Regularity and removal lemmas and their applications

by

László Miklós Lovász

Submitted to the Department of Mathematics
in partial fulfillment of the requirements for the degree of
Doctor of Philosophy

at the

MASSACHUSETTS INSTITUTE OF TECHNOLOGY

June 2017

© Massachusetts Institute of Technology 2017. All rights reserved.

Signature redacted

Author .................................................................

Department of Mathematics

May 5, 2017

Signature redacted

Certified by..... .........................................................

Jacob Fox
Professor of Mathematics, Stanford University
Thesis Supervisor

Signature redacted

Accepted by ................. ........................................

William Minicozzi
Chairman, Department Committee on Graduate Theses
Abstract

In this thesis, we analyze the regularity method pioneered by Szemerédi, and also discuss one of its prevalent applications, the removal lemma.

First, we prove a new lower bound on the number of parts required in a version of Szemerédi's regularity lemma, determining the order of the tower height in that version up to a constant factor. This addresses a question of Gowers.

Next, we turn to algorithms. We give a fast algorithmic Frieze-Kannan (weak) regularity lemma that improves on previous running times. We use this to give a substantially faster deterministic approximation algorithm for counting subgraphs. Previously, only an exponential dependence of the running time on the error parameter was known; we improve it to a polynomial dependence. We also revisit the problem of finding an algorithmic regularity lemma, giving approximation algorithms for some co-NP-complete problems. We show how to use the Frieze-Kannan regularity lemma to approximate the regularity of a pair of vertex sets. We also show how to quickly find, for each \( \epsilon' > \epsilon \), an \( \epsilon' \)-regular partition with \( k \) parts if there exists an \( \epsilon \)-regular partition with \( k \) parts.

After studying algorithms, we turn to the arithmetic setting. Green proved an arithmetic regularity lemma, and used it to prove an arithmetic removal lemma. The bounds obtained, however, were tower-type, and Green posed the problem of improving the quantitative bounds on the arithmetic triangle removal lemma, and, in particular, asked whether a polynomial bound holds. The previous best known bound was tower-type with a logarithmic tower height. We solve Green's problem, proving an essentially tight bound for Green's arithmetic triangle removal lemma in \( \mathbb{F}_p^m \).

Finally, we give a new proof of a regularity lemma for permutations, improving the previous tower-type bound on the number of parts to an exponential bound.

Thesis Supervisor: Jacob Fox
Title: Professor of Mathematics, Stanford University
Acknowledgments

I am deeply grateful to my advisor, Jacob Fox, for all his guidance and support throughout my graduate studies. I have learned much from him and grown greatly as a researcher by seeing his example. All through my studies, he has been selflessly available for guidance, mentoring, and support.

I would like to thank my high school math teacher László Surányi for being a wonderful math teacher and giving me a solid foundation for my mathematical knowledge.

I would like to thank Dan Král’ for multiple research visits and collaborations, mathematical discussions, and general advice and friendship.

I would like to thank Christian Borgs, Jennifer Chayes, and Henry Cohn for letting me intern with them at Microsoft Research for two valuable summers.

I would like to thank all my friends who have supported me throughout the years, including Lukas Brantner, Aditi Chandra, Irene Dedoussi, Tamás Iványi, Dániel Korándi, Przemyslaw Lasota, Gerry Makó, János Perczel, Lisa Sauermann, Boris Shteynas, Péter Szirmai, Fan Wei, Lawrence Wong, Richard Zhang, and Yufei Zhao. Additionally, a very special thanks to Danielle Bragg for always being there to encourage and support me.

Finally, I would like to thank my parents, my three sisters, and my entire family for all their love and support throughout my life.
## Contents

1 Introduction ........................................ 9

2 Regularity lemma lower bound ...................... 19
   2.1 Introduction ................................... 19
   2.2 Regularity Lemma Upper Bound ................. 23
   2.3 Construction .................................. 27
   2.4 Singular Values and Edge Distribution in Matrices .... 38
   2.5 Irregularity Between Parts .................... 41
   2.6 Proof of the Lower Bound ....................... 45
   2.7 Equitable partitions with small irregularity .... 64
   2.8 Concluding Remarks ........................... 68

3 Algorithmic regularity ............................ 71
   3.1 Introduction ................................... 71
   3.2 Algorithmic weak regularity .................... 78
   3.3 Approximation algorithm for subgraph counts .... 89
   3.4 Finding an irregular pair ...................... 92
   3.5 Approximating regularity ...................... 98

4 Arithmetic removal ................................ 103
   4.1 Introduction ................................... 103
   4.2 Proof of the lower bound ....................... 106
   4.3 Proof of the upper bound ...................... 112
4.4 Concluding remarks

5 Permutation regularity lemma

5.1 Interval regular partitions for functions

5.2 Dealing with equitable partitions
Chapter 1

Introduction

This thesis contains a collection of results concerning regularity lemmas and their applications. Szemerédi's regularity lemma [66] is one of the most important tools in graph theory. Roughly speaking, the lemma says that the vertices of any graph can be divided into a bounded number of parts (of almost equal size if needed) such that for most pairs of parts, the graph behaves in a probabilistic, random-like fashion between them, referred to as being $\epsilon$-regular for a parameter $\epsilon$. This is a rough structural result which holds for all graphs. At first, the lemma may seem surprising: even though a graph may have an arbitrarily large number of vertices, up to the error parameter, its structure can be described in a bounded way.

The regularity lemma and the regularity method revolutionized graph theory, shifting the focus from individual vertices and edges of a graph to global structure and properties. A weaker version of this lemma was used to prove a long-standing conjecture of Erdős and Turán [65], now known as Szemerédi's theorem, which states that any subset of the integers with positive upper density contains arbitrarily long arithmetic progressions. Later, the regularity lemma was used to solve many open problems in extremal combinatorics, and today it has numerous applications in graph theory, property testing, number theory, and theoretical computer science (see for example the surveys by Komlós and Simonovits [48], and Rödl and Schacht [61]). The regularity lemma is also an essential component in developing the theory of dense graph limits, where it can be thought of as a compactness result, see [52, 51].
The regularity lemma is often paired with the counting lemma (see for example [51, Lemma 10.22]). Roughly speaking, the counting lemma says that given a graph $H$, if we want to count the number of copies of $H$ between certain subsets of a graph $G$, and $G$ is $\epsilon$-regular between those subsets (for a sufficiently small $\epsilon$), then we can do so simply based on the densities between the subsets. This lemma, when paired with the regularity lemma, is very powerful, since it implies that we can estimate the distributions of copies of $H$ in $G$ based an $\epsilon$-regular partition.

One of the most fundamental applications of the regularity method is the triangle removal lemma of Ruzsa and Szemerédi [64]. Roughly speaking, it says that if a graph has few triangles (represented by a parameter $\delta$), then it is possible to delete a small number of edges (represented by a parameter $\epsilon$) so that the remaining graph is triangle-free. This lemma has many applications; for example, it can be used to prove Roth’s theorem, which says that any subset of the integers with positive upper density contains a three-term arithmetic progression (a special case of Szemerédi’s theorem on arithmetic progressions). The lemma is also true for other graphs $H$, not just triangles, and can be proven by analyzing a regularity partition of the graph, together with the counting lemma.

Removal lemmas are also very closely related to problems in the field of property testing. Property testing is a major area of study in theoretical computer science that has garnered a great amount of interest. The general problem was first formulated by Rubinfeld and Sudan [63], and later investigated for combinatorial objects by Goldreich, Goldwasser and Ron [37]. Suppose we have a “global” property of a graph, and we want to be able to quickly distinguish between graphs that satisfy the property, and graphs that are far from satisfying it. One method is to randomly sample a subset of the vertices, and look at the induced subgraph on them. Although this may give an incorrect answer, using regularity methods, one can show that for certain properties, the probability of error is very low. The following example illustrates this.

Suppose we are given a graph, and told that either it has no triangles, or even after deleting $\epsilon n^2$ edges it will still have a triangle. The triangle removal lemma implies that in the second case, there are at least $\delta n^3$ triangles for some $\delta > 0$. We can then sample
$k$ points from the graph, randomly, and check whether they contain a triangle. If they do, then clearly the graph is not triangle-free. Otherwise, our algorithm determines that the graph is triangle free. What is the probability of error? If the graph is not triangle-free, it has $\delta n^3$ triangles, so it is not too difficult to show that if we sample, say, $1000/\delta$ points, then the probability that there are no triangles between them is very small. Using the graph removal lemma, this method generalizes to testing $H$-freeness for any graph $H$, and there is a long line of work on property testing for other graph properties, see, for example, [1, 5, 6, 7, 8]. The number of sampled points needed depends on how small $\delta$ has to be as a function of $\epsilon$. If the removal lemma is proven using the regularity method, then this ultimately comes down to the number of parts in the regularity partition.

The number of parts needed in the regularity lemma depends on an error parameter $\epsilon$, which determines how close the graphs between parts have to be to being random, and how many “irregular” pairs of parts are allowed. As a function $M(\epsilon)$, it is incredibly large; the standard proof gives a bound that is a tower of twos of height polynomial in $\epsilon^{-1}$. In a breakthrough result, described by Bollobás [17] as a ‘tour de force’ in Gowers’ Fields Medal citation, Gowers [38] proved that indeed a tower-type number of parts is needed. In the paper, Gowers raised the problem of determining the correct order of the tower-height.

In Chapter 2, we address Gowers’ question by proving a lower bound on the number of parts that essentially matches the upper bound for a version of the lemma, showing that $M(\epsilon)$ is equal to a tower of twos of height $\Theta(\epsilon^{-2})$. Both Gowers’ construction and ours uses the probabilistic method in an intricate way, combining concrete structure with randomness, and in some sense reverse engineering the proof of the regularity lemma. In the standard proof of the upper bound for the original regularity lemma, one constructs, iteratively, a sequence of partitions $\mathcal{P}_0, \mathcal{P}_1, \ldots, \mathcal{P}_r$ that successively refine each other. Then one can show by a density increment argument that after at most $\epsilon^{-2}$ steps, one must obtain a regular partition. For our lower bound construction, we also take a sequence of partitions that refine each other, and we construct the graph based on this sequence, making sure each partition is not regular.
We then show that for any other partition \( Q \), if it is not “close” to being a refinement of \( \mathcal{P}_r \), then it cannot be \( \epsilon \)-regular. Therefore, if \( Q \) is \( \epsilon \)-regular, it cannot have significantly fewer parts than \( \mathcal{P}_r \). In previous works, the analysis of the constructions would always work with a single step; they would show that if \( Q \) is close to being a refinement of \( \mathcal{P}_i \) but not close to being a refinement of \( \mathcal{P}_{i+1} \), then there must be a certain amount of irregularity. Therefore, it would give a polynomial lower bound on the irregularity, but it was not optimal. For our construction and analysis, we obtain a tight bound by “collecting” irregularity carefully, over multiple steps. Roughly speaking, in step \( i \), the amount of irregularity we collect is proportional to how much further \( Q \) is from being a refinement of \( \mathcal{P}_i \) compared to \( \mathcal{P}_{i-1} \). Our construction and the corresponding analysis therefore shows that the proof of the regularity lemma is essentially best possible, and the density increment argument is optimal.

For our argument, we analyze the regularity lemma without adding requirement that the partition is equitable (i.e. that the parts have almost equal size). Although proving a lower bound for this version is stronger, to complete the analysis, we also present in Chapter 2 a theorem which directly shows that, for the version of the regularity lemma discussed, requiring the partition to be equitable has a negligible effect on the tower height.

Due to the enormous number of parts needed in Szemerédi’s regularity lemma, most of its applications also give very weak bounds. It is therefore often very useful to prove results that were originally proven using the regularity lemma by another method, as this can lead to greatly improved bounds. For many applications, a weaker property of the partition suffices. In this vein, Frieze and Kannan [32, 33] proved a “weak” regularity lemma (often referred to as the Frieze-Kannan regularity lemma). In a “Frieze-Kannan” \( \epsilon \)-regular, or weakly \( \epsilon \)-regular partition, we do not require any pair of parts to be regular. However, we require that between any two large sets, the number of edges is close to what we would expect based on how the sets intersect the parts of the partition, and the densities between parts. In exchange for only requiring this weaker property, the number of parts needed is much fewer: instead of the tower-type dependence, only \( 2^{2/\epsilon^2} \) parts are required (marginally increased if
an equitable partition is needed). In fact, in the later paper [33], Frieze and Kannan show that slightly more can be said: the graph is “close” to being the weighted sum of $O(1/e^2)$ “cut” graphs, graphs consisting of all edges between a single pair of parts. The also observe that given such an approximation, the common refinement of all the parts is a weakly regular partition.

Since the regularity lemma has numerous applications, there has been a great deal of interest in developing algorithmic versions. Chapter 3 discusses some of these, and gives some concrete applications. Szemerédi’s original proof of the regularity lemma was not algorithmic, because it involves checking, for every pair of parts, whether they are $\epsilon$-regular. In [3], it was shown that it is co-NP-complete to check whether a given pair of parts, and thus whether a partition, is $\epsilon$-regular. However, they do give an algorithm that takes as input a bipartite graph between two vertex sets and a parameter $\epsilon$, and either correctly states that a pair of sets is $\epsilon$-regular, or finds a pair of subsets which show that they are not $\epsilon^4/16$-regular. The running time of their algorithm is $O_\epsilon(n^{\omega+o(1)})$, where $\omega < 2.373$ is the matrix multiplication exponent [21, 50]. This can be used to give an algorithm that finds a regularity partition in time $O_\epsilon(n^{\omega+o(1)})$, with the number of parts at most an exponential tower of height $O(\epsilon^{-20})$. This was later improved by various authors [34, 47, 4]. In particular, in [47], the running time is $O_\epsilon(n^2)$, which is an optimal dependence on $n$ for a deterministic algorithm.

Frieze and Kannan [32, 34] give fast probabilistic algorithms for finding weakly regular partitions. In terms of deterministic algorithms, Dellamonica, Kalyanasundaram, Martin, Rödl, and Shapira [23, 24] give two algorithmic versions. The running times of the algorithms are $O(\epsilon^{-6} n^{\omega+o(1)})$, where $\omega < 2.373$ is the matrix multiplication exponent, and $O(2^{2^{-O(1)} n^2})$. We can see that there is a tradeoff between the two versions: in the second version, the exponent of $n$ is improved to 2, but the dependence on $\epsilon$ is much worse. In Chapter 3, we improve the running time to give on $\epsilon^{-O(1)} n^{2}$-time algorithm. In fact, we give an algorithm which finds an approximation of the graph as the weighted sum of $\epsilon^{-O(1)}$ cut graphs, something which previous algorithms did not do.
Although the number of parts guaranteed by the regularity lemma is huge, for many graphs that arise, there exists a regular partition that has a much smaller number of parts. One downside of these algorithms is that even for these graphs, they only guarantee a tower-type number of parts. Addressing this, Fischer, Matsliah, and Shapira [27] give a fast probabilistic algorithm which, given a graph that has an $\epsilon/2$-regular partition with $k$ parts, outputs an $\epsilon$-regular partition on at most $k$ parts. Their method builds on a folklore proof of the regularity lemma, written down by Tao [68], which gives a regularity partition by taking a random sample of the vertices and outputting the common refinement of their neighborhoods. However, it is still desirable to obtain a deterministic algorithm.

Using the algorithmic version of the Frieze-Kannan regularity lemma, we address these questions in Chapter 3. Specifically, we prove that there is a fast algorithm which, given a graph $G$ that admits an $\epsilon$-regular partition with $k$ parts, outputs a $(1 + \alpha)\epsilon$-regular partition into at most $k$ parts. The running time of the algorithm is $O_{\epsilon,\alpha,k}(n^2)$. We also give a fast algorithm that approximately tests whether a given pair of parts is regular, again with only a slight loss on the regularity parameter. In particular, this means that we can distinguish in time $O_{\epsilon,\alpha}(n^2)$ between an $\epsilon$-regular partition and a partition which is not $(1 - \alpha)\epsilon$-regular.

We also study in Chapter 3 the question of finding an algorithm which, given a (large) graph $G$ and a (smaller) graph $H$, counts the number of copies of $H$ in $G$. This is a well-known problem. It is not difficult to find a probabilistic algorithm which approximates the fraction of $k$-tuples (where $k$ is the number of vertices in $H$) which form a copy of $H$ in $G$, simply by sampling at random. However, this does not have a guarantee of success. It is desirable to find a deterministic algorithm which counts the number of copies of $H$, allowing a small error of $\epsilon n^k$, but guaranteeing with certainty that the error is not larger.

Together with the counting lemma and the algorithmic regularity lemma, one can obtain an algorithm to approximate the number of copies of $H$, by taking an $\epsilon$-regular partition, and counting based on the weighted graph obtained from the partition. This gives an $O_{\epsilon,k}(n^2)$ algorithm that counts the number of copies up to an error of at most
\( cn^k \), however, the dependence of the running time on \( \epsilon \) and \( k \) is tower-type, since the number of parts is tower-type. Duke, Lefmann, and Rödl [25] give a faster algorithm, by developing a weak regularity lemma with an exponential dependence instead of tower-type, and applying that. Their algorithm approximates the number of copies of \( H \) on \( k \) vertices, with an error of at most \( \epsilon n^k \), in time \( 2^{(k/\epsilon)^{O(1)}} n^{\omega+o(1)} \).

Using the algorithmic version of the Frieze Kannan weak regularity lemma, we obtain an even faster algorithm in Chapter 3. Specifically, given a graph \( H \) with \( k \) vertices and \( m \) edges, and a graph \( G \) on \( n \) vertices, and \( \epsilon > 0 \), we show that there is a deterministic algorithm that runs in time \( \epsilon^{-O(n^2)} n^{2} \), and finds the number of copies of \( H \) in \( G \) up to an error of at most \( \epsilon n^k \). Note that previously the dependence on \( 1/\epsilon \) was exponential, while here it is polynomial.

For a long time, the only known proof of the removal lemma used Szemerédi’s regularity lemma, and thus gave very weak bounds. Fox [28] gave a proof without using the (strong) regularity lemma, in which \( \delta^{-1} \) can be taken to be a tower of twos of height logarithmic in \( \epsilon^{-1} \). Even in the case of the triangle removal lemma, this is the best known upper bound. The best known lower bound, however, comes from Behrend’s construction giving a lower bound on Roth’s theorem in [10], which gives a quasi-polynomial bound of the form \( (1/\epsilon)^{\log(\epsilon^{-1})} \). Thus, there is an enormous gap between the upper and the lower bound. Closing this gap remains a major open problem.

Green [40] developed a regularity lemma in the arithmetic setting for abelian groups, and used it to prove some interesting applications. As a consequence, he obtained an arithmetic triangle removal lemma. Let \( p \) be a prime, \( n \) a positive integer, and suppose we have three subsets \( X, Y, Z \) of \( \mathbb{F}_p^n \). Call a triple of points \( x, y, z \) a triangle if \( x + y + z = 0 \). Green proved that if the number of triangles in \( X \times Y \times Z \) is small (again based on a parameter \( \delta \)), then it is possible to remove a small total number of elements (based on a parameter \( \epsilon \)) from \( X, Y, \) and \( Z \), and end up with no triangles. As in the graph theoretic case, this problem is very closely related to the property testing problem of testing whether a triple of sets is triangle-free. The removal lemma implies that if a triple of sets is far from being triangle-free
(in terms of edit distance), then there are many triangles, so if we start sampling triples randomly, with high probability we will see a triangle. However, the bounds obtained in the removal lemma when applying the regularity lemma are tower type. Green posed the problem of finding better bounds, and asked in particular whether a polynomial bound holds. This problem has received considerable attention by many researchers [11, 12, 13, 14, 15, 28, 35, 39, 40, 42, 43, 62].

Král’, Serra, and Vena [49] noticed that it follows from the graph triangle removal lemma, so Green’s arithmetic removal lemma is a special case. Until recently, the best known bound matched the best known bound for the general case due to Fox [28], which says that we can take $\delta = O((\log(1/\epsilon))^{-1})$, where $T$ is the tower function. This was also proven directly by Hatami, Sachdeva, and Tulsiani [42] in the case $p = 2$.

The case $p = 2$ in particular has received much attention due to its close connections to theoretical computer science, in particular, it is closely related to testing triangle freeness of boolean functions. Bhattacharyya and Xie [15] first proved a nontrivial upper bound on $\delta$ in terms of $\epsilon$, showing that we must have $\delta \leq \epsilon^c$ for $c \approx 4.847$. Later, Fu and Kleinberg [35] showed that $\delta \leq \epsilon^{C_2 - o(1)}$ for an explicitly defined $C_2 \approx 13.239$. Therefore, until recently, the best known upper bound for $\delta$ was polynomial in $\epsilon$, and the best known lower bound was tower-type. It was widely believed that the upper bound was far from optimal.

In Chapter 4, we prove a polynomial lower bound. For every $p$, we define a $C_p$, and prove that in $\mathbb{F}_p^n$, we must have $\delta \geq c^{C_p}$. The proof uses Kleinberg, Sawin, and Speyer’s essentially sharp bound on multicolored sum-free sets [46] (which involves a related sharp constant $c_p$ for each $p$), which builds on the breakthrough on the cap set problem by Croot-Lev-Pach [22], and further work by Ellenberg-Gijswijt [26], Blasiak-Church-Cohn-Grochow-Naslund-Sawin-Umans [16], and Alon. For $p = 2$, the exponent $C_2$ in our bound matches that of Fu and Kleinberg [35], and the work of Kleinberg, Sawin, and Speyer [46] implies that for every prime $p$ the exponent $C_p$ in our bound is tight.

Finally, in Chapter 5, we turn our attention to permutations. We give a new simple proof of a regularity lemma for permutations, first proven by Cooper [20] and
refined by Hoppen, Kohayakawa, and Sampaio [45]. Our proof improves the bound on the number of parts from tower-type to single exponential.

This thesis is based on a few papers at various stages of publication. Chapter 2 is mostly based on the paper [30], with the section regarding equitable partitions based on a section in the paper [31]. Chapters 3 and 5 are based on corresponding sections from the paper [31] and followup work. Finally, Chapter 4 is based on the paper [29].
Chapter 2

Regularity lemma lower bound

2.1 Introduction

As discussed in the introduction, Szemerédi’s regularity lemma [66] is one of the most powerful tools in graph theory. An early version was originally developed by Szemerédi [65] in his proof of Szemerédi’s theorem, and it has since become a central tool for studying graphs, with many applications (see the surveys [48], [61]). Roughly speaking, the lemma says that the vertex set of any graph may be partitioned into a small number of parts such that the bipartite subgraph between almost every pair of parts behaves in a random-like fashion.

We next describe more precisely the version of Szemerédi’s regularity lemma studied in this chapter. This version was first formulated by Lovász and Szegedy [52], and can easily be shown to be equivalent to Szemerédi’s original version, up to a polynomial loss. For a pair of vertex subsets $X$ and $Y$ of a graph $G$, let $e(X,Y)$ be the number of ordered pairs of vertices $(x, y) \in X \times Y$ that have an edge between them in the graph. Let $d(X,Y) = \frac{e(X,Y)}{|X||Y|}$ be the edge density between $X$ and $Y$. The irregularity of the pair $(X,Y)$ is defined to be

$$\text{irreg}(X,Y) = \max_{U \subseteq X, W \subseteq Y} |e(U,W) - |U||W||d(X,Y)|.$$ 

This is a value between 0 and $|X||Y|$. If this is a small fraction of $|X||Y|$, then the
edge distribution between $X$ and $Y$ is quite uniform, or random-like. The irregularity of a partition $\mathcal{P}$ of the vertex set of $G$ is defined to be

$$\text{irreg}(\mathcal{P}) = \sum_{X,Y \in \mathcal{P}} \text{irreg}(X,Y).$$

Szemerédi's regularity lemma, as stated in [52], is as follows.

**Theorem 2.1.1.** For any $\epsilon > 0$, there is a (least) $M(\epsilon)$ such that any graph $G = (V, E)$ has a vertex partition into at most $M(\epsilon)$ parts with irregularity at most $\epsilon |V|^2$.

We present for completeness the standard proof of the regularity lemma in Section 2.2, using a density increment argument with the mean square density. It shows that $M(\epsilon)$ is at most a tower of twos of height $O(\epsilon^{-2})$. By a careful argument, we obtain a tower height which is at most $2 + \epsilon^{-2}/16$. Formally, the *tower function* $T(n)$ of height $n$ is defined recursively by $T(1) = 2$ and $T(n) = 2^{T(n-1)}$.

A vertex partition of a graph is *equitable* if any two parts differ in size by at most one. In the statement of the regularity lemma, it is often added that the vertex partition is equitable. For many applications of the regularity lemma, this is a convenient property to have. However, there are several good reasons not to add this requirement to the regularity lemma. First, the main result of this section, which gives a lower bound on $M(\epsilon)$ whose height is on the same order as the upper bound, does not need this requirement. Second, the proof of the upper bound is cleaner without it. Finally, we show in section 2.7 that whether or not an equitable partition is required has a negligible effect on $M(\epsilon)$.

The original version of the regularity lemma is as follows. A pair of vertex subsets $(X, Y)$ is $\epsilon$-regular if for all $U \subset X$ and $W \subset Y$ with $|U| \geq \epsilon |X|$ and $|W| \geq \epsilon |Y|$, we have $|d(U, W) - d(X, Y)| \leq \epsilon$. It is not difficult to show that if a pair $(X, Y)$ is $\epsilon$-regular, then its irregularity is at most $\epsilon |X||Y|$. Conversely, if the irregularity is at most $\epsilon^2 |X||Y|$, then the pair is $\epsilon$-regular. A vertex partition with $k$ parts is $\epsilon$-regular if all but at most $\epsilon k^2$ pairs of parts are $\epsilon$-regular. The regularity lemma states that for every $\epsilon > 0$, there is a $K(\epsilon)$ such that there is an equitable $\epsilon$-regular partition into at most $K(\epsilon)$ parts. Again, it is not too difficult to see that if the irregularity of an
equitable partition is at most $\epsilon^4|V|^2$, then it is $\epsilon$-regular. Conversely, if an equitable partition is $\epsilon$-regular, it has irregularity at most $2\epsilon|V|^2$.

For many applications, it would be helpful to have a smaller bound on the number of parts in the regularity lemma. In other words, can the bound on $M(\epsilon)$ be significantly improved? A breakthrough result of Gowers [38] gave a negative answer to this problem, showing that $K(\epsilon)$ is at least a tower of twos of height on the order of $\epsilon^{-1/16}$, and hence $M(\epsilon)$ is at least a tower of height on the order of $\epsilon^{-1/64}$. Bollobás referred to this important work as a ‘tour de force’ in Gowers’ Fields Medal citation [17]. Gowers [38] further raises the problem of determining the correct order of the tower height in Szemerédi’s regularity lemma.

More recently, Conlon and Fox [19] estimated the number of irregular pairs in the regularity lemma, and Moshkovitz and Shapira [57] gave a simpler proof of a tower-type lower bound. However, these results still left a substantial gap in the order of the tower height.

The main result in this chapter is a tight lower bound on the tower height in the regularity lemma. It shows that $M(\epsilon)$ in Theorem 2.1.1, the regularity lemma, can be bounded from above and below by a tower of twos of height on the order of $\epsilon^{-2}$. Our lower bound construction shows that the density increment argument using the mean square density in the proof of the regularity lemma cannot be improved.

**Theorem 2.1.2.** For $\epsilon < 1/4$, there is a constant $c > 0$ such that the bound $M(\epsilon)$ on the number of parts in the regularity lemma is at least a tower of twos of height at least $c\epsilon^{-2}$.

The assumption $\epsilon < 1/4$ is needed because it is easy to check that $M(\epsilon) = 1$ for $\epsilon \geq 1/4$ and $M(\epsilon) \geq 2$ otherwise. The proof shows that we may take $c = 10^{-26}$, although we do not try to optimize this constant in order to give a clearer presentation.

All of the proofs of tower-type lower bounds on the regularity lemma, including that of Theorem 2.1.2, build on the basic framework developed by Gowers in [38]. It involves a probabilistic construction of a graph which requires many parts in any partition with small irregularity. A simple argument shows that it is sufficient for
the construction to be an edge-weighted graph, with edge weights in [0, 1]. The
construction begins with a sequence of equitable vertex partitions \( P_0, P_1, \ldots, P_s \) with
\( s = \Theta(\epsilon^{-2}) \) and \( P_{i+1} \) a refinement of \( P_i \) with exponentionally more parts than \( P_i \).
The weight of an edge will depend on which parts its vertices lie in. The proof further
shows that any partition of the constructed graph with irregularity at most \( \epsilon \) cannot
be too far from being a refinement of \( P_s \), which has many parts. The theorem quickly
follows from this result.

There are two novelties in the proof of Theorem 2.1.2. While the previous con-
structions could not give a tight bound, the lower bound construction here is carefully
chosen to mimic the upper bound. However, the main novelty is in how we analyze
the construction. To get a good lower bound on the irregularity of a partition, it is
helpful to see how each part \( S \) breaks into smaller pieces based on each partition \( P_i \).
A careful bookkeeping allows us to lower bound the irregularity by collecting how
much \( S \) splits at each step \( i \). Previous arguments could not collect the contributions
to the irregularity from a part splitting off in pieces over different steps.

To accomplish this, we need careful averaging arguments over many parts to obtain
the desired bound. This type of averaging argument is in fact necessary as we cannot
guarantee irregularity from a particular part \( S \). Indeed, it is not too difficult to show
that for each \( \epsilon > 0 \) there is a \( k = k(\epsilon) \) which is only exponential in \( \epsilon^{-O(1)} \) such that
every graph has an equitable vertex partition into parts \( V_1, \ldots, V_k \) such that \( V_1 \) is \( \epsilon \)-
regular with every other part \( V_j \). This justifies the need for the more global averaging
arguments we use to bound the irregularity.

We remark that Theorem 2.1.1 is often more convenient to work with than the
original version of the regularity lemma. There are several reasons for this. One reason
is that in various applications, the standard notion of regularity requires considering
extra case analysis depending on whether or not certain subsets are at least an \( \epsilon \)-
fraction of a part. This is so that one can apply the density conclusion in the definition
of an \( \epsilon \)-regular pair. A good example is the proof of the counting lemma, an important
tool in combination with the regularity lemma. See [18] for a nice proof of the
counting lemma using the notion of irregularity. Another reason is that the notion
of irregularity is closely related to the cut norm developed in the proof of the Frieze-Kannan weak regularity lemma, which shows the close relationship between these two important results; see [18, 19, 52, 61, 67] for details.

**Organization.** In the next section, we present a proof of Szemerédi’s regularity lemma, Theorem 2.1.1. In Section 2.3, we present our construction for the lower bound. In Section 2.4, we review some properties of weighted graphs that follow from their singular values, including a bipartite analogue of the expander mixing lemma. In Section 2.5, we use these results to show where most of the irregularity of the construction comes from. In Section 2.6, we use this to show that our construction requires a large number of parts for any partition of small irregularity. In Section 2.7, we show that adding the requirement that the partition is equitable has a negligible effect on the partition size. In the last section, we make some concluding remarks. For the sake of clarity of presentation, we do not make any serious attempt to optimize absolute constants in our statements and proofs.

### 2.2 Regularity Lemma Upper Bound

Here, for completeness, we present the proof of Szemerédi’s regularity lemma, showing that the bound $M(\varepsilon)$ on the number of parts is at most a tower of twos of height at most $2 + \varepsilon^{-2}/16$. The proof is similar to Szemerédi’s original proof [66], although it uses a probabilistic approach as in [9].

The key idea that makes the proof work is to use a density increment argument with the *mean square density*. Let $G = (V, E)$ be a graph and let $P$ be a vertex partition into parts $V_1, V_2, \ldots, V_k$. The mean square density of the partition $P$ is defined to be

$$q(P) := \sum_{i,j=1}^{k} \frac{|V_i||V_j|}{|V|^2} d(V_i, V_j)^2.$$

Let $x$ and $y$ be two vertices of $G$, chosen independently and uniformly at random. Let $Z$ be the random variable which is the density $d(V_i, V_j)$ between the pair of parts for which $x \in V_i$ and $y \in V_j$. Note that the mean square density $q(P)$ is exactly
$E[Z^2]$. Now, suppose that we have a refinement $\mathcal{P}'$ of $\mathcal{P}$. Let $Z'$ be the corresponding random variable for $\mathcal{P}'$. It is not difficult to see that for any fixed $i,j$, we have $E(Z'|x \in V_i, y \in V_j) = d(V_i, V_j)$, which is equal to $Z$ conditioned on $x \in V_i$ and $y \in V_j$. Thus,

$$E(Z'^2 - Z^2 | x \in V_i, y \in V_j) = E((Z' - Z)^2 + 2Z(Z' - Z)|x \in V_i, y \in V_j)$$

$$= E((Z' - Z)^2|x \in V_i, y \in V_j) + E(2Z(Z' - Z)|x \in V_i, y \in V_j)$$

$$= E((Z' - Z)^2|x \in V_i, y \in V_j). \quad (2.1)$$

This is clearly non-negative for any pair of parts $(V_i, V_j)$. This implies that the mean square density cannot decrease from taking a refinement. However, the next lemma shows that if we take the refinement $\mathcal{P}'$ carefully, we can find a lower bound on the difference in terms of the irregularity of $\mathcal{P}$.

**Lemma 2.2.1.** Suppose we have a partition $\mathcal{P}$ into $k$ parts, with irregularity $z|V|^2$. Then, there is a refinement $\mathcal{P}'$ of $\mathcal{P}$ with at most $k^2k+1$ parts and $q(\mathcal{P}') \geq q(\mathcal{P}) + 4z^2$.

**Proof.** Let $V_1, V_2, \ldots, V_k$ be the parts of $\mathcal{P}$. Fix a pair of parts $(V_i, V_j)$. First, assume $i \neq j$ (this just makes the argument a bit simpler). Let the irregularity of $V_i$ and $V_j$ be $z_{ij}|V_i||V_j|$, given by subsets $W_i \subset V_i$, $W_j \subset V_j$. Hence, we have

$$z_{ij} = \frac{1}{|V_i||V_j|}|e(W_i, W_j) - |W_i||W_j|d(V_i, V_j)| = \frac{|W_i||W_j|}{|V_i||V_j|}|d(W_i, W_j) - d(V_i, V_j)|.$$

For simplicity of notation, we will assume that the value on the right hand side in the absolute value is positive; the exact same proof works if it is negative. For now, let $\mathcal{P}'$ be the refinement of $\mathcal{P}$ obtained by dividing $V_i$ into $W_i$ and $U_i = V_i \setminus W_i$, and $V_j$ into $W_j$ and $U_j = V_j \setminus W_j$, and keeping the rest of the parts the same. Let $Z$ and $Z'$ be defined as before, and let $\tilde{Z}$ be the random variable equal to $Z' - Z$ conditioned on $x \in V_i, y \in V_j$. Using (2.1), we have

$$E(Z'^2 - Z^2|x \in V_i, y \in V_j) = E((Z' - Z)^2|x \in V_i, y \in V_j) = E(\tilde{Z}^2).$$

24
If \( x \in V_i, y \in V_j \), then \( Z = d(V_i, V_j) \). However, if \( x \in W_i, y \in W_j \), then \( Z' = d(W_i, W_j) \) and \( \tilde{Z} = d(W_i, W_j) - d(V_i, V_j) = z_{ij} \sqrt{\frac{|V_i||V_j|}{|W_i||W_j|}} \).

We use the following simple fact. If \( T \) is a random variable with \( \mathbb{E}(T) = 0 \) and \( T = a \neq 0 \) with probability \( p < 1 \), then \( \mathbb{E}(T^2) \geq \frac{p}{1-p} a^2 \). Indeed, if we let \( x = \mathbb{E}(T|T \neq a) \), then we have

\[
0 = \mathbb{E}(T) = pa + (1-p)x,
\]

which gives

\[
x = \frac{-pa}{1-p},
\]

and so

\[
\mathbb{E}(T^2) = pa^2 + (1-p)\mathbb{E}(T^2|T \neq a) \geq pa^2 + (1-p)x^2 = pa^2 + \frac{p^2 a^2}{1-p} = \frac{p}{1-p} a^2,
\]

where the inequality is by an application of the Cauchy-Schwarz inequality.

We know that if \( x \in W_i, y \in W_j \), then \( \tilde{Z} \) is equal to \( z_{ij} \sqrt{\frac{|V_i||V_j|}{|W_i||W_j|}} \). Let \( w = |W_i||W_j| \) and \( v = |V_i||V_j| \). Applying the above statement for \( T = \tilde{Z}, a = z_{ij} \frac{v}{w}, \) and \( p = \frac{w}{v} \), we obtain that

\[
\mathbb{E}(\tilde{Z}^2 - Z^2|x \in V_i, y \in V_j) = \mathbb{E}(\tilde{Z}^2) \geq z_{ij}^2 \left( \frac{v^2}{w^2} \right) \frac{w/v}{1-w/v} = z_{ij}^2 \frac{1}{v} \left( 1 - \frac{w}{v} \right) \geq 4z_{ij}^2.
\]

Note that the above remains true for any refinement of \( \mathcal{P}' \). Indeed, if \( \mathcal{P}'' \) is the refinement and \( Z'' \) the corresponding random variable, then

\[
\mathbb{E}(Z''^2 - Z^2|x \in V_i, y \in V_j)
\]

\[
= \mathbb{E}(Z''^2 - Z^2|x \in V_i, y \in V_j) + \mathbb{E}(Z''^2 - Z^2|x \in V_i, y \in V_j) \geq 0 + 4z_{ij}^2.
\]

We assumed above that \( i \neq j \). If \( i = j \) the same argument works by dividing \( V_i \) into four parts (the common refinement of the two bipartitions), and keeping the other parts the same.

Now, let us do this division for every pair of parts \( (V_i, V_j) \), and take the common
refinement. With a slight abuse of notation, call this common refinement $P'$, and take $Z'$ to be the corresponding random variable. Hence, the increase in the mean square density is

$$q(P') - q(P) = E(Z'^2) - E(Z^2) = \sum_{i,j=1}^{k} E(Z'^2 - Z^2 | x \in V_i, y \in V_j) \frac{|V_i||V_j|}{|V|^2}$$

$$\geq \sum_{i,j=1}^{k} \frac{4z_{ij}^2 |V_i||V_j|}{|V|^2}.$$ 

By the Cauchy-Schwarz inequality, we have

$$(z|V|^2)^2 = \left( \sum_{i,j=1}^{k} z_{ij}|V_i||V_j| \right)^2 \leq \left( \sum_{i,j=1}^{k} z_{ij}^2 |V_i||V_j| \right) \left( \sum_{i,j=1}^{k} |V_i||V_j| \right) = \left( \sum_{i,j=1}^{k} z_{ij}^2 |V_i||V_j| \right) |V|^2.$$ 

Dividing this by $|V|^4/4$, we obtain that the increase in the mean square density is at least

$$4 \sum_{i,j=1}^{k} \frac{z_{ij}^2 |V_i||V_j|}{|V|^2} \geq 4z^2.$$ 

The part $V_i$ is partitioned into two parts for each $j \neq i$, and into four parts if $j = i$. As there are $k$ parts in $P$, each part $V_i$ is divided into at most $2^{k+1}$ parts in $P'$, giving a total of at most $k2^{k+1}$ parts in $P'$, and completing the proof of the lemma.

Proof of Theorem 2.1.1. Let $d = d(V, V)$ be the edge density of the graph $G = (V, E)$. First, note that if we take the trivial partition with one part, then the mean square density is $d^2$. If we take a partition into parts of size one, then the mean square density is $d$. As taking a refinement of a partition can not decrease the mean square density, the mean square density of any vertex partition of $G$ is always between $d^2$ and $d$. Thus, the mean square density of every vertex partition lies in the interval $[d^2, d]$ of length $d - d^2 = d(1 - d) \leq 1/4$. Now, the proof is as follows.

Let $P_0$ be the trivial partition with one part. We will recursively define a sequence
of refinements $\mathcal{P}_0, \mathcal{P}_1, \ldots$ as follows. If the partition $\mathcal{P}_i$ has irregularity at most $\epsilon|V|^2$, then this is the desired partition, and we are done. Otherwise, letting $k$ denote the number of parts of $\mathcal{P}_i$, by Lemma 2.2.1 there is a refinement $\mathcal{P}_{i+1}$ into at most $k2^{k+1}$ parts such that $q(\mathcal{P}_{i+1}) \geq q(\mathcal{P}_i) + 4\epsilon^2$. Since the mean square density cannot be more than $q(\mathcal{P}_0) + 1/4$, this iteration can happen for at most $\lfloor \epsilon^{-2}/16 \rfloor$ steps, and we obtain a partition with irregularity at most $\epsilon|V|^2$. Thus, the number of parts is at most $k_s$ with $s = \lfloor \epsilon^{-2}/16 \rfloor$, where $k_i$ is defined recursively by $k_0 = 1$ and $k_{i+1} = k_i2^{k_i+1}$. At each step, we gain one exponential in the number of parts, and one can easily show by induction that $k_i \leq T(i+2)/4$, where $T$ is the tower function defined in the introduction. Indeed, $k_0 = 1 = T(2)/4$, and by induction

$$k_{i+1} = k_i2^{k_i+1} \leq 2^{2k_i} \leq 2^{4k_i-2} \leq 2^{T(i+2)-2} = \frac{T(i+3)}{4}.$$ 

Thus, the total tower height is at most $2 + \epsilon^{-2}/16$. 

2.3 Construction

We next give the construction of the graph which we use to prove Theorem 2.1.2. We will actually construct a weighted graph $G$ with edge weights in $[0,1]$. The following argument is based on a similar one in [38], but adapted to handle irregularity. It shows that constructing a weighted graph with the desired properties (rather than an unweighted graph) is sufficient. We first prove the following lemma.

**Lemma 2.3.1.** Suppose we have a weighted graph $G$ on $N$ vertices, with weights from the interval $[0,1]$. Let $\tilde{G}$ be a random, unweighted graph on the same set of vertices, where each pair of vertices forms an edge with probability equal to its weight in $G$, independently of the other pairs. Then, with positive probability, $|e_{\tilde{G}}(A,B) - e_G(A,B)| \leq 4N^{3/2}$ for every pair $(A,B)$ of vertex subsets.

**Proof.** We may assume $N \geq 3$ as otherwise the lemma is trivial. For a pair of vertices $(x,y)$, let $b(x,y) = \tilde{G}(x,y) - G(x,y)$. Then $\mathbb{E}(b(x,y)) = 0$, and $|b(x,y)| \leq 1$, for each
Fix two subsets $A$ and $B$. Then
\[ e_G(A, B) - e_G(A, B) = \sum_{\substack{x \in A \\ y \in B}} b(x, y). \]

We would like to apply Azuma’s inequality (see, eg, [9]). The inequality involves martingales, which are defined as follows. A sequence of random variables $X_0, X_1, X_2, ...$ is a martingale if for every $n \geq 1$, we have $E(X_n | X_1, X_2, ..., X_{n-1}) = X_{n-1}$. In our case, if we take an ordering of the edges, and let $X_i$ be the sum of $b(x, y)$ over the first $i$ edges, then this forms a martingale. Azuma’s inequality says the following. Suppose $\{X_k : k = 0, 1, ..., n\}$ is a martingale, and $|X_k - X_{k-1}| \leq c_k$ with probability 1. Then for all positive real numbers $t$, the following holds:

\[ P(X_n - X_0 \geq t) \leq e^{-t^2/(2 \sum_{k=1}^n c_k^2)}. \]

Let $c_{x,y}$ be 2 if $b(x, y)$ and $b(y, x)$ both appear in the sum (note that these numbers are the same), 1 if one of them appears, and 0 otherwise. Then, since $c_{x,y}^2 \leq 4$, we trivially have $\sum c_{x,y}^2 \leq 4 |A| |B|$ (one can actually replace the constant factor 4 by 2 using a more detailed analysis). Let $t = 4N^{3/2}$. By Azuma’s inequality, we have that

\[ P(|e_G(A, B) - e_G(A, B)| > t) = P(|\sum_{\substack{x \in A \\ y \in B}} b(x, y)| > t) \leq 2e^{-t^2/(2 \sum c_{x,y}^2)} \leq 2e^{-t^2/(8 |A| |B|)} \leq 2e^{-t^2/(8N^2)} = 2e^{-2N}. \]

As there are $2^{2N}$ pairs of sets of vertices $A$ and $B$, the probability that there is a pair of sets $(A, B)$ such that $|e_G(A, B) - e_G(A, B)| > 4N^{3/2}$ is at most

\[ 2^{2N}2e^{-2N} = e^{(2N+1)\ln 2 - 2N} = e^{\ln 2 - 2(1 - \ln 2)N} < 1. \]

Hence, with positive probability, we have $|e_G(A, B) - e_G(A, B)| \leq 4N^{3/2}$ for every pair of sets $(A, B)$. □
Next, we show that if we replace a weighted graph with an unweighted graph whose existence is guaranteed by the previous lemma, then the irregularity of any partition cannot decrease by too much. We note that an unweighted graph can be thought of as a weighted graph with weights 0 and 1.

**Lemma 2.3.2.** Suppose we have two weighted graphs $G$ and $G'$, on the same set $V$ of vertices, such that for any pair of subsets $(A, B) \subset V \times V$, we have $|e_G(A, B) - e_{G'}(A, B)| \leq t$. Then, the following holds:

1. For any pair of subsets $(U, W) \subset V \times V$, we have
   \[
   |\text{irreg}_G(U, W) - \text{irreg}_{G'}(U, W)| \leq 2t.
   \]

2. For any positive integer $k$ and partition $\mathcal{P}$ with at most $k$ parts,
   \[
   |\text{irreg}_G(\mathcal{P}) - \text{irreg}_{G'}(\mathcal{P})| \leq 2k^2t.
   \]

**Proof.** For part 1, by symmetry between $G$ and $G'$, it suffices to show that
   \[
   \text{irreg}_G(U, W) - \text{irreg}_{G'}(U, W) \leq 2t,
   \]
or equivalently, $\text{irreg}_{G'}(U, W) \geq \text{irreg}_G(U, W) - 2t$.

By the definition of irregularity, there are subsets $U_1 \subset U$ and $W_1 \subset W$ which satisfy
   \[
   \left| e_G(U_1, W_1) - \frac{|U_1||W_1|}{|U||W|} e_G(U, W) \right| = \text{irreg}_G(U, W).
   \]
As each of the terms on the left hand side changes by at most $t$ in changing $G$ to $G'$, the difference changes by at most $2t$. Therefore,
   \[
   \left| e_{G'}(U_1, W_1) - \frac{|U_1||W_1|}{|U||W|} e_{G'}(U, W) \right| \geq \text{irreg}_G(U, W) - 2t.
   \]
Since $\text{irreg}_{G'}(U, W)$ is the maximum of the left side over all sets $U_1 \subset U, W_1 \subset W$, the proof is complete.
this implies that
\[
\text{irreg}_{G'}(U, W) \geq \text{irreg}_G(U, W) - 2t,
\]
completing the proof of part 1.

For part 2, let \( \mathcal{P} \) partition \( V \) into parts \( V_1, V_2, \ldots, V_l \) with \( l \leq k \). Then
\[
\text{irreg}_G(\mathcal{P}) = \sum_{i,j=1}^{l} \text{irreg}_G(V_i, V_j).
\]
The definition for \( G' \) is analogous. Thus,
\[
|\text{irreg}_G(\mathcal{P}) - \text{irreg}_{G'}(\mathcal{P})| = \left| \sum_{i,j=1}^{l} \text{irreg}_G(V_i, V_j) - \text{irreg}_{G'}(V_i, V_j) \right| \\
\leq \sum_{i,j=1}^{l} \left| \text{irreg}_G(V_i, V_j) - \text{irreg}_{G'}(V_i, V_j) \right| \leq l^2 2t \leq 2k^2 t.
\]

Combining the previous two lemmas, we have the following immediate corollary.

**Corollary 2.3.3.** For every weighted graph \( G \) on \( N \) vertices there is an unweighted graph \( G' \) on the same set of vertices satisfying
\[
|\text{irreg}_{G'}(\mathcal{P}) - \text{irreg}_G(\mathcal{P})| \leq 8k^2 N^{3/2}
\]
for every vertex partition \( \mathcal{P} \) with at most \( k \) parts.

Now, we will construct a weighted graph on \( N \) vertices such that for any partition into at most \( k \) parts, where \( k \) is a tower of twos of height \( 10^{-26} \epsilon^{-2} \), the irregularity is at least \( \epsilon N^2 \). It will be clear from the construction that we may take \( N \) to be arbitrarily large. Let \( 0 < \gamma < 1/2 \). By Corollary 2.3.3, if we take \( N \geq 64 \gamma^{-2} \epsilon^{-2} k^4 \), then \( 8k^2 N^{3/2} \leq \gamma \epsilon N^2 \), and we obtain an unweighted graph such that the irregularity is at least \( (1 - \gamma) \epsilon N^2 \) in any partition into at most \( k \) parts. This justifies why it is sufficient to construct a weighted graph. If \( \epsilon \geq 10^{-13} \), the height of the tower is at most 1, and the result follows from the discussion immediately after Theorem 2.1.2. Hence, we can and will assume that \( \epsilon < 10^{-13} \).

The weighted graph \( G \) we construct to get a lower bound on \( M(\epsilon) \) is bipartite. This simplifies the analysis of the construction, and does not affect the constants by
too much. For the edge-weighted bipartite graph $G$ between vertex sets $V$ and $W$, each of equal size, we will have a sequence of equitable vertex partitions $\mathcal{P}_0, \mathcal{P}_1, \ldots, \mathcal{P}_s$ of $V$, and $\mathcal{Q}_0, \mathcal{Q}_1, \ldots, \mathcal{Q}_s$ of $W$ with $\mathcal{P}_{i+1}$ a refinement of $\mathcal{P}_i$, $\mathcal{Q}_{i+1}$ a refinement of $\mathcal{Q}_i$ for $0 \leq i \leq s-1$, $|\mathcal{P}_i| = |\mathcal{Q}_i|$ for $0 \leq i \leq s$, and the number of parts of $\mathcal{P}_{i+1}$ is exponential in the number of parts of $\mathcal{P}_i$. More precisely, we have a sequence $x_i$, and we will divide each part of $\mathcal{P}_{i-1}$ and $\mathcal{Q}_{i-1}$ into $2x_i$ equal parts. We let $k_i$ be the number of parts of $\mathcal{P}_i$, so $k_i = 2x_ik_{i-1}$. We start with $k_0 = 1$, $x_1 = 2^{10}$, and let $x_{i+1} = 2^{x_i/16}$ for $i \geq 1$. For example, it follows that $x_2 = 2^{26}$ and $x_3 = 2^{2^{26}}$. Note that we did not say anything about the number of vertices in the parts of the last partition, which can be any positive integer. Thus, the number of vertices of the graph can be arbitrarily large, and the argument in the previous paragraph does indeed work.

Let $\alpha$ be the minimum number with $\alpha > 2^{26} \cdot 10000\epsilon$ such that $\alpha^{-1}$ is a multiple of 6 (this will make our calculations later simpler). Since $\epsilon < 10^{-13}$, we have that $2^{26} \cdot 10000\epsilon < 1/6$, and this implies that $\alpha \leq 2^{27} \cdot 10000\epsilon$. We take $s = \alpha^{-2}/36$. Thus, $s \geq \frac{1}{36} \cdot 2^{24} \cdot 10^{6}\epsilon^{-2} \geq 10^{-26}\epsilon^{-2}$. Note that we only need to specify the edge weights between $V$ and $W$ because the non-edges (those pairs inside $V$ or inside $W$) have weight 0.

We begin with a weighted bipartite graph $G_0$ which has constant weight $1/2$ between $V$ and $W$, and take $\mathcal{P}_0$ and $\mathcal{Q}_0$ to be the trivial partitions of each side into a single part. For each $i$ from 1 to $s$, we will construct a weighted bipartite graph $G_i$ between $V$ and $W$ with every edge weight equal to $-\alpha$, 0, or $\alpha$. The graph $G_i$ will have the property that it is constant between any part in $\mathcal{P}_i$ and any part in $\mathcal{Q}_i$. Thus, $G_i$ is a blow-up of an edge-weighted graph between $\mathcal{P}_i$ and $\mathcal{Q}_i$. We will let $\tilde{G}_i = G_0 + G_1 + \cdots + G_i$, i.e., the edge weight of a pair in $\tilde{G}_i$ is the sum, over all $j \leq i$, of that pair’s edge weight in $G_j$. The weighted graph $G$ is defined as $G := \tilde{G}_s = G_0 + \cdots + G_s$, so that the edge weight in $G$ is the sum of the edge weights of the corresponding edge in each $G_i$.

For the construction, we have left to specify the edge weights in the $G_i$ for $i \geq 1$, and we do so recursively. We do the following for each $i$ from 1 to $s$. If the (constant) weight in $\tilde{G}_{i-1}$ across a pair of sets $X \in \mathcal{P}_{i-1}$ and $Y \in \mathcal{P}_{i-1}$ is 0 or 1, then we call the
pair \((X, Y)\) inactive, and the edge weight in \(G_i\) across the pair \((X, Y)\) is 0. Otherwise, we call the pair \((X, Y)\) active, and, since \(G_0\) had constant weight \(1/2\), and since \(\alpha^{-1}\) is an even integer, the (constant) weight between the pair \((X, Y)\) in \(\overline{G}_{i-1}\) is a multiple of \(\alpha\) which is at least \(\alpha\) and at most \(1 - \alpha\).

For every pair \(X \in \mathcal{P}_{i-1}, Y \in \mathcal{Q}_{i-1}\) of parts, we do the following. Recall that \(X\) and \(Y\) are each divided into \(2x_i\) parts in \(\mathcal{P}_i\) and \(\mathcal{Q}_i\), respectively. We randomly divide the \(2x_i\) parts in \(X\) into two groups of size \(x_i\), and let \(X^0_Y\) and \(X^1_Y\) be the vertices in each of these parts. Thus, \(X = X^0_Y \sqcup X^1_Y\) is an equitable partition of the vertices in \(X\). We define \(Y = Y^0_X \sqcup Y^1_X\) analogously. Thus, each \(Y \in \mathcal{Q}_{i-1}\) gives a random partition of \(X\), and we make these random partitions independently. If \((X, Y)\) is an active pair, then, for \(a = 0, 1\), the edge weight in \(G_i\) is \(\alpha\) between \(X^a_Y\) and \(Y^a_X\), and \(-\alpha\) between \(X^a_Y\) and \(Y^{1-a}_X\). See Figure 2-1. This completes the construction of \(G\).

![Figure 2-1: An active pair of parts \((X, Y)\)](image)

Note that if a pair of vertices goes across an inactive pair in step \(i\), then its weight
in $\tilde{G}_j$ for $j \geq i$ (and hence in $G$) is fixed at 0 or 1, and it will go across inactive pairs at each later step. We chose $s = \alpha^{-2}/36$ small enough to guarantee that only a small fraction of pairs will be inactive. In fact, the following lemma is true.

**Lemma 2.3.4.** In each step $i$, for any $X \in \mathcal{P}_i$, the proportion of $Y \in \mathcal{Q}_i$ such that the pair $(X, Y)$ is inactive is at most 0.05, that is, there are at most $0.05|\mathcal{Q}_i|$ such $Y$. The analogous statement holds for any part $Y \in \mathcal{Q}_i$.

**Proof.** Let $Y \in \mathcal{Q}_i$ be a part picked uniformly at random. For each $1 \leq j \leq i$, let $y_j$ be the number such that $y_j \alpha$ is the value of (any) edge between $X$ and $Y$ in the original $G_j$ (they all have the same value), if this is nonzero. This depends only on the partitions of the corresponding parts of step $j - 1$, thus, we can define this even if the pair from step $j - 1$ are inactive. With this extended definition, $y_j$ is a random variable that is 1 or $-1$, each with probability 1/2. In fact, these $y_j$ are independent. Indeed, for any $j$, the numbers $y_1, y_2, \ldots, y_{j-1}$ depend only on which part of $\mathcal{Q}_{j-1}$ is a superset of $Y$, but if we fix the part of $\mathcal{Q}_{j-1}$ which is a superset of $Y$, then $y_j$ is still 1 or $-1$ with probability 1/2. Let $S_j = y_1 + y_2 + \cdots + y_j$. The pair will be inactive if and only if for some $j \leq i$,

$$|S_j| \geq \frac{1}{2\alpha}.$$ 

Now, we use the following fact, which is a part of mathematical folklore: Let $t$ be a positive integer. Then

$$P(\exists j \leq i : S_j \geq t) = P(S_i \geq t) + P(S_i \geq t + 1).$$

We sketch a proof of this using the reflection principle. We take the following bijective map from the set of sequences $y_1, y_2, \ldots, y_i$ to itself. If there is no $j \leq i$ for which $S_j \geq t$, we map the sequence to itself. If there is such a $j$, let $j_0$ be the smallest $j$ with $S_j \geq t$ (note that necessarily $S_j = t$), and reverse the sign of all $y_j$ for $j > j_0$ (if we draw the path that $S_j$ takes as $j$ increases, we are "reflecting" the part of the path after the first point where it reaches $t$). Then it is not difficult to check that sequences with $S_i = t$ get mapped to such sequences, and we also obtain a bijection.
between sequences with \( S_i \geq t + 1 \), and sequences with \( S_i < t \) but \( S_j \geq t \) for some \( j \leq i \). From this, the claim quickly follows.

Now, since the \( y_i \) and hence the \( S_i \) are symmetrically distributed and \( P(S_i \geq t + 1) \leq P(S_i \geq t) \), we can conclude by the Chernoff bound that the probability that there is a \( j \leq i \) for which \( |S_j| \geq \frac{1}{2a} \) is at most

\[
4e^{-\left(\frac{\mu}{2a}\right)^2/(2i)} \leq 4e^{-1/(8a^2i)} \leq 4e^{-1/(8\alpha^2)}.
\]

Thus, substituting in \( s = \alpha^{-2}/36 \), then this will be at most \( 4e^{-1/(8/36)} \leq .05 \). That is, in any step, at most a .05 proportion of the pairs of parts containing any given part will be inactive. \( \square \)

Note that the conclusion of the above lemma is satisfied no matter what our choices are for the partitions in each step. In the following, we describe the properties that we want our graph to satisfy, and show that with positive probability, our construction will satisfy them. We will show that if the weighted graph \( G \) constructed above has the desired properties, then it will give a construction which verifies Theorem 2.1.2. In fact, in Theorem 2.3.5, we will show that it has the stronger property that any vertex partition of \( G \) with irregularity at most \( \epsilon |V(G)|^2 \) is not far from being a refinement of \( \mathcal{P} \).

Fix an \( i \geq 1 \), a pair \( X \in \mathcal{P}_{i-1} \), and \( Y \in \mathcal{Q}_{i-1} \), and \( a, b \in \{0, 1\} \), such that the weight of \( \tilde{G}_i \) between \( X^a_Y \) and \( Y^b_X \) is not 0 or 1 (note that it is constant). In other words, any part of \( \mathcal{P}_i \) which is a subset of \( X^a_Y \) and any part of \( \mathcal{Q}_i \) which is a subset of \( Y^b_X \) form an active pair. The number of parts of \( \mathcal{P}_i \) in \( X^a_Y \) is \( x_i \), and each such part is divided into \( 2x_{i+1} \) parts in \( \mathcal{P}_{i+1} \). Similarly, the number of parts of \( \mathcal{Q}_i \) in \( Y^b_X \) is \( x_i \), and each such part is divided into \( 2x_{i+1} \) parts in \( \mathcal{Q}_{i+1} \). Fix a part \( B \) of \( \mathcal{P}_i \) with \( B \subset X^0_Y \). Any part \( C \) of \( \mathcal{Q}_i \) with \( C \subset Y^0_X \) will split \( B \) into two collections of parts of \( \mathcal{P}_{i+1} \), where both collections have size \( x_{i+1} \). Now we want the bipartitions \( B = B^0_C \cup B^1_C \) of \( B \) to satisfy the following two properties (see Figure 2).

1. Given two parts \( C \) and \( C' \) of \( \mathcal{Q}_i \) that are subsets of \( Y^b_X \), we obtain two different partitions of \( B \), \( B^0_C \cup B^1_C \) and \( B^0_{C'} \cup B^1_{C'} \). Let \( z_{hj} \) be the number of parts
of $Q_{i+1}$ in $B_C^k \cap B_C^l$. Since $C$ and $C'$ both divide $B$ into two equal parts, we can see that $z_{00} = z_{11}$ and $z_{10} = z_{01}$. Also, $z_{00} + z_{01} = x_{i+1}$. Let $z = z(C, C') = z_B(C, C') = z_{00} - z_{01}$. We want every pair of $C$ and $C'$ to satisfy $|z(C, C')| \leq r = \sqrt{6x_{i+1} \ln x_i}$.

2. For any two vertices $u$ and $v$ in $B$ in different parts of $\mathcal{P}_{i+1}$, we say that a part $C$ separates $u$ and $v$ if $u$ and $v$ lie in different parts in the partition $B = B_C^0 \cup B_C^1$. Let $y(u, v)$ be the number of parts of $Q_i$ in $Y_X^b$ that do not separate $u$ and $v$ minus the number of parts of $Q_i$ in $Y_X^b$ that do separate $u$ and $v$. Then, for any $u$ and $v$, we want $|y(u, v)| \leq t = x_i/2$.

Figure 2-2: Part $B$ is divided into two parts by both $C$ and $C'$, the horizontal line shows how $C$ divides it, the other line how $C'$ divides it. We can see that $C$ separates $u$ and $v$, but $C'$ does not.

We will show that for each of these properties, the probability of failure, given a random set of $x_i$ partitions of $B$, is less than $1/2$, implying that the probability
that neither of them fails is positive. The partitions are independent for different choices of $B$, so with positive probability, the two properties are satisfied for each $B \in \mathcal{P}_{i-1}, B \subset X^a$. We require the analogous conditions for the bipartitions of each part in $Q_i$ in $Y^b_X$. For the sake of brevity, we omit explicitly stating these conditions for the other side. By symmetry between the two parts in the bipartite graph, the analogous conditions also hold with positive probability on the other side, and since the partitions on the two sides are independent, with positive probability they hold simultaneously for both sides. The conditions are also independent if we take a different pair $(X, Y)$ and $(a, b)$, or look at a different step $i$. This clearly implies that there exists a set of good partitions, that is, a set of partitions such that these two properties are satisfied for each $i$, for any of the required pairs $X \in \mathcal{P}_{i-1}, Y \in Q_{i-1}$, $a, b \in \{0, 1\}$, such that the pairs of parts in step $i$ between $X^a$ and $Y^b_X$ are active.

Thus, assume $i$ is fixed, $X \in \mathcal{P}_{i-1}, Y \in Q_{i-1}$, and $a, b \in \{0, 1\}$ such that any pair of parts of step $i$ between $X^a$ and $Y^b_X$ is active, and $B \in \mathcal{P}_i$ with $B \subset X^a$ is fixed. For fixed $C$ and $C'$, the value of $z_B(C, C')$ follows a hypergeometric distribution. That is, if we fix $B_0^0$ and $B_1^0$, and choose $B_0^1$ and $B_1^1$ randomly, this is the same as drawing half the vertices of $B$ randomly without replacement, and looking at how many we draw from $B_0^0$. As the hypergeometric distribution is at least as concentrated as the corresponding binomial distribution, that is, if we drew a random sequence of vertices from $B$ of the same length with replacement (for a proof, see Section 6 of [44]), we can apply the Chernoff bound to obtain

$$P(|z(C, C')| > r) \leq 2e^{-r^2/(2x_i+1)}.$$

Recall that $r = \sqrt{6x_{i+1} \ln x_i}$. Since the number of pairs is $x_i(x_i-1)/2$, the probability of the first property failing is at most

$$(x_i(x_i-1)/2)2e^{-r^2/(2x_i+1)} = x_i(x_i-1)e^{-3\ln x_i} = \frac{x_i-1}{x_i^2} < 1/2.$$

Now we show that the second property holds with probability of failure less than $1/2$. For a fixed pair $u$ and $v$, each bipartition separates them with probability
\[
\frac{x_{i+1}}{2x_{i+1}+1} = \frac{1}{2} + \frac{1}{4x_{i+1}-2}, \text{ independently of the other bipartitions.}
\]

For a real number \( p \) and positive integer \( n \), let \( B_{n,p} \) denote the binomial distribution with \( n \) trials and probability value \( p \), so if \( Z \) is a random variable with distribution \( B_{n,p} \), then \( P(Z = k) = \binom{n}{k} p^k (1 - p)^{n-k} \). If \( Z_1 \) is the random variable with distribution \( B_{n,\frac{1}{2}} \) and \( Z_2 \) is the random variable with distribution \( B_{n,\frac{1}{2} + \delta} \) with \( |\delta| < \frac{\ln 2}{2n} \), then \( P(Z_2 = k) \leq 2P(Z_1 = k) \) for each \( k \). Indeed,
\[
P(Z_2 = k) \over P(Z_1 = k) = (1 + 2\delta)^k (1 - 2\delta)^{n-k} \leq (1 + 2|\delta|)^n \leq e^{2|\delta|n} < 2.
\]

Note that we can write \( y(u, v) \) as \( 2x(u, v) - n \), where \( x(u, v) \sim B_{n,p} \), with \( n = x_i \) and \( p = \frac{1}{2} + \delta \) with \( \delta = -\frac{1}{4x_i+1-2} \). Since \( x_{i+1} = 2^{x_i/16} \), and \( x_i \geq x_1 = 2^{10} \), it is easy to check that \( |\delta| = \frac{1}{4x_{i+1}+1} < \frac{\ln 2}{2x_i} = \frac{\ln 2}{2n} \). If we instead take \( y'(u, v) = 2x'(u, v) - n \) with \( x'(u, v) \sim B_{n,1/2} \), by the Chernoff bound, we have
\[
P(|y'(u, v)| > t) < 2e^{-t^2/(2x_i)}.
\]

By the above analysis of binomial distributions with probability \( p \) near 1/2, we thus have
\[
P(|y(u, v)| > t) < 2 \cdot 2e^{-t^2/(2x_i)}.
\]

Recall that \( t = x_i/2 \), and \( x_{i+1} = 2^{x_i/16} \). We have \( \binom{2x_{i+1}}{2} < 2x_{i+1}^2 \) pairs of parts of \( P_{i+1} \) which are subsets of \( B \). Thus, since \( x_i \geq x_1 = 2^{10} \), the probability of the second property failing is less than
\[
2x_{i+1}^2 e^{-t^2/(2x_i)} = \exp((\ln 2)(3 + x_i/8 - x_i/8)) = \exp(x_i(\ln 2 - 1)/8 + 3 \ln 2) < 1/2.
\]

Thus, we indeed have a probability of failure less than 1/2 for each of the two properties. For the remainder of the proof we will fix such a graph \( G \) with these properties for each \( i \), for any pair \( X \in P_{i-1}, Y \in Q_{i-1} \), any \( a, b \in \{0, 1\} \), and any \( B \in P_i \) with \( B \subset X_j^a \), and the same conditions holds when we switch the two parts. We will show that the irregularity is at least \( \epsilon |V(G)|^2 \) in any partition with not too
many parts.

**Theorem 2.3.5.** The weighted graph $G$ constructed in the above manner has the property that any partition with fewer than $0.97(|P_s| + |Q_s|)/x_1$ parts has irregularity greater than $\epsilon |V(G)|^2$.

We will show that we can assume that our partition is a refinement of the first partition, $P_1 \cup Q_1$. This will give us a factor of $32x_1^2$ in the irregularity, and $4x_1$ in the size of the partition. Assuming this, we will show something a bit stronger. We will show that if the partition has irregularity at most $32x_1^2\epsilon |V(G)|^2$, then the partition is not too far from being a refinement of the joint partition $P_s \cup Q_s$, which implies that it has at least half as many parts.

### 2.4 Singular Values and Edge Distribution in Matrices

In this section, we prove some properties about the edge distribution of bipartite graphs. The key lemma gives a bipartite analogue of the expander mixing lemma of Alon and Chung [2]. While it is a modification of a standard proof, we give the proof here for completeness.

If we have any weighted bipartite graph (with potentially negative weights) between two sets $X$ and $Y$, each of size $n$, we can represent it by an $n \times n$ matrix $A$, with $X$ corresponding to the rows, $Y$ corresponding to the columns, and the entry value equal to the weight of the corresponding edge. Let $A = O\Lambda U^T$ be the singular value decomposition. Thus, $\Lambda = \text{diag}(\sigma_1, \sigma_2, \ldots, \sigma_n)$, $\sigma_i \geq 0$, and $O$ and $U$ are orthogonal real matrices (recall that $A$ has real entries). Let $\lambda = \max_i \sigma_i$. Also, for any $v \in X$, $C \subset Y$, define

$$d_C(v) = \sum_{w \in C} a_{vw}.$$ 

This is the sum of the edge weights between $v$ and the vertices in $C$. We next observe that if $\lambda$ is small, then the graph between $X$ and $Y$ is very uniform. The following
Lemma shows that for any subset of the vertices on one side, most of the vertices on the other side will have density about equal to the average density, which in our case is zero. This will imply that the number of edges between any two sets is small in absolute value compared to their sizes.

**Lemma 2.4.1.** Suppose $A$ is a matrix representing a weighted graph $G$ between $X$ and $Y$, and suppose $\lambda$ is the maximum of the singular values $\sigma_i$. We have the following.

1. For any $C \subset Y$,
$$\sum_{v \in X} d_C(v)^2 \leq \lambda^2 |C|.$$ 

2. For any $B \subset X$ and $C \subset Y$,
$$|e(B, C)| \leq \lambda \sqrt{|B||C|}.$$ 

**Proof.** We first prove part 1. Let $y$ be the vector with entries corresponding to elements of $Y$, with 1 at the elements of $C$, and 0 at the other elements. We claim that
$$\langle Ay, Ay \rangle \leq \lambda^2 \langle y, y \rangle.$$ 

Indeed, recall that $A = O\Lambda U^T$, and if we let $y' = U^T y$, then
$$\langle Ay, Ay \rangle = \langle O\Lambda U^T y, O\Lambda U^T y \rangle = y'^T \Lambda^2 y' \leq \lambda^2 \langle y', y' \rangle = \lambda^2 \langle y, y \rangle. \quad (2.2)$$

The first equality in (2.2) follows by substituting for $A$, the second by substituting $y'$ and using the fact that $O$ is orthogonal. The inequality follows from the fact that $\Lambda$ is a diagonal matrix containing the singular values, and the last equality follows from the fact that $U$ is orthogonal.

Moreover, we know that
$$Ay = (d_C(v), v \in X),$$
and thus
\[ \langle Ay, Ay \rangle = \sum_{v \in X} d_C(v)^2. \]  
(2.3)

However, it is easy to see that
\[ \langle y, y \rangle = |C|. \]  
(2.4)

Combining (2.2), (2.3), and (2.4), we obtain
\[ \sum_{v \in X} d_C(v)^2 \leq \lambda^2 |C|. \]

This completes the proof of part 1.

For part 2, we use the Cauchy-Schwarz inequality, and the statement of part 1. We can see that
\[ |e(B, C)| = \left| \sum_{v \in B} d_C(v) \right| \leq \sqrt{|B|} \sqrt{\sum_{v \in B} d_C(v)^2} \leq \sqrt{|B|} \sqrt{\sum_{v \in X} d_C(v)^2} \leq \lambda \sqrt{|B||C|}. \]

This completes the proof of part 2.

Suppose we have an \( n \times n \) matrix \( A \) with singular values \( \sigma_1, \sigma_2, \ldots, \sigma_n \). Consider the \( k \)-blow-up \( A' \) of \( A \), which is a \( kn \times kn \) matrix where we replace each entry \( a_{ij} \) of \( A \) with the \( k \times k \) matrix \( a_{ij}E \), where \( E \) is the constant matrix with 1’s everywhere. In other words, \( A' = A \otimes E \) is the tensor product (also known as the Kronecker product) of \( A \) and \( E \). Note that \( A' \) corresponds to the matrix we obtain by replacing each vertex by \( k \) (independent) vertices. We have the following lemma.

**Lemma 2.4.2.** Let \( A' \) be the \( k \)-blow-up of an \( n \times n \) matrix \( A \), and \( \lambda \) and \( \lambda' \) be the largest singular values of \( A \) and \( A' \) respectively. Then \( \lambda' = k\lambda \).

**Proof.** We will show that if the singular values of \( A \) are \( \sigma_1, \ldots, \sigma_n \), then the singular values of \( A' \) are \( k\sigma_1, k\sigma_2, \ldots, k\sigma_n, 0, 0, \ldots \). That is, we have each original singular value multiplied by \( k \) once, and the remaining singular values of \( A' \) are zeroes. This clearly implies the lemma. Let \( O_k \) be a \( k \times k \) matrix that is orthogonal, and the first column consists of \( \frac{1}{\sqrt{k}} \) everywhere (such an \( O_k \) clearly exists). Then
$O_k^T E O_k = \text{diag}(k, 0, 0, \ldots) = kE_{11}$, thus, the singular values of $E$ are $k, 0, \ldots, 0$. If $\Lambda = \text{diag}(\sigma_1, \sigma_2, \ldots, \sigma_n) = U^T A V$, then set $\bar{U} = U \otimes O_k$ and $\bar{V} = V \otimes O_k$. As the tensor product of orthogonal matrices is orthogonal, the matrices $\bar{U}$ and $\bar{V}$ are orthogonal matrices.

\[
\bar{U}^T A' \bar{V} = (U^T \otimes O_k^T)(A \otimes E)(V \otimes O_k) = (U^T A V) \otimes (O_k^T E O_k) = \Lambda \otimes kE_{11}
\]

\[
= \text{diag}(k\sigma_1, 0, \ldots, 0, k\sigma_2, 0, \ldots, 0, k\sigma_3, 0, \ldots, 0, k\sigma_4, \ldots, ).
\]

The second equality above is by the mixed-product property of the tensor product. Hence, the singular values are as claimed, which completes the proof. 

### 2.5 Irregularity Between Parts

Recall that a part of step $i$ refers to a part of $\mathcal{P}_i$ or $\mathcal{Q}_i$. Next, we establish the fact that while $G$ is quite irregular between any two parts $X$ and $Y$ of step $i - 1$ (on opposite sides), almost all of this irregularity comes from $\tilde{G}_i$. In fact, almost all of the irregularity comes from $G_i$, since the $G_j$ with $j < i$ are constant across $X$ and $Y$. Concretely, it is easy to check that $\text{irreg}_{G_i}(X, Y) = \frac{a}{4}|X||Y|$, coming from considering the subsets $X^0 \subset X$ and $Y^0 \subset Y$. In contrast, Lemma 2.5.3 below easily implies that $\text{irreg}_{G-\tilde{G}_i}(X, Y) = O(x_i^{-1/4}a|X||Y|)$. The following lemma is key in establishing this useful fact. The reader may ask why, if this is true for $i = 1$, we don’t just take a four element partition: $\{V^0_w, V^1_w, W^0_v, W^1_v\}$. The reason is that in this case the irregularity is larger than $\epsilon|V||W|$. The bound we get from the lemma will be useful for $i \geq 2$.

**Lemma 2.5.1.** If $X \in \mathcal{P}_{i-1}$ and $Y \in \mathcal{Q}_{i-1}$ are two parts in step $i - 1$, $a, b \in \{0, 1\}$, then for any $U \subset X^a_Y$ and $Z \subset Y^b_X$, we have $|e_{G_{i+1}}(U, Z)| \leq .8x_i^{-1/4}a n \sqrt{|U||Z|} \leq .8x_i^{-1/4}a n^2$, where $n = |X^b_Y| = |Y^a_X|$.

**Proof.** There are three cases to consider. The first case is that the pair $(X, Y)$ is already inactive. In this case, all the weights in $G_{i+1}$ between $X$ and $Y$ are zero (recall that if a pair of parts becomes inactive, then in later steps any pair of parts
coming from this pair is also inactive). The second case is that the pair \((X, Y)\) is active, but after adding the weights in \(G_i\), any edge between \(X^a_i\) and \(Y^b_i\) will have weight 0 or 1. In this case, for any two parts \(X' \subset X^a_i\) and \(Y' \subset Y^b_i\) of step \(i + 1\), the pair \((X', Y')\) will be inactive, thus, in \(G_{i+1}\), all edges between them will again have weight zero.

In the remaining case, any pair of parts of step \(i + 1\) coming from \(X^a_i\) and \(Y^b_i\) is active.

Let \(A'\) be the \(n \times n\) matrix corresponding to the weighted graph \(G_{i+1}\) restricted to \(X^a_i\) and \(Y^b_i\). We will show that the largest singular value \(\lambda'\) of \(A'\) is at most \(.8x_i^{-1/4}a\). This implies the desired result by Lemma 2.4.1.

Recall that the edge weights in \(G_{i+1}\) are equal between each part in \(P_{i+1}\) and each part in \(Q_{i+1}\). Let \(H_{i+1}\) be the bipartite graph with vertex sets \(P_{i+1}\) and \(Q_{i+1}\), so each part of step \(i + 1\) is a vertex, and each edge in \(H_{i+1}\) has weight equal to the density in \(G_{i+1}\) across that pair of parts. Thus, \(G_{i+1}\) is a blow-up of \(H_{i+1}\). Let \(\tilde{X}\) and \(\tilde{Y}\) be the sets of vertices of \(H_{i+1}\) that consists of those parts of step \(i + 1\) that come from \(X^a_i\) and \(Y^b_i\), respectively. Let \(A\) be the matrix corresponding to the weighted graph \(H_{i+1}\), restricted to \(\tilde{X}\) and \(\tilde{Y}\), and \(\lambda\) be the largest singular value of \(A\). Then, \(A\) is an \(m \times m\) matrix, where \(m = 2x_ix_{i+1}\), the number of vertices in each of \(\tilde{X}\) and \(\tilde{Y}\). By Lemma 2.4.2, we have \(\lambda' = \lambda \frac{m}{m}\), thus, it suffices to show that \(\lambda \leq .8x_i^{-1/4}a\).

Consider the matrix \(M = AA^T\). Note that

\[
\text{tr} M^2 = \text{tr}(AA^TAA^T) = \sum_{i=1}^{m} \sigma_i^4 \geq \max \sigma_i^4 = \lambda^4.
\]

To complete the lemma, it thus suffices to show that \(\text{tr} M^2 \leq (.8am)^4/x_i\).

The rows and columns of \(M\) both correspond to the parts of \(P_{i+1}\) that are subsets of \(X^a_i\). Any entry of \(M\) in the diagonal will be \(m\alpha^2\), but the other entries will be much smaller.

First, consider an element off the diagonal, corresponding to \(v\) and \(u\) in \(\tilde{X}\) that are in the same part \(B\) of \(P_i\). Then for any vertex \(w\) in any part that does not separate them, \(A_{vw} = A_{uw} = \pm \alpha\), and for a vertex in a part that does separate them,
\[ A_{vw} = -A_{uw} = \pm \alpha. \] Thus,

\[ M_{vu} = \sum_{w \in \bar{Y}} A_{vw}A_{uw} = y(v, u)2x_{i+1}\alpha^2, \]

where \( y(u, v) = y_B(u, v) \) is as defined at the end of Section 2.3, which is the difference between the number of parts that do not separate them and the number of parts that do separate them. By construction, \( y(u, v) \leq t = x_i/2 \), and hence

\[ |M_{vu}| \leq 2tx_{i+1}\alpha^2. \]

Now, if \( v \) and \( u \) come from two different parts \( B \) and \( B' \) of \( \mathcal{P}_i \), then for any part \( C \subset Y_X^b \) of \( Q_i \), it is divided "almost evenly" by \( B \) and \( B' \), that is, about a quarter of the vertices in \( C \) have an edge with \(+\alpha\) going to both \( v \) and \( u \), a quarter with \(+\alpha\) to \( v \), \(-\alpha\) to \( u \), a quarter with \(-\alpha\) to \( v \), \(+\alpha\) to \( u \), and a quarter with \(-\alpha\) to both. The difference is measured by \( z_C(B, B') \) as defined in Section 2.3. To be precise,

\[ M_{vu} = \sum_{w \in \bar{Y}} A_{vw}A_{uw} = \sum_{C \in Q_{i+1} \subset Y_X^b} 2z_C(B, B')\alpha^2. \]

Thus,

\[ |M_{vu}| \leq \sum_{C \in Q_{i+1} \subset Y_X^b} |2z_C(B, B')|\alpha^2 \leq 2x_i r \alpha^2, \]

where we recall that \( r = \sqrt{6x_{i+1} \ln x_i} \).

There are \( m = 2x_i x_{i+1} \) diagonal entries \( M_{vv} \) in \( M \), \( x_i(2x_{i+1})(2x_{i+1} - 1) \) entries \( M_{vu} \) in the same part of \( \mathcal{P}_i \) with \( u \neq v \), and \( x_i(x_i - 1)(2x_{i+1})^2 \) entries \( M_{vu} \) with \( u \) and
In different parts of \( P_i \). Hence,

\[
\text{tr} M^2 = \sum_{v, u \in X_p^a} M^2_{vu}
\]

\[
\leq m(m^2 + x_i (2x_{i+1}(2x_{i+1} - 1)) 4t^2 x_i x_{i+1}^2 \alpha^4 + x_i (x_i - 1)(2x_{i+1})^2 4x_i x_{i+1}^2 r^2 \alpha^4
\]

\[
\leq m^3 \alpha^4 + 4x_i^2 x_{i+1}^4 \alpha^4 + 96x_i^4 x_{i+1}^2 \ln x_i \alpha^4 = m^4 \alpha^4 \left( \frac{1}{m} + \frac{1}{4x_i} + \frac{6 \ln x_i}{x_{i+1}} \right)
\]

\[\leq \left( \frac{.8 \alpha m}{x_i} \right)^4.
\]

The last inequality is true because, if we recall that \( m = 2x_i x_{i+1}, x_{i+1} = 2x_i^{16} \), we have

\[
\frac{1}{2x_{i+1}} + \frac{1}{4} + \frac{6x_i \ln x_i}{x_{i+1}} \leq .4 \leq (.8)^4.
\]

(The expression on the left is maximized if \( i = 1 \), in which case \( x_i = x_1 = 2^{10} \), \( x_{i+1} = x_2 = 2^{06} \).)

This estimate completes the proof. \( \square \)

Note that if \( j \geq i \), we get the following corollary of the previous lemma, since we can divide \( X \) and \( Y \) into the parts of \( P_{j-1} \) and \( Q_{j-1} \) that are contained in them, and sum over all pairs.

**Corollary 2.5.2.** Let \( X \in P_{i-1}, Y \in Q_{i-1} \), and \( a, b \in \{0, 1\} \). If \( U \subset X_p^a \), \( Z \subset Y_p^b \), and \( j \geq i \), then \( |e_{G_{j+1}}(U, Z)| \leq .8x_j^{-1/4} \alpha |X_p^a||Y_p^b| \).

Now, from this, it is easy to obtain the following lemma, which is the main result in this section, and an important tool in the proof of the main result.

**Lemma 2.5.3.** Let \( X \in P_{i-1}, Y \in Q_{i-1} \), and \( a, b \in \{0, 1\} \). If \( U \subset X_p^a \) and \( Z \subset Y_p^b \), then

\[|e_{G-G_{i}}(U, Z)| \leq .9x_i^{-1/4} \alpha |X_p^a||Y_p^b|\).

**Proof.** Using the triangle inequality, and applying the previous corollary for each \( j \)
from $i$ to $s - 1$, we have

$$|e_{G_{-\bar{G}_1}}(U, Z)| = \left| \sum_{j=i}^{s-1} e_{G_{j+1}}(U, Z) \right| \leq \sum_{j=i}^{s-1} |e_{G_{j+1}}(U, Z)| \leq \sum_{j=i}^{s-1} \frac{.8x_j^{-1/4} \alpha n^2}{j(i)} \leq .9x_i^{-1/4} \alpha |X_\gamma| |Y_\beta|.$$  

Recall that $x_{j+1} = 2^{x_j/16}$ and $x_1 = 2^{10}$, thus the sequence $(x_j)_{j \geq 1}$ grows very rapidly, which justifies the last inequality. \hfill \Box

### 2.6 Proof of the Lower Bound

The goal of this section is to prove Theorem 2.1.2, by first proving Theorem 2.3.5. We therefore suppose we have a partition of the vertex set of $G$. We will show that if this partition is far from being a refinement of $P_s \cup Q_s$, then it has large irregularity. We first assume that this partition is a refinement of the bipartition, that is, we have two partitions, $S$ of $V$ and $T$ of $W$. In fact, we will further assume for now that $S$ is a refinement of $P_1$, and $T$ is a refinement of $Q_1$. We will later show (see Lemma 2.6.8) how to get rid of these assumptions. Our goal for now is to prove the following lemma.

**Lemma 2.6.1.** Assume $S$ is a refinement of $P_1$ and $T$ is a refinement of $Q_1$. If the vertex partition $S \cup T$ has irregularity at most $\frac{1}{5000} \alpha |V||W|$, then at least a .97 proportion of the vertices in $V$ are in parts in $S$ such that more than half of its vertices are in the same part of $P_s$, and at least a .97 proportion of the vertices in $W$ are in parts in $T$ such that more than half of its vertices are in the same part of $Q_s$. It follows that the number of parts in $S \cup T$ is at least $.97(|P_s| + |Q_s|)$.

First, we will establish a simple fact that makes future calculations easier. It uses the triangle inequality to show that the irregularity between a pair $S, T$ of parts is large if there are large subsets $S', S'' \subset S$ and $T', T'' \subset T$ for which the densities $d(S', T')$ and $d(S'', T'')$ are not close.
Lemma 2.6.2. Suppose we have, for two sets of vertices $S$ and $T$, subsets $S', S'' \subset S$ and $T', T'' \subset T$ for which $|S''||T''| \leq |S'||T'|$. Then

$$\text{irreg}(S, T) \geq \frac{1}{2} \left| e_G(S'', T'') - \frac{|S''||T''|}{|S'||T'|} e_G(S', T') \right|.$$  

Proof. By the triangle inequality, we have that either

$$\left| e_G(S'', T'') - \frac{|S''||T''|}{|S'||T'|} e_G(S, T) \right| \geq \frac{1}{2} \left| e_G(S'', T'') - \frac{|S''||T''|}{|S'||T'|} e_G(S', T') \right|,$$  

(2.5)

or

$$\left| \frac{|S''||T''|}{|S'||T'|} e_G(S', T') - \frac{|S''||T''|}{|S'||T'|} e_G(S, T) \right| \geq \frac{1}{2} \left| e_G(S'', T'') - \frac{|S''||T''|}{|S'||T'|} e_G(S', T') \right|.$$  

(2.6)

Indeed, the sum of the left-hand sides of (2.5) and (2.6) is at least the sum of the right-hand sides. If (2.5) holds, then we get the desired bound on the irregularity using the subsets $S''$ and $T''$. So we may suppose (2.6) holds.

Since $|S''||T''| \leq |S'||T'|$, multiplying the left side of (2.6) by $\frac{|S'||T'|}{|S''||T''|} \geq 1$ gives that

$$\left| e_G(S', T') - \frac{|S'||T'|}{|S''||T''|} e_G(S, T) \right| \geq \frac{1}{2} \left| e_G(S'', T'') - \frac{|S''||T''|}{|S'||T'|} e_G(S', T') \right|.$$  

Since $\text{irreg}(S, T)$ is the maximum of the expression on the left over all pairs of subsets of $S$ and $T$, this gives that

$$\text{irreg}(S, T) \geq \frac{1}{2} \left| e_G(S'', T'') - \frac{|S''||T''|}{|S'||T'|} e_G(S', T') \right|.$$  

\[ \square \]

Our goal is to give a lower bound on the irregularity of the partition $S \cup T$. Since $G$ is bipartite, the irregularity is zero between a pair of parts in $S$ or a pair of parts
in \( \mathcal{T} \). Hence, irreg\((S \cup \mathcal{T})\) is equal to

\[
\text{irreg}(S, \mathcal{T}) := \sum_{S \in \mathcal{S}, T \in \mathcal{T}} \text{irreg}(S, T).
\]

If we have such subsets for the previous lemma for certain pairs of parts, by collecting the irregularity between these pairs, we obtain the following corollary giving a lower bound on the irregularity of the partition.

**Corollary 2.6.3.** Suppose we have an \( \mathcal{R} \subset S \times \mathcal{T} \), and, for each pair of parts \((S, T) \in \mathcal{R}\), subsets \( S^0_S, S^1_T \subset S \) and \( T^0_S, T^1_T \subset T \) satisfying \( |S^0_S||T^0_S| \leq |S^1_T||T^1_T| \). Then,

\[
\text{irreg}(S, T) \geq \frac{1}{2} \sum_{(S, T) \in \mathcal{R}} \left| e_G(S^0_S, T^0_S) - \frac{|S^0_S||T^0_S|}{|S^1_T||T^1_T|} e_G(S^1_T, T^1_T) \right|.
\]

**Proof.** By the previous lemma, for any pair of \((S, T) \in \mathcal{R}\), we have

\[
\text{irreg}(S, T) \geq \frac{1}{2} \left| e_G(S^0_S, T^0_S) - \frac{|S^0_S||T^0_S|}{|S^1_T||T^1_T|} e_G(S^1_T, T^1_T) \right|.
\]

Adding this up for all \((S, T) \in \mathcal{R}\), we obtain

\[
\text{irreg}(S, \mathcal{T}) \geq \frac{1}{2} \sum_{(S, T) \in \mathcal{R}} \left| e_G(S^0_S, T^0_S) - \frac{|S^0_S||T^0_S|}{|S^1_T||T^1_T|} e_G(S^1_T, T^1_T) \right|,
\]

which is what we wanted to show. \( \square \)

For each step \( i \), we define a coloring of the vertices. For a vertex \( v \), look at the element \( S \) in \( S \) or \( \mathcal{T} \) containing \( v \), and the part \( X \) in step \( i \) containing \( v \). If \( |S \cap X| > |S|/2 \), color \( v \) blue, otherwise, color it red.

Since we keep refining partitions, a vertex can change from blue to red, but it can never change back to blue. By assumption, \( S \) refines \( \mathcal{P}_1 \) and \( \mathcal{T} \) refines \( \mathcal{Q}_1 \), so every vertex is blue in step 1.

For each \( i \), let \( R_i \) be the set of vertices that turn red in step \( i \). Let \( \tilde{R}_i \) be the set of vertices that are red in step \( i \). Thus, \( \tilde{R}_i = R_1 \cup R_2 \cup R_3 \cup \ldots \cup R_i \) is a partition of \( \tilde{R}_i \). We have \( R_1 = \emptyset \) as every vertex is blue in step 1. For any part \( S \) of \( S \cup \mathcal{T} \),
let $i_S$ be the last step when it contains blue vertices. Note that more than half of the vertices of $S$ are blue in step $i_S$, but in step $i_S + 1$ (and later steps), there are no blue vertices. For $j \leq i_S$, let $S_j$ be the set of vertices in $S$ that are blue in step $j$. Let $\tilde{S}_j$ be the set of vertices in $S$ that were blue in the previous step, but are red now, that is, $\tilde{S}_j = S_{j-1} \setminus S_j$. Also, for $S \in S$ and each $j \leq i_S$ there is a unique $X \in \mathcal{P}_j$ such that $S_j \subset X$. For each $Y \in Q_j$, this $X$ is divided into $X_Y^0$ and $X_Y^1$, and for $a = 0, 1$, let $S_{j,Y}^a = S_j \cap X_Y^a$. For each $S$ and $Y$, let $S_{j,Y}$ be the smaller and $S_{j,Y}^*$ be the larger of $S_{j,Y}^0$ and $S_{j,Y}^1$, breaking ties arbitrarily.

The next lemma provides an important estimate in establishing a lower bound on the irregularity. The set-up is that we have a fixed $S \in S$ and $X \in \mathcal{P}_{j-1}$ such that $S_{j-1} \subset X$. That is, $|S_{j-1}| = |S \cap X| \geq |S|/2$. It then roughly says that $S_{j-1,Y}$ is typically a large fraction of how much $S_{j-1}$ breaks off at the next step. Note that in the next lemma the averaging for a typical $Y$ is weighted by $|Y \cap W'|$, where $W'$ is a large subset of $W$. For a vector $\lambda = (\lambda_i) \in \mathbb{R}^k$ and $1 \leq p < \infty$, we write $\|\lambda\|_p$ for $(\sum_{i=1}^k |\lambda_i|^p)^{1/p}$ and $\|\lambda\|_\infty$ for $\max_{1 \leq i \leq k} |\lambda_i|$. For $X \in \mathcal{P}_{j-1}$, let $Q_X$ be the set of $Y \in Q_{j-1}$ such that $(X, Y)$ is an active pair.

**Lemma 2.6.4.** Let $2 \leq j \leq s$. Let $W' \subset W$ with $|W'| = C|W|$. For each $Y \in Q_{j-1}$, let $Y' = Y \cap W'$. Let $S \in S$ be such that $i_S \geq j - 1$, and let $X \in \mathcal{P}_{j-1}$ be the unique part with $S_{j-1} \subset X$.

1. If $i_S \geq j$, then
   \[ \sum_{Y \in Q_X} |S_{j-1,Y}| |Y'| \geq |\tilde{S}_j| (C - .8)|W|. \]

2. If $i_S = j - 1$, then
   \[ \sum_{Y \in Q_X} |S_{j-1,Y}| |Y'| \geq \frac{1}{2} (|S_{j-1}| - |S|/2) (C - .8)|W|. \]

**Proof.** In the case $C \leq .8$, the desired bound is trivial, and so we may (and will) assume $C > .8$.

First, note that by Lemma 2.3.4, at most a .05 proportion of the $Y$ are inactive with $X$, so at most $.05|W|$ vertices are in a $Y$ that is inactive with $X$. The set $S_{j-1}$
is divided into parts $A_1, \ldots, A_k$ by $P_j$. First, look at a fixed pair of distinct parts $A_h$ and $A_i$. By construction, at most a $\frac{3}{4}$ proportion of the $Y$ in $Q_{j-1}$ do not separate $A_h$ and $A_i$, and so at most $\frac{3}{4}|W|$ vertices are in a $Y$ that does not separate them. Call $Y \in Q_{j-1}$ good (with respect to the pair $(h, i)$) if the pair $(X, Y)$ is active and $Y$ separates $A_h$ and $A_i$. For each $h$ and $i$, let $\mathcal{R}(h, i)$ be the set of $Y$ which are good. It follows that

$$\sum_{Y \in \mathcal{R}(h, i)} |Y'| \geq |W'| - 0.05|W| - \frac{3}{4}|W| = \left(C - \frac{3}{4} - 0.05\right)|W| = (C - 0.8)|W|. \quad (2.7)$$

Now, we look at the first case. Since $i_S \geq j$, $S_j$ is well-defined, and is one of the $A_i$. Without loss of generality, we may assume $A_1 = S_j$. Then $\tilde{S}_j$ is the union of all $A_h$ with $h \geq 2$. Since $|S_j| \geq |S|/2$, for each $Y$, the partition of $S_{j-1}$ into two parts corresponding to $Y$ satisfies that $A_1$ is a subset of the larger part $S_{j-1,y}^*$ and is hence disjoint from the smaller part $S_{j-1,y}$. Thus $S_{j-1,y}$ is the union of those $A_h$ for which $Y$ separates the vertices in $A_1$ from the vertices in $A_h$. Using (2.7), we have

$$\sum_{Y \in Q_x} |S_{j-1,y}| |Y'| = \sum_{\substack{h \geq 2 \ Y \in \mathcal{R}(1,h) \ \text{for each} \ Y \text{for which} \ (X, Y) \text{is an active pair}}} |A_h| |Y'| \geq \sum_{h \geq 2} |A_h| (C - 0.8)|W| = |\tilde{S}_j| (C - 0.8)|W|.$$ 

This is the desired inequality, and completes this case.

We next consider the second case. Let $\lambda_h = |A_h|$. For each $Y$ for which $(X, Y)$ is an active pair, we have

$$|S_{j-1,y}| (|S_{j-1}| - |S_{j-1,y}|) = \sum_{\substack{h < i \ Y \in \mathcal{R}(h, i) \ \text{for each} \ Y \text{for which} \ (X, Y) \text{is an active pair}}} \lambda_h \lambda_i. \quad (2.8)$$
Thus,

\[
\sum_{Y \in Q_x} |S_{j-1,Y}||S_{j-1}||Y'| \geq \sum_{Y \in Q_x} |S_{j-1,Y}|(|S_{j-1}| - |S_{j-1,Y}|)|Y'|
\]

\[
= \sum_{Y \in Q_x} \sum_{h<i} \lambda_h \lambda_i |Y'| = \sum_{h<i} \lambda_h \lambda_i \sum_{Y \in R(h,i)} |Y'|
\]

\[
\geq \sum_{h<i} \lambda_h \lambda_i (C - .8)|W|
\]

\[
= (C - .8)|W| \sum_{h<i} \lambda_h \lambda_i = (C - .8)|W| (||\lambda||_1^2 - ||\lambda||_2) / 2
\]

\[
\geq (C - .8)|W||\lambda||_1 (||\lambda||_1 - ||\lambda||_\infty) / 2
\]

\[
= (C - .8)|W||S_{j-1}| (|S_{j-1}| - ||\lambda||_\infty) / 2.
\]

The first equation above is from substituting in (2.8), the second equation is by changing the order of summation, the second inequality is by substituting in (2.7), the last inequality uses the inequality $||\lambda||_2 \leq ||\lambda||_\infty ||\lambda||_1$, and the last equality is by substituting $||\lambda||_1 = |S_{j-1}|$. Dividing both sides by $|S_{j-1}|$, we obtain (using that $||\lambda||_\infty \leq |S|/2$)

\[
\sum_{Y \in Q_x} |S_{j-1,Y}||Y'| \geq (C - .8)|W| (|S_{j-1}| - ||\lambda||_\infty) / 2
\]

\[
\geq (C - .8)|W| (|S_{j-1}| - |S|/2) / 2,
\]

which completes the proof.

Our basic strategy is to show that if a significant proportion of the vertices turn red by step $s$, then we must have a certain amount of irregularity. There will be two approaches to show this. If the number of red vertices increases by a substantial amount in a single step, then we will look only at the partitions in that step, and analyzing these we will obtain enough irregularity. However, it is possible that the number of red vertices slowly increases on both sides, and eventually after many steps it adds up to a substantial amount, even though it never increases by much in any
single step. The following lemma will be applicable in this case. It obtains an amount of irregularity from each step proportional to the number of new red vertices \( j \leq i \). With careful bookkeeping, it manages to add these up to give a substantial lower bound on the irregularity.

**Lemma 2.6.5.** Suppose in step \( i \geq 2 \) we have \( C|W| \) blue vertices in \( W \) and \( r|V| \) red vertices in \( V \). Then

\[
\text{irreg}(S, T) \geq \frac{r}{12} (C - .8) \alpha |V||W| - \frac{1}{2} x^{-1/4} \alpha |V||W|.
\]

First, we will prove an auxiliary lemma, which will also be useful later. The lemma roughly says that in each step, we can collect irregularity proportional to the number of new red vertices in that step. In order to be able to do this over multiple steps, we have to do this in a careful way.

**Lemma 2.6.6.** Let \( j \geq 2 \) be a positive integer and suppose in step \( j \), we have a collection \( S^0_j \) of sets \( S \in S \), each of which has the property that they still have blue vertices in step \( j \), and for each of them we have a subset \( S'' \) of the blue vertices of size at least \( |S|/2 \). Suppose we also have a collection \( S^1_j \) of sets that have blue vertices in step \( j - 1 \) but not in step \( j \). Suppose we have a subset \( W' \subset W \) of vertices that are blue in step \( j \) of size \( |W'| = C|W| \). For each \( T \in T \), let \( T' = T \cap W \). Then we can find subsets \( S'(j, T) \subset \tilde{S}_j \) for each \( S \in S^0_j \), and subsets \( S'(T), S''(T) \subset S \) with \( |S'(T)| \leq |S''(T)| \) for each \( S \in S^1_j \), such that

\[
\sum_{S \in S^0_j, T \in T} \left| e_G(S'(j, T), T') - \frac{|S'(j, T)|}{|S''|} e_G(S'', T') \right| + \sum_{S \in S^1_j, T \in T} \left| e_G(S'(T), T') - \frac{|S'(T)|}{|S''(T)|} e_G(S'', T') \right| \quad (2.9)
\]

51
is at least

\[
2\alpha \left( \sum_{S \in \mathcal{S}_j} |\bar{S}_j| + \frac{1}{2} \sum_{S \in \mathcal{S}_j} \left( |S_{j-1}| - \frac{|S|}{2} \right) \right) (C - .8) |W| - 2 \cdot .9 \cdot x_j^{-1/4} |V||W|.
\]

If \(W'\) consists of vertices that are blue in step \(j - 1\) but not necessarily in step \(j\), then an analogous result is true, except for each term in (2.9), we replace \(T'\) with \(T'(S) \subset T'\) of size at least \(|T'|/2\) (for each \(S\)), and the first term in the lower bound decreases by a factor of 2.

**Proof.** Again, if \(C \leq .8\) the lemma is trivial, so assume \(C > .8\). We know that for any \(T \in \mathcal{T}\), \(T'\) is a subset of a single part of \(Q_j\), so it is in a single part of \(Q_{j-1}\). Let \(Y\) be any part of \(Q_{j-1}\), and \(T\) be such that \(T' \subset Y\).

![Figure 2-3: The disk is \(S_{j-1}\), the bottom right part is \(S_j\), the horizontal line divides it into \(S_{j-1,Y}^0\) and \(S_{j-1,Y}^1\). Recall that \(\bar{S}_j\) is \(S_{j-1} - S_j\).](image)

First, suppose \(S \in \mathcal{S}_j^0\), so \(i_S \geq j\) and \(S_j\) is nonempty. Then, \(S_j\) is a subset of a
part $X' \in \mathcal{P}_j$, and $S_{j-1}$ will be a subset of a part $X \in \mathcal{P}_{j-1}$ (where clearly $X' \subset X$).

Assume that the pair $(X, Y)$ is active. Recall that $X$ is divided into two parts by $Y$, $X^0_Y$ and $X^1_Y$. This divides $S_{j-1}$ into two parts, $S^0_{j-1,Y}$ and $S^1_{j-1,Y}$. The set $S_j$ is a subset of one of them (since it is inside $X'$), and, since $|S_j| > |S|/2$, it will be inside the bigger one. Recall that $S^0_{j-1,Y}$ is the smaller of the two. Thus, $S_j$ and $S^0_{j-1,Y}$ are both entirely inside one of $X^0_Y$ and $X^1_Y$, and we know that they are not inside the same one. See Figure 2-3. Similarly, since $T'$ is all blue in step $j$, it is entirely inside of a single part of $Q_j$. This means that one of $S_j$ and $S^0_{j-1,Y}$ has weight $\alpha$ on all edges going to $T'$ in $G_j$, and the other has weight $-\alpha$ on all such edges. Since they both have the same weight in $\tilde{G}_{j-1}$, this implies, since $S'' \subset S_j$, that

$$|d_{\tilde{G}_j}(S_{j-1,Y}, T') - d_{\tilde{G}_j}(S'', T')| = 2\alpha.$$  

In particular, we have

$$\left| e_{\tilde{G}_j}(S_{j-1,Y}, T') - \frac{|S^0_{j-1,Y}|}{|S''|-1} e_{\tilde{G}_j}(S'', T') \right| = 2\alpha |S^0_{j-1,Y}| |T'|.$$  

Note here that $|S''| > \frac{|S|}{2} > |S^0_{j-1,Y}|$.

Observe that, if $S_j \subset X^a_Y$, then the expression within the absolute value on the left is negative if $T' \subset Y^a_Y$ (and thus $S'' \subset X^b_Y$), and positive if $T' \subset Y^1_X - a$.

Now, suppose $S \in S^1_j$. Recall that in this case, $i_S = j - 1$, so $S_{j-1}$ is nonempty, but $S_j$ is empty. Thus there is a unique $X \in \mathcal{P}_{j-1}$ such that $S_{j-1} \subset X$. Recall that $Q_X$ is the set of $Y \in Q_{j-1}$ such that the pair $(X, Y)$ is active. For each $Y \in Q_X$, $S_{j-1}$ splits into $S^1_{j-1,Y}$ and $S^*_1_{j-1,Y}$. Recall that $S^1_{j-1,Y}$ is the smaller of these two sets, and $S^*_1_{j-1,Y}$ is the larger of these two sets. Note that $T'$ is a subset of $Y^b_Y$ for $b = 0$ or $b = 1$. Thus, as before, we have

$$\left| e_{\tilde{G}_j}(S^1_{j-1,Y}, T') - \frac{|S^1_{j-1,Y}|}{|S^*_1_{j-1,Y}|} e_{\tilde{G}_j}(S^*_1_{j-1,Y}, T') \right| = 2\alpha |S^1_{j-1,Y}| |T'|.$$  

Fix an active pair $(X, Y)$ with $X \in \mathcal{P}_{j-1}$ and $Y \in Q_{j-1}$. Adding the equations up for all pairs $(S, T)$ with $S \in S_j$, $S'' \subset X$, and $T' \subset Y$, we obtain that

53
\[ \sum_{S \in \mathcal{S}_j^0} \left| e_{\tilde{G}_j}(S_{j-1},Y,T') - \frac{|S_{j-1,Y}|}{|S'|} e_{\tilde{G}_j}(S',T') \right| + \sum_{S \in \mathcal{S}_j^1} \left| e_{\tilde{G}_j}(S_{j-1},Y,T') - \frac{|S_{j-1,Y}|}{|S_{j-1,Y}'|} e_{\tilde{G}_j}(S_{j-1,Y}',T') \right| \] (2.10)

is equal to

\[ 2\alpha \sum_{S \in \mathcal{S}_j^0} |S_{j-1,Y}| |T'|. \] (2.11)

It will therefore be helpful to find a lower bound on the sum of (2.11) over all active pairs \((X, Y)\) with \(X \in \mathcal{P}_{j-1}\) and \(Y \in \mathcal{Q}_{j-1}\), or equivalently, to find a lower bound on the following sum:

\[ \sum_{X \in \mathcal{P}_{j-1}} \sum_{Y \in \mathcal{Q}_X} |S_{j-1,Y}| |T'|. \] (2.12)

Fix an \(S \in \mathcal{S}_j\), and look at the terms of (2.12) corresponding to \(S\). Recall that \(|W'| = C|W|\). If \(S \in \mathcal{S}_j^0\), then \(i_S \geq j\), so we can apply case 1 of Lemma 2.6.4 to get that the terms with this \(S\) sum to at least \(|\tilde{S}_j|(C - .8)|W|\), where we recall that \(\tilde{S}_j = S_{j-1} \setminus S_j\). If \(S \in \mathcal{S}_j^1\), then we can apply case 2 of Lemma 2.6.4, to see that the sum of the terms containing \(S\) is at least \(\frac{1}{2}(|S_{j-1}| - |S|/2)(C - .8)|W|\). Thus, we have obtained that summing (2.10) over all active pairs \((X, Y)\) in step \(j - 1\) is at least

\[ 2\alpha \left( \sum_{S \in \mathcal{S}_j^0} |\tilde{S}_j| + \frac{1}{2} \sum_{S \in \mathcal{S}_j^1} \left( |S_{j-1}|-\frac{|S|}{2} \right) \right) (C - .8) |W|. \]

We would like to use Lemma 2.5.3 to show that if we change \(\tilde{G}_j\) to \(G\) everywhere
in (2.10), then it does not decrease by much. That is, if we look at

\[
\sum_{S \in S_d} \left| e_G(S_{j-1},Y, T') - \frac{|S_{j-1,Y}|}{|S''|} e_G(S'', T') \right|
\]

\[+
\sum_{S \in S_d} \left| e_G(S_{j-1},Y, T') - \frac{|S_{j-1,Y}|}{|S_{j-1,Y}^*|} e_G(S_{j-1,Y}^*, T') \right|,
\]

this is at least the value of (2.10) minus \(2 \cdot 0.9 \cdot x_j^{-1/4} |X||Y| \).

Fix \(a, b \in \{0, 1\} \). Let \(T' \) be the set of \(T \in T \) for which \(T' \subseteq Y^b \). For \(d \in \{0, 1\} \), let \(S'_d \) be the set of \(S \in S_d \) for which \(S_{j-1,Y} \subseteq X^a \). Let \(S' = S'_0 \cup S'_1 \). For \(S \in S'_0 \), let \(S' = S_{j-1,Y} \), and for \(S \in S'_1 \), let \(S' = S_{j-1,Y}^* \), \(S'' = S_{j-1,Y}^* \). Note that in both cases \(|S'| \leq |S''| \). Then (2.10) is just summing

\[
\sum_{S \in S'_d} \left| e_{\tilde{G}_j}(S', T') - \frac{|S'|}{|S''|} e_{\tilde{G}_j}(S'', T') \right|
\]

over \(a, b \in \{0, 1\} \). The key observation is that, if \(a \) and \(b \) are fixed, then every term above has the same sign, and hence the above sum is equal to

\[
\left| \sum_{S \in S'_d} \left( e_{\tilde{G}_j}(S', T') - \frac{|S'|}{|S''|} e_{\tilde{G}_j}(S'', T') \right) \right|.
\]

First, we want to say that if we change \(\tilde{G}_j \) to \(G \) in the expression inside the absolute values here, it changes by at most \(2 \cdot 0.9 \cdot x_j^{-1} |X^a||Y^b| \). This is equivalent to the statement

\[
\left| \sum_{S \in S'_d} \left( e_{G-\tilde{G}_j}(S', T') - \frac{|S'|}{|S''|} e_{G-\tilde{G}_j}(S'', T') \right) \right| \leq 2 \cdot 0.9 \cdot x_j^{-1} |X^a||Y^b|.
\]

For each \(S \in S'_d \), let \(S'' \) be a random subset of \(S'' \) of size \(|S'| \), chosen uniformly from all subsets of this size. Then \(\mathbb{E}(e_{G-\tilde{G}_j}(S'', T')) = \frac{|S'|}{|S''|} e_{G-\tilde{G}_j}(S'', T') \) and so the
above expression (without the absolute value signs) is the expected value of

\[ \sum_{S \in S', T \in T'} \left( e_{G-\tilde{G}_j}(S', T') - e_{G-\tilde{G}_j}(S'', T') \right). \]

Let \( T'_0 = \bigcup_{T \in T'} T' \), \( S'_0 = \bigcup_{S \in S'} S' \), \( S''_0 = \bigcup_{S \in S'} S'' \). The previous sum in absolute value equals

\[ \left| e_{G-\tilde{G}_j}(S'_0, T'_0) - e_{G-\tilde{G}_j}(S''_0, T'_0) \right| \leq 2 \cdot 0.9 \cdot x_j^{-1/4} |X^a_Y||Y^b_X|, \]

where the inequality is by Lemma 2.5.3. This holds for any choice of \( S'' \), and so it holds for the expected value. Hence,

\[ \left| \sum_{S \in S', T \in T'} \left( e_{G-\tilde{G}_j}(S', T') - \frac{|S'|}{|S''|} e_{G-\tilde{G}_j}(S'', T') \right) \right| \leq 2 \cdot 0.9 \cdot x_j^{-1/4} |X^a_Y||Y^b_X|. \]

This implies that

\[ \sum_{S \in S', T \in T'} \left| e_G(S', T') - \frac{|S'|}{|S''|} e_G(S'', T') \right| \geq \left| \sum_{S \in S', T \in T'} \left( e_G(S', T') - \frac{|S'|}{|S''|} e_G(S'', T') \right) \right| \]

\[ \geq \left| \sum_{S \in S', T \in T'} \left( e_{G_j}(S', T') - \frac{|S'|}{|S''|} e_{G_j}(S'', T') \right) \right| - 2 \cdot 0.9 \cdot x_j^{-1/4} |X^a_Y||Y^b_X| \]

\[ = \sum_{S \in S', T \in T'} \left| e_{G_j}(S', T') - \frac{|S'|}{|S''|} e_{G_j}(S'', T') \right| - 2 \cdot 0.9 \cdot x_j^{-1/4} |X^a_Y||Y^b_X| \]

Thus, the sum of (2.13) over each active pair \((X, Y)\) in step \( j - 1 \) and \( a, b \in \{0, 1\} \) is at least

\[ 2\alpha \left( \sum_{S \in S_j^0} |S_j| + \frac{1}{2} \sum_{S \in S_j^1} \left( |S_{j-1}| - \frac{|S|}{2} \right) \right) (C - 0.8) |W| - 2 \cdot 0.9 \cdot x_j^{-1/4} \alpha |V||W|. \] (2.14)
Setting $S'(T)$ and $S''(T)$ where applicable to the corresponding term for the set $Y$ with $T' \subset Y$, we have proven the first part of the lemma.

Suppose we just know that every vertex in $W'$ is blue in step $j - 1$. For any pair $(S, T)$, the set $S$ belongs to a single part $X \in \mathcal{P}_{j-1}$, which divides $T'$ into two parts. Setting $T'(S)$ to be the larger of the two, it is at least $\frac{|T'|}{2}$, and so we obtain that (2.10), with $T'$ replaced by $T'(S)$, is at least half of (2.11). Following the rest of the proof, we obtain the analogous lower bound decreased by a factor of two. \square

**Proof of Lemma 2.6.5.** Let $W' \subseteq W$ be the set of blue vertices in step $i$, and define $T'$ as in Lemma 2.6.6. Recall that for each $S \in \mathcal{S}$, we defined $i_S$ to be the largest $i$ for which $S$ has blue vertices in step $i$. We group the parts $S \in \mathcal{S}$ into three types:

A. $i_S \geq i$.

B. $i_S < i$ and $S_{i_S}$ contains at most $\frac{5}{6}|S|$ blue vertices.

C. $i_S < i$ and $S_{i_S}$ contains more than $\frac{5}{6}|S|$ blue vertices.

For each $2 \leq j \leq i$, let $\mathcal{S}_j^0$ be the collection of $S$ of type A, or of type B with $i_S \geq j$. For these sets, let $S'' = S_{i_S}$. Let $\mathcal{S}_j^1$ be the collection of $S$ of type C with $i_S = j - 1$. Let $\mathcal{S}_j$ be the union of these two collections. Let $R_j^0$ be the set of vertices in $R_j$ that belong to some $S \in \mathcal{S}_j^0$ (recall that $R_j$ is the set of vertices that became red in step $j$). Let $R_j^1$ be the set of vertices in $V$ that belong to some $S \in \mathcal{S}_j^1$. Note that it is not necessarily true that the vertices in $R_j^1$ are in $R_j$, however, all vertices in $R_j^1$ will be red by step $j$ (so they will be in $\tilde{R}_j$), and $R_j^1$ for different $j$ will be disjoint from each other (and from the various $R_j^0$). Also, note that while $R_j$ is a subset of $V \cup W$, $R_j^0$ and $R_j^1$ are subsets of just $V$.

We use Lemma 2.6.6 for each $2 \leq j \leq i$ to find, for each $S \in \mathcal{S}_j^0$, subsets $S'(j, T)$, and, for each $S \in \mathcal{S}_j^1$, subsets $S'(T)$ and $S''(T)$, such that
\[
\sum_{S \in S_0^i} \left| e_G(S'(j), T) - \frac{|S'(j)|}{|S''|} e_G(S'', T') \right| \\
+ \sum_{S \in S_1^j} \left| e_G(S'(T), T') - \frac{|S'(T)|}{|S''(T)|} e_G(S''(T), T') \right|
\]

is at least
\[
2\alpha \left( |R^0_j| + \frac{|R^1_j|}{6} \right) (C - .8) |W| - 2 \cdot .9 \cdot x^{-1/4} \alpha |V||W|.
\]

Here we use that
\[
\sum_{S \in S_j^i} |\tilde{S}_j| = |R^0_j|
\]
and that, since for any \( S \in S_j^i \), \(|S_{j-1}| \geq \frac{5}{6}|S|\), we have
\[
\frac{1}{2} \sum_{S \in S_j^i} \left( |S_{j-1}| - \frac{|S|}{2} \right) \geq \frac{\sum_{S \in S_j^i} |S|}{6} \geq \frac{|R^1_j|}{6}
\]

Let us look at the sum of the lower bounds for each \( 2 \leq j \leq i \). Recall that \( R_i \) is the set of vertices that turn red in step \( i \), and \( \tilde{R}_i \) is the set of vertices that are red in step \( i \). In the following discussion, the color of a vertex refers to its color in step \( i \). For an \( S \) of type A, each of its red vertices belong to \( R^0_j \) for some \( j \), since it turned red in some step. For an \( S \) of type B, each of its vertices are red. We know that \( i_S \) was the last step when it still had blue vertices, however, since \(|S_{i_S}| \leq \frac{5}{6}|S|\), we know that at least \( 1/6 \) of the vertices of \( S \) were red by step \( i_S \), so they belong to \( R_j \) for some \( j \leq i_S \). Also, by definition, for \( j \leq i_S \), \( S \) will belong to \( S^0_j \). Thus, all the vertices that turned red by step \( i_S \) will belong to \( R^0_j \) for some \( j \leq i_S \) (recall that \( R^0_j \) is the set of vertices in \( R_j \) that belong to an \( S \in S^0_j \)). To summarize, for an \( S \) of type B, at least \( 1/6 \) of its vertices will belong to \( R^0_j \) for some \( j \). This means that the sum of \(|R^0_j|\) over \( 2 \leq j \leq i \) is at least \( 1/6 \) of the number of red vertices that belong to an \( S \) of type A or B. If \( S \) is of type C, then by definition, each vertex in \( S \) is red, and belongs to \( R^1_i \), and so the sum of \(|R^1_i|\) over \( 2 \leq j \leq i \) is at least the number of red
vertices in an $S$ of type C. Since each $S \in S$ has type A, B, or C, we obtain that the sum of the lower bounds (2.14) over $2 \leq j \leq i$ is at least

$$\frac{1}{3} \alpha |\widetilde{R}_i \cap V| (C - .8) |W| - 2x_2^{-1/4} \alpha |V||W|.$$  

We will use Corollary 2.6.3 to get a lower bound on the irregularity in terms of the sum of (2.13) over active pairs $(X, Y)$ in step $j - 1$, for each $2 \leq j \leq i$. Fix a pair $S$ and $T$. We will find subsets $S^0_T, S^1_T \subset S$ and $T^0_S, T^1_S \subset T$ to obtain a lower bound on the irregularity. Suppose first that $S$ is of type A or B. Thus, we have the term

$$\left| e_G(S'(j, T), T') - \frac{|S'(j, T)|}{|S''|} e_G(S'', T') \right|$$  

(2.15)

for some (possible empty) set of indices $j \in J_S \subset [2, i] \cap \mathbb{Z}$. Let $J_S^+$ be the set of indices for which the expression inside the absolute values is positive, and let $J_S^-$ be the set for which it is negative. Now, if we take $S^0_T$ to be the (disjoint) union of the sets $S'(j, T), j \in J_S^+$, and $S^1_T = S''$, the sum of (2.15) over $j \in J_S^+$ is just

$$\left| e_G(S^0_T, T') - \frac{|S^0_T|}{|S^1_T|} e_G(S^1_T, T') \right|.$$  

The same holds if instead we take $S^0_T$ to be the union of $S'(j, T), j \in J_S^-$, and sum (2.15) over $j \in J_S^-$. Now, the sum of the expression over one of $J_S^+$ or $J_S^-$ is at least half the sum over all of $J_S$, so we let $S^0_T$ be the union over $J_S^+$ or $J_S^-$, whichever one has a larger sum, and $S^1_T = S''$. Notice that $|S^1_T| = |S''| > |S|/2$, and $S^0_T$ is disjoint from $S^1_T$, so $|S^0_T| \leq |S^1_T|$. If $S$ is of type C, then $S$ can only appear if $j = i_S + 1$, so there is at most one term which contains $S$ and $T$. Thus, we can just take $S^0_T = S'(T)$ and $S^1_T = S''(T)$, and in this case, by definition, $|S^0_T| \leq |S^1_T|$. Overall, we lose a factor of 2 in the lower bound (because of the sets $S$ of type A or B), and we obtain a choice of $S^0_T$ and $S^1_T$ with $|S^0_T| \leq |S^1_T|$ for certain pairs $S$ and $T$ such that

$$\sum_{S,T} \left| e_G(S^0_T, T') - \frac{|S^0_T|}{|S^1_T|} e_G(S^1_T, T') \right| \geq \frac{r}{6} (C - .8) \alpha |V||W| - x_2^{-1/4} \alpha |V||W|.$$  

59
Applying Corollary 2.6.3 with $T^0_S = T^1_S = T'$, we obtain the desired lower bound on the irregularity of

$$\frac{r}{12} (C - .8) \alpha |V||W| - \frac{1}{2} x_2^{-1/4} \alpha |V||W|.$$ 

Note that we can apply this corollary as $|S^0_T||T^0_S| \leq |S^1_T||T^1_S|$, which follows from $T^0_S = T^1_S = T'$ and $|S^0_T| \leq |S^1_T|$. This completes the proof. \qed

The next lemma establishes another lower bound on the irregularity between the two partitions. This will be useful if there is a single step in which the number of red vertices increases by a substantial amount. As this lemma obtains a bound on the irregularity by considering a single step, the proof of this lemma is simpler than that of the previous lemma.

**Lemma 2.6.7.** Suppose in step $i$, we have $C|W|$ blue vertices in $W$ and $\beta|V|$ blue vertices in $V$, and in step $i+1$, we only have $\beta'|V|$ blue vertices in $V$, where $\beta > \frac{\beta' + 1}{2}$. Then

$$\text{irreg}(S, T) \geq \frac{1}{4} \left( \beta - \frac{\beta' + 1}{2} \right) (C - .8) \alpha |V||W| - x_{i+1}^{-1/4} \alpha |V||W|.$$ 

**Proof.** For this lemma, we will apply Lemma 2.6.6 once with $j = i + 1$. Let $W' \subset W$ be the set of blue vertices in step $j - 1 = i$. Let $S^0_j = S^0$ be the collection of $S \in S$ with $i_S > i$, and $S^1_j = S^1$ be the collection of those $S \in S$ for which $i_S = i$. Let $S^*$ be the union of these two collections.

So we obtain choices $S'(T), S''(T) \subset S$ and $T'(S) \subset T$ such that

$$\sum_{S \in S^*} \left| e_G(S'(T), T'(S)) - \frac{|S'(T)|}{|S''(T)|} e_G(S''(T), T'(S)) \right|$$

is at least

$$\alpha \left( \sum_{S \in S^0_j} |S_{i+1}| + \frac{1}{2} \sum_{S \in S^1_j} \left( |S_i| - \frac{|S_i|}{2} \right) \right) (C - .8) |W| - 2 \cdot .9 \cdot x_{i+1}^{-1/4} \alpha |V||W|. $$
Note that $\sum_{S \in S^0} |\tilde{S}_{i+1}| + \sum_{S \in S^1} |S_i|$ is exactly the number of new red vertices in step $i + 1$, which is equal to $(\beta - \beta')|V|$. Indeed, any $S$ that has blue vertices still in step $i$ is in one of $S^0$ or $S^1$. If $S$ is in $S^0$, then it contains blue vertices in step $i + 1$, and so the new red vertices from $S$ are $S_i \setminus \tilde{S}_{i+1} = \bar{S}_{i+1}$. If $S$ is in $S^1$, then all of its vertices are red in step $i + 1$. Now, each vertex that belongs to an $S \in S^1$ is red in step $i + 1$, thus the number of such vertices is at most $(1 - \beta')|V|$. That is, $\sum_{S \in S^1} |S| \leq (1 - \beta')|V|$. This gives

$$\sum_{S \in S^0} |\tilde{S}_{i+1}| + \sum_{S \in S^1} \frac{|S_i|}{2} - \sum_{S \in S^1} \frac{|S|}{4} \geq \frac{1}{2} \left( \sum_{S \in S^0} |\tilde{S}_{i+1}| + \sum_{S \in S^1} |S_i| - \sum_{S \in S^1} \frac{|S|}{2} \right) \geq \frac{1}{2} \left( \beta - \beta' - \frac{1-\beta'}{2} \right) |V|.$$ 

Applying Corollary 2.6.3, we obtain

$$\text{irreg}(S, T) \geq \frac{1}{4} \left( \beta - \frac{1 + \beta'}{2} \right) (C - .8) \alpha|V||W| - .9x^{-1/4}_i \alpha|V||W|.$$ 

This completes the proof. \qed

Proof of Lemma 2.6.1. Now, we show that, assuming $S$ refines $P_1$, and $T$ refines $Q_1$, they must be close to being refinements of $P_s$ and $Q_s$, and cannot have too few parts. Let $v_i$ and $w_i$ be the proportion of vertices that are blue in step $i$ in $V$ and $W$, respectively. Then, since vertices can turn red but can never turn blue, this gives us two non-increasing sequences. We want to show that if the irregularity is less than $\frac{1}{5000} \alpha|V||W|$, then both sequences stay above $.97$, so in particular $v_s$ and $w_s$ are at least $.97$. If there is a step where $w_i \geq .9$, but $v_i \leq .97$, then using Lemma 2.6.5, and recalling that $x_2 = 2^{26}$, the irregularity is at least

$$\frac{.03}{12}(.9 - .8)\alpha|V||W| - \frac{1}{2} x_2^{-1/4} \alpha|V||W| = \frac{1}{4000} \alpha|V||W| - \frac{1}{2} x_2^{-1/4} \alpha|V||W| \geq \frac{1}{5000} \alpha|V||W|.$$ 

By the same logic, this also holds if $v_i \geq .9$ and $w_i \leq .97$. Thus, the only way it is
possible that one of \( v_s, w_s \) is below \(.97 \) is if there is an \( i \) such that \( v_i \) and \( w_i \) are both at least \(.97 \), but \( v_{i+1} \) and \( w_{i+1} \) are both at most \(.9 \). Then, using just the fact that \( w_i \geq .9 \), \( v_i \geq .97 \), \( v_{i+1} \leq .9 \), we can apply Lemma 2.6.7 to obtain that the irregularity is at least

\[
\frac{1}{4} \left( .97 - \frac{.9 + 1}{2} \right) (.9 - .8) \alpha |V| |W| - .9 x_{i+1}^{-1/4} \alpha |V| |W| \\
= \left( \frac{1}{2000} - .9 x_{i+1}^{-1/4} \right) \alpha |V| |W| \geq \frac{1}{5000} \alpha |V| |W|.
\]

Thus, if the irregularity is less than \( \frac{1}{5000} \alpha |V| |W| \), then both \( v_s \) and \( w_s \) are at least \(.97 \), and so \( S \) and \( T \) are in fact close to being refinements of \( P_s \) and \( Q_s \) respectively. To each part \( S \in S \) that has blue vertices, there is a unique part \( X \in P_s \) with \( S_s \subset X \). In other words, if two blue vertices belong to different parts of \( P_s \), then they belong to different parts of \( S \). Since at least a \(.97 \) proportion of the vertices are blue, at least \(.97 |P_s| \) parts of \( P_s \) have blue vertices (since each part in \( P_s \) has the same size), and so we have at least \(.97 |P_s| \) parts in \( S \). Applying the analogous argument, we have at least \(.97 |Q_s| \) parts in \( T \).

Now, in order to finish the proof of the main result, we need the following lemma.

**Lemma 2.6.8.** Let \( G \) be a graph with a vertex partition \( P \), and \( Q \) is a refinement of \( P \) such that each part of \( P \) is divided into at most \( k \) parts. If \( Q \) has irregularity \( I \), then \( P \) has irregularity at least \( \frac{I}{2k^2} \).

**Proof.** Let \( X \) and \( Y \) be two parts of \( P \). The idea is that since we divide each part into at most \( k \) parts, we can find a pair of parts of \( Q \) in \( X \) and \( Y \) that have large irregularity, and use this to show that \( X \) and \( Y \) have some irregularity. To be precise, let \( \mathcal{X} \) be the parts of \( P' \) that are subsets of \( X \), and \( \mathcal{Y} \) be the parts of \( P' \) that are subsets of \( Y \). Define

\[
I_{XY} = \sum_{\substack{X' \in \mathcal{X} \, Y' \in \mathcal{Y}}} \text{irreg}(X', Y').
\]

Now, since each part of \( P \) is divided into at most \( k \) parts, one of the terms of this sum is at least \( I_{XY} / k^2 \). Let \( X' \) and \( Y' \) be a pair of elements of \( \mathcal{X} \) and \( \mathcal{Y} \) for which
this holds. Now, this means that there are subsets \( X'' \subset X', Y'' \subset Y' \) such that

\[
\left| e(X'', Y'') - \frac{|X''||Y''|}{|X'||Y'|} e(X', Y') \right| \geq \frac{I_{XY}}{k^2}.
\]

Now, we can apply Lemma 2.6.2 to show that

\[
\text{irreg}(X, Y) \geq \frac{I_{XY}}{2k^2}.
\]

Adding this up for all \( X \) and \( Y \), we obtain

\[
\text{irreg}(P) = \sum_{X, Y \in P} \text{irreg}(X, Y) \geq \sum_{X, Y \in P} \frac{I_{XY}}{2k^2} = \frac{1}{2k^2} \sum_{X, Y \in P} \sum_{X' \in X} \sum_{Y' \in Y} \text{irreg}(X', Y')
\]

\[
= \frac{1}{2k^2} \sum_{X', Y' \in P'} \text{irreg}(X', Y') = \frac{1}{2k^2} \text{irreg}(P').
\]

\( \square \)

**Proof of Theorems 3.5 and 1.2.** Consider a partition \( P \) with less than \( \frac{.97}{4x_1}(|P_s| + |Q_s|) \) parts. Let \( P' \) be the common refinement of \( P \) and \( P_1 \cup Q_1 \). Then \( P' \) divides each part of \( P \) into at most \( 4x_1 \) parts. It follows that \( P' \) has less than \( 4x_1 \left( \frac{.97}{4x_1} \right) (|P_s| + |Q_s|) = .97(|P_s| + |Q_s|) \) parts. By Lemma 2.6.1, \( P' \) has irregularity at least \( \frac{1}{5000} |V||W| \). By Lemma 2.6.8, \( P \) has irregularity at least \( \frac{1}{5000} \frac{1}{32x_1^4} \alpha |V||W| = \frac{1}{241} \frac{1}{10000} \alpha |V||W| > \epsilon (|V| + |W|)^2 \) (recall that \( x_1 = 2^{10} \) and \( \alpha > 2^{26} 10^4 \epsilon \)). This is a contradiction, and so we have completed the proof of Theorem 2.3.5.

Now, to finish the proof of Theorem 1.2, we first prove by induction that \( x_k \geq 2^8 T(k + 1) \) holds for each positive integer \( k \). Indeed, this is true for the base case \( k = 1 \), since \( x_1 = 2^{10} = 2^8 T(2) \). Since \( x_k \geq x_1 = 2^{10} \), we have \( x_k/16 \geq 2^{-8} x_k + 8 \). So, by the induction hypothesis, we have

\[
x_{k+1} = 2^{x_k/16} \geq 2^{2^{-8} x_k + 8} \geq 2^{T(k+1)+8} = 2^8 T(k + 2),
\]

proving the claimed bound.
By Theorem 2.3.5, the number of parts any vertex partition with irregularity at most $\epsilon |V(G)|^2$ is at least $\frac{97}{4x_1}(|P_s|+|Q_s|) > \frac{1}{4x_1}|P_s| = 2^{-12}|P_s| \geq 2^{-12}x_s \geq 2^{-12}2^8T(s+1) \geq T(s)$, where the last inequality holds for $s \geq 3$. As $s \geq 10^{-26}\epsilon^2$, this proves Theorem 2.1.2.

2.7 Equitable partitions with small irregularity

The main result of this section shows that for any vertex partition, one can refine it a bit further to obtain a partition which is close to an equitable partition whose irregularity is not substantially larger.

**Theorem 2.7.1.** Let $0 < \alpha < 1/2$, and $m$ be a positive integer, and let $G$ be a graph on $n \geq 10^8\alpha^{-5}$ vertices. If $P$ is a vertex partition of $G$ into $m$ parts, then there is an equitable vertex partition $Q$ of $G$ into at most $4m/\alpha$ parts such that $\text{irreg}(Q) \leq \text{irreg}(P) + \alpha n^2$.

Let $M_{eq}(\epsilon)$ be the smallest $M$ such that, for any graph $G = (V, E)$, there is an equitable partition into at most $M$ parts with total irregularity is at most $\epsilon |V|^2$. We have $M_{eq}(\epsilon) \geq M(\epsilon)$ trivially. As a consequence of Theorem 2.7.1, we show directly that adding the condition that the partition is equitable has a very small effect on the size of the smallest partition with small irregularity.

**Theorem 2.7.2.** Let $0 < \epsilon < 1$ and $0 < \alpha < 1/2$. We have $M_{eq}(\epsilon+\alpha) \leq \alpha^{-O(1)}M(\epsilon)$.

In particular, taking $\alpha$ small but not too small, such as $\alpha = 2^{-1/\epsilon}$, we see that the tower height in Szemerédi’s regularity lemma is not significantly affected by adding the equitability requirement.

Note that Theorem 2.7.1 also applies to graphs $G$ whose number of parts in the regularity partition is not as large as the worst case $M(\epsilon)$. To prove Theorem 2.7.1, we randomly divide each part of the partition $P$ into parts of (essentially) equal size (apart from a small remaining subset), and then arbitrarily partition the relatively
few remaining vertices into parts of equal size to obtain an equitable partition. We show that this works with high probability.

As a first step, the following lemma shows that with high probability, a pair of random subsets \( X', Y' \) of a pair of parts \( X, Y \) induces roughly the same subgraph density as \( X \) and \( Y \).

**Lemma 2.7.3.** Let \( X \) and \( Y \) be vertex subsets of a graph \( G \). Let \( X' \subseteq X \) and \( Y' \subseteq Y \) be picked uniformly at random with \( |X'| = |Y'| = k \). Then

\[
\Pr(|d(X', Y') - d(X, Y)| < \delta) \geq 1 - 2e^{-\delta^2 k/4}.
\]

**Proof.** Consider picking the vertices of \( X' \) and \( Y' \) one at the time, starting with the ones in \( X' \). Let \( Z_0, \ldots, Z_{2k} \) be the martingale where \( Z_i \) is the expected value of \( e(X', Y') \) conditioned on knowing the first \( i \) vertices already chosen (this is sometimes referred to as the vertex-exposure martingale). We have \( |Z_i - Z_{i-1}| \leq k \) as the choice of each vertex in \( X' \) and \( Y' \) changes the final \( e(X', Y') \) by at most \( k \). By the Azuma–Hoeffding inequality (see Chapter 7 of Alon and Spencer [9]),

\[
\Pr(|Z_{2k} - Z_0| \geq t) \leq 2e^{-t^2/(4k^3)}.
\]

We have \( Z_0 = k^2 d(X, Y) \) and \( Z_{2k} = e(X', Y') \). Set \( t = \delta k^2 \), we obtain

\[
\Pr(|e(X', Y') - k^2 d(X, Y)| \geq \delta k^2) \leq 2e^{-\delta^2 k/4}.
\]

Noting that \( e(X', Y') = k^2 d(X', Y') \), the lemma follows. \( \square \)

The next lemma show that the irregularity parameter remains roughly the same when restricted to a random, much smaller, subset of vertices. Define the cut metric \( d_\Box \) between two graphs \( G \) and \( H \) on the same vertex set \( V = V(G) = V(H) \) by

\[
d_\Box(G, H) := \max_{U, W \subseteq V} |e_G(U, W) - e_H(U, W)| / |V|^2.
\]
When $G$ and $H$ are bipartite graphs on $V = X \cup Y$, we define the cut metric as
\[
d_{\square}(G, H) := \max_{U \subseteq X, W \subseteq Y} \frac{|e_G(U, W) - e_H(U, W)|}{|X||Y|}.
\]
We also use the same notation for edge-weighted graphs, where $e(U, W)$ denotes the sum of weights of all edges in $U \times W$.

**Lemma 2.7.4.** Let $X, Y$ be vertex subsets of a graph $G$. Let $X' \subseteq X$ and $Y' \subseteq Y$ be picked uniformly at random with $|X'| = |Y'| = k$. Then with probability at least $1 - 6e^{-\sqrt{k}/10}$,
\[
\left| \frac{\text{irreg}(X', Y')}{k^2} - \frac{\text{irreg}(X, Y)}{|X||Y|} \right| \leq \frac{9}{k^{1/4}},
\]
(2.16)

**Proof.** We use the so-called First Sampling Lemma [18, Theorem 2.10] (we quote the statement from [51, Lemma 10.5]): if $G$ and $H$ are weighted graphs with $V(G) = V(H)$ and edge weights in $[0, 1]$, and $S \subseteq V(G)$ is chosen uniformly at random with $|S| = k$, then with probability at least $1 - 4e^{-\sqrt{k}/10}$,
\[
|d_{\square}(G[S], H[S]) - d_{\square}(G, H)| \leq \frac{8}{k^{1/4}}.
\]
Let $G[X, Y]$ denote the bipartite (weighted) graph with vertex sets $X$ and $Y$, and whose edges are induced from $G$. A bipartite version of this sampling lemma holds true, that with probability at least $1 - 4e^{-\sqrt{k}/10}$,
\[
|d_{\square}(G[X', Y'], H[X', Y']) - d_{\square}(G[X, Y], H[X, Y])| \leq \frac{8}{k^{1/4}},
\]
(2.17)
and its proof is nearly identical to the first version stated above. Note that
\[
\text{irreg}(X, Y) = |X||Y|d_{\square}(G[X, Y], d(X, Y)),
\]
where the second argument denotes the complete graph with loops with all edge weights equal to $d(X, Y)$. Similarly,
\[
\text{irreg}(X', Y') = k^2d_{\square}(G[X', Y'], d(X', Y')).
\]
By letting $H$ in (2.17) be the complete graph with loops and all edge weights $d(X, Y)$, we obtain that with probability at least $1 - 4e^{-\sqrt{k}/10}$,

$$\left| \frac{\text{irreg}(X', Y')}{k^2} - \frac{\text{irreg}(X, Y)}{|X||Y|} \right| \leq \frac{8}{k^{1/4}} + |d(X, Y) - d(X', Y')|.$$ 

We then apply Lemma 2.7.3 with $\delta = 1/k^{1/4}$ to reach the desired conclusion. 

As a corollary of Lemma 2.7.3, and noting that the LHS of (2.16) is at most 1, we have

$$\mathbb{E}\left( \frac{\text{irreg}(X', Y')}{{k^2}} \right) \leq \frac{\text{irreg}(X, Y)}{|X||Y|} + \frac{9}{k^{1/4}} + 6e^{-\sqrt{k}/10} \leq \frac{\text{irreg}(X, Y)}{|X||Y|} + \frac{20}{k^{1/4}}. \quad (2.18)$$

Proof of Theorem 2.7.1. We shall omit floors and ceilings for the sake of clarity of presentation. Let $k = an/(4m)$. Let $V_1, \ldots, V_m$ be the parts of $\mathcal{P}$. Uniformly at random partition each $V_i$ into parts of size $k$, with possibly one remainder part of size less than $k$. Call $\mathcal{P}'$ be the resulting partition.

By (2.18),

$$\mathbb{E}\left( \sum_{X', Y' \in \mathcal{P}'} \text{irreg}(X', Y') \right) \leq \sum_{X, Y \in \mathcal{P}} \left( \text{irreg}(X, Y) + \frac{20}{k^{1/4}} |X||Y| \right) \leq \text{irreg}(\mathcal{P}) + \frac{20n^2}{k^{1/4}}.$$ 

So there exists some such partition $\mathcal{P}'$ such that

$$\sum_{X', Y' \in \mathcal{P}'} \text{irreg}(X', Y') \leq \text{irreg}(\mathcal{P}) + \frac{20n^2}{k^{1/4}}. \quad (2.19)$$ 

Fix $\mathcal{P}'$ to be this partition.

Let $S$ be the union of the parts of $\mathcal{P}'$ of size less than $k$. There is at most one such part of $\mathcal{P}'$ for each $V_i$, so $|S| < mk$. Arbitrarily partition $S$ into sets of size $k$, and let $\mathcal{Q}$ be the equitable vertex partition consisting of these parts of $S$ along with

\footnote{It is easy to modify the proof to address the case when $X'$ and $Y'$ are within the same part of $\mathcal{P}$.}
the parts of $\mathcal{P}'$ of size $k$.

Parts arising from $S$ contributes at most $2|S|n < 2mkn$ to $\text{irreg}(Q)$, whereas the other contributions to $\text{irreg}(Q)$ are bounded by (2.19). Thus

$$\text{irreg}(Q) \leq \text{irreg}(\mathcal{P}) + \frac{20n^2}{k^{1/4}} + 2mkn \leq \text{irreg}(\mathcal{P}) + \alpha n^2,$$

where the last step follows from

$$\frac{20}{k^{1/4}} = \frac{20(4m)^{1/4}}{(\alpha n)^{1/4}} \leq \frac{20(4m)^{1/4}}{(\alpha \cdot 10^8 m\alpha^{-5})^{1/4}} < \frac{\alpha}{2}$$

and

$$\frac{2mk}{n} \leq \frac{2m n}{4m} \leq \frac{\alpha}{2}.$$

Therefore $Q$ is the required equipartition. \hfill \Box

As a consequence of Theorem 2.7.1, we can prove Theorem 2.7.2.

**Proof of Theorem 2.7.2.** Let $G$ be a graph with a partition $\mathcal{P}$ of the vertex set into $m \leq M(\varepsilon)$ parts with irregularity at most $en^2$. If $n < 10^8 m\alpha^{-5}$, then we just partition the vertices into singleton sets, which has zero irregularity, using at most $10^8 m\alpha^{-5} \leq 10^8 \alpha^{-5} M(\varepsilon)$ parts. Otherwise apply Theorem 2.7.1 to obtain a partition with irregularity at most $(\varepsilon + \alpha)n^2$, and at most $4m/\alpha \leq 4\alpha^{-1} M(\varepsilon)$ parts. \hfill \Box

### 2.8 Concluding Remarks

In this chapter, we determined that the number $M(\varepsilon)$ of parts in Szemerédi’s regularity lemma, Theorem 2.1.1, grows as a tower of twos of height $\Theta(\varepsilon^{-2})$. The upper bound we proved is a tower of height $2 + \varepsilon^{-2}/16$, and the lower bound is a tower of twos of height $10^{-26}\varepsilon^{-2}$. We next sketch how to improve the constant factors in both the upper and lower bounds.

We first discuss the upper bound, whose proof is given in Section 2.2. Recall that the range of possible values for the mean square density is $[d^2, d]$, where $d$ is the edge density of the graph. To maximize the length of this interval, we should have $d = 1/2$. 

68
One can easily check that the upper bound argument is tight only if for every pair of parts \((X, Y)\) in a partition at a certain step, there are subsets \(X' \subset X\) and \(Y' \subset Y\) such that \(|X'||Y'| = |X||Y|/2\) and \(|c(X', Y') - d(X, Y)|X'||Y'|| = \epsilon|X||Y|\). It would then follow that \(|d(X', Y') - d(X, Y)| = 2\epsilon\). The distribution of densities between pairs of parts in the partition in step \(i\) would then follow a binomial distribution with mean \(1/2\) and \(i\) steps of size \(2\epsilon\). In other words, it would be \(1/2\) plus the sum of \(i\) independent random variables, where each random variable is \(2\epsilon\) or \(-2\epsilon\) with probability \(1/2\) for each outcome. This is not feasible, as the densities are in the interval \([0, 1]\), and after \(\Omega(\epsilon^{-2})\) steps, a constant fraction of the pairs go between pairs of parts whose density would be outside the interval \([0, 1]\). Optimizing this argument to get a best possible upper bound would be more complicated. We also remark that this issue of densities having to be between 0 and 1 is the same as one of the main technical difficulties of the lower bound proof, the fact that there are inactive pairs.

We next discuss the lower bound. Various parts of our lower bound proof do not attempt to optimize the constant factor in the tower height bound for \(M(\epsilon)\). One particularly large loss (a factor \(x_1^2\), where \(x_1\) is the size of the first partition in the construction) comes from picking a large \(x_1\) and assuming that the partition in question is a refinement of the partition in the first step. Instead, we could pick a first partition with \(x_1\) parts, and pick the graph in the first step to have densities bounded away from 0 and 1 and have the property that any refinement with small irregularity is close to being a refinement of this given partition. We chose the current presentation without optimal constant factors for the sake of clarity.

The original version of Szemerédi’s regularity lemma

The original version of the regularity lemma can be stated in a more general form as follows. We say a pair \((X, Y)\) of vertex subsets form an \((\epsilon, \delta)\)-regular pair if there is \(\alpha \in [0, 1]\) such that for any subsets \(U \subset X, W \subset Y\), such that \(|U| \geq \delta|X|, |W| \geq \delta|Y|\), we have \(d(U, W)\) lies between \(\alpha\) and \(\alpha + \epsilon\). Now, given \(\epsilon, \delta, \eta > 0\), call a partition into \(k\) parts \((\epsilon, \delta, \eta)\)-regular if all but at most \(\eta k^2\) pairs of parts are \((\epsilon, \delta)\)-regular. The lemma states that for any \(\epsilon, \delta, \eta\), there is a \(K = K(\epsilon, \delta, \eta)\) such that there is
an equitable partition into $k \leq K$ parts that is $(\varepsilon, \delta, \eta)$-regular. Again, it is not too difficult to see that if the irregularity of an equitable partition is at most $\frac{1}{2} \varepsilon \delta^2 \eta |V|^2$, then it is $(\varepsilon, \delta, \eta)$-regular. Conversely, if $\varepsilon = \delta = \eta$, then if an equitable partition is $\varepsilon$-regular, it has irregularity at most $2 \varepsilon |V|^2$.

Gowers [38] showed with $c = 1/16$ that there is a graph $G$ whose smallest equitable $(1 - \delta^c, \delta, 1 - \delta^c)$-regular partition is at least a tower of height $\delta^{-c}$. Addressing a question of Gowers, Conlon and Fox [19] showed that there are absolute constants $\varepsilon, \delta > 0$ such that there is a graph $G$ whose smallest equitable $(\varepsilon, \delta, \eta)$-regular partition is at least a tower of height $\Omega(\eta^{-1})$, which determines the right tower height as a function of $\eta$.

In the case $\varepsilon = \delta = \eta$, the upper bound proof gives a tower height of $O(\varepsilon^{-5})$. Our result in Theorem 2.1.2 directly implies a new lower bound on the tower height in this case of $\Omega(\varepsilon^{-2})$. In fact, one can combine the construction ideas in Theorem 2.1.2 and that in [19], and one should get a tower height of $\Omega(\varepsilon^{-3})$. The additional idea from [19] is that instead of adding and subtracting densities for every active pair at each step, we only do this for a random fraction (on the order of $\eta$) of the active pairs at each step. A further variant using ideas similar to those of Gowers by incorporating $\delta$ into the construction should improve the exponent beyond 3, but we do not yet see how to match the upper bound exponent of 5.

Acknowledgement: We would like to thank Noga Alon and David Conlon for helpful discussions.
Chapter 3

Algorithmic regularity

3.1 Introduction

As we have seen, Szemerédi’s regularity lemma [66] is one of the most powerful tools in graph theory. Szemerédi [65] used an early version in the proof of his celebrated theorem on long arithmetic progressions in dense subsets of the integers. Roughly speaking, the regularity lemma says that every large graph can be partitioned into a bounded number of parts such that the bipartite subgraph between almost every pair of parts is random-like.

In this chapter, we will study the classical version of the lemma, and the (weak) Frieze-Kannan regularity lemma. We first recall some terminology, and the statement of the lemma. Let $G$ be a graph, and let $X$ and $Y$ be (not necessarily disjoint) vertex subsets. Let $e(X, Y)$ denote the number of pairs vertices $(x, y) \in X \times Y$ that are edges of $G$. The edge density $d(X, Y) = e(X, Y)/(|X||Y|)$ between $X$ and $Y$ is the fraction of pairs in $X \times Y$ that are edges. The pair $(X, Y)$ is $\epsilon$-regular if for all $X' \subseteq X$ and $Y' \subseteq Y$ with $|X'| \geq \epsilon|X|$ and $|Y'| \geq \epsilon|Y|$, we have $|d(X', Y') - d(X, Y)| \leq \epsilon$. Qualitatively, a pair of parts is $\epsilon$-regular with small $\epsilon$ if the edge densities between pairs of large subsets are all roughly the same. A vertex partition $V = V_1 \cup \ldots \cup V_k$ is equitable if the parts have size as equal as possible, that is we have $||V_i| - |V_j|| \leq 1$ for all $i, j$. An equitable vertex partition with $k$ parts is $\epsilon$-regular if all but $\epsilon k^2$ pairs of parts $(V_i, V_j)$ are $\epsilon$-regular. The regularity lemma states that for every $\epsilon > 0$
there is a (least) integer \( K(\epsilon) \) such that every graph has an \( \epsilon \)-regular equitable vertex partition into at most \( K(\epsilon) \) parts. As we saw in Chapter 2, \( K(\epsilon) \) is an exponential tower of twos of height between \( \Omega(\epsilon^{-2}) \) and \( O(\epsilon^{-5}) \), arguably the main drawback of Szemerédi’s regularity lemma.

Due to the many applications of the regularity lemma, there has been a great deal of research on developing algorithmic versions of the regularity lemma and its applications (see the survey by Komlós and Simonovits [48]). We would like to be able to find an \( \epsilon \)-regular partition of a graph on \( n \) vertices in time polynomial in \( n \). Szemerédi’s original proof of the regularity lemma was not algorithmic. The reason for this is that it needs to be able to check if a pair of parts is \( \epsilon \)-regular, and if not, to use subsets of the parts that realize this. This is problematic because it is shown in [3] that determining whether a given pair of parts is \( \epsilon \)-regular is co-NP-complete. They use this to show that checking whether a given partition is \( \epsilon \)-regular is co-NP-complete.

However, Alon, Duke, Lefmann, Rödl, and Yuster [3] show how to find, if a given pair of vertex subsets of size \( n \) are not \( \epsilon \)-regular, a pair of subsets which realize that the pair is not \( \epsilon^4/16 \)-regular. The running time is \( O_\epsilon(n^\omega+o(1)) \), where \( \omega < 2.373 \) is the matrix multiplication exponent (multiplying two \( n \times n \) matrices in \( n^{\omega+o(1)} \) time) [21, 50]. Here we use the subscript \( \epsilon \) to mean that the hidden constants depend on \( \epsilon \). Finding a pair of subsets of vertices that detect irregularity is the key bottleneck for the algorithmic proof of the regularity lemma. It was shown [3] that one can find an \( \epsilon \)-regular partition with the number of parts at most an exponential tower of height \( O(\epsilon^{-20}) \) in an \( n \)-vertex graph in time \( O_\epsilon(n^{\omega+o(1)}) \). Thus, the following surprising fact holds: while checking whether a given partition is \( \epsilon \)-regular is co-NP-complete, finding an \( \epsilon \)-regular partition can be done in polynomial time.

Frieze and Kannan [34] later found a simple algorithmic proof of the regularity lemma based on a spectral approach. Using expander graphs, Kohayakawa, Rödl, and Thoma [47] gave a faster algorithmic regularity lemma with optimal running time of \( O_\epsilon(n^2) \). Alon and Naor [4] develop an algorithm which approximates the cut norm of a graph within a factor 0.56 using Grothendieck’s inequality and apply this to find a
polynomial time algorithm which finds, for a given pair of vertex subsets of order \( n \) which is not \( \epsilon \)-regular, a pair of subsets which realize that the pair is not \( \epsilon^3/2 \)-regular. They further observe that their approach gives an improvement on the tower height in the algorithmic regularity lemma to \( O(\epsilon^{-7}) \).

However, due to the tower-type dependence for the number of parts on the regularity parameter, these are not practical algorithms. While most graphs have a small regularity partition, the previous algorithmic proofs would not necessarily find it and would only guarantee to find a regular partition with a tower-type number of parts. Addressing this issue, Fischer, Matsliah, and Shapira \cite{27} give a probabilistic algorithm which runs in constant time (depending on \( \epsilon \) and \( k \)) which with high probability finds, in a graph which has an \( \epsilon/2 \)-regular partition with \( k \) parts, an \( \epsilon \)-regular partition with at most \( k \) parts (implicitly defined). Tao \cite{68} gives a probabilistic algorithm which with high probability in constant time (depending on \( \epsilon \)) produces an \( \epsilon \)-regular partition. The algorithm takes a random sample of vertices (the exact number of which is also random) and outputs the common refinement of the neighborhoods of these vertices.

Still, it is desirable to have a fast deterministic algorithm for finding a regularity partition, which we obtain here. We give several deterministic approximation algorithms for these co-NP-complete problems.

**Theorem 3.1.1.** There exists an \( O_{\epsilon,\alpha,k}(n^2) \) time algorithm, which, given \( 0 < \epsilon, \alpha < 1 \) and \( k \), and a graph \( G \) on \( n \) vertices that admits an equitable \( \epsilon \)-regular partition with \( k \) parts, outputs an equitable \( (1 + \alpha)\epsilon \)-regular partition of \( G \) into \( k \) parts.

In other words, if a graph has a regular partition with few parts, then we can quickly find a regular partition (losing very slightly on the regularity) with the same number of parts. In particular, we obtain an algorithmic regularity lemma which is optimal in terms of the number of parts as it is exactly the same as in the non-algorithmic version (with a very slight loss on the regularity parameter).

We also give an approximation algorithm for checking whether a given pair of vertex subsets is \( \epsilon \)-regular, in the sense that if the pair is not \( \epsilon \)-regular, then we can
algorithmically find a pair of vertex subsets that witness that its failure to be \((1 - \alpha)\varepsilon\)-regular. We will formulate this in terms of regularity of bipartite graphs. We say that bipartite graph \(G\) with bipartition \((X, Y)\) is \(\varepsilon\)-regular if the pair \((X, Y)\) is \(\varepsilon\)-regular.

**Theorem 3.1.2.** There exists an \(O_{\varepsilon, \alpha}(n^2)\) time algorithm, which, given \(\varepsilon, \alpha > 0\), and a bipartite graph \(G\) between vertex sets \(X\) and \(Y\), each of size at most \(n\), outputs one of the following:

1. Correctly states that \(G\) is \(\varepsilon\)-regular;
2. Finds a pair of vertex subsets \(U \subseteq X\) and \(W \subseteq Y\) which realize that \(G\) is not \((1 - \alpha)\varepsilon\)-regular, i.e., \(|U| \geq (1 - \alpha)\varepsilon|X|\), \(|W| \geq (1 - \alpha)\varepsilon|Y|\), and \(|d(U, W) - d(X, Y)| > (1 - \alpha)\varepsilon\).

Using this result, by checking the regularity of each pair of parts in a partition, we have the following corollary.

**Corollary 3.1.3.** Given \(\varepsilon, \alpha > 0\), we can distinguish in time \(O_{\varepsilon, \alpha}(n^2)\) between an \(\varepsilon\)-regular partition and a partition which is not \((1 - \alpha)\varepsilon\)-regular. \(\square\)

Sometimes, instead of requiring a partition to be \(\varepsilon\)-regular, the weaker notion of Frieze-Kannan, or weak, regularity suffices. To state this precisely, first, we extend the definition of \(e(X, Y)\) and \(d(X, Y)\) to weighted graphs. Below by *weighted graph* we mean a graph with edge-weights. Given two sets of vertices \(X\) and \(Y\), we let \(e(X, Y)\) denote the sum of the edge-weights over pairs \((x, y) \in X \times Y\) (taking 0 if a pair does not have an edge). Let \(d(X, Y) = e(X, Y)/(|X||Y|)\) as earlier. Recall that the cut metric \(d_{\Box}\) between two graphs \(G\) and \(H\) on the same vertex set \(V = V(G) = V(H)\) is defined by

\[
\max_{U, W \subseteq V} \frac{|e_G(U, W) - e_H(U, W)|}{|V|^2},
\]

and this extends to graphs with weighted edges, and can be adapted to bipartite graphs (with given bipartitions). Given any edge-weighted graph \(G\) and any partition \(\mathcal{P}: V = V_1 \cup V_2 \cup \cdots \cup V_t\) of the vertex set of \(G\) into \(t\) parts, let \(G_{\mathcal{P}}\) denote the weighted graph with vertex set \(V\) obtained by giving weight \(d_{ij} := d(V_i, V_j)\) to all
pairs of vertices in $V_i \times V_j$, for every $1 \leq i \leq j \leq t$. We say $\mathcal{P}$ is an $\epsilon$-regular Frieze–Kannan (or $\epsilon$-FK-regular) partition if $d_{\square}(G, G_{\mathcal{P}}) \leq \epsilon$. In other words, $\mathcal{P}$ is an $\epsilon$-regular Frieze–Kannan partition if

$$\left| e(S, T) - \sum_{i,j=1}^{t} d_{ij}|S \cap V_i||T \cap V_j| \right| \leq \epsilon|V|^2. \quad (3.1)$$

for all $S, T \subseteq V$. We say that sets $S$ and $T$ witness that $\mathcal{P}$ is not $\epsilon$-FK-regular if the above inequality is violated.

Frieze and Kannan [32, 33] proved the following regularity lemma.

**Theorem 3.1.4 (Frieze–Kannan).** Let $\epsilon > 0$. Every graph has an $\epsilon$-regular Frieze–Kannan partition with at most $2^{2/\epsilon^2}$ parts.

In addition to proving that the partitions exist, Frieze and Kannan gave probabilistic algorithms for finding weak regular partitions [32, 33]. A deterministic algorithm was given by Dellamonica, Kalyanasundaram, Martin, Rödl, and Shapira [23, 24]. Specifically, in [23], the authors gave an $\epsilon^{-6}n^{\omega+o(1)}$ time algorithm to generate an equitable $\epsilon$-regular Frieze–Kannan partition of a graph on $n$ vertices into at most $2^{O(\epsilon^{-7})}$ parts. Recall that $\omega < 2.373$ is the matrix multiplication exponent. In [24] a different algorithm was given which improved the dependence of the running time on $n$ from $O_\epsilon(n^{\omega+o(1)})$ to $O_\epsilon(n^2)$, while sacrificing the dependence of $\epsilon$. Namely, it was shown that there is a deterministic algorithm that finds, in $O(2^{\epsilon^{-O(1)}} n^2)$ time, an $\epsilon$-regular Frieze–Kannan partition into at most $2^{\epsilon^{-O(1)}}$ parts.

There is a variant of the weak regularity lemma, where the final output is not a partition of $V$ into $2^{\epsilon^{-O(1)}}$ parts, but rather an approximation of the graphs as a sum of $\epsilon^{-O(1)}$ complete bipartite graphs, each assigned some weight, see [33]. For $S, T \subseteq V$, we denote by $K_{S,T}$ the weighted graph where an edge $\{s, t\}$ has weight 1 if $s \in S$ and $t \in T$ (and weight 2 if $s, t \in S \cap T$) and weight zero otherwise. For any $c \in \mathbb{R}$, by $cG$ we mean the weighted graph obtained from $G$ by multiplying every edge-weight by $c$. For a pair of weighted graphs $G_1, G_2$ on the same set of vertices, we will use the notation $G_1 + G_2$ to denote the graph on the same vertex set with edge
weights summed (and weight 0 corresponding to not having an edge). Additionally, we write \( c \) to mean the constant graph with all edge-weights equal to \( c \).

**Theorem 3.1.5 (Frieze–Kannan).** Let \( \epsilon > 0 \). Let \( G \) be any weighted graph with \([-1,1]\)-valued edge weights. There exists an \( r \leq O(\epsilon^{-2}) \), and there exist subsets \( S_1, \ldots, S_r, T_1, \ldots, T_r \subseteq V \), and \( c_1, \ldots, c_k \in [-1,1] \), so that

\[
d_{\infty}(G, d(G) + c_1 K_{S_1,T_1} + \cdots + c_r K_{S_r,T_r}) \leq \epsilon.
\]

See [52, Lemma 4.1] for a simple proof (given there in a more general setting of arbitrary Hilbert spaces). It is well known (see, e.g., [33]) that given sets and numbers as in the theorem, the common refinement of all \( S_i, T_i \) must be a \( 2\epsilon \)-regular partition.

In Section 3.2, we give an optimal algorithm that provides the best of both worlds: We give an algorithm that runs in time \( \epsilon^{-O(1)} n^2 \), and finds a weakly regular partition. In fact, we prove something slightly stronger, which in some situations can be useful:

**Theorem 3.1.6.** There exists an algorithm which, given \( \epsilon > 0 \) and an \( n \)-vertex graph \( G \), outputs subsets \( S_1, S_2, \ldots, S_r, T_1, T_2, \ldots, T_r \subseteq V(G) \) and \( c_1, c_2, \ldots, c_r = \pm \frac{4}{\sqrt{n}} \) for some \( r \leq O(\epsilon^{-16}) \), such that

\[
d_{\infty}(G, d(G) + c_1 K_{S_1,T_1} + \cdots + c_r K_{S_r,T_r}) \leq \epsilon.
\]

The running time of the algorithm is \( \epsilon^{-O(1)} n^2 \).

**Remark.** As in the case of the usual regularity lemma, it is possible to obtain an equitable partition in the Frieze–Kannan regularity lemma, increasing the number of parts and the cut distance by a negligible amount. We will not need this for our algorithm, however.

Counting the number of copies of a graph \( H \) in another graph \( G \) is a famous problem in algorithmic graph theory. For example, a special case of this problem is to determine the clique number, the size of the largest clique, in a graph. This is a well-known NP-complete problem. In fact, Håstad [41] and Zuckerman [69] proved
that it is NP-hard to approximate the clique number of a $n$-vertex graph within a factor $n^{1-\epsilon}$ for any $\epsilon > 0$.

There is a fast *probabilistic* algorithm for approximating up to $\epsilon$ the fraction of $k$-tuples which make a copy of $H$. The algorithm takes $s = 10\epsilon^{-2}$ samples of $k$-tuples of vertices uniformly at random from $G$ and outputs the fraction of them that make a copy of $H$. The number of copies of $H$ is a binomial random variable with standard deviation at most $s^{1/2}/2$, and hence the fraction of $k$-tuples which make a copy of $H$ in this random sample is likely within $\epsilon$ of the fraction of $k$-tuples which makes copies of $H$. However, this algorithm has no guarantee of success. It is therefore desirable to have a *deterministic* algorithm for counting copies which gives an approximation for the subgraph count with complete certainty.

The algorithmic regularity lemma is useful for deterministically approximating the number of copies of any fixed graph in a graph. Indeed, the counting lemma shows that if $k$ parts $V_1, \ldots, V_k$ are pairwise regular, then the number of copies of a graph $H$ with $k$ vertices with the copy of the $i$th vertex in $V_i$ is close to what is expected in a random graph with the same edge densities between the pairs of parts. Adding up over all $k$-tuples of parts in an $\epsilon$-regular partition and noting that almost all $k$-tuples of parts have all its pairs $\epsilon$-regular, we get an algorithm which runs in time $O_{\epsilon,k}(n^2)$ which computes the number of copies of a graph $H$ on $k$ vertices in a graph on $n$ vertices up to an additive error of $\epsilon n^k$. The major drawback with this result is the tower-type dependence on $\epsilon$ and $k$, which comes from the number of parts in the regularity lemma.

Duke, Lefmann, and Rödl [25] gave a faster approximation algorithm for the number of copies of $H$ in a graph $G$. They first develop a weak regularity lemma which has an exponential dependence instead of a tower-type dependence. This gives an algorithm which runs in time $2^{(k/\epsilon)\Omega(1)} n^{\omega+o(1)}$ which computes the number of copies of a graph $H$ on $k$ vertices in a graph on $n$ vertices up to an additive error of $\epsilon n^k$.

In Section 3.3, we will use the algorithmic version of the Frieze-Kannan weak regularity lemma [23] to get the following even faster approximation algorithm for the subgraph counting problem. It improves the previous exponential dependence on
the error parameter to a polynomial dependence. Here \( v(H) \) and \( e(H) \) denote the number of vertices and edges in \( H \), respectively.

**Theorem 3.1.7.** Let \( H \) be a graph, and let \( \epsilon > 0 \) be given. There is a deterministic algorithm that runs in time \( \epsilon^{-O_H(1)}n^2 \), and finds the number of copies of \( H \) in \( G \) up to an error of at most \( \epsilon n^{v(H)} \).

An examination of the proof shows that the exponent is on the order of \( 9|E(H)| \). This implies, for example, that we can count the number of cliques of order 1000 in an \( n \)-vertex graph up to an additive error \( n^{1000-10^{1000000}} \) in time \( O(n^{2.1}) \).

### 3.2 Algorithmic weak regularity

In this section, we prove Theorem 3.1.6. We will apply these results in subsequent sections. We will prove the following, roughly equivalent form. In order to state it, we first give some notation. Given a matrix \( A \), we denote

\[
\|A\|_{p \to q} = \sup_{v \in \mathbb{R}^n \setminus \{0\}} \frac{\|Av\|_q}{\|v\|_p}.
\]

We also denote \( \|A\|_p = \|A\|_{p \to p} \). It is well known that \( \|A\|_2 \) is equal to the largest singular value of \( A \), i.e. the spectral norm. We also use the Frobenius norm

\[
\|A\|_F = \sqrt{\sum_{i,j} a_{i,j}^2}.
\]

and the entry-wise maximum norm

\[
\|A\|_{\text{max}} = \sup_{i,j} |a_{i,j}|.
\]

Throughout this section, given a set \( S \subseteq [n] \), we will denote by \( \chi_S \) the characteristic vector of \( S \). Also, \( \pm x \) denotes either \( x \) or \( -x \).

**Theorem 3.2.1.** There exists an algorithm which, given an \( \epsilon > 0 \) and a matrix \( A \in [-1,1]^{n \times n} \), outputs subsets \( S_1, \ldots, S_r, T_1, \ldots, T_r \subseteq [n] \) and real numbers \( c_1, \ldots, c_r = \ldots \)
\( \pm \frac{\epsilon^8}{300}, \) for some \( r \leq O(\epsilon^{-16}), \) such that for

\[
A' = \sum_{i=1}^{r} \chi_i \chi_{i}^T
\]

we have that each row and column of \( A - A' \) has square sum (the sum of the squares of the entries) at most \( n \), and

\[
\|A - A'\|_2 \leq \epsilon n.
\]

The running time of the algorithm is \( \epsilon^{-O(1)} n^2 \).

In Section 2.4, we essentially showed that if \( G \) and \( H \) are weighted bipartite graphs between two sets \( X, Y \) of size \( n \), and \( A_G, A_H \) are the adjacency matrices, with rows corresponding to \( X \) and columns corresponding to \( Y \), then

\[
d_{\square}(G, H) \leq \frac{\|A_G - A_H\|_2}{n}.
\]

We assumed that \( H \) is constant weight and \( G \) is an unweighted graph with the same average density, but this is not necessary for the proof. Therefore, this theorem indeed implies Theorem 3.1.6 (taking \( A \) to be \( A_G - d(G)11^T \)).

The proof of the Frieze–Kannan regularity lemma and its algorithmic versions, roughly speaking, run as follows:

- Given a partition (starting with the trivial partition with one part), either it is \( \epsilon \)-FK-regular (in which case we are done), or we can exhibit some pair of subsets \( S, T \) of vertices that witness the irregularity by violating (3.1) (in the algorithmic versions, one may only be guaranteed to find \( S \) and \( T \) that violate (3.1) for some smaller value of \( \epsilon \)).

- Refine the partition by using \( S \) and \( T \) to split each part into at most four parts, thereby increasing the total number of parts by a factor of at most 4.

- Repeat. Use a mean square density increment argument to upper bound the
number of possible iterations.

This can be modified to prove Theorem 3.1.5. Roughly speaking, to find the appropriate $S_i, T_i, c_i$, in the second step of the above outline of the proof of the weak regularity lemma, instead of using $S$ and $T$ to refine the existing partition, we subtract $c\chi_S\chi_T^T$ from the remaining matrix, for a carefully chosen $c$. We record the corresponding $S_i, T_i, c_i$ in step $i$ of this iteration. We can bound the number of iterations by observing that the $L^2$ norm of $A - c_1\chi_{S_1}\chi_{T_1}^T - \cdots - c_i\chi_{S_i}\chi_{T_i}^T$ must decrease by a certain amount at each step.

As for the algorithmic versions, the main challenge is checking whether a partition is regular, or a cut graph approximation is close in cut distance. Given a matrix $A$, up to a polynomial change in $\epsilon$, having small singular values as a fraction of $n$ is equivalent to $\text{tr} AA^T AA^T$ being small as a fraction of $n^4$, which roughly says that most scalar products of rows are small as a fraction of $n$. In [3], the authors use this fact to obtain an algorithm in $O(n^{4+o(1)})$ time which either correctly states that a pair of parts is regular, or gives a pair which is off by $\epsilon^{O(1)}$. This was adapted in [23] to the weak regular setting. In [47], the authors noticed that it suffices to check the scalar products along a well-chosen expander, which has a linear number of edges in terms of $n$, allowing them to obtain the $O(n^2)$-time algorithm. This was also the main idea in [24], but their algorithm is double exponential in $\epsilon^{-1}$. A further challenge in proving Theorem 3.2.1 with the cut matrix approximation is that the entries of the approximation matrices may not stay bounded, which was used in the algorithms for checking regularity. This is problematic, because for a general matrix $A$, the singular value (divided by $n$) and the cut norm may be different. To counter this, we give an algorithm which checks regularity effectively under a weaker assumption that simply the square sum of each row and each column stays bounded. Heuristically, the reason this property is useful is that it implies that if we have a singular vector (with norm 1) with a relatively large singular value, then no entry can be “too large”; it must be “spread out”, so a large singular value implies a large cut norm. We then show that if we are careful, we can make sure that this property holds throughout the process.

Let us state this more precisely. Given a matrix $A$, let $a_i$ be the $i$-th row of $A$
and $a^i$ the $i$-th column. Our main ingredient then is the following theorem. Note that in the algorithm below, the parameter $C$ affects the running time but not the discrepancy of the output sets $S, T$.

**Theorem 3.2.2.** There exists a $(C/\epsilon)^{O(1)} n^2$ algorithm which, given a matrix $A \in \mathbb{R}^{n \times n}$ such that $\|A\|_{\text{max}} \leq C$, and each $\|a_i\|_2^2 \leq n$, $\|a^i\|_2 \leq n$ (or equivalently $\|A^TA\|_{\text{max}}, \|AA^T\|_{\text{max}} \leq n$), either

- Correctly outputs that each singular value of $A$ is at most $cn$.
- Outputs sets $S, T \subseteq [n]$ such that

$$\left| \sum_{i \in S, k \in T} a_{i,k} \right| \geq \frac{c^8}{100} n^2.$$

(Which implies that $A$ has a singular value that is at least $\frac{c^8}{100} n$.)

In the next proposition, we construct the expander along which we will check the scalar products. Let $J$ denote the matrix with each entry equal to 1.

**Lemma 3.2.3.** There exists fixed $l$ and $0 < c < 1$ such that there is an algorithm which given $d_0$ and $n$, constructs a multigraph $G$ on $[n]$, such that $d_0 \leq d \leq l d_0$ and if $M$ is the adjacency matrix, then

$$\|\frac{d}{n} J - M\| \leq d^{1-c}.$$

In other words, for any vector $v = (v_i)_{i=1}^n \in \mathbb{R}^n$, we have

$$\left| \frac{n}{d^c} v^T J_d v - \left( \sum_i v_i \right)^2 \right| \leq \frac{n}{d^c} \|v\|_2^2. \quad (3.2)$$

The running time of the algorithm is $dn (\log n)^{O(1)}$.

**Proof.** Construct an $l$-regular two-sided expander $G_0$ on $[n_0]$ for some $n \leq n_0 \leq Kn$ with $K$ fixed. This can be done in $n (\log n)^{O(1)}$ time. For example, Margulis [54]
constructed an 8-regular expander on $\mathbb{Z}_m \times \mathbb{Z}_m$ for every $m$, and Gabber and Galil [36] showed that all other eigenvalues (besides 8 with multiplicity 1) are at most $5\sqrt{2} < 8$. For every vertex $(x, y) \in \mathbb{Z}_m \times \mathbb{Z}_m$, its eight neighbors are

$$(x \pm 2y, y), (x \pm (2y + 1), y), (x, y \pm 2x), (x, y \pm (2x + 1)).$$

Therefore we can compute, for each vertex, a list of neighbors in time $O(\log m) = O(\log n)$, which then takes $O(n \log n)$ time total. Alternatively, we can start with a Ramanujan graph for some fixed degree, constructed explicitly by Lubotzky, Phillips, and Sarnak [53]; Margulis [55]; and Morgenstern [56].

The adjacency matrix $A_{G_0}$ will have the property for an explicit $a < l$ (independent of $n$) that

$$A_{G_0} 1 = l 1,$$

and all the other eigenvalues have absolute value at most $a$. Let $M = A_{G_0}^k$ so that $d_0 \leq d = l^k < ld_0$. Note that $M$ is symmetric and has nonnegative integer entries, so it is the adjacency matrix of some graph $G$ (possibly with multiple edges). Clearly

$$M 1 = d 1.$$

and all other eigenvalues have absolute value at most $a^k = a^{\log_l (d)} = d^{\log_l (a)}$. Since $a < l$, $c := \log_l (a) < 1$.

We can construct $G_0$ in time $(\log n)^{O(1)} n$. We make sure, for each vertex, to keep a list of its neighbors. We then compute $A_{G_0}^i$ for $i = 1, 2, ..., k$. In each case, we make sure to keep a list of the $3^i$ neighbors of each vertex (with multiplicities). We can then compute $A_{G_0}^{i+1}$ in $O(3^i n)$ time by computing the list of $3^{i+1}$ neighbors for each vertex, by looking at its three neighbors in $G_0$ and taking the (multiset) union. The total running time is therefore $O(n + \sum_{i=1}^{k} 3^i n) = O(d n)$.

Proof of Theorem 3.2.2. The basic idea of the algorithm is the following. It is easy

82
to see that

\[ \text{tr}(AA^T A A^T) = \sum_{i,j,k,l} a_{i,k} a_{i,l} a_{j,k} a_{j,l}. \]  

(3.3)

In order to estimate this sum, we can use the expander to only compute the sum for pairs \((i, j)\) which form an edge of the expander (and then multiply by \(n/d\)). In fact, this is true even for the terms in (3.3) corresponding to a fixed \(k, l\). We can therefore use the expander to estimate the sum in (3.3), and if it is large, find a \(k\) for which the sum of the terms corresponding to \(k\) are large. This will allow us to find sets \(S, T\) as required.

Here is the algorithm.

1. Construct \(G\) for \(d_0 = (3C^2 \epsilon^{-4})^{1/c}\). Let \(M = (m_{i,j})_{i,j=1}^n\).

2. For each \(i, j\) with \(m_{i,j} > 0\), compute \(s_{i,j} = \langle a_i, a_j \rangle\).

3. For each \(k \in [n]\), compute

\[ b_k = \sum_{i,j=1}^n m_{i,j} a_{i,k} a_{j,k} s_{i,j}. \]

4. If each \(b_k \leq \frac{2}{3} \epsilon^4 d n^2\), return that \(\|A\|_2 \leq \epsilon n\).

5. If some \(b_k \geq \frac{2}{3} \epsilon^4 d n^2\), compute for each \(l\)

\[ c_l = \sum_i a_{i,k} a_{i,l}. \]

6. Let \(T\) be either the set of \(l\) such that \(c_l > 0\), or the set \(l\) such that \(c_l < 0\), whichever has a bigger sum in absolute value.

7. Let \(S\) be either the set of \(i \in [n]\) such that \(d_T(i) > 0\) or the set of \(i \in [n]\) such that \(d_T(i) < 0\), whichever has a bigger sum in absolute value.
Let us first analyze the running time. We can construct $G_d$ in time $(\log n)^{O(1)}dn$. We can compute each $s_{i,j}$ in $O(n)$ time, so computing all of them takes $O(dn^2)$ time in total. Computing each $b_k$ then similarly takes $O(dn)$ time (since we only need to sum the terms where $m_{i,j} > 0$, and we keep a list of these entries), so that takes $O(dn^2)$ total. If the algorithm says that $\|A\|_2 \leq \epsilon n$, then we are done. Otherwise, computing each $c_l$ can be done in time $O(n)$ so that takes $O(n^2)$ time in total. We then obtain $T$ in $O(n)$ time. Computing $S$ then similarly takes $O(n^2)$ time. Since $d = (C/\epsilon)^{O(1)}$, this shows that the algorithm runs in time $(C/\epsilon)^{O(1)}n^2$.

We now show that the algorithm is correct. First, we show the following lemma, which makes precise that we can use the expander to estimate the sum (3.3).

**Lemma 3.2.4.** For any $k, l \in [n]$, we have

$$\left| \sum_{i,j} a_{i,k}a_{i,l}a_{j,k}a_{j,l} - \frac{n}{d} \sum_{i,j} m_{i,j}a_{i,k}a_{i,l}a_{j,k}a_{j,l} \right| \leq \frac{C^2n^2}{d^2} \leq \frac{\epsilon^4}{3}n^2. \quad (3.4)$$

**Proof.** Let $a_{k,l}$ be the vector with entries $(a_{k,l})_i = a_{i,k}a_{i,l}$. Since each $|a_{i,j}| \leq C$, we have that $\|a_{k,l}\|_2^2 \leq C^2n$. Therefore, by (3.2),

$$\left| \left( \sum_i a_{i,k}a_{i,l} \right)^2 - \frac{n}{d} a_{k,l}^TMa_{k,l} \right| \leq \frac{C^2n^2}{d^2}. \quad .$$

Clearly

$$\left( \sum_i a_{i,k}a_{i,l} \right)^2 = \sum_{i,j} a_{i,k}a_{i,l}a_{j,k}a_{j,l},$$

and by the definition of $M$ and $a_{k,l}$, we have

$$a_{k,l}^TMa_{k,l} = \sum_{i,j} m_{i,j}a_{i,k}a_{i,l}a_{j,k}a_{j,l}.$$ 

\[\square\]

**Lemma 3.2.5.** If the algorithm returns that $\|A\|_2 \leq \epsilon n$ then it is correct.
Proof. We first claim that
\[
\sum_{i,j} m_{i,j} (a_i, a_j)^2 \leq \frac{2}{3} \epsilon^4 d n^3.
\]
Indeed, we have
\[
\sum_{i,j} m_{i,j} (a_i, a_j)^2 = \sum_{i,j} m_{i,j} a_i k a_j k (a_i, a_j) = \sum_k b_k \leq \frac{2}{3} \epsilon^4 d n^3.
\]
Now, if we add (3.4) up over all pairs \(k, l \in [n]\), we have
\[
\left| \sum_{k, l \in [n]} a_{i,k} a_{i,l} a_{j,k} a_{j,l} - \frac{n}{d} \sum_{k, l \in [n]} m_{i,j} a_{i,k} a_{i,l} a_{j,k} a_{j,l} \right| \leq \frac{\epsilon^4}{3} n^4.
\]
Note that
\[
\sum_{k, l \in [n]} m_{i,j} a_{i,k} a_{i,l} a_{j,k} a_{j,l} = \sum_{i,j} m_{i,j} (a_i, a_j)^2 \leq \frac{2}{3} \epsilon^4 d n^3.
\]
Therefore,
\[
\text{tr} A A^T A A^T = \sum_{i,j,k,l} a_{i,k} a_{i,l} a_{j,k} a_{j,l} \leq \frac{n}{d} \sum_{i,j} m_{i,j} (a_i, a_j)^2 + \frac{\epsilon^4}{3} n^4 \leq \epsilon^4 n^4.
\]
Since \(\text{tr} A A^T A A^T\) is the sum of the fourth powers of the singular values, this implies that each singular value is at most \(\epsilon n\).

Lemma 3.2.6. If the algorithm returns \(S, T\), then
\[
\left| \sum_{i,k \in S \times T} a_{i,k} \right| \geq \frac{\epsilon^8}{100} n^2.
\]
Proof. First, note that we have
\[
\frac{2}{3} \epsilon^4 d n^2 \leq b_k = \sum_{i,j} m_{i,j} a_{i,k} a_{j,k} b_{i,j} = \sum_{i,j,l} m_{i,j} a_{i,k} a_{j,k} a_{i,l} a_{j,l}.
\]
we claim that we have
\[ \sum_{i,j \in [n]} a_{i,k}a_{j,k}a_{i,l}a_{j,l} \geq \frac{n}{d} \sum_{i,j} m a_{i,k}a_{j,k}a_{i,l}a_{j,l} - \frac{\epsilon^4}{3} n^3 \geq \frac{\epsilon^4}{3} n^3. \]

Indeed, for any fixed \( l \), we have by (3.4)
\[ \left| \sum_{i,j \in [n]} a_{i,k}a_{j,k}a_{i,l}a_{j,l} - \frac{n}{d} \sum_{i,j} m a_{i,k}a_{j,k}a_{i,l}a_{j,l} \right| \leq \frac{\epsilon^4}{3} n^2, \]

and we can add this up over all \( l \in [n] \). Let \( u = (a_{i,k})_{i=1}^n \), and \( v \) be the vector with coordinates

\[ v_l = \sum_i a_{i,k}a_{i,l}. \]

Then \( \|v\|_\infty \leq n \), and \( \|u\|_2 \leq \sqrt{n} \) and we have
\[ u^T A v \geq \frac{\epsilon^4}{3} n^3. \]

This implies that the \( T \) that we obtain in step 6 has, if \( \chi_T \) is the characteristic vector,
\[ |u^T A \chi_T| \geq \frac{\epsilon^4}{6} n^2. \]

Since \( \|u\|_2 \leq \sqrt{n} \), this implies that
\[ \|A \chi_T\|_2^2 \geq (u^T A \chi_T)^2 \|u\|_2^2 \geq \frac{\epsilon^8}{36} n^3. \]

Since each row \( a_i \) of \( A \) has \( \|a_i\|_2 \leq \sqrt{n} \), we also have that
\[ \|A \chi_T\|_\infty \leq \sqrt{n} \|\chi_T\|_2 \leq n. \]

Therefore, we have
\[ \|A \chi_T\|_1 \geq \frac{\|A \chi_T\|_2^2}{\|A \chi_T\|_\infty} \geq \frac{\epsilon^8}{36} n^2. \]

This means that for the \( S \) that we obtain in step 7, we have, if \( \chi_S \) is the characteristic
vector,
\[ |x_S^T A x_T| \geq \frac{\epsilon^8}{72} n^2 \geq \frac{\epsilon^8}{100} n^2, \]
which is what we wanted to show.

We have seen that either output of the algorithm must be correct, so this completes the proof of Theorem 3.2.2.

Proof of Theorem 3.2.1. Let \( \epsilon' = \epsilon^8 / 100 \). Let \( A_0 = A \). We construct a sequence of matrices \( A_l \) such that for each \( l \geq 1 \), we have the following.

1. There exists \( S_l, T_l \) such that
   \[ A_{l+1} = A_l \pm \frac{\epsilon'}{3} K_{S_l, T_l}. \]

2. Each entry of \( A_l \) has absolute value at most \( 1 + l \epsilon' / 3 \).

3. The Frobenius norms of the matrices must decrease:
   \[ \|A_{l+1}\|_F^2 \leq \|A_l\|_F^2 - \frac{\epsilon'^2}{3} n^2. \]

4. If \( v \) is a row or column of \( A_l \), then
   \[ \|v\|_2^2 \leq n. \]

We claim that we can keep doing this until we can guarantee that for some \( l \), we have \( \|A_l\|_2 \leq \epsilon n \). Suppose that we have matrix \( A_l \). Since each entry of \( A_l \) is at most \( 1 + l \epsilon' / 3 \) in absolute value, by Theorem 3.2.2 (with \( C = 1 + l \epsilon' / 3 \)), there exists an algorithm that runs in \( (l / \epsilon)^{O(1)} n^2 \) time and either guarantees that \( \|A_l\|_2 \leq \epsilon n \), or
finds a pair of sets $S, T$ such that

$$\left| \sum_{i \in S, k \in T} a_{i,k} \right| \geq \epsilon' n^2.$$  

Without loss of generality, suppose that this sum is in fact positive. For each $i \in S$, and each $k \in T$, let us store the sum of the corresponding row or column in the restricted matrix. Check whether there is a row or column with sum less than $\epsilon' n$. If there is, delete it, and update the row or column sums by subtracting the corresponding element from each sum. Keep doing this until no such row or column remains.

In each step, the sum decreases by at most $\frac{\epsilon'}{6} n$, and there are at most $2n$ steps. Therefore, after this process, for the $S$ and $T$ that we kept, we must still have

$$\sum_{S \times T} a_{i,j} \geq \frac{2}{3} \epsilon' n^2.$$  

In particular, this implies that $S$ and $T$ cannot be empty. We furthermore now have the property that for any $i \in S$,

$$\sum_{j \in T} a_{i,j} \geq \frac{\epsilon'}{6} n,$$

and for any $j \in T$,

$$\sum_{i \in S} a_{i,j} \geq \frac{\epsilon'}{6} n.$$

Set $A_{t+1} = A_t - tK_{S,T}$ for $t = \frac{\epsilon'}{3}$. Clearly (1) is satisfied with $S_{t+1} = S, T_{t+1} = T$, and (2) follows by induction. For (3), note that

$$\|A_t\|_F - \|A_{t+1}\|_F = \sum_{i \in S, j \in T} a_{i,j}^2 - (a_{i,j} - t)^2 = 2t \sum_{i \in S, j \in T} a_{i,j} - |S||T|t^2 \geq \frac{4}{3} t\epsilon' n^2 - |S||T|t^2 \geq \left( \frac{4}{3} t\epsilon' - t^2 \right) n^2.$$
With our choice of $t = \frac{\epsilon'}{3}$, this implies that

$$||A_{i+1}||_F \leq ||A_i||_F - \frac{\epsilon'^2}{3}n^2.$$  

Finally, let $i$ be any row. If $i \notin S$, then the row doesn’t change, so the $L_2$ norm doesn’t change. If $i \in S$, then the difference is

$$\sum_{j \in T} a_{i,j}^2 - (a_{i,j} - t)^2 = 2t \sum_{j \in T} a_{i,j} - |T|t^2 \geq 2t \frac{\epsilon'}{6} n - t^2 n = t \left( \frac{\epsilon'}{3} - t \right) n = 0.$$  

Therefore, the square sum of each row cannot increase. Since for $A_0$ the square sum is at most $n$ (since each entry has absolute value at most 1), it must say below $n$.

If the sum in $S \times T$ is negative, then we first make sure the sum in each row or column is less than $-\frac{\epsilon'}{6} n$, then we add $tK_{S,T}$ and can prove the properties similarly.

Now, we must have $||A_0||_F^2 \leq n^2$. Since this decreases by at least $\frac{\epsilon'^2}{3}n^2 = \frac{\epsilon'^6}{30000}n^2$ in each step, the number of steps is at most $O(1/\epsilon^{16})$.

As for the running time, finding the pair $S, T$ takes at most $(l/\epsilon)^{O(1)n^2}$ time. Computing the sum of each row and column takes $O(n^2)$ time. Each time we delete a row or column, we can update the sums in linear time, therefore the whole process of row and column deletions takes $O(n^2)$ time. Finally, the number of steps is $O(\epsilon^{-16})$, so the whole process takes $\epsilon^{-O(1)n^2}$ time (since $l$ stays below $O(\epsilon^{-16})$ as well). 

### 3.3 Approximation algorithm for subgraph counts

Suppose that we are given a graph $G$, on $n$ vertices, and we would like to approximate the number of copies of a small graph $H$, on $k$ vertices, that are contained in $G$. We would like to count them up to an error at most $\epsilon n^k$. In this section, we will provide a deterministic algorithm that can do so. Specifically, we prove Theorem 3.1.7, reproduced below for convenience.

**Theorem.** Let $H$ be a graph, and let $\epsilon > 0$ be given. There is a deterministic algorithm that runs in time $\epsilon^{-O_H(1)n^2}$, that finds the number of copies of $H$ in $G$ up
to an error of at most $\epsilon n^{v(H)}$.

It will be cleaner to work instead with $\text{hom}(H, G)$, the number of graph homomorphisms from $H$ to $G$. This quantity differs from the number of (labeled) copies of $H$ in $G$ by a negligible $O(n^{v(H)-1})$ additive error (the hidden constants here and onward may depend on $H$). We extend the definition of $\text{hom}(H, G)$ to edge-weighted graphs $G$: if the edge $xy$ in $G$ has weight $G(x, y)$, then we define

$$\text{hom}(H, G) = \sum_{f:V(H) \rightarrow V(G)} \prod_{(u, v) \in E(H)} G(f(u), f(v)).$$

Note that here $G(x, y)$ is defined on all pairs, with $G(x, y) = 0$ if there is no edge between $x$ and $y$. We will actually work with a multi-partite version, as follows.

**Definition 3.3.1.** Let $H$ be a graph on $[k]$, let $V_i$ for $i \in [k]$ be sets of size at most $n$. Suppose for each $i, j$ we are given bipartite graphs $G_{i,j}$ between $V_i$ and $V_j$, let $G$ be the union of the graphs. We define

$$\text{hom}(H, G(V_1, V_2, \ldots, V_k)) = \sum_{v_i \in V_i} \prod_{(i, j) \in E(H)} G(v_i, v_j).$$

**Theorem 3.3.2.** Let $H, V_i, G_{i,j}, G$ be as above, and let $\epsilon > 0$ be given. There exists a deterministic algorithm that computes $\text{hom}(H, G(V_1, V_2, \ldots, V_k))$ up to an error of at most $\epsilon n^k$ in time $\epsilon^{-O_H(1)} n^2$.

*Proof.* We prove this by induction on the number of edges of $H$. If $H$ has no edges then the count is exactly $\prod_i |V_i|$. Let $e = \{i, j\}$ be an edge. We can obtain a weighted graph $G'_{i,j}$ such that $d_{\square}(G_{i,j}, G'_{i,j}) \leq \epsilon/2$ and there exists $r \leq 90000\epsilon^{-16}$ and subsets $S_l \subseteq V_i, T_l \subseteq V_j$ for $1 \leq l \leq r$ such that

$$G'_{i,j} = d(G_{i,j}) + \sum_{l=1}^r \pm \frac{\epsilon}{300} K_{S_l, T_l}.$$

Set $G'$ to be $G$ with $G_{i,j}$ replaced with $G'_{i,j}$. Recall that the cut metric has the
following equivalent form (see for example [51]).

\[ d_{\square}(G_{i,j}, G'_{i,j}) = \frac{1}{|V_i||V_j|} \sup_{u \in [0,1]^{|V_i|}, v \in [0,1]^{|V_j|}} \left| u^T A_{G_{i,j}} w - u^T A_{G'_{i,j}} w \right|. \]

Fix a choice of \( v_i \in V_i \) for \( l \neq i, j \). If we set \( u = (u_i)_{i=1}^n \) with

\[ u_i = \prod_{\{i',j'\} \in E(H)} G(v_i, v_{j'}) \]

and \( w = (w_i)_{i=1}^n \) with

\[ w_i = \prod_{\{i',j\} \in E(H)} G(v_{i'}, v_j), \]

then we have

\[
\left| \sum_{v_i \in V_i} \prod_{v_j \in V_j, \{i',j'\} \in E(H)} G'(v_{i'}, v_j) - \sum_{v_i \in V_i} \prod_{v_j \in V_j, \{i',j'\} \in E(H)} G(v_{i'}, v_j) \right|
\]

\[
= \left| \sum_{v_i \in V_i} \prod_{v_j \in V_j, \{i',j'\} \in E(H), i' \neq i} G(v_{i'}, v_j) (G'(v_{i'}, v_j) - G(v_{i'}, v_j)) \right|
\]

\[
\leq \prod_{\{i',j\} \in E(H)} G(v_{i'}, v_j).
\]

\[
\leq d_{\square}(G_{i,j}, G'_{i,j}) |V_i||V_j| \leq \frac{\epsilon}{2} n^2.
\]

Here we used the fact that \( 0 \leq G \leq 1 \) everywhere. Therefore, adding this up for all choices of \( v_l \in V_l \) for \( l \neq i, j \), we have

\[ |\hom(H, G(V_1, V_2, ..., V_k)) - \hom(H, G'(V_1, V_2, ..., V_k))| \leq \frac{\epsilon n^k}{2}. \]
Therefore, it suffices to compute \( \text{hom}(H, G'(V_1, V_2, ..., V_k)) \) up to an error of at most \( \frac{\epsilon}{2} n^k \). Let \( V_{i,l} \subseteq V_i = S_i \) and \( V_{j,l} \subseteq V_j = T_j \), and \( H' = H \setminus \{e\} \). Then

\[
\text{hom}(H, G'(V_1, V_2, ..., V_k)) = d(G_i, G_j) \text{hom}(H', G(V_1, V_2, ..., V_k)) + \sum_{l=1}^{r} \frac{\epsilon^8}{300} \text{hom}(H', G_i(V_1, V_2, ..., V_{i,l}, ..., V_{j,l}, ..., V_k)).
\]

By induction on the number of edges of \( H \), we can compute an estimate for the values \( \text{hom}(H', G(V_1, V_2, ..., V_k)) \) and each \( \text{hom}(H', G(V_1, V_2, ..., V_{i,l}, ..., V_{j,l}, ..., V_k)) \) with an error of at most \( \epsilon^9/100000 \), in time \( \epsilon^{-O_{\mathcal{H}(1)}} n^2 \). Adding these up, we obtain an estimate for \( \text{hom}(H, G'(V_1, V_2, ..., V_k)) \) with an error at most \( \frac{\epsilon}{2} n^k \), which is therefore an estimate for \( \text{hom}(H, G'(V_1, ..., V_k)) \) with error at most \( \epsilon n^k \). In terms of time, it takes \( \epsilon^{-O(1)} n^2 \) time (independent of \( H \)) to find \( G'_{i,j} \). We then have to compute an estimate for the counts of \( H' \) in various graphs \( O(\epsilon^{-O(1)}) \) times, each of which can be done in time \( \epsilon^{-O_{\mathcal{H}(1)}} n^2 \). \( \square \)

Proof of Theorem 3.1.7. Given a graph \( H \) (which we can assume to have vertex set \([k]\)), and a graph \( G \) on \( V \), take \( k \) copies of \( V, V_1, V_2, ..., V_k \). We define a graph \( G' = (G_{i,j})_{1 \leq i < j \leq k} \) as follows. For each edge \( \{u, v\} \in E(G) \), there are corresponding vertices \( u_i, v_i \in V_i, u_j, v_j \in V_j \). We add an edge between \( u_i \) and \( v_j \), and an edge between \( u_i \) and \( u_j \) (if \( G \) is weighted then with the corresponding weight). It is not difficult to check that then \( \text{hom}(H, G) = \text{hom}(H, G'(V_1, V_2, ..., V_k)) \), so our algorithm above gives an estimate for \( \text{hom}(H, G) \) that has an error of at most \( \epsilon n^k \). \( \square \)

3.4 Finding an irregular pair

In this section, we prove Theorem 3.1.2, reproduced below for convenience.

**Theorem.** There exists an \( O_{\epsilon, \alpha}(n^2) \) time algorithm, which, given \( \epsilon, \alpha > 0 \), and a bipartite graph \( G \) between vertex sets \( X \) and \( Y \), each of size at most \( n \), outputs one of the following:

1. Correctly states that \( G \) is \( \epsilon \)-regular;
2. Finds a pair of vertex subsets $U \subseteq X$ and $W \subseteq Y$ which realize that $G$ is not $(1 - \alpha)\epsilon$-regular, i.e., $|U| \geq (1 - \alpha)\epsilon|X|$, $|W| \geq (1 - \alpha)\epsilon|Y|$, and $|d(U, W) - d(X, Y)| > (1 - \alpha)\epsilon$.

We can assume that $\alpha < 1/2$ since for larger $\alpha$ we can just apply the algorithm with a lower value of $\alpha$. We shall give an $O(2^{2(\alpha)^{-O(1)}} n^2)$-time algorithm. Using Theorem 3.1.6, we approximate $G$ by $G' = d(G) + c_1K_{S_1,T_1} + \ldots + c_rK_{S_r,T_r}$ so that $r = (\alpha\epsilon)^{-O(1)}$ and $d_\square(G, G') \leq \alpha^3/4$. Here $S_1, \ldots, S_r \subseteq X$ and $T_1, \ldots, T_r \subseteq Y$. We shall assume that $r$ is small compared to $|X|$ and $|Y|$, namely,

$$100 \cdot r^2 \leq \alpha \epsilon^3 \min\{|X|, |Y|\}, \tag{3.5}$$

for otherwise we can accomplish the task by a complete search (say when $|X| \leq |Y|$) over all subsets of $X$ in $2^{O(|X|)} = 2^{O(\alpha^{-1}\epsilon^{-3}r^2r)} = 2^{O(\alpha\epsilon)^{-O(1)}}$ time, which is enough.

We say that a sequence of numbers $u, u_1, \ldots, u_r, w, w_1, \ldots, w_r$ is feasible if there exists a function $\mu : X \cup Y \to [0, 1]$ (we write $\mu(S) = \sum_{x \in S} \mu(x)$ from now on) such that the following quantities

$$\frac{\mu(X) - u}{|X|}, \frac{\mu(Y) - w}{|Y|}, \frac{|\mu(S_i) - u_i|}{|X|}, \frac{|\mu(T_i) - t_i|}{|Y|}, \text{ for all } 1 \leq i \leq r,$$

are each at most $\alpha \epsilon^3/(100r)$. One can think of $\mu$ as representing subsets $U \subseteq X$ and $W \subseteq Y$ with $[0, 1]$-valued weights attached to its elements. One can determine via a linear program if a given sequence is feasible (see Lemma 3.4.1 below).

Here is the algorithm. We perform a complete search through all sequences $u, u_1, \ldots, u_r, w, w_1, \ldots, w_r$ of nonnegative integers at most $n$, where $u$ and each $u_i$ are divisible by $[\alpha \epsilon^3 |X|/(100r)]$, and $w$ and each $w_i$ are divisible by $[\alpha \epsilon^3 |Y|/(100r)]$. For each such sequence, we check if it is feasible, and if so then we check whether the inequalities

$$\sum_{i=1}^{r} c_i u_i w_i > (1 - \alpha/2)\epsilon w, \quad u \geq (1 - \alpha/2)\epsilon |X|, \quad \text{and } w \geq (1 - \alpha/2)\epsilon |Y| \tag{3.6}$$

93
hold. If they never hold for any feasible sequence, then we state that $G$ is $\varepsilon$-regular. On the other hand, if they hold for some feasible sequence, then we can convert $f$ into actual sets $U$ and $W$ (as we shall explain) that witness that $G$ is not $(1 - \alpha)\varepsilon$-regular.

Next we prove the correctness of the algorithm if the output is that $G$ is $\varepsilon$-regular. Consider the partition of $X$ given by the common refinement by $S_1, \ldots, S_r$. For any index set $I \subseteq [r]$, let $S_I = (\bigcap_{i \in I} S_i) \cap (\bigcap_{i \notin I} (X \setminus S_i))$ denote the part in the common refinement indexed by $I$. We can compute the sizes $|S_I|$ for all $I \subseteq [r]$ in $O(2^n)$ time. With this information at hand:

**Lemma 3.4.1.** There exists a $2^{O(r)}$ time algorithm that determines whether a given sequence $u, u_1, \ldots, u_r, w, w_1, \ldots, w_r$ is feasible.

**Proof.** It suffices to show that one can determine in the required time whether there exists $\mu: X \to [0, 1]$ such that $|\mu(X) - u| \leq a$ and $|\mu(S_i) - u_i| \leq a_i$, for each $i$. Here $a = a_i = [\alpha^3 n/(100r)]$ is the required bound (though it could be chosen arbitrarily for the purpose of this lemma). The situation for $Y$ is analogous.

For the purpose of satisfying the inequalities $|\mu(X) - u| \leq a$ and $|\mu(S_i) - u_i| \leq a_i$, one only needs to know the sum of values of $\mu$ on parts in the partition of $X$ induced by the common refinement of $S_1, \ldots, S_r$.

For each $I \subseteq [r]$, the variable $x_I$ is supposed to correspond to the value of $\mu(S_I)$. Then $\mu$ exists if and only if there exists $(x_I)_{I \subseteq [r]} \in \mathbb{R}^{2^r}$ satisfying the following inequalities:

\[-a \leq \left( \sum_{I \subseteq [r]} x_I \right) - u \leq a,\]

\[-a_i \leq \left( \sum_{I \ni i} x_I \right) - u_i \leq a_i \quad \text{for all } i \in [r],\]

and $0 \leq x_I \leq |S_I|$ for all $I \subseteq [r]$.

This is a linear program in $2^r + 1$ variables, which can be solved in $2^{O(r)}$ time. The original sequence is feasible if and only if the above system of linear inequalities has some solution in $(x_I)$. \qed
Suppose the algorithm does not find any feasible sequence satisfying (3.6). We claim that $G$ is $\epsilon$-regular. Assume otherwise. Then there exist $U \subseteq X$ and $W \subseteq Y$ such that $|U| \geq \epsilon|X|$, $|W| \geq \epsilon|Y|$, and $|d(U,W) - d(X,Y)| > \epsilon$. Since $d_{\square}(G,G') \leq \alpha\epsilon^3/4$, we have $|e_G(U,W) - e_{G'}(U,W)| \leq (\alpha\epsilon^3/4)|X||Y|$. Thus

\[
|e_{G'}(U,W) - d_G(X,Y)||U||W|| \\
\quad \geq |e_G(U,W) - d_G(X,Y)||U||W|| - |e_G(U,W) - e_{G'}(U,W)| \\
\quad \geq |d_G(U,W) - d_G(X,Y)||U||W|| - \frac{1}{4}\alpha\epsilon^3|X||Y| \\
\quad \geq \epsilon|U||W| - \frac{1}{4}\alpha\epsilon|U||W| \\
\quad \geq (1 - \frac{1}{4}\alpha)\epsilon|U||W|.
\]

On the other hand, since $G' = d_G(X,Y) + c_1K_{S_1,T_1} + \cdots + c_rK_{S_r,T_r}$, we have

\[
e_{G'}(U,W) - d_G(X,Y)||U||W| = \sum_{i=1}^r c_i|U \cap S_i||W \cap T_i|.
\]

So

\[
\left| \sum_{i=1}^r c_i|U \cap S_i||W \cap T_i| \right| \geq (1 - \frac{1}{4}\alpha)\epsilon|U||W|.
\]

Let $u$ and $u_i$ be $|U|$ and $|U \cap S_i|$, each respectively rounded to the nearest integer multiple of $[\alpha\epsilon^3|X|/(100r)]$, for all $1 \leq i \leq r$. Similarly let $w, w_i$ be $|W|$ and $|W \cap S_i|$, each respectively rounded to the nearest integer multiple of $[\alpha\epsilon^3|Y|/(100r)]$, for all $1 \leq i \leq r$. The sequence $u, u_1, \ldots, u_r, w, w_1, \ldots, w_r$ is feasible as witnessed by $\mu = 1_{U \cup W}$. We claim that (3.6) holds. Indeed, we have

\[
u \geq |U| - \frac{1}{100}\alpha\epsilon^3|X| \geq (1 - \frac{1}{100}\alpha^2)\epsilon|X|,
\]

and

\[
w \geq |W| - \frac{1}{100}\alpha\epsilon^3|Y| \geq (1 - \frac{1}{100}\alpha^2)\epsilon|Y|.
\]
Furthermore, we have
\[ \left| \sum_{i=1}^{r} c_i u_i w_i \right| \geq \left| \sum_{i=1}^{r} c_i |U \cap S_i||W \cap T_i| \right| - \frac{3}{100} \alpha \epsilon^3 |X||Y| \]
\[ \geq (1 - \frac{1}{4} \alpha) \epsilon |U||W| - \frac{3}{100} \alpha \epsilon^3 |X||Y| \]
\[ \geq (1 - \frac{1}{4} \alpha - \frac{3}{100} \alpha) \epsilon |U||W| \]
\[ \geq (1 - \frac{1}{4} \alpha - \frac{3}{100} \alpha) \epsilon (1 + \frac{1}{100} \alpha \epsilon^2)^{-2} uw \]
\[ > (1 - \frac{1}{2} \alpha) \epsilon uw. \]

The first inequality above follows from the fact that for each i,
\[ |u_i - |U \cap S_i|| \leq \frac{\alpha \epsilon^3 |X|}{100r}, \]
\[ |u_i - |U \cap S_i|| \leq \frac{\alpha \epsilon^3 |Y|}{100r}, \]
and thus
\[ |u_i w_i - |U \cap S_i||W \cap T_i| \leq \frac{3 \alpha \epsilon^3 |X||Y|}{100r}. \]

The penultimate inequality follows from \( u \leq |U| + \frac{1}{100} \alpha \epsilon^3 |X| \leq (1 + \frac{1}{100} \alpha \epsilon^2)|U| \)
and similarly with w. So we have a feasible sequence satisfying (3.6), which is a contradiction.

Suppose now that the algorithm does find some feasible sequence that satisfies (3.6). By adjusting \( \mu \), we may assume that \( \mu \) takes \{0, 1\}-value on all but at most one element in each part in the common refinement partition of \( X \) by \( S_1, \ldots, S_r \), and likewise in \( Y \) by \( T_1, \ldots, T_r \). Let \( U \subseteq X \) and \( W \subseteq Y \) denote the elements where \( \mu \) is positive, we have
\[ ||U| - u| \leq \frac{\alpha \epsilon^3}{100r} |X| + 2^r \leq \frac{\alpha \epsilon^3}{50r} |X| \]
Here the extra \( 2^r \) term account for rounding up non-integral values of \( \mu \). We used the assumption (3.5) to bound \( 2^r \). It thus follows from above, and (3.6) that
\[ |U| \geq (1 - \frac{1}{2} \alpha - \frac{1}{50r} \alpha \epsilon^2) \epsilon |X| \geq (1 - \alpha) \epsilon |X|. \]
In particular, this means that

$$||U| - u| \leq \frac{\alpha \epsilon^3}{50r} |X| \leq \frac{\alpha \epsilon^2}{(1 - \alpha)50r} |U|.$$  

Similarly, we have

$$|W| \geq (1 - \frac{1}{2}\alpha - \frac{1}{50r}\alpha^2)\epsilon|Y| \geq (1 - \alpha)\epsilon|Y|,$$

and we have

$$||U \cap S_i| - u_i| \leq \frac{\alpha \epsilon^2}{(1 - \alpha)50r} |U|, \quad \text{for all } 1 \leq i \leq r,$$

and

$$||W| - w| \leq \frac{\alpha \epsilon^2}{(1 - \alpha)50r} |W|,$$

and

$$||W \cap T_i| - w_i| \leq \frac{\alpha \epsilon^2}{(1 - \alpha)50r} |W|, \quad \text{for all } 1 \leq i \leq r.$$

and

$$|d_G(U, W) - d_G(X, Y)|$$


$$\geq \frac{1}{|U||W|} \left| \sum_{i=1}^r c_i |U \cap S_i||W \cap T_i| - \frac{|X||Y|}{|U||W|} d_{\Box}(G, G') \right|$$

$$\geq \frac{1}{|U||W|} \left( \left| \sum_{i=1}^r c_i u_i w_i \right| - \frac{3}{(1 - \alpha)50} \alpha^2 |U||W| \right) \right) - \frac{1}{4} \alpha^3 \frac{|X||Y|}{|U||W|}$$

$$\geq \frac{1}{|U||W|} \left( \left( 1 - \frac{\alpha}{2} \right) \epsilon uw - \frac{3}{(1 - \alpha)50} \alpha^2 |U||W| \right) \right) - \frac{1}{4} \alpha \epsilon$$

$$\geq (1 - \alpha)\epsilon.$$

Hence the pair $(U, W)$ witnesses that $G$ is not $(1 - \alpha)\epsilon$-regular.

We will need the following easy corollary of Theorem 3.1.2 for the next section.

**Corollary 3.4.2.** There exists an $O_{\epsilon, \alpha, k}(n^2)$ time algorithm, which, given $\epsilon, \alpha, k > 0,$
a graph $G$ on $n$ vertices, and a partition $\mathcal{P}$ of the vertex set of $G$ into $k$ parts, does one of the following:

1. Correctly states that $\mathcal{P}$ is $(1 + \alpha)\varepsilon$-regular;

2. Correctly states that $\mathcal{P}$ is not $\varepsilon$-regular.

Note that sometimes both options are correct. The algorithm that we give runs in $O(k^2 2^{2(\alpha \varepsilon) - O(1)} n^2)$ time.

**Proof.** Let $\mathcal{P}$ be the partition of $V$ into $V_1, \ldots, V_k$. Apply the algorithm in Theorem 3.1.2 to each pair $V_i, V_j$ so that it either correctly states that $(V_i, V_j)$ is $(1 + \alpha)\varepsilon$-regular or that it is not $\varepsilon$-regular. If at least a $(1 - \varepsilon)$-fraction of pairs are seen to be $(1 + \alpha)\varepsilon$-regular, then we know that $\mathcal{P}$ is $(1 + \alpha)\varepsilon$-regular, otherwise, more than an $\varepsilon$-fraction of pairs fail to be $\varepsilon$-regular, so that $\mathcal{P}$ is not $\varepsilon$-regular. \qed

### 3.5 Approximating regularity

In this section, we prove Theorem 3.1.1, reproduced below for convenience.

**Theorem.** There exists an $O_{\varepsilon, \alpha, k}(n^2)$ time algorithm, which, given $0 < \varepsilon, \alpha < 1$ and $k$, and a graph $G$ on $n$ vertices that admits an equitable $\varepsilon$-regular partition with $k$ parts, outputs an equitable $(1 + \alpha)\varepsilon$-regular partition of $G$ into $k$ parts.

Here is the algorithm, which runs in $O(2^{k/(\alpha \varepsilon)} n^2)$ time. Using Theorem 3.1.6, we find $S_1, \ldots, S_r, T_1, \ldots, T_r \subseteq V$, with $r \leq (k/\alpha \varepsilon)^{O(1)}$, such that $d_G(G, G') \leq \alpha \varepsilon/(10k^2)$, where

$$G' = d(G) + c_1 K_{S_1, T_1} + \ldots + c_k K_{S_r, T_r}.$$ 

Let $\mathcal{Q}$ denote the partition of $V(G)$ given by the common refinement of the sets $S_1, \ldots, S_r, T_1, \ldots, T_r$. Let $\mathcal{Q}$ have $s \leq 4^r$ parts, with sizes $q_1, \ldots, q_s$. We shall search over all tuples $(q_{i,j})_{1 \leq i \leq s, 1 \leq j \leq k}$ of nonnegative integers satisfying all of the following requirements:

- $q_i = q_{i,1} + \ldots + q_{i,k}$ for each $1 \leq i \leq s$;
- each $q_{i,j}$ with $j < k$ is divisible by $\lfloor \alpha \epsilon n / (25sk) \rfloor$ (no divisibility requirements for $q_{i,k}$); and

- the sums $\sum_{i=1}^{s} q_{i,j}$ for different values of $j$ differ from $n/k$ by at most $\alpha \epsilon n / (50k)$.

For each eligible tuple $(q_{i,j})$, consider a partition $\mathcal{P} : V = V_1 \cup \cdots \cup V_k$ where $Q_i \cap V_j = q_{i,j}$ (there are many such partitions; pick an arbitrary one). Apply Corollary 3.4.2 to certify that either $\mathcal{P}$ is $(1 + 3\alpha/4)\epsilon$-regular or not $(1 + \alpha/2)\epsilon$-regular. It turns out that the latter option cannot always be true for all $\mathcal{P}$ searched, as we assume that $G$ admits some $\epsilon$-regular partition with $k$ parts (we will justify this claim). From this search, we find a $(1 + 3\alpha/4)\epsilon$-regular partition $\mathcal{P}$ which is almost equitable in the sense that its parts have sizes differing from $n/k$ by at most $\alpha \epsilon n / (50k)$. We modify $\mathcal{P}$ by moving a minimum number of vertices to make it equitable. We claim that the resulting partition is $(1 + \alpha)\epsilon$-regular.

We next analyze the running time of this algorithm. Theorem 3.1.6 finds the cut norm decomposition in $O\left( \left( \frac{k}{\alpha \epsilon} \right)^{O(1)} n^2 \right)$ time. The number of tuples $(q_{i,j})$ is at most $(25sk\alpha^{-1}\epsilon^{-1})^{ks} \leq 2^{2(k/(\alpha \epsilon))^{O(1)}}$. For each $(q_{i,j})$, the algorithm in Corollary 3.4.2 takes $O(k^2 2^{2(\alpha \epsilon)} - O(1) n^2)$ time. Therefore, the entire algorithm takes $O\left( 2^{2(k/(\alpha \epsilon))^{O(1)}} n^2 \right) = O_{\alpha,\epsilon,k}(n^2)$ time.

Now we verify correctness. We shall prove the following claims, which together imply the result. Indeed, (1) shows that the algorithm always finds some $(1 + 3\alpha/4)\epsilon$-regular partition $\mathcal{P}$, and (2) shows that making $\mathcal{P}$ equitable by moving a minimum number of vertices between parts results in a $(1 + \alpha)\epsilon$-regular partition.

1. If a partition $\mathcal{P} = \{V_1, V_2, ..., V_k\}$ of $V$ is $\epsilon$-regular, then we can modify it slightly to obtain $\mathcal{P}' = \{V'_1, V'_2, ..., V'_k\}$ such that $q_{i,j} = |Q_i \cap V'_j|$ form an eligible tuple, and $\mathcal{P}'$ is $(1 + \alpha/2)\epsilon$-regular for $G$ (so the search would not pass over this $(q_{i,j})$).

2. If a partition $\mathcal{P}$ of $V$ is $(1 + 3\alpha/4)\epsilon$-regular for $G$, then by modifying $\mathcal{P}$ by adding or deleting at most $\alpha \epsilon n / (50k)$ vertices from each part, the resulting partition is $(1 + \alpha)\epsilon$-regular.

In order to show these claims, we first establish a few simple lemmas.
Lemma 3.5.1. Let $X, X', Y$ be vertex subsets of a graph with $X \subseteq X'$ and $|X| \geq (1 - \delta)|X'|$. Then $|d(X',Y) - d(X,Y)| \leq \delta$.

Proof. We have the identity
\[
d(X',Y) - d(X,Y) = \frac{e(X' \setminus X,Y) + e(X,Y)}{|X' \setminus X||Y| + |X||Y|} - \frac{e(X,Y)}{|X||Y|} = (d(X' \setminus X,Y) - d(X,Y)) \frac{|X' \setminus X|}{|X'|}.
\]
The lemma follows from noting that densities are between 0 and 1 and $|X' \setminus X| \leq \delta|X'|$.

Recall that $A \Delta B := (A \setminus B) \cup (B \setminus A)$ denotes the symmetric difference between $A$ and $B$.

Lemma 3.5.2. If $U, U', W, W'$ are vertex subsets of a graph with $|U \Delta U'| \leq \delta|U \cup U'|$ and $|W \Delta W'| \leq \delta|W \cup W'|$, then $|d(U,W) - d(U',W')| \leq 2\delta$.

Proof. It suffices to prove the lemma in the case $W = W'$ and with the bound $2\delta$ replaced by $\delta$. Indeed, the lemma would then follow by applying this case twice and the triangle inequality. By the triangle inequality and applying Lemma 3.5.1 twice with $X' = U \cup U'$, first with $\delta_1 = \frac{|U \cup U'|-|U|}{|U \cup U'|}$ and then with $\delta_2 = \frac{|U \cup U'|-|U'|}{|U \cup U'|}$, and finally using $\delta_1 + \delta_2 = \frac{|U \Delta U'|}{|U \cup U'|} \leq \delta$, we have
\[
|d(U,W) - d(U',W)| \leq |d(U,W) - d(U \cup U',W)| + |d(U \cup U',W) - d(U',W)| \\
\leq \delta_1 + \delta_2 \leq \delta.
\]

Lemma 3.5.3. Suppose $(V_1, V_2)$ is an $\epsilon$-regular pair of vertex subsets of a graph. Suppose we modify them slightly to $V_1'$ and $V_2'$, with $|V_i \Delta V_i'| \leq \delta|V_i|$ for $i = 1, 2$. Then $V_1'$ and $V_2'$ are $\epsilon + 4\delta$-regular.

Proof. Clearly we may assume that $\epsilon + 4\delta \leq 1$. Let $U' \subseteq V_1'$ and $W' \subseteq V_2'$ with $|U'| \geq (\epsilon + 4\delta)|V_1'|$ and $|W'| \geq (\epsilon + 4\delta)|V_2'|$. Let $U = U' \cap V_1$ and $W = W' \cap V_2$. Then
we have

\[ |U| = |U'| - |U' \setminus V_1| \geq |U'| - |V'_1 \setminus V_1| \geq (\epsilon + 4\delta)|V'_1| - \delta|V_1| \geq (\epsilon + 4\delta)(|V_1| - \delta|V_1|) - \delta|V_1| = \epsilon|V_1| + 4\delta|V_1| - (1 + \epsilon + 4\delta)\delta|V_1| \geq \epsilon|V_1|. \]

Similarly \( |W| \geq \epsilon|V_2| \). Thus by the regularity of the pair \( (V_1, V_2) \), we have

\[ |d(U, W) - d(V_1, V_2)| \leq \epsilon. \]

Now, we have that

\[ |U \Delta U'| \leq \delta|V_1| \leq \delta|U| \leq \delta|U \cup U'|, \]

and similarly \( |W \Delta W'| \leq \delta|W \cup W'| \), and thus \( |d(U, W) - d(U', W')| \leq 2\delta \) by Lemma 3.5.2. Similarly \( |d(V'_1, V'_2) - d(V_1, V_2)| \leq 2\delta \leq 2\epsilon \). By the triangle inequality, we have \( |d(U', W') - d(V'_i, V'_j)| \leq \epsilon + 4\delta \), showing that \( (V'_i, V'_j) \) is \((\epsilon + 4\delta)\)-regular.

As a corollary, we have the following:

**Corollary 3.5.4.** Let \( 0 < \epsilon, \delta < 1 \). Let \( G \) be a graph with \( n \) vertices. Let \( P \) be a partition of \( V(G) \) into \( k \) parts, with each part having size at least \( n/(2k) \). Suppose that \( P \) is \( \epsilon \)-regular for \( G \). If we modify \( P \) by adding or deleting at most \( \delta|V|/k \) vertices from each part of \( P \), then the resulting partition is \((\epsilon + 8\delta)\)-regular for \( G \).

**Proof.** Indeed, for any part \( V_i \) of \( P \), if we let \( V'_i \) be its modification, then \( |V_i \Delta V'_i| \leq \delta|V_i|/k \leq 2\delta|V_i| \). This means that if a pair \( (V_i, V_j) \) was \( \epsilon \)-regular, then after the modification it is \((\epsilon + 8\delta)\)-regular, and so the proportion of pairs that are not \((\epsilon + 8\delta)\)-regular is at most \( \epsilon \leq \epsilon + 8\delta \).

Now we prove claim (1) above. Let \( P \) be an equitable \( \epsilon \)-regular partition of \( G \). Since \( d_2(G, G') \leq \alpha \epsilon/(10k^2) \), \( P \) is \((1 + \alpha/10)\epsilon\)-regular for \( G' \). In \( G' \), edges between the same parts of \( Q \) have equal weights, and we can take \( P' \) such that \( |Q_i \cap V_j| \) differs from \( |Q_i \cap V'_j| \) by at most \( \alpha \epsilon n/(50sk) \) for each \( i, j \). This means that \( P' \) can be taken so that \( V_j \) and \( V'_j \) differ by at most \( \alpha \epsilon n/(50k) \) for each \( j \), so it follows from the lemma.
above that $\mathcal{P}'$ is $(1 + 3\alpha/10)\epsilon$-regular for $G'$. Therefore, $\mathcal{P}'$ must be $(1 + \alpha/2)\epsilon$-regular for $G$.

The claim (2) follows immediately from the corollary above.
Chapter 4

Arithmetic removal

4.1 Introduction

Let $p$ be a fixed prime. A triangle in $\mathbb{F}_p^n$ is a triple $(x, y, z)$ of points with $x + y + z = 0$. We use $N = |\mathbb{F}_p^n| = p^n$ throughout this chapter. The arithmetic triangle removal lemma for vector spaces over finite fields states that for each $\epsilon > 0$ and prime $p$, there is a $\delta = \delta(\epsilon, p) > 0$ such that if $X, Y, Z \subset \mathbb{F}_p^n$, and the number of triangles in $X \times Y \times Z$ is at most $\epsilon N^2$, then we can delete $\epsilon N$ points from $X$, $Y$, and $Z$ so that no triangle remains. It was originally proved by Green [40] using a Fourier analytic arithmetic regularity lemma, with a bound on $1/\delta$ which is a tower of twos of height polynomial in $1/\epsilon$. Král’, Serra, and Vena [49] showed that the arithmetic triangle removal lemma follows from the triangle removal lemma for graphs. Green’s proof shows the result further holds in any abelian group, and the Král’-Serra-Vena proof shows that an analogue holds in any group.

Green [40] posed the problem of improving the quantitative bounds on the arithmetic triangle removal lemma, and, in particular, asked whether a polynomial bound holds. Green’s problem has received considerable attention by many researchers [11, 12, 13, 14, 15, 28, 35, 39, 40, 42, 43, 62]. This is in part due to its applications and connections to several major open problems in number theory, combinatorics, and computer science. Fox [28] gave an improved bound on the triangle removal lemma for graphs. Together with the Král’-Serra-Vena reduction, it gives a bound on $1/\delta$ in
the arithmetic triangle removal lemma which is a tower of twos of height logarithmic in $1/\epsilon$, see [28] for details. An alternative, Fourier-analytic proof of this bound in the case when $p = 2$ was given by Hatami, Sachdeva, and Tulsiani [42]. Prior to this work, this tower-type bound was the best known bound for the arithmetic triangle removal lemma.

In the case $p = 2$, obtaining better upper bounds for $\delta$ has also received much attention due in part to its close connection to property testing, and, in particular, testing triangle-freeness of Boolean functions. Bhattacharyya and Xie [15] were the first to give a nontrivial upper bound, showing that we must have $\delta \leq e^{1.847}$. Fu and Kleinberg [35] provided a simple construction showing that we must have $\delta \leq e^{C_2 - o(1)}$, where $C_2 \approx 13.239$ is defined below. It is based on a construction from Coppersmith and Winograd’s famous matrix multiplication algorithm [21]. It was widely conjectured that the exponent in the bound is not optimal (and that the bound is perhaps even superpolynomial), and Haviv and Xie [43] introduced an approach to obtaining better upper bounds on the arithmetic triangle removal lemma.

In this chapter, we solve Green’s problem by proving an essentially tight bound for the arithmetic triangle removal lemma in vector spaces over finite fields. We show that a polynomial bound holds, and further determine the best possible exponent for every prime $p$. It involves, for every $p$, constants $c_p$ which will be between 0 and 1 and will be defined below, and $C_p = 1 + 1/c_p$. In particular, we have\(^1\) $c_2 = 5/3 - \log 3 \approx .0817$ and $C_2 = 1 + 1/c_2 \approx 13.239$ as in Fu and Kleinberg [35], and $c_3 = 1 - \frac{\log b}{\log 3}$, where $b = a^{-2/3} + a^{1/3} + a^{4/3}$, and $a = \sqrt[3]{3} - 1$. This gives $c_3 \approx .0775$, and $C_3 \approx 13.901$.

**Theorem 4.1.1.** Let $0 < \epsilon \leq 1/2$ and $\delta = (\epsilon/3)C_p$. If $X, Y, Z \subset F_p^n$ with less than $\delta N^2$ triangles in $X \times Y \times Z$, then we can remove $\epsilon N$ elements from $X \cup Y \cup Z$ so that no triangle remains. Furthermore, this bound is tight in that we must have $\delta \leq e^{C_p - o(1)}$.

Following a breakthrough by Croot, Lev, and Pach [22] on 3-term arithmetic progressions in $\mathbb{Z}_4^n$, Ellenberg and Gijswijt [26] showed that their method can be used

\(^1\)Here, and throughout this chapter, all logarithms are base 2.
to greatly improve the upper bound on the cap set problem, and more generally to bound the size of a set in $\mathbb{F}_p^n$ with no 3-term arithmetic progressions. Further work by Blasiak-Church-Cohn-Grochow-Naslund-Sawin-Umans [16], and independently Alon, established that their proof also shows the following multicolored sum-free theorem over $\mathbb{F}_p^n$. Later work by Kleinberg, Sawin, and Speyer [46] showed that there exists a computable exponent $c_p$ that is sharp. They conjecture that their bound is equal to the bound of Blasiak-Church-Cohn-Grochow-Naslund-Sawin-Umans, and Alon, and verify this for $p \leq 7$. Later, Norin [58], and independently Pebody [59] verified this conjecture for all $p$.

**Theorem 4.1.2.** [46] Given a collection of ordered triples $\{(x_i, y_j, z_k)\}_{i=1}^m$ in $\mathbb{F}_p^n$ such that $x_i + y_j + z_k = 0$ holds if and only if $i = j = k$, the size of the collection satisfies the bound

$$m \leq p^{(1-c_p)n}.$$ 

Furthermore, there exists such a collection with $m \geq p^{(1-c_p)-o(n)}$.

We note that the lower bound in the case of $p = 2$ was established earlier by Fu and Kleinberg [35].

The proof of Kleinberg, Sawin, and Speyer gives the value of $c_p$ as follows. For a probability distribution $\psi$ on the set $I = \{0, 1, \ldots, p-1\}$ given by probabilities $\{\psi(0), \psi(1), \ldots, \psi(p-1)\}$, define its entropy as

$$\eta(\psi) = \sum_i -\psi(i) \log \psi(i).$$

Here we define $0 \log 0$ to be 0. Now, suppose we have a probability distribution $\pi$ on $(a, b, c) \in T \subset I^3 : a + b + c = p - 1$. We then obtain marginal distributions $\pi_a, \pi_b, \pi_c$ for each coordinate. If $\pi$ is $S_3$-invariant, then these three marginal distributions are the same. Let $\gamma$ be the maximum of $\eta(\pi_a)$ over $S_3$-invariant probability distributions $\pi$ over $T$. Then we have

$$c_p = 1 - \frac{\gamma}{\log p}.$$ 

We note that while our proof of Theorem 4.1.1, determining the exponential con-
stant, relies on Theorem 4.1.2, the existence of the exponential constant follows from
the existence of a polynomial bound (which is implied by the earlier work of Ellenberg
and Gijswijt [26]), and a simple product trick, discussed in Section 4.2.

We give our proof of the lower bound in the next section. In the following section,
for completeness, we give the argument that shows the matching upper bound. We
finish with some concluding remarks.

4.2 Proof of the lower bound

Recall that \( N = p^n \). We prove the following theorem, which is roughly equivalent to
Theorem 4.1.1.

**Theorem 4.2.1.** Suppose we have \( m = \epsilon N \) disjoint triangles \( \{(x_i, y_i, z_i)\}_{i=1}^m \) in \( \mathbb{F}_p^n \).
Let \( \delta = \epsilon^{C_p} \). Then, for \( X = \{x_i\}_{i=1}^m \), \( Y = \{y_i\}_{i=1}^m \), and \( Z = \{z_i\}_{i=1}^m \), we have at least
\( \delta N^2 \) triangles in \( X \times Y \times Z \).

*Proof that Theorem 4.2.1 implies Theorem 4.1.1.* Suppose we have \( X, Y, \) and \( Z \) such
that we cannot remove \( \epsilon N \) points to remove all triangles. In this case, we can find
a set of at least \( \delta N^2 \) disjoint triangles \( (x_i, y_i, z_i) \). Indeed, start taking out disjoint
triangles greedily from \( X \times Y \times Z \). If at some point we cannot take any more and we
have taken less than \( \epsilon N \) triangles, then remove any point in any of these triangles.
We would have removed less than \( \epsilon N \) points and we would have no triangles left, a
contradiction. So we can find \( \delta N^2 \) disjoint triangles, and so we can apply Theorem
4.2.1 with \( \epsilon/3 \) to find \( \delta N^2 \) triangles, a contradiction. \( \square \)

To obtain Theorem 4.2.1, we first prove the following weaker, asymptotic version. We will later use a simple product amplification trick to remove the \( o(1) \) in the
exponent and obtain Theorem 4.2.1.

**Theorem 4.2.2.** Suppose we have \( m = \epsilon N \) disjoint triangles \( \{(x_i, y_i, z_i)\}_{i=1}^m \) in \( \mathbb{F}_p^n \).
Then, for \( X = \{x_i\}_{i=1}^m \), \( Y = \{y_i\}_{i=1}^m \), and \( Z = \{z_i\}_{i=1}^m \), we have at least \( \delta N^2 \) triangles
in \( X \times Y \times Z \), where \( \delta = \epsilon^{C_p + o(1)} \).
We first prove two lemmas. The following lemma bounds the total number of triangles based on the maximal degree of a point in the union of the sets. In particular, this proves the theorem in the case when the maximum degree is not much larger than the average degree.

**Lemma 4.2.3.** Suppose $0 < \rho \leq 1/(5p^3)$ and we have disjoint sets $X, Y, Z \subset \mathbb{F}_p^n \setminus \{0\}$, such that any two vectors from their union are linearly independent, any two-dimensional subspace contains at most one triangle in $X \times Y \times Z$, and each element in $X \cup Y \cup Z$ is in at most $\rho N$ triangles in $X \times Y \times Z$. Let $\delta N^2$ be the total number of triangles in $X \times Y \times Z$. Then $\delta \leq 125p^2\rho^{1+c_\rho}$.

**Proof.** Let $d = \lceil \log(1/(5p)) \rceil$, so $1/(5p) < \rho p^d \leq 1/5$. Note that $d \geq 3$. Take a uniformly random subspace $U$ of $\mathbb{F}_p^n$ of dimension $d$. Let $X' = X \cap U$, and define $Y'$ and $Z'$ analogously. We call a triangle $(x, y, z) \in X' \times Y' \times Z'$ good if it is the unique triangle in $X' \times Y' \times Z'$ containing $x$, $y$, or $z$.

**Claim.** Given a triangle $(x, y, z) \in X \times Y \times Z$, conditioned on $x \in X'$, $y \in Y'$, $z \in Z'$, the probability that it is good is at least $2/5$.

**Proof.** We know that $x$ is in at most $\rho N$ triangles in $X \times Y \times Z$. These are each formed with a different element of $Y$. Let $Y_1$ be the set of elements in $Y$ that together with $x$ are in a triangle in $X \times Y \times Z$. By assumption, they each generate a two-dimensional subspace with $x$, and different elements of $Y$ generate different two-dimensional subspaces. For any element in $Y_1$ not equal to $y$, the probability that it is in $U$, conditioned on $x, y, z \in U$, is

$$\frac{p^{d-2} - 1}{p^{n-2} - 1}.$$ 

Thus, using the union bound, the probability that any other element of $Y_1$ is in $U$ is at most

$$\rho N \frac{p^{d-2} - 1}{p^{n-2} - 1} = \rho p^n \frac{p^{d-2} - 1}{p^{n-2} - 1} \leq \rho p^n \frac{p^{d-2}}{p^{n-2}} = \rho p^d \leq \frac{1}{5}.$$

So, conditioned on $x \in X'$, $y \in Y'$, $z \in Z'$, the probability that either $x$, $y$, or $z$ is in
another triangle in \( X' \times Y' \times Z' \) that is not in the two-dimensional subspace generated by \( \{x, y, z\} \) is at most \( \frac{3}{5} \).

\[ \square \]

**Claim.** Let \( T \) be the number of good triangles. Then \( \mathbb{E}[T] \geq \frac{\delta}{125p^2\rho^2} \).

**Proof.** For each triangle in \( X \times Y \times Z \), it has a probability of \( \frac{(p^d-1)(p^d-1)}{(p^n-1)(p^n-1)} \) of being in \( X' \times Y' \times Z' \), and conditioned on this, it has a probability of at least \( \frac{2}{5} \) of being good. Since there are \( \delta N^2 \) triangles in total,

\[
\mathbb{E}[T] \geq \delta N^2 \frac{2(p^d-1)(p^d-1)}{5(p^n-1)(p^n-1)} \geq \frac{1}{5} \delta p^2 n p^{2d} \geq \frac{\delta}{125p^2\rho^2}.
\]

\[ \square \]

Observe that the set of good triangles satisfies the hypothesis of Theorem 4.1.2. Therefore, \( T \) is always at most \( p^{(1-c_p)d} \).

We thus have

\[
p^{(1-c_p)d} \geq \mathbb{E}[T] \geq \frac{\delta}{125p^2\rho^2},
\]

which, combined with the fact that \( p^d \leq \frac{1}{5p} \leq \frac{1}{\rho} \), implies that

\[
\delta \leq 125p^2\rho^2(p^d)^{1-c_p} \leq 125p^2\rho^{1+c_p}.
\]

This completes the proof of Lemma 4.2.3. \( \square \)

Let \( 0 < a_p \leq 5/p^4 \) and \( g : (0, a_p] \rightarrow \mathbb{R}^+ \) be any function such that \( g(\beta) \) increases as \( \beta \) decreases, but \( g(\beta)\beta \) and \( \beta^{a_p}g(\beta)^{1+c_p} \) both decrease as \( \beta \) decreases, and \( \sum_{i=1}^{\infty} \frac{1}{g(2^{-i}a_p)} \leq 1/4 \). We will later take for convenience \( g(\beta) = \log^2(1/\beta) \), which satisfies the conditions for some \( a_p > 0 \).

**Lemma 4.2.4.** Let \( 0 < \delta \leq a_p \) and \( \epsilon \geq (125p^2)^{1/1+c_p} g(\delta) \). Suppose \( m = \epsilon N \) and we have a collection of pairwise disjoint triangles \( (x_i, y_i, z_i) \) for \( i \in [m] \), and let \( X = \{x_1, x_2, \ldots, x_m\} \), \( Y = \{y_1, y_2, \ldots, y_m\} \), and \( Z = \{z_1, z_2, \ldots, z_m\} \). Suppose that any two vectors from \( X \cup Y \cup Z \) are linearly independent, and any two-dimensional
Subspace contains at most one triangle. Then there must be at least \( \delta N^2 \) triangles in \( X \times Y \times Z \).

Proof. Suppose for contradiction that we have less than \( \delta N^2 \) triangles. We would like to obtain large subsets \( X_1 \subset X, Y_1 \subset Y, Z_1 \subset Z \) such that no element of \( X_1, Y_1, \) or \( Z_1 \) is contained in significantly more triangles in \( X_1 \times Y_1 \times Z_1 \) than the average. This will allow us to apply Lemma 4.2.3 to obtain a contradiction.

We do this by removing bad points from the sets one at a time, based on the density of triangles. First, we remove points that are in too many triangles. Suppose that after removing a certain number of points, we are left with \( \delta' N^2 \) triangles. If there exists a point that is in at least \( g(\delta') \frac{\delta}{\epsilon} N \) triangles, we remove it. We then update \( \delta' \) and repeat. We claim that when this process ends, we only removed at most \( m/2 \) points. Indeed, for any \( \beta \), while the number of triangles is between \( 2\beta N^2 \) and \( \beta N^2 \), we remove at least \( g(\beta) \frac{\beta}{\epsilon} N \) triangles in each step, but since throughout this period the number of triangles is at most \( 2\beta N^2 \), this happens at most \( 2\beta N^2 / (g(\beta) \frac{\beta}{\epsilon} N) = \frac{2\epsilon}{g(\beta)} \) times. Overall, this implies that we end up removing at most

\[
\frac{2\epsilon}{g(\delta/2)} N + \frac{2\epsilon}{g(\delta/4)} N + \cdots \leq \frac{\epsilon}{2} N
\]

points. So after this process, we are left with sets \( X', Y', Z' \) which contain at least \( \frac{\delta}{2} N \) disjoint triangles between them, and for some \( \delta' \), we have \( \delta' N^2 \) triangles in \( X' \times Y' \times Z' \), and every point in each set is in at most \( g(\delta') \frac{\delta}{\epsilon} N \) such triangles.

We can now apply Lemma 4.2.3 with

\[
\rho = g(\delta') \frac{\delta'}{\epsilon} \leq g(\delta) \frac{\delta}{\epsilon} \leq \left( \frac{\delta}{125p^2} \right)^{1+\epsilon} p \leq \frac{1}{5p^3}.
\]

Here the first inequality follows from the fact that \( g(\beta) \beta \) decreases as \( \beta \) decreases, the second from the bound on \( \epsilon \), and the third from the bound \( \delta \leq a_p \leq 5/p^4 \) and the fact that \( 0 < c_p < 1 \). The lemma implies that

\[
\delta' \leq 125p^2 \left( g(\delta') \frac{\delta'}{\epsilon} \right)^{1+c_p}
\]
which means that

$$\epsilon \leq (125p^2)^{\frac{1}{1+cp}} \delta'/\delta g(\delta') \leq (125p^2)^{\frac{1}{1+cp}} \delta g(\delta),$$

where we used for the second inequality that if $\delta' < a_p$, then $\delta g(\delta')^{1+cp}$ increases as $\delta'$ increases.

\[\Box\]

**Proof of Theorem 4.2.2.** We are now ready to prove Theorem 4.2.2. First, we would like to argue that we may assume that the sets $X$, $Y$, and $Z$ are pairwise disjoint, any two vectors are linearly independent, and any two-dimensional subspace contains at most one triangle in $X \times Y \times Z$. Let us work in $\mathbb{F}_p^{n+2}$, so we add two coordinates to each vector. For each $x_i$, we define $x'_i = (x_i, 1, 0)$. We also take $y'_i = (y_i, -1, 1)$, and $z'_i = (z_i, 0, -1)$. Then, for each $i$, we still have $x'_i + y'_i + z'_i = 0$, and if we take $X' = \{x'_i\}_{i=1}^m$, $Y' = \{y'_i\}_{i=1}^m$, and $Z' = \{z'_i\}_{i=1}^m$, then the triangles in $X' \times Y' \times Z'$ correspond exactly to the triangles in $X \times Y \times Z$. It is easy to see that the three new sets are disjoint, simply because of their last two coordinates. The last two coordinates also imply that if we multiply a point from one of the sets by a scalar not equal to 1, we cannot be in the union of the sets. Thus, any two vectors from $X \cup Y \cup Z$ are linearly independent. Finally, fix a triangle $(x_i, y_j, z_k)$, and let $U$ be the two-dimensional subspace generated by them. Suppose that for some $j'$ with $j \neq j'$, the subspace generated by $x_i$ and $y_j$ contains $y_{j'}$. This means that

$$y_{j'} = \alpha x_i + \beta y_j.$$

Just by looking at the last two coordinates, we clearly must have $\alpha = 0$ and $\beta = 1$, but then $y_j = y_{j'}$, a contradiction. This means that any triangle contained in $U$ must contain $y_j$. It is easy to analogously show that any triangle in $U$ must contain $x_i$. This implies that the only triangle contained in $U$ is $(x_i, y_j, z_k)$. Therefore, any two-dimensional subspace contains at most one triangle.

Since the size of the underlying space is now $N' = p^2 N$, we obtain $\epsilon' N'$ triangles with $\epsilon' = \epsilon/p^2$. Using Lemma 4.2.4, if we take $\delta' = \delta/p^4$, we have that there must be
at least $\delta'N^2 = \delta N^2$ triangles in total, provided that we have

$$\frac{\epsilon}{p^2} > (125p^2)^{\frac{1}{1+\epsilon_p}} (\delta/p^4)^{\frac{\epsilon_p}{1+\epsilon_p}} g(\delta/p^4).$$

With our choice of $g(\rho) = (\log(1/\rho))^2$, if we take $\delta = \Omega_p\left(\epsilon^{1+1/\epsilon_p}(\log(1/\epsilon))^{\frac{2+2\epsilon_p}{\epsilon_p}}\right)$, the conditions of the lemma are satisfied if $\epsilon$ (and thus $\delta$) are small enough. This gives $\delta \geq \epsilon^{C_p+o(1)}$.

We now tighten the asymptotic bound in Theorem 4.2.2 to prove Theorem 4.2.1 with a product argument, inspired by a similar one Kleinberg, Sawin, and Speyer [46] used for the multicolor sum-free problem.

**Proof of Theorem 4.2.1.** Suppose we have $m = \epsilon N$ disjoint triangles $\{(x_i, y_i, z_i)\}_{i=1}^m$ in $(\mathbb{F}_p^n)^3$. Define $X = \{x_i\}_{i=1}^m$, $Y = \{y_i\}_{i=1}^m$, $Z = \{z_i\}_{i=1}^m$, and suppose for the sake of contradiction there are only $\delta N^2$ triangles in $X \times Y \times Z$ with $\delta < \epsilon^{C_p}$, so we can write $\delta = \epsilon^{(1+\alpha)C_p}$ for some $\alpha > 0$. We define for every positive integer $k$ a collection of $m^k$ disjoint triangles in $(\mathbb{F}_p^n)^3$. For $k$ vectors $x_1, x_2, \ldots, x_k$ in $\mathbb{F}_p^n$, we can take their concatenation $(x_1, x_2, \ldots, x_k) \in \mathbb{F}_p^{nk}$. Then, we can take for any $k$-tuple $i_1, i_2, \ldots, i_k$ the triangle $((x_{i_1}, x_{i_2}, \ldots, x_{i_k}), (y_{i_1}, y_{i_2}, \ldots, y_{i_k}), (z_{i_1}, z_{i_2}, \ldots, z_{i_k}))$. It is easy to see that these are indeed triangles and disjoint, and if we define $X_k$, $Y_k$, and $Z_k$ as subsets of $\mathbb{F}_p^{nk}$ analogously to $X$, $Y$, and $Z$, then $X_k = X^k$, $Y_k = Y^k$, and $Z_k = Z^k$. Then any triangle must be a triangle in the first $n$ coordinates, in the second $n$ coordinates, and so on. Thus, the number of triangles in $X^k \times Y^k \times Z^k$ is $\delta^k N^{2k} = \epsilon^{(1+\alpha)C_p k} N^{2k}$. However, by Theorem 4.2.2 applied to $X^k \times Y^k \times Z^k$, we must have

$$(\epsilon^k)^{C_p+o(1)} \leq \delta^k = (\epsilon^k)^{(1+\alpha)C_p},$$

where the $o(1)$ term tends to 0 as $\epsilon^k$ tends to 0. Letting $k \to \infty$, we obtain a contradiction. We conclude that $\delta \geq \epsilon^{C_p}$, completing the proof of Theorem 4.2.1. \qed

111
4.3 Proof of the upper bound

For completeness, we now give a proof of the last part of Theorem 4.1.1, which shows that the lower bound is tight.

For \( p = 2 \), using their lower bound construction discussed in Theorem 4.1.2, Fu and Kleinberg [35] provided a simple construction that implies an upper bound on \( \delta \) in terms of \( \epsilon \), which shows that the exponent \( C_2 \) in Theorem 4.1.1 is tight. That a lower bound for the multicolor sum-free problem can be turned into an upper bound for the arithmetic removal lemma was first observed by Bhattacharyya and Xie [15]; see also [43]. The argument works in general for \( \mathbb{F}_p \), so by the result of Kleinberg, Sawin, and Speyer [46], the exponent \( C_p \) is tight for each \( p \).

We now present a construction based on that of Bhattacharyya and Xie [15]. Recall that the multicolored sum-free problem asks to find a collection of ordered triples \( \{(x_i, y_i, z_i)\}_{i=1}^m \) in \( \mathbb{F}_p^n \) such that \( x_i + y_j + z_k = 0 \) holds if and only if \( i = j = k \).

We show that if there is such a collection of \( m = p^{(1-c_p)n-o(n)} \) triples, then there are examples that show that in the removal lemma over \( \mathbb{F}_p \), we must take

\[
\delta \leq \epsilon^{C_p-o(1)}.
\]

Given such a set of \( m \) triples \( (x_i, y_i, z_i) \), with \( X = \{x_i\}_{i=1}^m \), \( Y = \{y_i\}_{i=1}^m \), \( Z = \{z_i\}_{i=1}^m \), define, for any positive integer \( l \), subsets \( X', Y', Z' \subset \mathbb{F}_{p^{n+l}} \) by taking \( X' = X \times \mathbb{F}_p^l \), \( Y' = Y \times \mathbb{F}_p^l \), and \( Z' = Z \times \mathbb{F}_p^l \).

First, we claim that if we want to delete elements from \( X', Y', \) and \( Z' \) such that no triangles remain, we must delete at least \( mp^l \) elements in total. Fix \( i \in [m] \), and look at \( x_i \times \mathbb{F}_p^l, y_i \times \mathbb{F}_p^l, z_i \times \mathbb{F}_p^l \). These are all contained in \( X', Y', \) and \( Z' \) respectively, so suppose that we have removed less than \( p^l \) of these \( 3 \cdot p^l \) elements. Define \( X_i, Y_i, \) and \( Z_i \) so that we have kept \( x_i \times X_i, y_i \times Y_i, \) and \( z_i \times Z_i, \) and note that they must each be nonempty. Since \( x_i + y_i + z_i = 0 \), we must have that \( X_i + Y_i \) is disjoint from \( Z_i \). However, if we take any \( z \in Z_i \), the set \( z - Y_i \) must intersect the set \( X_i \), since \( |X_i| + |Y_i| > p^l \). This means that we must have at least one triangle, a contradiction.

Furthermore, we claim that the total number of triangles in \( X' \times Y' \times Z' \) is
Indeed, if $(x_i, x') + (y_j, y') + (z_k, z') = 0$, then in particular, we must have $x_i + y_j + z_k = 0$, so $i = j = k$, and in this case, we can choose anything in $\mathbb{F}_p^l$ for $x'$ and for $y'$, and then the choice of $z'$ is determined.

This means that taking

$$
\epsilon = \frac{mp^l}{p^{n+l}} = \frac{m}{p^n}, \quad \delta = \frac{mp^{2l}}{p^{2n+2l}} = \frac{m}{p^{2n}},
$$

we obtain a construction of an infinite family of sets $X_i, Y_i, Z_i \subset \mathbb{F}_p^n$ such that there are at most $\delta p^{2n+2l}$ triangles between the sets, but we have to remove at least $\epsilon p^{n+l}$ points to remove all triangles.

Since

$$m = p^{(1-c_p)n-o(n)},$$

we have

$$\epsilon = p^{-n(c_p+o(1))}, \quad \delta = p^{-n((1+c_p)+o(1))}.$$  

Thus, as $n \to \infty$, we have that $\epsilon \to 0$, and (recalling that $C_p = 1 + 1/c_p$), we have

$$\delta \leq e^{C_p-o(1)}.$$

4.4 Concluding remarks

We expect that these methods will also apply to similarly give tight polynomial bounds for removal lemmas of linear equations in more variables in vector spaces over $\mathbb{F}_p$. Sets that contain no arithmetic progressions have also been studied over other abelian groups, see [16] and [60]. We expect these methods can be used in those settings as well to improve bounds on the corresponding removal lemma. We also expect further applications in property testing.

The triangle removal lemma states that every graph on $n$ vertices with $o(n^3)$ triangles can be made triangle-free by removing $o(n^2)$ edges. One consequence of this, known as the diamond-free lemma, is that every graph on $n$ vertices in which each edge is in precisely one triangle has $o(n^2)$ edges. The diamond-free lemma is also
equivalent to the Ruzsa-Szemerédi induced matching theorem, or the extremal $(6, 3)$-theorem. The argument presented here for getting Green’s arithmetic removal lemma from the tri-colored sum-free result can be adapted, by replacing random subspaces by random subsets, to show that the diamond-free lemma implies the triangle removal lemma if the bound on the diamond-free lemma is sufficiently good, and any such bound on the diamond-free lemma can be turned into a similar bound on the triangle removal lemma.

**Acknowledgement:** We would like to thank Lisa Sauermann for several helpful comments, including two simplifications of our proof. In particular, she pointed out that a lower bound on the degrees in Lemma 4.2.3 is not necessary. We would also like to thank Noga Alon, Thomas Church, Robert Kleinberg, Will Sawin, Benny Sudakov, Madhu Sudan, and Terence Tao for helpful comments.
Chapter 5

Permutation regularity lemma

In this chapter, we turn our attention to a regularity lemma for permutations. Cooper [20] proved a permutation regularity lemma which was later refined by Hoppen, Kohayakawa, and Sampaio [45]. Here we give a new short proof of the permutation regularity lemma, improving the number of parts from tower-type to single exponential, and further extend it to an interval regularity lemma for graphs and matrices.

To define regular partitions for permutations, it is natural to state it as a special case in a more general setting for matrices. Let $Y = (y_{ij})$ be a $n \times n$ matrix. We use interval to mean a subset of $[n]$ of consecutive integers. For any intervals $I, J$ of $[n]$, we write

$$d_Y(I, J) := \frac{1}{|I||J|} \sum_{i \in I, j \in J} y_{ij}.$$

**Definition 5.0.1.** Let $Y$ be a $n \times n$ square matrix. Let $I, J \subseteq [n]$ be intervals. We say that $(I, J)$ is interval $\epsilon$-regular for $Y$ if for all subintervals $A \subseteq I$ and $B \subseteq J$ with $|A| \geq \epsilon |I|$ and $|B| \geq \epsilon |J|$ one has

$$|d_Y(A, B) - d_Y(I, J)| \leq \epsilon.$$

Let $\mathcal{P}$ be a partition of $[n]$ into $k$ intervals. We say that $\mathcal{P}$ is interval $\epsilon$-regular for $Y$ if all except at most $\epsilon k^2$ pairs of intervals $(I, J)$ of $\mathcal{P}$ are interval $\epsilon$-regular for $Y$.

**Definition 5.0.2.** We say that $\mathcal{P}$ is an equipartition of $[n]$ if every pair of parts in $\mathcal{P}$
differ in size by at most one.

Here is the regularity lemma for interval regular partitions.

**Theorem 5.0.3** (Interval regularity lemma). *For every $\epsilon > 0$ and positive integer $m$ there is some $M = m^{O(1)}\epsilon^{-O(\epsilon^{-5})}$ with the following property. For every $n \in \mathbb{N}$, and $n \times n$ matrix $Y = (y_{ij})$ with $[0,1]$-valued entries, there is some integer $k \in [m, M]$ so that every equipartition of $[n]$ into $k$ intervals is interval $\epsilon$-regular for $Y$.*

Remark. If $n \leq M$, then we can take the partition of $[n]$ into singletons. Otherwise, our proof will show that one can pick $k$ from a small set of choices: one can take $k = mq^i$, where $q = [16\epsilon^{-3}]$ and $0 \leq i < [4\epsilon^{-5}]$ is some integer.

Theorem 5.0.3 has the following immediate consequence on permutation regularity. Given a permutation $\sigma : [n] \rightarrow [n]$, associate to it the $n \times n$ matrix $Y^\sigma$ defined by

\[
y_{ij} = \begin{cases} 
1 & \text{if } \sigma(i) < j \\
0 & \text{otherwise.}
\end{cases}
\]

A partition of $[n]$ into intervals is said to be $\epsilon$-regular for $\sigma$ if it is interval $\epsilon$-regular for the associated matrix $Y^\sigma$.

**Theorem 5.0.4** (Permutation regularity lemma). *For every $\epsilon > 0$ and positive integer $m$, there exist $M = m^{O(1)}\epsilon^{-O(\epsilon^{-5})}$ with the following property. Let $n \neq n_0$ and $\sigma$ be a permutation of $[n]$. Then for some integer $k \in [m, M]$, every equitable partition of $[n]$ into $k$ intervals is $\epsilon$-regular for $\sigma$.*

An early form of this permutation regularity lemma was first proved by Cooper in [20]. The above form was proved in [45] with $M$ being a tower exponential of height $O(\epsilon^{-5})$. Our version requires a much smaller $M$.

### 5.1 Interval regular partitions for functions

We first prove the interval regularity lemma for functions. It is somewhat cleaner to work with partitions of the real interval $[0,1]$ into exactly equal-length subintervals,
instead of equitable partitions of \([n]\). The measure theoretic approach has the slight advantage that it allows us to defer divisibility issues of \(n\) until the end.

Let \(f: [0, 1]^2 \to [0, 1]\) be a measurable function. For any intervals \(I, J \subseteq [0, 1]\) we write
\[
d_f(I, J) := \frac{1}{\lambda(I)\lambda(J)} \int_{I \times J} f(x, y) \, dx \, dy.
\]
Here \(\lambda\) denotes the Lebesgue measure.

**Definition 5.1.1.** Let \(f: [0, 1]^2 \to [0, 1]\) be a measurable function. Let \(I, J \subseteq [0, 1]\) be intervals. We say that \((I, J)\) is interval \(\epsilon\)-regular for \(f\) if for all subintervals \(A \subseteq I\) and \(B \subseteq J\) with \(\lambda(A) \geq \epsilon\lambda(I)\) and \(\lambda(B) \geq \epsilon\lambda(J)\) one has
\[
|d_f(A, B) - d_f(I, J)| \leq \epsilon.
\]

Let \(P\) a partition of \([0, 1]\) into \(k\) intervals. We say that \(P\) is interval \(\epsilon\)-regular for \(f\) if all except at most \(\epsilon k^2\) pairs of intervals \((I, J)\) of \(P\) are interval \(\epsilon\)-regular for \(f\).

**Theorem 5.1.2.** For every \(\epsilon > 0\) and positive integer \(m\) there is some \(M = me^{-O(\epsilon^{-5})}\) with the following property. For every measurable function \(f: [0, 1]^2 \to [0, 1]\), there is some integer \(k \in [m, M]\) such that the partition of \([0, 1]\) into \(k\) equal-length intervals \([0, 1/k) \cup [1/k, 2/k) \cup \cdots \cup [(k - 1)/k, 1]\) is interval \(\epsilon\)-regular for \(f\).

**Remark.** In Theorem 5.1.2, it is possible to take \(k = mq^i\), where \(q = \lceil 16\epsilon^{-3}\rceil\) and \(0 \leq i < \lfloor 4\epsilon^{-5}\rfloor\) is some integer.

Before proving Theorem 5.1.2, we first prove a lemma showing that the density \(d_f(A, B)\) does not change very much if \(A\) and \(B\) are changed only slightly.

**Lemma 5.1.3.** Let \(f: [0, 1]^2 \to [0, 1]\) be a measurable function. For any four intervals \(A, A', B, B' \subseteq [0, 1]\) we have
\[
|d_f(A, B) - d_f(A', B')| \leq \frac{2\lambda((A \times B) \Delta (A' \times B'))}{\lambda(A)\lambda(B)}.
\]
Proof. By the triangle inequality,

\[
\lambda(A)\lambda(B)|d_f(A, B) - d_f(A', B')| \\
\leq |\lambda(A)\lambda(B)d_f(A, B) - \lambda(A')\lambda(B')d_f(A', B')| + d_f(A', B')|\lambda(A)\lambda(B) - \lambda(A')\lambda(B')| \\
\leq \left| \int_{A \times B} f \, d\lambda - \int_{A' \times B'} f \, d\lambda \right| + |\lambda(A)\lambda(B) - \lambda(A')\lambda(B')| \\
\leq 2\lambda((A \times B)\Delta(A' \times B')).
\]

The bound in Lemma 5.1.3 can be improved by a factor 2 by following the proof of Lemma 3.5.2.

Proof of Theorem 5.1.2. Let \( f_k \) denote the function obtained from \( f \) by replacing its value inside each box \([i/k, (i+1)/k) \times [j/k, (j+1)/k)\) by its average inside that box, i.e.,

\[
f_k(x, y) := k^2 \int_{[i/k, i+1/k) \times [j/k, j+1/k)} f \, d\lambda \quad \text{if } (x, y) \in \left[ \frac{i}{k}, \frac{i+1}{k} \right) \times \left[ \frac{j}{k}, \frac{j+1}{k} \right)
\]

for \( i, j = 0, 1, \ldots, k - 1 \) (when \( i \) or \( j \) equals \( k - 1 \), the corresponding interval should be modified to be closed on the right). Write

\[
\|f\|_2 := \left( \int_{[0,1]^2} |f|^2 \, d\lambda \right)^{1/2}
\]

for the \( L^2 \) norm.

Let \( q = \lceil 16\epsilon^{-3} \rceil \). Consider the sequence \( f_m, f_{mq}, f_{mq^2}, \ldots \). Since \( 0 \leq \|f_k\|_2 \leq 1 \) for all \( k \), there exists some \( k = mq^i \) for \( 0 \leq i < \lceil 4\epsilon^{-5} \rceil \) such that

\[
\|f_{kq}\|_2^2 \leq \|f_k\|_2^2 + \frac{\epsilon^5}{4}. \tag{5.1}
\]

We will show that the partition of \([0,1]\) into \( k \) equal-length intervals is interval \( \epsilon \)-regular. Indeed, if this were not the case, then there exists more than \( \epsilon k^2 \) irregular
pairs of intervals \((I, J)\), where \(I = [i/k, (i + 1)/k)\) and \(J = [j/k, (j + 1)/k)\) for some integers \(i\) and \(j\). Due to the irregularity, there exist subintervals \(A \subseteq I\) and \(B \subseteq J\) such that \(\lambda(A) \geq \epsilon \lambda(I), \lambda(B) \geq \epsilon \lambda(J)\), and

\[
|d_f(I, J) - d_f(A, B)| > \epsilon. \tag{5.2}
\]

Let \(A'\) be the smallest interval containing \(A\) with both ends being multiples of \(1/(kq)\). Note that \(A' \subseteq I\). Similarly define \(B'\). We see that \(A' \times B'\) contains \(A \times B\), and the difference in area is at most \(4/(kq)\). By Lemma 5.1.3,

\[
|d_f(A, B) - d_f(A', B')| \leq \frac{2(4/(kq))}{(\epsilon/k)^2} = \frac{8}{q\epsilon^2} = \frac{8}{16\epsilon^{-3}\epsilon^2} \leq \frac{\epsilon}{2}.
\]

By (5.2) we have

\[
|d_f(I, J) - d_f(A', B')| > \frac{\epsilon}{2}.
\]

Since the endpoints of \(I\) and \(J\) are multiples of \(1/k\) and those of \(A'\) and \(B'\) are multiples of \(1/(kq)\), the function \(f_k - f_{kq}\) has average value \(d_f(I, J) - d_f(A', B')\) over the box \(A' \times B'\). So the contribution to \(\|f_k - f_{kq}\|_2^2\) from \(A' \times B'\) is at least \(\lambda(A')\lambda(B')(\epsilon/2)^2 \geq \epsilon^4/(4k^2)\). As there are more than \(\epsilon k^2\) irregular pairs \((I, J)\), and all the rectangles \(I \times J\) are disjoint, we have

\[
\|f_k - f_{kq}\|_2 > \frac{\epsilon^5}{4}.
\]

Note that

\[
\int_{[0,1]^2} (f_k - f_{kq})f_k d\lambda = 0
\]

since \(f_k\) is constant over each box \([i/k, (i+1)/k) \times [j/k, (j+1)/k)\), and \(f_{kq}\) averages to \(f_k\) on this box. Thus \(f_k\) and \(f_k - f_{kq}\) are orthogonal, so by the Pythagorean theorem,

\[
\|f_{kq}\|_2^2 = \|f_k - (f_k - f_{kq})\|_2^2 = \|f_k\|_2^2 + \|f_k - f_{kq}\|_2^2 > \|f_k\|_2^2 + \frac{\epsilon^5}{4},
\]

which contradicts (5.1). It follows that the partition of \([0,1]\) into \(k\) equal-length
5.2 Dealing with equitable partitions

Here is a lemma that will be useful for the proof of Theorem 5.0.3. It says that $(I, J)$ being interval regular is robust under changing $I$ and $J$ by a small amount.

**Lemma 5.2.1.** Let $f : [0, 1]^2 \to [0, 1]$ be a measurable function. Let $I, I', J, J' \subseteq [0, 1]$. Let $0 < \epsilon \leq 1$. Let $\epsilon' > 0$ be a quantity less than each of

\[
\epsilon - \frac{4\lambda((I \times J) \Delta (I' \times J'))}{\epsilon^2 \lambda(I) \lambda(J)}, \quad \frac{\lambda(I) \epsilon - \lambda(I \setminus I')}{\lambda(I')}, \quad \frac{\lambda(J) \epsilon - \lambda(J \setminus J')}{\lambda(J')}.
\]

If $(I', J')$ is interval $\epsilon'$-regular for $f$, then $(I, J)$ is interval $\epsilon$-regular for $f$.

**Proof.** Let $A \subseteq I$ and $B \subseteq J$ be subintervals such that $\lambda(A) \geq \epsilon \lambda(I)$ and $\lambda(B) \geq \epsilon \lambda(J)$. Let $A' = A \cap I'$ and $B' = B \cap J'$. The second and third hypotheses about $\epsilon'$ above imply that $\lambda(A') \geq \epsilon' \lambda(I')$ and $\lambda(B') \geq \epsilon' \lambda(J')$. Since $(I', J')$ is $\epsilon'$-regular for $f$, we have

\[|d_f(A', B') - d_f(I', J')| \leq \epsilon'.\]

By Lemma 5.1.3, we have

\[|d_f(I, J) - d_f(I', J')| \leq \frac{2\lambda((I \times J) \Delta (I' \times J'))}{\lambda(I) \lambda(J)}\]

and

\[|d_f(A, B) - d_f(A', B')| \leq \frac{2\lambda((A \times B) \Delta (A' \times B'))}{\lambda(A) \lambda(B)} \leq \frac{2\lambda((I \times J) \Delta (I' \times J'))}{\epsilon^2 \lambda(I) \lambda(J)}.
\]

It follows by the triangle inequality and the first hypotheses on $\epsilon'$ that

\[|d_f(A, B) - d_f(I, J)| \leq \epsilon,
\]

which proves that $(I, J)$ is $\epsilon$ regular for $f$. \qed
Proof of Theorem 5.0.3. Let \( f : [0, 1]^2 \to [0, 1] \) be the function that takes constant value \( y_{ij} \) on the rectangle \( ((i - 1)/n, i/n) \times ((j - 1)/n, j/n) \), for each \( 1 \leq i, j \leq n \). By Theorem 5.1.2, there is some \( k \in [m, m\varepsilon^{-O(\varepsilon^{-5})}] \) so that the partition of \([0, 1]\) into \( k \) equal-length intervals \( \epsilon/2 \)-regular.

Any equitable partition \( \mathcal{P} \) of \([n]\) into sets of sizes \( c_1, \ldots, c_k \), gives rise to a partition \( \mathcal{Q} \) of \([0, 1]\) into intervals of length \( c_1/n, \ldots, c_k/n \). Since \( \mathcal{P} \) is an equitable partition, the \( i \)-th interval \( I_i \) of \( \mathcal{Q} \) differs, in terms of symmetric difference, from \( ((i - 1)/k, i/k) \) by at most \( k/n \) in measure. It follows from Lemma 5.2.1 that if \( n \) is large enough, say, \( n \geq 100k^3\varepsilon^{-3} \), then \( (I_i, J_i) \) is interval \( \epsilon \)-regular for \( f \) whenever \( ((i - 1)/k, i/k) \times ((j - 1)/k, j/k) \) is interval \( \epsilon/2 \)-regular for \( f \). It follows that \( \mathcal{Q} \) is interval \( \epsilon \)-regular for \( f \).

When \( n < 100k^3\varepsilon^{-3} \), we can take the partition of \([n]\) into singletons, which is trivially interval \( \epsilon \)-regular. \( \square \)
Bibliography


<table>
<thead>
<tr>
<th>Reference</th>
<th>Citation</th>
</tr>
</thead>
</table>


