Electronic Colloquium on Computational Complexity, Report No. 58 (2016)

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

Boaz Barak [*] Harvard University
Samuel B. Hopkins [†] Cornell University
Jonathan Kelner [‡] MIT
Pravesh K. Kothari [§] UT Austin
Ankur Moitra [¶] MIT
Aaron Potechin [‖] Cornell University

April 12, 2016

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 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 *pseudo-calibration* to construct Sum of Squares lower bounds. This framework is inspired by taking a computational analog of Bayesian probability theory. It yields a general recipe for constructing good pseudo-distributions (i.e., dual certificates for the Sum-of-Squares semidefinite program), and sheds further light on the ways in which this hierarchy differs from others.

[*] Harvard John A. Paulson School of Engineering and Applied Sciences, b@boazbarak.org. Part of the work was done while the author was at Microsoft Research New England.

[†] samhop@cs.cornell.edu. Partially supported by an NSF GRFP under grant no. 1144153. Part of this work was done while the author was at Microsoft Research New England.

[‡] kelner@mit.edu. Partially supported by NSF Award 1111109.

[§] kothari@cs.utexas.edu Part of the work was done while the author was at Microsoft Research New England.

[¶] moitra@mit.edu Partially supported by NSF CAREER Award CCF-1453261, a grant from the MIT NEC Corporation and a Google Faculty Research Award.

[‖] aaronpotechin@gmail.com Part of the work was done while the author was at Microsoft Research New England.

ISSN 1433-8092

Contents

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 is solvable by brute force search (which in this case takes quasipolynomial time) whenever $\omega > c \log n$ for any constant $c > 2$. However, despite intense effort, the best polynomial-time algorithms only work for $\omega = \varepsilon \sqrt{n}$, for any constant $\varepsilon > 0$ [AKS98].

Over the years the planted clique problem and related problems have been connected to many other questions in a variety of areas including finding communities [HWX15], finding signals in molecular biology [PS000], discovering network motifs in biological networks [MSOI+02, JM15], computing Nash equilibrium [HK11, ABC13], property testing [AAK+07], sparse principal component analysis [BR13], compressed sensing [KZ14], cryptography [JP00, ABW10] and even mathematical finance [DBL10].

Thus, the question of whether the currently known algorithms can be improved is of great interest. Unfortunately, it is unlikely that lower bounds for planted clique (because it is an average-case problem) can be derived from conjectured complexity class separations such as $P \neq NP$ [FF93, BT06]. Our best evidence for the difficulty of this problem comes from works showing limitations on particular *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 algorithm are of significant interest. One such negative result was shown by Feige and Krauthgamer [FK03a] 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]. Thus it was conceivable that the SOS hierarchy might be able to find cliques that are much smaller than $\sqrt{n}$ in polynomial time.

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 on 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:

[1] As we discuss in Remark 1.2 below, 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.

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 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 to a certain extent 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 **NP**, the planted clique problem with parameter ω has three variants of *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 Sum-of-Squares hierarchy are 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 the planted clique problem 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, 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 between the planted clique problems and the problems tackled by previous results is that lower bounds for the planted clique problem boil down to handling **weak global constraints** as opposed to **strong local** ones. That is, while 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 the planted clique setting every variable has a weak *global* effect on all 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 that vertex 17 is not in S, then it makes it slightly less likely that 17's neighbors are in S and slightly more likely that 17's non-neighbors 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 17^{th} variable (the variables that are involved in constraints with 17 or one that have a short path of constraints to it) but there is an exponential decay of these correlations and so this information basically tells us essentially nothing about the distribution of most of the variables x_i (that are far away from 17 in the sparse graph induced by φ).x[2] Thus, in the random 3SAT setting information about the assignments of individual variables has a *strong local effect*. Indeed, previous Sum-of-Squares lower bounds for random 3SAT and 3XOR [Gri01, Sch08], could be interpreted as producing "distribution like" objects in which, conditioned on the value of a small set of variables S, some of the variables "close" to S in the formula were completely 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 setting of random constraint satisfaction problems need to be tackled head-on when dealing with planted cliques. However to make this clearer, we need to take a detour and discuss Bayesian probabilities and their relation to the Sum of Square Algorithm.

2.1 Computational Bayesian Probabilities and Pseudo-distributions

Strictly speaking, if a graph G contains a unique clique S of size ω, 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 i is in the clique or not, and we could try to quantify this ignorance using probabilities. These can be thought of as a computational analogs of *Bayesian probabilities*, that, rather than aiming to measure the frequency at which an event occurs in some sample space, attempt 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 E happened or not, she 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, she could

[2]This 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 overconstrainted 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 Sum-of-Squares or belief propagation. Both these algorithms exhibit this sort of exponential decay we talk about; see also Remark 2.1

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, etc.

The Sum-of-Squares 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 certain restricted proof system.

Bayesian estimates for planted clique. To get a sense for 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 n-vertex graphs G and vectors $x \in \mathbb{R}^n$ that 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 clique. Let $f : \{0,1\}^{\binom{n}{2}} \times \mathbb{R}^n \to \mathbb{R}$ 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 needs 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 will simply let $\tilde{\mathbb{E}}_G f_G$ 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

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

Now if Bob is computationally bounded, then he might not be able to compute the value of $E_{x|G} f_G(x)$ 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 0 otherwise. However, note that we don't need to compute the true conditional expectation to obtain a calibrated estimate. In particular, in the above example, simply setting $\tilde{\mathbb{E}} x_{17} = \omega/n$ will satisfy (2.1).

Our Sum-of-Squares lower bound amounts to coming up with some reasonable "pseudo-expectation" 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 pseudo-expectation will not be even close to the true conditional expectations, but will at least be internally consistent in the sense that for "simple" functions f it satisfies (2.1). It will also satisfy some basic sanity checks such as that for every graph G and "simple" f, $\tilde{\mathbb{E}}_G f_G^2 \geqslant 0$. In fact, since the pseudo-expectation will not distinguish between a graph G drawn from $G(n, 1/2, \omega)$ and a random G from $G(n, 1/2)$ it will also satisfy the following *pseudo-calibration* condition:

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

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

[3] The 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".

even in the support of $G(n, 1/2, \omega)$. Nevertheless, our pseudo-distributions 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 "pseudo-calibration" condition (2.2) might seem innocent, but it turns out to imply many useful properties. In particular 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 pseudo-expectation 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 non-neighbors are more likely to be in it.

Indeed, the key conceptual insight of this paper is to phrase the calibration property (2.2) as a desiderata for our pseudo-distributions. Namely, we define 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 require (2.2) to hold for all simple functions. It turns out that once you do so, the choice for the pseudo-distribution 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 some of the ideas involved in this proof.

Remark 2.1 (Planted Clique vs 3XOR). In the 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 $\approx 1/2$ the constraints can be satisfied by any assignment) with high probability. On the other hand, Grigorev [Gri01] constructed a sum of squares pseudoexpectation that pretends that such instances instances are satisfiable with high probability, proving a sum of squares 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 pseudo-calibrating with respect to this planted distribution a la (2.2) produces precisely the pseudoexpectation that Grigoriev constructed. However, unlike in the planted clique case, in the case of 3XOR, the pseudo-calibration 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 [FK03b, MPW15, DM15]) are similar to Grigoriev's construction, in the sense that they essentially respect only strong constraints (e.g., that 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 pseudo-calibration condition (2.2) and the pseudoexpectation of [FK03b, MPW15, DM15] can be shown to violate weak probabilistic constraints imposed by (2.2) even at degree four. See Observation 2.4 for an

example.

2.2 From Calibrated Pseudo-distributions to Sum-of-Squares Lower Bounds

What do these Bayesian inferences and calibrations have to do with Sum-of-Squares? In this section, we show how calibration is almost forced on any pseudodistribution feasible for the sum of squares algorithm. Specifically, 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. (Booleanity constraint) $\tilde{\mathbb{E}}_G x_i^2 p = \tilde{\mathbb{E}} x_i p$ 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 p of degree at most $d-2$.
5. (Size constraint) $\tilde{\mathbb{E}}_G \sum_{i=1}^n x_i = \omega$.
6. (Positivity) $\tilde{\mathbb{E}}_G p^2 \geqslant 0$ for every p of degree at most $d/2$.

Definition 2.2. A map $p \mapsto \tilde{\mathbb{E}}_G p$ satisfying the above constraints 1–6 is called a *degree d pseudo-distribution* (w.r.t. the planted clique problem with parameter ω).

We can restate our main result as follows:

Theorem 2.3 (Theorem 1.1, restated)**.** *There is some constant c such that if $\omega \leqslant n^{1/2 - c(d/\log n)^{1/2}}$ 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 not event a $2.1 \log n$-sized clique in G (and let alone a $\approx \sqrt{n}$ sized one) so we will need a completely different mechanism to obtain such a pseudo-distribution.

Previously, the choice of the pseudo-distribution 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 planted clique as well as some related problems [MW15] the choice of the pseudo-distribution seemed to have more freedom, and more than one choice appeared in the literature.

For example, Feige and Krauthgamer [FK03b] (henceforth FK) defined a very natural pseudo-distribution $\tilde{\mathbb{E}}^{FK}$ for a weaker hierarchy. For a graph G on n vertices, and subset $A \subseteq [n]$, $\tilde{\mathbb{E}}^{FK}_G x_A$ 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 extended to degree d polynomials using linearity.[4] [FK03b] showed that that for every d, and $\omega < O(\sqrt{n/2^d})$,

[4] The actual pseudo-distribution used by [FK03b] (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 $\ell \in \{1,\ldots,d\}$. This property is sometimes described as satisfying the constraint $\{\sum_i x_i = \omega\}$.

this pseudo-distribution satisfies the constraints 1–5 as in Definition 2.2 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 pseudo-distribution satisfies all the constraints 1–6 (and hence is a valid degree d pseudo-distribution) 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 [DM15] and to $\tilde{O}(n^{(\lfloor d/2 \rfloor + 1)^{-1}})$ for a general d by [HKP15].

Interestingly, the FK pseudo-distribution does *not* satisfy the full positivity constraint for larger values of ω. The issue is that while that while the FK pseudo-distribution satisfies the "strong" constraints that $\tilde{\mathbb{E}}^{FK}_G x_A = 0$ 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 pseudo-distribution does not make this correction. In particular, for every ℓ, there's a simple polynomial that shows that the FK pseudoexpectation is not calibrated.

Observation 2.4. Fix $i \in [n]$ and let ℓ be some constant. If $p_G = (\sum_j G_{i,j} x_j)^\ell$ then **(i)** $\mathbb{E}_{G \sim G(n,1/2)} \tilde{\mathbb{E}}^{FK}_G [p_G^2] \leqslant \omega^\ell$ and **ii** $\mathbb{E}_{(G,x) \sim G(n,1/2,\omega)} [p_G(x)^2] \geqslant \frac{\omega^{2\ell+1}}{n}$. In particular, when $\omega \gg n^{\frac{1}{\ell+1}}$, $\mathbb{E}_{G \sim G(n,1/2)} \tilde{\mathbb{E}}^{FK}_G [p_G^2] \ll \mathbb{E}_{(G,x) \sim G(n,1/2,\omega)} p_G(x)$.

Proof sketch. For 2 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 $(\omega/n)\omega^{2\ell}$. To compute 1, we open up the expectation and the definition to get (up to a constant depending on ℓ) $\sum_{j_1,\ldots,j_{2\ell}} G_{i,j_1} \ldots G_{i,j_{2\ell}} (\omega/n)^{2\ell} \mathbb{E}_{G \sim G(n,1/2)} 1_{\{i_1,\ldots,i_{2\ell}\} \text{ is clique}}$. 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, which means the sum is at most $n^\ell (\omega/n)^\ell = \omega^\ell$. □

Observation 2.4 notes 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 above can also be massaged to obtain a proof (due to Kelner, see [HKP15]) that degree d $\tilde{\mathbb{E}}^{FK}$ does not satisfy the positivity constraint at degree d for $\omega \gg n^{\frac{1}{\frac{d}{2}+1}}$.

Fact 2.5. *Let p_G be as in the Observation 2.4. 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 pseudo-distribution 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 2.5.

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 2.4 and Fact 2.5. Moreover, "fixing" these families by the correction as in case of degree four leads to new families of polynomials that fail to achieve their calibrated value and exhibit negative pseudoexpectation for their squares etc.

The *coefficients* of the polynomial p_G of Observation 2.4 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 pseudo-distribution must satisfy (2.2) for every such polynomial, and using that as our implicit definition of the pseudo-distribution. Indeed it turns our that once we do so, the pseudo-distribution is essentially determined. Moreover, (2.2) guarantees that it satisfies many of the "weak global constraints" that can be shown using Bayesian calculations.

Pseudo-calibrating polynomials whose coefficient are low-degree in G amounts to restricting the pseudo-distribution to satisfy that the map $G \mapsto \tilde{\mathbb{E}}_G$ is itself a low degree polynomial in G. Why is it OK to make such a restriction? One justification is the heuristic that the pseudo-distribution 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 pseudo-distribution to be low-degree we are essentially making it *smooth* or "high entropy", which is consistent with the Jaynes *maximum entropy principle* [Jay57b, Jay57a]. Most importantly – and this is the bulk of the technical work of this paper and the subject of the next subsection – this pseudo-distribution can be shown to satisfy *all* the constraints 1–6 of Definition 2.2 including the positivity constraint.

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

Remark 2.6 (*Where does the planted distribution arise from?*). Theorem 2.3 (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 pseudo-distribution 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 pseudo-distribution can be viewed as correcting this planted distribution to a posterior that respects simple inferences from the observed graph G.

2.3 Proving Positivity

Now we have seen that pseudocalibration is desirable both *a priori* and in light of the failure of previous lower-bound attempts. We turn to the question: how do we formally define a pseudo-calibrated linear map $\tilde{\mathbb{E}}_G$, and how do we show that it satisfies constraints (1) – (6) with high probability, to prove Theorem 2.3?

Recall that our goal is to give a map from G to $\tilde{\mathbb{E}}_G$ such that when G is taken from $G(n, 1/2)$ then with high probability $\tilde{\mathbb{E}}_G$ satisfies constraints 1–6 of Definition 2.2. Our strategy is to define $\tilde{\mathbb{E}}_G$ in a way that it satisfies the pseudo-calibration requirement (2.2) with respect to all functions $f = f(G, x)$ that are low degree polynomials in both the G and x variables. The above requirements determine all the low-degree Fourier coefficients of the map $G \mapsto \tilde{\mathbb{E}}_G$. Indeed, instantiating (2.2) with every particular function $f = f(G, x)$ defines a linear constraint on the pseudo-expectation operator. If

we 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 $S \subseteq [n]$, $|S| \leqslant d$, using the Fourier transform it is not too hard to compute $\tilde{\mathbb{E}}_G[x_S]$ as an explicit low degree polynomial in G_e:

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

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 $\tau \gtrapprox d/\varepsilon$. 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 a sum of squares 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 $\binom{n}{\leqslant d/2} \times \binom{n}{\leqslant d/2}$ 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 very simple random matrices–for example, a matrix with independent ± 1 entries—are not explicitly understood. Our moment matrix $\mathcal{M}$ is a much more complicated random matrix, with intricate dependencies among the entries. However, as the next example demonstrates, it is sometimes possible to prove PSDness for a random matrix using an *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 2.2, but only for $d = 2$. In this simple case, it turns out that the subtleties of (pseudo)calibration may safely be ignored, but it is still instructive to revisit the proof of PSDness. It will be 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 let $\tilde{\mathbb{E}} x_i x_j$ equal $\left(\frac{\omega}{n}\right)^2$ if (i, j) is an an edge in G and equal zero otherwise. It's not hard to show that positivity of this pseudo-expectation 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 (with eigenvector very close to the vector $\vec{u} = (1/\sqrt{n}, \ldots, 1/\sqrt{n})$) of value ω^2/n, while all others are distributed in the interval $\frac{\omega}{n} \pm O\left(\frac{\omega^2}{n^2}\sqrt{n}\right)$ 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 $\mathcal{N} \approx \frac{\omega^2}{n} \vec{u}\vec{u}^\dagger + \frac{\omega}{n} Id + E$ for some sufficiently small (in spectral norm) "error matrix" E.

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, 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 $\vec{u} = (1/\sqrt{n}, \ldots, 1/\sqrt{n})$.) As we saw in Observation 2.4, 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 $\mathcal{M} \approx \mathcal{L} Q \mathcal{L}^\dagger$ for some approximately-diagonal matrix 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 so small that we can ignore them 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 $f : \{0,1\}^{\binom{n}{2}} \times \mathbb{R}^n \to \mathbb{R}$ 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 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 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 to be find an approximate diagonalization for all the non-trivial 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 pseudo-calibration 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 G, 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 $x_1, \ldots, x_n$, 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 of 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 pseudo-calibration condition (2.2) in particular implies that for every f that is low degree in both G and x, $\mathbb{E}_G \tilde{\mathbb{E}}_G f_G^2 \geqslant 0$. 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 cause of much of the technical complications of this paper.

Remark 2.7 (*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 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* (e.g., see [Tao15, Gra95]). These can be thought of 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., that if i and j are non-neighbors 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 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 pseudo-distribution is positive semidefinite. Recall that this is the $\binom{n}{\leqslant d/2} \times \binom{n}{\leqslant d/2}$ 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

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

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 different 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 $\mathcal{M}$ 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 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. [HKP15] show that this construction satisfies positivity for $\omega \lessapprox n^{1/(\frac{d}{2}+1)}$.

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 $F = F_0 + F_1 + \cdots + F_d$ 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 diagonals blocks must be "small enough" to be charged to the on diagonal block. Thus the main question before us is 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.

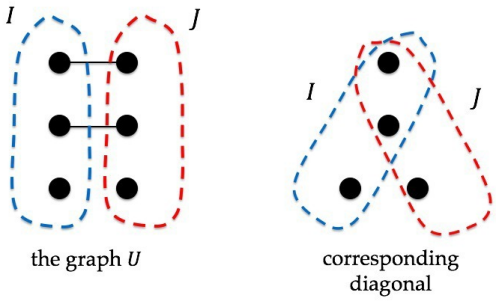
Definition 3.1 (Graphical Matrices (see Def 7.6 for a formal version))**.** Let U be a graph on $[2d]$ with specially identified subsets left and right subsets $[d]$ and $[2d] \setminus [d]$. For any $I, J \in \binom{[n]}{d}$, $I \cap J = \varnothing$, let $\pi_{I,J}$ be an injective map that takes $[d]$ 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 index 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.

[5]The 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.

In particular, this implies that when U has a perfect matching, M_U is pseudorandom 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 thought of as obtained by "collapsing" each matching edge into a vertex in U. In particular, this collapsing produces a matrix that lies in the decomposition of F_t.



There are a two main takeaways from this analysis that would serve as inspiration in the analysis of our actual construction. First is the decomposition into graphical matrices in order to have a coarse handle on the spectrum of the moment matrix. Second, the "charging" of negative eigenvalues against appropriate "diagonals" 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 Def 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 warmup). 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 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 digonal mass" that needs to be charged to it together and build an approximately PSD 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 $\mathsf{left-sep}(T) = S_\ell$ to be the q-sized separator that is closest to I and the *rightmost* separator $\mathsf{right-sep}(T) = S_r$ 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.

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

We can then partition T into three subsets $\mathcal{R}_\ell$, $\mathcal{R}_m$ 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 6.4). We thus immediately obtain that

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

Thus:

$$\mathcal{M}(I,J) = \sum_{1 \leqslant q \leqslant |I|,|J|} \sum_{S_\ell, S_R : |S_\ell| = |S_r| = q} \sum_{\substack{T \subseteq \binom{[n]}{2} \\ |\mathcal{V}(T) \cup I \cup J| \leqslant \tau \\ \mathsf{left-sep}(T) = S_\ell \\ \mathsf{right-sep}(T) = S_r}} \left(\left(\frac{\omega}{n}\right)^{|\mathcal{V}(\mathcal{R}_\ell)|} \chi_{\mathcal{R}_\ell}(G)\right)\left(\left(\frac{\omega}{n}\right)^{|\mathcal{V}(\mathcal{R}_m)|-2q} \chi_{\mathcal{R}_m}(G)\right)\left(\left(\frac{\omega}{n}\right)^{|\mathcal{V}(\mathcal{R}_r)|} \chi_{\mathcal{R}_r}(G)\right) \tag{3.3}$$

One could hope that we could replace the RHS of (3.3) by

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

In fact, it turns out we can focus attention (up to sufficiently small error in the spectral norm) to the case $\tau_1 \leqslant \tau/3$, $\tau_2 \leqslant \tau/3$, $\tau_3 \leqslant \tau/3$ 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 $|I| \leqslant d$ 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 $\mathcal{R}_\ell$ 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 $\mathcal{R}_\ell$, $\mathcal{R}_m$ 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 $\mathcal{R}_\ell$, $\mathcal{R}_m$ 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

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

The error matrices $\xi_0, \xi_1, \ldots, \xi_{2d}$ 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 $\mathcal{Q}_1, \ldots \mathcal{Q}_{2d-1}$ 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 [MP] showed that

the spectral norm of 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) tradeoff 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 Q_i for $i > 1$ as $-\frac{D}{8d} \succeq Q_i \succeq \frac{D}{8d}$, and this completes our proof.

4 Preliminaries

4.1 General Notation

- We use small Greek letters indicate constants/parameters.
- $\mathcal{P}^n_d$ 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 $C_q = C_q(G) = \{I \subseteq [n] \,:\, I \text{ is a } q\text{-clique in } G\}$, and let $C_{\leqslant q} = \bigcup_{q' \leqslant q} C_{d'}$. Let $C(G) = C_{\leqslant \infty}$ 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 (w.h.p.)* if $\mathbb{P}[E] \geqslant 1 - \Omega(1)/n^{10 \log n}$ 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 $n \geqslant n_0$, $f(n) \leqslant Cg(n)$.

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 $e \in [n] \times [n]$ is an edge (indicated by $G_e = +1$) in the graph G or not. 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 $G^1, G^2, \ldots, G^k$ on the vertex set $[n]$, we define $\Delta(G^1, G^2, \ldots, G^k)$ to be the graph G satisfying $G_e = \Pi_{i \leqslant k} G^i_e$.

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 $f : G \to \mathbb{R}$ 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:

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

The parity function χ_W are 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 identity:*

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

4.4 The Sum-of-Squares Algorithm

The sum of squares 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 $\tilde{\mathbb{E}} : \mathcal{P}^n_d \to \mathbb{R}$ is said to be a *degree d-pseudoexpectation* if it satisfies:

1. Normalization: $\tilde{\mathbb{E}}[\mathbf{1}] = 1$.

2. Positive Semidefiniteness: $\tilde{\mathbb{E}}[p^2] \geqslant 0$ for every polynomial $p \in \mathcal{P}^n_d$.

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

Given a set of constraints $\{p_i = 0\}$ for $1 \leqslant i \leqslant m$ and an objective polynomial p, degre sum of squares 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 $1 \leqslant i \leqslant m$.

5 The Pseudo-expectation

We now define our pseudo-distribution 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 $Cd/\varepsilon \leqslant \tau \leqslant (\varepsilon/C)\log n$ and $\varepsilon \geqslant C \log\log n/\log n$ 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 $|I| \leqslant d$. $\tilde{\mathbb{E}}[x_I]$ 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-size ω, truncation τ)**.** Let $S \subseteq [n]$ be a set of vertices of size $|S| \leqslant d$. Let $T \subseteq \binom{[n]}{2}$ be a set of edges. Let $\chi_T = \prod_{e \in T} G_e$. Let

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

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*

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

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

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

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

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

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

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

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 $f_G(x) = \sum_{|S|\leqslant 2d} c_S(G) \cdot x_S$ 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 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 $\tilde{\mathbb{E}}[x_S] = 0$.

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 $\mathbf{1}_{S \text{ is a clique in } G} = 2^{-\binom{|S|}{2}} \sum_{T\subseteq\binom{S}{2}} \chi_T$. Becasue the Fourier expansion of $\tilde{\mathbb{E}}[x_S]$ is truncated using the threshold $|\mathcal{V}(T) \cup S| \leqslant \tau$, 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) PSD. That is:

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

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

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

Proof of Theorem 1.1. By Lemma 5.4, Lemma 5.5, and Lemma 5.6, there is a universal C so that if $Cd/\varepsilon \leqslant \tau \leqslant (1/C)\varepsilon \log n$, (by a union bound) with high probability the following all 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. $\tilde{\mathbb{E}}[\sum_i x_i] \geqslant (1 - n^{-\Omega(\varepsilon)})\omega$.
4. $\tilde{\mathbb{E}}[p(x)^2] \geqslant 0$ 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 pseudo-distribution with $\tilde{\mathbb{E}}[\sum_i x_i] \geqslant \Omega(n^{1/2-\Theta(d/\log n)^{1/2}})$ 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 $\mathcal{M} \in \mathbb{R}^{\binom{[n]}{\leqslant d} \times \binom{[n]}{\leqslant d}}$ be given by $\mathcal{M}(I, J) = \tilde{\mathbb{E}}[x_I x_J]$.

Thus, Lemma 5.6 is equivalent to showing:

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 $\mathcal{M} = \mathcal{L} Q_0 \mathcal{L}^\dagger +$"*error*" for appropriately defined matrices $\mathcal{L}$ and Q_0. This step is the key technical part of the proof - given such a factorization, our task reduces to showing that Q_0 and $\mathcal{L}\mathcal{L}^\dagger$ has large enough positive eigenvalues to compensate for the error. The first approximate factorization step will occupy us in Section 6. The technical work in second step involves showing upper bounds on the spectral norms of appropriately defined pieces of 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.

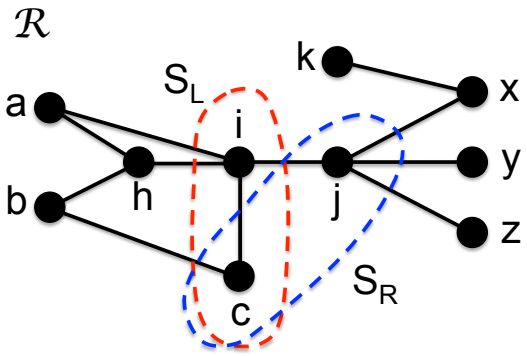
Ultimately, we will want to partition a ribbon into three subribbons in such a way that we can express the moment matrix as the sum of positive semidefinite matrices, and some error terms. Our partitioning will be based on minimum vertex separators.

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 *rightmost* vertex separator of minimum size. We defer its proof to Appendix A.3.

Lemma 6.3 (Leftmost/Rightmost 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 Q 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 illustrate the notion of a leftmost and rightmost vertex separator in the example below.



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 $\mathcal{M}(I, J)$.

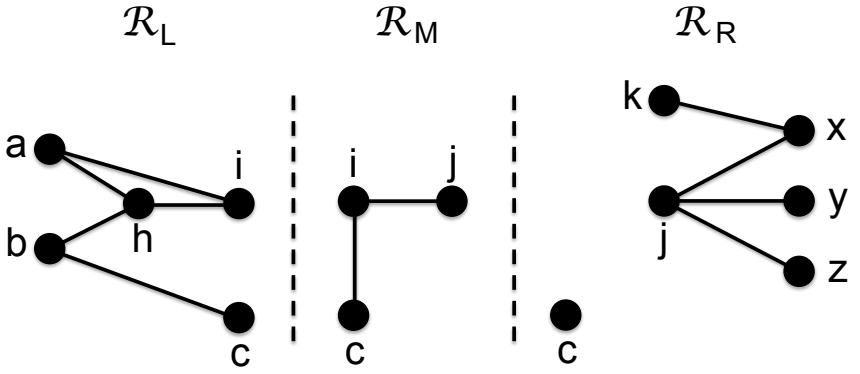
Definition 6.4 (Canonical Factorization)**.** Let $\mathcal{R}$ be an (I, J)-ribbon with edge set $W_{\mathcal{R}}$ and vertex set $V_{\mathcal{R}}$. Let V_ℓ be the vertices reachable from I without passing through $S_L(\mathcal{R})$, and similarly for V_r, and let $V_m = V_{\mathcal{R}} \setminus (V_\ell \cup V_r)$. Let $W_\ell \subseteq W_{\mathcal{R}}$ be 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 $\mathcal{R}_\ell$ 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 $\mathcal{R}_\ell$ anyways. This is why we can say that $\mathcal{R}_\ell$ is an $(I, S_L(\mathcal{R}))$-ribbon. The following illustrates what the canonical factorization would look like in our earlier example:



We chose this example to illustrate a subtle point. The edge (i, c) has both its endpoints in both $\mathcal{R}_\ell$ 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 $\mathcal{R}_\ell$ or within S_R in $\mathcal{R}_m$. Finally, note that there can be isolated vertices in $\mathcal{R}_\ell$ or $\mathcal{R}_r$ but such vertices need to be in I or J respectively.

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 6.5. 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. □

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 6.6. *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-size vertex separator for $\mathcal{R}_0$, then it is also not a minimum-size 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-size separator, it must also be the right-most minimum-size 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 their 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. □

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 6.5, for any $I, J \subseteq [n]$ of size at most d we can write

$$\mathcal{M}(I,J) = \sum_{\substack{\mathcal{R} \text{ an } (I,J)\text{-ribbon with edge set } W,\\ |\mathcal{V}(W)|\leqslant\tau \\ \text{canonical factorization } (\mathcal{R}_\ell,\mathcal{R}_m,\mathcal{R}_r)}} \left(\frac{\omega}{n}\right)^{|\mathcal{V}(\mathcal{R})|} \cdot \chi_{\mathcal{R}_\ell} \cdot \chi_{\mathcal{R}_m} \cdot \chi_{\mathcal{R}_r}$$

$$= \sum_{\substack{S_\ell,S_r\subseteq[n]\\ |S_\ell|=|S_r|\leqslant d}} \left(\frac{\omega}{n}\right)^{-\frac{|S_\ell|+|S_r|}{2}} \sum_{\substack{\mathcal{R}_\ell,\mathcal{R}_m,\mathcal{R}_r\subseteq\binom{[n]}{2} \\ \text{satisfying } S_\ell,S_r \text{ factorization conditions} \\ \text{and } |\mathcal{V}(\mathcal{R}_\ell)\cup\mathcal{V}(\mathcal{R}_m)\cup\mathcal{V}(\mathcal{R}_r)|\leqslant\tau}} \left(\frac{\omega}{n}\right)^{|\mathcal{V}(\mathcal{R}_\ell)|+|\mathcal{V}(\mathcal{R}_m)|+|\mathcal{V}(\mathcal{R}_r)|-\frac{|S_\ell|+|S_r|}{2}} \cdot \chi_{\mathcal{R}_\ell} \cdot \chi_{\mathcal{R}_m} \cdot \chi_{\mathcal{R}_r}$$

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

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

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

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

6.4 Factorization of the Matrix $\mathcal{M}$

In lines 6.2 and 6.3 we have defined two error matrices, $\xi_0, E_0 \in \mathbb{R}^{\binom{[n]}{\leqslant d} \times \binom{[n]}{\leqslant d}}$. Inspired by the factorization of $\mathcal{M}(I,J)$ in line 6.1, we define another pair of matrices as follows:

$$Q_0 \in \mathbb{R}^{\binom{[n]}{d} \times \binom{[n]}{d}} \quad \text{given by} \quad Q_0(S_\ell, S_r) = \sum_{\substack{\mathcal{R}_m \text{ having } 3 \\ |\mathcal{V}(\mathcal{R}_m)| \leqslant \tau}} \left(\frac{\omega}{n}\right)^{|\mathcal{V}(\mathcal{R}_m)| - \frac{|S_\ell| + |S_r|}{2}} \chi_{\mathcal{R}_m}$$

$$\mathcal{L} \in \mathbb{R}^{\binom{[n]}{d} \times \binom{[n]}{d}} \quad \text{given by} \quad \mathcal{L}(I,S) = \left(\frac{\omega}{n}\right)^{-\frac{|S|}{2}} \sum_{\substack{\mathcal{R}_\ell \text{ having } 1 \\ |\mathcal{V}(\mathcal{R}_\ell)| \leqslant \tau}} \left(\frac{\omega}{n}\right)^{|\mathcal{V}(\mathcal{R}_\ell)|} \chi_{\mathcal{R}_\ell} .$$

The powers of (ω/n) are split between Q_0 and $\mathcal{L}$ so that the typical of eigenvalue of Q_0 will be approximately 1 (although it will be some time before we are prepared to prove that).

The equation in lines 6.1, 6.2, and 6.3 can be written succinctly as

$$\mathcal{M} = \mathcal{L} Q_0 \mathcal{L}^\dagger - \xi_0 - E_0 .$$

As we will see later, with high probability $Q_0 \succeq 0$, and thus also $\mathcal{L} Q_0 \mathcal{L}^\dagger \succeq 0$. 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 Q_0:

Lemma 6.7. *Let Π be the projector to $\mathsf{Span}\{e_C : C \in C_{\leqslant d}\}$. Then $Q_0 = \Pi Q_0 = Q_0 \Pi$.*

Proof. Suppose S is not a clique in G. We need to show that the row $Q_0(S, \cdot)$ is zero. For every entry $Q_0(S, S')$, notice that the Fourier coefficients $\widehat{Q_0(S,S')}(T) = \widehat{Q_0(S,S')}(T')$ 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 $Q_0(S,S') = \mathbf{1}_{S \text{ is a clique in } G} \cdot f_{S,S'}(G)$ for some function $f_{S,S'}$. □

6.5 Iterative Factorization of E_0

We recall now the definition of the matrix $E_0 \in \mathbb{R}^{\binom{[n]}{\leqslant d}\times\binom{[n]}{\leqslant d}}$.

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

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 6.8 (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 $\mathcal{Z}(\cdot) = \varnothing$, 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 (S_ℓ, S_r)-ribbon.

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

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

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 $E_0 = \mathcal{E}_{c_0}$. 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 $\mathcal{Q}_{c'}, \xi_c \in \mathbb{R}^{\binom{[n]}{\leqslant d}\times\binom{[n]}{\leqslant d}}$ 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 4, there must be some vertices occurring in more than one of $\mathcal{V}(\mathcal{R}_\ell), \mathcal{V}(\mathcal{R}_m), \mathcal{V}(\mathcal{R}_\ell)$. 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 6.9 (Separating Factorization)**.** Let $\mathcal{R}_\ell, \mathcal{R}_m, \mathcal{R}_r$ be ribbons satisfying S_ℓ, S_r factorization conditions 1, 3*, 2 but not 4, with $|\mathcal{V}(\mathcal{R}_\ell)|, |\mathcal{V}(\mathcal{R}_m)|, |\mathcal{V}(\mathcal{R}_r)| \leqslant \tau$. 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-size 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-size 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 $\mathcal{R}'_\ell$ 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

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

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

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

and have that

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

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 $|\mathcal{V}(\mathcal{R}'_\ell)|, |\mathcal{V}(\mathcal{R}'_m)|, |\mathcal{V}(\mathcal{R}'_r)| \leqslant \tau$. We define

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

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 $\mathcal{L}^\dagger \mathcal{Q}_{c'} \mathcal{L}^\dagger - \mathcal{E}_{c'}$ for the correct choice of c'.

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

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

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

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

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 $\gamma_{\mathcal{R}'_\ell,\mathcal{R}'_m,\mathcal{R}'_r,I,J,S'_\ell,S'_r}$. It is useful to define:

Definition 6.10. 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 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 $\mathcal{V}(\mathcal{R}_\ell) \cup \mathcal{V}(\mathcal{R}_m) \cup \mathcal{V}(\mathcal{R}_\ell)$ 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 multiplicity higher than they should 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

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

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,

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

Note that for this to happen, the right hand side must be independent of $\mathcal{R}'_\ell$ 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 6.11. Let $\mathcal{R}'_\ell, \mathcal{R}'_m, \mathcal{R}'_r$ satisfy 1, 3*, 2, 4 for some $S'_\ell, S'_r \subseteq [n]$. Let $\mathcal{R}''_\ell$ and $\mathcal{R}''_r$ also satisfy 1 and 2, respectively, for S'_ℓ, S'_r, respectively. Then

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

(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 $\mathcal{R}_\ell$ between I and S'_ℓ must be $\mathcal{R}'_\ell$ while the part of $\mathcal{R}_\ell$ between S'_ℓ and S_ℓ becomes part of $\mathcal{R}'_m$. To shift from $\mathcal{R}'_\ell$ to $\mathcal{R}''_\ell$, we simply replace $\mathcal{R}'_\ell$ by $\mathcal{R}''_\ell$ within $\mathcal{R}_\ell$. 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 $\mathcal{R}'_\ell$ to $\mathcal{R}''_\ell$ 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 $\mathcal{R}_\ell$ 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 $\mathcal{R}_\ell$ but whether or not they do so is unaffected by the shift 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 $\mathcal{R}''_r$ □

Remark 6.12. 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 a situation where there is a vertex v which appears in $\mathcal{R}_\ell$ and $\mathcal{R}_m$ but disappears from $\mathcal{R}'_m$ and is not in S'_ℓ. Since v is no longer in $\mathcal{R}'_m$, $\mathcal{R}''_\ell$ could contain v. If so, then we cannot shift from $\mathcal{R}'_\ell$ to $\mathcal{R}''_\ell$ 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 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 6.13. *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*

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

Proof. We claim that $|S_\ell| + |S_r| + 1 \leqslant |S'_\ell| + |S'_r|$ 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 $\mathcal{R}_\ell$ and $\mathcal{R}_\ell$ 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, so the difference must be at least 1 and we are done. □

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 $E_0 = \mathcal{E}_{c_0}$. 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 6.14.

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

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 appliying 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 6.13).

However, the size of these separators is always at most d, 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. □

7 $\mathcal{M}$ is PSD

In this section we combine the factorization of $\mathcal{M}$ in terms of the matrices $\mathcal{L}, Q_i, \xi_i$ that we obtained in Section 6 with estimates on the eigenvalues of the Qs and ξs. The starting point is the following PSDness claim for Q_0.

Lemma 7.1. *Let $D \in \mathbb{R}^{\binom{[n]}{\leqslant d}\times\binom{[n]}{\leqslant d}}$ be the diagonal matrix with $D(S,S) = 2^{\binom{|S|}{2}}/4$ if S is a clique in G and 0 otherwise. With high probability, $Q_0 \succeq D$.*

We also need to bound $\|Q_i\|$ for $i > 0$.

Lemma 7.2. *Let $D \in \mathbb{R}^{\binom{[n]}{\leqslant d}\times\binom{[n]}{\leqslant d}}$ 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 Q_i for $i \in [1, 2d]$ satisfies*

$$\frac{-D}{8d} \preceq Q_i \preceq \frac{D}{8d}\,.$$

The preceding lemmas are enough to obtain $Q_0 - \ldots + Q_{2d} \succeq D/2$, but in the end we need to work with the matrix $\mathcal{L}(Q_0 - \ldots + Q_{2d})\,\mathcal{L}^\dagger - (\xi_0 - \ldots + \xi_{2d})$. 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 $\mathsf{Span}\{e_C : C \in C_{\leqslant d}\}$.*

Finally, we need a bound on the ξ matrices.

Lemma 7.4. *With high probability, $\|\xi_0 - \ldots + \xi_{2d}\| \leqslant n^{-16d}$.*

We can now prove Lemma 5.8.

Proof of Lemma 5.8. By Claim 6.14,

$$\mathcal{M} = \mathcal{L}(Q_0 - Q_1 + Q_2 - \ldots - Q_{2d-1} + Q_{2d})\,\mathcal{L}^\dagger - (\xi_0 - \xi_1 + \xi_2 - \ldots - \xi_{2d-1} + \xi_{2d})\,.$$

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

$$Q_0 - Q_1 + Q_2 - \ldots - Q_{2d-1} + Q_{2d} \succeq \frac{D}{2} \succeq \frac{\Pi}{2}\,.$$

where as usual Π is the projector to $\mathsf{Span} e_C : C \in C_{\leqslant d}$. Thus by Lemma 7.3, we obtain $\mathcal{L}(Q_0 - \ldots + Q_{2d})\,\mathcal{L}^\dagger \succeq \Omega(\omega/n)^{d+1} \cdot \Pi$. Finally, by Lemma 7.4 we have

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

In the next subsections, we prove the foregoing lemmas.

7.1 Ribbons and Spectral Norms

Our PSDness arguments 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 as *graphical matrices*, that are are defined to capture the matrices are invariant under permutation of vertices in the graph G and are in fact "minimal" such matrices.

We first identify the *shape* of a ribbon that basically identifies the structure of a ribbon up to renaming.

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 $\mathsf{shape}(\mathcal{R}) = U$.

We record some observations on shapes of ribbons.

- If $\mathcal{R}$ is a ribbon (not an improper ribbon), its shape satisfies the assumptions of Lemma 7.8 (namely, 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 (I', J')-ribbon $\mathcal{R}'$ only has the same shape as $\mathcal{R}$ if $|I' \cap J'| = 1$ and contains only the least element in I and J. 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 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

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

Example 7.7. When U is a graph on 2 vertices with distinguished sets $\{1\}$ and $\{2\}$ of size 1 each and a single edge connecting vertex 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 [MP] 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 $t \leqslant O(\log n)$ vertices, with two distinguished subsets of vertices A and B, and suppose:*

- *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, whp, $\|M\| \leqslant n^{\frac{t-p-r}{2}} \cdot 2^{O(t)} \cdot (\log n)^{O(t-r+p)}$.

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 adjacency matrix is a graphical matrix with a shape U on 2 vertices with a single edge connecting them, thus, $t = 2$ and $r = 1$. Lemma 7.8 thus shows an upper bound of $\sqrt{n}\,\mathrm{poly}\log(n)$ on the spectral norm of the adjacency matrix which is tight up to a $\mathrm{poly}\log(n)$ factor.

7.2 PSDness for Q_0—Proof of Lemma 7.1

In this section we prove Lemma 7.1, which we restate here.

Lemma (Restatement of Lemma 7.1)**.** *Let $D \in \mathbb{R}^{\binom{[n]}{\leqslant d} \times \binom{[n]}{\leqslant d}}$ be the diagonal matrix with $D(S,S) = 2^{\binom{|S|}{2}}/4$ if S is a clique in G and 0 otherwise. With high probability, $Q_0 \succeq D$.*

Proof of Lemma 7.1. To begin, we split Q_0 into its diagonal Q_0^{diag} and its off-diagonal $Q_0^{\text{off-diag}}$ parts.

$$Q_0^{\text{diag}}(S_\ell, S_r) = \begin{cases} Q_0(S_\ell, S_r) \text{ if } S_\ell = S_r \\ 0 \text{ otherwise.} \end{cases} \qquad Q_0^{\text{off-diag}}(S_\ell, S_r) = \begin{cases} Q_0(S_\ell, S_r) \text{ if } S_\ell \neq S_r \\ 0 \text{ otherwise.} \end{cases}$$

Then $Q_0 = Q_0^{\text{diag}} + Q_0^{\text{off-diag}}$. Expanding Q_0^{diag},

$$Q_0^{\text{diag}}(S,S) = 2^{\binom{|S|}{2}} \cdot \mathbf{1}_{S \text{ is a clique}} \cdot \left(1 + \sum_{\substack{\mathcal{R} \text{ nonempty, having 3} \\ \text{and no edges inside } S \\ |S| < |\mathcal{V}(\mathcal{R})| \leqslant \tau}} \left(\frac{\omega}{n}\right)^{|\mathcal{V}(\mathcal{R})| - |S|} \cdot \chi_{\mathcal{R}} \right) = 2^{\binom{|S|}{2}} \cdot \mathbf{1}_{S \text{ is a clique}} \cdot (1 \pm n^{-\Omega(\varepsilon)})$$

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

Next, we bound $\|Q_0^{\text{off-diag}}\|$ be decomposing it according to ribbon shape. Fix $s, t \leqslant \tau$. Let $U_1^{(s,t)}, \ldots, U_q^{(s,t)}$ be all the graphs on vertex set $[t]$ with two distinguished sets of vertices A, B, both of

size s, with $|A \cap B| \leqslant s-1$, 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

$$Q_0^{\text{off-diag}} = \sum_{\substack{s \leqslant d \\ t \leqslant \tau \\ i \leqslant q}} \left(\frac{\omega}{n}\right)^{t-s} \cdot M_i^{(s,t)} \,.$$

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

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

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

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

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

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

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

Notice that the above matrix is exactly the block of $Q_0^{\text{off-diag}}$ corresponding to subsets of size s. Together with our bound on Q_0^{diag}, this proves the lemma. □

7.3 Norm Bounds for Q_i—Proof of Lemma 7.2

In this section we prove Lemma 7.2, restated here.

Lemma (Restatement of Lemma 7.2). *Let $D \in \mathbb{R}^{\binom{[n]}{\leqslant d} \times \binom{[n]}{\leqslant d}}$ 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 Q_i for $i \in [1, 2d]$ satisfies*

$$\frac{-D}{8d} \preceq Q_i \preceq \frac{D}{8d} \,.$$

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

Lemma 7.10. *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 $i \leqslant 2d$, writing $s = \frac{|S_\ell|+|S_r|}{2}$,*

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

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 $0 < i \leqslant 2d$. We will use Lemma 7.8, which requires that we first decompose each Q_i into simpler matrices. First of all, 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 7.11. $\tilde{c}_i(\mathcal{R}_m) \leqslant 2(\omega/n)^s \cdot n^{\frac{p-i/2}{2}+\varepsilon s}$, 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 7.10, $(\omega/n)^k c_i(\mathcal{R}'_m) \leqslant \left(\frac{\omega}{n}\right)^{k+s} \cdot n^{\frac{p-k-i/2}{2}+\varepsilon s}$. 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 $k \geqslant 0$ gives the result. □

Now fix $s_\ell, s_r \leqslant d$ and $t \leqslant \tau$ 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 $1 - O(n^{-100\log n})$,

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

$$= n^{-\varepsilon(t-s)-i/4} \cdot 2^{O(t)} \cdot (\log n)^{O(t-r+\tilde{p})}$$
$$= n^{-\varepsilon(t-s)-i/4} \cdot 2^{O(t)} \cdot (\log n)^{O(t-s)},$$

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

By inspection,

$$Q_i = \sum_{\substack{s_\ell, s_r \leqslant d \\ t \leqslant \tau \\ j \leqslant q}} M_j^{(s_\ell, s_r, t)}.$$

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

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

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

$$\begin{aligned}
|v^\dagger Q_i v| &\leqslant \sum_{k_1=0}^{d} \sum_{k_2=0}^{d} \|v_{k_1}\| \left\| \sum_{\substack{t \leqslant \tau \\ j \leqslant q}} M_j^{(k_1, k_2, t)} \right\| \|v_{k_2}\| \\
&\leqslant \sum_{k_1=0}^{d} \sum_{k_2=0}^{d} \frac{1}{100d^3} \left(2^{\binom{k_1}{2}/2+\binom{k_2}{2}/2} \|v_{k_1}\| \|v_{k_2}\| \right) \\
&\leqslant \sum_{k_1=0}^{d} \sum_{k_2=0}^{d} \frac{1}{200d^3} \left(2^{\binom{k_1}{2}} \|v_{k_1}\|^2 + 2^{\binom{k_2}{2}} \|v_{k_2}\|^2 \right) \\
&\leqslant \sum_{k=0}^{d} \frac{2^{\binom{k}{2}}}{100d^2} \|v_k\|^2 \leqslant v^\dagger (D/8d) v
\end{aligned}$$

□

7.3.1 Coefficient Decay in the Factorization: Proof of Lemma 7.10

We turn to the proof of Lemma 7.10, 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 7.12.

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 factorization 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. $r = (|\mathcal{V}(\mathcal{R}_\ell)| + |\mathcal{V}(\mathcal{R}_m)| + |\mathcal{V}(\mathcal{R}_\ell)| - |S_\ell| - |S_r|) - (|\mathcal{V}(\mathcal{R}'_\ell)| + |\mathcal{V}(\mathcal{R}'_m)| + |\mathcal{V}(\mathcal{R}'_\ell)| - |S'_\ell| - |S'_r|)$, the number of intersections among R_ℓ, R_m, 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 7.12.

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

The following series of claims will help us in the proof of Lemma 7.12

Claim 7.13. $I \cap \mathcal{V}(\mathcal{R}'_m) \subseteq S'_\ell$ 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. □

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

Claim 7.14. $\mathfrak{D} \subseteq \mathfrak{U}$.

Proof of claim. Let $u \in \mathfrak{D}$. If $u \in S_\ell$ or $u \in S_r$ 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 factorization.

If neither of these cases hold, 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. □

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

Claim 7.15. By Menger's theorem, there are $|S'_\ell|$ vertex-disjoint paths from $\mathfrak{U} \cap \mathcal{V}(\mathcal{R}_\ell)$ to I in $\mathcal{R}_\ell$. Let $u_\ell^{(1)}, \ldots, u_\ell^{(|S'_\ell|)}$ be distinct vertices so that $u^{(i)}$ is the last vertex in $\mathfrak{U}$ along the i-th vertex disjoint path. Let $u_r^{(1)}, \ldots, u_r^{(|S'_r|)}$ 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 $u \notin \mathfrak{D}$. □

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 7.16. 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. By inspection. □

We are prepared to prove Lemma 7.12.

Proof of Lemma 7.12. We start by bounding the number of vertices in $\mathfrak{U} \setminus \mathfrak{D}$. By Claim 7.15, there are at least $|\{u_\ell^{(1)}, \ldots, u_\ell^{(|S'_\ell|)}, u_r^{(1)}, \ldots, u_r^{|S'_r|}\}|$ 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 $w_1, \ldots, w_a$ 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 $\mathcal{U}$, ending at $u_\ell^{(i)}$, 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

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

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), 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 $q_{\ell,m,r}$, $S_l \cap \mathcal{V}(\mathcal{R}_r)$, or $S_r \cap \mathcal{V}(\mathcal{R}_l)$ we are done, so suppose otherwise. If it is contained entirely in $q_{\ell,m} \cup q_{r,m} \cup (S_\ell \setminus \mathcal{V}(\mathcal{R}_r)) \cup (S_r \setminus \mathcal{V}(\mathcal{R}_l))$, 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's time to put things together. By Claim 7.14, we can bound $|\mathfrak{D}|$ by

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

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

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

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

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

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

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

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

Now we can prove Lemma 7.10.

Proof of Lemma 7.10. First of all, we note that $c_i(\mathcal{R}_m)$ depends only on the shape of $\mathcal{R}_m$ by 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 $\mathcal{R}_m$ which have $\mathcal{Z}(\mathcal{R}_m) = \varnothing$ 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 $c_0(\mathcal{R}_m) \leqslant 1$, 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 1 and 2,

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

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 7.12,

$$\begin{aligned}
\left(\frac{\omega}{n}\right)^r \cdot c_i(\mathcal{R}_m) &\leqslant \left(\frac{\omega}{n}\right)^r \cdot \left(\frac{\omega}{n}\right)^s \cdot n^{\frac{p-|\mathcal{Z}(\mathcal{R}_m)|-i/2}{2}+\varepsilon s} \quad \text{by induction} \\
&= \left(\frac{\omega}{n}\right)^{s'} \cdot \left(\frac{\omega}{n}\right)^{r-s'+s} \cdot \cdot n^{\frac{p-|\mathcal{Z}(\mathcal{R}_m)|-i/2}{2}+\varepsilon s} \\
&= \left(\frac{\omega}{n}\right)^{s'} \cdot n^{-\varepsilon(r-s'+s)} \cdot n^{-\frac{1}{2}(r-s'+s)} \cdot n^{\frac{p-|\mathcal{Z}(\mathcal{R}_m)|-i/2}{2}+\varepsilon s} \quad \text{using } \omega = n^{1/2-\varepsilon} \\
&\leqslant \left(\frac{\omega}{n}\right)^{s'} \cdot n^{-\varepsilon(r-s'+s)} \cdot n^{-\frac{1}{2}(s'-s+p-p'+|\mathfrak{D}|)} \cdot n^{\frac{p-|\mathcal{Z}(\mathcal{R}_m)|-i/2}{2}+\varepsilon s} \quad \text{by Lemma 7.12} \\
&= \left(\frac{\omega}{n}\right)^{s'} \cdot n^{-\varepsilon(r-s'+s)} \cdot n^{\frac{p'-|\mathcal{Z}(\mathcal{R}'_m)|-i/2-s'+s}{2}+\varepsilon s} \quad \text{canceling terms, using } |\mathcal{Z}(\mathcal{R}'_m)| = |\mathfrak{D}| + |\mathcal{Z}(\mathcal{R}_m)| \\
&= n^{-\varepsilon r} \cdot \left(\frac{\omega}{n}\right)^{s'} \cdot n^{\frac{p'-|\mathcal{Z}(\mathcal{R}'_m)|-i/2-(s'-s)}{2}+\varepsilon s'} \\
&\leqslant n^{-\varepsilon r} \cdot \left(\frac{\omega}{n}\right)^{s'} \cdot n^{\frac{p'-|\mathcal{Z}(\mathcal{R}'_m)|-(i+1)/2}{2}+\varepsilon s'} \quad \text{using } s'-s \geqslant 1/2, \text{ by Lemma 6.13}
\end{aligned}$$

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 7 possibilities for which ribbon(s) it came from in $\{\mathcal{R}_\ell, \mathcal{R}_m, \mathcal{R}_r\}$ so there are at most 7^τ choices overall (recall that $\mathcal{R}'_m$ has at most τ vertices for the terms we are looking at). Once we have chosen which ribbon(s) each vertex of $\mathcal{R}'_m$ came from, everything is fixed except for possible edges of $\mathcal{R}'_m$ which appear at least twice in $\mathcal{R}_\ell$, $\mathcal{R}_m$, and $\mathcal{R}_r$. There are two possibilities for each possible edge of $\mathcal{R}'_m$ which appears twice in $\mathcal{R}_\ell$, $\mathcal{R}_m$, and $\mathcal{R}_r$ and four possibilities for each possible edge of $\mathcal{R}'_m$ which appers three times in $\mathcal{R}_\ell$, $\mathcal{R}_m$, 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 $\mathcal{R}_\ell$, $\mathcal{R}_m$, 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 $\tau \leqslant (\varepsilon/C)\log n$ for some universal constant C, we have $2^{O(r\tau)} \cdot n^{-\varepsilon r} \ll 1/\tau$ for all $r \geqslant 1$. All in all, we obtain

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

which completes the induction. □

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 (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* $\mathsf{Span}\{e_C : C \in C_{\leqslant d}\}$*.*

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

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

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 $s_\ell, s_r \leqslant d$. For a fixed $t \leqslant \tau$, 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 size 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-size 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 $1 - O(n^{-100\log n})$,

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

where in the last step we have used that $t \geqslant s_\ell + s_r - r$ and $s_r \leqslant s_\ell$, 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 triangle inequality

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

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

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

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 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. □

7.5 High-Degree Matrices Have Small Norms

In this section we prove Lemma 7.4, restated here:

Lemma (Restatement of Lemma 7.4)**.** *With high probability,* $\|\xi_0 - \ldots + \xi_{2d}\| \leqslant n^{-16d}$.

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

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

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

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

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

We will actually prove a bound on the Frobenious 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 A.1, 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 7.17. *Suppose c_T are a collection of coefficients, one for each $T \subseteq \binom{[n]}{2}$, and there is a constant C such that*

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

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

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

Lemma 7.18. *Every triple $\mathcal{R}_\ell, \mathcal{R}_m, \mathcal{R}_r$ appearing with 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 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

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

or

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

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

We are prepared to prove Lemma 7.4.

Proof of Lemma 7.4. We will apply Lemma 7.17 to $\xi_i(I,J)$ for each $i \leqslant 2d$ and $I, J \subseteq [n]$ with $|I|, |J| \leqslant d$. 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 7.18, 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

$$|c_T| \leqslant \sum_{\substack{S_\ell, S_r \subseteq [n] \\ |S_\ell|, |S_r| \leqslant d}} \left(\frac{\omega}{n}\right)^{-\frac{|S_\ell|+|S_r|}{2}} \sum_{\substack{\mathcal{R}_\ell, \mathcal{R}_m, \mathcal{R}_r \\ \text{nonzero in } \xi_i(I,J) \text{ as in } 7.18 \\ \chi_{\mathcal{R}_\ell}\cdot\chi_{\mathcal{R}_m}\cdot\chi_{\mathcal{R}_r} = \chi_T}} c_{i-1}(\mathcal{R}_m) \left(\frac{\omega}{n}\right)^{|\mathcal{V}(\mathcal{R}_\ell)|+|\mathcal{V}(\mathcal{R}_r)|+|\mathcal{V}(\mathcal{R}_m)|-\frac{|S_\ell|+|S_r|}{2}} \tag{7.3}$$

By Lemma 7.10, we have $c_{i-1}(\mathcal{R}_m) \leqslant n^d \leqslant (\omega/n)^{-2d}$. At the same time, there are at most $2^{O(\tau^2)}$ nonzero terms in the sum (7.3). Thus by Lemma 7.18 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 7.17, we obtain $|\xi_i(I,J)| \leqslant n^{-20d}$ with probability $1 - O(n^{-100\log n})$. Taking a union bound over all $n^{2d} \leqslant n^{2\log n}$ entries of ξ_i, and over all $i \leqslant 2d$, we obtain that $\|\xi_0 - \ldots + \xi_{2d}\| \leqslant \|\xi_0 - \ldots + \xi_{2d}\|_F \leqslant n^{-16d}$ with probability $1 - O(n^{-96\log n})$. □

Acknowledgements

We thank Raghu Meka, Ryan O'Donnell, Prasad Raghavendra, Tselil Schramm, David Steurer, and Avi Wigderson for many useful discussions related to this paper.

References

[AAK+07] Noga Alon, Alexandr Andoni, Tali Kaufman, Kevin Matulef, Ronitt Rubinfeld, and Ning Xie, *Testing k-wise and almost k-wise independence*, STOC, 2007, pp. 496–505.

[ABC13] Per Austrin, Mark Braverman, and Eden Chlamtac, *Inapproximability of np-complete variants of nash equilibrium*, Theory of Computing **9** (2013), 117–142.

[ABW10] Benny Applebaum, Boaz Barak, and Avi Wigderson, *Public-key cryptography from different assumptions*, STOC, 2010, pp. 171–180.

[AKS98] Noga Alon, Michael Krivelevich, and Benny Sudakov, *Finding a large hidden clique in a random graph*, SODA, 1998, pp. 594–598.

[BBH+12] Boaz Barak, Fernando GSL Brandao, Aram W Harrow, Jonathan Kelner, David Steurer, and Yuan Zhou, *Hypercontractivity, sum-of-squares proofs, and their applications*, Proceedings of the forty-fourth annual ACM symposium on Theory of computing, ACM, 2012, pp. 307–326.

[BCK15] Boaz Barak, Siu On Chan, and Pravesh K. Kothari, *Sum of squares lower bounds from pairwise independence*, Proceedings of the Forty-Seventh Annual ACM on Symposium on Theory of Computing, STOC 2015, Portland, OR, USA, June 14-17, 2015, 2015, pp. 97–106.

[BKS14] Boaz Barak, Jonathan A Kelner, and David Steurer, *Rounding sum-of-squares relaxations*, Proceedings of the 46th Annual ACM Symposium on Theory of Computing, ACM, 2014, pp. 31–40.

[BKS15] ———, *Dictionary learning and tensor decomposition via the sum-of-squares method*.

[BR13] Quentin Berthet and Philippe Rigollet, *Complexity theoretic lower bounds for sparse principal component detection*, COLT 2013 - The 26th Annual Conference on Learning Theory, June 12-14, 2013, Princeton University, NJ, USA, 2013, pp. 1046–1066.

[BS14] Boaz Barak and David Steurer, *Sum-of-squares proofs and the quest toward optimal algorithms*, Proceedings of the International Congress of Mathematicians (2014).

[BT06] Andrej Bogdanov and Luca Trevisan, *On worst-case to average-case reductions for NP problems*, SIAM J. Comput. **36** (2006), no. 4, 1119–1159.

[DBL10] *Computational complexity and information asymmetry in financial products (extended abstract)*, 2010.

[DM15] Yash Deshpande and Andrea Montanari, *Improved sum-of-squares lower bounds for hidden clique and hidden submatrix problems*, COLT (2015).

[FF93] Joan Feigenbaum and Lance Fortnow, *Random-self-reducibility of complete sets*, SIAM J. Comput. **22** (1993), no. 5, 994–1005.

[FK96] A. Frieze and R. Kannan, *The regularity lemma and approximation schemes for dense problems*, Foundations of Computer Science, 1996. Proceedings., 37th Annual Symposium on, Oct 1996, pp. 12–20.

[FK03a] Uriel Feige and Robert Krauthgamer, *The probable value of the lovász–schrijver relaxations for maximum independent set*, SIAM J. Comput. **32** (2003), no. 2, 345–370.

[FK03b] _____, *The probable value of the lovász–schrijver relaxations for maximum independent set*, SIAM J. Comput. **32** (2003), no. 2, 345–370.

[Gra95] Andrew Granville, *Harald cramér and the distribution of prime numbers*, Scandinavian Actuarial Journal **1995** (1995), no. 1, 12–28.

[Gri01] Dima Grigoriev, *Complexity of positivstellensatz proofs for the knapsack*, Computational Complexity **10** (2001), no. 2, 139–154.

[HK11] Elad Hazan and Robert Krauthgamer, *How hard is it to approximate the best nash equilibrium?*, SIAM J. Comput. **40** (2011), no. 1, 79–91.

[HKP15] Samuel B. Hopkins, Pravesh K. Kothari, and Aaron Potechin, *Sos and planted clique: Tight analysis of MPW moments at all degrees and an optimal lower bound at degree four*, CoRR **abs/1507.05230** (2015).

[HKP+16] Samuel B. Hopkins, Pravesh Kothari, Aaron Henry Potechin, Prasad Raghavendra, and Tselil Schramm, *On the integrality gap of degree-4 sum of squares for planted clique*, Proceedings of the Twenty-Seventh Annual ACM-SIAM Symposium on Discrete Algorithms, SODA 2016, Arlington, VA, USA, January 10-12, 2016, 2016, pp. 1079–1095.

[HWX15] Bruce E. Hajek, Yihong Wu, and Jiaming Xu, *Computational lower bounds for community detection on random graphs*, Proceedings of The 28th Conference on Learning Theory, COLT 2015, Paris, France, July 3-6, 2015, 2015, pp. 899–928.

[Jay57a] E. T. Jaynes, *Information theory and statistical mechanics. ii*, Phys. Rev. **108** (1957), 171–190.

[Jay57b] Edwin T Jaynes, *Information theory and statistical mechanics*, Physical review **106** (1957), no. 4, 620.

[Jer92] Mark Jerrum, *Large cliques elude the metropolis process*, Random Struct. Algorithms **3** (1992), no. 4, 347–360.

[JM15] Hamid Javadi and Andrea Montanari, *The hidden subgraph problem*, arXiv preprint arXiv:1511.05254 (2015).

[JP00] Ari Juels and Marcus Peinado, *Hiding cliques for cryptographic security*, Des. Codes Cryptography **20** (2000), no. 3, 269–280.

[Kar76] Richard M. Karp, *Probabilistic analysis of some combinatorial search problems*, Algorithms and Complexity: New Directions and Recent Results (1976).

[Kuc95] Ludek Kucera, *Expected complexity of graph partitioning problems*, Discrete Applied Mathematics **57** (1995), no. 2-3, 193–212.

[KZ14] Pascal Koiran and Anastasios Zouzias, *Hidden cliques and the certification of the restricted isometry property*, IEEE Trans. Information Theory **60** (2014), no. 8, 4999–5006.

[Las01] Jean B. Lasserre, *An explicit exact sdp relaxation for nonlinear 0-1 programs*, IPCO, 2001, pp. 293–303.

[MP] Dhruv Medarametla and Aaron Potechin, *Bounds on the norms of uniform low degree graph matrices*, Preprint.

[MPW15] Raghu Meka, Aaron Potechin, and Avi Wigderson, *Sum-of-squares lower bounds for planted clique*, 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), no. 5594, 824–827.

[MW15] Tengyu Ma and Avi Wigderson, *Sum-of-squares lower bounds for sparse pca*, Advances in Neural Information Processing Systems, 2015, pp. 1603–1611.

[Par00] Pablo A. Parrilo, *Structured semidefinite programs and semialgebraic geometry methods in robustness and optimization*, Ph.D. thesis, California Institute of Technology, May 2000.

[PS000] *Combinatorial approaches to finding subtle signals in dna sequences.*, vol. 8, 2000.

[Sch08] Grant Schoenebeck, *Linear level lasserre lower bounds for certain k-csps*, FOCS, 2008, pp. 593–602.

[Sho87] N. Z. Shor, *Class of global minimum bounds of polynomial functions*, Cybernetics **23** (1987), no. 6, 731–734, (Russian orig.: Kibernetika, No. 6, (1987), 9–11).

[Sze78] Endre Szemerédi, *Regular partitions of graphs*, Problèmes combinatoires et théorie des graphes (1978).

[Tao05] Terence Tao, *The dichotomy between structure and randomness, arithmetic progressions, and the primes*, arXiv preprint math/0512114 (2005).

[Tao15] *Probabilistic models and heuristics for the primes*, 2015, Available at https://terrytao.wordpress.com/2015/01/04/254a-supplement-4-probabilistic-models-and-heuristics-for-the-primes-optional/.

A Omitted Proofs

A.1 Calibration of $\tilde{\mathbb{E}}$

In this subsection we prove Lemma 5.3, restated here.

Lemma (Restatement of Lemma 5.3)**.** *Let $f_G(x) = \sum_{|S|\leqslant 2d} c_S(G)\cdot x_S$ 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 $f_G(x) = \sum_{S\subseteq[n]} c_S\cdot x_S$.

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

□

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 warmup for Lemma 7.8.)

Lemma A.1. *Let $\mathcal{T}$ be a family of subsets of $\binom{[n]}{2}$ so that for every $T, T'\in\mathcal{T}$ there exists $\sigma : [n]\to[n]$ a permutation 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 $s\geqslant 0$ and every even ℓ,*

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

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

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

In order to have $\mathbb{E}\prod_{j\leqslant\ell}\chi_{T_j} = 1$, every edge in the multiset $\bigcup_{j\leqslant\ell} T_j$ must appear at least twice, so every vertex in the multiset $\bigcup_{j\leqslant\ell}\mathcal{V}(T_j)$ 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 $(T_1,\dots,T_\ell)$, so

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

For even ℓ, by Markov's inequality,

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

□

Lemma (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

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

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 $\{\mathcal{T}_i^t\}_{i=1}^p$ by placing T and T' in the same family iff there exists a permutation $\sigma:[n]\to[n]$ of vertices so that $\sigma(T) = T'$. Thus,

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

By Lemma A.1 (taking $\ell = (\log n)^2$), and since $t\leqslant\tau\leqslant\log n$, each $\mathcal{T}_i^t$ satisfies

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

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

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

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

.

A.3 Combinatorial Proofs about Ribbons

In this section we prove Lemma 6.3, restated here:

Lemma (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 Q 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.2. We write $Q_1 \leqslant Q_2$ 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.3. *For any set of minimum vertex separators Q_1, Q_2, Q_3 an (I, J)-ribbon, we have:*

1. $Q_1 \leqslant Q_1$.
2. *If $Q_1 \leqslant Q_2$ and $Q_2 \leqslant Q_3$, then, $Q_1 \leqslant Q_3$.*
3. *If $Q_1 \leqslant Q_2$ and $Q_2 \leqslant Q_1$, 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 $Q_2 \leqslant Q_3$, P passes through a vertex in Q_2. Thus, P contains a subpath that connects I and Q_2. But since $Q_1 \leqslant Q_2$, this subpath must pass through Q_1. Thus, any such P must pass through Q_1 and thus, $Q_1 \leqslant Q_3$.

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 $P_1, P_2, \ldots, P_k$ 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 $Q_1 \leqslant Q_2$ and $Q_2 \leqslant Q_1$ if every P_i intersects Q_1, Q_2 in the same vertex. □

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 $Q_L \leqslant Q_1 \leqslant Q_R$ and $Q_L \leqslant Q_2 \leqslant Q_R$. 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 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:

Claim A.4. $W_L \cap W_R = \varnothing$.

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. □

Next, we have:

Claim A.5. 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 is similar. Assume there is a path P from I to J that does not 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. □

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

Claim A.6. $|Q_L| = |Q_R| = |Q_1| = |Q_2| = k$

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

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

Claim A.7. $Q_L \leqslant Q_1$ and $Q_2 \leqslant Q_R$.

Proof of Claim. Let P be a path from I to Q_1, 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, $Q_L \leqslant Q_1$. The other case is similar. □

This concludes the proof of the lemma. □

B Spectral Norms

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

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

- *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, whp, $\|M\| \leqslant n^{\frac{t-p-r}{2}} \cdot 2^{O(t)} \cdot (\log n)^{O(t-r+p)}$.

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 $\sigma(\bigcup_{j \leqslant p} q_j) \subseteq V_1^k$ and $\sigma(U \setminus (\bigcup_{j \leqslant p} q_j)) \subseteq V_2^k$ 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 preceeding 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.1. There is $m = O(2^t \cdot t \cdot \log n)$ so that $\bigcup_{k=1}^m S_k$ contains every labeling $\sigma : U \to G$.

Proof. For a fixed σ,

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

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

Henceforth, let $S_1, \dots, S_m$ be the partition guaranteed by the preceeding claim. For $k \leqslant m$, let $M_k(I, J) = \sum_{\sigma \in S_k \,:\, \sigma(A)=I, \sigma(B)=J} \operatorname{val}(\sigma)$. 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, $\|M\| \leqslant \sum_{k=1}^m \|M_k\|$. Fix k. We expand $\mathbb{E}_G \operatorname{Tr}(M_k^\dagger M_k)^\ell$ as

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

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

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

Notice that for all $\sigma_1, \dots, \sigma_{2\ell}$, the expectation $\mathbb{E} \prod_{j=1}^{2\ell} \operatorname{val}(\sigma_j)$ is either 0 or 1. We will bound the number of $\sigma_1, \dots, \sigma_{2\ell}$ for which $\mathbb{E} \prod_{j=1}^{2\ell} \operatorname{val}(\sigma_j) = 1$ 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 $2\ell \sum_{j \leqslant p} |q_j|$ 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 $\mathbb{E}\prod_{i=1}^{2\ell} \mathrm{val}(\sigma_i) = 1$; 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 $\ell \sum_{j\leqslant p} |q_j|$ edges, and so at most $\ell \sum_{j\leqslant p} |q_j| + p$ vertices. Thus, the vertices in $q_1, \ldots, q_p$ together receive at most $\ell \sum_{j\leqslant p} |q_j| + p$ distinct labels among all $\sigma_1, \ldots, \sigma_{2\ell}$.

Next we account for labels of $v \notin (\bigcup_{j\leqslant p} q_j \cup A \cup B)$. If $\mathbb{E}_G \prod_{i=1}^{2\ell} \mathrm{val}(\sigma_i) = 1$ then the 2ℓ-size multiset $\{\sigma_i(v)\}_{i\leqslant 2\ell}$ 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 $A \setminus (B \cup \bigcup_{j\leqslant p} q_j)$ and $B \setminus (A \cup \bigcup_{j\leqslant p} q_j)$. 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

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

unique labels.

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

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

Stage 1. Choose the labels $\sigma_i(j_1), \ldots, \sigma_i(j_r)$ 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 $\sigma_i(j)$ it will be the first appearance of the index $\sigma_i(j) \in [n]$ or if there is 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 apperance of a label. Here there are at most $x^{2\ell t - 2\ell r - x}$ options.

All together, there are at most $n^r \cdot 2^{2\ell t} \cdot n^x \cdot x^{2\ell(t-r)-x} \leqslant n^r \cdot 2^{2\ell t} \cdot n^x \cdot (2\ell t)^{2\ell(t-r)-x}$ choices for a given x. Since $4lt \ll n$, summing up over all $x \leqslant \ell(t - p - r) + p$ 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,

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

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

$$\mathbb{P}\{\|M_k\| \geqslant s\} = \mathbb{P}\{\|M_k^\dagger M_k\|^\ell \geqslant s^{2\ell}\}$$

$$
\begin{aligned}
&\leqslant \frac{\mathbb{E}\,\|(M_k^\dagger M_k)^\ell\|}{s^{2\ell}} \quad \text{by Markov's} \\
&\leqslant \frac{\mathbb{E}\,\mathrm{Tr}(M_k^\dagger M_k)^\ell}{s^{2\ell}} \quad \text{since } \|(M_k^\dagger M_k)^\ell\| \leqslant \mathrm{Tr}(M_k^\dagger M_k)^\ell \\
&\leqslant \frac{2n^r \cdot 2^{2\ell t} \cdot n^{\ell(t-p-r)+p} \cdot (2\ell t)^{\ell(t-r+p)-p}}{s^{2\ell}}
\end{aligned}
$$

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

ECCC ISSN 1433-8092

http://eccc.hpi-web.de