Bounds on the Norms of Uniform Low Degree Graph Matrices

Dhruv Medarametla[1] and Aaron Potechin[2]

1 **Stanford University, Stanford, CA, USA**
dhruvm321@gmail.com
2 **Cornell University, Ithaca, NY, USA**
aaronpotechin@gmail.com

Abstract

The Sum Of Squares hierarchy is one of the most powerful tools we know of for solving combinatorial optimization problems. However, its performance is only partially understood. Improving our understanding of the sum of squares hierarchy is a major open problem in computational complexity theory.

A key component of analyzing the sum of squares hierarchy is understanding the behavior of certain matrices whose entries are random but not independent. For these matrices, there is a random input graph and each entry of the matrix is a low degree function of the edges of this input graph. Moreoever, these matrices are generally invariant (as a function of the input graph) when we permute the vertices of the input graph. In this paper, we bound the norms of all such matrices up to a polylogarithmic factor.

1998 ACM Subject Classification G.2.2 Graph Theory

Keywords and phrases sum of squares hierarchy, matrix norm bounds

Digital Object Identifier 10.4230/LIPIcs.APPROX-RANDOM.2016.40

1 Introduction

1.1 Background and Motivation

The sum of squares hierarchy, independently developed by Shor, Nesterov, Parrillo, and Lasserre [26, 22, 23, 19], is a powerful tool for solving combinatorial optimization problems. The first level of the sum of squares hierarchy corresponds to semidefinite programming on the input variables, which is extremely useful on its own, and each subsequent level of the sum of squares hierarchy gives a larger but more accurate semidefinite program for the problem.

However, the performance of the sum of squares hierarchy is only partially understood. It is known that the sum of squares hierarchy is strictly more powerful than the Lovasz-Schrijver Hierarchy and the Sherali-Adams hierarchy. It is also known that the sum of squares hierarchy captures the best known algorithms for many problems. For example, the sum of squares hierarchy captures the Goemans-Williamson algorithm for max-cut [11] and the Goemans-Linial relaxation for sparsest cut (which was shown to give an $O(\sqrt{\log n})$ approximation by Arora, Rao, and Vazirani [3]). Also, as shown by Barak, Raghavendra, and Steurer [5] and by Guruswami and Sinop[14], the sum of squares hierarchy captures the sub-exponential algorithm for unique games found by Barak et. al. [2]. That said, for all we know, the sum of squares hierrarchy may do even better than these algorithms on max-cut, sparsest cut, and/or unique games; determining the exact performance of the sum of squares hierrarchy on max-cut, sparsest cut, and unique games is a major open problem.

On the lower bound side, it is known that the sum of squares hierarchy cannot solve NP-hard problems. Such lower bounds generally follow from the result of Grigoriev [12, 13], which was independently rediscovered by Schoenebeck [25], that the sum of squares hierarchy cannot distinguish between a random 3-XOR instance and a random 3-XOR instance with a planted solution. This problem can be reduced to 3-SAT and other NP-hard problems, which implies sum of squares lower bounds for these problems. However, until recently, few lower bounds were known for the sum of squares hierarchy for problems which are not NP-hard. For more information about the sum of squares hierarchy, see the survey of Barak and Steurer [6].

Recently, there have been several papers proving lower bounds for the performance of the Sum Of Squares Hierarchy on the planted clique problem [20, 15, 7, 24]. In the planted clique problem, introduced by Jerrum [16] and Kucera [18], we are given a graph which was created by first choosing a random graph and then randomly planting a clique of size k by choosing k vertices and making them all adjacent to each other. The goal of the problem is to recover the planted clique. Although with high probability the size of the largest clique in a random graph is only around $2 \lg n$, the current best polynomial time algorithm, a spectral algorithm due to Alon et. al. [1], can only solve the planted clique problem for $k = \Theta(\sqrt{n})$. In fact, we have strong reason to believe that doing better than $\Theta(\sqrt{n})$ in polynomial time is hard. It has been shown [16, 8, 9] that several classes of algorithms, including Monte-Carlo Markov chains, the Lovasz-Schrijver Hierarchy, and statistical algorithms, cannot do better than $\Theta(\sqrt{n})$ in polynomial time.

The papers [20, 15, 7, 24] show partial lower bounds on the sum of squares hierarchy for the planted clique problem, proving that the second level of the sum of squares hierarchy cannot solve planted clique if k is much smaller than $\sqrt{n}$ and that the rth level of the sum of squares hierarchy cannot solve planted clique if k is much smaller than $n^{\frac{1}{r+1}}$. While these papers use many different techniques, a crucial part of all of them is probabilistically bounding the norms of certain matrices. In these matrices, the entries are not completely independent of each other, but are low degree in the edges of the input graph and are highly symmetric, so we call them uniform low degree graph matrices.

Here, inspired by these papers [20, 15, 7, 24], we investigate the norms of uniform low degree graph matrices. While special cases of these matrices have been analyzed, here we generalize this analysis, proving bounds on the norms of all uniform low degree graph matrices.

Concurrently with this work, a nearly tight lower bound was proved for the sum of squares hierarchy on the planted clique problem [4], showing that the sum of squares hierarchy cannot solve the planted clique problem in polynomial time if k is much smaller than $\sqrt{n}$. Coming full circle, it turns out that this general analysis of uniform low degree graph matrices is a key component of proving the full lower bound. We have good reason to believe that this analysis of uniform low degree graph matrices will be useful in analyzing the sum of squares hierarchy on other problems and it may also be of independent interest.

Finally, we note that this work can be viewed as progress towards matrix concentration inequalities. In random matrix theory, finding concentration inequalities for the norms of matrix-valued functions is a longstanding open problem. This work gives bounds for the case when the matrix function is highly symmetric and has a random graph as input.

1.2 Preliminaries

In this paper, we use the following standard linear algrebra definitions.

▶ **Definition 1.**
1. Given a matrix M, let $M(i,j)$ be the element in the ith row and jth column of M. We use $M(i,j)$ rather than M_{ij} because we will often want to give our matrices subscripts and superscripts.
2. Given a matrix M, we take $||M||$ to be the induced norm of M, i.e. $||M|| = \max_{||v||=1} ||Mv||$.

Throughout this paper, we will be bounding the norms of matrices whose entries depend on a random graph $G \sim G(n, \frac{1}{2})$. To avoid writing G repeatedly, we make this dependence implicit rather than writing it explicitly.

To bound the norms of our matrices, we will use the moment method. In particular, we use the following fact.

▶ **Lemma 2.** *For any real matrix M, for all $k \geq 1$, $\sqrt[2k]{\mathrm{tr}((MM^T)^k)} \geq ||M||$.*

For completeness, we give a short proof of this fact in Appendix A.

Finally, we recall König's Theorem and Menger's Theorem as they will play a crucial role in our analysis.

▶ **Definition 3.** Given a graph G, a vertex cover of G is a set of vertices $V \subseteq V(G)$ such that all edges of G are incident with at least one vertex in V.

▶ **Theorem 4** (König's Theorem)**.** *If G is a bipartite graph with partite sets U and V then the minimal size of a vertex cover of G is equal to the maximal size of a matching between U and V.*

▶ **Definition 5.** If G is a graph and $U, V \subseteq V(G)$, we define a vertex separator S of U and V to be a set of vertices such that all paths from U to V intersect S.

▶ **Theorem 6** (Menger's Theorem)**.** *If G is a graph and $U, V \subseteq V(G)$ then the minimal size of a vertex separator of U and V is equal to the maximal number of vertex disjoint paths between U and V.*

1.3 Definitions for Uniform Low Degree Graph Matrices

We now rigorously define what uniform low degree graph matrices are. For the remainder of the paper, we assume that $V(G) = [1, n]$ so that the vertices of G have a natural ordering.

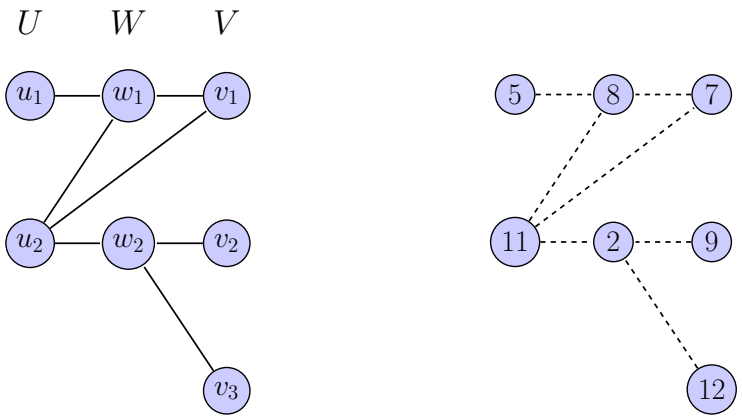
▶ **Definition 7.** Given an input graph G and a possible edge e, we define the edge variable $e = (i,j)$ to be 1 if $(i,j) \in E(G)$ and -1 otherwise. Given a set of edges E, we define $\chi_E = \prod_{e \in E} e$.

▶ Remark. We can think of the χ_E as Fourier characters on the input graph.

▶ **Definition 8.** We say that a matrix R is a graph matrix if its entries are all functions of the edge variables of some input graph G. We say that R has degree d if the maximum degree among all of these functions is d.

Uniformity says that the matrix is the same (as a function of the input graph G) when we permute the vertices of G. More precisely, we have the following definitions.

▶ **Definition 9.** Given a permutation σ of $V(G)$,
1. If $e = (u,v)$ is a possible edge of G then define $\sigma(e) = (\sigma(u), \sigma(v))$.
2. Given a set E of possible edges of G, define $\sigma(E) = \{\sigma(e) : e \in E\}$.



(a) H.

(b) The subgraph of G that $\chi_{H,A,B,C}$ is calculated from.

Figure 1

▶ **Definition 10.** We say that a graph matrix R is uniform if the following conditions hold:

1. R has rows and columns indexed by subsets A and B of $V(G)$.
2. Letting $c_{A,B,E}$ be the coefficient of χ_E in $R(A,B)$, whenever $A, A', B, B' \subseteq V(G)$, $|A'| = |A|$, $|B'| = |B|$, and σ is a permutation of $V(G)$ which maps the ith element of A to the ith element of A' and maps the jth element of B to the jth element of B', $c_{A',B',\sigma(E)} = c_{A,B,E}$.

In this paper, we focus on the following type of uniform graph matrix.

▶ **Definition 11.** Let H be a graph with two distinguished subsets of vertices $U = \{u_1, u_2, \dots u_x\}$ and $V = \{v_1, v_2, \dots v_y\}$. Let $W = \{w_1, \dots, w_z\}$ be the remaining vertices of H. Given $A = \{a_1, \dots, a_x\}$, $B = \{b_1, \dots, b_y\}$, and $C = \{c_1, \dots, c_z\}$ such that $a_i = b_j$ if and only if $u_i = v_j$, C is disjoint from $A \cup B$, and A and B are in increasing order but C may be in any order (though still with no duplicates), define $\chi_{H,A,B,C} = \chi_{\pi(E(H))}$ where π is the mapping from $V(H)$ to $V(G)$ such that $\forall i \in [1,x], \pi(u_i) = a_i$, $\forall j \in [1,y], \pi(v_j) = b_j$, $\forall k \in [1,z], \pi(w_k) = c_k$ and we take $\pi(E(H)) = \{(\pi(u), \pi(v)) : (u,v) \in E(H)\}$

We define the matrix R_H to be the $\binom{n}{x} \times \binom{n}{y}$ matrix with entries $R_H(A,B) = \sum_C \chi_{H,A,B,C}$ whenever $a_i = b_j$ if and only if $u_i = v_j$ and we take $R_H(A,B) = 0$ otherwise.

▶ **Example 12.** The following is an example of $\chi_{H,A,B,C}$ for a particular H, A, B, and C. If H is the graph shown below in Figure 1a, $A = \{5, 11\}$, $B = \{7, 9, 12\}$, and $C = \{8, 2\}$, then $\chi_{H,A,B,C}$ is calculated from the subgraph of G displayed in Figure 1b. In particular, $\chi_{H,A,B,C}$ is the product of the edge variables of the seven possible edges of G that are displayed in Figure 1b.

▶ Remark. If H is a bipartite graph with partite sets U and V then $R_H(A,B) = 0$ if $A \cap B \neq \emptyset$ and whenever $A \cap B = \emptyset$, $R_H(A,B)$ is ± 1. Moreover, $R(A,B)$ only depends on the edges between A and B in G.

▶ **Example 13.** If H consists of a single edge from u_1 to v_1 then R_H is a ± 1 symmetric random matrix with zeros on the diagonal.

▶ Remark. All uniform graph matrices can be expressed as a linear combination of matrices of the form R_H. Thus, to upper bound the norms of all uniform low degree graph matrices,

it is sufficient to upper bound norms of the matrices R_H for small H. To lower bound the norms of all uniform low degree graph matrices, a priori it is insufficient to lower bound the norms of the matrices R_H for small H, as if we take a linear combination of different R_H it is possible that there is almost perfect cancellation between them. That said, it turns out that the probility of such a cancellation is negligible, so the norms of all uniform low degree graph matrices can be understood in terms of the norms of their component R_H. For details on how this can be shown, see Section 6.

1.4 Paper Outline and Results

Our main result is the following theorem.

▶ **Theorem 14.** *Let H be a graph with distinguished sets of vertices U and V such that U and V are disjoint and all vertices in $H(V) \setminus (U \cup V)$ have degree at least one. Let $t = |V(H)|$, let $z = |V(H) \setminus (U \cup V)|$, and let q be the size of the minimal separator between U and V. If $q \geq 1$ then for all $\epsilon \in (0,1)$,*

$$\mathbb{P}\left[||R_H|| \geq 2(t^t)\left(e(t+z)\left(\frac{\ln(8n^q/\epsilon)}{2(q+z)}+1\right)\right)^{q+z} n^{\frac{t-q}{2}}\right] \leq \epsilon\,.$$

In Section 2, we introduce our main techniques by applying them to the simple and well-studied case of a symmetric ± 1 random matrix. We then give a brief technical overview of the proof for the general case in Section 3. In Section 4 we prove the result for all bipartite graphs H with partite sets U and V. In Section 5 we generalize our techniques and prove the full result. The case where U and V have non-trivial intersection is considered in Appendix B. Finally, in Section 6 we show that this theorem is tight up to a polylog(n) factor.

1.5 Comparison with Previous Work

This paper can be compared to the recent body of work [20, 15, 7, 24] showing planted clique lower bounds and to previous work in random matrix theory. In the planted clique lower bounds, $||R_H||$ is bounded for several special cases of H, but only the ones that are needed for the sum of squares lower bounds. In this paper, we use many of the same ideas (constraint graphs, looking at cycles, vertex partitioning), but we consider bounding $||R_H||$ as a mathematical problem independent of its applications to the sum of squares hierarchy, obtaining bounds for all possible H and greatly generalizing the previous work.

In terms of random matrix theory, our results are much less precise than classical results such as Wigner's semicircle law [27] and Girko's circular law [10]. While these results give an exact distribution for the eigenvalues of symmetric random matrices and asymmetric random matrices respectively, we only give a norm bound and this norm bound is off by polylogarithmic factor. That said, the matrices we are considering are much more complicated as the entries are no longer independent and may behave in complex ways on the input graph G. To the best of our knowledge, uniform low degree graph matrices have not previously been studied in random matrix theory. Indeed, as noted in the introduction, obtaining general norm bounds when we have a matrix valued function of random inputs rather than a matrix with independent entries is a longstanding open problem in random matrix theory.

2 Warm-up: Bounding the Norm of a ± 1 Random Matrix

As a warmup, we consider the case of a ± 1 symmetric random matrix. This type of matrix and its norm have already been studied extensively, in particular Wigner's semicircle law

[27] says that with high probability, the norm of an $n \times n$ symmetric random ± 1 matrix is $2\sqrt{n}(1 \pm o(1))$. While our upper bound will not be as strong, it will illustrate the general ideas involved.

▶ **Definition 15.** Given a random graph $G \sim G(n, \frac{1}{2})$ with vertices $1, \cdots, n$, let R be the matrix with the following entries:

$$R(i,j) = \begin{cases} 0 & i = j \\ 1 & (i,j) \in E(G) \\ -1 & (i,j) \notin E(G). \end{cases}$$

▶ Remark. As noted in the introduction, $R = R_H$ where H consists of a single edge from u_1 to v_1.

Note that R is closely related to the adjacency matrix of G; in fact, it is the additive inverse of the Seidel adjacency matrix. Further note that for all $1 \leq i, j \leq n$, $\mathbb{E}[R(i,j)] = 0$, as any edge (i,j) has probability $\frac{1}{2}$ of being included in G. We now show the following probabilistic bound on the norm of R. Note that this bound has an extra factor of $\ln(n)$, but this is fine for our purposes as in this paper we are only aiming to get the correct norm bounds to within a factor of polylog(n).

▶ **Theorem 16.** *For all $\epsilon \in (0,1)$,*

$$\mathbb{P}\left[||R|| \geq e\sqrt{n}(\ln(n/\epsilon) + 2)\right] \leq \epsilon\,.$$

Proof. In order to find a probabilistic bound for $||R||$, we bound $\mathbb{E}\left[\sqrt[2k]{\mathrm{tr}((RR^T)^k)}\right]$. Notice that

$$\mathrm{tr}((RR^T)^k) = \mathrm{tr}(R^{2k}) = \sum_{i_1, i_2, \ldots, i_{2k} \in [1,n]} \left(\prod_{j=1}^{2k} R(i_j, i_{j+1}) \right)$$

where $i_{2k+1} = i_1$ and $[1,n] = \{1, 2, \ldots, n\}$. Therefore,

$$\begin{aligned} \mathbb{E}[\mathrm{tr}((RR^T)^k)] = \mathbb{E}[\mathrm{tr}(R^{2k})] &= \mathbb{E}\left[\sum_{i_1, i_2, \ldots, i_{2k} \in [1,n]} \left(\prod_{j=1}^{2k} R(i_j, i_{j+1}) \right) \right] \\ &= \sum_{i_1, i_2, \ldots, i_{2k} \in [1,n]} \mathbb{E}\left[\prod_{j=1}^{2k} R(i_j, i_{j+1}) \right] \end{aligned}$$

by linearity of expectation. Now, note that because $\mathbb{E}[R(i,j)] = 0$, the vast majority of the terms $\mathbb{E}\big[\prod_{j=1}^{2k} R(i_j, i_{j+1})\big]$ are 0; in fact, the only time the expected value is non-zero is when each consecutive pair of i's is distinct and when each $R(i,j)$ term appears an even number of times, in which case the expected value will be 1. Therefore, we can calculate the number of choices for $i_1, i_2, \ldots, i_{2k}$ that yield a non-zero value for $\mathbb{E}\big[\prod_{j=1}^{2k} R(i_j, i_{j+1})\big]$ and use that number to bound $\mathbb{E}[\mathrm{tr}((RR^T)^k)]$. We can think of the sum $\mathbb{E}\big[\prod_{j=1}^{2k} R(i_j, i_{j+1})\big]$ graphically as a sum over length $2k$ cycles in the vertex set $[1,n]$ where some vertices in the cycle may be equal to each other. We use what we call a **constraint graph** to represent each such cycle (similar graphs appeared in [20] and [15]). In this case, the constraint graph consists of $2k$

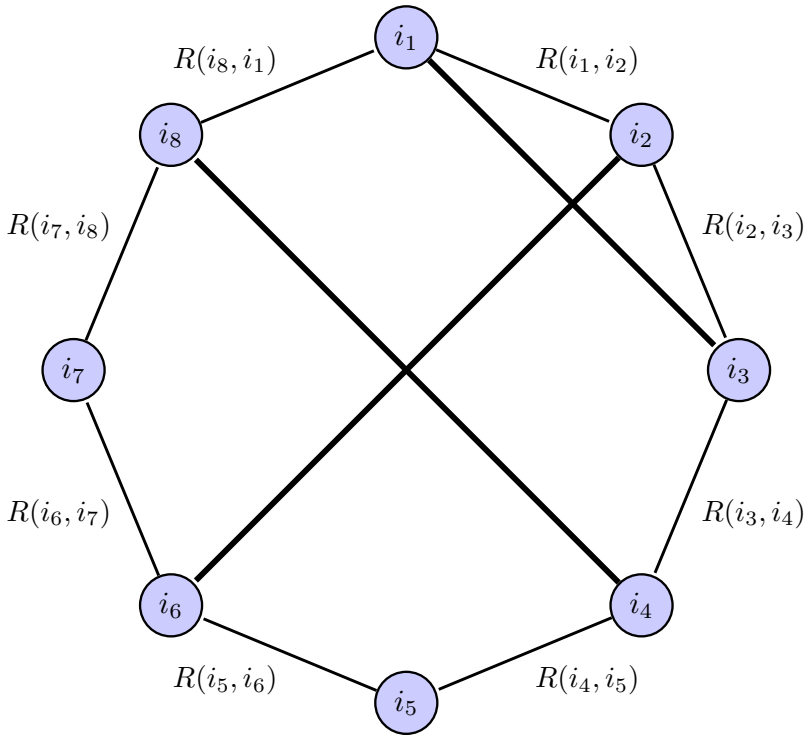


Figure 2 An example of a constraint graph where $k = 4$, $i_1 = i_3$, $i_2 = i_6$, and $i_4 = i_8$.

vertices, each labeled from i_1 to i_{2k}; vertex i_j is connected to vertex i_{j+1} for all $1 \leq j \leq 2k$ to represent the term $R(i_j, i_{j+1})$, and a bold constraint edge is drawn between i_r and i_s whenever $i_r = i_s$ to signify that they are equal.

In the case where j of $2k$ variables are equal, we only draw $j - 1$ constraint edges to represent that equality, rather than $\binom{j}{2}$. This is because each constraint edge essentially represents a restriction; the extra constraint edges do not add to these restrictions, so they are not included.

▶ **Proposition 17.** *In order for* $\mathbb{E}\big[\prod_{j=1}^{2k} R(i_j, i_{j+1})\big]$ *to have a non-zero value, there must be at least* $k - 1$ *constraint edges in the respective constraint graph; in addition, this bound is sharp.*

Proof. We prove the first statement by induction on k. When $k = 1$, the statement is vacuously true; $\mathbb{E}\big[\prod_{j=1}^{2k} R(i_j, i_{j+1})\big] = \mathbb{E}[R(i_1, i_2)^2]$, which has a non-zero value regardless of constraint edges.

Now, assume that the statement is true for $k = r$, and consider $k = r + 1$. Assume $\mathbb{E}\big[\prod_{j=1}^{2k} R(i_j, i_{j+1})\big] \neq 0$, and consider the constraint graph. If each vertex is adjacent to at least one constraint edge, then because each constraint edge is incident to two vertices, there are at least $\frac{2r+2}{2} = r + 1$ constraint edges, and we are done. Therefore, we only need to consider the case where there exists a vertex that is not adjacent to any constraint edges. Call this vertex i_j. Then, note that the statement $i_{j-1} = i_{j+1}$ must be true; if it was not, then the values $R(i_{j-1}, i_j)$ and $R(i_j, i_{j+1})$ have no corresponding equal terms, which means

$\mathbb{E}\big[\prod_{j=1}^{2k} R(i_j, i_{j+1})\big] = 0$. But if $i_{j-1} = i_{j+1}$, then $R(i_{j-1}, i_j) = R(i_j, i_{j+1})$, meaning that we no longer need to consider the vertex i_j and its adjacent edges. Therefore, we can treat the vertices i_{j-1} and i_{j+1} as the same vertex, as they are equal, meaning that we have essentially reduced the constraint graph to one on $2r$ vertices. Then, by our induction hypothesis, this constraint graph requires at least $r-1$ constraint edges to create a nonzero expected value, which means that our total constraint graph requires at least r constraint edges, completing the proof.

In order to prove the sharpness of the bound, simply consider the case where $i_j = i_{2k+2-j}$ for all $2 \leq j \leq k$. Then, $R(i_l, i_{l+1}) = R(i_{2k+1-l}, i_{2k+2-l})$ for all $1 \leq l \leq k$, which creates a non-zero expected value. ◀

We now use Proposition 17 to bound the maximum number of times that $\mathbb{E}\big[\prod_{j=1}^{2k} R(i_j, i_{j+1})\big]$ can take a non-zero value, and use that information to bound $\mathbb{E}[\mathrm{tr}(R^{2k})]$.

▶ **Proposition 18.** *Given a constraint graph on b vertices such that at least c constraint edges are required to create a non-zero expectation value, where each vertex has n possible values, let N represent the number of choices for the b vertices such that the expectation value of the product is non-zero. Then, $N \leq \binom{b}{c} n^{b-c}(b-c)^c \leq b^{2c} n^{b-c}$.*

Proof. Treat the set of vertices as an ordered set $S = \{d_1, d_2, \ldots, d_b\}$.

Because there must be at least c constraint edges, there must be at least c elements of S that are duplicates of other elements, so we can choose a set $I \subseteq S_b$ of c indices such that for all $j \in I$, there exists $m \notin I$ such that $d_j = d_m$. There are $\binom{b}{c}$ choices for I. We can then choose the elements $\{d_j \mid j \notin I\}$. Each element has at most n possible values so there are at most n^{b-c} choices for these elements. Finally, we choose the elements $\{d_j \mid j \in I\}$. To determine each d_j it is enough to specify the $m \notin I$ such that $d_j = d_m$. Each such d_j has $b-c$ choices, so there are at most $(b-c)^c$ choices for these elements. Therefore, $N \leq \binom{b}{c} n^{b-c}(b-c)^c$.

Now, note that $\binom{b}{c} \leq b^c$, as $\binom{b}{c} = \frac{b!}{(b-c)!c!} \leq \frac{b!}{(b-c)!} \leq b^c$. As $(b-c)^c \leq b^c$, this completes the proof. ◀

▶ **Corollary 19.** *Let N represent the number of choices for the variables $(i_1, i_2, \ldots, i_{2k})$ such that $\mathbb{E}\big[\prod_{j=1}^{2k} R(i_j, i_{j+1})\big] \neq 0$. Then, $N \leq (2k)^{2k-2} n^{k+1}$.*

Proof. Apply Proposition 18. Note that $b = 2k$ and $c = k-1$ by Proposition 17. This implies the desired result. ◀

▶ **Corollary 20.** $\mathbb{E}[\mathrm{tr}(R^{2k})] \leq (2k)^{2k-2} n^{k+1}$.

Proof. Recall $\mathbb{E}[\mathrm{tr}(R^{2k})] = \sum_{i_1, i_2, \ldots, i_{2k} \in [1,n]} \mathbb{E}\big[\prod_{j=1}^{2k} R(i_j, i_{j+1})\big]$. By Corollary 19, the number of choices for $(i_1, i_2, \ldots, i_{2k})$ that yield a non-zero value for $\mathbb{E}\big[\prod_{j=1}^{2k} R(i_j, i_{j+1})\big]$ is at most $(2k)^{2k-2} n^{k+1}$; in addition, $\mathbb{E}\big[\prod_{j=1}^{2k} R(i_j, i_{j+1})\big] \leq 1$ for all choices of $(i_1, i_2, \ldots, i_{2k})$. These two observations complete the proof. ◀

Now, note that for any matrix R, $\text{tr}(R^{2k})$ must take on a nonnegative value. By Markov's inequality, for all $\epsilon \in (0,1)$ and all $k \geq 1$, $\mathbb{P}[\text{tr}(R^{2k}) \geq \frac{\mathbb{E}[\text{tr}(R^{2k})]}{\epsilon}] \leq \epsilon$

Using Corollary 20, $\mathbb{P}[\text{tr}(R^{2k}) \geq (2k)^{2k-2}n^{k+1}/\epsilon] \leq \epsilon$. Since $||R|| \leq \sqrt[2k]{\text{tr}((RR^T)^k)} = \sqrt[2k]{\text{tr}(R^{2k})}$ for all $k \geq 1$, this implies that for all $\epsilon \in (0,1)$ and all $k \geq 1$,

$$\mathbb{P}\left[||R|| \geq \sqrt[2k]{(2k)^{2k-2}n^{k+1}/\epsilon}\right] \leq \epsilon\,.$$

Choosing $k = \lceil \ln(n/\epsilon)/2 \rceil$, we have that

$$\sqrt[2k]{(2k)^{2k-2}n^{k+1}/\epsilon} \leq \sqrt[2k]{(2k)^{2k}n^{k+1}/\epsilon} = 2k\sqrt{n}(n/\epsilon)^{\frac{1}{2k}} = 2k\sqrt{n}e^{\frac{\ln(n/\epsilon)}{2k}} \leq e\sqrt{n}(\ln(n/\epsilon)+2)\,.$$

Thus, $\mathbb{P}[||R|| \geq e\sqrt{n}(\ln(n/\epsilon)+2)] \leq \epsilon$, as needed. ◀

In the following sections, we generalize these techniques for matrices whose entries depend on the random graph in more complex ways.

3 Technical Overview of the General Norm Bounds

For the general bounds on $||R_H||$, we use similar ideas. The following is almost correct, but there is a technical issue that needs to be dealt with which we discuss afterwards. We express $E[tr((R_H {R_H}^T)^k)]$ as a sum of many different terms, each of which can be represented with a constraint graph. We upper bound the number of terms which have nonzero expectation by showing a lower bound on the number of constraint edges needed. We then use this to probabilistically bound $||R_H||$.

In the case where H is bipartite, each vertex of H has k copies in the constraint graph so the total number of vertices is kt where $t = V(H)$. The number of constraint edges that are needed to make a term have non-zero expectation is $q(k-1)$ where q is the size of a minimal vertex cover of H. One way we can achieve this is as follows. We take a minimal vertex cover S of H and set all copies of a vertex in S to be equal to each other. Since each vertex in H is copied k times, this requires $q(k-1)$ constraint edge. It turns out that this is tight. Using this bound, there are at most $O(n^{tk-q(k-1)})$ nonzero terms in $E[tr((R_H {R_H}^T)^k)]$ (where the constant hides a function of k). Taking this to the power $\frac{1}{2k}$ for an appropriately chosen k, we obtain that with high probability, $||R_H||$ is at most $O(n^{\frac{t-q}{2}}polylog(n))$. The general case is more complicated but similar ideas apply. It turns out that the key object is a minmal separator S of U and V in H.

However, there is a technical issue in the analysis. In order to obtain the lower bounds on the number of constraint edges needed, we need to assume that the constraint edges behave nicely, namely that we don't have constraint edges between copies of two different vertices in H. This makes part of the constraint graph decompose into disjoint cycles, allowing us to use Proposition 17 (without this restriction, we could have constraint edges between different cycles, which invalidates the analysis). To handle this, we use a vertex partitioning argument. In particular, given a partition $V_1, \ldots, V_t$ of $[1,n]$ we consider the part of R_H where for all i, vertex i is in V_i. This gives us a matrix R' where when we look at $E[tr((R'{R'}^T)^k)]$, the constraint edges behave nicely and we can obtain a probabilistic bound on $||R'||$. We then bound $||R_H||$ using the bound on $||R'||$.

4 Bounding the Norms of Uniform Locally Random Matrices

In this section, we generalize the techniques used in Section 2 to prove Theorem 14 whenever H is a bipartite graph with partite sets U and V. We call these matrices locally random because the value of the entry in row A and column B only depends on the behavior of the input graph G on the vertices $A \cup B$.

▶ **Theorem 21.** *If H is a bipartite graph with t vertices and minimal vertex cover of size q then*

1. $||R_H|| \leq n^{\frac{t}{2}}$
2. *If $q \geq 1$, for all $\epsilon \in (0,1)$,*

$$\mathbb{P}\left[||R_H|| > 2t^t \left(et\left(\frac{\ln(8n^q/\epsilon)}{2q}+1\right)\right)^q n^{\frac{t-q}{2}}\right] < \epsilon$$

▶ Remark. As we will show in Section 6, this bound is tight up to a factor of polylog(n).

Proof. For the first statement, recall that for any matrix M, $||M|| \leq ||M||_{Fr}$, where $||M||_{Fr} = \sqrt{\sum_{i,j} M(i,j)^2}$ is the Frobenius norm of M. To see this, note that if u and v are unit vectors then

$$u^T M v = \sum_{i,j} u_i M(i,j) v_j \leq \sqrt{\sum_{i,j} {u_i}^2 {v_j}^2}\sqrt{\sum_{i,j} M(i,j)^2} = \sqrt{\sum_{i,j} M(i,j)^2}$$

by the Cauchy-Schwarz inequality. Since every entry of R_H has magnitude at most 1, the result follows.

For the second statement, as described in the technical overview, we first bound the norms of closely related matrices where we restrict which vertices H can map into. We will then use this bound to bound $||R_H||$.

▶ **Definition 22.** Given a partition $V_1, \ldots, V_t$ of the vertices of $V(G)$, we define $R_{H,V_1,\ldots,V_t}$ be the $\binom{n}{x} \times \binom{n}{y}$ matrix such that

$$R_{H,V_1,\ldots,V_t}(A,B) = \begin{cases} R_H(A,B) = \chi_{H,A,B} & A \cap B = \emptyset, \\ & \forall i \in [1,x], a_i \in V_i,\ \forall j \in [1,y], b_j \in V_{x+j} \\ 0 & \textit{otherwise} \end{cases}$$

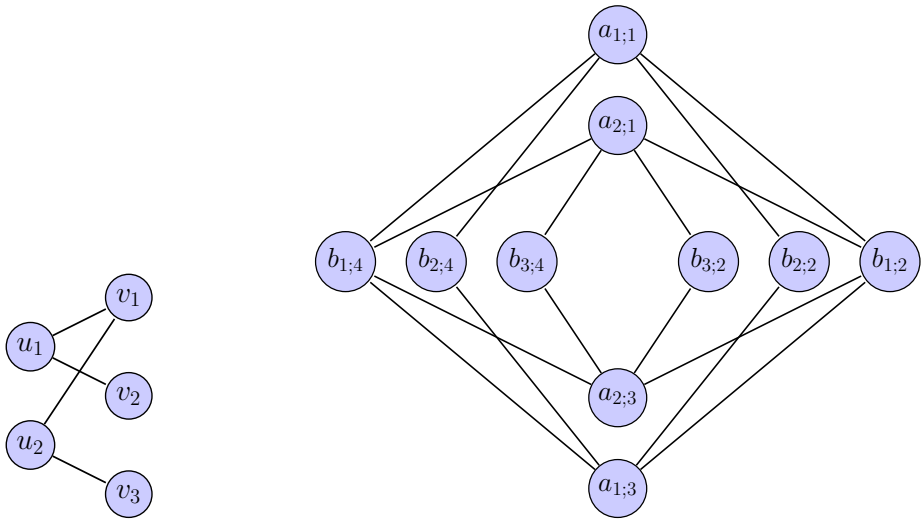
▶ **Lemma 23.** *Let $R' = R_{H,V_1,\ldots,V_t}$. For all $\epsilon \in (0,1)$,*

$$\mathbb{P}\left[||R'|| \geq \left(et\left(\frac{\ln(n^q/\epsilon)}{2q}+1\right)\right)^q n^{\frac{t-q}{2}}\right] \leq \epsilon\,.$$

Proof. As before, we probabilistically bound $||R'||$ by bounding $\mathbb{E}[\sqrt[2k]{\text{tr}((R'R'^T)^k)}]$. Define $\binom{[n]}{i}$ to be the set of all subsets of $[1,n]$ of size i. Now note that

$$\mathbb{E}[\text{tr}((R'R'^T)^k)] = \mathbb{E}\left[\sum_{\substack{A_1,A_3,\ldots,A_{2k-1} \in \binom{[n]}{x} \\ B_2,B_4,\ldots,B_{2k} \in \binom{[n]}{y}}} \left(\prod_{j=1}^{k} R'(A_{2j-1},B_{2j}) R'^T(B_{2j},A_{2j+1})\right)\right]$$

$$= \sum_{\substack{A_1,A_3,\ldots,A_{2k-1} \in \binom{[n]}{x} \\ B_2,B_4,\ldots,B_{2k} \in \binom{[n]}{y}}} \mathbb{E}\left[\prod_{j=1}^{k} R'(A_{2j-1},B_{2j}) R'^T(B_{2j},A_{2j+1})\right]$$



(a) H.

(b) An example of the constraint graph for the given example of H, where $k = 2$.

Figure 3

by linearity of expectation. Denote $\prod_{j=1}^{k} R'(A_{2j-1}, B_{2j})R'^{T}(B_{2j}, A_{2j+1})$ as $P(A_1, \ldots, B_{2k})$. Similarly to the previous case, because $\mathbb{E}[R'(A, B)] = 0$, the vast majority of the terms $\mathbb{E}[P(A_1, \ldots, B_{2k})]$ are 0; the only time the expected value can be non-zero is when each consecutive pair of A's and B's is disjoint and every edge of G involved in the product appears an even number of times. In this case, the expected value will be 1. So, we can bound the number of choices for $A_1, B_2, \ldots, A_{2k-1}, B_{2k}$ that yield a non-zero value for $\mathbb{E}[P(A_1, \ldots, B_{2k})]$ and use that number to bound $\mathbb{E}[\mathrm{tr}((R'R'^{T})^k)]$. In order to represent $\mathbb{E}[P(A_1, \ldots, B_{2k})]$, we use another constraint graph.

This constraint graph is similar to the constraint graph in Proposition 17. In this constraint graph, there are $k(x + y)$ vertices sorted into $2k$ sets. These vertices are labeled $A_1 = \{a_{1;1}, a_{2;1}, \ldots, a_{x;1}\}, B_2 = \{b_{1;2}, b_{2;2}, \ldots, b_{y;2}\}, A_3 = \{a_{1;3}, a_{2;3}, \ldots, a_{x;3}\}, \ldots, B_{2k} = \{b_{1;2k}, b_{2;2k} \ldots, b_{y;2k}\}$. Two vertices $a_{p;q}$ and $b_{r;s}$ are adjacent in the constraint graph if and only if $|q - s| = 1$ and u_p and v_r are adjacent in H, where $a_{p;1} = a_{p;2k+1}$.

Now, in order to bound the number of choices for $A_1, B_2, \ldots, A_{2k-1}, B_{2k}$ that yield a non-zero expectation value, we can introduce the constraint edges again. However, note that due to the definition of R', constraint edges can only exist between vertices of the constraint graph that are created by the same vertex of H, as it is impossible for two vertices that are not created by the same vertex of H to be equal, as they correspond to different disjoint sets V_i and the value of each variable must be in its respective set.

▶ **Lemma 24.** *In order for* $\mathbb{E}[P(A_1, \ldots, B_{2k})]$ *to have a non-zero value, there must be at least* $q(k-1)$ *constraint edges in the respective constraint graph, where* q *is the size of a minimal vertex cover of* H*; in addition, this bound is sharp.*

Proof. In order to prove this lemma, we first show that the given bound is an upper bound then show that it is sharp by König's Theorem [17].

First, note that in order for $\mathbb{E}[P(A_1, \ldots, B_{2k})]$ to have a non-zero value, every edge in the constraint graph must have an equal counterpart by virtue of the constraint edges; this

ensures that any edge that appears in the product appears an even number of times, creating a non-zero expected value.

It is easy to see that at most $q(k-1)$ constraint edges are required; namely, if V is a minimal vertex cover of H, then if $x_i \in V$, set $a_{i;1} = a_{i;3} = \cdots = a_{i;2k-1}$, and if $y_j \in V$, set $b_{j;2} = b_{j;4} = \cdots = b_{j;2k}$. Each such set of equalities corresponds to $k-1$ constraint edges, meaning that there are $q(k-1)$ constraint edges total. In addition, every edge in the constraint graph will have an equal counterpart by this method. If $(x_i, y_j) \in H$, at least one of x_i and y_j is in V by definition; without loss of generality $y_j \in V$. Then, this implies that for the edges in the constraint graph of the form $(a_{i;1}, b_{j;2}), (b_{j;2}, a_{i;3}), (a_{i;3}, b_{j;4}), \ldots, (b_{j;2k}, a_{i;1})$, each edge $(a_{i;2m-1}, b_{j,2m-2})$ has the equal counterpart $(a_{i;2m-1}, b_{j,2m})$ for $1 \leq m \leq k$, as $b_{j,2m-2} = b_{j,2m}$. Thus, this set of constraint edges is sufficient to create a non-zero expected value.

Now, we must show at least $q(k-1)$ constraint edges are required. Because H is a bipartite graph, we can apply König's Theorem, which states that there exists a matching of size q in H. Consider the q disjoint cycles of length $2k$ in the constraint graph that are created by the q edges in the matching of H. Because R' is defined so that constraint edges can only exist between vertices $a_{p;q}$ and $a_{p;q'}$ or between $b_{r;s}$ and $b_{r;s'}$, as two vertices not of this form do not belong to the same set V_i, any constraint edge created can affect at most 1 of the q cycles, due to the fact that all the cycles are disjoint and thus are impossible to link with a constraint edge. Therefore, each cycle requires at least $k-1$ constraint edges by Proposition 17, implying that the q cycles require at least $q(k-1)$ constraint edges total, completing the proof. ◀

▶ **Corollary 25.** *Let N represent the number of choices for the sets $A_1, B_2, \ldots, A_{2k-1}, B_{2k}$ such that $\mathbb{E}[P(A_1, \ldots, B_{2k})] \neq 0$. Then, $N \leq (tk)^{2(k-1)q} n^{(t-q)k+q}$ where $t = |V(H)| = x + y$.*

Proof. Apply Proposition 18. In this situation, $b = kt$ and $c = q(k-1)$. This implies the desired result. ◀

▶ **Corollary 26.** $\mathbb{E}\left[\mathrm{tr}((R'R'^T)^k)\right] \leq (tk)^{2(k-1)q} n^{(t-q)k+q}$.

Proof. Recall $\mathbb{E}[\mathrm{tr}((R'R'^T)^k)] = \sum_{\substack{A_1, A_3, \ldots A_{2k-1} \in \binom{[n]}{x} \\ B_2, B_4, \ldots B_{2k} \in \binom{[n]}{y}}} \mathbb{E}\left[\prod_{j=1}^{k} R'(A_{2j-1}, B_{2j}) R'^T(B_{2j}, A_{2j+1})\right]$.

Then, by Proposition 25, the number of choices for $A_1, B_2, \ldots A_{2k-1}, B_{2k}$ that yield a non-zero value for $\mathbb{E}\left[\prod_{j=1}^{k} R'(A_{2j-1}, B_{2j}) R'^T(B_{2j}, A_{2j+1})\right]$ is at most $(tk)^{2(k-1)q} n^{(t-q)k+q}$; in addition,
$\mathbb{E}\left[\prod_{j=1}^{k} R'(A_{2j-1}, B_{2j}) R'^T(B_{2j}, A_{2j+1})\right] \leq 1$. These two observations complete the proof. ◀

Now, note that for any graph G on n vertices, $\mathrm{tr}((R'R'^T)^k)$ must take on a nonnegative value. Then, by Markov's inequality and Corollary 26, for all $\epsilon \in (0, 1)$,

$$\mathbb{P}\left[\mathrm{tr}((R'R'^T)^k) \geq \frac{\mathbb{E}[\mathrm{tr}((R'R'^T)^k)]}{\epsilon}\right] \leq \mathbb{P}\left[\mathrm{tr}((R'R'^T)^k) \geq (tk)^{2(k-1)q} n^{(t-q)k+q}/\epsilon\right] \leq \epsilon.$$

Since $||R'|| \leq \sqrt[2k]{\mathrm{tr}((R'R'^T)^k)}$, this implies that for all $k \geq 1$ and all $\epsilon \in (0, 1)$,

$$\mathbb{P}\left[||R'|| \geq \sqrt[2k]{(tk)^{2(k-1)q} n^{(t-q)k+q}/\epsilon}\right] \leq \mathbb{P}\left[||R'|| \geq (tk)^q n^{\frac{t-q}{2}} (n^q/\epsilon)^{1/2k}\right] \leq \epsilon\,.$$

Setting $k = \lceil \frac{1}{2q} \ln(n^q/\epsilon) \rceil$ we have that $(tk)^q n^{\frac{t-q}{2}} (n^q/\epsilon)^{1/2k} \leq \left(t\left(\frac{\ln(n^q/\epsilon)}{2q} + 1\right)\right)^q n^{\frac{t-q}{2}} e^q$

Therefore, $\mathbb{P}[||R'|| \geq \left(et\left(\frac{\ln(n^q/\epsilon)}{2q} + 1\right)\right)^q n^{\frac{t-q}{2}}] \leq \epsilon$, as needed. ◀

We can now use our bounds for $||R'||$ to bound $||R||$ through the following lemma.

▶ **Lemma 27.** *Let M be a matrix and B, p be positive numbers such that:*

1. *$M = \frac{1}{N} \sum_{V_1,\cdots,V_t} M_{V_1,\cdots,V_t}$ for some matrices $\{M_{V_1,\cdots,V_t}\}$ where N is the number of possible $V_1, \cdots, V_t$.*
2. *For each choice of $V_1, \cdots, V_t$, for all $x \in [\frac{1}{2}, N]$, $\mathbb{P}(||M_{V_1,\cdots,V_t}|| > Bx) \leq \frac{p}{64x^3}$.*

then $\mathbb{P}(||M|| \geq B) < p$.

▶ Remark. Unlike in [15], we use all possible partitions so we have that $N = t^n$

Proof. The result follows from the following proposition.

▶ **Proposition 28.** *For all $j \in [0, \lg N]$, the probability that there are more than $\frac{N}{2^{2j+2}}$ matrices $M_{V_1,\cdots,V_t}$ such that $||M_{V_1,\cdots,V_t}|| > 2^{j-1}B$ is at most $\frac{p}{2^{j+1}}$.*

Proof. We prove this by contradiction. If the probability that there are more than $\frac{N}{2^{2j+2}}$ matrices $M_{V_1,\cdots,V_t}$ such that $||M_{V_1,\cdots,V_t}|| > 2^{j-1}B$ is greater than $\frac{p}{2^{j+1}}$ then the probability that $||M_{V_1,\cdots,V_t}|| > 2^{j-1}B$ must be greater than $\frac{p}{2^{3j+3}}$. Plugging in $x = 2^{j-1}$, this gives a contradiction. ◀

Using this proposition, with probability at least $1 - \sum_{j=0}^{\lfloor \lg n \rfloor} \frac{p}{2^{j+1}}$, for all integers j such that $0 \leq j \leq \lg N$, there are at most $\frac{N}{2^{2j+2}}$ matrices $M_{V_1,\cdots,V_t}$ such that $||M_{V_1,\cdots,V_t}|| > 2^{j-1}B$. When this occurs, for all integers j such that $0 \leq j \leq \lfloor \lg N \rfloor - 1$, there are at most $\frac{N}{2^{2j+2}}$ matrices $M_{V_1,\cdots,V_m}$ such that $2^{j-1}B < ||M_{V_1,\cdots,V_m}|| \leq 2^j B$. Moreover, there are no matrices such that $||M_{V_1,\cdots,V_t}|| > 2^{\lfloor \lg N \rfloor - 1}B$. This implies that with probability at least $1 - \sum_{j=0}^{\lfloor \lg n \rfloor} \frac{p}{2^{j+1}}$, $||M|| \leq \frac{B}{2} + \sum_{j=0}^{\lfloor \lg N \rfloor} \frac{2^j B}{2^{2j+2}} < B$, as needed. Since $1 - \sum_{j=0}^{\lfloor \lg n \rfloor} \frac{p}{2^{j+1}} > 1 - p$, the result follows. ◀

▶ **Proposition 29.** *Set $M = R_H$ and $M_{V_1,\ldots,V_t} = t^t R_{H,V_1,\ldots,V_t}$. For all $\epsilon \in (0,1)$, take $p = \epsilon$ and let $B = 2t^t \left(et\left(\frac{\ln(8n^q/\epsilon)}{2q} + 1\right)\right)^q n^{\frac{t-q}{2}}$. Then, conditions (1) and (2) of Lemma 27 holds true for these values of M, $M_{V_1,\ldots,V_t}$, B, and p.*

Proof. We must show both (1) and (2).

Note that $M = \frac{1}{N} \sum_{V_1,\ldots,V_t} M_{V_1,\ldots,V_t}$ because given a non-zero term $R_H(A,B)$, $R'(A,B)$ has probability $\frac{1}{t^t}$ of equaling $R_H(A,B)$ among all possible $V_1, V_2, \ldots, V_t$. Thus, part (1) of the Lemma holds.

Now, we must show (2); that $\mathbb{P}[t^t||R'|| > Bx] \leq \frac{p}{64x^3}$ for all $x \in [\frac{1}{2}, N]$. Plugging in $\epsilon' = \frac{\epsilon}{64x^3}$ to Lemma 23, we have that

$$\mathbb{P}\left[||R'|| \geq \left(et\left(\frac{\ln(64x^3n^q/\epsilon)}{2q} + 1\right)\right)^q n^{\frac{t-q}{2}}\right] \leq \frac{\epsilon}{64x^3}.$$

We need to show that for all $x \in [\frac{1}{2}, N]$, $Bx \geq t^t \left(et\left(\frac{\ln(64x^3n^q/\epsilon)}{2q} + 1\right)\right)^q n^{\frac{t-q}{2}}$. Note that if $x = \frac{1}{2}$ then $Bx = t^t \left(et\left(\frac{\ln(64x^3n^q/\epsilon)}{2q} + 1\right)\right)^q n^{\frac{t-q}{2}}$ so it is sufficient to show that

$\frac{\left(\frac{\ln(64x^3n^q/\epsilon)}{2q}+1\right)^q}{x}$ is a decreasing function for $x \geq \frac{1}{2}$. Taking the derivative of this function yields

$$\frac{\left(\frac{\ln(64x^3n^q/\epsilon)}{2q}+1\right)^q}{x^2}\left(\frac{3}{2\left(\frac{\ln(64x^3n^q/\epsilon)}{2q}+1\right)}-1\right)$$

This is negative for $x \geq \frac{1}{2}$ if $n \geq e$. If $n \leq e$ then it only makes sense to have $q \leq 1$ and we again have that this is negative for $x \geq \frac{1}{2}$. This completes the proof. ◀

Now that we know our particular values of M, $M_{V_1,\ldots,V_t}$, B, and p satisfy the conditions of Lemma 27, we apply the aforementioned lemma, obtaining that

$$\mathbb{P}\left[||R_H|| > 2t^t\left(et\left(\frac{\ln(8n^q/\epsilon)}{2q}+1\right)\right)^q n^{\frac{t-q}{2}}\right] < \epsilon\,.$$

as needed. ◀

5 Bounding the Norms of Uniform Low Degree Graph Matrices

In this section, we generalize our techniques further to prove our main result, Theorem 14, which we restate here.

▶ **Theorem 30.** *Let H be a graph with distinguished sets of vertices U and V such that U and V are disjoint and all vertices in $H(V)\setminus(U\cup V)$ have degree at least one. Let $t = |V(H)|$, let $z = |V(H)\setminus(U\cup V)|$, and let q be the size of the minimal separator between U and V.*

1. *If $q + z \geq 1$ then for all $\epsilon \in (0,1)$,*

$$\mathbb{P}\left[||R_H|| \geq 2(t^t)\left(e(t+z)\left(\frac{\ln(8n^q/\epsilon)}{2(q+z)}+1\right)\right)^{q+z} n^{\frac{t-q}{2}}\right] \leq \epsilon\,.$$

2. *If $q = z = 0$ then $||R_H|| \leq n^{\frac{t}{2}}$.*

Proof. For the second statement, note that if $q = z = 0$ then every entry of R_H has magnitude at most 1, so we can again use the fact that $||R_H|| \leq ||R_H||_{Fr} = \sqrt{\sum_{A,B} R_H(A,B)^2}$. For the first statment, similar to before, we first bound the norm of a closely related matrix where we restrict which entries the vertices of H can map into.

▶ **Definition 31.** Let $W = w_1,\ldots,w_z$ be the vertices of H outside of U and V. Given a partition $V_1,\ldots,V_t$ of $V(G)$, define $R_{H,V_1,\ldots,V_t}$ to be the $\binom{n}{x}\times\binom{n}{y}$ matrix with entries

$$R_{H,V_1,\ldots,V_t}(A,B) = \begin{cases} \sum_{C:\forall k, c_k\in V_{x+y+k}} \chi_{H,A,B,C} & A\cap B = \emptyset, \\ & \forall i\in[1,x], a_i\in V_i, \forall j\in[1,y], b_j\in V_{x+j} \\ 0 & otherwise \end{cases}$$

where the sum is over all $C = \{c_1,\ldots,c_z\}$ where the $c_1,\ldots,c_z$ are disjoint but not necessarily in order.

Let $R' = R_{H,V_1,\ldots,V_t}$. In order to find a probabilistic bound for $||R'||$, we bound $\mathbb{E}[\sqrt[2k]{\mathrm{tr}((R'R'^T)^k)}]$.

▶ **Lemma 32.** $\mathbb{P}\left[||R'|| \geq \left(e(t+z)\left(\frac{\ln(n^q/\epsilon)}{2(q+z)}+1\right)\right)^{q+z} n^{\frac{t-q}{2}}\right] \leq \epsilon.$

Proof.

▶ **Definition 33.** Define $S_{n,z}$ to be the set of all ordered tuples of z distinct elements of $[1,n]$.

▶ **Definition 34.** For all $A \in \binom{[n]}{x}$, $B \in \binom{[n]}{y}$, $C \in S_{n,z}$, define

$$Q(A,C,B) = \begin{cases} \chi_{H,A,B,C} & A \cap B = \emptyset, \\ & \forall i \in [1,x], a_i \in V_i, \forall j \in [1,y], b_j \in V_{x+j}, \forall k \in [1,z], c_k \in V_{x+y+k} \\ 0 & otherwise \end{cases}$$

Now note that

$$\mathbb{E}[\mathrm{tr}((R'R'^T)^k)] = \sum_{\substack{A_1,A_5,\ldots,A_{4k-3} \in \binom{[n]}{x} \\ B_3,B_7,\ldots,B_{4k-1} \in \binom{[n]}{y}}} \mathbb{E}\left[\prod_{j=1}^{k} R'(A_{4j-3},B_{4j-1})R'^T(B_{4j-1},A_{4j+1})\right]$$

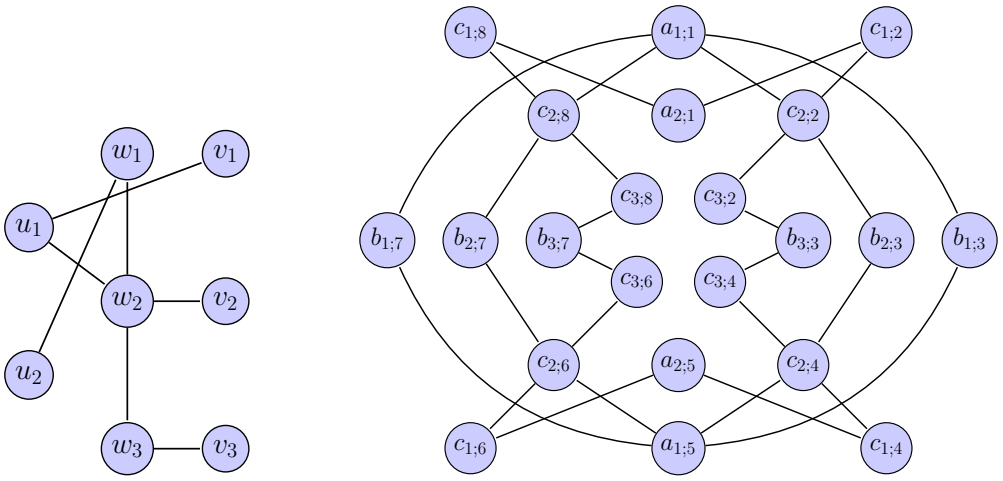
$$= \sum_{\substack{A_1,A_5,\ldots,A_{4k-3} \in \binom{[n]}{x} \\ B_3,B_7,\ldots,B_{4k-1} \in \binom{[n]}{y} \\ C_2,C_4,\ldots C_{4k} \in S_{n,z}}} \mathbb{E}\left[\prod_{j=1}^{k} Q(A_{4j-3},C_{4j-2},B_{4j-1})Q^T(B_{4j-1},C_{4j},A_{4j+1})\right]$$

by linearity of expectation, where $A_{4k+1} = A_1$.

Denote $\prod_{j=1}^{k} Q(A_{4j-3},C_{4j-2},B_{4j-1})Q^T(B_{4j-1},C_{4j},A_{4j+1})$ as $P(A_1,C_2,\ldots,B_{4k-1},C_{4k})$. Similar to before, because $\mathbb{E}[Q(A,C,B)] = 0$ for randomly chosen A, B, and C, the vast majority of the terms $\mathbb{E}[P(A_1,C_2,\ldots,B_{4k-1},C_{4k})]$ are 0; in fact, the only time the expected value can be non-zero is when each consecutive pair of sets of variables is disjoint and every edge of G involved in the product appears an even number of times, in which case the expected value is at most 1. Again, we can upper bound the number of choices for $A_1,C_2,\ldots,B_{4k-1},C_{4k}$ that yield a non-zero value for $\mathbb{E}[P(A_1,C_2,\ldots,B_{4k-1},C_{4k})]$ and use that number to bound $\mathbb{E}[\mathrm{tr}((M'M'^T)^k)]$. In order to represent $\mathbb{E}[P(A_1,C_2,\ldots,B_{4k-1},C_{4k})]$, we use another constraint graph. This constraint graph is similar to the earlier one; however, it is slightly more complicated, as there is an extra set of vertices involved.

In this constraint graph, there are $k(x+2z+y)$ vertices sorted into $4k$ sets. These vertices are labeled $A_1 = \{a_{1;1},a_{2;1},\ldots,a_{x;1}\}, C_2 = \{c_{1;2},c_{2;2},\ldots,c_{z;2}\}, B_3 = \{b_{1;3},b_{2;3},\ldots,b_{y;3}\}$, $C_4 = \{c_{1;4},c_{2;4},\ldots,c_{z;4}\}, A_5 = \{a_{1;5},a_{2;5},\ldots,a_{x;5}\},\ldots,C_{4k} = \{c_{1;4k},c_{2;4k}\ldots,c_{z;4k}\}$. Note that, in particular, the sets describing the sets of vertices are labeled A_1, C_2, B_3, C_4,..., and repeat this pattern exactly k times. Two vertices $a_{p;q}$ and $b_{r;s}$ are adjacent if and only if $|q-s| = 2$ and u_p and v_r are adjacent in H, where $a_{p;1} = a_{p;4k+1}$. Similarly, $a_{p;q}$ and $c_{t;o}$ are adjacent if and only if $|q-o| = 1$ and u_p and w_t are adjacent in H, and $b_{r;s}$ and $c_{t;o}$ are adjacent if and only if $|s-o| = 1$ and v_r and w_t are adjacent in H. Finally, $c_{t;o}$ and $c_{t';o'}$ are adjacent if and only if $o = o'$ and w_t and $w_{t'}$ are adjacent in H.

Now, in order to bound $\mathbb{E}[\mathrm{tr}((R'R'^T)^k)]$, we must calculate the minimum number of constraint edges required in our constraint graph to bring about a non-zero expectation value, as that will help bound the number of choices for $A_1,C_2,B_3,C_4,A_5,\ldots,B_{4k-1},C_{4k}$ that yield a non-zero expectation value for the product $\mathbb{E}[P(A_1,C_2,\ldots,B_{4k-1},C_{4k})]$.



(a) H.

(b) An example of the constraint graph for the given example of H, where $k = 2$.

Figure 4

▶ **Lemma 35.** *In order for* $\mathbb{E}[P(A_1, C_2, \ldots, B_{4k-1}, C_{4k})]$ *to have a non-zero value, there must be at least* $q(k-1) + zk$ *constraint edges in the respective constraint graph, where* q *is the maximal number of vertex-independent paths from* X *to* Y *in* H*. In addition, this bound is sharp.*

Proof. Choose q vertex-independent paths from U to V. Let l_i be the length of the ith such path.

Note that any path of length l_i in H from U to V corresponds to a cycle of size $2kl_i$ in our constraint graph; this cycle requires $kl_i - 1$ constraint edges by Proposition 17. Since the cycles are disjoint and by the definition of R' we cannot have constraint edges between them, the total number of constraint edges required by just the q vertex-independent paths is $\sum_{i=1}^{q}(kl_i - 1) = k(\sum_{i=1}^{q} l_i) - q$.

Consider the vertices in W which are not in the q paths. Of the z vertices in W, because each pair of paths is disjoint and a path of length l_i corresponds to exactly $l_i - 1$ vertices in C, exactly $\sum_{i=1}^{q}(l_i - 1) = (\sum_{i=1}^{q} l_i) - q$ vertices of C are included in paths, so there are $z - ((\sum_{i=1}^{q} l_i) - q)$ vertices of W not included in the paths. Each of these vertices is incident with an edge in H, and that edge is repeated $2k$ times in the constraint graph; therefore, as each edge of G that appears in the constraint graph must appear an even number of times in the constraint graph, any particular vertex of positive degree can, in all of its appearances in the constraint graph, only take on at most k values. However, the particular vertex appears $2k$ times in the constraint graph, once in each set of vertices of the form C_i, so it must have at least k constraint edges between those $2k$ appearances. Therefore, there are required to be a minimum of $(z - ((\sum_{i=1}^{q} l_i) - q))k + (k(\sum_{i=1}^{q} l_i) - q) = q(k-1) + zk$ constraint edges in the constraint graph.

In order to prove the sharpness, we utilize Menger's Theorem [21]. First, note that the existence of q vertex-independent paths from U to V implies that there exists q vertex-

independent paths from U to V, with first vertex in U, last vertex in V, and all internal vertices in W. This statement is true because given these q vertex-independent paths, if any of them have internal vertices in U or V, we can simply shorten these paths until they have only first and last vertices in U and V.

Now, Menger's Theorem states that because there are a maximum of q vertex-indepedent paths from U to V in H with all internal vertices in W, there is a set $S \in V(H)$ with $|S| = q$ such that all paths from U to V pass through at least one vertex of S. Consider such a set S. Using S, we will create $q(k-1) + zk$ constraint edges that yield a non-zero expectation value for $\mathbb{E}[P(A_1, C_2, \ldots, B_{4k-1}, C_{4k})]$ as follows.

For all vertices $u_i \in X$ such that $u_i \in S$, set $a_{i;1} = a_{i;5} = \cdots = a_{i;4k-3}$. Note that this requires $k-1$ constraint edges per vertex in S. Similarly, for all vertices $v_i \in Y$ such that $v_i \in S$, set $b_{i;3} = b_{i;7} = \cdots = b_{i;4k-1}$. In addition, for all vertices $w_i \in Z$ such that $w_i \in S$, set $c_{i;2} = c_{i;4} = \cdots = c_{i;4k}$. This requires $2k-1$ constraint edges per vertex in S.

Now, consider all vertices $w_i \in Z$ such that $w_i \notin S$. Note that if there existed a path from w_i to U that passed through no vertices in S, there cannot exist a path from w_i to V that passes through no vertices in S, as that would imply that there was a path from U to V not passing through any vertices on S. So, if there exists a path from w_i to U passing through no vertices of S, set $c_{i;4} = c_{i;6}, c_{i;8} = c_{i;10}, \ldots, c_{i;4k} = c_{i;2}$. Otherwise, set $c_{i;2} = c_{i;4}, c_{i;6} = c_{i;8}, \ldots, c_{i;4k-2} = c_{i;4k}$. This requires k constraint edges per vertex.

Now given an edge in the constraint graph $(a_{p;q}, b_{r;s})$ with $|q-s| = 2$, then either u_p or v_r is in S or else there would exist a path from U to V not in S; therefore, either $(a_{p;q}, b_{r;q+2}) = (a_{p;q}, b_{r;q-2})$ or $(a_{p;s-2}, b_{r;s}) = (a_{p;s+2}, b_{r;s})$ by the constraint edges. Similarly, given $(a_{p;q}, c_{t;o})$ with $|p-o| = 1$, either $c_{t;q-1} = c_{t;q+1}$, which implies $(a_{p;q}, c_{t;q-1}) = (a_{p;q}, c_{t;q+1})$, or $w_t \notin S$ and $u_p \in S$, which implies $(a_{p;2o-q-2}, c_{t;2o-q-1}) = (a_{p;2o-q+2}, c_{t;2o-q+1})$. A similar argument applies to edges of the form $(c_{t;o}, b_{r;s})$. For edges of the form $(c_{t;o}, c_{t';o})$, it can be shown that either $(c_{t;o}, c_{t';o}) = (c_{t;o-2}, c_{t';o-2})$ or $(c_{t;o}, c_{t';o}) = (c_{t;o+2}, c_{t';o+2})$. Finally, note that edges in the constraint graph of the form $(a_{p_1;q}, a_{p_2;q})$ and $(b_{r_1;s}, b_{r_2;s})$ are automatically doubled and have no effect. Thus, this construction makes every edge appear with an even multiplicity, as needed.

If S contains exactly j vertices in W, then the total number of constraint edges used in this construction is $(k-1)(|S|-j) + (2k-1)j + (k)(z-j) = q(k-1) + zk$, meaning that the bound given is sharp. ◀

▶ **Corollary 36.** *Let N represent the number of choices for A_1, C_2, …, B_{4k-1}, B_{4k} such that*
$\mathbb{E}[P(A_1, C_2, \ldots, B_{4k-1}, C_{4k})] \neq 0$. *Then, $N \leq ((t+z)k)^{2k(z+q)-2q} n^{(t-q)k+q}$.*

Proof. Apply Proposition 18. In this situation, $b = k(t+z)$ and $c = q(k-1) + zk$. This implies the desired result. ◀

▶ **Corollary 37.** $\mathbb{E}[\mathrm{tr}((R'R'^T)^k)] \leq ((t+z)k)^{2k(z+q)-2q} n^{(t-q)k+q}$.

Proof. Recall that

$$\mathbb{E}[\mathrm{tr}((R'R'^T)^k)] = \sum_{\substack{A_1, A_5, \ldots, A_{4k-3} \in S_{n,u} \\ B_3, B_7, \ldots, B_{4k-1} \in S_{n,v} \\ C_2, C_4, \ldots C_{4k} \in S'_{n,w}}} \mathbb{E}\left[\prod_{j=1}^{k} Q(A_{4j-3}, C_{4j-2}, B_{4j-1}) Q^T(B_{4j-1}, C_{4j}, A_{4j+1})\right].$$

Then, by Proposition 36, there are at most $((t+z)k)^{2k(z+q)-2q} n^{(t-q)k+q}$ choices for $A_1, C_2, \ldots, B_{4k-1}, C_{4k}$ that yield a non-zero value for

$\mathbb{E}\left[\prod_{j=1}^{k} Q(A_{4j-3}, C_{4j-2}, B_{4j-1})Q^T(B_{4j-1}, C_{4j}, A_{4j+1})\right]$; in addition, $\mathbb{E}\left[\prod_{j=1}^{k} Q(A_{4j-3}, C_{4j-2}, B_{4j-1})Q^T(B_{4j-1}, C_{4j}, A_{4j+1})\right] \leq 1$. These two observations complete the proof. ◀

Now for any graph G on n vertices, $\text{tr}((R'R'^T)^k)$ must take on a nonnegative value. By Markov's inequality and Corollary 37, for all $\epsilon \in (0,1)$

$$\mathbb{P}\left[\text{tr}((R'R'^T)^k) \geq \frac{\mathbb{E}[\text{tr}((R'R'^T)^k)]}{\epsilon}\right] \leq \mathbb{P}\left[\text{tr}((R'R'^T)^k) \geq ((t+z)k)^{2k(z+q)-2q} n^{(t-q)k+q}/\epsilon\right] \leq \epsilon.$$

Since $||R'|| \leq \sqrt[2k]{\text{tr}((R'R'^T)^k)}$, this implies that for all $k \geq 1$ and all $\epsilon \in (0,1)$,

$$\mathbb{P}\left[||R'|| \geq \sqrt[2k]{((t+z)k)^{2k(z+q)-2q} n^{(t-q)k+q}/\epsilon}\right] \leq \mathbb{P}\left[||R'|| \geq ((t+z)k)^{z+q} n^{\frac{t-q}{2}} (n^q/\epsilon)^{1/2k}\right] \leq \epsilon.$$

Setting $k = \lceil \frac{1}{2(q+z)} \ln(n^q/\epsilon) \rceil$ we have that

$$((t+z)k)^q n^{\frac{t-q}{2}} (n^q/\epsilon)^{1/2k} \leq \left((t+z)\left(\frac{\ln(n^q/\epsilon)}{2(q+z)} + 1\right)\right)^{q+z} n^{\frac{t-q}{2}} e^{q+z}.$$

Therefore, $\mathbb{P}[||R'|| \geq \left(e(t+z)\left(\frac{\ln(n^q/\epsilon)}{2(q+z)} + 1\right)\right)^{q+z} n^{\frac{t-q}{2}}] \leq \epsilon$, as needed. ◀

Using Lemma 27 with $p = \epsilon$ and $B = 2(t^t)\left(e(t+z)\left(\frac{\ln(8n^q/\epsilon)}{2(q+z)} + 1\right)\right)^{q+z} n^{\frac{t-q}{2}}$ and following the same logic as in the proof of Theorem 21, we obtain that for all $\epsilon \in (0,1)$,

$$\mathbb{P}[||R_H|| \geq 2(t^t)\left(e(t+z)\left(\frac{\ln(8n^q/\epsilon)}{2(q+z)} + 1\right)\right)^{q+z} n^{\frac{t-q}{2}}] \leq \epsilon\,.$$ ◀

6 Lower Bounds

In this section, we show that the bounds we have obtained on the norms of uniform low degree graph matrices are tight up to a factor of polylog(n). This makes intuitive sense as our bounds on $E[tr((R'R'^T)^k)]$ were tight up to a polylog factor. Unfortunately, these lower bounds on $E[tr((R'R'^T)^k)]$ are insufficient for two reasons. First, they do not rule out the possibility that $||R'||$ is sometimes very small. Second, lower bounds on the norms of the matrices R' do not imply a lower bound on $||R_H||$. We now show how these obstacles can be overcome (though we only give a proof sketch as a full proof would be long and technical).

▶ **Theorem 38.** *Let H be a graph with distinguished sets of vertices U and V where U and V are disjoint and for all vertices w in $V(H) \setminus (U \cup V)$, there is a path from w to either U or V in H. Letting $t = |V(H)|$ and letting q be the minimal size of a vertex separator between U and V, with high probability $||R_H||$ is $\Omega(n^{\frac{t-q}{2}})$.*

▶ Remark. If H has non-isolated vertices which are not connected to U or V, there is a non-negligible chance that R_H has considerably smaller norm than expected. To see this, let H_0 be the part of H which is connected to U and/or V and let H_1 be the remainder of H. We have that $R_H \approx R_{H_1} R_{H_0}$ where R_{H_1} is a constant depending on the input graph G which has some chance of being close to 0.

Before giving a proof sketch for Theorem 38, we first consider the case when H is bipartite with partite sets U and V, which can be analyzed directly.

▶ **Theorem 39.** *Let H be a bipartite graph with partite sets U and V. Letting $t = |V(H)|$ and letting q be the minimal size of a vertex cover of H, $||R_H||$ is $\Omega(n^{\frac{t-q}{2}})$.*

Proof. We show this by finding vectors u and v such that $u^T R_H v$ is $\Theta(n^{\frac{t-q}{2}})||u|| \cdot ||v||$. The idea is to take a minimal vertex cover S of U and V and fix the vertices that S maps to. This essentially separates the dependence of $R_H(A, B)$ on A and B which means that we can choose u to match the dependence on A and choose v to match the dependence on B.

▶ **Definition 40.**

1. Define $E_L \subseteq E(H)$ to be the edges in H between $U \setminus S$ and $S \cap V$.
2. Define $E_M \subseteq E(H)$ to be the edges in H between $S \cap U$ and $S \cap V$
3. Define $E_R \subseteq E(H)$ to be the edges in H between $S \cap U$ and $V \setminus S$.

Now choose disjoint vertices $A_S \cup B_S$ for the vertices of S to map into and define u and v as follows.

▶ **Definition 41.** Given an A which is disjoint from B_S, letting π be the map which maps U into A and maps $S \cap V$ into B_S, define u_A to be $\chi_{\pi(E_L)}$ if $\pi(S \cap U) = A_S$ and 0 otherwise.

Given a B which is disjoint from A_S, letting π be the map which maps V into B and maps $S \cap U$ into A_S, define v_B to be $\chi_{\pi(E_R)}$ if $\pi(S \cap V) = B_S$ and 0 otherwise.

Note that $u_A R_H(A, B) v_B$ is only nonzero if the following conditions hold

1. A and B are disjoint
2. If π is the map that maps U into A and V into B then $\pi(S \cap U) = A_S$ and $\pi(S \cap V) = B_S$.

When these conditions hold, $u_A R_H(A, B) v_B = \chi_{\pi(E_L)} \chi_{\pi(E(H))} \chi_{\pi(E_R)} = \chi_{\pi(E_M)}$ which will always be the same as A_S and B_S are fixed. There are $\Theta(n^{|U|-|S\cap U|})$ A such that $u_A \neq 0$, there are $\Theta(n^{|V|-|S\cap V|})$ B such that $v_B \neq 0$, and there are $\Theta(n^{|U|-|S\cap U|} n^{|V|-|S\cap V|}) = \Theta(n^{t-q})$ choices for A and B for which these conditions hold. This implies that $||u||$ is $\Theta(n^{\frac{|U|-|S\cap U|}{2}})$, $||v||$ is $\Theta(n^{\frac{|V|-|S\cap V|}{2}})$, and $|u^T R_H v|$ is $\Theta(n^{t-q})$. Putting everything together, $|u^T R_H v|$ is $\Theta(n^{\frac{t-q}{2}})||u|| \cdot ||v||$, so $||R_H||$ is $\Omega(n^{\frac{t-q}{2}})$, as needed. ◀

In the general case, we could use similar ideas, choosing a minimal vertex seperator S of U and V in H, fixing the vertices that S maps to, and then choosing u to match the dependence on A and v to match the dependence on B. However, the analysis is tricky for two reasons. First, it is non-trivial to bound $||u||$ and $||v||$ as the entries of u and v are the sums of many terms. Second, $u^T R_H v$ will have additional terms coming from different choices for the vertices in $S \setminus (U \cup V)$ in the sums describing the entries of R_H. To deal with these issues, we use an argument involving $||R_H||_{Fr}$, the Frobenius norm of R_H, which is somewhat cleaner to analyze.

Proof Sketch of Theorem 38.

▶ **Definition 42.** Given two matrices M_1 and M_2 with the same dimensions, we define $\langle M_1, M_2 \rangle = \sum_{i,j} M_1(i,j) M_2(i,j)$. Note that for any matrix M, $\langle M, M \rangle = ||M||_{Fr}^2$

Given a matrix M, we can bound $||M||$ as follows.

▶ **Proposition 43.** *If $M = \sum_i c_i u_i v_i^T$ for some vectors u_i, v_i and some positive coefficients c_i then $||M|| \geq \frac{||M||_{Fr}^2}{\sum_i c_i ||u_i|| \cdot ||v_i||}$*

Proof. Note that

$$||M||_{Fr}^2 = \langle M, M\rangle = \langle M, \sum_i c_i u_i v_i^T\rangle = \sum_i c_i u_i^T M v_i \leq ||M|| \sum_i c_i ||u_i|| \cdot ||v_i|| \quad \blacktriangleleft$$

With this proposition in mind, we first describe how we can decompose R_H. We then describe how to bound $||R_H||_{Fr}^2$ and the norms of the vectors involved.

▶ **Definition 44.** Given a graph H where all vertices are connected to U or V and a minmal separator S of U and V,

1. Let L be the set of vertices which are connected to U once we remove S from H
2. Let R be set of vertices connected to V once we remove S from H.
3. Let H_L be the graph with vertices $L \cup S$ and all edges in H which are incident with a vertex in L.
4. Let H_R be the graph with vertices $R \cup S$ and all edges in H which are between two vertices in S or are incident with a vertex in R.
5. Let $l = |L|$, $q = |S|$, and $r = |R|$. Note that $t = l + q + r$.

▶ **Proposition 45.**

$$R_{H,V_1,\ldots,V_t} = \sum_{\substack{v_{l+1},\ldots,v_{l+q}: \\ \forall i \in [l+1,l+q], v_i \in V_i}} R_{H_L,V_1,\ldots,V_l,\{v_{l+1}\},\ldots,\{v_{l+q}\}} R_{H_R,\{v_{l+1}\},\ldots,\{v_{l+q}\},V_{l+q+1},\ldots,V_t}$$

where we take $V = \emptyset$ for H_L and we take $U = \emptyset$ for H_R (so $R_{H_L,V_1,\ldots,V_l,\{v_{l+1}\},\ldots,\{v_{l+q}\}}$ and $R^T_{H_R,\{v_{l+1}\},\ldots,\{v_{l+q}\},V_{l+q+1},\ldots,V_t}$ are in fact vectors)

Proof. This proposition follows directly from the definitions of the matrices involved. ◀

Writing $R_H = \frac{t^t}{t^n} \sum_{V_1,\ldots,V_t} R_{H,V_1,\ldots,V_t}$, we have that $R_H = \sum_{c_i} u_i v_i^T$ where each vector u_i is of the form $R_{H_L,V_1,\ldots,V_l,\{v_{l+1}\},\ldots,\{v_{l+q}\}}$ and each vector v_i is of the form $R^T_{H_R,\{v_{l+1}\},\ldots,\{v_{l+q}\},V_{l+q+1},\ldots,V_t}$. Note that the sum of the coefficients of the vector products in this sum is $O(n^q)$. We now probabilistically bound $||R_H||_{Fr}^2$, $||u_i||$, and $||v_i||$.

▶ **Lemma 46.** *Let H_1 and H_2 be two graphs such that both H_1 and H_2 have distinguished sets of vertices U and V, U and V are disjoint, every vertex in H_1 is connected to a vertex in U or V, and every vertex in H_2 is connected to a vertex in U or V. Let $t_1 = |V(H_1)|$ and let $t_2 = |V(H_2)|$.*

1. *If $H_1 = H_2 = H$ (after permuting the vertices not in $U \cup V$) then letting $t = t_1 = t_2$, $E[\langle R_H, R_H\rangle]$ is $\Theta(n^t)$ and with high probability, $\langle R_H, R_H\rangle - E[\langle R_H, R_H\rangle]$ is $O(n^{t-\frac{1}{2}} polylog(n))$*
2. *If H_1 and H_2 are different graphs then with high probability, $\langle R_{H_1}, R_{H_2}\rangle$ is $O(n^{\frac{t_1+t_2-1}{2}} polylog(n))$*

Proof. Consider the terms in $\langle R_{H_1}, R_{H_2}\rangle$. Each such term is determined by mappings $\pi_1 : V(H_1) \to G$, $\pi_2 : V(H_2) \to G$ where $\pi_2(U) = \pi_1(U)$, $\pi_2(V) = \pi_1(V)$, and these maps preserve the ordering of U and V. For a given term, let H' be the graph with vertices $\pi_1(V(H_1)) \cup \pi_2(V(H_2))$ and edges $\pi_1(E(H_1))\Delta\pi_2(E(H_2))$ where Δ is the symmetric difference. If we group the terms with the same graph H' (up to a permutation of the vertices) together, then we obtain sums of the following form.

▶ **Definition 47.** Given a graph H' with two distinguished subsets of vertices U and V, define $f_{H'}(G) = \sum_{A,B} R_{H'}(A, B)$

We have the following probabilistic bound on these sums.

▶ **Lemma 48.** *If x is the number of non-isolated vertices in H' and y is the number of isolated vertices in H', with high probability $f_{H'}(G)$ is $O(n^{\frac{x}{2}+y}polylog(n))$*

Proof Sketch. We show the result for a related function $f_{H',V_1,\dots,V_m}(G)$ where we restrict where the vertices of H' map to. To rigorously show the result, we would then use Lemma 27 (which applies just as well to scalars).

▶ **Definition 49.** Letting $t' = |V(H')|$ and given disjoint sets of vertices $V_1, \dots, V_m$, define $f_{H',V_1,\dots,V_{t'}}(G) = \sum_{A,B} R_{H',V_1,\dots,V_{t'}}(A,B)$

▶ **Lemma 50.** *If x be the number of non-isolated vertices in H' and y is the number of isolated vertices in H;, with high probability $f_{H',V_1,\dots,V_m}(G)$ is $O(n^{\frac{x}{2}+y})$*

Each term in $f_{H',V_1,\dots,V_m}(G)$ can be described by choosing a vertex v_i from each V_i. To prove this result, we start by considering each term individually, thinking of all the vertices v_i as being fixed. We then iteratively group these terms together by choosing one or two vertices which are still fixed and summing over all possibilities for these vertices.

When we sum over the possibilities for some vertex v_i, we randomly fix all of the edges of G which are incident with a vertex in V_i and call v_i free. By doing this, we always know the magnitudes of all our sums (but not their signs!). At each point, we have the following bound.

▶ **Lemma 51.** *We can choose the order in which we make vertices free so that if we have made x_1 non-isolated vertices free and y_1 isolated vertices free then with high probability our sums are all $O(n^{\frac{x_1}{2}+y_1}polylog(n))$. Moreover, at all times, for every non-isolated vertex v_i, there is an edge between v_i and another fixed vertex v_j in $\pi(E(H))$ (where π is the mapping from $V(H)$ to $V(G)$).*

Proof. We prove this result by induction. The base case $x_1 = y_1 = 0$ is trivial. If there is an isolated vertex which is fixed, there are at most n possibilities for this vertex, so this grouping increases y_1 by 1 and multiplies our bound by a factor of at most $O(n)$. If there is a non-isolated fixed vertex v_i which can be made free while maintaining the invariant that every fixed vertex has an edge to another fixed vertex in $\pi(E(H))$, sum over the possibilities for this vertex. In this sum, we know the magnitude of each term, but the signs of each term are random and completely independent from each other (as they depend on edge(s) between the different possibilities for v_i and other fixed vertices). This summation increases x_1 by 1 and since the signs are independent, with high probability this summation increases our bound by a factor of at most $O(\sqrt{n}log(n))$. The remaining case is when $\pi(E(H))$ contains a perfect matching between the fixed vertices and no other edges between fixed vertices. In this case, choose one such edge (v_i, v_j) and sum up over the possibilities for v_i and v_j. Again, we know the magnitudes of each term in this sum, but the signs of each term are independent (as they depend on the different edges (v_i, v_j)). This summation increases x_1 by 2 and since the signs are independent, with high probability this summation increases our bound by a factor of at most $O(nlog(n))$. ◀

▶ Remark. The reason we needed to use $f_{H',V_1,\dots,V_{t'}}(G)$ rather than $f_{H'}(G)$ in this argument is that it allowed us to consistently choose which edges of G we fixed and which edges of G were still undetermined at any given point. ◀

We now bound the terms in $\langle R_{H_1}, R_{H_2} \rangle$ using Lemma 48. We consider all of the terms $f_{H'}$ which appear in this sum. If $\pi_1(H_1) = \pi_2(H_2)$ then H' consists of t_1 isolated vertices and we can see directly that $f_{H'}(G)$ is $\Theta(n^{t_1})$.

For a given $f_{H'}$ such that $\pi_1(H_1) \neq \pi_2(H_2)$, let $r = |\pi_1(V(H_1)) \cap \pi_2(V(H_2))|$. Letting x be the number of non-isolated vertices in H' and y be the number of isolated vertices in H', we have that $x + y = |V(H')| = t_1 + t_2 - |U| - |V| - r$. Note that only the vertices in $\pi_1(U) \cup \pi_1(V) \cup \pi_1(V(H_1)) \cap \pi_2(V(H_2))$ can be isolated in H' because all other vertices are incident with an edge in $\pi_1(E(H_1)) \cup \pi_2(E(H_2))$ which only appears once. This tells us that $y \leq |U| + |V| + r$. Applying Lemma 48, we have that $f_{H'}$ is $O(n^{\frac{x}{2}+y} polylog(n)) = O(n^{\frac{(x+y)+y}{2}} polylog(n))$ which is $O(n^{\frac{t_1+t_2}{2}} polylog(n))$.

To improve this bound and obtain our result, it is sufficient to show that if $\pi_1(H_1) \neq \pi_2(H_2)$, there must be some vertex in $\pi_1(U) \cup \pi_1(V) \cup \pi_1(V(H_1)) \cap \pi_2(V(H_2))$ which is not isolated, as this reduces our upper bound on y by 1. To show this, first note that if $\pi_1(V(H_1)) = \pi_2(V(H_2))$ yet $\pi_1(H_1) \neq \pi_2(H_2)$ then H' must have an edge so not all of the vertices of H' can be isolated. If $\pi_1(V(H_1)) \neq \pi_2(V(H_2))$ then either $\pi_1(V(H_1)) \setminus \pi_2(V(H_2))$ is nonempty or $\pi_2(V(H_2)) \setminus \pi_1(V(H_1))$ is nonempty. Without loss of generality, we may assume that $\pi_1(V(H_1)) \setminus \pi_2(V(H_2))$ is nonempty. Let v_i be a vertex in $\pi_1(V(H_1)) \setminus \pi_2(V(H_2))$. Since every vertex in H_1 is connected to either U or V, if we look at the edges $\pi_1(E(H_1)) \cup \pi_2(E(H_2))$, there must be a path from v_i to some vertex v_j in $\pi_1(U \cup V)$. There must be an edge in this path between a vertex in $\pi_1(V(H_1)) \setminus \pi_2(V(H_2))$ and a vertex in $\pi_1(U) \cup \pi_1(V) \cup \pi_1(V(H_1)) \cap \pi_2(V(H_2))$, let (v_k, v_l) be the first such edge. $v_l \in \pi_1(U) \cup \pi_1(V) \cup \pi_1(V(H_1)) \cap \pi_2(V(H_2))$ and cannot be isolated in H', so the result follows. ◀

We now give a sketch for how to probabilistically bound the norms of the vectors u_i and the vectors v_i and complete the proof. Recall that each vector u_i is of the form $R_{H_L,V_1,\ldots,V_{l+q}}$ where the V for H_L is empty and $V_{l+1}, \ldots, V_{l+q}$ have size 1. We now use similar reasoning as we used to prove Lemma 46 except that there is only one possibility for the vertices corresponding to $V_{l+1}, \ldots, V_{l+q}$ so we always keep these vertices fixed (and don't sum over their possibilities). Applying this reasoning, we obtain that for each i, with high probability, $||u_i||^2$ is $\Theta(n^l)$ so $||u_i||$ is $\Theta(n^{\frac{l}{2}})$. By symmetry, for each i, with high probability $||v_i||$ is $\Theta(n^{\frac{r}{w}})$. Putting everything together, with high probability $||R_H||_{Fr}^2$ is $\Theta(n^t)$ and $\sum_i c_i ||u_i|| \cdot ||v_i||$ is $O(n^{\frac{l+r}{2}+q}) = O(n^{\frac{t+s}{2}})$. $||R_H|| \geq \frac{||R_H||_{Fr}^2}{\sum_i c_i ||u_i|| \cdot ||v_i||}$, which is $\Omega(n^{\frac{t-q}{2}})$, as needed. ◀

▶ Remark. While Theorem 38 was only stated for a single R_H, the techniques used to prove Theorem 38 apply just as well to a linear combination of such matrices. In particular, if we have distinct graphs $H_1, \ldots, H_k$ which all satisfy the conditions of Theorem 38 then for any coefficients $c_1, \ldots, c_k$, with high probability, $||\sum_{i=1}^k c_k R_{H_k}||$ is $\Theta(\max_i\{||c_i R_{H_i}||\})$

7 Conclusion and Further Studies

In this paper, we analyzed the norms of uniform low degree graph matrices, which appear naturally when analyzing the sum of squares hierarchy. While special cases of these matrices were analyzed in previous works on sum of squares lower bounds for the planted clique problem [20], we generalized this analysis, proving an upper bound on the norms of all such matrices which is tight up to a polylogarithmic factor. This general analysis is a key component of the work [4] proving almost tight lower bounds for sum of squares on planted clique and will very likely be useful for further analysis of the sum of squares hierarchy.

That said, there are several open problems raised by this work. First, to what extent can these norm bounds be improved? It is very likely that with a more careful analysis, the polylog factors can be reduced or removed and the dependence of $||R_H||$ on the size of the graph H can be improved. Can we go further and determine the distribution of these

matrices' eigenvalues? Second, what can we say about the non-uniform case? How much structure do we need our matrices to have to obtain interesting norm bounds?

Acknowledgements. This research was supported by the Program for Research in Mathematics, Engineering, and Science (PRIMES-USA) at MIT. We thank the PRIMES-USA faculty, including Dr. Tanya Khovanova, Dr. Pavel Etingof, and Dr. Slava Gerovitch for helpful feedback on this work.

References

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

2 Sanjeev Arora, Boaz Barak, and David Steurer. Subexponential algorithms for unique games and related problems. In *Foundations of Computer Science (FOCS), 2010 51st Annual IEEE Symposium on*, pages 563–572. IEEE, 2010.

3 Sanjeev Arora, Satish Rao, and Umesh Vazirani. Expander flows, geometric embeddings and graph partitioning. *Journal of the ACM (JACM)*, 56(2):5, 2009.

4 Boaz Barak, Sam Hopkins, Jonathan Kelner, Ankur Moitra, Pravesh Kothari, and Aaron Potechin. A Nearly Tight Sum-of-Squares Lower Bound for the Planted Clique Problem. *ArXiv e-prints*, April 2016. `arXiv:1503.06447`.

5 Boaz Barak, Prasad Raghavendra, and David Steurer. Rounding semidefinite programming hierarchies via global correlation. In *Foundations of Computer Science (FOCS), 2011 IEEE 52nd Annual Symposium on*, pages 472–481. IEEE, 2011.

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

7 Yash Deshpande and Andrea Montanari. Improved sum-of-squares lower bounds for hidden clique and hidden submatrix problems. *CoRR*, abs/1502.06590, 2015. URL: `http://arxiv.org/abs/1502.06590`.

8 Uriel Feige and Robert Krauthgamer. The probable value of the lovász–schrijver relaxations for maximum independent set. *SIAM J. Comput.*, 32(2):345–370, 2003. `doi:10.1137/S009753970240118X`.

9 Vitaly Feldman, Elena Grigorescu, Lev Reyzin, Santosh Vempala, and Ying Xiao. Statistical algorithms and a lower bound for detecting planted cliques. In *Proceedings of the forty-fourth annual ACM symposium on Theory of Computing*. ACM, 2013.

10 Vyacheslav L. Girko. Circular law. *Theory of Probability and its Applications*, 29:694–706, 1984.

11 Michel X Goemans and David P Williamson. Improved approximation algorithms for maximum cut and satisfiability problems using semidefinite programming. *Journal of the ACM (JACM)*, 42(6):1115–1145, 1995.

12 Dima Grigoriev. Complexity of positivstellensatz proofs for the knapsack. *Computational Complexity*, 10(2):139–154, 2001.

13 Dima Grigoriev. Linear lower bound on degrees of positivstellensatz calculus proofs for the parity. *Theoretical Computer Science*, 259(1):613–622, 2001.

14 Venkatesan Guruswami and Ali Kemal Sinop. Lasserre hierarchy, higher eigenvalues, and approximation schemes for graph partitioning and quadratic integer programming with psd objectives. In *Foundations of Computer Science (FOCS), 2011 IEEE 52nd Annual Symposium on*, pages 482–491. IEEE, 2011.

15 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. URL: `http://arxiv.org/abs/1507.05230`.

16 Mark Jerrum. Large cliques elude the metropolis process. *Random Structures & Algorithms*, 3(4):347–359, 1992.

17 Denés Konig. Gráfok és mátrixok. matematikai és fizikai lapok, 38: 116–119, 1931.

18 Luděk Kučera. Expected complexity of graph partitioning problems. *Discrete Applied Mathematics*, 57(2):193–212, 1995.

19 Jean B Lasserre. Global optimization with polynomials and the problem of moments. *SIAM Journal on Optimization*, 11(3):796–817, 2001.

20 R. Meka, A. Potechin, and A. Wigderson. Sum-of-squares lower bounds for planted clique. *ArXiv e-prints*, March 2015. `arXiv:1503.06447`.

21 Karl Menger. Zur allgemeinen kurventheorie. *Fundamenta Mathematicae*, 10(1):96–115, 1927.

22 Yurii Nesterov. Squared functional systems and optimization problems. In *High performance optimization*, pages 405–440. Springer, 2000.

23 Pablo A Parrilo. *Structured semidefinite programs and semialgebraic geometry methods in robustness and optimization*. PhD thesis, Citeseer, 2000.

24 Prasad Raghavendra and Tselil Schramm. Tight lower bounds for planted clique in the degree-4 SOS program. *CoRR*, abs/1507.05136, 2015. URL: `http://arxiv.org/abs/1507.05136`.

25 Grant Schoenebeck. Linear level lasserre lower bounds for certain k-csps. In *Foundations of Computer Science, 2008. FOCS'08. IEEE 49th Annual IEEE Symposium on*, pages 593–602. IEEE, 2008.

26 NZ Shor. Class of global minimum bounds of polynomial functions. *Cybernetics and Systems Analysis*, 23(6):731–734, 1987.

27 Eugene P Wigner. On the distribution of the roots of certain symmetric matrices. *Annals of Mathematics*, pages 325–327, 1958.

A Justification of the moment method

In this appendix, we provide a proof that the moment method gives an upper bound on the norm. To make the proof easier, we consider the case of positive semidefinite matrices separately.

▶ **Lemma 52.**
1. *For any positive semidefinite matrix A, for all $k \geq 1$, $\sqrt[k]{\mathrm{tr}(A^k)} \geq ||A||$.*
2. *For any real matrix M, for all $k \geq 1$, $\sqrt[2k]{\mathrm{tr}((MM^T)^k)} \geq ||M||$.*

Proof. To show the first statement, we recall the following fact about positive semidefinite matrices.

▶ **Proposition 53.** *For any positive semidefinite matrix A, $||A|| = \lambda_{max}(A)$ where $\lambda_{max}(A)$ is the maximum eigenvalue of A.*

Proof. Since A is positive semidefinite, there is an orthonomal basis $v_1, \cdots, v_n$ of eigenvectors of A with eigenvalues $\lambda_1, \cdots, \lambda_n$. Without loss of generality, we may assume that $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_n \geq 0$. Given a unit vector v, $v = \sum_{i=1}^n c_i v_i$ where $\sum_{i=1}^n c_i^2 = 1$. $Av = \sum_{i=1}^n \lambda_i c_i v_i$ so we have that

$$||Av||^2 \leq \sum_{i=1}^n \lambda_i^2 c_i^2 \leq \sum_{i=1}^n \lambda_1^2 c_i^2 = \lambda_1^2 .$$

This implies that $||A|| \leq \lambda_1$. $||Av_1|| = \lambda_1$ so $||A|| \geq \lambda_1$ and thus $||A|| = \lambda_1$, as needed. ◀

With this in mind, for any $k \geq 1$, the eigenvalues of A^k are $\lambda_1^k, \cdots, \lambda_n^k$. Thus,

$$trace(A^k) = \sum_{i=1}^{n} \lambda_i^k \geq \lambda_1^k \geq ||A||^k$$

and the first statement follows. The second statement follows immediately from the first statement and the following proposition.

▶ **Proposition 54.** *For any matrix M, $||MM^T|| = ||M||^2$.*

Proof. Note that for any matrix M, $||M^T|| = ||M|| = \max_{||u||=1,||v||=1} u^T M v$. For any unit vectors u and v,

$$u^T M M^T v = (M^T u)^T (M^T v) = M^T u \cdot M^T v \leq ||M^T u|| \cdot ||M^T v|| \leq ||M^T||^2 = ||M||^2 .$$

Thus, $||M^T M|| \leq ||M||^2$. On the other hand, letting v be a unit vector which maximizes $||M^T v||$, giving $||M^T v|| = ||M^T|| = ||M||$, we have that

$$v^T M M^T v = (M^T v)^T (M^T v) = ||M^T v||^2 = ||M^T||^2 = ||M||^2 .$$

Thus, $||M^T M|| \geq ||M||^2$ and the result follows. ◀◀

B Norm bounds with left/right intersections

In this appendix, we consider the case when $U \cap V$ is nonempty in H.

▶ **Theorem 55.** *Let H be a graph with distinguished sets of vertices U and V such that $|U \cap V| = r$ and all vertices in $H(V) \setminus (U \cup V)$ have degree at least one. Let $t = |V(H)|$, let $z = |V(H) \setminus (U \cup V)|$, and let q be the size of the minimal separator between U and V (which must include $U \cap V$). Let $t' = t - r$ and let $q' = q - r$.*

1. *If $q' + z > 0$ then for all $\epsilon \in (0, 1)$,*

$$\mathbb{P}[||R_H|| \geq 2(t')^{t'} \left(e(t'+z) \left(\frac{\ln(8n^{q'+r}/\epsilon)}{2(q'+z)} + 1 \right) \right)^{q'+z} n^{\frac{t'-q'}{2}}] \leq \epsilon .$$

2. *If $q' = z = 0$ then $||R_H|| \leq n^{\frac{t}{2}}$.*

Proof Sketch. The key idea is to reduce the case when to the case when $U \cap V = \emptyset$. We first choose the elements which $U \cap V$ map to. These will be the elements of $A \cap B$ and they must occur in fixed positions within A and B. Thus, this choice partitions R_H into blocks which share no rows or columns. It is now sufficient to obtain a probabilistic bound for each block and use a union bound.

For a particular block, we can now use the same proof we used to prove Theorem 14. Let R' be a matrix where we have restricted ourselves to a particular block and have partitioned the vertices in $[1, n] \setminus (A \cap B)$ into $V_1, \ldots, V_{t'}$, restricting where the vertices in $V(H) \setminus (U \cap V)$ can map to accordingly. Consider $tr((R'R'^T)^k)$. Roughly speaking, we ignore the vertices in $A \cap B$ (in fact we could replace n by $n' = n - r$ in the bound). Since the vertices in $A \cap B$ are fixed, they appear in every copy of R' and are already determined, so they do not contribute to the number of terms with nonzero expectation.

For the other vertices, it is still true that if we have vertex-independent paths of lengths $l_1, \ldots, l_{q'}$ in H from $U \setminus V$ to $V \setminus U$, in the constraint graph the path of length l_i becomes a

cycle of length $2kl_i$, requiring $kl_i - 1$ constraint edges. It is also still true that every vertex in W which is not on one of these paths requires k constraint edges. Thus, we have the same norm bound on each block as we did in Theorem 14. To take a union bound over all the blocks, since there are at most n^r blocks, we use this bound with ϵ replaced by $\frac{\epsilon}{n^r}$ and the result follows. ◀