT Faculty has made this article openly available. **Please share** how this access benefits you. Your story matters.

Citation: Barak, Boaz, Hopkins, Samuel, Kelner, Jonathan, Kothari, Pravesh K, Moitra, Ankur et al. 2019. "A Nearly Tight Sum-of-Squares Lower Bound for the Planted Clique Problem." SIAM Journal on Computing, 48 (2).

As Published: 10.1137/17M1138236

Publisher: Society for Industrial & Applied Mathematics (SIAM)

Persistent URL: <https://hdl.handle.net/1721.1/145722>

Version: Final published version: final published article, as it appeared in a journal, conference proceedings, or other formally published context

Terms of Use: Article is made available in accordance with the publisher's policy and may be subject to US copyright law. Please refer to the publisher's site for terms of use.



A NEARLY TIGHT SUM-OF-SQUARES LOWER BOUND FOR THE PLANTED CLIQUE PROBLEM*

BOAZ BARAK[†], SAMUEL HOPKINS[‡], JONATHAN KELNER[§], PRAVESH K. KOTHARI[¶],
ANKUR MOITRA[§], AND AARON POTECHIN^{||}

Abstract. We prove that with high probability over the choice of a random graph G from the Erdős–Rényi distribution $G(n, 1/2)$, the $n^{O(d)}$ -time degree d sum-of-squares (SOS) semidefinite programming relaxation for the clique problem will give a value of at least $n^{1/2-c(d/\log n)^{1/2}}$ for some constant $c > 0$. This yields a nearly tight $n^{1/2-o(1)}$ bound on the value of this program for any degree $d = o(\log n)$. Moreover, we introduce a new framework that we call *pseudocalibration* to construct SOS lower bounds. This framework is inspired by taking a computational analogue of Bayesian probability theory. It yields a general recipe for constructing good pseudodistributions (i.e., dual certificates for the SOS semidefinite program) and sheds further light on the ways in which this hierarchy differs from others.

Key words. sum-of-squares, planted clique, lower bound

AMS subject classification. 68Q25

DOI. 10.1137/17M1138236

1. Introduction. The *planted clique* (also known as *hidden clique*) problem is a central question in average-case complexity. Arising from the 1976 work of Karp [Kar76], the problem was formally defined by Jerrum [Jer92] and Kucera [Kuc95] as follows: given a random Erdős–Rényi graph G from the distribution $G(n, 1/2)$ where every edge is chosen to be included with probability $1/2$ independently of all others in which we *plant* an additional clique (i.e., set of vertices that are all neighbors of one another) S of size ω , find S . It is not hard to see that the problem can be solved by brute force search, which in this case takes quasi-polynomial time whenever $\omega > c \log n$ for any constant $c > 2$. Despite considerable effort, the best polynomial-time algorithms only work when $\omega = \varepsilon \sqrt{n}$ for any constant $\varepsilon > 0$ [AKS98].

Over the years, planted clique and related problems have found applications to important questions in a variety of areas, including community detection [HWX15], finding signals in molecular biology [PS00], discovering motifs in biological networks [MSOI+02, JM15], computing the Nash equilibrium [HK11, ABC13], property testing [AAK+07], sparse principal component analysis [BR13], compressed sensing [KZ14], cryptography [JP00, ABW10], and even mathematical finance [ABBG11].

Thus, the question of whether the currently known algorithms can be improved

*Received by the editors July 10, 2017; accepted for publication (in revised form) January 22, 2019; published electronically April 30, 2019. An extended abstract of this work appeared in *Proceedings of the 57th Annual IEEE Symposium on Foundations of Computer Science*, 2016.

<http://www.siam.org/journals/sicomp/48-2/M113823.html>

Funding: The second author’s research was partially supported by NSF GRFP grant 1144153 and by NSF grants 1408673 and 1350196. The third author’s research was partially supported by NSF award 1111109. The fifth author’s research was partially supported by NSF CAREER award CCF-1453261, a grant from the MIT NEC Corporation, and a Google Faculty Research Award.

[†]Microsoft Research New England, Cambridge, MA 02142 (b@boazbarak.org).

[‡]Computer Science, Cornell University, Ithaca, NY 14853 (samhop@cs.cornell.edu).

[§]MIT, School of Engineering, Cambridge, MA 02139 (kelner@mit.edu, moitra@mit.edu).

[¶]Princeton University and Institute for Advanced Study, Princeton, NJ 08540 (kotpravesh@gmail.com).

^{||}Institute for Advanced Study, Princeton, NJ 08540 (apotechin@ias.edu).

is of great interest. Unfortunately, it is unlikely that lower bounds for planted clique can be derived from conjectured complexity class separations such as $\mathbf{P} \neq \mathbf{NP}$, precisely because it is an average-case problem [FF93, BT06]. Our best evidence for its difficulty comes from showing limitations on powerful *classes* of algorithms. In particular, since many of the algorithmic approaches for this and related problems involve spectral techniques and convex programs, limitations for these types of algorithms are of significant interest. One such negative result was shown by Feige and Krauthgamer [FK03], who proved that the $n^{O(d)}$ -time *degree d Lovász–Schrijver semidefinite programming hierarchy* (LS_+ for short) can only recover the clique if its size is at least $\sqrt{n/2^d}$.¹

However, recently it was shown that in several cases, the *sum-of-squares (SOS) hierarchy* [Sho87, Par00, Las01]—a stronger family of semidefinite programs which can be solved in time $n^{O(d)}$ for degree parameter d —can be significantly more powerful than other algorithms such as LS_+ [BBH+12, BKS14, BKS15]. In particular, it was conceivable that the SOS hierarchy might be able to find planted cliques that are much smaller than \sqrt{n} in polynomial time, or at least be able to beat brute force search.

The first SOS lower bound for planted clique was shown by Meka, Potechin, and Wigderson [MPW15], who proved that the degree d SOS hierarchy cannot recover a clique of size $\tilde{O}(n^{1/d})$. This bound was later improved upon by Deshpande and Montanari [DM15] and then Hopkins et al. [HKP+16] to $\tilde{O}(n^{1/2})$ for degree $d = 4$ and $\tilde{O}(n^{1/(\lceil d/2 \rceil + 1)})$ for general d . However, this still left open the possibility that the constant degree (and hence polynomial time) SOS algorithm can significantly beat the \sqrt{n} bound, perhaps even being able to find cliques of size n^ε for any fixed $\varepsilon > 0$. This paper answers this question negatively by proving the following theorem.

THEOREM 1.1 (main theorem). *There is an absolute constant c so that for every $d = d(n)$ and large enough n , the SOS relaxation of the planted clique problem has integrality gap at least $n^{1/2 - c(d/\log n)^{1/2}}$.*

Beyond improving the previously known results, our proof is significantly more general and we believe provides a better intuition behind the limitations for SOS algorithms by viewing them from a “computational Bayesian probability” lens that is of its own interest. Moreover, there is some hope (as we elaborate on below) that this view could be useful not just for more negative results but for SOS *upper bounds* as well. In particular, our proof elucidates some aspects of the way in which the SOS algorithm is more powerful than the LS_+ algorithm.

Remark 1.2 (the different variants of the planted clique problem). Like other average-case problems in \mathbf{NP} , the planted clique problem with parameter ω has three variants: *search*, *refutation*, and *decision*. The *search* variant is the task of recovering the clique from a graph in which it was planted. The *refutation* variant is the task of *certifying* that a random graph in $G(n, 1/2)$ (where with high probability the largest clique has size $(2 + o(1)) \log n$) does not have a clique of size ω . The *decision* problem is to distinguish between a random graph from $G(n, 1/2)$ and a graph in which an ω -sized clique has been planted. The decision variant can be reduced to either the search or the refutation variant, but we know of no reduction between the latter two variants. Integrality gaps for mathematical relaxations such as the SOS hierarchy are

¹As we discuss in Remark 1.2, formally such results apply to the incomparable *refutation* problem, which is the task of certifying that there is no ω -sized clique in a random $G(n, 1/2)$ graph. However, our current knowledge is consistent with these variants having the same computational complexity.

most naturally stated as negative results for the *refutation* variant, as they show that such relaxations cannot certify that a random graph has no ω -sized clique by looking at the maximum value of the objective function. Our result can also be viewed as showing that the natural SOS-based algorithm for the *decision* problem (which attempts to distinguish on the objective value) also fails. Moreover, our result also rules out some types of SOS-based algorithms for the *search* problem, as it shows that in a graph with a planted clique, there exists a solution with an objective value of ω based only on the random part, which means that it does not contain any information about which nodes participate in the clique and hence is not useful for rounding algorithms.

2. Planted clique and probabilistic inference. We now discuss the ways in which planted clique differs from problems for which strong SOS lower bounds have been shown before and how this relates to a “computational Bayesian” perspective. There have been several strong lower bounds for the SOS algorithm before, in particular for problems such as 3SAT and 3XOR, and other constraint satisfaction problems, as well as the knapsack problem [Gri01, Sch08, BCK15]. However, obtaining strong lower bounds for the planted clique problem seems to have required different techniques. A high-level way to describe the difference is that lower bounds for planted clique require accounting for *weak global constraints* rather than *strong local* ones. In the random 3SAT/3XOR setting, the effect of one variable on another is either extremely strong (if they are “nearby” in the formula) or essentially zero. In contrast, in planted clique each variable has a weak *global* effect on all of the other variables. We now explain this in more detail.

Consider a random graph G in which a clique S of size ω has been planted. If someone tells us some simple statistics of G and then tells us that vertex 17 is not in S , this new information makes it slightly less likely that 17’s neighbors are in S and slightly more likely that 17’s nonneighbors are in S . So, this information has a *weak global* effect. In contrast, when we have a random sparse 3SAT formula φ in which an assignment x has been planted, if someone tells us that $x_{17} = 0$, then it gives us a lot of information about the local neighborhood of the 17th variable (the variables that are involved in constraints with 17 or ones that have a short path of constraints to it), but there is an exponential decay of these correlations, and so this information tells us almost nothing about the distribution of most of the variables x_i (that are far away from 17 in the sparse graph induced by φ).² Thus, in the random 3SAT setting, information about the assignments of individual variables has a *strong local effect*. Indeed, previous SOS lower bounds for random 3SAT and 3XOR [Gri01, Sch08] could be interpreted as producing a “distribution-like” object in which, conditioned on the value of a small set of variables S , some of the variables “close” to S in the formula were fixed, and the rest were completely independent.

This difference between the random SAT and the planted clique problems means that some subtleties that can be ignored in the setting of random constraint satisfaction problems need to be tackled head-on when dealing with planted cliques. However, to make this more clear, we need to take a detour and discuss Bayesian probabilities and their relation to SOS.

²This exponential decay can be shown formally for the case of satisfiable random 3SAT or 3XOR formulas whose clause density is sufficiently smaller than the threshold. In our regime of overconstrained random 3SAT/3XOR formulas, there will not exist any satisfying assignments, and so to talk about “correlations” in the distributions of assignments we need to talk about the “Bayesian estimates” that arise from algorithms such as SOS or belief propagation. Both of these algorithms exhibit the sort of exponential decay that we are talking about; see also Remark 2.1.

2.1. Computational Bayesian probabilities and pseudodistributions. If a graph G contains a unique clique S of size ω , then, strictly speaking, for every vertex i the probability that i is in S is either zero or one. But, a computationally bounded observer may not know whether or not i is in the clique, and we could try to quantify this ignorance using probabilities. These can be thought of as a computational analogue of *Bayesian probabilities* that, rather than aiming to measure the frequency at which an event occurs in some sample space, attempts to capture the subjective beliefs of some observer.

That is, the Bayesian probability that an observer B assigns to an event E can be thought of as corresponding to the odds at which B would make the bet that E holds. Note that this probability could be strictly between zero and one, even if the event E is fully determined, depending on the evidence available to B . While typically Bayesian analysis does not take into account computational limitations, one could imagine that even if B has access to information that fully determines whether or not E happened, he could still rationally assign a subjective probability to E that is strictly between zero and one if making the inferences from this information is computationally infeasible. In particular, in the example above, even if a computationally bounded observer has access to the graph G , which information-theoretically fully determines the planted ω -sized clique, he could still assign a probability strictly between zero and one to the event that vertex 17 is in the planted ω -sized clique, based on some simple to compute statistics, such as how many neighbors 17 has.

The SOS algorithm can be thought of as giving rise to an internally consistent set of such “computational probabilities.” These probabilities may not capture *all* possible inferences that a computationally bounded observer could make, but they do capture all inferences that can be made via a powerful proof system.

Bayesian estimates for planted clique. To get a sense of our results and techniques, it is instructive to consider the following scenario. Let $G(n, 1/2, \omega)$ be the distribution over pairs (G, x) of an n -vertex graph G and a vector $x \in \mathbb{R}^n$ which is obtained by sampling a random graph in $G(n, 1/2)$, planting an ω -sized clique in it, and letting G be the resulting graph and x the 0/1 characteristic vector of the planted clique. Let $f : \{0, 1\}^{\binom{n}{2}} \times \mathbb{R}^n \rightarrow \mathbb{R}$ be some function that maps a graph G and a vector x into some real number $f_G(x)$. Now imagine two parties, Alice and Bob (where Bob can also stand for “Bayesian”), that play the following game: Alice samples (G, x) from the distribution $G(n, 1/2, \omega)$ and sends G to Bob, who wants to output the expected value of $f_G(x)$. We denote this value by $\tilde{\mathbb{E}}_G f_G$.

If we have no computational constraints, then it is clear that Bob can simply output $\tilde{\mathbb{E}}_G f_G$ to be equal to $\mathbb{E}_{x|G} f_G(x)$, by which we mean the expected value of $f_G(x)$, where x is chosen according to the conditional distribution on x given the graph G .³ In particular, the value $\tilde{\mathbb{E}}_G f_G$ will be *calibrated* in the sense that

$$(2.1) \quad \mathbb{E}_{G \in_R G(n, 1/2, \omega)} \tilde{\mathbb{E}}_G f_G = \mathbb{E}_{(G, x) \in_R G(n, 1/2, \omega)} f_G(x).$$

Now if Bob is computationally bounded, then he will not necessarily be able to compute the value of $\mathbb{E}_{x|G} f_G(x)$, even for a simple function such as $f_G(x) = x_{17}$. Indeed, as we mentioned, since with high probability the clique x is uniquely determined by G , $\mathbb{E}_{x|G} x_{17}$ will simply equal 1 if vertex 17 is in the clique and equal

³The astute reader might note that this expectation is somewhat degenerate since with very high probability the graph G will uniquely determine the vector x , but please bear with us, as in the computational setting we will be able to treat x as “undetermined.”

to 0 otherwise. However, note that we don't need to compute the true conditional expectation to obtain a calibrated estimate. In the above example, if Bob simply outputs $\tilde{\mathbb{E}}x_{17} = \omega/n$, then his estimate will satisfy (2.1).

Our SOS lower bound amounts to coming up with some reasonable “pseudoexpectation” that can be efficiently computed, where $\tilde{\mathbb{E}}_G$ is meant to capture a “best effort” of a computationally bounded party of approximating the Bayesian conditional expectation $\mathbb{E}_{x|G}$. Our pseudoexpectation will be far from the true conditional expectations but will be internally consistent in the sense that for all “simple” functions f it will satisfy (2.1). The key property is that our pseudoexpectation will not distinguish between a graph G drawn from $G(n, 1/2, \omega)$ and a random G from $G(n, 1/2)$. In particular, it will also satisfy the following *pseudocalibration* condition:

$$(2.2) \quad \mathbb{E}_{G \in_R G(n, 1/2)} \tilde{\mathbb{E}}_G f_G = \mathbb{E}_{(G, x) \in_R G(n, 1/2, \omega)} f_G(x)$$

for all “simple” functions $f = f(G, x)$. Note that (2.2) does not make sense for the estimates of a truly Bayesian (i.e., computationally unbounded) Bob since almost all graphs G in $G(n, 1/2)$ are not even in the support of $G(n, 1/2, \omega)$. Nevertheless, our pseudodistributions will be well defined even for a random graph and hence will yield estimates for the probabilities over this hypothetical object (i.e., the ω -sized clique) that does not exist. The “pseudocalibration” condition (2.2) might seem innocent, but it turns out that it implies many useful properties. In particular, it is not hard to see that (2.2) implies that for every *simple strong constraint* of the clique problem—a function f such that $f(G, x) = 0$ for every x that is a characteristic vector of an ω -clique in G —it must hold that $\tilde{\mathbb{E}}_G f_G = 0$. But even beyond these “strong constraints,” (2.2) implies that the pseudoexpectation satisfies many *weak constraints* as well, such as the fact that a vertex of high degree is more likely to be in the clique and that if i is not in the clique, then its neighbors are less likely and its nonneighbors are more likely to be in it.

Indeed, the key conceptual insight of this paper is to phrase the pseudocalibration property (2.2) as a desiderata for our pseudodistributions. Namely, we say that a function $f = f(G, x)$ is “simple” if it is a low-degree polynomial in both the entries of G 's adjacency matrix and the variables x , and then we require (2.2) to hold for all simple functions. It turns out that once you do so, the choice for the pseudodistribution is essentially determined, and hence proving the main result amounts to showing that it satisfies the constraints of the SOS algorithm. In the next section, we will outline the main ideas of our proof.

Remark 2.1 (planted clique vs. 3XOR). In light of the discussion above, it is instructive to consider the case of random 3XOR discussed before. Random 3XOR instances on n variables and $\Theta(n)$ constraints are easily seen to be maximally unsatisfiable (that is, at most $\approx 1/2$ constraints can be satisfied by any assignment) with high probability. On the other hand, Grigoriev [Gri01] constructed an SOS pseudoexpectation that pretends that such instances are satisfiable with high probability, proving an SOS lower bound for refuting random 3XOR formulas.

Analogous to the planted distribution $G(n, 1/2, \omega)$, one can define a natural planted distribution over 3XOR instances—roughly speaking, this corresponds to first choosing a random Boolean assignment x^* to n variables and then sampling random 3XOR constraints conditioned on being consistent with x^* . It is not hard to show that pseudocalibrating with respect to this planted distribution à la (2.2) produces precisely the pseudoexpectation that Grigoriev constructed. However, unlike in the

planted clique case, in the case of 3XOR, the pseudocalibration condition implies that for every low-degree monomial x_S , either the value of x_S is completely fixed (if it can be derived via low-width resolution from the 3XOR equations of the instance) or it is completely unconstrained.

The pseudoexpectations considered in previous works [FK03, MPW15, DM15] are similar to Grigoriev's construction in the sense that they essentially respect only strong constraints (e.g., if A is not a clique in the graph, then the probability that it is contained in the planted clique is zero) but other than that assume that variables are independent. However, unlike the 3XOR case, in the planted clique problem, respecting these strong constraints is not enough to achieve the pseudocalibration condition (2.2) and the pseudoexpectation of [FK03, MPW15, DM15] can be shown to violate weak probabilistic constraints imposed by (2.2) even at degree four. See Observation 2.5 for an example.

2.2. From calibrated pseudodistributions to SOS lower bounds. What does Bayesian inference and calibration have to do with SOS? In this section, we show how calibration is almost forced on any pseudodistribution feasible for the SOS algorithm. In order to show that the degree d SOS algorithm fails to certify that a random graph does not contain a clique of size ω , we need to show that for a random G , with high probability we can come up with an operator that maps a degree at most d , n -variate polynomial p to a real number $\tilde{\mathbb{E}}_G p$ satisfying the following constraints:

1. (Linearity) The map $p \mapsto \tilde{\mathbb{E}}_G p$ is linear.
2. (Normalization) $\tilde{\mathbb{E}}_G 1 = 1$.
3. (Boolean constraint) $\tilde{\mathbb{E}}_G x_i^2 p = \tilde{\mathbb{E}}_G x_i p$ for every p of degree at most $d - 2$ and $i \in [n]$.
4. (Clique constraint) $\tilde{\mathbb{E}}_G x_i x_j p = 0$ for every (i, j) that is not an edge and for every p of degree at most $d - 2$.
5. (Size constraint) $\tilde{\mathbb{E}}_G \sum_{i=1}^n x_i = \omega$.
6. (Positivity) $\tilde{\mathbb{E}}_G p^2 \geq 0$ for every p of degree at most $d/2$.

Remark 2.2 (“optimization” vs. “feasibility” formulation of planted clique). The above requirements on the operator correspond to showing that the SDP relaxation of the natural polynomial optimization formulation of the Max-Clique problem has the appropriate integrality gap for random instances drawn according to the $G(n, 1/2)$ distribution: find x satisfying (i) $x_i^2 = x_i$ for every $i \leq n$, and (ii) $x_i x_j = 0$ if and only if $\{i, j\}$ is not an edge in the graph that maximizes $\sum_{i \leq n} x_i$.

In the feasibility polynomial formulation of the problem, we replace the objective by the constraint that $\sum_{i \leq n} x_i = \omega$. Formally, this feasibility program is tighter than the optimization version. We note that our lower bound formally applies only to the optimization version at present.

DEFINITION 2.3. A map $p \mapsto \tilde{\mathbb{E}}_G p$ satisfying constraints 1–6 above is called a degree d pseudodistribution (with respect to the planted clique problem with parameter ω).

We can now restate our main result as follows.

THEOREM 2.4 (Theorem 1.1, restated). *There is some constant c such that if $\omega \leq n^{1/2 - c(d/\log n)^{1/2}}$, then with high probability over G sampled from $G(n, 1/2)$, there is a degree d pseudodistribution $\tilde{\mathbb{E}}_G$ satisfying constraints 1–6 above.*

Note that all of these constraints would be satisfied if $\tilde{\mathbb{E}}_G p$ was obtained by taking the expectation of p over a distribution on ω -sized cliques in G . However,

with high probability there is no $2.1 \log n$ -sized clique in G (and let alone a roughly \sqrt{n} -sized one), and so we will need a completely different mechanism to obtain such a pseudodistribution.

Previously, the choice of the pseudodistribution seemed to require a “creative guess” or an “ansatz.” For problems such as random 3SAT, this guess was fairly natural and almost “forced,” while for planted clique as well as some related problems [MW15] the choice of the pseudodistribution seemed to have more freedom, and more than one choice appeared in the literature.

For example, Feige and Krauthgamer [FK03] (henceforth, Feige–Krauthgamer (FK)) defined a very natural pseudodistribution $\tilde{\mathbb{E}}^{FK}$ for a weaker hierarchy. For a graph G on n vertices and for a subset $A \subseteq [n]$, $\tilde{\mathbb{E}}_G^{FK} x_A$ is equal to zero if A is not a clique in G and equal to $2^{\binom{|A|}{2}} \left(\frac{\omega}{n}\right)^{|A|}$ if A is a clique, and it is extended to degree d polynomials using linearity.⁴ Feige and Krauthgamer [FK03] showed that for every d and for $\omega < O(\sqrt{n/2^d})$, this pseudodistribution satisfies constraints 1–5 as in Definition 2.3 as well as a weaker version of positivity (this amounts to the so-called Lovász–Schrijver + SDP). Meka, Potechin, and Wigderson [MPW15] proved that the same pseudodistribution satisfies all constraints among 1–6 (and hence is a valid degree d pseudodistribution) as long as $\omega < \tilde{O}(n^{1/d})$. This bound on ω was later improved to $\tilde{O}(n^{1/3})$ for $d = 4$ by Deshpande and Montanari [DM15] and to $\tilde{O}(n^{(\lfloor d/2 \rfloor + 1)^{-1}})$ for a general d by Hopkins, Kothari, and Potechin [HKP15].

Interestingly, the FK pseudodistribution does *not* satisfy the full positivity constraint for larger values of ω . The issue is that while the FK pseudodistribution satisfies the “strong” constraint that $\tilde{\mathbb{E}}_G^{FK} x_A = 0$ if A is not a clique, it does not satisfy weaker constraints that are implied by (2.2). For example, for every constant ℓ , if vertex i participates in \sqrt{n} more ℓ -cliques than the expected number, then one can compute that the conditional probability of i belonging in the clique should be a factor $1 + c\omega/\sqrt{n}$ larger for some constant $c > 0$. However, the FK pseudodistribution does not make this correction. In particular, for every ℓ , there is a simple polynomial that shows that the FK pseudoexpectation is not calibrated.

OBSERVATION 2.5. Fix $i \in [n]$, and let ℓ be some constant. If $p_G = (\sum_j G_{i,j} x_j)^\ell$, then (i) $\mathbb{E}_{G \sim G(n, 1/2)} \tilde{\mathbb{E}}_G^{FK}[p_G^2] \leq \omega^\ell$ and (ii) $\mathbb{E}_{(G,x) \sim G(n, 1/2, \omega)}[p_G(x)^2] \geq \frac{\omega^{2\ell+1}}{n}$. In particular, when $\omega \gg n^{\frac{1}{2\ell+1}}$, $\mathbb{E}_{G \sim G(n, 1/2)} \tilde{\mathbb{E}}_G^{FK}[p_G^2] \ll \mathbb{E}_{(G,x) \sim G(n, 1/2, \omega)} p_G(x)$.

Proof sketch. For (ii), note that with probability (ω/n) vertex i is in the clique, in which case $\sum_j G_{i,j} x_j = \omega$, and hence the expectation of p_G^2 is at least $(\omega/n)\omega^{2\ell}$. For (i), we open up the expectation and the definition to get (up to a constant depending on ℓ)

$$\sum_{j_1, \dots, j_{2\ell}} G_{i,j_1} \dots G_{i,j_{2\ell}} (\omega/n)^{2\ell} \mathbb{E}_{G \sim G(n, 1/2)} 1_{\{i_1, \dots, i_{2\ell}\} \text{ is clique}}.$$

Since this expectation is zero unless every variable $G_{i,j}$ is squared, in which case the number of distinct j 's is at most ℓ , we can bound the sum by $n^\ell (\omega/n)^\ell = \omega^\ell$. This completes the proof sketch. \square

Observation 2.5 captures the failure of calibration for a specific polynomial $p_G(x)$ where the coefficients are low-degree functions of the graph G . The polynomial p_G

⁴The actual pseudodistribution used in [FK03] (and the followup works [MPW15, DM15]) was slightly different so as to satisfy $\tilde{\mathbb{E}}_G(\sum_{i=1}^m x_i)^\ell = \omega^\ell$ for every $\ell \in \{1, \dots, d\}$. This property is sometimes described as satisfying the constraint $\{\sum_i x_i = \omega\}$.

above can be used to show that degree d $\tilde{\mathbb{E}}^{FK}$ does not satisfy the positivity constraint for $\omega \gg n^{1/(\frac{d}{2}+1)}$. This observation is originally due to Kelner; see [HKP15]

FACT 2.6. *Let p_G be as in the Observation 2.5. Then there exists a C such that for $q = q_G = (C\omega^\ell x_S - p_G)$ with high probability over the graph $G \sim G(n, 1/2)$, $\tilde{\mathbb{E}}^{FK}[q_G^2] < 0$ for $\omega \gg n^{\frac{1}{\ell+1}}$.*

For the case $d = 4$, Hopkins et al. [HKP+16] proposed an “ad hoc” fix for the FK pseudodistribution that satisfies positivity up to $\omega = \tilde{O}(\sqrt{n})$ by explicitly adding a correction term to essentially calibrate for the low-degree polynomials q_G from Fact 2.6. However, their method did not extend even for $d = 6$ because of the sheer number of corrections that would need to be added and analyzed. Specifically, there are multiple families of polynomials such that their $\tilde{\mathbb{E}}^{FK}$ value departs significantly from their calibrated value in expectation and gives multiple points of failure of positivity in a manner similar to Observation 2.5 and Fact 2.6. Moreover, “fixing” these families by the correction as in the case of degree four leads to new families of polynomials that fail to achieve their calibrated value and exhibit negative pseudoexpectation for their squares and so on.

The *coefficients* of the polynomial p_G of Observation 2.5 are themselves low-degree polynomials in the adjacency matrix of G . This turns out to be a common feature in all the families of polynomials one encounters in the above works. Thus our approach is to fix all these polynomials *by fiat* by placing the constraint that the pseudodistribution must satisfy (2.2) for every such polynomial and using that as our implicit definition of the pseudodistribution. Indeed, it turns out that once we do so, the pseudodistribution is essentially determined. Moreover, (2.2) guarantees that it satisfies many of the “weak global constraints” that can be shown using Bayesian calculations.

Ultimately, we will construct the map $G \mapsto \tilde{\mathbb{E}}_G$ as a low-degree polynomial in G . Why is it OK to make such a restriction? One justification is the heuristic that the pseudodistribution itself must be simple since we know that it is efficiently computable (via the SOS algorithm) from the graph G . Another justification is that by forcing the pseudodistribution to be low-degree we are essentially making it *smooth* or “high entropy,” which is consistent with the Jaynes *maximum entropy principle* [Jay57a, Jay57b]. Most importantly—and this is the bulk of the technical work of this paper and the subject of the next subsection—this pseudodistribution can be shown to satisfy *all* constraints among 1–6 of Definition 2.3, including the positivity constraint.

We believe that this principled approach to designing pseudodistributions elucidates the power and limitations of the SOS algorithm in cases such as planted clique, where accounting for weak global correlations is a crucial aspect of the problem.

Remark 2.7 (*where does the planted distribution arise from?*). Theorem 2.4 (as well as Theorem 1.1) makes no mention of the planted distribution $G(n, 1/2, \omega)$ and only refers to an actual random graph. Thus it might seem strange that we base our pseudodistribution on the planted distribution via (2.2). One way to think about the planted distribution is that it corresponds to a *Bayesian prior* distribution on the clique. Note that this is the *maximum entropy* distribution on cliques of size ω , and so it is a natural choice for a prior per Jaynes’s principle of maximum entropy. Our actual pseudodistribution can be viewed as correcting this planted distribution to a posterior that respects simple inferences from the observed graph G .

2.3. Towards proving positivity: Structure vs. randomness. We have seen that pseudocalibration is desirable both a priori and in light of the failure of

previous lower-bound attempts. Now we turn to the following question: How do we formally define a pseudocalibrated linear map $\tilde{\mathbb{E}}_G$, and show that it satisfies constraints 1–6 with high probability, to yield Theorem 2.4?

We will require (2.2) to hold with respect to every function $f = f(G, x)$ that has degree at most τ in the entries of the adjacency matrix G and degree at most d in the variables x , and in addition we require that the map $G \mapsto \tilde{\mathbb{E}}_G$ is itself of degree at most τ in G ; then this completely determines $\tilde{\mathbb{E}}_G$. For any $S \subseteq [n]$, $|S| \leq d$, using the Fourier transform we can write $\tilde{\mathbb{E}}_G[x_S]$ as an explicit low-degree polynomial in G_e :

$$(2.3) \quad \tilde{\mathbb{E}}_G[x_S] = \sum_{\substack{T \subseteq \binom{[n]}{2} \\ |\mathcal{V}(T) \cup S| \leq \tau}} \left(\frac{\omega}{n}\right)^{|\mathcal{V}(T) \cup S|} \chi_T(G),$$

where $\mathcal{V}(T)$ is the set of nodes incident to the subset of edges (i.e., graph) T and $\chi_T(G) = \prod_{e \in T} G_e$. We carry out this computation in section 5. For $\omega \approx n^{0.5-\varepsilon}$, we will need to choose the truncation threshold $\tau \gtrsim d/\varepsilon$. It turns out that constraints 1–5 are easy to verify, and thus we are left with proving the *positivity constraint*. Indeed, this is not surprising, as verifying this constraint is always the hardest part of an SOS lower bound.

As is standard, to analyze this positivity requirement we work with the *moment matrix* of $\tilde{\mathbb{E}}_G$. Namely, let \mathcal{M} be the $\binom{n}{\leq d/2} \times \binom{n}{\leq d/2}$ matrix where $\mathcal{M}(I, J) = \tilde{\mathbb{E}}_G \prod_{i \in I} x_i \prod_{j \in J} x_j$ for every pair of subsets $I, J \subseteq [n]$ of size at most $d/2$. Our goal can be rephrased as showing that $\mathcal{M} \succeq 0$ (i.e., \mathcal{M} is positive semidefinite).

Given a (symmetric) matrix N , to show that $N \succeq 0$ our first hope might be to diagonalize N . That is, we would hope to find a matrix V and a diagonal matrix D so that $N = VDV^\dagger$. Then, as long as every entry of D is nonnegative, we would obtain $N \succeq 0$. Unfortunately, carrying this out directly can be far too complicated. Even the eigenvectors of simple random matrices are not completely understood, let alone matrices like ours with intricate dependencies among the entries. However, as the next example demonstrates, it is sometimes possible to prove positivity for a random matrix using what we call *approximate diagonalization*.

Example: Planted clique lower bound for $d = 2$ (a.k.a. basic SDP). Consider the problem of producing a pseudodistribution $\tilde{\mathbb{E}}$ satisfying constraints 1–6 of Definition 2.3, with $d = 2$. In this simple case, many subtleties can be safely ignored but can still provide some intuition. For $d = 2$, it is enough to define $\tilde{\mathbb{E}}x_i$ and $\tilde{\mathbb{E}}x_i x_j$ for every $i \in [n]$ and $\{i, j\} \subseteq [n]$. Let $\tilde{\mathbb{E}}x_i = (\omega/n)$ for every i , and set $\tilde{\mathbb{E}}x_i x_j$ to be $(\frac{\omega}{n})^2$ if (i, j) is an edge in G and zero otherwise. It is not hard to show that positivity reduces to showing that $\mathcal{N} \succeq 0$, where \mathcal{N} is the $n \times n$ matrix with $\mathcal{N}_{i,j} = \tilde{\mathbb{E}}x_i x_j$. Using standard results on random matrices, \mathcal{N} has one eigenvalue (whose corresponding eigenvector is close to the vector $u = (1/\sqrt{n}, \dots, 1/\sqrt{n})$) of value ω^2/n , while all others are distributed in the interval $\frac{\omega}{n} \pm O(\frac{\omega^2}{n^2}\sqrt{n})$, which is strictly positive as long as $\omega \ll \sqrt{n}$. Thus, while we cannot explicitly diagonalize \mathcal{N} , we have enough information to conclude that it is positive semidefinite. In other words, it was enough for us to get an *approximate diagonalization* for \mathcal{N} of the form $\mathcal{N} \approx \frac{\omega^2}{n}uu^\dagger + \frac{\omega}{n}Id + E$ for some sufficiently small (in the spectral norm) “error matrix” E . Ultimately, we will need to do something similar, but with many eigenvalues and many error matrices that are interdependent.

Approximate factorization for \mathcal{M} . We return now to the moment matrix \mathcal{M} for our (pseudo)calibrated pseudodistribution. Our goal is to give an approximate diagonalization of \mathcal{M} . There are several obstacles to doing so:

1. In the case $d = 2$, there was just one rank-1 approximate eigenspace to be handled. The number of these approximate eigenspaces will grow with d , and so we will need a more generic way to handle them.
2. Each approximate eigenspace corresponds to a family of polynomials $\{p\}$ whose calibrated pseudoexpectations are all roughly equal. (In the case $d = 2$, the only interesting polynomial was the polynomial $\sum_j x_j$, whose coefficients are proportional to the vector $u = (1/\sqrt{n}, \dots, 1/\sqrt{n})$.) As we saw in Observation 2.5, if p_G is a polynomial whose coefficients depend on the graph G , even in simple ways, the calibrated value $\tilde{\mathbb{E}}_G p_G$ may also depend substantially on the graph. Thus, when we write $\mathcal{M} \approx \mathcal{L} \mathcal{Q} \mathcal{L}^\dagger$ for some approximately diagonal matrix \mathcal{Q} , we will need the structured part $\mathcal{L} = \mathcal{L}(G)$ to itself be graph-dependent.
3. The errors in our diagonalization of \mathcal{M} —corresponding in our $d = 2$ example to the matrix E —will not be small enough to ignore as we did above. Instead, each error matrix will itself have to be approximately diagonalized, recursively until these errors are driven down sufficiently far in magnitude.

We now discuss at a high level our strategy to address items 1 and 2. The resolution to item 3 is the most technical element of our proof, and we leave it for later. Consider the vector space of all polynomials $f : \{0, 1\}^{\binom{n}{2}} \times \mathbb{R}^n \rightarrow \mathbb{R}$ which take a graph and an n -dimensional real vector and yield a real number. (We write $f_G(x)$, where G is the graph and $x \in \mathbb{R}^n$.) If we restrict our attention to the subspace of those of degree at most d in x , we obtain the polynomials in the domain of our operator $\tilde{\mathbb{E}}_G$. If we additionally restrict ourselves to the subspace of polynomials which are low-degree in G , we obtain the family of polynomials so that $\mathbb{E}_G \tilde{\mathbb{E}}_G f_G(x)$ is calibrated. Call this subspace \mathcal{V} .

Our goal would be to find an approximate diagonalization for all the nontrivial eigenvalues of \mathcal{M} using only elements from \mathcal{V} . The advantage of doing so is that for every $f \in \mathcal{V}$, we can calculate $\mathbb{E}_G \tilde{\mathbb{E}}_G f_G^2$ using the pseudocalibration condition (2.2). In particular, it means that if we find a function f such that f_G is with high probability an approximate eigenvector of \mathcal{M} , then we can compute the corresponding expected eigenvalue $\lambda(f)$.

A crucial tool in finding such an approximate eigenbasis is the notion of *symmetry*. For every f , if f' is obtained from f via a permutation of the variables x_1, \dots, x_n , then $\mathbb{E}_G \tilde{\mathbb{E}}_G f_G^2 = \mathbb{E}_G \tilde{\mathbb{E}}_G f'_G{}^2$. The result of this symmetry, for us, is that our approximate diagonalization requires only a constant (depending on d) number of eigenspaces. This argument allows us to restrict our attention to a constant number of classes of polynomials, where each class is determined by some finite graph U that we call its *shape*. For every polynomial f with shape U , we compute (approximately) the value of $\mathbb{E}_G \tilde{\mathbb{E}}_G f_G^2$ as a function of a simple combinatorial property of U , and our approximate eigenspaces correspond to polynomials with different shapes.

We can show that that in expectation our approximate eigenspaces will have non-negative eigenvalues since the pseudocalibration condition (2.2) in particular implies that for every f that is low-degree in both G and x , $\mathbb{E}_G \tilde{\mathbb{E}}_G f_G^2 \geq 0$. However, the key issue is to deal with the error terms that arise from the fact that these are only approximate eigenspaces. One could hope that, like in other “structure vs. randomness” partitions, this error term is small enough to ignore. Alas, this is not the case, and we need to handle it recursively, which is the crux of item 3 and the cause of much of the technical complications of our paper.

Remark 2.8 (structure vs. randomness). At a high level, our approach can be viewed as falling into the general paradigm of “structure vs. randomness,” as discussed by Tao [Tao05]. The general idea of this paradigm is to separate an object O into a “structured” part that is simple and predictable, and a “random” part that is unpredictable but has small magnitude or has some global statistical properties.

One example of this is the Szemerédi regularity lemma [Sze78], as well as variants such as [FK96] that partition a matrix into a sum of a low rank and pseudorandom components. Another example arises from the random models for the *primes* (see, e.g., [Tao15, Gra95]). These can be thought of as positing that, as far as certain simple statistics are concerned, (large enough) primes can be thought of as being selected randomly, conditioned on not being divisible by 2, 3, 5, etc., up to some bound w .

All these examples can be viewed from a computationally bounded Bayesian perspective. For every object O , we can consider the part of O that can be inferred by a computationally bounded observer to be O ’s *structured* component, while the remaining uncertainty can be treated as if it is *random*, even if in actuality it is fully determined. Thus, in our case, even though for almost every particular graph G from $G(n, 1/2, \omega)$ the clique x is fully determined by G , we still think of x as having a “structured” part which consists of all the inferences a “simple” observer can make from G (e.g., if i and j are nonneighbors, then $x_i x_j = 0$) and a “random” part that consists of the remaining uncertainty. As in other cases of applying this paradigm, part of the technical work is bounding the magnitude (in our case in the spectral norm) that arises from the “random” part, though as mentioned above in our case we need a particularly delicate control of the error terms, which ends up causing much of the technical difficulty.

3. Proving positivity: A technical overview. We now discuss in more detail how we prove that the *moment matrix* \mathcal{M} corresponding to our pseudodistribution is positive semidefinite. Recall that this is the $\binom{n}{\leq d/2} \times \binom{n}{\leq d/2}$ matrix \mathcal{M} such that $\mathcal{M}(I, J) = \tilde{\mathbb{E}}_G \prod_{i \in I} x_i \prod_{j \in J} x_j$ for every pair of subsets $I, J \subseteq [n]$ of size at most $d/2$ and that it is defined via (2.3) as

$$(3.1) \quad \mathcal{M}(I, J) = \sum_{\substack{T \subseteq \binom{[n]}{2} \\ |V(T) \cup I \cup J| \leq \tau}} \left(\frac{\omega}{n}\right)^{|V(T) \cup I \cup J|} \chi_T(G).$$

The matrix \mathcal{M} is generated from the random graph G , but its entries are *not* independent. Rather, each entry is a polynomial in G_e , and there are some fairly complex dependencies between them. Indeed, these dependencies will create a spectral structure for \mathcal{M} that is very different from the spectrum of standard random matrices with independent entries and makes proving that \mathcal{M} is positive semidefinite challenging. Our approach to showing that \mathcal{M} is positive semidefinite is through a type of “symbolic factorization” or “approximate diagonalization,” which we explain next.

3.1. Warm up. It is instructive to begin with the tight analysis presented in [HKP15] of the moments constructed in [MPW15].⁵ These moments can in fact be obtained by using truncation threshold $\tau = |S|$ in (2.3). This choice of τ is the smallest possible for which the resulting construction satisfies the hard clique constraints. The authors of [HKP15] show that this construction satisfies positivity for $\omega \lesssim n^{1/(\frac{d}{2}+1)}$.

⁵The construction in [MPW15] actually also satisfies $\sum x_i = \omega$ as a constraint which causes the precise form to differ. We ignore this distinction here.

For the purpose of this overview, let us work with the principal submatrix F indexed by subsets I and J of size exactly d . The analysis in [HKP15] proceeds by first splitting F into $d+1$ components $F = F_0 + F_1 + \dots + F_d$, where $F_i(I, J) = F(I, J)$ if $|I \cap J| = i$ and 0 otherwise. Below, we discuss two of the key ideas involved that will serve as an inspiration for us.

As discussed before, we must approximately diagonalize the matrix F in the sense that the off-diagonal blocks must be “small enough” to be charged to the on-diagonal blocks. Thus the main question before us is how to obtain an (approximate) understanding of the spectrum of F that allows us to come up with a “change of basis” in which the off-diagonal blocks are small enough to be charged to the positive eigenmass in the on-diagonal blocks.

Let us consider the piece F_0 for our discussion here. As alluded to in section 3, we want to break F into minimal pieces so that each piece is symmetric under the permutation of vertices. We can hope that each piece will then essentially have a single dominating eigenvalue that can be determined relatively easily. Below, we will essentially implement this plan.

First, we need to decide what kind of “pieces” we will need. These are the *graphical matrices* that we define next.

DEFINITION 3.1 (graphical matrices (see Definition 7.6 for a formal version)). *Let U be a graph on $[2d]$ with specially identified left and right subsets $[d]$ and $[2d] \setminus [d]$. For any $I, J \in \binom{[n]}{d}$, $I \cap J = \emptyset$, let $\pi_{I,J}$ be an injective map that takes $[d]$ into I and $[2d] \setminus [d]$ into J using a fixed convention. The graphical matrix M_U with graph U is then defined by $M_U(I, J) = \chi_{\pi_{I,J}(U)}(G)$.*

The starting point of the analysis is to decompose $F_0 = \sum_U \left(\frac{\omega}{n}\right)^{2d} M_U$, where M_U is the graphical matrix with shape U . Graphical matrices as above turn out to be the right building blocks for spectral analysis of our moment matrix. This is because a key observation in [HKP15] shows that a simple combinatorial parameter, the size of the maximum bipartite matching between the left and right indexes in U (i.e., between $[d]$ and $[2d] \setminus [d]$), determines the spectral norm of M_U . Specifically, when U has a maximum matching of size $t < d$, the spectral norm of M_U is $\tilde{O}(n^{d-\frac{t}{2}})$, with high probability. Observe that when $d = 2$ and U is a single edge connecting the left vertex with the right, M_U is just the $\{-1, 1\}$ -adjacency matrix of the underlying random graph and it is well known that the spectral norm in this case is $\Theta(\sqrt{n})$, matching the more general claim above.

In particular, this implies that when U has a perfect matching, M_U is pseudo-random in the sense that F_U essentially has the spectral norm $\approx n^{d/2}$, the same as that of an independent $\{-1, 1\}$ random matrix of the same dimensions. This allows M_U to be bounded against the positive eigenvalue $\left(\frac{\omega}{n}\right)^d$ of the diagonal matrix F_d as $\left(\frac{\omega}{n}\right)^d \gg \left(\frac{\omega}{n}\right)^{2d} n^{d/2}$ (even for ω approaching $\sqrt{n}(!)$). However, for M_U when U has a maximum matching of size $t < d$, one can't bound against the diagonal matrix F_d anymore.

The next main idea is to note that for every M_U there's an appropriate “diagonal” against which we must charge the negative eigenvalues of M_U . When U has a perfect matching, this is literally the diagonal matrix F_d , as done above. However, when, say, U is a (bipartite) matching of size $t < d$, we should instead charge against the “diagonal” matrix that can be thought of as obtained by “collapsing” each matching edge into a vertex in U (see Figure 3.1). In particular, this collapsing produces a matrix that lies in the decomposition of F_t .

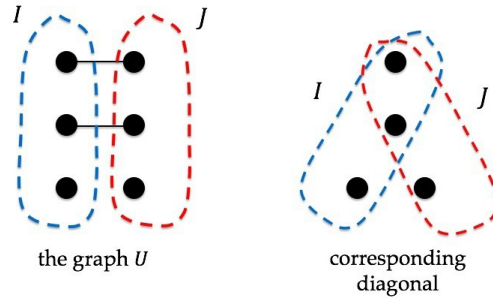


FIG. 3.1.

There are two main takeaways from this analysis that would serve as inspiration in the analysis of our actual construction. The first is the decomposition into graphical matrices in order to have a coarse handle on the spectrum of the moment matrix. The second is the “charging” of negative eigenvalues against appropriate “diagonals,” which is essentially governed by the combinatorics of matchings in U .

3.2. The main analysis. We can now try to use the lessons from the warm up analysis to inspire our actual analysis. To begin with, we recall that each graphical matrix was obtained by choosing an appropriate (set of) Fourier monomials for any entry indexed by I, J . However, since for our actual construction we have monomials of much higher degree, we need to extend the notion of graphical matrices with *shapes* corresponding to larger graphs U . See Definition 7.6 for a formal definition.

It turns out that the right combinatorial idea to generalize the size of the maximum matching and control the spectral norm of the graphical matrices \mathcal{M}_U is the maximum number of *vertex disjoint paths* between specially designated left and right endpoints of U (themselves the generalization of the bipartition we had in the warm up). Using Menger’s theorem, this is equal to the size of a minimal collection of vertices that separates the left and right sets in the graph U , which we call the *separator size* of U .

Finally, we need a “charging” argument to work with the approximate diagonalization we end up with. Generalizing the idea in the warm up here is the hardest part of our proof, but it relates again to the notion of vertex separators defined above. In the warm up, we used a naive charging scheme, breaking the moment matrix into simpler (graphical) matrices, each of which was either a “positive diagonal” mass or a “negative off-diagonal mass,” and pairing up the terms. Such a crude association doesn’t work out immediately in the general setting. Instead, large groups of graphical matrices must be treated all at once. In each subspace of our approximate diagonalization of the moment matrix \mathcal{M} , we collect the “positive diagonal mass” and the “negative off-diagonal mass” that need to be charged to it together and build an approximately positive semidefinite matrix out of it. As alluded to before, the error in this approximation is not negligible, and thus we must further recurse on the error terms. In what follows, we discuss the factorization process that accomplishes the charging scheme implicitly and the recursive factorization for the error terms in some more detail. Consider some graph $T \subseteq \binom{[n]}{2}$ that corresponds to one term in the sum in (3.1) above, and let q be the minimum size of a set that separates I from J in T . Such a set is not necessarily unique, but we can define the *leftmost* separator

$\text{left-sep}(T) = S_\ell$ to be the q -sized separator that is “closest” to I and the *rightmost* separator $\text{right-sep}(T) = S_r$ to be the q -sized separator that is “closest” to J .

We can rewrite the (I, J) entry moment matrix \mathcal{M} (3.1) by collecting monomials T with a fixed choice of the leftmost and rightmost separators S_ℓ and S_r . This step corresponds to collecting terms with similar spectral norms together, accomplishing the goal of collecting together into a term the “positive diagonal mass” and the “negative off-diagonal mass” that are implicitly charged to each other in the intended approximate diagonalization.

$$(3.2) \quad \mathcal{M}(I, J) = \sum_{1 \leq q \leq |I|, |J|} \sum_{S_\ell, S_r: |S_\ell| = |S_r| = q} \sum_{\substack{T \subseteq \binom{[n]}{2} \\ |\mathcal{V}(T) \cup I \cup J| \leq \tau \\ \text{left-sep}(T) = S_\ell, \text{right-sep}(T) = S_r}} \left(\frac{\omega}{n}\right)^{|\mathcal{V}(T) \cup I \cup J|} \chi_T(G).$$

We can then partition T into three subsets \mathcal{R}_ℓ , \mathcal{R}_m , and \mathcal{R}_r that represent the part of the graph T between I and S_ℓ , the part between S_ℓ and S_r , and the part between S_r and J , respectively (where edges within S_ℓ and edges within S_r are all placed in \mathcal{R}_m ; see Definition 6.5). We thus immediately obtain that

$$\chi_T(G) = \chi_{\mathcal{R}_\ell}(G) \chi_{\mathcal{R}_m}(G) \chi_{\mathcal{R}_r}(G).$$

Thus,

$$(3.3) \quad \begin{aligned} \mathcal{M}(I, J) = & \sum_{1 \leq q \leq |I|, |J|} \sum_{S_\ell, S_r: |S_\ell| = |S_r| = q} \sum_{\substack{T \subseteq \binom{[n]}{2} \\ |\mathcal{V}(T) \cup I \cup J| \leq \tau \\ \text{left-sep}(T) = S_\ell \\ \text{right-sep}(T) = S_r}} \left(\left(\frac{\omega}{n}\right)^{|\mathcal{V}(\mathcal{R}_\ell)|} \chi_{\mathcal{R}_\ell}(G) \right) \\ & \times \left(\left(\frac{\omega}{n}\right)^{|\mathcal{V}(\mathcal{R}_m)| - 2q} \chi_{\mathcal{R}_m}(G) \right) \left(\left(\frac{\omega}{n}\right)^{|\mathcal{V}(\mathcal{R}_r)|} \chi_{\mathcal{R}_r}(G) \right). \end{aligned}$$

One could hope that we could replace the right-hand side of (3.3) by

$$(3.4) \quad \begin{aligned} & \sum_{\substack{1 \leq q \leq |I|, |J| \\ \tau_1 + \tau_2 + \tau_3 \leq \tau}} \sum_{\substack{S_\ell \subseteq \binom{[n]}{q} \\ S_r \subseteq \binom{[n]}{q}}} \left(\sum_{\substack{\mathcal{R}_\ell \\ \mathcal{V}(\mathcal{R}_\ell) \supseteq I \cup S_\ell \\ |\mathcal{V}(\mathcal{R}_\ell)| = \tau_1}} \left(\frac{\omega}{n}\right)^{|\mathcal{V}(\mathcal{R}_\ell)|} \chi_{\mathcal{R}_\ell}(G) \right) \\ & \times \left(\sum_{\substack{\mathcal{R}_m \\ \mathcal{V}(\mathcal{R}_m) \supseteq S_\ell \cup S_r \\ |\mathcal{V}(\mathcal{R}_m)| = \tau_2}} \left(\frac{\omega}{n}\right)^{|\mathcal{V}(\mathcal{R}_m)| - 2q} \chi_{\mathcal{R}_m}(G) \right) \left(\sum_{\substack{\mathcal{R}_r \\ \mathcal{V}(\mathcal{R}_r) \supseteq S_r \cup J \\ |\mathcal{V}(\mathcal{R}_r)| = \tau_3}} \left(\frac{\omega}{n}\right)^{|\mathcal{V}(\mathcal{R}_r)|} \chi_{\mathcal{R}_r}(G) \right). \end{aligned}$$

In fact, it turns out we can focus attention (up to sufficiently small error in the spectral norm) to the case $\tau_1 \leq \tau/3$, $\tau_2 \leq \tau/3$, $\tau_3 \leq \tau/3$, in which case if $M(I, J)$ was equal to (3.4), we could simply write

$$\mathcal{M} = \sum_q \mathcal{L}_q \mathcal{Q}_q \mathcal{L}_q^\dagger,$$

where for $I, S \subseteq [n]$ with $|I| \leq d$ and $|S| = q$ we let $\mathcal{L}_q(I, S)$ be the sum of $(\omega/n)^{|V(\mathcal{R}_\ell)|} \chi_{\mathcal{R}_\ell}(G)$ over all graphs \mathcal{R}_ℓ of at most $\tau/3$ vertices connecting I to S , and for S, S' of size q , we let $\mathcal{Q}_q(S, S')$ be the sum of $(\omega/n)^{|\mathcal{R}_m|-2q} \chi_{\mathcal{R}_m}(G)$ over all graphs \mathcal{R}_m of at most $\tau/3$ vertices connecting S to S' .

Thus, in this case, this reduces our task of showing that \mathcal{M} is positive semidefinite to showing that for every q , the matrix $\mathcal{Q} = \mathcal{Q}_q$ is positive semidefinite. However, the main complication is that there are cross terms in the product $\mathcal{L}_q \mathcal{Q}_q \mathcal{L}_q^\dagger$ that correspond to repeating the same vertex (not in S_ℓ and S_r) in more than one of \mathcal{R}_ℓ , \mathcal{R}_m , and \mathcal{R}_r . There is no matching term in the Fourier decomposition of $\mathcal{M}(I, J)$. So, at best, for every fixed q , we can write the part of \mathcal{M} corresponding to indices I, J with minimal vertex separator equal to q as

$$\mathcal{L} \mathcal{Q}_0 \mathcal{L}^\dagger - \mathcal{E}_1$$

for some error matrix \mathcal{E}_1 that exactly cancels out the extra terms contributed by cross terms with repeated vertices. Unfortunately, the spectral norm of this error matrix \mathcal{E}_1 is not small enough that we could simply ignore it. Luckily, however, we can recurse and factorize \mathcal{E}_1 approximately as well. We can form a new graph T' by taking the parity of the edge sets in \mathcal{R}_ℓ , \mathcal{R}_m , and \mathcal{R}_r . Now we find the leftmost and rightmost separators that separate I and J from each other and from all repeated vertices. This gives us another decomposition of a graph into three pieces, from which we can write

$$\mathcal{E}_1 = \mathcal{L} \mathcal{Q}_1 \mathcal{L}^\dagger - \mathcal{E}_2$$

for some other matrix \mathcal{Q}_1 . Continuing this argument gives us for every q a factorization of \mathcal{M}_q as

$$\mathcal{L}(\mathcal{Q}_0 - \mathcal{Q}_1 + \mathcal{Q}_2 - \cdots - \mathcal{Q}_{2d-1} + \mathcal{Q}_{2d}) \mathcal{L}^\dagger - (\xi_0 - \xi_1 + \xi_2 - \cdots - \xi_{2d-1} + \xi_{2d}).$$

The error matrices $\xi_0, \xi_1, \dots, \xi_{2d}$ arise from truncation issues, which we have ignored in the argument above, and turn out to be negligible.

It is not hard to show that $\mathcal{Q}_0 \succeq D$ for some positive semidefinite matrix D that we define later. What remains is to bound the remaining matrices $\mathcal{Q}_1, \dots, \mathcal{Q}_{2d-1}$ in order to conclude that \mathcal{M} is positive semidefinite. Next, we elaborate on the structure of these matrices. It turns out that we can define the “shape” of a graph \mathcal{R}_m in an appropriate way so that

$$\mathcal{Q}_i^U(S_\ell, S_r) = \sum_{\text{shape}(\mathcal{R}_m)=U} c_i(\mathcal{R}_m) \chi_{\mathcal{R}_m},$$

where U is a finite (for constant d) sized graph with vertex set $A \cup B \cup C$, where we call A the “left” side of U and B the “right” side of U . Moreover, $\mathcal{Q}_i = \sum_U \mathcal{Q}_i^U$. Now \mathcal{Q}_i^U is a random matrix and special cases of this general family of matrices (for particular choices of U) arise in several earlier works on lower bounds for planted clique. Medarametla and Potechin [MP16] showed that the spectral norm of \mathcal{Q}^U can be controlled by a bound on its coefficients and a few combinatorial parameters of U , namely $|\mathcal{V}(U)|$, $|A \cap B|$, and the number of vertex disjoint paths between A/B and B/A .

A major challenge in our work is to understand and analyze the coefficients c_i . In the course of decomposing \mathcal{M} , we are able to characterize $c_i(\mathcal{R}_m)$ as an appropriately weighted sum over $c_{i-1}(\mathcal{R}'_m)$, where \mathcal{R}'_m ranges over the middle piece of all graphs

with leftmost and rightmost separators S_ℓ and S_r that could have resulted in \mathcal{R}_m due to repeated vertices. Recall that when there are repeated vertices, we take the parity of the edge sets of the three pieces and compute a new set of left and rightmost vertex separators. The set of \mathcal{R}'_m 's that could result in \mathcal{R}_m is complicated. Instead, our approach is to show that the various combinatorial parameters of \mathcal{R}'_m (which affect the spectral norm bounds) trade off against each other when accounting for the effect of repeated vertices. This allows us to bound their contribution and ultimately show that the coefficients c_i decay quickly enough for all values of $\omega < n^{1/2-\varepsilon}$ that we can bound each \mathcal{Q}_i for $i > 1$ as $-\frac{D}{8d} \succeq \mathcal{Q}_i \succeq \frac{D}{8d}$, and this completes our proof.

4. Preliminaries.

4.1. General notation.

- We use small Greek letters to indicate constants/parameters.
- \mathcal{P}_d^n denotes the linear space of all *multilinear* polynomials of degree at most d on $\{0, 1\}^n$.
- We write $\mathbf{1}_Q$ for any event Q to be the 0–1 indicator of whether Q happens.
- For a subset $T \subseteq \binom{[n]}{2}$ of edges of a graph on vertex set $[n]$, we write $\mathcal{V}(T) \subseteq [n]$ to denote the vertices that have at least one edge incident on them in T .
- For a matrix $Q \in \mathbb{R}^{N \times N}$, $\|Q\|$ denotes its spectral norm (or the largest singular value) and $\|Q\|_F = \sqrt{\sum_{x,y \in [N]} Q(x,y)^2}$ denotes its Frobenius norm.
- For a graph G , let $\mathcal{C}_q = \mathcal{C}_q(G) = \{I \subseteq [n] : I \text{ is a } q\text{-clique in } G\}$, and let $\mathcal{C}_{\leq q} = \bigcup_{q' \leq q} \mathcal{C}_{q'}$. Let $\mathcal{C}(G) = \mathcal{C}_{\leq \infty}$ be the collection of all cliques in G . We count the empty set and all singletons as cliques.
- We write $\mathcal{G}(n, \frac{1}{2})$ to denote the distribution on graphs on the vertex set $[n]$ where each edge is included with probability $1/2$ independently of others.
- We say that an event E with respect to the probability distribution $\mathcal{G}(n, \frac{1}{2})$ happens *with high probability* if $\mathbb{P}[E] \geq 1 - \Omega(1)/n^{10 \log n}$ for large enough n .
- We write $f(n) \ll g(n)$ to mean that for every constant c there is an n_0 such that if $n \geq n_0$, then $f(n) \leq Cg(n)$.

4.2. Graphs. We identify a graph G with its $\{-1, 1\}$ adjacency matrix and write $G_e \in \{-1, 1\}$ for the $\{-1, 1\}$ -indicator of whether or not $e \in [n] \times [n]$ is an edge (indicated by $G_e = +1$) in the graph G . When $G \sim \mathcal{G}(n, \frac{1}{2})$, G_e are independent $\{-1, 1\}$ -random variables.

A *graph function* is a real-valued function of the variables $G_e \in \{-1, 1\}$ for $e \in \binom{[n]}{2}$. For graphs G^1, G^2, \dots, G^k on the vertex set $[n]$, we define $\Delta(G^1, G^2, \dots, G^k)$ to be the graph G satisfying $G_e = \prod_{i \leq k} G_e^i$.

DEFINITION 4.1 (vertex separator). *For a graph G on $[n]$ and vertex sets $I, J \subseteq [n]$, a set of vertices $S \subseteq [n]$ is said to be a minimal vertex separator if S is a set of smallest possible size such that every path between I and J in G passes through some vertex of S .*

Often, I and J will be allowed to intersect, in which case any vertex separator must contain $I \cap J$.

FACT 4.2 (Menger's theorem). *For a graph G on $[n]$ and two subsets of vertices $I, J \subseteq [n]$, the maximum number of vertex disjoint paths between I and J in G is equal to the size of any minimal vertex separator between I and J in G .*

4.3. Fourier analysis. Any graph function f can be represented as a Fourier polynomial in the variables G_e :

$$f(G) = \sum_{W \subseteq \binom{[n]}{2}} \widehat{f}(W) \chi_W(G),$$

where $\chi_W(G)$ is the *parity* function on edges in W :

$$\chi_W(G) = \prod_{e \in W} G_e.$$

The parity function χ_W is an orthonormal basis for functions on G under the inner product defined by $\langle f, h \rangle = \mathbb{E}_{G \sim G(n, \frac{1}{2})} [f(G)h(G)]$ for any graph functions f and h .

The following fact is easy to verify.

FACT 4.3. *Let G be a graph on n described by the vector $G \in \{-1, 1\}^{\binom{[n]}{2}}$. For any subset $S \subseteq [n]$ of the vertices, we have the following identity:*

$$\sum_{W \subseteq \binom{[n]}{2}} \chi_W(G) = \begin{cases} 2^{\binom{|S|}{2}} & \text{if } S \text{ is a clique in } G, \\ 0 & \text{otherwise.} \end{cases}$$

4.4. The SOS algorithm. The SOS algorithm has several equivalent definitions. We follow the notation of *pseudoexpectations* as in the survey of Barak and Steurer [BS14].

DEFINITION 4.4 (pseudoexpectation). *A linear operator $\tilde{\mathbb{E}} : \mathcal{P}_d^n \rightarrow \mathbb{R}$ is said to be a degree d -pseudoexpectation if it satisfies the following:*

1. *Normalization:* $\tilde{\mathbb{E}}[\mathbf{1}] = 1$.
2. *Positive semidefiniteness:* $\tilde{\mathbb{E}}[p^2] \geq 0$ for every polynomial $p \in \mathcal{P}_d^n$.

A pseudoexpectation operator $\tilde{\mathbb{E}}$ on \mathcal{P}_d^n is said to satisfy a constraint $\{p = 0\}$ for any $p \in \mathcal{P}_d^n$ if for every polynomial $q \in \mathcal{P}_d^n$ such that $p \cdot q \in \mathcal{P}_d^n$, $\tilde{\mathbb{E}}[pq] = 0$.

Given a set of constraints $\{p_i = 0\}$ for $1 \leq i \leq m$ and an objective polynomial p , the SOS algorithm of degree d solves the problem

$$\arg \max \tilde{\mathbb{E}}[p]$$

over all degree d pseudoexpectations $\tilde{\mathbb{E}}$ that satisfy $\{p_i = 0\}$ for $1 \leq i \leq m$.

5. The pseudoexpectation. We now define our pseudodistribution operator $\tilde{\mathbb{E}}_G$. As discussed in section 2.2, it is based on requiring (2.2) to hold for every f that has degree at most τ in G and d in x .

Important parameters. The following parameters will be fixed for the rest of the paper:

- $\varepsilon \in (0, 1/2)$, which determines the size $\omega = n^{1/2-\varepsilon}$ of the planted clique.
- $d = d(n) \in \mathbb{N}$, the degree of the SOS relaxation against which we prove a lower bound.
- $\tau = \tau(n) \in \mathbb{N}$, the degree of our pseudoexpectation $\tilde{\mathbb{E}}$ as a function of $G \sim G(n, 1/2)$.

We always assume that $Cd/\varepsilon \leq \tau \leq (\varepsilon/C) \log n$ and $\varepsilon \geq C \log \log n / \log n$ for a sufficiently large constant C . Eventually, we will set $d = (\varepsilon/C)^2 \log n$ (this yields the parameters stated in Theorem 1.1 since then $n^{1/2-\varepsilon} = n^{1/2-\Omega(d/\log n)^{1/2}}$), which implies that $\varepsilon \gg \log \log n / \log n$.

5.1. Definition of $\tilde{\mathbb{E}}$. As discussed previously, $\tilde{\mathbb{E}}$ is completely specified by its *multilinear moments*: $\tilde{\mathbb{E}}[x_I]$ for $I \subseteq [n]$ and $|I| \leq d$. $\tilde{\mathbb{E}}[x_I]$ is a function of G_e for $e \in \binom{[n]}{2}$ and can be written as a polynomial in G_e with coefficients $\widehat{\tilde{\mathbb{E}}[x_S]}(T)$ for each $T \subseteq \binom{[n]}{2}$ (the “Fourier coefficients”). These Fourier coefficients will be fixed by our insistence on the pseudoexpectation being pseudocalibrated with respect to the planted distribution $G(n, 1/2, \omega)$.

DEFINITION 5.1 ($\tilde{\mathbb{E}}$ of degree d , clique-sized ω , truncation τ). *Let $S \subseteq [n]$ be a set of vertices of size $|S| \leq d$. Let $T \subseteq \binom{[n]}{2}$ be a set of edges. Let $\chi_T = \prod_{e \in T} G_e$. Let*

$$\widehat{\tilde{\mathbb{E}}[x_S]}(T) = \begin{cases} \mathbb{E}_{(G,x) \sim G(n, 1/2, \omega)}[\chi_T(G)x_S] & \text{if } |\mathcal{V}(T) \cup S| \leq \tau, \\ 0 & \text{otherwise.} \end{cases}$$

As usual, $\tilde{\mathbb{E}}[x_S] = \sum_{T \subseteq \binom{[n]}{2}} \widehat{\tilde{\mathbb{E}}[x_S]}(T) \cdot \chi_T(G)$.

The Fourier coefficients can in fact be explicitly computed easily.

LEMMA 5.2. *Let $T \subseteq \binom{[n]}{2}$, $S \subseteq [n]$, and $\mathcal{V}(T) \subseteq [n]$ be the vertices incident to edges in T . Then*

$$\mathbb{E}_{(H,x) \sim G(n, 1/2, \omega)}[\chi_T \cdot x_S] = \left(\frac{\omega}{n}\right)^{|\mathcal{V}(T) \cup S|}.$$

Proof. Throughout this proof, we suppress explicit notation for the underlying random variable, which is $(H, x) \sim G(n, \frac{1}{2}, \omega)$. We claim that $\mathbb{E}[\chi_T \cdot x_S] = \mathbb{P}[x_{\mathcal{V}(T) \cup S} = 1]$. To see this, note that

$$(5.1) \quad \mathbb{E}[\chi_T \cdot x_S] = \mathbb{P}[x_{\mathcal{V}(T) \cup S} = 1] \cdot \mathbb{E}[\chi_T \cdot x_S \mid x_{\mathcal{V}(T) \cup S} = 0] \\ + (1 - \mathbb{P}[x_{\mathcal{V}(T) \cup S} = 1]) \cdot \mathbb{E}[\chi_T \cdot x_S \mid x_{\mathcal{V}(T) \cup S} = 1].$$

We note that the second term above is 0. It is easy to see whether $x_S = 0$. Otherwise, $x_{\mathcal{V}(T)} = 0$, and there is an edge $e \in T$ but not contained in the clique x . Thus,

$$\mathbb{E}[\chi_e \chi_{T \setminus e} \cdot x_S \mid x_{\mathcal{V}(T) \cup S} = 0] = 0.$$

If $x_{\mathcal{V}(T) \cup S} = 1$, then $\chi_T = 1$, and so $\mathbb{E}[\chi_T \cdot x_S \mid x_{\mathcal{V}(T) \cup S} = 1] = 1$. By a simple computation,

$$\mathbb{P}[x_{\mathcal{V}(T) \cup S} = 1] = \left(\frac{\omega}{n}\right)^{|\mathcal{V}(T) \cup S|}. \quad \square$$

As discussed in section 2.3, our construction of $\tilde{\mathbb{E}}$ is pseudocalibrated. The following lemma captures this formally. We include the (straightforward) proof in Appendix A.1.

LEMMA 5.3. *Let $f_G(x) = \sum_{|S| \leq 2d} c_S(G) \cdot x_S$ be a real-valued polynomial on $\{0, 1\}^n$ whose coefficients have degree at most τ when expressed in the ± 1 indicators G_e for edges in G . Then $\mathbb{E}_{G \sim G(n, \frac{1}{2})}[\tilde{\mathbb{E}}[f_G(x)]] = \mathbb{E}_{(H,x) \sim G(n, 1/2, \omega)}[f_H(x)]$.*

5.2. $\tilde{\mathbb{E}}$ satisfies constraints. We now show that the $\tilde{\mathbb{E}}$ defined in the previous section satisfies all linear constraints among (1)–(6) in section 2.2 and has an objective value of ω . That is, (1) $\tilde{\mathbb{E}}[1] \approx 1$, (2) $\tilde{\mathbb{E}}[\sum_{i \in [n]} x_i] \approx \omega$, and (3) $\tilde{\mathbb{E}}[x_S] = 0$ for every $S \subseteq [n]$ which is not a clique in G .

We analyze $\tilde{\mathbb{E}}[1]$ and $\tilde{\mathbb{E}}[\sum_{i \in [n]} x_i]$ in the next lemma and include a proof based on the moment method in Appendix A.2.

LEMMA 5.4. *With high probability, $\tilde{\mathbb{E}}[1] = 1 \pm n^{-\Omega(\varepsilon)}$ and $\tilde{\mathbb{E}}[\sum_{i \in [n]} x_i] = \omega \cdot (1 \pm n^{-\Omega(\varepsilon)})$.*

The next lemma shows that $\tilde{\mathbb{E}}[x_S] = 0$ whenever S is not a clique.

LEMMA 5.5. *With probability 1, if $S \subseteq [n]$ of size at most d is not a clique in G , then $\tilde{\mathbb{E}}[x_S] = 0$.*

Proof. Let $S \subseteq [n]$ have size at most d . Recall that

$$\mathbf{1}_{S \text{ is a clique in } G} = 2^{-\binom{|S|}{2}} \sum_{T \subseteq \binom{S}{2}} \chi_T.$$

Because the Fourier expansion of $\tilde{\mathbb{E}}[x_S]$ is truncated using the threshold $|\mathcal{V}(T) \cup S| \leq \tau$, two Fourier characters $\chi_T, \chi_{T'}$ have the same coefficient in $\tilde{\mathbb{E}}[x_S]$ if $T \oplus T' \subseteq \binom{S}{2}$. So we can factor $\tilde{\mathbb{E}}[x_S] = \mathbf{1}_{S \text{ is a clique in } G} \cdot f_S(G)$ for some function f_S . \square

5.3. Proof of main theorem. Our main technical claim is that $\tilde{\mathbb{E}} = \tilde{\mathbb{E}}_G$ is (approximately) positive semidefinite. That is, we have the following.

LEMMA 5.6. *With high probability over G from $G(n, 1/2)$, every $p \in \mathcal{P}_d$ satisfies*

$$\tilde{\mathbb{E}}_G[p(x)^2] \geq 0.$$

It is now easy to complete the proof of Theorem 1.1.

Proof of Theorem 1.1. By Lemmas 5.4, 5.5, and 5.6, there is a universal C so that if $Cd/\varepsilon \leq \tau \leq (1/C)\varepsilon \log n$ (by a union bound), then with high probability, the following hold:

1. $\tilde{\mathbb{E}}[1] = 1 \pm n^{-\Omega(\varepsilon)}$.
2. $\tilde{\mathbb{E}}[x_S] = 0$ for every S of size at most d not a clique in G .
3. $\tilde{\mathbb{E}}[\sum_i x_i] \geq (1 - n^{-\Omega(\varepsilon)})\omega$.
4. $\tilde{\mathbb{E}}[p(x)^2] \geq 0$ for every $p \in \mathcal{P}_d$.

Thus, choose $\varepsilon = (C^2 d / \log n)^{1/2}$ and $\tau = (1/C)\varepsilon \log n$. The operator given by $\tilde{\mathbb{E}}^*[p(x)] = \tilde{\mathbb{E}}[p(x)]/\tilde{\mathbb{E}}[1]$ is a valid degree d pseudodistribution with $\tilde{\mathbb{E}}[\sum_i x_i] \geq \Omega(n^{1/2 - \Theta(d/\log n)^{1/2}})$, as desired.

5.4. Proof plan. As is standard, we can reduce Lemma 5.6 to showing that the associated *moment matrix* is positive semidefinite.

DEFINITION 5.7 (moment matrix). *Let $\mathcal{M} \in \mathbb{R}^{\binom{[n]}{\leq d} \times \binom{[n]}{\leq d}}$ be given by $\mathcal{M}(I, J) = \tilde{\mathbb{E}}[x_I x_J]$.*

Thus, Lemma 5.6 is equivalent to showing the following.

LEMMA 5.8. *With high probability, $\mathcal{M} \succeq 0$.*

At a high level, our plan involves first getting an approximate factorization of the moment matrix $\mathcal{M} = \mathcal{L} \mathcal{Q}_0 \mathcal{L}^\dagger + \text{“error”}$ for appropriately defined matrices \mathcal{L} and \mathcal{Q}_0 . This step is the key technical part of the proof—given such a factorization, our task reduces to showing that \mathcal{Q}_0 and $\mathcal{L} \mathcal{L}^\dagger$ have large enough positive eigenvalues to compensate for the error. The first approximate factorization step will occupy us in section 6. The technical work in the second step involves showing upper bounds on the spectral norms of appropriately defined pieces of \mathcal{Q}_0 and is the content of section 7.

6. Approximate factorization of the moment matrix.

6.1. Ribbons and vertex separators. In this section, we get set up for the first step in the proof of Lemma 5.8 by setting up some definitions. *Ribbons* will play a crucial role in our analysis.

DEFINITION 6.1 (ribbon). An (I, J) -ribbon \mathcal{R} is a graph with edge set $W_{\mathcal{R}} \subseteq \binom{[n]}{2}$ and vertex set $V_{\mathcal{R}} \supseteq \mathcal{V}(W_{\mathcal{R}}) \cup I \cup J$ for two specially identified subsets $I, J \subseteq [n]$, each of size at most d , called the left and the right ends, respectively. We sometimes write $\mathcal{V}(\mathcal{R}) \stackrel{\text{def}}{=} V_{\mathcal{R}}$ and call $|\mathcal{V}(\mathcal{R})|$ the size of \mathcal{R} . Also, we write $\chi_{\mathcal{R}}$ for the monomial $\chi_{W_{\mathcal{R}}}$, where $W_{\mathcal{R}}$ is the edge set of the ribbon \mathcal{R} .

In our analysis, (I, J) -ribbons arise as the terms in the Fourier decomposition of the entry $\mathcal{M}(I, J)$ in the moment matrix. It is important to emphasize that the subsets I and J in an (I, J) -ribbon are allowed to intersect. Also, $\mathcal{V}(\mathcal{R})$ can contain vertices that are not in $\mathcal{V}(W_{\mathcal{R}})$ if there are isolated vertices in the ribbon.

Ultimately, we will want to partition a ribbon into three subribbons in such a way that we can express the moment matrix as the sum of positive semidefinite matrices and some error terms. Our partitioning will be based on minimum vertex separators.

DEFINITION 6.2 (vertex separator). For an (I, J) -ribbon \mathcal{R} with edge set $W_{\mathcal{R}}$, a subset $Q \subseteq \mathcal{V}(\mathcal{R})$ of vertices is a vertex separator if Q separates I and J in $W_{\mathcal{R}}$. A vertex separator is minimum if there are no other vertex separators with strictly fewer vertices. The separator size of \mathcal{R} is the cardinality of any minimum vertex separator of \mathcal{R} .

The following elementary lemma establishes that a ribbon has a unique *leftmost* and a *rightmost* vertex separator of minimum size. We defer its proof to Appendix A.3.

LEMMA 6.3 (leftmost/rightmost minimum vertex separator). Let \mathcal{R} be an (I, J) -ribbon. There is a unique minimum vertex separator S of \mathcal{R} such that S separates I and J for any vertex separator Q of \mathcal{R} . We call S the leftmost separator in \mathcal{R} . We define the rightmost separator analogously, and we denote them by $S_L(\mathcal{R})$ and $S_R(\mathcal{R})$, respectively.

Remark 6.4. Note that while we omit writing “minimum” explicitly, leftmost and rightmost vertex separators are, in particular, minimum vertex separators. This shouldn’t cause any confusion because in what follows, we will exclusively talk about minimum vertex separators.

We illustrate the notion of a leftmost and a rightmost vertex separator in the example below (see Figure 6.1).

Let $I = \{a, b, c\}$, and let $J = \{c, x, y, z\}$. The maximum number of vertex disjoint paths from I to J is 2—for example, we could take the path $\{c\}$ and the path $\{b, h, i, j, z\}$. The leftmost and rightmost separators are $S_L = \{c, i\}$ and $S_R = \{c, j\}$, respectively. This example illustrates an important point that when I and J intersect, S_L and S_R must both contain $I \cap J$.

6.2. Factorization of monomials. Our factorization of \mathcal{M} will rely on an iterative argument for grouping and factoring the Fourier characters in the decomposition of $\mathcal{M}(I, J)$.

DEFINITION 6.5 (canonical factorization). Let \mathcal{R} be an (I, J) -ribbon with edge set $W_{\mathcal{R}}$ and vertex set $V_{\mathcal{R}}$. Let V_ℓ be the vertices reachable from I without passing through $S_L(\mathcal{R})$, and similarly for V_r , and let $V_m = V_{\mathcal{R}} \setminus (V_\ell \cup V_r)$. Let $W_\ell \subseteq W_{\mathcal{R}}$ be

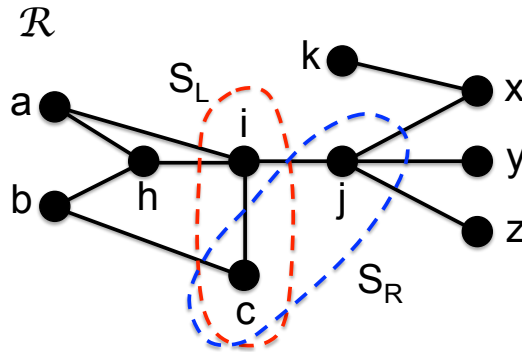


FIG. 6.1.

given by

$$W_\ell = \{(u, v) \in W_{\mathcal{R}} : u \in V_\ell \text{ and } v \in V_\ell \cup S_L\},$$

and similarly for W_r . Finally, let $W_m = W_{\mathcal{R}} \setminus (W_\ell \cup W_r)$.

Let \mathcal{R}_ℓ be the $(I, S_L(\mathcal{R}))$ -ribbon with vertex set $V_\ell \cup S_L(\mathcal{R})$ and edge set W_ℓ , and similarly for \mathcal{R}_r . Let \mathcal{R}_m be the $(S_L(\mathcal{R}), S_R(\mathcal{R}))$ -ribbon with vertex set V_m and edge set W_m . The triple $(\mathcal{R}_\ell, \mathcal{R}_m, \mathcal{R}_r)$ is the canonical factorization of \mathcal{R} .

Some facts about the canonical factorization are worth emphasizing. First, W_ℓ , W_m , and W_r are disjoint and are a partition of $W_{\mathcal{R}}$ by construction. Hence, $\chi_{\mathcal{R}} = \chi_{W_\ell} \cdot \chi_{W_m} \cdot \chi_{W_r}$. Second, some vertices in I may not be in V_ℓ at all. However, any such vertices that are in I but not V_ℓ are necessarily in S_L and thus will be contained in \mathcal{R}_ℓ anyway. This is why we can say that \mathcal{R}_ℓ is an $(I, S_L(\mathcal{R}))$ -ribbon. Figure 6.2 illustrates what the canonical factorization would look like in our earlier example.

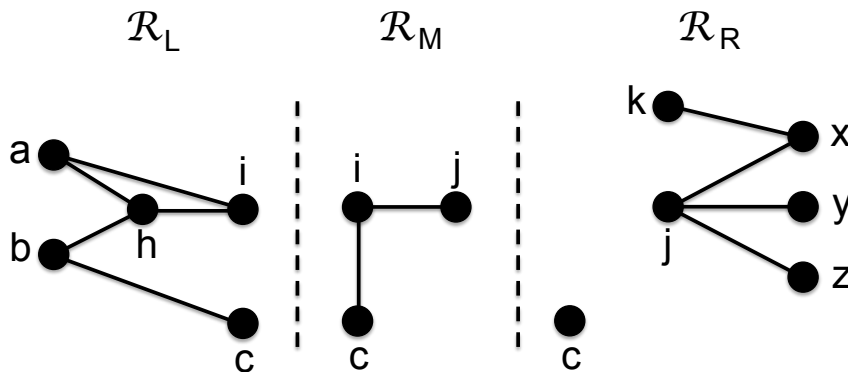


FIG. 6.2.

We chose this example to illustrate a subtle point. The edge (i, c) has both its endpoints in both \mathcal{R}_ℓ and \mathcal{R}_m . We could in principle choose to place it in either, but we have adopted the convention that because both of its endpoints are in S_L we place it in \mathcal{R}_m . In this way, there are no edges within S_L in \mathcal{R}_ℓ or within S_R in \mathcal{R}_m . Finally, note that there can be isolated vertices in \mathcal{R}_ℓ or \mathcal{R}_r but such vertices need to be in I or J .

With the definition of the canonical factorization in hand, we will collect some important properties about it that we will make use of later.

CLAIM 6.6. *Let \mathcal{R} be an (I, J) -ribbon with canonical factorization $(\mathcal{R}_\ell, \mathcal{R}_m, \mathcal{R}_r)$. Then*

$$|\mathcal{V}(\mathcal{R})| = |\mathcal{V}(\mathcal{R}_\ell)| + |\mathcal{V}(\mathcal{R}_m)| + |\mathcal{V}(\mathcal{R}_r)| - |S_L(\mathcal{R})| - |S_R(\mathcal{R})|.$$

Proof. It is important to note that $S_L(\mathcal{R})$ and $S_R(\mathcal{R})$ are not necessarily disjoint (indeed, this happens in the example above). Nevertheless, we know that by construction, V_ℓ , V_m , and V_r are disjoint and that $S_L(\mathcal{R}) \cup S_R(\mathcal{R}) \subseteq V_m$. Every vertex that appears just once in $S_L(\mathcal{R})$ and $S_R(\mathcal{R})$ appears twice in the canonical factorization. And every vertex that is in $S_L(\mathcal{R}) \cap S_R(\mathcal{R})$ appears three times. Thus,

$$|\mathcal{V}(\mathcal{R})| = |\mathcal{V}(\mathcal{R}_\ell)| + |\mathcal{V}(\mathcal{R}_m)| + |\mathcal{V}(\mathcal{R}_r)| - |S_L(\mathcal{R})/S_R(\mathcal{R})| - |S_R(\mathcal{R})/S_L(\mathcal{R})| - 2|S_L(\mathcal{R}) \cap S_R(\mathcal{R})|,$$

which completes the proof. \square

In the discussion above, we established some properties that a canonical factorization must satisfy. Next, we show the reverse direction, that any collection of ribbons that satisfies the below properties must be a canonical factorization. Consider a collection of ribbons $\mathcal{R}_0, \mathcal{R}_1, \mathcal{R}_2$ and the following list of properties.

- S_ℓ, S_r factorization conditions for $\mathcal{R}_0, \mathcal{R}_1, \mathcal{R}_2$. (Here $S_\ell, S_r \subseteq [n]$.)
1. \mathcal{R}_0 is an (I, S_ℓ) -ribbon with $S_L(\mathcal{R}_0) = S_R(\mathcal{R}_0) = S_\ell$, and all vertices in $\mathcal{V}(\mathcal{R}_0)$ are either reachable from I without passing through S_ℓ or are in I or S_ℓ . Finally, \mathcal{R}_0 has no edges between vertices in S_ℓ .
 2. \mathcal{R}_2 is an (S_r, J) -ribbon with $S_L(\mathcal{R}_2) = S_R(\mathcal{R}_2) = S_r$, and all vertices in $\mathcal{V}(\mathcal{R}_2)$ are either reachable from J without passing through S_r or are in J or S_r . Finally, \mathcal{R}_2 has no edges between vertices in S_r .
 3. \mathcal{R}_1 is an (S_ℓ, S_r) -ribbon with $S_L(\mathcal{R}_1) = S_\ell$ and $S_R(\mathcal{R}_1) = S_r$. Every vertex in $\mathcal{V}(\mathcal{R}_1) \setminus (S_\ell \cup S_r)$ has degree at least 1.
 4. $W_{\mathcal{R}_0}, W_{\mathcal{R}_1}, W_{\mathcal{R}_2}$ are pairwise disjoint. Also, $V_{\mathcal{R}_0} \cap V_{\mathcal{R}_1} = S_\ell$, $V_{\mathcal{R}_1} \cap V_{\mathcal{R}_2} = S_r$, and $V_{\mathcal{R}_0} \cap V_{\mathcal{R}_2} = S_\ell \cap S_r$.

LEMMA 6.7. *Let $\mathcal{R}_0, \mathcal{R}_1, \mathcal{R}_2$ be ribbons. Then $(\mathcal{R}_0, \mathcal{R}_1, \mathcal{R}_2)$ is the canonical factorization of the (I, J) -ribbon \mathcal{R} with edge set $W_{\mathcal{R}_0} \oplus W_{\mathcal{R}_1} \oplus W_{\mathcal{R}_2}$ and vertex set $\mathcal{V}(\mathcal{R}_0) \cup \mathcal{V}(\mathcal{R}_1) \cup \mathcal{V}(\mathcal{R}_2)$ if and only if the S_ℓ, S_r factorization conditions hold for $\mathcal{R}_0, \mathcal{R}_1, \mathcal{R}_2$ for some $S_\ell, S_r \subseteq [n]$.*

Proof. If \mathcal{R} is a ribbon with leftmost and rightmost vertex separators S_ℓ and S_r and canonical factorization $(\mathcal{R}_0, \mathcal{R}_1, \mathcal{R}_2)$, then many of the conditions above are automatically satisfied. By construction, $W_{\mathcal{R}_0}, W_{\mathcal{R}_1}, W_{\mathcal{R}_2}$ are pairwise disjoint. Because any edge with both endpoints in S_ℓ is included in \mathcal{R}_m , we have that there are no edges between vertices in S_ℓ in \mathcal{R}_0 , and similarly for \mathcal{R}_2 . Finally, suppose there is a vertex u in \mathcal{R}_0 . If u is not reachable from I without passing through S_ℓ and is not in I or S_ℓ , then it would not be included in \mathcal{R}_0 . An identical argument holds for \mathcal{R}_2 .

All that remains is to verify that $S_L(\mathcal{R}_0) = S_R(\mathcal{R}_0) = S_\ell$, and similarly for $\mathcal{R}_1, \mathcal{R}_2$. If $S_\ell = S_L(\mathcal{R})$ is not a minimum-sized vertex separator for \mathcal{R}_0 , then it is also not a minimum-sized vertex separator for \mathcal{R} , which is impossible. Similarly, if it is not the leftmost separator for \mathcal{R}_0 , then it was not the leftmost separator for \mathcal{R} . Since \mathcal{R}_0 is an (I, S_ℓ) -ribbon and S_ℓ is a minimum-sized separator, it must also be the rightmost minimum-sized separator.

Now, in the reverse direction, suppose that $\mathcal{R}_0, \mathcal{R}_1, \mathcal{R}_2$ are ribbons that meet the S_ℓ, S_r factorization conditions. We claim that S_ℓ is the leftmost separator for \mathcal{R} . If

not, then either there is a smaller vertex separator, or there is a vertex separator S'_ℓ of the same size that separates I and S_ℓ . To rule out the former case, note that since S_ℓ and S_r are both minimum vertex separators for \mathcal{R}_1 , we must have $|S_\ell| = |S_r|$. Then it follows from the S_ℓ, S_r factorization conditions that there are $|S_\ell|$ vertex disjoint paths from I to J , but this would contradict the fact that there is a vertex separator with fewer than $|S_\ell|$ vertices. In the latter case, any other vertex separator S'_ℓ of the same size that separates I and S_ℓ would contradict the condition $S_L(\mathcal{R}_0) = S_\ell$. An identical argument shows that S_r is the rightmost separator for \mathcal{R} .

Finally, by assumption, all the vertices in $\mathcal{V}(\mathcal{R}_0)$ are either reachable from I without passing through S_ℓ or are in I or S_ℓ and hence would be included in \mathcal{R}_0 . Similarly, there are no edges in $W_{\mathcal{R}_0}$ with both endpoints in S_ℓ . Thus, if we were to compute the canonical factorization for \mathcal{R} , we would get the same set of vertices in each ribbon and the same partition of the edges. \square

6.3. Factorization of matrix entries. This leads to our first factorization of the entries $\mathcal{M}(I, J)$ of \mathcal{M} . Unfortunately, the error terms in this first attempt will be too large. Using canonical factorizations and Claim 6.6, for any $I, J \subseteq [n]$ of size at most d we can write

$$\begin{aligned} \mathcal{M}(I, J) &= \sum_{\substack{\mathcal{R} \text{ an } (I, J)\text{-ribbon with edge set } W, \\ |\mathcal{V}(W)| \leq \tau \\ \text{canonical factorization } (\mathcal{R}_\ell, \mathcal{R}_m, \mathcal{R}_r)}} \left(\frac{\omega}{n}\right)^{|\mathcal{V}(\mathcal{R})|} \cdot \chi_{\mathcal{R}_\ell} \cdot \chi_{\mathcal{R}_m} \cdot \chi_{\mathcal{R}_r} \\ &= \sum_{\substack{S_\ell, S_r \subseteq [n] \\ |S_\ell| = |S_r| \leq d}} \left(\frac{\omega}{n}\right)^{-\frac{|S_\ell| + |S_r|}{2}} \sum_{\substack{\mathcal{R}_\ell, \mathcal{R}_m, \mathcal{R}_r \subseteq \binom{[n]}{2} \\ \text{satisfying } S_\ell, S_r \text{ factorization conditions} \\ \text{and } |\mathcal{V}(\mathcal{R}_\ell) \cup \mathcal{V}(\mathcal{R}_m) \cup \mathcal{V}(\mathcal{R}_r)| \leq \tau}} \\ &\quad \times \left(\frac{\omega}{n}\right)^{|\mathcal{V}(\mathcal{R}_\ell)| + |\mathcal{V}(\mathcal{R}_m)| + |\mathcal{V}(\mathcal{R}_r)| - \frac{|S_\ell| + |S_r|}{2}} \cdot \chi_{\mathcal{R}_\ell} \cdot \chi_{\mathcal{R}_m} \cdot \chi_{\mathcal{R}_r}. \end{aligned}$$

Notice that except for the disjointness condition, the S_ℓ, S_r factorization conditions can be separated into condition 1 for \mathcal{R}_ℓ , condition 3 for \mathcal{R}_m , and condition 2 for \mathcal{R}_r . We use this to rewrite as

$$(6.1) \quad = \sum_{\substack{S_\ell, S_r \subseteq [n] \\ |S_\ell| = |S_r| \leq d}} \left(\frac{\omega}{n}\right)^{-\frac{|S_\ell| + |S_r|}{2}} \left(\sum_{\substack{\mathcal{R}_\ell \text{ having 1} \\ |\mathcal{V}(\mathcal{R}_\ell)| \leq \tau}} \left(\frac{\omega}{n}\right)^{|\mathcal{V}(\mathcal{R}_\ell)|} \chi_{\mathcal{R}_\ell} \right) \left(\sum_{\substack{\mathcal{R}_m \text{ having 3} \\ |\mathcal{V}(\mathcal{R}_m)| \leq \tau}} \left(\frac{\omega}{n}\right)^{|\mathcal{V}(\mathcal{R}_m)| - \frac{|S_\ell| + |S_r|}{2}} \chi_{\mathcal{R}_m} \right) \left(\sum_{\substack{\mathcal{R}_r \text{ having 2} \\ |\mathcal{V}(\mathcal{R}_r)| \leq \tau}} \left(\frac{\omega}{n}\right)^{|\mathcal{V}(\mathcal{R}_r)|} \chi_{\mathcal{R}_r} \right)$$

$$(6.2) \quad - \underbrace{\sum_{\substack{S_\ell, S_r \subseteq [n] \\ |S_\ell| = |S_r| \leq d}} \left(\frac{\omega}{n}\right)^{-\frac{|S_\ell| + |S_r|}{2}} \sum_{\substack{\mathcal{R}_\ell, \mathcal{R}_m, \mathcal{R}_r \\ \text{satisfying } S_\ell, S_r \text{ conditions} \\ |\mathcal{V}(\mathcal{R}_\ell)|, |\mathcal{V}(\mathcal{R}_m)|, |\mathcal{V}(\mathcal{R}_r)| \leq \tau, \\ |\mathcal{V}(\mathcal{R}_\ell) \cup \mathcal{V}(\mathcal{R}_m) \cup \mathcal{V}(\mathcal{R}_r)| > \tau}} \left(\frac{\omega}{n}\right)^{|\mathcal{V}(\mathcal{R}_\ell)| + |\mathcal{V}(\mathcal{R}_m)| + |\mathcal{V}(\mathcal{R}_r)| - \frac{|S_\ell| + |S_r|}{2}} \cdot \chi_{\mathcal{R}_\ell} \cdot \chi_{\mathcal{R}_m} \cdot \chi_{\mathcal{R}_r}}_{\stackrel{\text{def}}{=} \xi_0(I, J), \text{ the error from ribbon size}}$$

$$(6.3) \quad - \underbrace{\sum_{\substack{S_\ell, S_r \subseteq [n] \\ |S_\ell| = |S_r| \leq d}} \left(\frac{\omega}{n}\right)^{-\frac{|S_\ell| + |S_r|}{2}} \sum_{\substack{\mathcal{R}_\ell, \mathcal{R}_m, \mathcal{R}_r \text{ satisfying} \\ 1, 3, 2 \text{ and not 4} \\ |\mathcal{V}(\mathcal{R}_\ell)|, |\mathcal{V}(\mathcal{R}_m)|, |\mathcal{V}(\mathcal{R}_r)| \leq \tau}} \left(\frac{\omega}{n}\right)^{|\mathcal{V}(\mathcal{R}_\ell)| + |\mathcal{V}(\mathcal{R}_m)| + |\mathcal{V}(\mathcal{R}_r)| - \frac{|S_\ell| + |S_r|}{2}} \cdot \chi_{\mathcal{R}_\ell} \cdot \chi_{\mathcal{R}_m} \cdot \chi_{\mathcal{R}_r}}_{\stackrel{\text{def}}{=} E_0(I, J), \text{ the error from ribbon nondisjointness}}$$

6.4. Factorization of the matrix \mathcal{M} . In (6.2) and (6.3), we have defined two error matrices, $\xi_0, E_0 \in \mathbb{R}^{\binom{[n]}{\leq d} \times \binom{[n]}{\leq d}}$. Inspired by the factorization of $\mathcal{M}(I, J)$ in (6.1),

we define another pair of matrices as follows:

$$\begin{aligned} \mathcal{Q}_0 \in \mathbb{R}^{\binom{[n]}{d} \times \binom{[n]}{d}} \quad \text{given by} \quad \mathcal{Q}_0(S_\ell, S_r) &= \sum_{\substack{\mathcal{R}_m \text{ having 3} \\ |\mathcal{V}(\mathcal{R}_m)| \leq \tau}} \left(\frac{\omega}{n}\right)^{|\mathcal{V}(\mathcal{R}_m)| - \frac{|S_\ell| + |S_r|}{2}} \chi_{\mathcal{R}_m}, \\ \mathcal{L} \in \mathbb{R}^{\binom{[n]}{d} \times \binom{[n]}{d}} \quad \text{given by} \quad \mathcal{L}(I, S) &= \left(\frac{\omega}{n}\right)^{-\frac{|S|}{2}} \sum_{\substack{\mathcal{R}_\ell \text{ having 1} \\ |\mathcal{V}(\mathcal{R}_\ell)| \leq \tau}} \left(\frac{\omega}{n}\right)^{|\mathcal{V}(\mathcal{R}_\ell)|} \chi_{\mathcal{R}_\ell}. \end{aligned}$$

The powers of (ω/n) are split between \mathcal{Q}_0 and \mathcal{L} so that the typical eigenvalue of \mathcal{Q}_0 will be approximately 1 (although it will be some time before we are prepared to prove that).

Equations (6.1), (6.2), and (6.3) can be written succinctly as

$$\mathcal{M} = \mathcal{L} \mathcal{Q}_0 \mathcal{L}^\dagger - \xi_0 - E_0.$$

As we will see later, with high probability $\mathcal{Q}_0 \succeq 0$, and thus also $\mathcal{L} \mathcal{Q}_0 \mathcal{L}^\dagger \succeq 0$. So long as τ is sufficiently large, the spectral norm $\|\xi_0\|$ of the error term that accounts for ribbons whose size is too large will be negligible. However, the error E_0 does not turn out to be negligible. To overcome this, we will apply a similar factorization approach to E_0 as we did for \mathcal{M} ; iterating this factorization will push down the error from ribbon nondisjointness.

We record an elementary fact about \mathcal{Q}_0 .

LEMMA 6.8. *Let Π be the projector to $\text{Span}\{e_C : C \in \mathcal{C}_{\leq d}\}$. Then $\mathcal{Q}_0 = \Pi \mathcal{Q}_0 = \mathcal{Q}_0 \Pi$.*

Proof. Suppose S is not a clique in G . We need to show that the row $\mathcal{Q}_0(S, \cdot)$ is zero. For every entry $\mathcal{Q}_0(S, S')$, notice that the Fourier coefficients $\widehat{\mathcal{Q}_0(S, S')}(T) = \widehat{\mathcal{Q}_0(S, S')}(T')$ if $T, T' \subseteq \binom{[n]}{2}$ disagree only on edges inside S . (That is, $T \oplus T' \subseteq \binom{S}{2}$.) This means that $\mathcal{Q}_0(S, S') = \mathbf{1}_S$ is a clique in $G \cdot f_{S, S'}(G)$ for some function $f_{S, S'}$. \square

6.5. Iterative factorization of E_0 . We recall now the definition of the matrix $E_0 \in \mathbb{R}^{\binom{[n]}{\leq d} \times \binom{[n]}{\leq d}}$:

$$\begin{aligned} E_0(I, J) &= \sum_{\substack{S_\ell, S_r \subseteq [n] \\ |S_\ell| = |S_r| \leq d}} \left(\frac{\omega}{n}\right)^{-\frac{|S_\ell| + |S_r|}{2}} \sum_{\substack{\mathcal{R}_\ell, \mathcal{R}_m, \mathcal{R}_r \text{ satisfying} \\ 1, 3, 2 \text{ and not } 4 \\ |\mathcal{V}(\mathcal{R}_\ell)|, |\mathcal{V}(\mathcal{R}_m)|, |\mathcal{V}(\mathcal{R}_r)| \leq \tau}} \\ &\quad \times \left(\frac{\omega}{n}\right)^{|\mathcal{V}(\mathcal{R}_\ell)| + |\mathcal{V}(\mathcal{R}_r)| + |\mathcal{V}(\mathcal{R}_m)| - \frac{|S_\ell| + |S_r|}{2}} \cdot \chi_{\mathcal{R}_\ell} \cdot \chi_{\mathcal{R}_m} \cdot \chi_{\mathcal{R}_r}. \end{aligned}$$

In what follows, we will show how to factor a slightly more general sort of matrix; this factorization will be applicable iteratively, starting with E_0 .

6.5.1. The matrix \mathcal{E}_c and its factorization. To express the family of matrices we will factor, we introduce a relaxation of our definition of ribbon and a corresponding relaxation 3* of condition 3 of the S_ℓ, S_r factorization conditions.

DEFINITION 6.9 (improper ribbon). *An improper (I, J) -ribbon \mathcal{R} is an (I, J) -ribbon \mathcal{R}_0 together with a set $\mathcal{Z}(\mathcal{R}) \subseteq [n]$ of vertices disjoint from $\mathcal{V}(\mathcal{R}_0)$. (Think of adding the vertices $\mathcal{Z}(\mathcal{R})$ to the ribbon \mathcal{R}_0 as degree-0 nodes.) We write $\mathcal{V}(\mathcal{R}) = \mathcal{V}(\mathcal{R}_0) \cup \mathcal{Z}(\mathcal{R})$. When we need to distinguish, we sometimes call ordinary ribbons “proper.”*

Every ribbon is also an improper ribbon by taking $\mathcal{Z}(\cdot) = \emptyset$, and every improper ribbon has a corresponding ribbon given by deleting its degree-0 vertices.

Relaxed factorization condition for ribbon \mathcal{R}_1 with $\mathcal{S}_\ell, \mathcal{S}_r \subseteq [n]$.

3*. \mathcal{R}_1 is an improper $(\mathcal{S}_\ell, \mathcal{S}_r)$ -ribbon.

Let c be an \mathbb{R} -valued function $c(\mathcal{R})$ on (possibly improper) ribbons. Let $\mathcal{E}_c \in \mathbb{R}^{\binom{[n]}{\leq d} \times \binom{[n]}{\leq d}}$ be given by

$$(6.4) \quad \mathcal{E}_c(I, J) = \sum_{\substack{\mathcal{S}_\ell, \mathcal{S}_r \subseteq [n] \\ |\mathcal{S}_\ell|, |\mathcal{S}_r| \leq d}} \left(\frac{\omega}{n} \right)^{-\frac{|\mathcal{S}_\ell| + |\mathcal{S}_r|}{2}} \sum_{\substack{\mathcal{R}_\ell, \mathcal{R}_m, \mathcal{R}_r \text{ satisfying} \\ 1, 3^*, 2 \text{ and not } 4 \\ |\mathcal{V}(\mathcal{R}_\ell)|, |\mathcal{V}(\mathcal{R}_m)|, |\mathcal{V}(\mathcal{R}_r)| \leq \tau}} \times c(\mathcal{R}_m) \left(\frac{\omega}{n} \right)^{|\mathcal{V}(\mathcal{R}_\ell)| + |\mathcal{V}(\mathcal{R}_r)| + |\mathcal{V}(\mathcal{R}_m)| - \frac{|\mathcal{S}_\ell| + |\mathcal{S}_r|}{2}} \cdot \chi_{\mathcal{R}_\ell} \cdot \chi_{\mathcal{R}_m} \cdot \chi_{\mathcal{R}_r}.$$

Note that 3 is a strictly more restrictive condition than 3*. Hence, we can define the function c_0 by $c_0(\mathcal{R}_m) = 1$ if \mathcal{R}_m satisfies 3 and $c_0(\mathcal{R}_m) = 0$ otherwise. Then $\mathcal{E}_0 = \mathcal{E}_{c_0}$. In this subsection, we will show how to factor any matrix of the form \mathcal{E}_c as

$$\mathcal{E}_c = \mathcal{L} \mathcal{Q}_{c'} \mathcal{L}^\dagger - \mathcal{E}_{c'} - \xi_c$$

for some function c' on ribbons and matrices $\mathcal{Q}_{c'}, \xi_c \in \mathbb{R}^{\binom{[n]}{\leq d} \times \binom{[n]}{\leq d}}$, where $\|\xi_c\|$ is negligible with high probability.

Just as our initial factorization of \mathcal{M} began with a factorization of each ribbon appearing in the Fourier expansion, our factorization of \mathcal{E}_c depends on a factorization for each triple $(\mathcal{R}_\ell, \mathcal{R}_m, \mathcal{R}_r)$ appearing in (6.4). Since they do not satisfy condition 4, there must be some vertices occurring in more than one of $\mathcal{V}(\mathcal{R}_\ell), \mathcal{V}(\mathcal{R}_m), \mathcal{V}(\mathcal{R}_r)$. Before, the canonical factorization depended on the leftmost and rightmost vertex separators in an (I, J) -ribbon \mathcal{R} separating I from J . But now we will be interested in leftmost and rightmost separators that separate both I and J from each other and from these repeated vertices.

DEFINITION 6.10 (separating factorization). *Let $\mathcal{R}_\ell, \mathcal{R}_m, \mathcal{R}_r$ be ribbons satisfying $\mathcal{S}_\ell, \mathcal{S}_r$ factorization conditions 1, 3*, 2 but not 4, with $|\mathcal{V}(\mathcal{R}_\ell)|, |\mathcal{V}(\mathcal{R}_m)|, |\mathcal{V}(\mathcal{R}_r)| \leq \tau$. Let \mathcal{R} be the (I, J) -ribbon with edge set $W_{\mathcal{R}_\ell} \oplus W_{\mathcal{R}_m} \oplus W_{\mathcal{R}_r}$ and vertex set $\mathcal{V}(\mathcal{R}_\ell) \cup \mathcal{V}(\mathcal{R}_m) \cup \mathcal{V}(\mathcal{R}_r)$. (Thus, $\chi_{\mathcal{R}_\ell} \cdot \chi_{\mathcal{R}_m} \cdot \chi_{\mathcal{R}_r} = \chi_{\mathcal{R}}$.)*

Let S'_ℓ be the leftmost minimum-sized vertex separator in \mathcal{R} which separates I from J and any vertices appearing in more than one of $\mathcal{V}(\mathcal{R}_\ell), \mathcal{V}(\mathcal{R}_m), \mathcal{V}(\mathcal{R}_r)$. Similarly, let S'_r be the rightmost minimum-sized vertex separator in \mathcal{R} separating J from I and these repeated vertices. (Notice that S'_ℓ and S'_r could have different sizes.)

Let V'_ℓ be the vertices reachable from I without passing through S'_ℓ , and similarly for V'_r . Let $V'_m = V_{\mathcal{R}} \setminus (V'_\ell \cup V'_r)$. Let $W'_\ell = \{(u, v) \in W_{\mathcal{R}} : u \in V_\ell, v \in V_\ell \cup S'_\ell\}$, and similarly for W'_r , and let $W'_m = W_{\mathcal{R}} \setminus (W'_\ell \cup W'_r)$.

Let \mathcal{R}'_ℓ be the (I, S'_ℓ) -ribbon with vertex set $V'_\ell \cup S'_\ell$ and edge set W'_ℓ , and let \mathcal{R}'_r be the (S'_r, J) -ribbon with vertex set $V'_r \cup S'_r$ and edge set W'_r . Finally, let \mathcal{R}'_m be the improper (S'_ℓ, S'_r) -ribbon with edge set W'_m and vertex set $(\mathcal{V}(\mathcal{R}) \setminus (V'_\ell \cup V'_r)) \cup (S'_\ell \cup S'_r)$.

Note that $\chi_{\mathcal{R}_\ell} \cdot \chi_{\mathcal{R}_m} \cdot \chi_{\mathcal{R}_r} = \chi_{\mathcal{R}'_\ell} \cdot \chi_{\mathcal{R}'_m} \cdot \chi_{\mathcal{R}'_r}$ if $\mathcal{R}'_\ell, \mathcal{R}'_m, \mathcal{R}'_r$ is the separating factorization for $\mathcal{R}_\ell, \mathcal{R}_m, \mathcal{R}_r$. We can use this to rewrite \mathcal{E}_c as

$$\begin{aligned}
& \mathcal{E}_c(I, J) \\
&= \sum_{\substack{S_\ell, S_r \subseteq [n] \\ |S_\ell|, |S_r| \leq d}} \left(\frac{\omega}{n} \right)^{-\frac{|S_\ell| + |S_r|}{2}} \sum_{\substack{\mathcal{R}_\ell, \mathcal{R}_m, \mathcal{R}_r \text{ satisfying} \\ 1, 3^*, 2 \text{ and not } 4 \\ |\mathcal{V}(\mathcal{R}_\ell)|, |\mathcal{V}(\mathcal{R}_m)|, |\mathcal{V}(\mathcal{R}_r)| \leq \tau \\ \text{separating factorization} \\ \mathcal{R}'_\ell, \mathcal{R}'_m, \mathcal{R}'_r, S'_\ell, S'_r}} \\
(6.5) \quad & \times c(\mathcal{R}_m) \left(\frac{\omega}{n} \right)^{|\mathcal{V}(\mathcal{R}_\ell)| + |\mathcal{V}(\mathcal{R}_r)| + |\mathcal{V}(\mathcal{R}_m)| - \frac{|S_\ell| + |S_r|}{2}} \cdot \chi_{\mathcal{R}'_\ell} \cdot \chi_{\mathcal{R}'_m} \cdot \chi_{\mathcal{R}'_r}.
\end{aligned}$$

Our goal is to find some coefficient function c' on (improper) ribbons and a matrix $\mathcal{Q}_{c'}$ so that this is approximately equal to $\mathcal{L} \mathcal{Q}_{c'} \mathcal{L}^\dagger - \mathcal{E}_{c'}$. For c' yet to be chosen, we take

$$\mathcal{Q}_{c'}(S'_\ell, S'_r) \stackrel{\text{def}}{=} \sum_{\substack{\mathcal{R}'_m \text{ having } 3^* \\ |\mathcal{V}(\mathcal{R}'_m)| \leq \tau}} c'(\mathcal{R}'_m) \left(\frac{\omega}{n} \right)^{|\mathcal{V}(\mathcal{R}'_m)| - \frac{|S'_\ell| + |S'_r|}{2}} \chi_{\mathcal{R}'_m}$$

and have that

$$\begin{aligned}
& \mathcal{L} \mathcal{Q}_{c'} \mathcal{L}^\dagger(I, J) - \mathcal{E}_{c'}(I, J) \\
&= \sum_{\substack{S'_\ell, S'_r \subseteq [n] \\ |S'_\ell|, |S'_r| \leq d}} \left(\frac{\omega}{n} \right)^{-\frac{|S'_\ell| + |S'_r|}{2}} \sum_{\substack{\mathcal{R}'_\ell, \mathcal{R}'_m, \mathcal{R}'_r \text{ satisfying} \\ 1, 3^*, 2, \text{ and } 4 \\ |\mathcal{V}(\mathcal{R}'_\ell)|, |\mathcal{V}(\mathcal{R}'_m)|, |\mathcal{V}(\mathcal{R}'_r)| \leq \tau}} \\
(6.6) \quad & \times c'(\mathcal{R}'_m) \left(\frac{\omega}{n} \right)^{|\mathcal{V}(\mathcal{R}'_\ell)| + |\mathcal{V}(\mathcal{R}'_r)| + |\mathcal{V}(\mathcal{R}'_m)| - \frac{|S'_\ell| + |S'_r|}{2}} \cdot \chi_{\mathcal{R}'_\ell} \cdot \chi_{\mathcal{R}'_m} \cdot \chi_{\mathcal{R}'_r}.
\end{aligned}$$

We will compare (6.5) and (6.6) by collecting like terms, but first we handle the discrepancy in the size bounds on the ribbons with a corresponding error term ξ_c . The following matrix is similar to \mathcal{E}_c but places a size bound on the ribbons in the separating factorization $|\mathcal{V}(\mathcal{R}'_\ell)|, |\mathcal{V}(\mathcal{R}'_m)|, |\mathcal{V}(\mathcal{R}'_r)| \leq \tau$. We define

$$\begin{aligned}
& \mathcal{E}'_c(I, J) \\
&= \sum_{\substack{S_\ell, S_r \subseteq [n] \\ |S_\ell|, |S_r| \leq d}} \left(\frac{\omega}{n} \right)^{-\frac{|S_\ell| + |S_r|}{2}} \sum_{\substack{\mathcal{R}_\ell, \mathcal{R}_m, \mathcal{R}_r \text{ satisfying} \\ 1, 3^*, 2 \text{ and not } 4 \\ \text{separating factorization} \\ \mathcal{R}'_\ell, \mathcal{R}'_m, \mathcal{R}'_r, S'_\ell, S'_r \\ |\mathcal{V}(\mathcal{R}'_\ell)|, |\mathcal{V}(\mathcal{R}'_m)|, |\mathcal{V}(\mathcal{R}'_r)| \leq \tau}} \\
& \times c(\mathcal{R}_m) \left(\frac{\omega}{n} \right)^{|\mathcal{V}(\mathcal{R}_\ell)| + |\mathcal{V}(\mathcal{R}_r)| + |\mathcal{V}(\mathcal{R}_m)| - \frac{|S_\ell| + |S_r|}{2}} \cdot \chi_{\mathcal{R}'_\ell} \cdot \chi_{\mathcal{R}'_m} \cdot \chi_{\mathcal{R}'_r}.
\end{aligned}$$

We take $\xi_c = \mathcal{E}'_c - \mathcal{E}_c$, and we will show below that with high probability the error $\|\xi_c\|$ is negligible. Before doing this, we show that \mathcal{E}'_c is exactly equal to $\mathcal{L}^\dagger \mathcal{Q}_{c'} \mathcal{L} - \mathcal{E}_{c'}$ for the correct choice of c' .

To collect like terms, it helps to define the following quantity $\gamma_{\mathcal{R}'_\ell, \mathcal{R}'_m, \mathcal{R}'_r, I, J, S'_\ell, S'_r}$:

$$\begin{aligned} \gamma_{\mathcal{R}'_\ell, \mathcal{R}'_m, \mathcal{R}'_r, I, J, S'_\ell, S'_r} &\stackrel{\text{def}}{=} \sum_{\substack{\mathcal{R}_\ell, \mathcal{R}_m, \mathcal{R}_r \text{ satisfying} \\ 1, 3^*, 2 \text{ and not } 4 \text{ for some } S_\ell, S_r \\ \text{separating factorization } \mathcal{R}'_\ell, \mathcal{R}'_m, \mathcal{R}'_r, S'_\ell, S'_r}} \\ &\times c(\mathcal{R}_m) \left(\frac{\omega}{n} \right)^{|\mathcal{V}(\mathcal{R}_\ell)| + |\mathcal{V}(\mathcal{R}_m)| + |\mathcal{V}(\mathcal{R}_r)| + \frac{|S'_\ell| + |S'_r|}{2} - |S_\ell| - |S_r|}. \end{aligned}$$

Then we can rewrite $\mathcal{E}'_c(I, J)$ again as

$$\begin{aligned} \mathcal{E}'_c(I, J) &= \sum_{\substack{S'_\ell, S'_r \subseteq [n] \\ |S'_\ell|, |S'_r| \leq d}} \left(\frac{\omega}{n} \right)^{-\frac{|S'_\ell| + |S'_r|}{2}} \sum_{\substack{\mathcal{R}'_\ell, \mathcal{R}'_m, \mathcal{R}'_r \\ \text{satisfying } 1, 3^*, 2, 4 \text{ for } S'_\ell, S'_r \\ |\mathcal{V}(\mathcal{R}'_\ell)|, |\mathcal{V}(\mathcal{R}'_m)|, |\mathcal{V}(\mathcal{R}'_r)| \leq \tau}} \\ &\times \gamma_{\mathcal{R}'_\ell, \mathcal{R}'_m, \mathcal{R}'_r, I, J, S'_\ell, S'_r} \cdot \chi_{\mathcal{R}'_\ell} \cdot \chi_{\mathcal{R}'_m} \cdot \chi_{\mathcal{R}'_r}. \end{aligned}$$

We will obtain $\mathcal{E}'_c = \mathcal{L}^\dagger \mathcal{Q}_{c'} \mathcal{L}^\dagger - \mathcal{E}_{c'}$ if we define $c'(\mathcal{R}'_m)$ so that

$$c'(\mathcal{R}'_m) \left(\frac{\omega}{n} \right)^{|\mathcal{V}(\mathcal{R}'_\ell)| + |\mathcal{V}(\mathcal{R}'_r)| + |\mathcal{V}(\mathcal{R}'_m)| - \frac{|S'_\ell| + |S'_r|}{2}} = \gamma_{\mathcal{R}'_\ell, \mathcal{R}'_m, \mathcal{R}'_r, I, J, S'_\ell, S'_r}.$$

To express this in terms of the function c , we expand out $\gamma_{\mathcal{R}'_\ell, \mathcal{R}'_m, \mathcal{R}'_r, I, J, S'_\ell, S'_r}$. The following definition is useful.

DEFINITION 6.11. *Let*

$$r = (|\mathcal{V}(\mathcal{R}_\ell)| + |\mathcal{V}(\mathcal{R}_m)| + |\mathcal{V}(\mathcal{R}_r)| - |S_\ell| - |S_r|) - (|\mathcal{V}(\mathcal{R}'_\ell)| + |\mathcal{V}(\mathcal{R}'_m)| + |\mathcal{V}(\mathcal{R}'_r)| - |S'_\ell| - |S'_r|).$$

(The ribbons $\mathcal{R}_\ell, \mathcal{R}_m, \mathcal{R}_r, \mathcal{R}'_\ell, \mathcal{R}'_m, \mathcal{R}'_r$ will always be clear from the context.)

Note that $(|\mathcal{V}(\mathcal{R}_\ell)| + |\mathcal{V}(\mathcal{R}_m)| + |\mathcal{V}(\mathcal{R}_r)| - |S_\ell| - |S_r|)$ is the total number of vertices we would have in the (I, J) -ribbon with vertex set $\mathcal{V}(\mathcal{R}_\ell) \cup \mathcal{V}(\mathcal{R}_m) \cup \mathcal{V}(\mathcal{R}_r)$ if $\mathcal{R}_\ell, \mathcal{R}_m, \mathcal{R}_r$ satisfied condition 4 (which they do not!). Similarly, $(|\mathcal{V}(\mathcal{R}'_\ell)| + |\mathcal{V}(\mathcal{R}'_m)| + |\mathcal{V}(\mathcal{R}'_r)| - |S'_\ell| - |S'_r|)$ is the total number of vertices in the (I, J) -ribbon with edge set $\mathcal{W}(\mathcal{R}'_\ell) \cup \mathcal{W}(\mathcal{R}'_m) \cup \mathcal{W}(\mathcal{R}'_r)$ and vertex set $\mathcal{V}(\mathcal{R}'_\ell) \cup \mathcal{V}(\mathcal{R}'_m) \cup \mathcal{V}(\mathcal{R}'_r)$. Thus, r is the number of vertices occurring with multiplicities higher than they should be in $\mathcal{V}(\mathcal{R}_\ell) \cup \mathcal{V}(\mathcal{R}_m) \cup \mathcal{V}(\mathcal{R}_r)$.

We can rewrite the γ 's as

$$\begin{aligned} \gamma_{\mathcal{R}'_\ell, \mathcal{R}'_m, \mathcal{R}'_r, I, J, S'_\ell, S'_r} &= \left(\frac{\omega}{n} \right)^{|\mathcal{V}(\mathcal{R}'_\ell)| + |\mathcal{V}(\mathcal{R}'_m)| + |\mathcal{V}(\mathcal{R}'_r)| - \frac{|S'_\ell| + |S'_r|}{2}} \\ &\times \sum_{\substack{\mathcal{R}_\ell, \mathcal{R}_m, \mathcal{R}_r \text{ satisfying} \\ 1, 3^*, 2 \text{ and not } 4 \text{ for some } S_\ell, S_r \\ r \text{ intersections outside } S_\ell, S_r \\ \text{separating factorization } \mathcal{R}'_\ell, \mathcal{R}'_m, \mathcal{R}'_r, S'_\ell, S'_r}} c(\mathcal{R}_m) \left(\frac{\omega}{n} \right)^r. \end{aligned}$$

Thus, we will have that $\mathcal{E}'_c = \mathcal{L} \mathcal{Q}_{c'} \mathcal{L}^\dagger - \mathcal{E}_{c'}$ if and only if for every (S'_ℓ, S'_r) -ribbon \mathcal{R}'_m and every $\mathcal{R}'_\ell, \mathcal{R}'_r$ satisfying 1, 2,

$$\begin{aligned} c'(\mathcal{R}'_m) &= \sum_{\substack{\mathcal{R}_\ell, \mathcal{R}_m, \mathcal{R}_r \text{ satisfying} \\ 1, 3^*, 2 \text{ and not } 4 \text{ for some } S_\ell, S_r \\ r \text{ intersections outside } S_\ell, S_r \\ \text{separating factorization } \mathcal{R}'_\ell, \mathcal{R}'_m, \mathcal{R}'_r, S'_\ell, S'_r}} c(\mathcal{R}_m) \left(\frac{\omega}{n} \right)^r. \end{aligned}$$

Note that for this to happen, the right-hand side must be independent of \mathcal{R}'_ℓ and \mathcal{R}'_r . If this is the case, then we can define

$$c'(\mathcal{R}'_m) \stackrel{\text{def}}{=} \sum_{\substack{\mathcal{R}_\ell, \mathcal{R}_m, \mathcal{R}_r \text{ satisfying} \\ 1, 3^*, 2 \text{ and not } 4 \text{ for some } S'_\ell, S'_r \\ r \text{ intersections outside } S'_\ell, S'_r \\ \text{separating factorization } \mathcal{R}'_\ell, \mathcal{R}'_m, \mathcal{R}'_r, S'_\ell, S'_r}} c(\mathcal{R}_m) \left(\frac{\omega}{n} \right)^r \text{ for some } \mathcal{R}'_\ell, \mathcal{R}'_r \text{ satisfying } 1, 2.$$

The next claim shows that, indeed, the choice of $\mathcal{R}'_\ell, \mathcal{R}'_r$ does not matter. (This would not have been true without passing from \mathcal{E}_c to \mathcal{E}'_c .)

CLAIM 6.12. *Let $\mathcal{R}'_\ell, \mathcal{R}'_m, \mathcal{R}'_r$ satisfy 1, 3^* , 2, 4 for some $S'_\ell, S'_r \subseteq [n]$. Let \mathcal{R}''_ℓ and \mathcal{R}''_r also satisfy 1 and 2, respectively, for S'_ℓ, S'_r , respectively. Then*

$$\begin{aligned} & \sum_{\substack{\mathcal{R}_\ell, \mathcal{R}_m, \mathcal{R}_r \text{ satisfying} \\ 1, 3^*, 2 \text{ and not } 4 \text{ for some } S'_\ell, S'_r \\ r \text{ intersections outside } S'_\ell, S'_r \\ \text{separating factorization } \mathcal{R}'_\ell, \mathcal{R}'_m, \mathcal{R}'_r, S'_\ell, S'_r}} c(\mathcal{R}_m) \left(\frac{\omega}{n} \right)^r \\ &= \sum_{\substack{\mathcal{R}_\ell, \mathcal{R}_m, \mathcal{R}_r \text{ satisfying} \\ 1, 3^*, 2 \text{ and not } 4 \text{ for some } S'_\ell, S'_r \\ r \text{ intersections outside } S'_\ell, S'_r \\ \text{separating factorization } \mathcal{R}''_\ell, \mathcal{R}''_m, \mathcal{R}''_r, S'_\ell, S'_r}} c(\mathcal{R}_m) \left(\frac{\omega}{n} \right)^r. \end{aligned}$$

(Notice that the left-hand sum refers to $\mathcal{R}'_\ell, \mathcal{R}'_r$ and the right-hand one to $\mathcal{R}''_\ell, \mathcal{R}''_r$.)

Proof. We prove this by showing that there is an exact match between terms on the left-hand side and terms on the right-hand side. Consider a term on the left-hand side. Note that the part of \mathcal{R}_ℓ between I and S'_ℓ must be \mathcal{R}'_ℓ , while the part of \mathcal{R}_ℓ between S'_ℓ and S_ℓ becomes part of \mathcal{R}'_m . To shift from \mathcal{R}'_ℓ to \mathcal{R}''_ℓ , we simply replace \mathcal{R}'_ℓ by \mathcal{R}''_ℓ within \mathcal{R}_ℓ . Similarly, to shift from \mathcal{R}'_r to \mathcal{R}''_r , we simply replace \mathcal{R}'_r by \mathcal{R}''_r within \mathcal{R}_r .

To show that this gives an exact match, we need to show that r is unaffected by these shifts. To see that shifting from \mathcal{R}'_ℓ to \mathcal{R}''_ℓ does not affect r , note that all vertices in $\mathcal{V}(\mathcal{R}'_\ell) \setminus S'_\ell$ or $\mathcal{V}(\mathcal{R}'_\ell) \setminus S'_\ell$ must appear in the corresponding \mathcal{R}_ℓ and cannot appear in \mathcal{R}_m or \mathcal{R}_r . Thus, these vertices always have multiplicity 1 in $\mathcal{V}(\mathcal{R}_\ell) \cup \mathcal{V}(\mathcal{R}_m) \cup \mathcal{V}(\mathcal{R}_r)$. All other vertices (including the ones in S'_ℓ) may appear in \mathcal{R}_m or \mathcal{R}_r , as well as \mathcal{R}_ℓ , but whether or not they do so is unaffected by the shift, and so their multiplicities in $\mathcal{V}(\mathcal{R}_\ell) \cup \mathcal{V}(\mathcal{R}_m) \cup \mathcal{V}(\mathcal{R}_r)$ are unaffected by the shift and r remains the same. A similar argument holds for shifting from \mathcal{R}'_r to \mathcal{R}''_r . \square

Remark 6.13. For this argument, it was important to keep track of the isolated vertices in \mathcal{R}'_m . If we did not keep track of isolated vertices and instead had them disappear, we could have had a situation where there is a vertex v which appears in \mathcal{R}_ℓ and \mathcal{R}_m but disappears from \mathcal{R}'_m and is not in S'_ℓ . Since v is no longer in \mathcal{R}'_m , \mathcal{R}''_ℓ could contain v . If so, then we cannot shift from \mathcal{R}'_ℓ to \mathcal{R}''_ℓ , as this would create a copy of v to the left of S'_ℓ , but v should be to the right of S'_ℓ .

Putting everything together, $\mathcal{E}'_c = \mathcal{L} \mathcal{Q}_{c'} \mathcal{L}^\dagger - \mathcal{E}_{c'}$. Since we defined $\xi_c = \mathcal{E}'_c - \mathcal{E}_c$, we get that $\mathcal{E}_c = \mathcal{L} \mathcal{Q}_c \mathcal{L}^\dagger - \mathcal{E}_{c'} - \xi_c$, as needed.

The remaining step will be to show that with high probability the error term ξ_c has a negligible norm, which we will accomplish in section 7.5.

Finally, we record the following easy lemma about separating factorizations, which will be useful in the application of the foregoing to factor \mathcal{E}_0 .

LEMMA 6.14. Suppose $\mathcal{R}_\ell, \mathcal{R}_m, \mathcal{R}_r$ satisfy conditions 1, 3*, 2 but not 4. Let $\mathcal{R}'_\ell, \mathcal{R}'_m, \mathcal{R}'_r$ be their separating factorization, with separators S'_ℓ, S'_r . Then

$$\frac{|S'_\ell| + |S'_r|}{2} - \frac{|S_\ell| + |S_r|}{2} \geq \frac{1}{2}.$$

Proof. We claim that $|S_\ell| + |S_r| + 1 \leq |S'_\ell| + |S'_r|$. By the violation of condition 4, we cannot have $S_\ell = S'_\ell$ and $S_r = S'_r$. But since S'_ℓ separates I from S_ℓ in \mathcal{R}_ℓ and \mathcal{R}_ℓ is an (I, S_ℓ) -ribbon whose rightmost vertex separator is also S_ℓ , if $S_\ell \neq S'_\ell$, then $|S_\ell| < |S'_\ell|$, and similarly for S_r and S'_r . So either $|S_\ell| < |S'_\ell|$ or $|S_r| < |S'_r|$, and since the separator sizes are integers, the difference must be at least 1 and we are done. \square

6.5.2. Application to E_0 and \mathcal{M} . We are ready to define our recursive factorization of E_0 . Recall that $c_0(\mathcal{R}_m) = 1$ if \mathcal{R}_m satisfies 3 and $c_0(\mathcal{R}_m) = 0$ otherwise and that $E_0 = \mathcal{E}_{c_0}$. Applying the factorization above to \mathcal{E}_{c_0} , we obtain matrices $\xi_1 = \xi_{c_0}$, \mathcal{Q}_1 , and \mathcal{E}_{c_1} . Then, of course, we can apply the factorization again to \mathcal{E}_{c_1} .

Proceeding inductively, for all $i \in [1, 2d]$, let $\xi_i = \xi_{c_{i-1}}$, \mathcal{Q}_i , and \mathcal{E}_{c_i} be the matrices given by applying the factorization to $\mathcal{E}_{c_{i-1}}$ at step i .

CLAIM 6.15.

$$\mathcal{M} = \mathcal{L}(\mathcal{Q}_0 - \mathcal{Q}_1 + \mathcal{Q}_2 - \cdots - \mathcal{Q}_{2d-1} + \mathcal{Q}_{2d}) \mathcal{L}^\dagger - (\xi_0 - \xi_1 + \xi_2 - \cdots - \xi_{2d-1} + \xi_{2d}).$$

Proof. We have that $\mathcal{M} = \mathcal{L}(\mathcal{Q}_0) \mathcal{L}^\dagger - \mathcal{E}_0 - \xi_0$ and $\mathcal{E}_{i-1} = \mathcal{L} \mathcal{Q}_i \mathcal{L}^\dagger - \mathcal{E}_i - \xi_{c_{i-1}} = \mathcal{L} \mathcal{Q}_i \mathcal{L}^\dagger - \mathcal{E}_i - \xi_i$. We prove the claim by starting with the first formula and applying the second formula for each $i \in [1, 2d]$. At the end, we are left with an extra term \mathcal{E}_{2d} . We must show that $\mathcal{E}_{2d} = 0$.

To see why $\mathcal{E}_{2d} = 0$, note that every time we have a separating factorization $\mathcal{R}'_\ell, \mathcal{R}'_m, \mathcal{R}'_r$ for $\mathcal{R}_\ell, \mathcal{R}_m, \mathcal{R}_r$ the size of either the left separator or the right separator must increase (see Lemma 6.14). However, the size of these separators is always at most d , and so the only way we can do this for $2d$ steps is if we started with the empty set as the separators and increased the size of either the left or right separator by 1 each time, but not both. However, this too is impossible, as if we start with the empty set as the separators, after the first step both the new left separator and the new right separator must have size at least 1. \square

7. \mathcal{M} is positive semidefinite. In this section, we combine the factorization of \mathcal{M} in terms of the matrices $\mathcal{L}, \mathcal{Q}_i, \xi_i$ that we obtained in section 6 with estimates on the eigenvalues of the \mathcal{Q} 's and ξ 's. The starting point is the following positive semidefiniteness claim for \mathcal{Q}_0 .

LEMMA 7.1. Let $D \in \mathbb{R}^{\binom{[n]}{\leq d} \times \binom{[n]}{\leq d}}$ be the diagonal matrix with $D(S, S) = 2^{\binom{|S|}{2}}/4$ if S is a clique in G and is 0 otherwise. With high probability, $\mathcal{Q}_0 \succeq D$.

We also need to bound $\|\mathcal{Q}_i\|$ for $i > 0$.

LEMMA 7.2. Let $D \in \mathbb{R}^{\binom{[n]}{\leq d} \times \binom{[n]}{\leq d}}$ be the diagonal matrix with $D(S, S) = 2^{\binom{|S|}{2}}/4$ if S is a clique and is otherwise zero. With high probability, every \mathcal{Q}_i for $i \in [1, 2d]$ satisfies

$$\frac{-D}{8d} \preceq \mathcal{Q}_i \preceq \frac{D}{8d}.$$

The preceding lemmas are enough to obtain $\mathcal{Q}_0 - \cdots + \mathcal{Q}_{2d} \succeq D/2$, but in the end we need to work with the matrix $\mathcal{L}(\mathcal{Q}_0 - \cdots + \mathcal{Q}_{2d}) \mathcal{L}^\dagger - (\xi_0 - \cdots + \xi_{2d})$. The next two lemmas allow us to make this last step.

LEMMA 7.3. *With high probability, $\Pi \mathcal{L} \Pi \mathcal{L}^\dagger \Pi \succeq \Omega((\omega/n)^{d+1}) \cdot \Pi$, where as usual, Π is the projector to $\text{Span}\{e_C : C \in \mathcal{C}_{\leq d}\}$.*

Finally, we need a bound on the ξ matrices.

LEMMA 7.4. *With high probability, $\|\xi_0 - \dots + \xi_{2d}\| \leq n^{-16d}$.*

We can now prove Lemma 5.8.

Proof of Lemma 5.8. By Claim 6.15,

$$\mathcal{M} = \mathcal{L}(\mathcal{Q}_0 - \mathcal{Q}_1 + \mathcal{Q}_2 - \dots - \mathcal{Q}_{2d-1} + \mathcal{Q}_{2d}) \mathcal{L}^\dagger - (\xi_0 - \xi_1 + \xi_2 - \dots - \xi_{2d-1} + \xi_{2d}).$$

By a union bound, with high probability the conclusions of Lemmas 7.1, 7.2, 7.3, and 7.4 all hold. By Lemmas 7.1 and 7.2,

$$\mathcal{Q}_0 - \mathcal{Q}_1 + \mathcal{Q}_2 - \dots - \mathcal{Q}_{2d-1} + \mathcal{Q}_{2d} \succeq \frac{D}{2} \succeq \frac{\Pi}{2},$$

where as usual, Π is the projector to $\text{Span}_{e_C} : C \in \mathcal{C}_{\leq d}$. Thus, by Lemma 7.3, we obtain $\mathcal{L}(\mathcal{Q}_0 - \dots + \mathcal{Q}_{2d}) \mathcal{L}^\dagger \succeq \Omega((\omega/n)^{d+1}) \cdot \Pi$. Finally, by Lemma 7.4 we have

$$\mathcal{M} = \Pi \cdot \mathcal{M} \cdot \Pi \succeq \Omega\left(\left(\frac{\omega}{n}\right)^{d+1}\right) \cdot \Pi - n^{-16d} \cdot \Pi \succeq 0. \quad \square$$

In the next subsections, we prove the foregoing lemmas.

7.1. Ribbons and spectral norms. We will require bounds on the spectral norm of certain random matrices. Our random matrices arise out of decompositions of the moment matrix from Definition 5.7 and are functions of a graph G on vertex set $[n]$. Our norm bounds will hold for what we call *graphical matrices*, which are defined to capture the matrices that are invariant under a permutation of the vertices of G and are in fact “minimal” such matrices.

We first define the *shape* of a ribbon that identifies the structure of a ribbon up to relabeling.

DEFINITION 7.5 (shape of a ribbon). *For an (I, J) -ribbon \mathcal{R} , consider the graph U on the vertex set $[\mathcal{V}(\mathcal{R})]$ whose edges are*

$$E(U) = \{(i, j) : \text{there is an edge in } \mathcal{R} \text{ from the } i\text{th to the } j\text{th least element of } \mathcal{V}(\mathcal{R})\}.$$

(Here we are considering $\mathcal{V}(\mathcal{R})$ to have the usual ordering inherited from $[n]$.) Also, let U have two distinguished subsets of vertices A and B , where

$$A = \{i : \text{the } i\text{th element of } \mathcal{V}(\mathcal{R}) \text{ is in } I\},$$

and similarly for B and J . We call U the *shape* of \mathcal{R} and write $\text{shape}(\mathcal{R}) = U$.

We record some observations on shapes of ribbons:

- If \mathcal{R} is a ribbon (not an improper ribbon), its shape satisfies the condition that every vertex outside $A \cup B$ has degree at least 1.
- If, for example, \mathcal{R} is an (I, J) ribbon where $I \cap J = \{1\}$ (which must be the least element in both I and J), then in order for the (I', J') -ribbon \mathcal{R}' to have the same shape as \mathcal{R} it is necessary that $|I' \cap J'| = 1$. More broadly, specifying the shape of a ribbon in particular specifies the pattern of intersection of its endpoints.

- A matrix $M \in \mathbb{R}^{\binom{n}{\leq d} \times \binom{n}{\leq d}}$ whose entries are given by

$$M(I, J) = \sum_{\mathcal{R} \text{ an } (I, J)\text{-ribbon with shape } U} \chi_{\mathcal{R}}$$

satisfies the assumptions of Lemma 7.8. In the following sections, our main strategy will be to decompose the matrices \mathcal{Q}_i into matrices of this form.

We are now ready to define graphical matrices.

DEFINITION 7.6 (graphical matrices). *Let U be a graph on the vertex set $[t]$ with two distinguished sets of vertices $A, B \subseteq [t]$. Let $\mathcal{T}(U)$ be the collection of all I, J ribbons with shape U . The graphical matrix $M \in \mathbb{R}^{\binom{[n]}{A} \times \binom{[n]}{B}}$ of shape U is defined by*

$$M(I, J) = \sum_{\mathcal{R}: \mathcal{R} \text{ is an } (I, J)\text{-ribbon and } \text{shape}(\mathcal{R})=U} \chi_{\mathcal{R}}.$$

Example 7.7. When U is a graph on two vertices with distinguished sets $\{1\}$ and $\{2\}$ of size 1 each and a single edge connecting vertices 1 and 2, the graphical matrix of shape U is just the standard $\{-1, 1\}$ -adjacency matrix of the graph G .

The following lemma will be our main tool. It is in essence due to Medarametla and Potechin [MP16], and special cases of the bound have been proven and used in [HKP+16, HKP15, DM15]. We give a proof in the appendix for completeness.

LEMMA 7.8. *Let U be a graph on $t \leq O(\log n)$ vertices, with two distinguished subsets of vertices A and B , and suppose the following:*

- U admits p vertex-disjoint paths from $A \setminus B$ to $B \setminus A$.
- $|A \cap B| = r$.
- Every vertex outside $A \cup B$ has degree at least 1.

Let $M = M(G)$ be the graphical matrix with shape U . Then, with high probability, $\|M\| \leq n^{\frac{t-p-r}{2}} \cdot 2^{O(t)} \cdot (\log n)^{O(t-r+p)}$.

Remark 7.9. Lemma 7.8 can be seen as a generalization of the standard upper bound on the spectral norm of the adjacency matrix. Example 7.7 shows how the adjacency matrix is a graphical matrix with a shape U on two vertices with a single edge connecting them, and thus $t = 2$, $r = 0$, and $p = 1$. Lemma 7.8 thus shows an upper bound of $\sqrt{n} \text{poly log}(n)$ on the spectral norm of the adjacency matrix which is tight up to a $\text{poly log}(n)$ factor.

7.2. Positivity for \mathcal{Q}_0 —Proof of Lemma 7.1. In this section, we prove Lemma 7.1, which we restate here.

LEMMA 7.10 (restatement of Lemma 7.1). *Let $D \in \mathbb{R}^{\binom{[n]}{\leq d} \times \binom{[n]}{\leq d}}$ be the diagonal matrix with $D(S, S) = 2^{\binom{|S|}{2}}/4$ if S is a clique in G and is 0 otherwise. With high probability, $\mathcal{Q}_0 \succeq D$.*

Proof of Lemma 7.1. To begin, we split \mathcal{Q}_0 into its diagonal $\mathcal{Q}_0^{\text{diag}}$ and its off-diagonal $\mathcal{Q}_0^{\text{off-diag}}$ parts.

$$\mathcal{Q}_0^{\text{diag}}(S_\ell, S_r) = \begin{cases} \mathcal{Q}_0(S_\ell, S_r) & \text{if } S_\ell = S_r, \\ 0 & \text{otherwise,} \end{cases} \quad \mathcal{Q}_0^{\text{off-diag}}(S_\ell, S_r) = \begin{cases} \mathcal{Q}_0(S_\ell, S_r) & \text{if } S_\ell \neq S_r, \\ 0 & \text{otherwise.} \end{cases}$$

Then $\mathcal{Q}_0 = \mathcal{Q}_0^{\text{diag}} + \mathcal{Q}_0^{\text{off-diag}}$. Expanding $\mathcal{Q}_0^{\text{diag}}$,

$$\mathcal{Q}_0^{\text{diag}}(S, S) = 2^{\binom{|S|}{2}} \cdot \mathbf{1}_{S \text{ is a clique}} \cdot \left(1 + \sum_{\substack{\mathcal{R} \text{ nonempty, having 3} \\ \text{and no edges inside } S \\ |S| < |\mathcal{R}| \leq \tau}} \left(\frac{\omega}{n} \right)^{|\mathcal{V}(\mathcal{R})| - |S|} \cdot \chi_{\mathcal{R}} \right) = 2^{\binom{|S|}{2}} \cdot \mathbf{1}_{S \text{ is a clique}} \cdot (1 \pm n^{-\Omega(\varepsilon)})$$

for all $S \in \binom{[n]}{d}$ with high probability by an argument similar to that in Lemma 5.4 and a union bound.

Next, we bound $\|\mathcal{Q}_0^{\text{off-diag}}\|$ by decomposing it according to ribbon shape. Fix $s, t \leq \tau$. Let $U_1^{(s,t)}, \dots, U_q^{(s,t)}$ be all the graphs on vertex set $[t]$ with two distinguished sets of vertices A, B , both of size s , with $|A \cap B| \leq s-1$, and where there are $s - |A \cap B|$ vertex-disjoint paths from $A \setminus B$ to $B \setminus A$. Let $M_i^{(s,t)}$ be given by

$$M_i^{(s,t)}(S_\ell, S_r) = \sum_{\mathcal{R} \text{ an } (S_\ell, S_r)\text{-ribbon with shape } U_i^{(s,t)}} \chi_{\mathcal{R}}.$$

Then

$$\mathcal{Q}_0^{\text{off-diag}} = \sum_{\substack{s \leq d \\ t \leq \tau \\ i \leq q}} \left(\frac{\omega}{n} \right)^{t-s} \cdot M_i^{(s,t)}.$$

We can apply Lemma 7.8 to conclude that with probability at least $1 - O(n^{-100 \log n})$,

$$\begin{aligned} \left\| \left(\frac{\omega}{n} \right)^{t-s} \cdot M_i^{(s,t)} \right\| &\leq \left(\frac{\omega}{n} \right)^{t-s} \cdot n^{\frac{t-s}{2}} \cdot 2^{O(t)} \cdot (\log n)^{O(t - |A \cap B| + |A \setminus B|)} \\ &\leq n^{-\varepsilon(t-s)} \cdot 2^{O(t)} \cdot (\log n)^{O(t-s)}, \end{aligned}$$

where to conclude the bound on the exponent in $(\log n)^{O(t - |A \cap B| + |A \setminus B|)}$ we have used that $t \geq 2s - |A \cap B|$.

Notice that for fixed s and t , there are at most $2^{\binom{t}{2} + O(t)}$ unique shapes $U_1^{(s,t)}, \dots, U_q^{(s,t)}$. Thus, with a union bound followed by the triangle inequality, we obtain that for fixed s and t , with probability at least $1 - O(n^{-99 \log n})$,

$$\left\| \left(\frac{\omega}{n} \right)^{t-s} \sum_{i \leq q} M_i^{(s,t)} \right\| \leq 2^{\binom{t}{2} + O(t)} \cdot n^{-\varepsilon(t-s)} \cdot 2^{O(t)} \cdot (\log n)^{O(t-s)}.$$

Under our assumptions on the parameters d, τ , and ε , this is at most $2^{\binom{s}{2}} / (100\tau)$. Summing over all $t \leq \tau$, for a fixed s , we have

$$\left\| \left(\frac{\omega}{n} \right)^{t-s} \sum_{\substack{t \leq \tau \\ i \leq q}} M_i^{(s,t)} \right\| \leq \frac{2^{\binom{s}{2}}}{100}.$$

Notice that the above matrix is exactly the block of $\mathcal{Q}_0^{\text{off-diag}}$ corresponding to subsets of size s . Together with our bound on $\mathcal{Q}_0^{\text{diag}}$, this proves the lemma. \square

7.3. Norm bounds for \mathcal{Q}_i —Proof of Lemma 7.2. In this section, we prove Lemma 7.2, restated here.

LEMMA 7.11 (restatement of Lemma 7.2). *Let $D \in \mathbb{R}^{\binom{[n]}{\leq d} \times \binom{[n]}{\leq d}}$ be the diagonal matrix with $D(S, S) = 2^{\binom{S}{2}}/4$ if S is a clique and is otherwise zero. With high probability, every \mathcal{Q}_i for $i \in [1, 2d]$ satisfies*

$$\frac{-D}{8d} \preceq \mathcal{Q}_i \preceq \frac{D}{8d}.$$

We will need to bound the coefficients $c_i(\mathcal{R}'_m)$ used to define the matrices \mathcal{Q}_i which we set up in section 6.

LEMMA 7.12. *Let c_1, \dots, c_{2d} be the coefficient functions defined in section 6. For all improper (S_ℓ, S_r) -ribbons \mathcal{R}_m admitting exactly p vertex-disjoint paths from S_ℓ to S_r , and all $i \leq 2d$, writing $s = \frac{|S_\ell| + |S_r|}{2}$,*

$$c_i(\mathcal{R}_m) \leq \left(\frac{\omega}{n}\right)^s \cdot n^{\frac{p - |\mathcal{Z}(\mathcal{R}_m)| - i/2}{2} + \varepsilon s},$$

recalling that $\omega = n^{1/2 - \varepsilon}$. Furthermore, if \mathcal{R}_m and \mathcal{R}'_m have the same shape, then $c_i(\mathcal{R}_m) = c_i(\mathcal{R}'_m)$.

With this lemma in hand, we can prove Lemma 7.2.

Proof of Lemma 7.2. Fix some $0 < i \leq 2d$. We will use Lemma 7.8, which requires that we first decompose each \mathcal{Q}_i into simpler matrices. First, for a proper ribbon \mathcal{R}_m , let

$$\tilde{c}_i(\mathcal{R}_m) = \sum_{\mathcal{R}'_m \text{ an improper ribbon whose largest proper subribbon is } \mathcal{R}_m} \left(\frac{\omega}{n}\right)^{|\mathcal{Z}(\mathcal{R}'_m)|} \cdot c_i(\mathcal{R}'_m).$$

Note that we include \mathcal{R}_m itself in this sum, as a proper ribbon is also an improper ribbon.

CLAIM 7.13. $\tilde{c}_i(\mathcal{R}_m) \leq 2(\omega/n)^s \cdot n^{\frac{p-i/2}{2} + \varepsilon s}$, where p is the number of vertex-disjoint paths from S_ℓ to S_r in \mathcal{R}_m .

Proof. Consider all of the improper ribbons \mathcal{R}'_m with k isolated vertices whose largest proper subribbon is \mathcal{R}_m . For each such ribbon \mathcal{R}'_m , by Lemma 7.12,

$$(\omega/n)^k c_i(\mathcal{R}'_m) \leq \left(\frac{\omega}{n}\right)^{k+s} \cdot n^{\frac{p-k-i/2}{2} + \varepsilon s}.$$

There are at most n^k such improper ribbons. Adding all of their contributions together gives at most

$$\left(\frac{\omega}{\sqrt{n}}\right)^k \left(\frac{\omega}{n}\right)^s \cdot n^{\frac{p-i/2}{2} + \varepsilon s} < 2^{-k} (\omega/n)^s \cdot n^{\frac{p-i/2}{2} + \varepsilon s}.$$

Summing this up over all $k \geq 0$ gives the result. \square

Now fix $s_\ell, s_r \leq d$ and $t \leq \tau$, and let $U_1^{(s_\ell, s_r, t)}, \dots, U_q^{(s_\ell, s_r, t)}$ be all graphs on the vertex set $[t]$ with two distinguished subsets of vertices: A of size s_ℓ and B of size s_r . Let

$$\begin{aligned}
M_j^{(s_\ell, s_r, t)}(S_\ell, S_r) &= \sum_{\mathcal{R} \text{ is an } (S_\ell, S_r)\text{-ribbon with shape } U_j^{(s_\ell, s_r, t)}} \tilde{c}_i(\mathcal{R}) \cdot \left(\frac{\omega}{n}\right)^{t-s} \cdot \chi_{\mathcal{R}} \\
&= \tilde{c}_i(U_j^{(s_\ell, s_r, t)}) \sum_{\mathcal{R} \text{ is an } (S_\ell, S_r)\text{-ribbon with shape } U_j^{(s_\ell, s_r, t)}} \left(\frac{\omega}{n}\right)^{t-s} \cdot \chi_{\mathcal{R}},
\end{aligned}$$

where $s = \frac{s_\ell + s_r}{2}$ and we have used the fact that $\tilde{c}_i(\mathcal{R})$ depends only on the shape of \mathcal{R} .

Let $r = |A \cap B|$, where A, B are the distinguished sets of vertices for $U_j^{(s_\ell, s_r, t)}$, and let \tilde{p} be the number of vertex-disjoint paths from $A \setminus B$ to $B \setminus A$, so that $p = r + \tilde{p}$. We can apply Lemma 7.8 and our bound on \tilde{c}_i to get that with probability $1 - O(n^{-100 \log n})$,

$$\begin{aligned}
\|M_j^{(s_\ell, s_r, t)}\| &\leq \left(\frac{\omega}{n}\right)^{t-s} \cdot n^{\frac{\tilde{p}+r-i/2}{2}+\varepsilon s} \cdot n^{\frac{t-\tilde{p}-r}{2}} \cdot 2^{O(t)} \cdot (\log n)^{O(t-r+\tilde{p})} \\
&= n^{-\varepsilon(t-s)-i/4} \cdot 2^{O(t)} \cdot (\log n)^{O(t-r+\tilde{p})} \\
&= n^{-\varepsilon(t-s)-i/4} \cdot 2^{O(t)} \cdot (\log n)^{O(t-s)},
\end{aligned}$$

where in the last step we have used that $t \geq 2s - r$ and $\tilde{p} \leq s - r$.

By inspection,

$$\mathcal{Q}_i = \sum_{\substack{s_\ell, s_r \leq d \\ t \leq \tau \\ j \leq q}} M_j^{(s_\ell, s_r, t)}.$$

For a fixed t , there are at most $2^{\binom{t}{2}+O(t)}$ choices for U , and so $q \leq 2^{\binom{t}{2}+O(t)}$. Now we fix s_ℓ, s_r and sum over t to obtain the block of \mathcal{Q}_i corresponding to size- s_ℓ and size- s_r subsets. By the triangle inequality and a union bound, with probability at least $1 - O(n^{-97 \log n})$,

$$\left\| \sum_{\substack{t \leq \tau \\ j \leq q}} M_j^{(s_\ell, s_r, t)} \right\| \leq 2^{\binom{t}{2}+O(t)} \cdot n^{-\varepsilon(t-s)-i/4} \cdot 2^{O(t)} \cdot (\log n)^{O(t-s)}.$$

From our assumptions on d, τ , and ε , this is at most $2^{(\binom{s_\ell}{2})/2 + (\binom{s_r}{2})/2} / 100d^3$.

As usual, let Π be the projector to $\text{Span}\{e_C : C \in \mathcal{C}_{\leq d}\}$. Note that $\Pi \mathcal{Q}_i = \mathcal{Q}_i \Pi = \mathcal{Q}_i$ since $\mathcal{Q}_i(I, J) = 0$ whenever I or J is not a clique. So, to show that $D/8d \succeq \mathcal{Q}_i \succeq -D/8d$, it is sufficient to show that for all vectors v with $v = \Pi v$ it happens that $|v^\dagger \mathcal{Q}_i v| \leq v^T (D/8d) v$. To see this, let v_k be the part of v indexed by cliques of size exactly k . Now

$$\begin{aligned}
|v^\dagger \mathcal{Q}_i v| &\leq \sum_{k_1=0}^d \sum_{k_2=0}^d \|v_{k_1}\| \left\| \sum_{\substack{t \leq \tau \\ j \leq q}} M_j^{(k_1, k_2, t)} \right\| \|v_{k_2}\| \\
&\leq \sum_{k_1=0}^d \sum_{k_2=0}^d \frac{1}{100d^3} \left(2^{(\binom{k_1}{2})/2 + (\binom{k_2}{2})/2} \|v_{k_1}\| \|v_{k_2}\| \right) \\
&\leq \sum_{k_1=0}^d \sum_{k_2=0}^d \frac{1}{200d^3} \left(2^{(\binom{k_1}{2})} \|v_{k_1}\|^2 + 2^{(\binom{k_2}{2})} \|v_{k_2}\|^2 \right)
\end{aligned}$$

$$\leq \sum_{k=0}^d \frac{2^{\binom{k}{2}}}{100d^2} \|v_k\|^2 \leq v^\dagger (D/8d) v. \quad \square$$

7.3.1. Coefficient decay in the factorization: Proof of Lemma 7.12. We turn to the proof of Lemma 7.12, for which we want the following characterization of the effect of the separating factorization on the underlying graph of a ribbon.

We require the following combinatorial quantities.

Definitions for Lemma 7.14.

1. $I, J, S_\ell, S_r \subseteq [n]$ of size at most d .
2. Ribbons $\mathcal{R}_\ell, \mathcal{R}_m, \mathcal{R}_r$ satisfying 1,3*,2 but not 4 for $S_\ell, S_r, I, J \subseteq [n]$. (Remember that \mathcal{R}_m may be improper.)
3. Ribbons $\mathcal{R}'_\ell, \mathcal{R}'_m, \mathcal{R}'_r$ which are the separating factorizations of $\mathcal{R}_\ell, \mathcal{R}_m, \mathcal{R}_r$, with separators S'_ℓ, S'_r . (Remember that \mathcal{R}'_m may be improper.)
4. p , the number of vertex-disjoint paths from S_ℓ to S_r in \mathcal{R}_m .
5. p' , the number of vertex-disjoint paths from S'_ℓ to S'_r in \mathcal{R}'_m .
6. $r = (|\mathcal{V}(\mathcal{R}_\ell)| + |\mathcal{V}(\mathcal{R}_m)| + |\mathcal{V}(\mathcal{R}_r)| - |S_\ell| - |S_r|) - (|\mathcal{V}(\mathcal{R}'_\ell)| + |\mathcal{V}(\mathcal{R}'_m)| + |\mathcal{V}(\mathcal{R}'_r)| - |S'_\ell| - |S'_r|)$, the number of intersections among $\mathcal{R}_\ell, \mathcal{R}_m, \mathcal{R}_r$.
7. $\mathfrak{D} = \mathcal{Z}(\mathcal{R}'_m) \setminus \mathcal{Z}(\mathcal{R}_m)$, the newly degree-0 (we write *isolated*) vertices in \mathcal{R}'_m .
8. $\mathfrak{U} \subseteq \mathcal{V}(\mathcal{R}_\ell) \cup \mathcal{V}(\mathcal{R}_m) \cup \mathcal{V}(\mathcal{R}_r)$, the set of vertices appearing in more than one of $\mathcal{V}(\mathcal{R}_\ell), \mathcal{V}(\mathcal{R}_m)$, and $\mathcal{V}(\mathcal{R}_r)$. (Note that $\mathfrak{U} \subseteq \mathcal{V}(\mathcal{R}'_m)$.)

LEMMA 7.14.

$$\underbrace{|S'_\ell| + |S'_r| - (|S_\ell| + |S_r|)}_{\text{increase in separator size}} + \underbrace{p - p'}_{\text{lost paths between separators}} + \underbrace{|\mathfrak{D}|}_{\text{new isolated vertices}} \leq \underbrace{r}_{\text{number of intersections}}.$$

The following series of claims will help us in the proof of Lemma 7.14.

CLAIM 7.15. $I \cap \mathcal{V}(\mathcal{R}'_m) \subseteq S'_\ell$ and $J \cap \mathcal{V}(\mathcal{R}'_m) \subseteq S'_r$.

Proof of claim. If $u \in I \cap \mathcal{V}(\mathcal{R}'_m)$, then since $I \subseteq \mathcal{V}(\mathcal{R}'_\ell)$, we have $u \in \mathcal{V}(\mathcal{R}'_\ell) \cap \mathcal{V}(\mathcal{R}'_m) = S'_\ell$, and similarly for the second part. \square

Next, we have a simple analysis of which vertices may possibly be newly isolated.

CLAIM 7.16. $\mathfrak{D} \subseteq \mathfrak{U}$.

Proof of claim. Let $u \in \mathfrak{D}$. If $u \in S_\ell$ or $u \in S_r$, then we are done. Otherwise, if $u \in I$ or $u \in J$, then u appeared in more than one of $\mathcal{V}(\mathcal{R}_\ell), \mathcal{V}(\mathcal{R}_m), \mathcal{V}(\mathcal{R}_r)$ by the definition of the canonical separating factorization.

If neither of these cases holds, then u was incident to an edge in at least one of $\mathcal{R}_\ell, \mathcal{R}_m, \mathcal{R}_r$. Since that edge does not exist in \mathcal{R}'_m , it must have appeared at least twice among the edge sets of $\mathcal{R}_\ell, \mathcal{R}_m, \mathcal{R}_r$, and therefore u appeared at least twice among the vertex sets, thus proving the claim. \square

Next, we show that some vertices in \mathfrak{U} cannot become isolated.

CLAIM 7.17. *By Menger's theorem, there are $|S'_\ell|$ vertex-disjoint paths from $\mathfrak{U} \cap \mathcal{V}(\mathcal{R}_\ell)$ to I in \mathcal{R}_ℓ . Let $u_\ell^{(1)}, \dots, u_\ell^{(|S'_\ell|)}$ be distinct vertices so that $u^{(i)}$ is the last vertex*

in \mathfrak{U} along the i th vertex disjoint path. Let $u_r^{(1)}, \dots, u_r^{(|S'_r|)}$ be similarly defined. None of the vertices u may be in \mathfrak{D} .

Proof of claim. Fix one of these vertices u , and consider its neighbor v one step farther along the path to I (or J). By definition, the vertex v does not appear in more than one of $\mathcal{V}(\mathcal{R}_\ell), \mathcal{V}(\mathcal{R}_m), \mathcal{V}(\mathcal{R}_r)$. If $v \in \mathcal{R}'_m$, then the edge (u, v) must be in \mathcal{R}'_m , and so u is not isolated in \mathcal{R}'_m . If $v \notin \mathcal{R}'_m$, then u must be in $S'_\ell \cup S'_r$, in which case by definition, $u \notin \mathfrak{D}$. \square

We set up sets q of vertices to divide up the intersecting vertices among $\mathcal{R}_\ell, \mathcal{R}_m, \mathcal{R}_r$ according to which ribbons witness the intersection.

CLAIM 7.18. *Let*

$$\begin{aligned} q_{\ell,m,r} &\stackrel{\text{def}}{=} (\mathcal{V}(\mathcal{R}_r) \cap \mathcal{V}(\mathcal{R}_m) \cap \mathcal{V}(\mathcal{R}_\ell)) \setminus (S_\ell \cup S_r), \\ q_{\ell,r} &\stackrel{\text{def}}{=} (\mathcal{V}(\mathcal{R}_\ell) \cap \mathcal{V}(\mathcal{R}_r)) \setminus \mathcal{V}(\mathcal{R}_m), \\ q_{\ell,m} &\stackrel{\text{def}}{=} (\mathcal{V}(\mathcal{R}_\ell) \cap \mathcal{V}(\mathcal{R}_m)) \setminus (S_\ell \cup \mathcal{V}(\mathcal{R}_r)), \\ q_{r,m} &\stackrel{\text{def}}{=} (\mathcal{V}(\mathcal{R}_r) \cap \mathcal{V}(\mathcal{R}_m)) \setminus (S_r \cup \mathcal{V}(\mathcal{R}_\ell)). \end{aligned}$$

The sets q are pairwise disjoint, and

$$r = 2|q_{\ell,m,r}| + |q_{\ell,r}| + |q_{\ell,m}| + |q_{r,m}| + |S_\ell \cap (\mathcal{V}(\mathcal{R}_r) \setminus S_r)| + |S_r \cap (\mathcal{V}(\mathcal{R}_\ell) \setminus S_\ell)|.$$

Also, $\mathfrak{U} = q_{\ell,m,r} \cup q_{\ell,r} \cup q_{\ell,m} \cup q_{r,m} \cup S_\ell \cup S_r$.

Proof. The proof is by inspection. \square

We are prepared to prove Lemma 7.14.

Proof of Lemma 7.14. We start by bounding the number of vertices in $\mathfrak{U} \setminus \mathfrak{D}$. By Claim 7.17, there are at least $|\{u_\ell^{(1)}, \dots, u_\ell^{(|S'_\ell|)}, u_r^{(1)}, \dots, u_r^{(|S'_r|)}\}|$ such vertices.

Let a be the number of pairs i, j so that $u_\ell^{(i)} = u_r^{(j)}$. Then there are vertex-disjoint paths w_1, \dots, w_a from S'_ℓ to S'_r . The path w corresponding to $u_\ell^{(i)} = u_r^{(j)}$ is given by following $u_\ell^{(i)}$'s path from I to \mathfrak{U} , ending at $u_\ell^{(i)}$, and then following $u_r^{(j)}$'s path from \mathfrak{U} to J . This gives a path from I to J , which must have a subpath from S'_ℓ to S'_r .

Now consider the p vertex-disjoint paths from S_ℓ to S_r in \mathcal{R}_m . We claim that

$$\begin{aligned} p - |S_\ell \cap S_r| &\leq |q_{\ell,m,r}| + |S_\ell \cap \mathcal{V}(\mathcal{R}_r) \setminus S_r| + |S_r \cap \mathcal{V}(\mathcal{R}_\ell) \setminus S_\ell| \\ (7.1) \quad &+ |\mathfrak{U} \setminus (\{u_\ell^{(1)}, \dots, u_\ell^{(|S'_\ell|)}, u_r^{(1)}, \dots, u_r^{(|S'_r|)}\} \cup \mathfrak{D})| + (p' - a). \end{aligned}$$

In words, every nontrivial path from S_ℓ to S_r contributes to at least one of

- $|q_{\ell,m,r}|$, the number of 3-way intersections;
- intersections between S_ℓ and $\mathcal{V}(\mathcal{R}_r)$ (but not S_r), and intersections between $\mathcal{V}(\mathcal{R}_\ell)$ and S_r (but not S_ℓ);
- vertices in \mathfrak{U} which are guaranteed not to become isolated (and which we have not yet accounted for); or
- vertex-disjoint paths from S'_ℓ to S'_r (which we have not yet accounted for).

Fix one such path. If it intersects $q_{\ell,m,r}$, $S_\ell \cap \mathcal{V}(\mathcal{R}_r)$, or $S_r \cap \mathcal{V}(\mathcal{R}_\ell)$, we are done. Suppose otherwise. If it is contained entirely in $q_{\ell,m} \cup q_{r,m} \cup (S_\ell \setminus \mathcal{V}(\mathcal{R}_r)) \cup (S_r \setminus \mathcal{V}(\mathcal{R}_\ell))$, then there is some edge along the path connecting a vertex in $\mathcal{V}(\mathcal{R}_\ell) \cap \mathcal{V}(\mathcal{R}_m) \setminus \mathcal{V}(\mathcal{R}_r)$ with one in $\mathcal{V}(\mathcal{R}_r) \cap \mathcal{V}(\mathcal{R}_m) \setminus \mathcal{V}(\mathcal{R}_\ell)$. That edge can occur nowhere else among $\mathcal{R}_\ell, \mathcal{R}_m, \mathcal{R}_r$, and so the incident vertices must not be in \mathfrak{D} . At the same time, if there

is any vertex along the path which is outside \mathfrak{U} , then the nearest vertices along the path to either side which do lie in \mathfrak{U} also must be outside \mathfrak{D} .

In either case, there are two vertices along the path in $\mathfrak{U} \setminus \mathfrak{D}$. If either of these is not among the u vertices, we are done. If both are, then by definition of the u vertices this creates a path from I to J and so from S'_ℓ to S'_r . Furthermore, this path must be vertex disjoint from the paths w_1, \dots, w_a previously constructed since the u vertices involved in those paths were $\mathcal{V}(\mathcal{R}_\ell) \cap \mathcal{V}(\mathcal{R}_r)$. This proves (7.1).

It is time to put things together. By Claim 7.16, we can bound $|\mathfrak{D}|$ by

$$|\mathfrak{D}| \leq |\mathfrak{U}| - |\mathfrak{U} \setminus \mathfrak{D}|.$$

We have $|\mathfrak{U} \setminus \mathfrak{D}| \geq |S'_\ell| + |S'_r| - a + |\mathfrak{U} \setminus (\{u_\ell^{(1)}, \dots, u_\ell^{(|S'_\ell|)}, u_r^{(1)}, \dots, u_r^{(|S'_r|)}\} \cup \mathfrak{D})|$, and $|\mathfrak{U}| = |q_{\ell,m,r}| + |q_{\ell,r}| + |q_{\ell,m}| + |q_{r,m}| + |S_\ell \cup S_r|$. This gives

$$\begin{aligned} |\mathfrak{D}| &\leq |q_{\ell,m,r}| + |q_{\ell,r}| + |q_{\ell,m}| + |q_{r,m}| + |S_\ell \cup S_r| - |S'_\ell| \\ &\quad - |S'_r| + a - |\mathfrak{U} \setminus (\{u_\ell^{(1)}, \dots, u_\ell^{(|S'_\ell|)}, u_r^{(1)}, \dots, u_r^{(|S'_r|)}\} \cup \mathfrak{D})|. \end{aligned}$$

Adding (7.1) to both sides and rearranging, we get

$$\begin{aligned} p - p' + |\mathfrak{D}| &\leq 2|q_{\ell,m,r}| + |S_\ell \cap (\mathcal{V}(\mathcal{R}_r) \setminus S_r)| + |S_r \cap (\mathcal{V}(\mathcal{R}_\ell) \setminus S_\ell)| \\ &\quad + |q_{\ell,r}| + |q_{\ell,m}| + |q_{r,m}| + |S_\ell \cup S_r| - |S'_\ell| - |S'_r| + |S_\ell \cap S_r|, \end{aligned}$$

and substituting $r = 2|q_{\ell,m,r}| + |S_\ell \cap (\mathcal{V}(\mathcal{R}_r) \setminus S_r)| + |S_r \cap (\mathcal{V}(\mathcal{R}_\ell) \setminus S_\ell)| + |q_{\ell,r}| + |q_{\ell,m}| + |q_{r,m}|$ gives

$$p - p' + |\mathfrak{D}| \leq r + |S_\ell \cup S_r| - |S'_\ell| - |S'_r| + |S_\ell \cap S_r|.$$

Notice that $|S_\ell \cup S_r| + |S_\ell \cap S_r| = |S_\ell| + |S_r|$, and so we can rearrange to obtain the lemma. \square

Now we can prove Lemma 7.12.

Proof of Lemma 7.12. First, we note that $c_i(\mathcal{R}_m)$ depends only on the shape of \mathcal{R}_m by the symmetry of our construction. We turn to the quantitative bound.

The proof is by induction. The coefficients $c_0(\mathcal{R}_m)$ are nonzero only for ribbons \mathcal{R}_m having $\mathcal{Z}(\mathcal{R}_m) = \emptyset$ and admitting $|S_\ell| = |S_r| = p$ paths from S_ℓ to S_r . Thus, in the case that $i = 0$, the statement reduces to $c_0(\mathcal{R}_m) \leq 1$, which is true by definition.

Suppose the lemma holds for c_i , and consider c_{i+1} . By definition, for an (improper) S'_ℓ, S'_r -ribbon \mathcal{R}'_m and ribbons $\mathcal{R}'_\ell, \mathcal{R}'_r$ satisfying conditions 1 and 2,

$$(7.2) \quad c_{i+1}(\mathcal{R}'_m) = \sum_{\substack{\mathcal{R}_\ell, \mathcal{R}_m, \mathcal{R}_r \text{ satisfying} \\ 1, 3^*, 2 \text{ and not } 4 \text{ for some } S_\ell, S_r \\ r \text{ intersections outside } S_\ell, S_r \\ \text{separating factorization } \mathcal{R}'_\ell, \mathcal{R}'_m, \mathcal{R}'_r, S'_\ell, S'_r}} c_i(\mathcal{R}_m) \left(\frac{\omega}{n}\right)^r.$$

We introduce the shorthand $s' = \frac{|S'_\ell| + |S'_r|}{2}$. Consider first a particular term in the sum, $c_i(\mathcal{R}_m)(\omega/n)^r$, where \mathcal{R}_m is an improper S_ℓ, S_r ribbon, and let $|\mathfrak{D}| = |\mathcal{Z}(\mathcal{R}'_m) \setminus \mathcal{Z}(\mathcal{R}_m)|$. By induction and Lemma 7.14,

$$\begin{aligned}
\left(\frac{\omega}{n}\right)^r \cdot c_i(\mathcal{R}_m) &\leq \left(\frac{\omega}{n}\right)^r \cdot \left(\frac{\omega}{n}\right)^s \cdot n^{\frac{p-|\mathcal{Z}(\mathcal{R}_m)|-i/2}{2}+\varepsilon s} \quad (\text{by induction}) \\
&= \left(\frac{\omega}{n}\right)^{s'} \cdot \left(\frac{\omega}{n}\right)^{r-s'+s} \cdot n^{\frac{p-|\mathcal{Z}(\mathcal{R}_m)|-i/2}{2}+\varepsilon s} \\
&= \left(\frac{\omega}{n}\right)^{s'} \cdot n^{-\varepsilon(r-s'+s)} \cdot n^{-\frac{1}{2}(r-s'+s)} \cdot n^{\frac{p-|\mathcal{Z}(\mathcal{R}_m)|-i/2}{2}+\varepsilon s} \\
&\quad (\text{using } \omega = n^{1/2-\varepsilon}) \\
&\leq \left(\frac{\omega}{n}\right)^{s'} \cdot n^{-\varepsilon(r-s'+s)} \cdot n^{-\frac{1}{2}(s'-s+p-p'+|\mathfrak{D}|)} \cdot n^{\frac{p-|\mathcal{Z}(\mathcal{R}_m)|-i/2}{2}+\varepsilon s} \\
&\quad (\text{by Lemma 7.14}) \\
&= \left(\frac{\omega}{n}\right)^{s'} \cdot n^{-\varepsilon(r-s'+s)} \cdot n^{\frac{p'-|\mathcal{Z}(\mathcal{R}'_m)|-i/2-s'+s}{2}+\varepsilon s} \\
&\quad (\text{canceling terms, using } |\mathcal{Z}(\mathcal{R}'_m)| = |\mathfrak{D}| + |\mathcal{Z}(\mathcal{R}_m)|) \\
&= n^{-\varepsilon r} \cdot \left(\frac{\omega}{n}\right)^{s'} \cdot n^{\frac{p'-|\mathcal{Z}(\mathcal{R}'_m)|-i/2-(s'-s)}{2}+\varepsilon s'} \\
&\leq n^{-\varepsilon r} \cdot \left(\frac{\omega}{n}\right)^{s'} \cdot n^{\frac{p'-|\mathcal{Z}(\mathcal{R}'_m)|-(i+1)/2}{2}+\varepsilon s'} \\
&\quad (\text{using } s'-s \geq 1/2, \text{ by Lemma 6.14}).
\end{aligned}$$

Next, we assess how many nonzero terms are in the sum (7.2) for a fixed r and a fixed \mathcal{R}'_m . For each vertex of \mathcal{R}'_m , there are seven possibilities for which ribbon(s) it came from in $\{\mathcal{R}_\ell, \mathcal{R}_m, \mathcal{R}_r\}$, and so there are at most 7^τ choices overall (recall that \mathcal{R}'_m has at most τ vertices for the terms at which we are looking). Once we have chosen from which ribbon(s) each vertex of \mathcal{R}'_m came, everything is fixed, except for possible edges of \mathcal{R}'_m which appear at least twice in \mathcal{R}_ℓ , \mathcal{R}_m , and \mathcal{R}_r . There are two possibilities for each possible edge of \mathcal{R}'_m which appears twice in \mathcal{R}_ℓ , \mathcal{R}_m , and \mathcal{R}_r and four possibilities for each possible edge of \mathcal{R}'_m which appears three times in \mathcal{R}_ℓ , \mathcal{R}_m , and \mathcal{R}_r . However, note that any such edge must be between an intersected vertex and either another intersected vertex or a vertex in $S_\ell \cup S_r$. Thus, there are at most $r\tau$ possible edges of \mathcal{R}'_m which appear at least twice in \mathcal{R}_ℓ , \mathcal{R}_m , and \mathcal{R}_r and the total number of possibilities for these edges is at most $4^{r\tau}$.

All together, there are at most $2^{O(r\tau)}$ nonzero terms for fixed r . This means that the total contribution from such terms is at most

$$2^{O(r\tau)} \cdot n^{-\varepsilon r} \cdot \left(\frac{\omega}{n}\right)^{s'} \cdot n^{\frac{p'-|\mathcal{Z}(\mathcal{R}'_m)|-(i+1)/2}{2}+\varepsilon s'}.$$

As long as $\tau \leq (\varepsilon/C) \log n$ for some universal constant C , we have $2^{O(r\tau)} \cdot n^{-\varepsilon r} \ll 1/\tau$ for all $r \geq 1$. All in all, we obtain

$$c_{i+1}(\mathcal{R}'_m) \leq \left(\frac{\omega}{n}\right)^{s'} \cdot n^{\frac{p'-|\mathcal{Z}(\mathcal{R}'_m)|-(i+1)/2}{2}+\varepsilon s'},$$

which completes the induction. \square

7.4. $\mathcal{L}\mathcal{L}^\dagger$ is well-conditioned—Proof of Lemma 7.3. In this section, we prove Lemma 7.3, restated here.

LEMMA 7.19 (restatement of Lemma 7.3). *With high probability, $\Pi \mathcal{L} \Pi \mathcal{L}^\dagger \Pi \succeq \Omega((\omega/n)^{d+1}) \cdot \Pi$, where as usual, Π is the projector to $\text{Span}\{e_C : C \in \mathcal{C}_{\leq d}\}$.*

Proof of Lemma 7.3. We recall the definition of \mathcal{L} :

$$\mathcal{L}(I, S) = \left(\frac{\omega}{n}\right)^{-\frac{|S|}{2}} \sum_{\substack{\mathcal{R} \text{ having } 1 \\ |\mathcal{V}(\mathcal{R}_\ell)| \leq \tau}} \left(\frac{\omega}{n}\right)^{|\mathcal{V}(\mathcal{R}_\ell)|} \chi_{\mathcal{R}_\ell}.$$

Consider a diagonal entry $\mathcal{L}(S, S)$. Since every ribbon \mathcal{R} appearing in its expansion must have 1, in particular it has no edges inside S . Thus, by the same argument as in Lemma 5.4, with probability at least $1 - O(n^{-10 \log n})$,

$$\mathcal{L}(S, S) = \left(\frac{\omega}{n}\right)^{\frac{|S|}{2}} (1 \pm n^{-\Omega(\varepsilon)}).$$

Let $\mathcal{L}^{\text{off-diag}}$ be given by

$$\mathcal{L}^{\text{off-diag}}(I, S) = \begin{cases} \mathcal{L}(I, S) & \text{if } I \neq S, \\ 0 & \text{otherwise.} \end{cases}$$

We will consider the block of $\mathcal{L}^{\text{off-diag}}$ with rows indexed by sets of size s_ℓ and columns indexed by sets of size s_r for some $s_\ell, s_r \leq d$. For a fixed $t \leq \tau$, let $U_1^{(s_\ell, s_r, t)}, \dots, U_q^{(s_\ell, s_r, t)}$ be all the graphs on vertex set $[t]$ with distinguished subsets of vertices A, B of sizes s_ℓ, s_r , respectively, and where

- $A \neq B$,
- there are no edges inside B ,
- every vertex in U outside $A \cup B$ is reachable from A without passing through B , and
- B is the unique minimum-sized vertex separator in U separating A from B .

Then let $M_i^{(s_\ell, s_r, t)}$ be given by

$$M_i^{(s_\ell, s_r, t)}(I, S) = \left(\frac{\omega}{n}\right)^{t - \frac{s_r}{2}} \cdot \sum_{\mathcal{R} \text{ an } (I, S)\text{-ribbon with shape } U_i^{(s_\ell, s_r, t)}} \chi_{\mathcal{R}}.$$

By assumption on $U_i^{(s_\ell, s_r, t)}$, there are s_r vertex-disjoint paths from A to B . Let $r = |A \cap B|$. By Lemma 7.8, with probability at least $1 - O(n^{-100 \log n})$,

$$\begin{aligned} \|M_i^{(s_\ell, s_r, t)}\| &\leq \left(\frac{\omega}{n}\right)^{\frac{s_r}{2}} \cdot \left(\frac{\omega}{n}\right)^{t-s_r} \cdot n^{\frac{t-s_r}{2}} \cdot 2^{O(t)} \cdot (\log n)^{O(t-r+(s_r-r))} \\ &= \left(\frac{\omega}{n}\right)^{\frac{s_r}{2}} \cdot n^{-\varepsilon(t-s_r)} \cdot 2^{O(t)} \cdot (\log n)^{O(t-s_r)}, \end{aligned}$$

where in the last step we have used that $t \geq s_\ell + s_r - r$ and $s_r \leq s_\ell$, which holds by the vertex-separator requirement on B . There are at most $2^{\binom{t}{2} - \binom{s_r}{2} + O(t)}$ choices for $U_i^{(s_\ell, s_r, t)}$ when s_ℓ, s_r, t are fixed by the requirement that U have no edges inside B . Summing over all q for a fixed t , we get by the triangle inequality

$$\left\| \sum_{i \leq q} M_i^{(s_\ell, s_r, t)} \right\| \leq \left(\frac{\omega}{n}\right)^{\frac{s_r}{2}} \cdot 2^{\binom{t}{2} - \binom{s_r}{2} + O(t)} \cdot n^{-\varepsilon(t-s_r)} \cdot (\log n)^{O(t-s_r)}$$

with probability $1 - O(n^{-99 \log n})$. By our assumptions on d, τ , and ε , this is at most $(\omega/n)^{s_r/2} \cdot 1/d^4$.

The following standard manipulations now prove the lemma. Let $D' \in \mathbb{R}^{\binom{[n]}{\leq d}}$ be the diagonal matrix with $D'(S, S) = (\omega/n)^{|S|/2}$ if S is a clique in G and is 0 otherwise. Then we can decompose $\mathcal{L} = D + E + \mathcal{L}^{\text{off-diag}}$, where E is a diagonal matrix with $|E(S, S)| \leq n^{-\Omega(\varepsilon)} \cdot (\omega/n)^{|S|/2}$. Then we have

$$\begin{aligned} \Pi \mathcal{L} \Pi \mathcal{L}^\dagger \Pi &= D^2 \\ &\quad + \Pi(D \Pi \mathcal{L}^{\text{off-diag}} + D \Pi E \\ &\quad + E \Pi D + E \Pi \mathcal{L}^{\text{off-diag}} + \mathcal{L}^{\text{off-diag}} \Pi D + \mathcal{L}^{\text{off-diag}} \Pi E \\ &\quad + E \Pi E + \mathcal{L}^{\text{off-diag}} \Pi \mathcal{L}^{\text{off-diag}}) \Pi. \end{aligned}$$

Each of the above matrices aside from D^2 is a $d \times d$ block matrix, where the (s_ℓ, s_r) block is $\binom{[n]}{s_\ell} \times \binom{[n]}{s_r}$ dimensional and has a norm at most $(\omega/n)^{(s_\ell+s_r)/2} \cdot d^{-4}$. By the same argument as in the proof of Lemma 7.2, using Cauchy–Schwarz to combine the d^2 blocks, we obtain the lemma. \square

7.5. High-degree matrices have small norms. In this section, we prove Lemma 7.4, restated here.

LEMMA 7.20 (restatement of Lemma 7.4). *With high probability, $\|\xi_0 - \dots - \xi_{2d}\| \leq n^{-16d}$.*

We recall the definition of ξ_i . For a coefficient function on ribbons $c_{i-1}(\mathcal{R}_m)$, we have a matrix \mathcal{E} given by

$$\begin{aligned} \mathcal{E}(I, J) &= \sum_{\substack{S_\ell, S_r \subseteq [n] \\ |S_\ell|, |S_r| \leq d}} \left(\frac{\omega}{n} \right)^{-\frac{|S_\ell|+|S_r|}{2}} \sum_{\substack{\mathcal{R}_\ell, \mathcal{R}_m, \mathcal{R}_r \text{ satisfying} \\ 1, 3^*, 2 \text{ and not } 4 \\ |\mathcal{V}(\mathcal{R}_\ell)|, |\mathcal{V}(\mathcal{R}_m)|, |\mathcal{V}(\mathcal{R}_r)| \leq \tau \\ \text{separating factorization} \\ \mathcal{R}'_\ell, \mathcal{R}'_m, \mathcal{R}'_r, S'_\ell, S'_r}} \\ &\quad \times c_{i-1}(\mathcal{R}_m) \left(\frac{\omega}{n} \right)^{|\mathcal{V}(\mathcal{R}_\ell)|+|\mathcal{V}(\mathcal{R}_r)|+|\mathcal{V}(\mathcal{R}_m)|-\frac{|S_\ell|+|S_r|}{2}} \cdot \chi_{\mathcal{R}'_\ell} \cdot \chi_{\mathcal{R}'_m} \cdot \chi_{\mathcal{R}'_r} \end{aligned}$$

and another one, \mathcal{E}' , given by

$$\begin{aligned} \mathcal{E}'(I, J) &= \sum_{\substack{S_\ell, S_r \subseteq [n] \\ |S_\ell|, |S_r| \leq d}} \left(\frac{\omega}{n} \right)^{-\frac{|S_\ell|+|S_r|}{2}} \sum_{\substack{\mathcal{R}_\ell, \mathcal{R}_m, \mathcal{R}_r \text{ satisfying} \\ 1, 3^*, 2 \text{ and not } 4 \\ \text{separating factorization} \\ \mathcal{R}'_\ell, \mathcal{R}'_m, \mathcal{R}'_r, S'_\ell, S'_r \\ |\mathcal{V}(\mathcal{R}'_\ell)|, |\mathcal{V}(\mathcal{R}'_m)|, |\mathcal{V}(\mathcal{R}'_r)| \leq \tau}} \\ &\quad \times c_{i-1}(\mathcal{R}_m) \left(\frac{\omega}{n} \right)^{|\mathcal{V}(\mathcal{R}_\ell)|+|\mathcal{V}(\mathcal{R}_r)|+|\mathcal{V}(\mathcal{R}_m)|-\frac{|S_\ell|+|S_r|}{2}} \cdot \chi_{\mathcal{R}'_\ell} \cdot \chi_{\mathcal{R}'_m} \cdot \chi_{\mathcal{R}'_r}. \end{aligned}$$

Then the matrix ξ_i is given by $\mathcal{E} - \mathcal{E}'$.

We will actually prove a bound on the Frobenius norm of each matrix ξ_i . The following will allow us to control the magnitude of the entries. It follows immediately from our concentration bound lemma (Lemma A.2), which is proved via the moment method. (Under the slightly stronger assumption $\tau \ll \varepsilon \log n / \log \log n$, it would also follow from standard hypercontractivity.)

LEMMA 7.21. Suppose c_T are a collection of coefficients, one for each $T \subseteq \binom{[n]}{2}$, and there is a constant C such that the following hold:

1. If $|T| > C\tau$, then $c_T = 0$.
2. Otherwise, $|c_T| \leq (\omega/n)^{|T|/C - Cd}$.

Then, with probability at least $1 - O(n^{-100 \log n})$, it occurs that $|\sum_{T \subseteq \binom{[n]}{2}} c_T \cdot \chi_T| \leq n^{-20d}$.

We will also need several facts about the coefficients of ribbons in the expansion of each matrix ξ_i .

LEMMA 7.22. Every triple $\mathcal{R}_\ell, \mathcal{R}_m, \mathcal{R}_r$ appearing with a nonzero coefficient in ξ_c satisfies $|\mathcal{V}(\mathcal{R}_\ell)| + |\mathcal{V}(\mathcal{R}_m)| + |\mathcal{V}(\mathcal{R}_r)| = \Theta(\tau)$.

Proof. To appear with a nonzero coefficient, the triple $\mathcal{R}_\ell, \mathcal{R}_m, \mathcal{R}_r$ with separating factorization $\mathcal{R}'_\ell, \mathcal{R}'_m, \mathcal{R}'_r$ must either have

$$|\mathcal{V}(\mathcal{R}_\ell)|, |\mathcal{V}(\mathcal{R}_m)|, |\mathcal{V}(\mathcal{R}_r)| \leq \tau \quad \text{but} \quad |\mathcal{V}(\mathcal{R}'_\ell)| > \tau \text{ or } |\mathcal{V}(\mathcal{R}'_m)| > \tau \text{ or } |\mathcal{V}(\mathcal{R}'_r)| > \tau$$

or

$$|\mathcal{V}(\mathcal{R}'_\ell)|, |\mathcal{V}(\mathcal{R}'_m)|, |\mathcal{V}(\mathcal{R}'_r)| \leq \tau \quad \text{but} \quad |\mathcal{V}(\mathcal{R}_\ell)| > \tau \text{ or } |\mathcal{V}(\mathcal{R}_m)| > \tau \text{ or } |\mathcal{V}(\mathcal{R}_r)| > \tau.$$

In the first case, we must have one of $|\mathcal{V}(\mathcal{R}_\ell)| \geq \tau/3$ or $|\mathcal{V}(\mathcal{R}_m)| \geq \tau/3$ or $|\mathcal{V}(\mathcal{R}_r)| \geq \tau/3$. In the second, we must have $|\mathcal{V}(\mathcal{R}_\ell)|, |\mathcal{V}(\mathcal{R}_m)|, |\mathcal{V}(\mathcal{R}_r)| \leq 3\tau$. \square

We are prepared to prove Lemma 7.4.

Proof of Lemma 7.4. We will apply Lemma 7.21 to $\xi_i(I, J)$ for each $i \leq 2d$ and $I, J \subseteq [n]$ with $|I|, |J| \leq d$. So consider the Fourier expansion of $\xi_i(I, J)$, given by

$$\xi_i(I, J) = \sum_{T \subseteq \binom{[n]}{2}} c_T \cdot \chi_T.$$

From Lemma 7.22, we obtain that if $|T| > C\tau$, then $c_T = 0$ for some absolute constant C . For smaller T , we need a bound on the magnitude $|c_T|$. The coefficient c_T is bounded by

$$(7.3) \quad |c_T| \leq \sum_{\substack{S_\ell, S_r \subseteq [n] \\ |S_\ell|, |S_r| \leq d}} \left(\frac{\omega}{n}\right)^{-\frac{|S_\ell| + |S_r|}{2}} \sum_{\substack{\mathcal{R}_\ell, \mathcal{R}_m, \mathcal{R}_r \\ \text{nonzero in } \xi_i(I, J) \text{ as in Lemma 7.22} \\ \chi_{\mathcal{R}_\ell} \cdot \chi_{\mathcal{R}_m} \cdot \chi_{\mathcal{R}_r} = \chi_T}} \times c_{i-1}(\mathcal{R}_m) \left(\frac{\omega}{n}\right)^{|\mathcal{V}(\mathcal{R}_\ell)| + |\mathcal{V}(\mathcal{R}_r)| + |\mathcal{V}(\mathcal{R}_m)| - \frac{|S_\ell| + |S_r|}{2}}.$$

By Lemma 7.12, we have $c_{i-1}(\mathcal{R}_m) \leq n^d \leq (\omega/n)^{-2d}$. At the same time, there are at most $2^{O(\tau^2)}$ nonzero terms in the sum (7.3). Thus, by Lemma 7.22 and our assumptions on d, τ , and ε , the coefficient c_T is at most $(\omega/n)^{\tau/C - Cd}$ for some absolute constant C .

Applying Lemma 7.21, we get $|\xi_i(I, J)| \leq n^{-20d}$ with probability $1 - O(n^{-100 \log n})$. Taking a union bound over all $n^{2d} \leq n^{2 \log n}$ entries of ξ_i and over all $i \leq 2d$, we get $\|\xi_0 - \dots + \xi_{2d}\| \leq \|\xi_0 - \dots + \xi_{2d}\|_F \leq n^{-16d}$ with probability $1 - O(n^{-96 \log n})$. \square

Appendix A. Omitted proofs.

A.1. Calibration of $\tilde{\mathbb{E}}$. In this subsection, we prove Lemma 5.3, restated here.

LEMMA A.1 (restatement of Lemma 5.3). *Let $f_G(x) = \sum_{|S| \leq 2d} c_S(G) \cdot x_S$ be a real-valued polynomial on $\{0, 1\}^n$ whose coefficients have degree at most τ when expressed in the ± 1 indicators G_e for edges in G . Then $\mathbb{E}_{G \sim G(n, \frac{1}{2})}[\tilde{\mathbb{E}}[f_G(x)]] = \mathbb{E}_{(H, x) \sim G(n, 1/2, \omega)}[f_H(x)]$.*

Proof. The proof is straightforward by expanding the coefficients f in the Fourier basis. For $S \subseteq [n]$, let $c_S : G \mapsto \mathbb{R}$ be maps so that $f_G(x) = \sum_{S \subseteq [n]} c_S \cdot x_S$:

$$\begin{aligned}
 \mathbb{E}_{G \sim G(n, \frac{1}{2})}[\tilde{\mathbb{E}}[f_G(x)]] &= \mathbb{E}_{G \sim G(n, \frac{1}{2})} \left[\tilde{\mathbb{E}} \left[\sum_{S \subseteq [n]} c_S \cdot x_S \right] \right] \\
 &= \sum_{S \subseteq [n]} \mathbb{E}_{G \sim G(n, \frac{1}{2})} [c_S \tilde{\mathbb{E}}[x_S]] \\
 &= \sum_{S \subseteq [n]} \mathbb{E}_{G \sim G(n, \frac{1}{2})} \left[\sum_{T, T' \subseteq \binom{[n]}{2}} \widehat{c_S}(T) \widehat{\tilde{\mathbb{E}}[x_S]}(T') \cdot \chi_T \chi_{T'} \right] \\
 &= \sum_{S \subseteq [n]} \sum_T \widehat{c_S}(T) \mathbb{E}_{(H, x) \sim G(n, 1/2, \omega)} [\chi_T(H) \cdot x_S] \\
 &= \mathbb{E}_{(H, x) \sim G(n, 1/2, \omega)} \left[\sum_{S \subseteq [n]} \sum_T \widehat{c_S}(T) \chi_T(H) \prod_{i \in S} x_i \right] \\
 &= \mathbb{E}_{(H, x) \sim G(n, 1/2, \omega)} \left[\sum_{S \subseteq [n]} c_S \prod_{i \in S} x_i \right] \\
 &= \mathbb{E}_{(H, x) \sim G(n, 1/2, \omega)} [f_H(x)]. \quad \square
 \end{aligned}$$

A.2. Concentration bounds for linear constraints. In this section, we prove Lemma 5.4. We will use the following elementary concentration bound repeatedly. (It is the scalar version of the matrix concentration bound Lemma 7.8; we state and prove a scalar version here because it is a good warm up for Lemma 7.8.)

LEMMA A.2. *Let \mathcal{T} be a family of subsets of $\binom{[n]}{2}$ so that for every $T, T' \in \mathcal{T}$ there exists a permutation $\sigma : [n] \rightarrow [n]$ of vertices so that $\sigma(T) = T'$. Let t be the number of vertices incident to edges in any $T \in \mathcal{T}$. For every $s \geq 0$ and every even ℓ ,*

$$\mathbb{P}_{G \sim G(n, 1/2)} \left\{ \left| \sum_{T \in \mathcal{T}} \chi_T(G) \right| \leq s \right\} \geq 1 - \frac{n^{t\ell/2} \cdot (t\ell)^{t\ell}}{s^\ell}.$$

Proof. Let $\ell \in \mathbb{N}$ be a parameter that will be chosen later. We will estimate $\mathbb{E}_{G \sim G(n, 1/2)}[(\sum_{T \in \mathcal{T}} \chi_T)^{\ell}]$:

$$\begin{aligned} \mathbb{E}_{G \sim G(n, 1/2)} \left[\left(\sum_{T \in \mathcal{T}} \chi_T \right)^\ell \right] &= \sum_{T_1, \dots, T_\ell \in \mathcal{T}} \mathbb{E}_{G \sim G(n, 1/2)} \prod_{j \leq \ell} \chi_{T_j} \\ &= \left| \left\{ (T_1, \dots, T_\ell) : \mathbb{E} \prod_{j \leq \ell} \chi_{T_j} = 1 \right\} \right|. \end{aligned}$$

In order to have $\mathbb{E} \prod_{j \leq \ell} \chi_{T_j} = 1$, every edge in the multiset $\bigcup_{j \leq \ell} T_j$ must appear at least twice, and so every vertex in the multiset $\bigcup_{j \leq \ell} \mathcal{V}(T_j)$ also appears at least twice. Thus, this multiset contains at most $t\ell/2$ distinct vertices. Since each $T_j \in \mathcal{T}$, each is uniquely determined by an ordered tuple of t elements of $[n]$. Thus, there are at most $n^{t\ell/2} \cdot (t\ell)^{t\ell}$ distinct choices for (T_1, \dots, T_ℓ) , and so

$$\mathbb{E}_{G \sim G(n, 1/2)} \left[\left(\sum_{T \in \mathcal{T}} \chi_T \right)^\ell \right] \leq n^{t\ell/2} \cdot (t\ell)^{t\ell}.$$

For even ℓ , by Markov's inequality,

$$\begin{aligned} \mathbb{P} \left\{ \left| \sum_{T \in \mathcal{T}} \chi_T \right| > s \right\} &= \mathbb{P} \left\{ \left| \sum_{T \in \mathcal{T}} \chi_T \right|^\ell > s^\ell \right\} \\ &\leq \frac{n^{t\ell/2} \cdot (t\ell)^{t\ell}}{s^\ell}. \end{aligned} \quad \square$$

LEMMA A.3 (restatement of Lemma 5.4). *With high probability, $\tilde{\mathbb{E}}[1] = 1 \pm n^{-\Omega(\varepsilon)}$ and $\tilde{\mathbb{E}}[\sum_{i \in [n]} x_i] = \omega \cdot (1 \pm n^{-\Omega(\varepsilon)})$.*

Proof. We will prove the statement regarding $\tilde{\mathbb{E}}[1]$; the bound for $\tilde{\mathbb{E}}[\sum_{i \in [n]} x_i]$ is almost identical.

Recall the Fourier expansion

$$\tilde{\mathbb{E}}[1] - 1 = \sum_{\substack{T \subseteq \binom{[n]}{2} \\ 2 \leq |\mathcal{V}(T)| \leq \tau}} \left(\frac{\omega}{n} \right)^{|\mathcal{V}(T)|} \cdot \chi_T.$$

Considering each $T \subseteq \binom{[n]}{2}$ as a graph, we partition $\{T \subseteq \binom{[n]}{2} : |\mathcal{V}(T)| = t\}$ into p_t families $\{\mathcal{T}_i^t\}_{i=1}^{p_t}$ by placing T and T' in the same family if and only if there exists a permutation $\sigma : [n] \rightarrow [n]$ of vertices so that $\sigma(T) = T'$. Thus,

$$\tilde{\mathbb{E}}[1] - 1 = \sum_{t=2}^{\tau} \left(\frac{\omega}{n} \right)^t \sum_{i=1}^{p_t} \sum_{T \in \mathcal{T}_i^t} \chi_T \leq \sum_{t=2}^{\tau} \left(\frac{\omega}{n} \right)^t \sum_{i=1}^{p_t} \left| \sum_{T \in \mathcal{T}_i^t} \chi_T \right|.$$

By Lemma A.2 (taking $\ell = (\log n)^2$), and since $t \leq \tau \leq \log n$, each \mathcal{T}_i^t satisfies

$$\mathbb{P} \left\{ \left| \sum_{T \in \mathcal{T}_i^t} \chi_T \right| < O(n^{t/2} \cdot (\log n)^{3t}) \right\} \geq 1 - (\tau \cdot 2^{t^2} \cdot n^{\log n})^{-1}.$$

By a union bound over all $p_t \leq 2^{t^2}$ families \mathcal{T}_i^t , we get that with high probability

$$|\tilde{\mathbb{E}}[1] - 1| \leq \tau \cdot \max_{t \leq \tau} \left(2^{t^2} \cdot \left(\frac{\omega}{\sqrt{n}} \right)^t \right).$$

For $\tau \leq (\varepsilon/2) \log n$ and $\omega = n^{1/2-\varepsilon}$, this is at most $n^{-\Omega(\varepsilon)}$. \square

A.3. Combinatorial proofs about ribbons. In this section, we prove Lemma 6.3, restated here.

LEMMA A.4 (restatement of Lemma 6.3). *Let \mathcal{R} be an (I, J) -ribbon. There is a unique minimum vertex separator S of \mathcal{R} such that S separates I and J for any vertex separator Q of \mathcal{R} . We call S the leftmost separator in \mathcal{R} . We define the rightmost separator analogously, and we denote them by $S_L(\mathcal{R})$ and $S_R(\mathcal{R})$, respectively.*

We start by defining a natural partial order on the set of vertex separators in a ribbon \mathcal{R} .

DEFINITION A.5. *We write $Q_1 \leq Q_2$ for two vertex separators Q_1 and Q_2 of an (I, J) -ribbon \mathcal{R} if Q_1 separates I and Q_2 .*

Next, we check that the definition above indeed is a partial order.

LEMMA A.6. *For any set of minimum vertex separators Q_1, Q_2, Q_3 of an (I, J) -ribbon, we have the following:*

1. $Q_1 \leq Q_1$.
2. If $Q_1 \leq Q_2$ and $Q_2 \leq Q_3$, then $Q_1 \leq Q_3$.
3. If $Q_1 \leq Q_2$ and $Q_2 \leq Q_1$, then $Q_1 = Q_2$.

Proof. The first statement is immediate from the definition. For the second, consider a path P from I to Q_3 in \mathcal{R} . Since $Q_2 \leq Q_3$, P passes through a vertex in Q_2 . Thus, P contains a subpath that connects I and Q_2 . But since $Q_1 \leq Q_2$, this subpath must pass through Q_1 . Thus, any such P must pass through Q_1 , and so $Q_1 \leq Q_3$.

Finally, for the third statement, let $k = |Q_1| = |Q_2|$. Then, using Menger's theorem (Fact 4.2), there is a set of k vertex-disjoint paths P_1, P_2, \dots, P_k between I and J . By virtue of Q_1, Q_2 being *minimum* vertex separators of \mathcal{R} , Q_1 and Q_2 must intersect each P_i in exactly one vertex. It is then immediate that the only way $Q_1 \leq Q_2$ and $Q_2 \leq Q_1$ is if every P_i intersects Q_1, Q_2 in the same vertex. \square

Now we can prove Lemma 6.3.

Proof of Lemma 6.3. It is enough to show that for any two minimum separators Q_1, Q_2 of size k in R there are separators Q_L, Q_R such that $Q_L \leq Q_1 \leq Q_R$ and $Q_L \leq Q_2 \leq Q_R$. We now construct Q_L and Q_R , as required.

Let $U = Q_1 \cap Q_2$ and $V = Q_1 \Delta Q_2$. Let $W_L \subseteq V$ be the set of vertices w such that there is a path from I to w that doesn't pass through $Q_1 \cup Q_2$. Similarly, let $W_R \subseteq V$ be the set of vertices w such that there is a path from w to some vertex in J that doesn't pass through any vertex in $Q_1 \cup Q_2$. Then we first observe the following.

CLAIM A.7. $W_L \cap W_R = \emptyset$.

Proof of claim. Assume otherwise, and let $w \in W_L \cap W_R$. Then there is a path between I and J that doesn't go through any vertex in at least one of Q_1 or Q_2 , contradicting that both are in fact vertex separators. \square

Next, we have the following.

CLAIM A.8. *Let $Q_L = U \cup W_L$ and $Q_R = U \cup W_R$. Then Q_L, Q_R are both vertex separators in R .*

Proof of claim. We only give the argument for Q_L , the other case being similar. Assume there is a path P from I to J that doesn't pass through Q_L . P must intersect $Q_1 \cup Q_2$. Then there is a vertex $v \in Q_1 \cup Q_2$ such that there is a path I to v which

intersects no other vertices in $Q_1 \cup Q_2$. This implies that either $v \in U$ or $v \in W_L$. But by our construction of W_L this is a contradiction. \square

Finally, we note that both Q_L and Q_R must in fact be *minimum* vertex separators.

CLAIM A.9. $|Q_L| = |Q_R| = |Q_1| = |Q_2| = k$.

Proof of claim. Let $|Q_1| = |Q_2| = k$. Then $2k = |Q_1| + |Q_2| = 2|U| + |V| \geq 2|U| + |W_L| + |W_R| = |U \cup W_L| + |U \cup W_R| = |Q_L| + |Q_R|$. Since Q_L and Q_R are vertex separators, $|Q_L|, |Q_R| \geq k$. Thus, $|Q_L| = |Q_R| = k$. \square

Finally, we have the ordering requirement on Q_L and Q_R .

CLAIM A.10. $Q_L \leq Q_1$ and $Q_2 \leq Q_R$.

Proof of claim. Let P be a path from I to Q_1 , and let v be the first vertex on this path which is in $Q_1 \cup Q_2$. Then $v \in U$ or $v \in W_L$. Thus, $Q_L \leq Q_1$. The other case is similar.

This concludes the proof of the lemma. \square

Appendix B. Spectral norms. The results in this section are in essence due to Medarametla and Potechin [MP16]. For completeness, we state and prove them here in the language and notation of the current paper, with minor modifications as needed.

LEMMA B.1 (restatement of Lemma 7.8). *Let U be a graph on $t \leq O(\log n)$ vertices, with two distinguished subsets of vertices A and B , and suppose the following:*

- *U admits p vertex-disjoint paths from $A \setminus B$ to $B \setminus A$.*
- *$|A \cap B| = r$.*
- *Every vertex outside $A \cup B$ has degree at least 1.*

Let $M = M(G)$ be the graphical matrix with shape U . Then, with high probability, $\|M\| \leq n^{\frac{t-p-r}{2}} \cdot 2^{O(t)} \cdot (\log n)^{O(t-r+p)}$.

Proof of Lemma 7.8. We proceed by the trace power method, with a dependence-breaking step beforehand.

Breaking dependence. Let q_1, \dots, q_p be vertex-disjoint paths from $A \setminus B$ to $B \setminus A$ in U . Without loss of generality, we can take each to intersect $A \setminus B$ and $B \setminus A$ only at its endpoints. We will partition the space of labelings σ into disjoint sets S_1, \dots, S_m . For each S_k , there will be a partition V_1^k, V_2^k of $[n]$ so that $\sigma(\bigcup_{j \leq p} q_j) \subseteq V_1^k$ and $\sigma(U \setminus (\bigcup_{j \leq p} q_j)) \subseteq V_2^k$ for every $\sigma \in S_k$. Let $(V_1^1, V_2^1), \dots, (V_1^m, V_2^m)$ be a sequence of independent uniformly random partitions of $[n]$. Call a labeling σ *good* at k if the preceding conditions apply to σ for the partition V_1^k, V_2^k and not for any $V_1^{k'}, V_2^{k'}$ for some $k' < k$. Let $S_k = \{\sigma : \sigma \text{ is good at } k\}$.

CLAIM B.2. *There is $m = O(2^t \cdot t \cdot \log n)$ so that $\bigcup_{k=1}^m S_k$ contains every labeling $\sigma : U \rightarrow G$.*

Proof. For a fixed σ ,

$$\mathbb{P}\{\sigma \text{ not good for some } k \leq m\} \leq (1 - 2^{-t})^m$$

since every vertex $u \in U$ is in V_i with probability $1/2$. If $m \geq 10t2^t \log n$, then by a union bound over all $\sigma : U \rightarrow G$ (of which there are at most n^t), we get $\mathbb{P}\{\text{all } \sigma \text{ good for some } k \leq m\} > 0$. \square

Henceforth, let S_1, \dots, S_m be the partition guaranteed by the preceding claim. For $k \leq m$, let $M_k(I, J) = \sum_{\sigma \in S_k : \sigma(A)=I, \sigma(B)=J} \text{val}(\sigma)$. Then $M = \sum_{k=1}^m M_k$.

Moment calculation. Let $\ell = \ell(n)$ be a parameter to be chosen later. By the triangle inequality, $\|M\| \leq \sum_{k=1}^m \|M_k\|$. Fix k . We expand $\mathbb{E}_G \text{Tr}(M_k^\dagger M_k)^\ell$ as

$$\mathbb{E} \text{Tr}(M_k^\dagger M_k)^\ell = \mathbb{E} \sum_{\substack{\sigma_1, \dots, \sigma_{2\ell} \in S_k \\ \sigma_{2i}(A) = \sigma_{2i-1}(A) \\ \sigma_{2i}(B) = \sigma_{2i+1}(B)}} \prod_{j=1}^{2\ell} \text{val}(\sigma_j).$$

(Here arithmetic with indices i is modulo 2ℓ , and so, for example, we take $2i+1 = 1$.) For any σ ,

$$\text{val}(\sigma) = \prod_{(i,j) \in U} G_{\sigma(i), \sigma(j)}.$$

Notice that for all $\sigma_1, \dots, \sigma_{2\ell}$, the expectation $\mathbb{E} \prod_{j=1}^{2\ell} \text{val}(\sigma_j)$ is either 0 or 1. We will bound the number of $\sigma_1, \dots, \sigma_{2\ell}$ for which $\mathbb{E} \prod_{j=1}^{2\ell} \text{val}(\sigma_j) = 1$ by bounding the number of distinct labels such a family of labelings may assign to vertices in U .

Fix $\sigma_1, \dots, \sigma_{2\ell} \in S_k$. Consider the family q_1, \dots, q_p of vertex-disjoint paths. Every edge in every q_j receives one pair of labels from each σ_i . Consider these labels arranged on 2ℓ adjoined copies of each q_j , one for each σ (giving p paths with $2\ell \sum_{j \leq p} |q_j|$ edges in total, where $|q_j|$ is the number of edges in q_j). Every pair of labels $\{\sigma_i(v), \sigma_i(w)\}$ appearing on an edge (v, w) in this graph must also appear on some distinct edge (v', w') in order to have $\mathbb{E} \prod_{i=1}^{2\ell} \text{val}(\sigma_i) = 1$; otherwise, the disjointness of V_1^k, V_2^k would be violated. Merging edges which received the same pair of labels, we arrive at a graph with at most p connected components and at most $\ell \sum_{j \leq p} |q_j|$ edges, and so at most $\ell \sum_{j \leq p} |q_j| + p$ vertices. Thus, the vertices in q_1, \dots, q_p together receive at most $\ell \sum_{j \leq p} |q_j| + p$ distinct labels among all $\sigma_1, \dots, \sigma_{2\ell}$.

Next, we account for labels of $v \notin (\bigcup_{j \leq p} q_j \cup A \cup B)$. If $\mathbb{E}_G \prod_{i=1}^{2\ell} \text{val}(\sigma_i) = 1$, then the 2ℓ -sized multiset $\{\sigma_i(v)\}_{i \leq 2\ell}$ of labels for such v contains at most ℓ distinct labels since by assumption, v has degree at least 1 in U .

Next, we account for labels of vertices in $A \setminus (B \cup \bigcup_{j \leq p} q_j)$ and $B \setminus (A \cup \bigcup_{j \leq p} q_j)$. Every such vertex receives a label from every σ_i , but σ_{2i} and σ_{2i-1} must agree on A -labels and σ_{2i} and σ_{2i+1} must agree on B -labels. So in total there are at most $\ell(|A| + |B| - 2p - 2r)$ distinct labels for such vertices.

This means that among the labels $\sigma_i(j)$ for all $j \notin A \cap B$, there are at most

$$\underbrace{\ell \sum_{j \leq p} |q_j| + p}_{\text{labels from paths}} + \underbrace{\ell(|A| + |B| - 2p - 2r)}_{\text{additional vertices in } A \cup B \setminus (A \cap B)} + \underbrace{\ell(t - (|A| + |B| - r) - (\sum_j |q_j| - p))}_{\text{vertices in } U \setminus (\bigcup_j q_j \cup A \cup B)} = \ell(t - p - r) + p$$

unique labels.

Finally, consider the labels of the r vertices j_1, \dots, j_r in $A \cap B$. The first labeling σ_1 assigns these vertices some $\sigma_1(j_1), \dots, \sigma_1(j_r)$ labels in G . Since σ_2 agrees with σ_1 on A -vertices, we must have $\sigma_2(j_1) = \sigma_1(j_1), \dots, \sigma_1(j_r) = \sigma_2(j_r)$. Since σ_3 agrees with σ_2 on B -vertices, we must have $\sigma_3(j_1) = \sigma_2(j_1), \dots, \sigma_3(j_r) = \sigma_2(j_r)$. And so on. So there are at most r unique labels for such vertices.

Now we can assess how many choices there are for $\sigma_1, \dots, \sigma_{2\ell} \in S_k$ so that $\mathbb{E} \prod_{i \leq 2\ell} \text{val}(\sigma_i) = 1$. To choose such a collection $\sigma_1, \dots, \sigma_{2\ell}$, we proceed in stages.

Stage 1. Choose the labels $\sigma_i(j_1), \dots, \sigma_i(j_r)$ of all the vertices in $A \cap B$. Here there are at most n^r options.

Stage 2. For each pair (i, j) , where $j \notin A \cap B$, choose whether $\sigma_i(j)$ will be the first appearance of the index $\sigma_i(j) \in [n]$ or if there are some $i' < i$ and j' so that $\sigma_{i'}(j') = \sigma_i(j)$. Here there are $2^{2\ell t}$ options.

Stage 3. Choose the labels $\sigma_i(j) \in [n]$ for all $j \notin A \cap B$ and pairs (i, j) which in Stage 2 we chose to be the first appearance of a label. If there are x such vertices, there are at most n^x options.

Stage 4. Choose the labels $\sigma_i(j) \in [n]$ for all the pairs (i, j) , with $j \notin A \cap B$, which in Stage 2 we chose not to be the first appearance of a label. Here there are at most $x^{2\ell t - 2\ell r - x}$ options.

All together, there are at most $n^r \cdot 2^{2\ell t} \cdot n^x \cdot x^{2\ell(t-r)-x} \leq n^r \cdot 2^{2\ell t} \cdot n^x \cdot (2\ell t)^{2\ell(t-r)-x}$ choices for a given x . Since $4\ell t \ll n$, summing up over all $x \leq \ell(t-p-r) + p$, the total number of choices is at most $2n^r \cdot 2^{2\ell t} \cdot n^{\ell(t-p-r)+p} \cdot (2\ell t)^{\ell(t-r+p)-p}$. Putting it together,

$$\mathbb{E} \text{Tr}(M_k^\dagger M_k)^\ell \leq 2n^r \cdot n^{\ell(t-p-r)+p} \cdot (2\ell t)^{\ell(t-r+p)-p}.$$

Now using Markov's inequality and standard manipulations, for any s ,

$$\begin{aligned} \mathbb{P}\{\|M_k\| \geq s\} &= \mathbb{P}\{\|M_k^\dagger M_k\|^\ell \geq s^{2\ell}\} \\ &\leq \frac{\mathbb{E} \|(M_k^\dagger M_k)^\ell\|}{s^{2\ell}} \quad (\text{by Markov's inequality}) \\ &\leq \frac{\mathbb{E} \text{Tr}(M_k^\dagger M_k)^\ell}{s^{2\ell}} \quad (\text{since } \|(M_k^\dagger M_k)^\ell\| \leq \text{Tr}(M_k^\dagger M_k)^\ell) \\ &\leq \frac{2n^r \cdot 2^{2\ell t} \cdot n^{\ell(t-p-r)+p} \cdot (2\ell t)^{\ell(t-r+p)-p}}{s^{2\ell}}. \end{aligned}$$

By taking $\ell = (\log n)^3$ and by using $p \leq t \leq O(\log n)$, there exists $s = 2^t \cdot n^{(t-p-r)/2} (\log n)^{O(t-r+p)}$ so that $\mathbb{P}\{\|M_k\| \geq s\} \leq n^{-100 \log n} m^{-1}$. By a union bound, $\mathbb{P}\{\|M_k\| \leq s \text{ for all } k\} \geq 1 - n^{-100 \log n}$, and so $\|M\| \leq sm$ with probability $1 - n^{-100 \log n}$. Since $m \leq 2^{O(t)} \cdot \log(n)^{O(1)}$, this completes the proof. \square

Acknowledgments. We thank Raghu Meka, Ryan O'Donnell, Prasad Raghavendra, Tselil Schramm, David Steurer, and Avi Wigderson for many useful discussions related to this paper. We thank the anonymous referees for FOCS 2016 and SICOMP for their careful reading, pointing out several typos and concrete suggestions for the improvement of this manuscript.

REFERENCES

- [AAK+07] N. ALON, A. ANDONI, T. KAUFMAN, K. MATULEF, R. RUBINFELD, AND N. XIE, *Testing k -wise and almost k -wise independence*, in Proceedings of the 39th Annual ACM Symposium on Theory of Computing, 2007, pp. 496–505.
- [ABBG11] S. ARORA, B. BARAK, M. BRUNNERMEIER, AND R. GE, *Computational complexity and information asymmetry in financial products*, Commun. ACM, 54 (2011), pp. 101–107.
- [ABC13] P. AUSTRIN, M. BRAVERMAN, AND E. CHLAMTÁČ, *Inapproximability of NP-complete variants of Nash equilibrium*, Theory Comput., 9 (2013), pp. 117–142.
- [ABW10] B. APPLEBAUM, B. BARAK, AND A. WIGDERSON, *Public-key cryptography from different assumptions*, in Proceedings of the 42nd Annual ACM Symposium on Theory of Computing, 2010, pp. 171–180.
- [AKS98] N. ALON, M. KRIVELEVICH, AND B. SUDAKOV, *Finding a large hidden clique in a random graph*, in Proceedings of the Ninth Annual ACM-SIAM Symposium on Discrete Algorithms, 1998, pp. 594–598.
- [BBH+12] B. BARAK, F. G. S. L. BRANDAO, A. W. HARROW, J. KELNER, D. STEURER, AND Y. ZHOU, *Hypercontractivity, sum-of-squares proofs, and their applications*, in

- Proceedings of the 44th annual ACM Symposium on Theory of Computing, 2012, pp. 307–326.
- [BCK15] B. BARAK, S. O. CHAN, AND P. K. KOTHARI, *Sum of squares lower bounds from pairwise independence*, in Proceedings of the 47th Annual ACM Symposium on Theory of Computing, 2015, pp. 97–106.
- [BKS14] B. BARAK, J. A. KELNER, AND D. STEURER, *Rounding sum-of-squares relaxations*, in Proceedings of the 46th Annual ACM Symposium on Theory of Computing, 2014, pp. 31–40.
- [BKS15] B. BARAK, J. A. KELNER, AND D. STEURER, *Dictionary learning and tensor decomposition via the sum-of-squares method*, in Proceedings of the 47th Annual ACM Symposium on Theory of Computing, 2015, pp. 143–151.
- [BR13] Q. BERTHET AND P. RIGOLLET, *Complexity theoretic lower bounds for sparse principal component detection*, in Proceedings of the 26th Annual Conference on Learning Theory, 2013, pp. 1046–1066.
- [BS14] B. BARAK AND D. STEURER, *Sum-of-squares proofs and the quest toward optimal algorithms*, in Proceedings of the International Congress of Mathematicians, 2014, pp. 509–533.
- [BT06] A. BOGDANOV AND L. TREVISAN, *On worst-case to average-case reductions for NP problems*, SIAM J. Comput., 36 (2006), pp. 1119–1159, <https://doi.org/10.1137/S0097539705446974>.
- [DM15] Y. DESHPANDE AND A. MONTANARI, *Improved sum-of-squares lower bounds for hidden clique and hidden submatrix problems*, in Proceedings of the 28th Annual Conference on Learning Theory, 2015, pp. 523–562.
- [FF93] J. FEIGENBAUM AND L. FORTNOW, *Random-self-reducibility of complete sets*, SIAM J. Comput., 22 (1993), pp. 994–1005, <https://doi.org/10.1137/0222061>.
- [FK96] A. FRIEZE AND R. KANNAN, *The regularity lemma and approximation schemes for dense problems*, in Proceedings of the 37th Annual IEEE Symposium on Foundations of Computer Science, 1996, pp. 12–20.
- [FK03] U. FEIGE AND R. KRAUTHGAMER, *The probable value of the Lovász–Schrijver relaxations for maximum independent set*, SIAM J. Comput., 32 (2003), pp. 345–370, <https://doi.org/10.1137/S009753970240118X>.
- [Gra95] A. GRANVILLE, *Harald Cramér and the distribution of prime numbers*, Scand. Actuar. J., 1995 (1995), pp. 12–28.
- [Gri01] D. GRIGORIEV, *Complexity of Positivstellensatz proofs for the knapsack*, Comput. Complexity, 10 (2001), pp. 139–154.
- [HK11] E. HAZAN AND R. KRAUTHGAMER, *How hard is it to approximate the best Nash equilibrium?*, SIAM J. Comput., 40 (2011), pp. 79–91, <https://doi.org/10.1137/090766991>.
- [HKP15] S. B. HOPKINS, P. K. KOTHARI, AND A. POTECHIN, *SoS and Planted Clique: Tight Analysis of MPW Moments at All Degrees and an Optimal Lower Bound at Degree Four*, preprint, <https://arxiv.org/abs/1507.05230>, 2015.
- [HKP+16] S. B. HOPKINS, P. KOTHARI, A. H. POTECHIN, P. RAGHAVENDRA, AND T. SCHRAMM, *On the integrality gap of degree-4 sum of squares for planted clique*, in Proceedings of the Twenty-Seventh Annual ACM-SIAM Symposium on Discrete Algorithms, 2016, pp. 1079–1095, <https://doi.org/10.1137/1.9781611974331.ch76>.
- [HWX15] B. E. HAJEK, Y. WU, AND J. XU, *Computational lower bounds for community detection on random graphs*, in Proceedings of the 28th Conference on Learning Theory, 2015, pp. 899–928.
- [Jay57a] E. T. JAYNES, *Information theory and statistical mechanics*, Phys. Rev. (2), 106 (1957), pp. 620–630.
- [Jay57b] E. T. JAYNES, *Information theory and statistical mechanics. II*, Phys. Rev. (2), 108 (1957), pp. 171–190.
- [Jer92] M. JERRUM, *Large cliques elude the metropolis process*, Random Structures Algorithms, 3 (1992), pp. 347–360.
- [JM15] H. JAVADI AND A. MONTANARI, *The Hidden Subgraph Problem*, preprint, <https://arxiv.org/abs/1511.05254v1>, 2015.
- [JP00] A. JUELS AND M. PEINADO, *Hiding cliques for cryptographic security*, Des. Codes Cryptogr., 20 (2000), pp. 269–280.
- [Kar76] R. M. KARP, *The probabilistic analysis of some combinatorial search problems*, in Algorithms and Complexity: New Directions and Recent Results, Academic Press, New York, 1976, pp. 1–19.
- [Kuc95] L. KUCERA, *Expected complexity of graph partitioning problems*, Discrete Appl. Math.,

- 57 (1995), pp. 193–212.
- [KZ14] P. KOIRAN AND A. ZOUZIAS, *Hidden cliques and the certification of the restricted isometry property*, IEEE Trans. Inform. Theory, 60 (2014), pp. 4999–5006.
- [Las01] J. B. LASSERRE, *An explicit exact SDP relaxation for nonlinear 0–1 programs*, in Proceedings of the Eighth Conference on Integer Programming and Combinatorial Optimization, 2001, pp. 293–303.
- [MP16] D. MEDARAMETLA AND A. POTECHIN, *Bounds on the norms of uniform low degree graph matrices*, in Approximation, Randomization, and Combinatorial Optimization. Algorithms and Techniques, LIPIcs. Leibniz Int. Proc. Inform. 60, Schloss Dagstuhl. Leibniz-Zent. Inform., Wadern, Germany, 2016, article 40.
- [MPW15] R. MEKA, A. POTECHIN, AND A. WIGDERSON, *Sum-of-squares lower bounds for planted clique*, in Proceedings of the 47th Annual ACM Symposium on Theory of Computing, 2015, pp. 87–96.
- [MSOI+02] R. MILO, S. SHEN-ORR, S. ITZKOVITZ, N. KASHTAN, D. CHKLOVSKII, AND U. ALON, *Network motifs: Simple building blocks of complex networks*, Science, 298 (2002), pp. 824–827.
- [MW15] T. MA AND A. WIGDERSON, *Sum-of-squares lower bounds for sparse PCA*, in Advances in Neural Information Processing Systems, MIT Press, Cambridge, MA, 2015, pp. 1603–1611.
- [Par00] P. A. PARRILO, *Structured Semidefinite Programs and Semialgebraic Geometry Methods in Robustness and Optimization*, Ph.D. thesis, California Institute of Technology, Pasadena, CA, 2000.
- [PS00] P. A. PEVZNER AND S.-H. SZE, *Combinatorial approaches to finding subtle signals in DNA sequences*, in Proceedings of the Eighth International Conference on Intelligent Systems for Molecular Biology, pp. 269–278.
- [Sch08] G. SCHOENEBECK, *Linear level Lasserre lower bounds for certain k -CSPs*, in Proceedings of the 49th Annual IEEE Symposium on Foundations of Computer Science, 2008, pp. 593–602.
- [Sho87] N. Z. SHOR, *Class of global minimum bounds of polynomial functions*, Kibernetika, 6 (1987), pp. 9–11 (in Russian); Cybernetics, 23 (1987), pp. 731–734 (in English).
- [Sze78] E. SZEMERÉDI, *Regular partitions of graphs*, in Problèmes combinatoires et théorie des graphes, CNRS, Paris, 1978, pp. 399–401.
- [Tao05] T. TAO, *The Dichotomy Between Structure and Randomness, Arithmetic Progressions, and the Primes*, preprint, <https://arxiv.org/abs/math/0512114>, 2005.
- [Tao15] T. TAO, 254A, *Supplement 4: Probabilistic Models and Heuristics for the Primes (Optional)*; available online from <https://terrytao.wordpress.com/2015/01/04/254a-supplement-4-probabilistic-models-and-heuristics-for-the-primes-optional/>.