Master in Applied Mathematics

Degree Thesis

The Regularity Lemma in Additive Combinatorics

Author: Lluís Vena Cros

Advisor: Oriol Serra Albó

Department: Applied Mathematics IV

Facultat de Matemàtiques i Estadística

Universitat Politècnica de Catalunya

Academic year: 2007-2008 1Q

Abstract. The Szemerédi Regularity Lemma (SzRL) was introduced by Endré Szemerédi in his celebrated proof of the density version of Van der Waerden Theorem, namely, that a set of integers with positive density contains arbitrarily long arithmetic progressions. The SzRL has found applications in many areas of Mathematics, including of course Graph Theory and Combinatorics, but also in Number Theory, Analysis, Ergodic Theory and Computer Science.

One of the consequences of the SzRL are the so–called 'Counting Lemma' and 'Removing Lemma', which roughly says that a sufficiently large graph G which contains not many copies of a fixed graph H can be made H–free by removing a small number of edges.

Recently Ben Green gave an algebraic version of both, the SzRL and the Removal Lemma for groups. In this algebraic version the structural result fits into the algebraic structure in terms of subgroups. On the other hand, the Removal Lemma has its algebraic counterpart in the estimation of the number of solutions of equations in groups.

The purpose of this Master Thesis is to give a detailed account on the SzRL and some of its applications, particularly to Additive Combinatorics. We particularly focuss on the consequences of the SzRL related to the Counting Lemma. By combining the version by Alon and Shapira of the directed version of the SzRL with the version of Simonovits for edge–colored graphs, we state and prove a Counting Lemma for arc–colored directed graphs.

The methods used by Green heavily rely on Fourier Analysis, and as such, his results are applicable only to Abelian groups. By using our general version of the Counting Lemma we prove a generalization of Ben Green's Removal Lemma which is applicable to finite groups, non necessarily abelian.

Contents

Introduction

The Szemerédi Regularity Lemma (SzRL) was introduced by Endré Szemerédi [**22**] in his celebrated proof of the density version of Van der Waerden Theorem, namely, that a set of integers with positive density contains arbitrarily long arithmetic progressions.

The SzRL is a general structural result which states that every sufficiently large graph admits a partition into equal parts such that most of the pairs of parts behave 'regularly', essentially like a random graph. The SzRL has found applications in many areas of Mathematics, including of course Graph Theory and Combinatorics, but also in Number Theory, Analysis, Ergodic Theory and Computer Science. The surveys of Komlós and Simonovits [**15**] and Komlós, Simonovits and Szmereédi [**14**] give a good account on the SzRL and its applications.

One of the consequences of the SzRL is the so–called 'Removing Lemma', which can be traced back to the paper of Ruzsa and Szemerédi [**21**] where they give a simple proof of Roth's Theorem, the case of progressions of length 3 in Szemerédi's Theorem. The Removal Lemma roughly says that a sufficiently large graph G which contains not many copies of a fixed graph H can be made H–free by removing a small number of edges. The known proofs of the Removal Lemma rely on the SzRL and thus suffer from the drawback of the SzRL: being so general, the constants involved in the statement are necessarily large. The problem of finding an independent proof of the Removal Lemma, which may take advantage of regularity properties of the host graph G with the benefit of depending on more reasonable constants, is already mentioned by Tao and Vu in his book on Additive Combinatorics [**24**].

Recently, Ben Green [**12**] gave an algebraic version of both, the SzRL and the Removal Lemma for groups. In this algebraic version the structural result fits into the algebraic structure in terms of subgroups. On the other hand, the Removal Lemma has its algebraic counterpart in the estimation of the number of solutions of equations in groups. Two applications of these algebraic versions are the fact that every set which contains few Schur triples can be set sum–free by removing a small portion, and the fact that a set of integers with positive density contains a large number of 3–term arithmetic progressions with a common difference.

The purpose of this Master Thesis is to give a detailed account on the SzRL and some of its applications, particularly to Additive Combinatorics. We particularly focus on the consequences of the SzRL related to the Counting Lemma. By combining the version by Alon and Shapira [**1**] of the directed version of the SzRL with the version of Simonovits for edge–colored graphs, we state and prove a Counting Lemma for edge colored directed graphs. One of the motivations of this statement is the potential application to Additive Combinatorics through Cayley graphs (which are directed and naturally arc–colored).

The methods used by Green to prove his versions of the Szemerédi Regularity Lemma and the Removal Lemma for groups heavily rely on Fourier Analysis, and as such, his results are applicable only to Abelian groups. By using our general version of the Counting Lemma we prove a generalization of Ben Green's Removal Lemma which is applicable to general finite groups, non necessarily abelian. This general version of the Removal Lemma is the object of a Research Note with Dan Kral [**16**] which was prepared during my participation in the Spring Combinatorics School [**28**] held in the Czech Republic in the Spring of 2007. The same strategy can be used to prove a Removal Lemma for a class of systems of equations.

This Thesis is organized as follows. In Chapter 1 we give the general notation and definitions which will be used throughout the work and we review the Szemerédi Regularity Lemma. Chapter 2 contains the proofs of the directed and edge colored versions of the SzRL. The statements of proofs for the Counting Lemma and Removal Lemma are detailed in Chapter 3. Chapter 4 describes some of the classical applications of the SzRL, particularly the Erdős-Stone Theorem, the $(6, 3)$–Theorem and the proof of Roth's Theorem on 3–term arithmetic progressions. Finally, in Chapter 5 we give the version of the Removal Lemma for arbitrary groups and also its extension to systems of equations. We also describe some of its applications. The final section contains some future work which arose in the preparation of this work and some open problems.

CHAPTER 1

Definitions and the Szemerédi Regularity Lemma

1. Definitions

In the first part of this work, mainly in chapters 2, 3 and 4 G will denote a graph, but also, in some parts will denote a group (Chapter 5 mostly): in both cases it is specified. It is also specified if the graph is directed or undirected. If $G = (V, E)$ will denote that $V = V(G)$ is the set of vertices and $E = E(G)$ the set of edges of the graph G. Usually $|V| = n$ and H will denote a subgraph, with $|V(H)| = h$.

Let $G = (V, E)$ be an undirected graph and let $X, Y \subseteq V$ be disjoint subsets of vertices. We denote by $\|X, Y\|$ the number of edges of G that connect one vertex in X with one vertex in Y. With this we define the edge-*density* of the pair (X, Y) as:

$$d(X, Y) := \frac{\|X, Y\|}{|X|\,|Y|}$$

Notice that it is a number between 0 (there are no edges) and 1 (the pair is full of edges), and represents the proportion of edges we have in between the pair.

We will denote $[N]$ as the set of the first N natural numbers.

If G is a graph and v a vertex, $N(v)$ is the *neighbourhood* of v, this is, the set of vertices v is connected to.

$\mathbb{N}$ will denote the natural numbers. $\mathbb{Z}$ will denote the integers.

2. The Szemerédi Regularity Lemma

In this section we present one of the many versions that one can find on the Szmerédi Regularity Lemma (SzRL). All these versions are equivalent and some of them can be found in the survey of Komlós and Simonovits [**15**]. We here present and prove one of the most popular forms of the SzRL which can be found in the book of Diestel [**7**] as well as in the above mentioned reference [**15**].

The SzRL describes an inherent structure of all graphs which becomes meaningful when we consider very large graphs (graphs with many vertices). Even if the structural result can be applied when the graph is sparse (sparse-specific versions of the lemma have been developed; see e.g. [**15**]), it is more efficient when applied to dense graphs (graphs whose set of edges has size $O(n^2)$).

The SzRL tell us that we can arrange the vertices of a graph in clusters with equal size (except for a residual small part) such that the graph behaves like a random graph: between most of pairs of clusters the degrees of the vertices are roughly equal, and the neighbourhoods are fairly uniformly distributed. These facts are precisely stated in the context of the regularity notions inherent to the SzRL.

2.1. Statement of the Lemma.

2.1.1. *The regularity concept.* In 1978 Szemerédi proved the Regularity Lemma as we know today. He used a weaker statement only for bipartite graphs to prove, in 1975, the celebrated Szemerédi Theorem for the integers: every set of integers with positive density contains arbitrarily long arithmetic progressions (see [**22**]). Before presenting the SzRL we must first introduce the notion of regularity.

Definition 1.1 (Regularity pair). *Let $\epsilon > 0$. Given a graph $G = (V, E)$ and two disjoint sets of vertices $A \subset V$ and $B \subset V$, we say that the pair (A, B) is ϵ-regular if for every $X \subset A$ and $Y \subset B$ such that*

$$|X| > \epsilon|A| \qquad \textit{and} \qquad |Y| > \epsilon|B|$$

we have

$$|d(X, Y) - d(A, B)| < \epsilon.$$

So the distribution of the edges of the whole pair behaves uniformly (with an ϵ-error), as we compare every pair of big enough subsets.

A partition $\{V_0, V_1, \ldots, V_k\}$ of the vertex set V, $|V| = n$, is *ϵ-regular* if:

- $|V_0| < \epsilon n$: we will refer to V_0 as the exceptional set.
- $|V_1| = |V_2| = \ldots = |V_k|$: all have the same size,
- all but at most ϵk^2 pairs (V_i, V_j), with $1 \leq i < j \leq k$, are ϵ-regular: most of the pairs are ϵ-regular.

The presence of V_0 is merely technical, as we want the other parts to have the same size.

2.1.2. *The Regularity Lemma.* Now we state the SzRL:

LEMMA 1.2 (Szemerédi Regularity Lemma, 1978, [**23**]). *For every $\epsilon > 0$ and every integer $m > 1$ there exist an integer $M = M(m, \epsilon)$ such that every undirected graph G of order at least m admits an ϵ-regular partition $\{V_0, V_1, \ldots, V_k\}$ of the set of vertices, with $m \leq k \leq M$.*

So, once the minimum number of sets in the partition and the ϵ are chosen, every graph can be partitioned in a bounded number of sets (clusters) such that the majority of pairs are ϵ-regular. This means that almost all the pairs are such that the edge density of every pair of large subsets is close to the edge density of the pair itself (with an ϵ-error): the pair is highly uniform.

A key feature of the lemma is related with upper bound of M: on the cardinality of the partition: although it may be huge, it only depends on ϵ and m. Notice that we could set the the partition to be the trivial one where $|V_1| = \ldots = |V_n| = 1$. In this extremal case the partition will be trivially ϵ-regular: the densities will be 0 or 1 depending whether there is an edge or not. The size of such a partition grows with n, hence is n-dependent whereas the one ensured by the lemma is not.

The lower bound m helps us in knowing the proportion of edges which are outside the cluster-sets: if m increases, then the proportion of edges that can be inside the clusters decreases and we have more edges outside.

2.2. Proof.

2.2.1. *Sketch of the proof.* The proof we present here is based on Diestel [**7**]. Different proofs which can be found for instance in Bollobás [**5**] or in Komlós and Simonovits [**15**] rely on similar strategies: they are all based on a potential-like function, which is defined slightly different in each version.

The proof is quite technical but the general idea consist in the following: first one should define one potential-like function that will be positive and bounded from above. The process will be iterative: since the very first step we will have one partition of the set of vertices with the claimed properties except for the ϵ-regularity (all but one of the sets with the same size, and the exceptional one small enough). At each step we will ask whether the partition is ϵ-regular or not. In case the answer is negative, we will manage to find another partition, with smaller sets (we must pay some price, and grow the number of sets, but not by "that much"), such that this new partition make the potential function grow (once ϵ is fixed it will grow by a constant amount). Thus, we reach an ϵ-regular partition as we should not violate the upper bound on the potential function.

2.2.2. *Naming and first lemmas.* Let G be a graph with $V = V(G)$ its vertex set, $|V| = n$. For a pair (A, B), A and B disjoint subsets of V we define the function q, that will be finally the potential-energy like function as:

$$q(A, B) := \frac{|A|\,|B|}{n^2} \cdot d^2(A, B) = \frac{\|A, B\|^2}{|A|\,|B|\,n^2}$$

First we extend the definition of q given above to partitions of those sets. If $\mathcal{A}$ is a partition of A and $\mathcal{B}$ is a partition of B, with $A, B \subset V$ disjoint:

$$q(\mathcal{A}, \mathcal{B}) := \sum_{A' \in A, B' \in B} q(A', B')$$

that is, the sum of q over all the possible pairs.

Now we define q for partitions of the set V without any distinguished set V_0. Let $\mathcal{P} = \{V_1, \ldots, V_k\}$ be a partition of the vertex set V, $|V| = n$. We extend q to $\mathcal{P}$ as follows:

$$q(\mathcal{P}) := \sum_{i<j} q(V_i, V_j)$$

Finally, if we have $\mathcal{P} = \{V_0, V_1, \ldots, V_k\}$ a partition with V_0 as the exceptional, distinguished set, we define:

$$q(\mathcal{P}) := q(\overline{\mathcal{P}})$$

where $\overline{\mathcal{P}} := \{V_1, \ldots, V_k\} \cup \{\{v\}, v \in V_0\}$, that is, we consider V_0 as the union of singletons and we apply q as we have no significate set. This leads us to define q without paying attention to the peculiar properties of the exceptional set V_0. This set will increase its size as the iteration process go on while the others parts get smaller. Note that each individual vertex is always ϵ-regular with respect to any other set; since we do not care about the ϵ-regularity of V_0 we simply make V_0 ϵ-regular by considering it as a union of parts of cardinality one. As it will be shown, the function q is monotone under refinement. Therefore, by considering the exceptional set as the union of singletons it will always be a refinement, avoiding the size growing problem.

As we will use q extensively, we will show some of its properties to help us to prove the Regularity Lemma.

LEMMA 1.3. *Let q be defined as above. Then*

(i) *q is bounded.*

(ii) *q is monotone increasing under partition refinement.*

PROOF. Let $\mathcal{P} = \{V_1, \ldots, V_k\}$ be a partition of V, $|V(G)| = n$. We have:

$$\begin{aligned} q(\mathcal{P}) &= \sum_{i<j} q(V_i, V_j) = \frac{1}{n^2}\sum_{i<j} |V_i||V_j| d^2(V_i, V_j) \\ &\overset{d\leq 1}{\leq} \frac{1}{n^2}\sum_{i<j} |V_i||V_j| \leq \frac{1}{n^2}\sum_{i,j} |V_i||V_j| = \frac{1}{n^2}\sum_{i} |V_i| \sum_{j} |V_j| \\ &= 1 \end{aligned}$$

Therefore, for a partition $\mathcal{P} = \{V_0, V_1, \ldots, V_k\}$ with a distinguished set V_0, we have

$$q(\mathcal{P}) = q(\overline{\mathcal{P}}) \leq 1.$$

On the other hand, since all the quantities we are adding up are always nonnegative, we trivially have

$$q(\mathcal{P}) \geq 0$$

This proves (i).

Let $\mathcal{C}$ and $\mathcal{D}$ be partitions of the sets C and D respectively. We shall show that:

$$q(\mathcal{C}, \mathcal{D}) \geq q(C, D) \tag{1}$$

To prove this we will use the Cauchy-Schwarz inequality:

$$\begin{aligned} q(\mathcal{C}, \mathcal{D}) &= \sum_{i,j} q(C_i, D_j) \\ &= \frac{1}{n^2}\sum_{i,j} \frac{\|C_i, D_j\|^2}{|C_i||D_j|} \\ &\overset{\textit{Cauchy-Schwarz},*}{\geq} \frac{1}{n^2} \frac{\left(\sum_{i,j} \|C_i, D_j\|\right)^2}{\sum_{i,j} |C_i||D_j|} \\ &\overset{\textit{we have all the products}}{=} \frac{1}{n^2} \frac{\|C, D\|^2}{\left(\sum_i |C_i|\right)\left(\sum_j |D_j|\right)} \\ &= q(C, D) \end{aligned}$$

$*$: by using Cauchy-Schwarz we know that $\sum_k a_k^2 \sum_k b_k^2 \geq \sum_k (a_k b_k)^2$. For the inequality choose $a_k = \sqrt{|C_i|D_j|}$ and $b_k = \|C_i, D_j\|/\sqrt{|C_i|D_j|}$ where k runs over all the unordered pairs $\{i, j\}$.

Let us now show (ii). Let $\mathcal{P}' = \{V_0', V_1', \ldots, V_{k'}'\}$ be a refinement of $\mathcal{P} = \{V_0, V_1, \ldots, V_k\}$. Let $\mathcal{V}_i$ be the partition induced by the sets of $\mathcal{P}'$ over $V_i \in \mathcal{P}$. Then,

$$\begin{aligned} q(\mathcal{P}) &= \sum_{i<j} q(V_i, V_j) \\ &\overset{\text{(1), seen before}}{\leq} \sum_{i<j} q(\mathcal{V}_i, \mathcal{V}_j) \\ &\overset{*}{\leq} \sum_{i<j} q(V_i', V_j') \\ &= q(\mathcal{P}') \end{aligned}$$

$*$: in the sums that arise from each term $q(\mathcal{V}_i, \mathcal{V}_j)$, there will be, maybe, some terms that are not in $q(\mathcal{P}')$ since $q(\mathcal{P}') = \sum_{i<j} q(V_i', V_j') = \underline{\sum_i q(\mathcal{V}_i)} + \sum_{i<j} q(\mathcal{V}_i, \mathcal{V}_j)$, ($q$ is symmetric). This shows (ii). ☐

The next step is to show that, for fixed ϵ, if we have a partition which is not ϵ-regular, we will be able to build up a new partition which is 'more regular' in such a way that q increases by a constant (depending on ϵ). So, because of Lemma 1.3 (i) (q is bounded from above), we should get some ϵ-regular partition before going beyond the upper bound.

Fist let us see what happens when there is just one ϵ–irregular pair (a pair which is not ϵ-regular). We can manage to find a partition which increases our potential function a bit (although it will not be a constant). This way, if we have many irregular pairs we will be able to increase q much more. However some care is needed to keep the partitions with precise properties, namely with parts of the same size, except one bounded-size exceptional set.

LEMMA 1.4. *Let $\epsilon > 0$ and let $C, D \subseteq V$ disjoint. If (C, D) is an ϵ-irregular pair, we can partition C and D in two parts $\mathcal{C} = \{C_1, C_2\}$ and $\mathcal{D} = \{D_1, D_2\}$ such that*

$$q(\mathcal{C}, \mathcal{D}) \geq q(C, D) + \epsilon^4 \frac{|C||D|}{n^2}$$

PROOF. If the pair (C, D) is not ϵ-regular there will be two sets $C_1 \subset C$ i $D_1 \subset D$ with $|C_1| > \epsilon|C|$ and $|D_1| > \epsilon|D|$ such that

$$|\mu| := |d(C_1, D_1) - d(C, D)| > \epsilon$$

Let $C_2 = C \setminus C_1$ and $D_2 = D \setminus D_1$, and define $\mathcal{C} := \{C_1, C_2\}$ and $\mathcal{D} := \{D_1, D_2\}$. Let us show that these partitions meet the statement of the lemma:

$$\begin{aligned}
q(\mathcal{C},\mathcal{D}) &= \frac{1}{n^2}\sum_{i,j}\frac{\|C_i,D_j\|^2}{|C_i||D_j|}\\
&\overset{\textit{take apart the interesting}}{=} \frac{1}{n^2}\left(\frac{\|C_1,D_1\|^2}{|C_1||D_1|}+\sum_{i+j>2}\frac{\|C_i,D_j\|^2}{|C_i||D_j|}\right)\\
&\overset{\textit{same argument used in Lemma 1.3(ii)}}{\geq} \frac{1}{n^2}\left(\frac{\|C_1,D_1\|^2}{|C_1||D_1|}+\frac{\left(\sum_{i+j>2}\|C_i,D_j\|\right)^2}{\sum_{i,j}|C_i||D_j|-|C_1||D_1|}\right)\\
&= \frac{1}{n^2}\left(\frac{\|C_1,D_1\|^2}{|C_1||D_1|}+\frac{(\|C,D\|-\|C_1,D_1\|)^2}{|C||D|-|C_1||D_1|}\right).
\end{aligned}$$

By the definition of $\mu := d(C_1,D_1)-d(C,D)$ we have

$$\|C_1,D_1\| \;=\; |C_1||D_1|\left(\mu+\frac{\|C,D\|}{|C||D|}\right).$$

To simplify the size of the formulae lets introduce some notation: $d_i=|D_i|$, $c_j=|C_j|$, $d=|D|$, $c=|C|$, $e_{ij}=\|C_i,D_j\|$, $e=\|C,D\|$. With this notation

$$\begin{aligned}
n^2q(\mathcal{C},\mathcal{D}) &\geq \frac{1}{c_1d_1}\left(\frac{c_1d_1e}{cd}+\mu c_1d_1\right)^2+\\
&\quad+\frac{1}{cd-c_1d_1}\left(\frac{cd-c_1d_1}{cd}e+\mu c_1d_1\right)^2\\
&= \frac{c_1d_1e^2}{c^2d^2}+\frac{2e\mu c_1d_1}{cd}+\mu^2c_1d_1+\\
&\quad+\frac{cd-c_1d_1}{c^2d^2}e^2-\frac{2e\mu c_1d_1}{cd}+\frac{\mu^2c_1^2d_1^2}{cd-c_1d_1}\\
&= \frac{e^2}{c^2d^2}+\mu^2c_1d_1\frac{cd}{cd-c_1d_1}\\
&\geq \frac{e^2}{c^2d^2}+\mu^2c_1d_1\\
&\overset{|\mu|>\epsilon,\ c_1\geq\epsilon c,\ d_1\geq\epsilon d}{\geq} \frac{e^2}{c^2d^2}+\epsilon^2\epsilon c\epsilon d
\end{aligned}$$

Hence,

$$q(\mathcal{C},\mathcal{D}) \;\geq\; \frac{e^2}{n^2c^2d^2}+\frac{\epsilon^4cd}{n^2}=q(C,D)+\frac{\epsilon^4cd}{n^2} \tag{2}$$

as desired. □

Once we know what happens when we refine a pair of sets, let us see what we can say in the general case. According to the statement of the Regularity Lemma we should eventually get a regular partition, this is, we should have, at most, ϵk^2 ϵ–irregular pairs. If the partition is not ϵ-regular we have, at least, ϵk^2 pairs that we want to refine in order to try to find an ϵ-regular partition, while increasing q. It is intuitively clear that, if by refining a pair we obtain a growth of $\frac{\epsilon^4 cd}{n^2}$, if we refine more than ϵk^2 (as $k \approx n/c$) this would imply that we should be able to increase q by $\approx \epsilon^5$; but, we should remember that we want to have nearly uniform partitions and some control over the size of the exceptional set. So we should proceed with care and see what we can get. Thus:

LEMMA 1.5. *Let $0 < \epsilon \leq 1/4$ and let $\mathcal{P} = \{V_0, V_1, \ldots, V_k\}$ be a partition of V, the set of vertices, with V_0 the exceptional set, verifying $|V_0| < \epsilon n$ and $|V_1| = |V_2| = \ldots = |V_k| =: c$. If $\mathcal{P}$ is not an ϵ-regular partition then there exists another partition $\mathcal{P}' = \{V_0', V_1', \ldots, V_{k'}'\}$ with exceptional set V_0', $k \leq k' \leq k4^k$, $|V_0'|$ is so that $|V_0'| \leq |V_0| + n/2^k$, the rest of V_i''s have the same size and*

$$q(\mathcal{P}') \geq q(\mathcal{P}) + \frac{\epsilon^5}{2}$$

Note : As the proof will show, the new partition $\mathcal{P}'$ will be a refinement of the old partition $\mathcal{P}$, so we can use the monotonicity of q.

PROOF. As for the one pair case, we will define a partition that allow us to increase q. For every pair of subscripts (i, j), $1 \leq i < j \leq k$, we define a partition $\mathcal{V}_{ij}$ of V_i and $\mathcal{V}_{ji}$ a partition of V_j as follows:

- If the pair (V_i, V_j) is already ϵ-regular then $\mathcal{V}_{ij} = \{V_i\}$ and $\mathcal{V}_{ji} = \{V_j\}$ (as the pair is already ϵ-regular, there is no need to change the partition locally).
- If the pair (V_i, V_j) is ϵ-irregular then we use the Lemma 1.4 for the one-pair case: we know there are partitions $\mathcal{V}_{ij}$ of V_i and $\mathcal{V}_{ji}$ of V_j into two sets ($|\mathcal{V}_{ij}| = |\mathcal{V}_{ji}| = 2$) such that:

$$q(\mathcal{V}_{ij}, \mathcal{V}_{ji}) \geq q(V_i, V_j) + \frac{\epsilon^4 c^2}{n^2} \tag{3}$$

Now with these locally-fine partitions, pair by pair, we build a partition for every set V_i such that it is consistent with the ones found pair by pair. So we take $\mathcal{V}_i$ as the partition that refines every partition $\mathcal{V}_{ij}$ with $|\mathcal{V}_i|$ minimum (the least partition that refines them all, and so we retain the partitions we have build pair by pair). Since we can build a partition of this kind by taking all the possible intersections between the sets in the partitions $\mathcal{V}_{ij}$, and since in each partition V_{ij} there are at most two parts, we have $|\mathcal{V}_i| \leq 2^{k-1}$. So the partition (of V) that we take to start with is:

$$\mathcal{C} := \{V_0\} \cup \bigcup_{i=1}^{k} \mathcal{V}_i,$$

with V_0 as the exceptional set (the same that we have for $\mathcal{P}$, the original partition we start the proof with). By the way we have defined $\mathcal{C}$ we know that it is a refinement of $\mathcal{P}$ and that

$$k \le |\mathcal{C}| \le k2^k \tag{4}$$

The partition of V_0 will be taken as a set of singletons: $\mathcal{V}_0 = \{\{v\}, v \in V_0\}$. Now, by the hypothesis of the lemma, $\mathcal{P}$ was not an ϵ-regular partition (if so we would have finished!) and, consequently, there exist ϵk^2 pairs (V_i, V_j), with $1 \le i < j \le k$ that have created some non-trivial partitions. Let us look at the value of q:

$$\begin{aligned} q(\mathcal{C}) &= \sum_{1\le i<j} q(\mathcal{V}_i, \mathcal{V}_j) + \sum_{1\le i} q(\mathcal{V}_0, \mathcal{V}_i) + \sum_{0\le i} q(\mathcal{V}_i) \\ &\overset{q\ monotony}{\ge} \sum_{1\le i<j} q(\mathcal{V}_{ij}, \mathcal{V}_{ji}) + \sum_{1\le i} q(\mathcal{V}_0, V_i) + q(\mathcal{V}_0) \\ &\overset{\ge \epsilon k^2\ pairs\ +\ (3)}{\ge} \sum_{1\le i<j} q(V_i, V_j) + \epsilon k^2 \frac{\epsilon^4 c^2}{n^2} + \sum_{1\le i} q(\mathcal{V}_0, V_i) + q(\mathcal{V}_0) \\ &= q(\mathcal{P}) + \epsilon^5 \left(\frac{kc}{n}\right)^2 \\ &\ge q(\mathcal{P}) + \epsilon^5/2. \end{aligned}$$

In the last inequality we take into account that $|V_0| \le \epsilon n \le \frac{1}{4}n$, and so (the rest): $kc \ge \frac{3}{4}n$. Notice also that the V_0 partition ($\mathcal{V}_0$) is the same when computing $q(\mathcal{C})$ as $q(\mathcal{P})$.

At this point we just have to transform $\mathcal{C}$ into a valid partition. To do this we simply cut the sets in $\mathcal{C}$ into parts with the same size and throw away the rest in the exceptional set. Thus these equal parts should be large enough (because the remaining exceptional set cannot grow much). We will take $V'_1, V'_2, \ldots, V'_{k'}$ disjoint subsets of V with size $c' := \lfloor c/4^k \rfloor$ such that every V'_i is a subset of one $C \in \mathcal{C} \setminus V_0$. We will take as many V'_i's as we can. The new exceptional set will be formed with the remaining parts: $V'_0 = V \setminus \bigcup V'_i$ (note that $V_0 \subseteq V'_0$). So the new partition will be: $\mathcal{P}' = \{V'_0, V'_1, \ldots, V'_{k'}\}$. Since we consider the exceptional set as partitioned into singletons, the resulting partition $\mathcal{P}'$ is a refinement of both $\mathcal{P}$ and also of $\mathcal{C}$, so that

$$q(\mathcal{P}') \ge q(\mathcal{C}) \ge q(\mathcal{P}) + \epsilon^5/2$$

At this point we have the sets of the new partition with the same size, but we have to look after the size of the exceptional set and after the total number of sets. We have that each V'_i, $i \neq 0$ is included in some V_j, $j \neq 0$. As we have taken $c' := \lfloor c/4^k \rfloor$, we have at most 4^k sets V'_i (they are pairwise-disjoint) inside every V_j. So we know there are $k \le k' \le k4^k$ as we want. On the other hand, as we have taken k' to be maximal we know that the number of vertices left over which are added to V_0' in every set of $\mathcal{C} \setminus \{V_0\}$ is less than c'. Thus:

$$\begin{aligned} |V_0'| &\leq |V_0| + c' \, |\mathcal{C} \setminus \{V_0\}| \\ &\leq |V_0| + \frac{c}{4^k}\left(k2^k\right) \\ &= |V_0| + ck/2^k \\ &\leq |V_0| + n/2^k \end{aligned} \tag{5}$$

concluding the proof of the lemma. □

2.2.3. *Regularity Lemma proof:* Now we are nearly finish, we just have to adjust the constants and find the value of M:

Proof of the Szemerédi Regularity Lemma 1.2. Without loss of generality we take $0 < \epsilon \leq 1/4$ because if it works for small ϵ it is also true for larger ones. Take the $m \geq 1$ given. As we have already seen it is enough with $s := 2/\epsilon^5$ iterations of Lemma 1.5. Indeed, if in each iteration the partition is *not* ϵ-regular, then we can find a refinement which increases the value of q by $\epsilon^5/2$; as we start in the worst case with $q = 0$ and we cannot exceed $q = 1$, we should find an ϵ-regular partition before $2/\epsilon^5$ iterations.

We have to fulfill the requirements involving the size of the exceptional set V_0, namely $|V_0| < \epsilon n$, and this should be valid for every iteration. We know that at each step, the exceptional set grows by at most $n/2^k$, where $k+1$ is the size of the partition. In the next iteration it will grow by $n/2^{k'}$ but, as $k \leq k'$, we have $n/2^k \leq n/2^{k'}$. Therefore, we can bound the growth of the exceptional set by $n/2^{k_0}$, where k_0 will be the initial partition.

Accordingly, we should choose an initial k_0 large enough in order to be sure that, in case of doing s iteration, we never exceed a bound, say for example, $\frac{1}{2}\epsilon n$. For instance, take the initial V_0 such that $|V_0| \leq \frac{1}{2}\epsilon n$ and thus $|V_0|$, is bounded by ϵn through $s = 2/\epsilon^5$ iterations.

To do this we will take n (the order of the graph) large enough to allow $|V_0| \leq \frac{1}{2}\epsilon n$ and also allow $V_1, V_2, \ldots, V_k$ to have the same cardinality. If we let $|V_0| \leq k$ (at most) we will be able to build k sets with the same cardinality: this will be just the starting point, as the other partitions will be given by the iteration procedure described in Lemma 1.5.

So we must have $k \geq m$ large enough to allow $s\frac{n}{2^k} \leq \frac{1}{2}\epsilon n$ so the inequality $s/2^k \leq \epsilon/2$ allows us to find the initial value for k: k is such that $s/2^k \leq \epsilon/2$. By letting k be large enough we will be able to achieve this.

If we want that the initial $|V_0| \leq k$, and that with $s\frac{n}{2^k}$ increments it does not go beyond ϵn, it suffices to set $k/n \leq \epsilon/2$. This can be achieved whenever $n \geq 2k/\epsilon$.

To find the value of M we will examine the growth of the size of the partitions through the iteration procedure. If the partition i-th partition has r sets then in the next step we will have a maximum of $r4^r$ sets. Let $f(r) := r4^r$, if we let $M = \max\{f^s(k), 2k/\epsilon\}$ we are done: the $f^s(k)$ solves the general iteration process and the $2k/\epsilon$ solves the "extremal" case where, if n is so small that we cannot be sure to have $n \geq 2k/\epsilon$ and, because of that, we cannot ensure a small enough exceptional set, then we simply use the trivial partition which is ϵ-regular for every ϵ (in fact, the key element on the Regularity Lemma is the bounded number of sets, once an ϵ is given).

Let us put all in order: let $G = (V, E)$ an undirected graph $|V| = n$, with $n \geq m$. We must see that we can find an ϵ-regular partition $\mathcal{P} = \{V_0, V_1, \ldots, V_k\}$ with $m \leq k \leq M$. We have a threshold for n: $n \leq M$ or $n > M$.

In the first case we take $k = n \leq M$ and the trivial partition (the singleton partition), that is ϵ-regular: $V_0 = \emptyset$, $|V_1| = \ldots = |V_k| = 1$.

In the more general case, $n > M$, we take the V_0, exceptional set, as a set with minimal cardinality such that k (k is the minimal that allow us to make s iterations without problems on the size of the exceptional set: say is such that $s/2^k \leq \epsilon/2$) divide $|V \setminus V_0|$ and let $\{V_1, \ldots, V_k\}$ a partition of $|V \setminus V_0|$ in k parts such that $|V_1| = \ldots = |V_k|$ (it does not matter the regularity!). The fact that $n > M \geq 2k/\epsilon \geq k$ implies $|V_0| < k$ and $|V_0| < \epsilon n/2 < \epsilon n$. Now we check whether the partition is ϵ-regular. If we are lucky we have finished, in the other case we just have to apply the Lemma 1.5 till we found one. We are sure we will finish before the number of sets go beyond the M barrier (because we have choose the constants to do so) and at every point $|V_0|$ will be always below ϵn, and all this before s iterations on Lemma 1.5. Obviously, once we have chosen ϵ and m, M is a constant (very big, but a constant). $\square$

Note : The grow of M given by this proof is what is called a *tower type* grow as we will have a tower $4^{4^{4^{\cdots}}}$ with a height of s.

CHAPTER 2

The Regularity Lemma for directed graphs and for colored directed graphs

In this chapter we will prove two versions of the Regularity Lemma: the Alon and Shapira's directed version of the lemma from 2003 [**1**], and a colored-digraph version. The edge-colored version for undirected graphs can also be found in [**15**, Theorem 1.18].

In the next chapter we prove a lemma which is called either *Counting Lemma* or also *Key Lemma* in the literature (see Chapter 3), which follows from the Regularity Lemma and which will be very helpful in various applications. The Regularity Lemma and the Counting Lemma together are often referred to as the Regularity Method.

1. The Regularity Lemma for directed graphs

Since the formulation of the original Regularity Lemma a number of different versions and generalizations have been considered. Here we present a generalization to the directed case due to Alon and Shapira in 2003 (see [**1**]) along with a Removal Lemma for directed graphs (presented and proved in Chapter 3, also see [**1**]). This will allow us to present a different proof of an extension to arbitrary groups of a theorem proved by Green [**12**] in Chapter 5.

We first introduce, like in the undirected graph version of the Regularity Lemma, an ϵ-regularity notion for digraphs. We will then formulate the statement of the Regularity Lemma for directed graphs and finally give the proof of this generalization.

1.1. The directed regularity notion and statement of the theorem. Here we present the regularity notion presented by Alon an Shapira in [**1**] for digraphs. We first give the notion of density of sets of vertices in a directed graph. Let $G = (V, E)$ be a directed graph and let $X, Y \subseteq V$ be disjoint subsets of vertices. We denote by $\overrightarrow{E}(X, Y)$ the set of edges going from X to Y and by $\overleftarrow{E}(X, Y)$ the set of edges from Y to X. We also denote by $\overline{E}(X, Y)$ the set of pairs of edges which form a 2–cycle between X and Y, that is, the pairs of edges $(x, y), (y, x) \in E(G)$ with $x \in X$ and $y \in Y$. Note that $|E(X, Y)| \leq |\overrightarrow{E}(X, Y)| + |\overleftarrow{E}(X, Y)| + |\overline{E}(X, Y)|$ (we count the edges $\overline{E}(X, Y)$ twice). With these notations we can define the directed edge-*densities* of the pair (X, Y) as:

$$\overrightarrow{d}(X,Y) := \frac{|\overrightarrow{E}(X,Y)|}{|X|\,|Y|}, \qquad \overleftarrow{d}(X,Y) := \frac{|\overleftarrow{E}(X,Y)|}{|X|\,|Y|}, \qquad \overline{d}(X,Y) := \frac{|\overline{E}(X,Y)|}{|X|\,|Y|}.$$

We can observe that, as in the undirected case, all three are real numbers between 0 and 1.

Now the definition of ϵ–regular pairs is defined as follows.

DEFINITION 2.1 (Digraph Regularity [**1**]). *Let $\epsilon > 0$. Given a digraph $G = (V, E)$ and two disjoint vertex sets $A \subset V$ and $B \subset V$, we say that the pair (A, B) is ϵ-regular if for every $X \subset A$ and $Y \subset B$ such that*

$$|X| > \epsilon|A| \qquad \textit{and} \qquad |Y| > \epsilon|B|$$

we have

$$|\overrightarrow{d}(X,Y) - \overrightarrow{d}(A,B)| < \epsilon, \qquad |\overleftarrow{d}(X,Y) - \overleftarrow{d}(A,B)| < \epsilon, \qquad |\overline{d}(X,Y) - \overline{d}(A,B)| < \epsilon,$$

the three at the same time.

Notice that the above definition is essentially the same as in the undirected case: a pair (A, B) will be ϵ-regular, if and only if, the pair is ϵ-regular in the original sense applied to each of the three graphs obtained by selecting only one class of the edges, $\overrightarrow{E}(X, Y)$, $\overleftarrow{E}(X, Y)$ or $\overline{E}(X, Y)$ and ignoring the directions.

On the other hand, the definition is an extension of the undirected regularity notion: we can convert an undirected graph into a digraph exchanging every edge by a 2–cycle. In that case the classes $\overrightarrow{E}(A, B)$, $\overleftarrow{E}(A, B)$ and $\overline{E}(A, B)$ will represent the same edges if we think them as undirected edges. So if a pair is ϵ-regular in the undirected graph sense, it will also be ϵ-regular in the digraph sense (if the digraph comes from an undirected graph).

With the notion of ϵ-regularity for a pair of disjoint vertices sets, we define an ϵ-regular partition in the digraph case the same way as for the undirected graph case. Thus a partition $\{V_0, V_1, \ldots, V_k\}$ of the vertex set of a digraph G of order n is ϵ–regular if $|V_0| < \epsilon n$, $|V_1| = |V_2| = \cdots = |V_k|$ and all but at most ϵk^2 pairs (V_i, V_j) are ϵ–regular.

Once defined the ϵ-regularity we can continue with the statement of the Digraph Regularity Lemma.

THEOREM 2.2 (Directed Szemerédi Regularity Lemma, 2003 [**1**]). *For every $\epsilon > 0$ and every $m \geq 1$ there exists an integer $DM = DM(m, \epsilon)$ such that every digraph of order at least m admits an ϵ-regular partition $\mathcal{P} = \{V_0, V_1, \ldots, V_k\}$ with $m \leq k \leq DM$.*

1.2. The proof for the directed case. The proof of Lemma 2.2 follows the same lines as the original SzRL. The strategy will consist in showing that, essentially, we can treat the digraph as three related undirected graphs. To prove Lemma 1.2 we have proved before Lemma 1.5 which shows the relation between the refinement of a non-ϵ-regular partition (in a certain, cleverly enough, way) and the function q.

Let $G = (V, E)$ be the digraph of order n and let $\mathcal{P}' = \{V_1, \dots, V_{k'}\}$ be a partition of V. We define a $\mathcal{P}'$-related partition of E into three sets, one for each "type" of directed edge:

$$\overrightarrow{E} = \{(u, v) \in E : u \in V_i, v \in V_j, i < j\}$$
$$\overleftarrow{E} = \{(u, v) \in E : u \in V_i, v \in V_j, i > j\}$$
$$\overline{E} = \{(u, v) \in E : (v, u) \in E, u \in V_i, v \in V_j, i \neq j\}$$

Notice that, whenever we have a partition $\mathcal{P}$ with an exceptional set, we will consider the suitable partition to define the sets $\overrightarrow{E}$, $\overleftarrow{E}$ and $\overrightarrow{E}$ as the partition where V_0 is considered as a set of singletons ordered by some prefixed order (so to compute q).

Notice also that these three sets are usually not necessarily disjoint because of $\overline{E}$. Note also that the union of the three sets do not cover all the edges in G: the edges inside each V_j are not covered. This is however a suitable choice since the ϵ-regularity only considers edges between parts of the partition.

Remark : Note that the definition of $\overrightarrow{E}$ and $\overleftarrow{E}$ depends on the ordering of sets within the partition. In the refinement process that will be considered in the proof, this ordering must be respected in order to keep track of the sets $\overrightarrow{E}$ and $\overleftarrow{E}$. On the other hand, the exceptional set V_0 will grow with additions from the other parts, and it will be partitioned into singletons. Thus these additions are not just placed in the 'first' set of the partition as this would alter the sets of edges we are considering from these singletons. In order to preserve the monotonicity of the function q under refinement, the relative order of the vertices which go to the 'bargain' set V_0 should be preserved so this exceptional set eventually consists of singletons scattered in the ordering established by the initial partition. This will be done in this way because:

- If we arrange V_0 "on the fly" during the, possibly many, iterations, the edges that connect V_0 with other vertices in $\overrightarrow{E}$, $\overleftarrow{E}$ will change from one set to the other and then we can have difficulties in computing q (we can loose the monotonicity).
- If we change the order on the new sets it may happen that an edge that was from the set $\overrightarrow{E}$, with the new iteration, went to $\overleftarrow{E}$ causing a change in the graphs $(V, \overrightarrow{E})$ and $(V, \overleftarrow{E})$ along the way, and therefore we loose control on the value of q.

Now we can see the partition $\mathcal{P}$ with its edges between V_i and V_j as three partitions $\overrightarrow{\mathcal{P}}$, $\overleftarrow{\mathcal{P}}$ and $\overline{\mathcal{P}}$ of three graphs on the same vertex set: in $\overrightarrow{\mathcal{P}}$ we just look at the edges in $\overrightarrow{E}$ and not in all E (in fact not in all $E \setminus \bigcup_{1\leq j\leq k} E(V_j)$, where $E(V_j)$ are the edges inside V_j); in $\overleftarrow{\mathcal{P}}$ we just look at the edges in $\overleftarrow{E}$ and in $\overline{\mathcal{P}}$ we just look at the edges in $\overline{E}$ (or the 2–cycles). Once the respective edge set is selected we will consider the edges to be undirected so we can apply the lemmas proved for the undirected case. In the case of $\overline{\mathcal{P}}$ we will consider each 2–cycle of $\overline{E}$ as one undirected edge.

Once we have $\overrightarrow{\mathcal{P}}$, $\overleftarrow{\mathcal{P}}$ and $\overline{\mathcal{P}}$, we can define $q(\overrightarrow{\mathcal{P}})$, $q(\overleftarrow{\mathcal{P}})$ and $q(\overline{\mathcal{P}})$ as the function $q(\cdot)$ over V (as $\mathcal{P}$ is a partition of V) by using the sets of edges $\overrightarrow{E}$, $\overleftarrow{E}$ and $\overline{E}$ as undirected edges respectively.

Remark : Note that, if each of $\overrightarrow{\mathcal{P}}$, $\overleftarrow{\mathcal{P}}$ and $\overline{\mathcal{P}}$ are ϵ–regular partitions in the undirected graph sense over the same vertex partition $\mathcal{P}$, then $\mathcal{P}$ is a 3ϵ–regular partition of V in the directed graph sense, that is:

- $|V_0| < \epsilon|V|$: because each of $\overrightarrow{\mathcal{P}}$, $\overleftarrow{\mathcal{P}}$ and $\overline{\mathcal{P}}$ share the same partition $\mathcal{P}$ of V.
- $|V_1| = |V_2| = \ldots = |V_k|$: by the same reason, the three partitions share the same partition $\mathcal{P}$.
- all but at most $3\epsilon k^2$ pairs with $1 \leq i < j \leq k$ are ϵ–regular: we need to put a 3 because the $\leq \epsilon k^2$ pairs that fail to be regular for $\overrightarrow{\mathcal{P}}$ may well not be the same pairs that fail to be ϵ-regular for $\overleftarrow{\mathcal{P}}$ or for $\overline{\mathcal{P}}$; but we can be sure about this $3\epsilon k^2$ bound on the irregular pairs.

Therefore, if $\overrightarrow{\mathcal{P}}$, $\overleftarrow{\mathcal{P}}$ and $\overline{\mathcal{P}}$ are $(\epsilon/3)$–regular partitions in the undirected graph sense, then $\mathcal{P}$ (the underlying vertex-set partition in $\overrightarrow{\mathcal{P}}$, $\overleftarrow{\mathcal{P}}$ and $\overline{\mathcal{P}}$) is an ϵ–regular partition in the directed graph sense.

With this remark in mind we just have to find an $(\epsilon/3)$–regular partition compatible with the three sets of edges: this is, a partition $\mathcal{P}$ of V that makes $\overrightarrow{\mathcal{P}}$, $\overleftarrow{\mathcal{P}}$ and $\overline{\mathcal{P}}$ to be $(\epsilon/3)$–regular partitions with the undirected graph regularity notion.

Proof of Theorem 2.2. Let ϵ be $0 < \epsilon \leq 1/4$ without loss of generality. Let $G = (V, E)$ be the given digraph. We can follow the same scheme as to prove the Szemerédi Regularity Lemma. Once a partition $\mathcal{P} = \{V_0, V_1, \ldots, V_k\}$ of V is given we can define $\overrightarrow{\mathcal{P}}$, $\overleftarrow{\mathcal{P}}$ and $\overline{\mathcal{P}}$ as described above and apply $q(\cdot)$ to each of them. By Lemma 1.3 (i), as we are considering them as undirected partitions, the three values $q(\overrightarrow{\mathcal{P}})$, $q(\overleftarrow{\mathcal{P}})$ and $q(\overline{\mathcal{P}})$ should be ≤ 1 (and also ≥ 0) for any $\mathcal{P}$.

We ask if $\overrightarrow{\mathcal{P}}$, $\overleftarrow{\mathcal{P}}$ and $\overline{\mathcal{P}}$ are $(\epsilon/3)$–regular partitions. If the answer is no we will apply Lemma 1.5 on $\mathcal{P}$ till $\overrightarrow{\mathcal{P}}$, $\overleftarrow{\mathcal{P}}$ and $\overline{\mathcal{P}}$ are $(\epsilon/3)$–regular partitions. Let us show how to do it.

If the partition $\mathcal{P}$ is not $(\epsilon/3)$–regular it means that there is one of the three partitions which is not $(\epsilon/3)$–regular. Let us suppose, for example, that $\overrightarrow{\mathcal{P}}$ is not a $(\epsilon/3)$–regular partition. We can apply Lemma 1.5 on $\mathcal{P}$ and find a new partition, say $\mathcal{P}'$, with $q(\overrightarrow{\mathcal{P}'}) \geq q(\overrightarrow{\mathcal{P}}) + (\epsilon/3)^5/2$. When building the new partition $\mathcal{P}'$ in Lemma 1.5 we have some aspects to consider:

- If we are looking for the new $\overrightarrow{\mathcal{P}'}$ we should take in consideration the set of edges $\overrightarrow{E}$ and update $\mathcal{P}'$ accordingly, without considering neither $\overleftarrow{E}$ nor $\overline{E}$.
- Because the definition of $\overrightarrow{E}$, $\overleftarrow{E}$ and $\overline{E}$ rely on the ordering of the partition sets V_i we should keep the ordering in the new partition $\mathcal{P}'$ compatible with the original one in $\mathcal{P}$. To do this we can imagine the sets V_i as boxes over an ordered line: when we partition each vertex set we can just place new separation panels into the box and deliver the vertices accordingly to the set-partition: if V_i and V_j are parts of $\mathcal{P}$ such that $i > j$ then, for each $V'_s \subset V_i$ and $V'_t \subset V_j$ belonging to the new partition $\mathcal{P}'$ we must have $s > t$. In doing so we know that, if an edge is from $\overrightarrow{E}(\mathcal{P})$ it will also be from $\overrightarrow{E}(\mathcal{P}')$ as the relative order in the sets remains unchanged. Of course this is also the case for the edges in $\overleftarrow{E}(\mathcal{P})$ and $\overleftarrow{E}(\mathcal{P}')$.
- Also the set V_0 will be, after some iterations, the remaining parts of the partitions $\mathcal{V}_i$ of the sets V_i which have to be neglected to make the parts of $\mathcal{P}$ but V'_0 of equal size: those parts will remain in the same place (spread all over the ordered line, as singletons, in the same place they first turned to be elements of V_0), that is, we will not rearrange them in the first position of V_0. We should keep the set $\overrightarrow{E}$ with the same edges and add the new ones which where inner edges of V_i and now appear connecting parts in this set which belong to the new partition $\mathcal{P}'$. Maintaining the new singletons in the relative order of the set they belonged to will not affect the regularity of the partition as we do not look at the edges that come from or enter into V_0 as a set when asking if a partition is regular.

By taking the above remarks into consideration we just have to notice that, when updating $\mathcal{P}$ because $\overrightarrow{\mathcal{P}}$ is not $(\epsilon/3)$–regular, say, we will build, by using Lemma 1.5, a refinement $\mathcal{P}'$ of $\mathcal{P}$ (see the note after Lemma 1.5) that will increase $q(\overrightarrow{\mathcal{P}})$ to $q(\overrightarrow{\mathcal{P}'}) \geq q(\overrightarrow{\mathcal{P}}) + (\epsilon/3)^5/2$. Since $\mathcal{P}'$ is a refinement of $\mathcal{P}$ and the graphs induced by the edges in $\overleftarrow{E}$ and $\overline{E}$ have not changed (except for the addition of new edges from the original digraph connecting parts of $\mathcal{P}'$ within parts of $\mathcal{P}$, by the monotonicity of $q(\cdot)$ we also have $q(\overleftarrow{\mathcal{P}'}) \geq q(\overleftarrow{\mathcal{P}})$ and $q(\overline{\mathcal{P}'}) \geq q(\overline{\mathcal{P}})$.

Thus we can be sure that within a maximum of $3 \cdot 2/(\epsilon/3)^5$ iterations of Lemma 1.5 (which corresponds to the maximum of $2/(\epsilon/3)^5$ iterations for each $\overrightarrow{\mathcal{P}}$, $\overleftarrow{\mathcal{P}}$ and $\overline{\mathcal{P}}$) we will find a partition $\mathcal{P}$ that will be a $(\epsilon/3)$–regular partition for the each $\overrightarrow{\mathcal{P}}$, $\overleftarrow{\mathcal{P}}$ and $\overline{\mathcal{P}}$. This will be an ϵ–regular partition in the digraph sense.

It just remains to check how are the constants. Following the SzRL proof we should change ϵ for $\epsilon/3$ (the ϵ we have used) and $s = 3 \cdot s'(\epsilon/3) = 3 \cdot 2/(\epsilon/3)^5 = 6/(\epsilon/3)^5$. Let $f(r) := r4^r$:

- k is such that $s/2^k \leq \epsilon/6$ and $k \geq m$ the m given.
- $DM = \max\{f^s(k), 6k/\epsilon\}$

If $n \leq DM$ then $k = n$ the partition is the trivial one: $|V_1| = \ldots = |V_k| = 1$

If $n > DM$ take an V_0 with minimal cardinality such that k (with $s/2^k \leq \epsilon/6$) divide $|V \setminus V_0|$ and let $\{V_1, \ldots, V_k\}$ be a partition of $|V \setminus V_0|$ in k parts such that $|V_1| = \ldots = |V_k|$. The fact that $n > DM \geq 6k/\epsilon \geq k$ implies $|V_0| < k$ and $|V_0| < \epsilon n/6 < (\epsilon/3)n$. We can proceed with Lemma 1.5 applied to the three partitions $\overrightarrow{\mathcal{P}}$, $\overleftarrow{\mathcal{P}}$ and $\overline{\mathcal{P}}$ and we know we will find an ϵ–regular partition in the digraph sense with $m \leq k \leq DM = DM(m, \epsilon)$. $\square$

2. The Regularity Lemma for directed an edge–colored graphs

The same strategy used in the proof of the directed version of the SzRL in Section 2 can be applied to obtain a more general version for edge–colored directed graphs. This more general statement will be used to state and prove a corresponding version for the Removal Lemma in Chapter 3, which in turn will be applied in Chapter 5 to deal with some systems of equations in finite groups.

The class of edge colored directed graphs occurs naturally in many applications. For instance it is considered by Nešetřil and Raspaud [**19**] where a generalized version of chromatic number is introduced for the so–called (n, m)–mixed graphs. An (n, m)–mixed graph has a set of undirected edges partitioned into n color classes and a set of directed edges partitioned into m color classes. We shall keep the convention of treating undirected edges as 2–cycles, so that we can restrict ourselves, without loss of generality, to directed graphs in which the set of arcs is partitioned into some number of color classes. Another natural source of edge–colored directed graphs are Cayley graphs, defined on a base group G which is also the vertex set of the graph and whose edges are of the form (x, xs) for $x \in G$ and s belonging to a fixed subset $S \subset G$. The edge (x, xs) is colored s. In an edge–colored directed graph we may look at some edge–colored subgraphs, distinguished not only by their graph structure but also by the colors on their edges.

2.1. The directed and edge–colored regularity notion and the statement of the theorem. Now we extend the regularity notion for digraphs to edge-colored digraphs. We first introduce some notation.

Let $\mathcal{L}$ be a set of colors and set $l := 3 \cdot |\mathcal{L}|$. For a pair of disjoint subsets A, B of vertices, the notation $E^*_\alpha(A,B)$ will denote the set of edges with color $\alpha \in \mathcal{L}$ with some direction, $\overrightarrow{E_\alpha}(A,B)$, $\overleftarrow{E_\alpha}(A,B)$ or $\overline{E_\alpha}(A,B)$. Here $\overrightarrow{E_\alpha}(A,B)$ denotes the set of edges from A to B with color α, $\overleftarrow{E_\alpha}(A,B)$ the ones from B to A with color α and $\overline{E_\alpha}(A,B)$ the set of 2–cycles between A and B in which both edges have the same color α.

We denote the coloring by $L : E(G) \to \mathcal{L}$ where $L(u,v)$ gives the color in $\mathcal{L}$ of the edge (u,v). We also define the edge–colored density of a pair A, B of disjoint subsets of vertices as

$$d^*_\alpha(A,B) = \frac{|E^*_\alpha(A,B)|}{|A| \cdot |B|}.$$

Definition 2.3 (Edge–colored Digraph Regularity). *Let $\epsilon > 0$ and let $L : E \to \mathcal{L}$ be an edge–coloring of a digraph $G = (V,E)$. For two disjoint sets $A \subset V$ and $B \subset V$, we say that the pair (A,B) is ϵ-regular if, for every $X \subset A$ and $Y \subset B$ such that*

$$|X| > \epsilon|A| \qquad \text{and} \qquad |Y| > \epsilon|B|$$

we have, for each color α from $\mathcal{L}$,

$$|\overrightarrow{d_\alpha}(X,Y) - \overrightarrow{d_\alpha}(A,B)| < \epsilon, \qquad |\overleftarrow{d_\alpha}(X,Y) - \overleftarrow{d_\alpha}(A,B)| < \epsilon, \qquad |\overline{d_\alpha}(X,Y) - \overline{d_\alpha}(A,B)| < \epsilon,$$

all three at the same time for each $\alpha \in \mathcal{L}$.

With this notation, the edge–colored digraph version of the Regularity Lemma is stated as follows:

Theorem 2.4 (Edge–colored Directed Regularity Lemma). *For every $\epsilon > 0$, every $m \geq 1$ and every set $\mathcal{L}$ of colors, there exists an integer $DMM = DMM(m, \epsilon, |\mathcal{L}|)$ such that every edge–colored digraph G with set of colors $\mathcal{L}$ and order at least m admits an ϵ–regular partition $\{V_0, V_1, \ldots, V_k\}$ with $m \leq k \leq DMM$.*

The proof of Theorem 2.4 follow the lines of the directed version Theorem 2.2.

2.2. Proof for the colored and directed case. Let ϵ be without loss of generality $0 < \epsilon \leq 1/4$. Let $G = (V,E)$ a directed graph and let $L : E \to \mathcal{L}$ be an edge–coloring of G. Let $\mathcal{P}' = \{V_1, \ldots, V_{k'}\}$ be a partition of the set V of vertices. We define $\mathcal{P}'$-related sets of edges, similarly as we did in the directed case, where any edge between V_i and V_j (for every $i \neq j$) will be in some of the (not necessarily disjoint) sets. For each color α in $\mathcal{L}$ we define the three sets:

$$\overrightarrow{E_\alpha} = \{(u,v) \in E : u \in V_i, v \in V_j, i < j, L(u,v) = \alpha\}$$
$$\overleftarrow{E_\alpha} = \{(u,v) \in E : u \in V_i, v \in V_j, i > j, L(u,v) = \alpha\}$$
$$\overline{E_\alpha} = \{(u,v) \in E : (v,u) \in E, u \in V_i, v \in V_j, i \neq j, L(u,v) = L(v,u) = \alpha\}$$

Notice that, whenever we have the partition $\mathcal{P} = \{V_0, V_1, \ldots, V_k\}$ with an exceptional set V_0, we will consider the suitable partition to define the sets $\overrightarrow{E_\alpha}$, $\overleftarrow{E_\alpha}$ and $\overleftrightarrow{E_\alpha}$ as the partition where V_0 is considered as a set of singletons ordered by some prefixed order (so to compute q).

Let $\mathcal{P} = \{V_0, V_1, \ldots, V_k\}$ be a partition with exceptional set V_0.

The sets $\overrightarrow{E_\alpha}$, $\overleftarrow{E_\alpha}$ and $\overline{E_\alpha}$ capture all the edges between V_i and V_j but never see the edges inside V_t, for $t \neq 0$. This is not the case for the exceptional set because we consider it as a set of singletons and all the edges inside V_0 will be in some set E^*_α. As done in the previous cases, we consider the V_0 as a set of its singletons to compute q, the function that will help us in proving the Theorem. Also, the V_0 will be, eventually, spread between the different new V_i' because the definition on $\overrightarrow{E_\alpha}$, for example, depends on the order of the sets in the partition: we want that, if an edge is from $\overrightarrow{E_\alpha}$, it remains in it under refinement of the partition. All the other remarks concerning the ordering of parts in refinements that are applicable in the uncolored directed case are applicable here as well.

We define, as before, $\overrightarrow{\mathcal{P}_\alpha}$ as the partition defined by $\mathcal{P}$ with respect to the uncolored directed graph whose edges are the ones in $\overrightarrow{E_\alpha}$ seen as undirected an uncolored edges. We do the same for all the possible directions and colors.

Now we just have to mimic the proof for the uncolored directed case.

Remark : If we find a partition $\mathcal{P}$ that makes, for every color $\alpha \in \mathcal{L}$, the three partitions $\overrightarrow{\mathcal{P}_\alpha}$, $\overleftarrow{\mathcal{P}_\alpha}$ and $\overline{\mathcal{P}_\alpha}$ ϵ–regular in the undirected graph sense we will have that, the $\mathcal{P}$ is an $l\epsilon$–regular partition in the edge colored digraph sense as we have:

- $|V_0| < \epsilon|V|$: because, for every color α, the three partitions $\overrightarrow{\mathcal{P}_\alpha}$, $\overleftarrow{\mathcal{P}_\alpha}$ and $\overline{\mathcal{P}_\alpha}$ share the same partition of V, $\mathcal{P}$.
- $|V_1| = |V_2| = \ldots = |V_k|$: by the same reason, the $l = 3|\mathcal{L}|$ partitions share the same partition $\mathcal{P}$.
- all but at most $l\epsilon k^2$ pairs with $1 \leq i < j \leq k$ are ϵ-regular: we need to put an l because we are not sure if the $\leq \epsilon k^2$ pairs that fail to be regular for $\overrightarrow{\mathcal{P}_\alpha}$ are the same pairs that fail to be ϵ-regular for $\overleftarrow{\mathcal{P}_\alpha}$ or for $\overline{\mathcal{P}_\alpha}$, so the l-factor is needed as the pairs can be all different.

As before we will achieve the ϵ–regularity partition in the edge–colored digraph case by finding an (ϵ/l)–regular partition that makes all the $\overrightarrow{\mathcal{P}_\alpha}$, $\overleftarrow{\mathcal{P}_\alpha}$ and $\overline{\mathcal{P}_\alpha}$, for every α, (ϵ/l)–regular in the undirected graph version. Notice that we will have a sure-much smaller exceptional set, but this is fine since if $|V_0| < (\epsilon/l)n$ then $|V_0| < \epsilon n$.

In case that some partition, $\overleftarrow{\mathcal{P}_\alpha}$ for exemple, fail to be an (ϵ/l)–regular partition, we simply apply Lemma 1.5. We should maintain the relative order on the new partition of V, $\mathcal{P}'$, that is a refinement of $\mathcal{P}$ as explained in the uncolored directed version in order to guarantee the monotonicity of q under refinements.

As $\mathcal{P}'$ is a refinement of $\mathcal{P}$ and the edges remain in place (if $e \in \overrightarrow{E_\alpha}(\mathcal{P})$ then $e \in \overrightarrow{E_\alpha}(\mathcal{P}')$), for every color α, $q(\overrightarrow{\mathcal{P}'_\alpha}) \geq q(\overrightarrow{\mathcal{P}_\alpha})$, $a(\overleftarrow{\mathcal{P}'_\alpha}) \geq q(\overleftarrow{\mathcal{P}_\alpha})$ and $q(\overline{\mathcal{P}'_\alpha}) \geq q(\overline{\mathcal{P}_\alpha})$.

Because we have chosen $\overleftarrow{\mathcal{P}_\alpha}$ (it were not (ϵ/l)–regular) we have increased $q(\overleftarrow{\mathcal{P}'_\alpha})$ to $q(\overleftarrow{\mathcal{P}_\alpha})+(\epsilon/l)^5/2$ (at least) while, as is a refinement partition, all the other $q(\cdot)$ evaluations do not decrease (for all other colors and directions). Maybe we have turned some already-(ϵ/l)–regular partitions into irregular ones but, we have not decreased $q(\cdot)$. As $q(\cdot)$ is bounded from above by 1 we can be sure that, before doing a maximum of $s = 2l/(\epsilon/l)^5$ iterations (l times the usual one-color case number) we will reach an (ϵ/l)–regularity partition $\mathcal{P}$ for all the "graphs" $\overrightarrow{\mathcal{P}_\alpha}$, $\overleftarrow{\mathcal{P}_\alpha}$ and $\overline{\mathcal{P}_\alpha}$ (and for every color!).

So we just have to adjust the constants, k and DMM, on the color and directed ϵ–regular sense partition $\mathcal{P}$. Let $f(r) := r4^r$ and $s = 2l/(\epsilon/l)^5$ (the number of iterations we know):

- k is such that $s/2^k \leq \epsilon/(2l)$ and $k \geq m$ the m given.
- $DMM = \max\{f^s(k), 2lk/\epsilon\}$

If $n \leq DMM$ then $k = n$ the partition is the trivial one: $|V_1| = \ldots = |V_k| = 1$.

If $n > DMM$ take an V_0 with minimal cardinality such that k (with $s/2^k \leq \epsilon/(2l)$) divide $|V \setminus V_0|$ and let $\{V_1, \ldots, V_k\}$ a partition of $|V \setminus V_0|$ in k parts such that $|V_1| = \ldots = |V_k|$. The fact that $n > DMM \geq 2lk/\epsilon \geq k$ implies $|V_0| < k$ and $|V_0| < \epsilon n/(2l) < (\epsilon/l)n$. We can proceed with Lemma 1.5 applying to one of the l cases: $\overrightarrow{\mathcal{P}_\alpha}$, $\overleftarrow{\mathcal{P}_\alpha}$ and $\overline{\mathcal{P}_\alpha}$ (for all $\alpha \in \mathcal{L}$) whenever that partition is not ϵ/l–regular. We know we will find an ϵ-regular partition in the edge–colored digraph sense with $m \leq k \leq DMM = DMM(m, \epsilon, |\mathcal{L}|)$ after a maximum of s iterations. This concludes the proof.

Notice that we could have composite colors and deal with them just by considering each composite color as a new one. An edge with composite color $r + b$ will be in the set E^*_b in E^*_r and also in E^*_{r+b}.

CHAPTER 3

The Removal Lemma

One of the most useful applications of the SzRL is the so-called *Counting Lemma*. The Counting Lemma provides information on the (asymptotic) number of copies of a given subgraph H contained in a graph G. This information is obtained from regular partitions of G whose existence is guaranteed by the SzRL.

In this chapter we present a Counting Lemma for directed arc-colored graphs (see Lemma 3.1). This Counting Lemma is in turn used to prove three versions of the so–called *Removal Lemma*. The Removal Lemma is a type of result which states that in a graph G which contains not many copies of a given graph H, these copies can be eliminated by removing a small number of edges in the graph. We first give the classical version of the Removal Lemma for graphs. We then present the version for directed graphs and for edge–colored directed graphs. The versions of the Removal Lemma we give in this chapter will be the core for the proofs of various applications in Chapter 4 and Chapter 5.

Together the SzRL plus the Counting Lemma (also known as Key Lemma, see e.g. [**15**]) are often known as the *Regularity Method*, as they form the basic machinery for many applications mostly related with graph theory and combinatorics.

1. Counting Lemma for edge–colored digraphs

The SzRL assures that, given a positive integer m and an $\epsilon > 0$, we can find in every sufficiently large graph an ϵ-regular partition into parts, or clusters of vertices, $\{V_0, V_1, \ldots, V_k\}$. With this partition we can consider the so–called *reduced graph*, R. The reduced graph associated to an ϵ–regular partition $\{V_0, V_1, \ldots, V_k\}$ of G is the graph with k vertices, one for each set V_i but the exceptional set. There is an edge between two vertices whenever the corresponding pair of clusters is an ϵ-regular pair with density above a fixed d; because we want edges that represent significantly the original graph, we safely can ignore edges joining regular pairs of low density, and since there are not many irregular pairs, the total number of edges neglected is relatively small. Thus the reduced graph captures the main structural properties of the original graph. The Counting Lemma provides quantitative measures of the above generic statements.

We will state and prove a Counting Lemma for edge–colored directed graphs. This general version includes the uncolored (one only color) and undirected (every directed edge belongs to a 2–cycle) versions.

Let $L : E \to \mathcal{L}$ be an edge–coloring of a directed graph $G = (V, E)$. We will use the notation introduced in Chapter 2. Thus $\overrightarrow{E_\alpha}(A, B)$, $\overleftarrow{E_\alpha}(A, B)$ and $\overline{E_\alpha}(A, B)$ denotes the sets of edges of color $\alpha \in \mathcal{L}$ directed from A to B, from B to A and belonging to a 2–cycle between A and B respectively. We also denote by $E^*_\alpha(A, B)$ one of the sets $\overrightarrow{E_\alpha}(A, B)$, $\overleftarrow{E_\alpha}(A, B)$ or $\overline{E_\alpha}(A, B)$. By an α^*–edge joining vertices u and v we mean either an edge from u to v, or an edge from v to u or a 2–cycle with these two vertices, in all three cases colored by α. We define $l := |\mathcal{L}|$ if the graph is undirected and $l := 3 \cdot |\mathcal{L}|$ if G is directed.

Given a directed graph $G = (V, E)$ we will construct the *reduced graph* $R := R(G, d, \epsilon, \mathcal{L})$ of G associated to a regular partition (in the sense of edge–colored directed graphs) $\mathcal{V} := \{V_0, V_1, \ldots, V_k\}$, with exceptional set V_0, as follows:

- Delete the exceptional set and any edge that has a vertex in V_0.
- Delete the edges inside each cluster.
- Delete all the edges between ϵ-irregular pairs.
- Take the quotient graph by the relation induced by the partition and name its vertices $v_1, \ldots, v_k$, where v_i stands for the cluster V_i.
- Put an edge labeled α^* from v_i to v_j if there is a set of surviving edges $E'^*_\alpha(V_i, V_j)$ with density $d^*_\alpha(V_i, V_j)$ more than d. If $|\overrightarrow{E_\alpha}'(V_i, V_j)| > d|V_i||V_j|$ then put an edge labeled $\overrightarrow{\alpha}$, if $|\overleftarrow{E_\alpha}'(V_i, V_j)| > d|V_i||V_j|$ put an edge labeled $\overleftarrow{\alpha}$ and if $|\overline{E_\alpha}'(V_i, V_j)| > d|V_i||V_j|$ then put another edge labeled $\overline{\alpha}$.

The above procedure gives an undirected multigraph with at most l parallel edges where an α^*-labeled edge between joining vertices v_i and v_j corresponds to at least $d\,|V_i|\,|V_j|$ edges in $E^*_\alpha(V_i, V_j)$ and we know that (V_i, V_j) is an ϵ–regular pair.

Let H be a subgraph of G with h vertices. Let $R := R(G, d, \epsilon, \mathcal{L})$ be the reduced graph of G associated to an ϵ–regular partition $\mathcal{V} := \{V_0, V_1, \ldots, V_k\}$. We say that a map $\phi : V(H) \to V(R)$ is an homomorphism from an edge-colored digraph to an edge-colored multigraph if, whenever there is an α^* edge between v and w then there is an α^*-labeled edge between $\phi(v)$ and $\phi(w)$. We write $H \to R$ if there is such an homomorphism from H to R. We also denote by $|H \subset G|$ the number of subgraphs isomorphic to H in G. We can now state a Counting Lemma that is a mixture from the [**15**, Theorem 2.1] and the [**1**, Lemma 4.1].

Lemma 3.1 (Counting Lemma for edge–colored directed graphs). *Let $d \in (0, 1)$ and $\epsilon > 0$ be given. Let H be an edge–colored directed graph of order h. Let G be a graph of order n and let*

$R = R(G, d, \epsilon, \mathcal{L})$ be a reduced graph of G corresponding to an ϵ–regular partition with cluster size m. Let $\delta := d - \epsilon$ and $\epsilon_0 := \delta^h/(l\,h+2)$.

If $\epsilon \le \epsilon_0$, $h < \epsilon_0 m$ and

$$H \to R,$$

then

$$|H \subset G| > \left(\frac{\epsilon_0}{\sqrt[h]{h!}} m\right)^h .$$

Remark : The conditions on ϵ_0 are not too restrictive and they can be achieved by

- concerning h, we just need to take n large enough,
- concerning the relationship between ϵ and ϵ_0, once d is arbitrarily small but fixed, we just have to choose ϵ sufficiently small: the difference between d and ϵ will make δ^h large enough so to achieve $\epsilon \le \delta^h/(l\,h+2)$. See the proof of the Corollary 3.3 as an example.

The inequalities $\epsilon \le \epsilon_0$, $h < \epsilon_0 m$ are just technical conditions, but important for everything to work, as the proof will show.

1.1. Proof. First we prove an interesting property about the regular pairs. Let $N_Y^{\alpha^*}(v)$ count the number of neighbours of v in Y connected with an α^*–edge.

PROPOSITION 3.2. *Let (A, B) be an ϵ-regular pair of an edge–colored directed graph with densities $d^*_\alpha := d^*_\alpha(A, B)$.*

If $Y \subset B$ has cardinality at least $\epsilon|B|$, then for all but at most $l\epsilon|A|$ vertices $v \in A$, the inequalities $N_Y^{\alpha^*}(v) \ge (d^*_\alpha - \epsilon)|Y|$ hold for every color and direction α^*.

PROOF. If (A, B) is ϵ-regular we know that, for each pair $X \subset A$ and $Y \subset B$ such that

$$|X| > \epsilon|A| \qquad \text{and} \qquad |Y| > \epsilon|B|$$

we have

$$|d^*_\alpha(X, Y) - d^*_\alpha| < \epsilon, \qquad \text{for all } \alpha^*.$$

Suppose that there is a subset $X_{\alpha^*} \subset A$ with $|X_{\alpha^*}| > \epsilon|A|$ such that, for all $v \in X_{\alpha^*}$, we have $N_Y^{\alpha^*}(v) < (d^*_\alpha - \epsilon)|Y|$, then, $d^*_\alpha - N_Y^{\alpha^*}(v)/|Y| > \epsilon$. But

$$\begin{aligned} d^*_\alpha(X_{\alpha^*}, Y) - d^*_\alpha &= \frac{\sum_{v \in X_{\alpha^*}} N_Y^{\alpha^*}(v)}{|X_{\alpha^*}|\,|Y|} - d^*_\alpha \\ &< \frac{(d^*_\alpha - \epsilon)|Y|\,|X_{\alpha^*}|}{|X_{\alpha^*}|\,|Y|} - d^*_\alpha \\ &= -\epsilon \end{aligned}$$

which makes the pair (A, B) ϵ-irregular. Therefore, a set X_{α^*} for which the inequality

$$N_Y^{\alpha^*}(v) \geq (d_{\alpha^*} - \epsilon)|Y|$$

does not hold has size $|X_{\alpha^*}| \leq \epsilon|A|$. Since $|\bigcup_{\alpha^*} X_{\alpha^*}| \leq \sum_{\alpha^*} |X_{\alpha^*}| \leq l\epsilon|A|$, all the inequalities $N_Y^{\alpha^*}(v) \geq (d_{\alpha^*} - \epsilon)|Y|$ hold for all $v \in A$ except for the vertices in a subset of A of size at most $l\epsilon|A|$. □

With this proposition proved we begin the proof of Lemma 3.1. Take the given d and ϵ. Define $\delta := d - \epsilon$. Let $incr_H : H \to R$ be an homomorphism. Let $u_1, \ldots, u_h$ be the vertices of H and denote by $v_{\sigma(i)} = incr_H(u_i)$.

First replace each vertex v_i in R by an m-cluster V_i (a cluster with m vertices) with no edges inside. If between the vertices v_i and v_j in R we have an α^*-labeled edge then between the corresponding clusters (V_i, V_j) we will build an ϵ-regular pair with at least dm^2 edges of type α^*. Let us call G' to that graph.

Let us show that there are many copies of H in G'. We proceed iteratively: first we will find some candidates to be v_1, some others to be v_2, and so on till find a set of candidates to be v_h such that the subgraph spanned by $v_1, \ldots, v_h$ in G' contains H as a subgraph.

In fact, the process will be dynamic: the size of the set where v_i belongs to will depend on (because the set itself will depend on) the choice of v_j for all $j < i$. The dependence is on the number of edges that have some v_j as one of its ends and v_i in the other one, and, since the choice of v_j had itself depended on the preceding v_k's, v_i depend on the choice of the preceding v_j's for all $j < i$. But we will show that, if we choose them in a certain way, thanks to the ϵ-regularity, we will be able to ensure that the size of this set will be large enough, no matter the choice of v_j's.

We will see that we can choose v_1 from a set such that, for any choice of v_1 inside this set, we will be able to build sets for v_i (for all $i = 2, \ldots, h$) such that the size of those sets are large enough and, even if the sets depend on the possible choice of v_1, the size of the sets has a lower bound which is independent of that choice.

Let us define these sets. Let $C_{i,j}$ be the set where v_i will belong to at step j (this is, after choosing the first j elements of the copy of H inside G'). Initially $C_{i,0} = V_{\sigma(i)}$ with $|C_{i,0}| = m$, and in fact $C_{i,j} \subseteq V_{\sigma(i)}$ for all i and $j < i$. If we select $v_j \in C_{j,j-1}$ to be the j–th vertex from one copy of H inside G' we should update the existing $C_{i,j-1}$: we must intersect the current set $C_{i,j-1}$ with a proper neighbourhood of v_j: if u_i and u_j are connected with an edge α^* in H then $C_{i,j} = C_{i,j-1} \cap N_{G'}^{\alpha^*}(v_j) = N_{C_{i,j-1}}^{\alpha^*}(v_j)$. This is: we intersect the current set $C_{i,j-1}$ containing the candidates for v_i with the α^*-neighbourhood of v_j in G'; if there is no edge joining u_i with

u_j then we set $C_{i,j} = C_{i,j-1}$. We should do this intersections with all the H-neighbours of u_j that have an index bigger than j (as the other vertices has been already chosen) because, if we have selected v_j, all his H-neighbours should be reflected in G'-neighbours if we want a copy of H inside G'.

Once we know what we will do, we want to do it right, this is, we want $C_{i,j}$ to be quite big for all $j < i$ so that we have many choices when the time i to select v_i arrives. At this point we will use the ϵ-regularity. As all the edges in R are now many edges (more than dm^2 per type) that form an ϵ-regular and high-density pair, we can try to apply Proposition 3.2 that tells us about minimum size neighbourhoods in "high" density ϵ-regular pairs. To apply Proposition 3.2 we must have that $|C_{i,j-1}| > \epsilon|V_{\sigma(i)}|$ (as is the "arrival set" we want to intersect with the neighbourhood of v_j). We should manage to achieve this lower bound for all $i > j$ such that $(v_j, v_i) \sim (u_j, u_i)$ is an edge of H. Let us suppose we are in the hypothesis of Proposition 3.2: we know that all but at most $l\epsilon|V_{\sigma(j)}|$ vertices in $V_{\sigma(j)}$ are such that:

$$|C_{i,j}| = |N^{\alpha^*}_{C_{i,j-1}}(v_j)| \geq (d^*_\alpha - \epsilon)|C_{i,j-1}| \geq (d - \epsilon)|C_{i,j-1}| = \delta|C_{i,j-1}|.$$

As H has h vertices it can happen that at each choice with $i > j$ we have to exclude $l\epsilon|V_{\sigma(j)}|$ as candidates for v_j. Since it may happen that many vertices belong to the same cluster $V_{\sigma(j)}$, we have at least $|C_{j,j-1}| - l\epsilon h|V_{\sigma(j)}|$ candidates for v_j.

So we have, for any $i > j$ (as we have done a maximum of h iterations):

$$|C_{i,j}| - l\epsilon h|V_{\sigma(j)}| - h \geq \delta^h m - l\epsilon hm - h = (\delta^h - l\epsilon h)m - h.$$

In order to apply Proposition 3.2 we should have $\delta^h m - l\epsilon hm - h > \epsilon m$. At this point we need that:

$$(\delta^h - l\epsilon h)m - h = (\epsilon_0(l\ h + 2) - l\epsilon h)m - h \geq 2\epsilon_0 m - h > \epsilon_0 m \geq \epsilon m = \epsilon|V_{\sigma(p)}|.$$

The first equality comes from the definition of ϵ_0 in the lemma statement, the other inequalities come from the relation of ϵ_0 with ϵ and h.

Now we are nearly finished: once we know the size of each $C_{i,j}$ at every step we know that, with independence of the previous selections, we have $> \epsilon_0 m \geq \epsilon m$ possible choices for v_j, for every j. Therefore,

$$|H \subset G'| > (\epsilon_0 m)^h \Longrightarrow |H \subset G| > (\epsilon_0 m)^h$$

Because we have just used the elements that share both G and G': the ϵ-regularity condition and the Regularity Lemma. We have built a graph G' that, maybe, it is not a subgraph of G but, as we have not used properties of G but the cluster configuration (that both of them share) we have proved the claim.

Remark : Here we have counted the vertex-labelled copies of H as subgraphs of G, if we want vertex-unlabeled copies, as is the case, we can be sure that $|H \subset G| > (\epsilon_0 m)^h/h! = (\epsilon_0/\sqrt[h]{h!})^h m^h$. We can simply divide by $h!$: the maximum order of the automorphism group of H.

With this remark we have finished the proof.

Notice that this $h!$ factor can be better if the map $incr_H$ has nice properties: if $incr_H$ is an homomorphism, the sets $V_{\sigma(i)}$ do not intersect and then we can reduce it to 1. This is the case when we are finding complete subgraphs in G.

2. The Removal Lemma: undirected case

Once we have the Counting Lemma 3.1 it is fairly easy to proof the Removal Lemma for the various cases: undirected, directed and multicolored graphs. The Removal Lemma has many applications, both the direct statement and the reciprocal one. It will allow us to prove the $(6,3)$ Theorem as well as the Roth's theorem (see Chapter 4 and [**21**], for the Roth Theorem see [**20**]). Also, the Removal Lemma for directed graphs will allow us to prove the [**12**, Theorem 1.5] in another way (Theorem 5.2 in this work).

We start with the Removal Lemma for the undirect case which was an observation made by Füredi.

COROLLARY 3.3 (Removal Lemma for Graphs). *For every $\beta > 0$. Let G_n be a graph with n vertices and a subgraph $H \subset G_n$. If there is a $\gamma = \gamma(\beta, H) > 0$ such that G_n is a graph with at most γn^h copies of H, then by deleting at most βn^2 edges one can make G_n H-free.*

Note : The relation between β and γ is explicit and if we have a graph with $o(n^h)$ H-subgraphs, then we can let $\beta \to_{n\to\infty} 0$, so we can delete $o(n^2)$ edges.

This will be a corollary of the Counting Lemma 3.1 with one color and for the undirected graph case (directed case with all the edges as 2–cycles). This is: $l = 1$. We can invest the βn^2-deletable edges in converting a graph into its reduced graph: once find a regular partition, remove all of the edges that the reduced graphs does not represent and then use the Counting Lemma. As the maximum number of clusters is bounded, we should be able to put the size of the cluster as a function of n, and hence, find many copies of H inside G_n.

PROOF. Let d be an edge density and let $\epsilon > 0$ be a constant with $d > \epsilon$. Let G be a graph with n vertices. Let m be the minimum number of clusters allowed. Find an ϵ-regular partition for G, $\mathcal{P} = \{V_0, V_1, \ldots, V_k\}$, with $p = |V_1|$. Let $M = M(\epsilon, m)$ be the upper bound for the number of partition sets given by the Regularity Lemma with m as the lower bound.

Remark : As $m \leq k \leq M$ we have that $kp \leq n$ and $p \geq n/M$

Lets see how many edges should be removed to get the Reduced Graph:

- Remove the edges that touch V_0: if we suppose the vertices in V_0 has maximum degree then $\leq |V_0|n < \epsilon n^2$ total edges.
- Remove the edges inside each V_i, $i \neq 0$: those are $\leq k\binom{p}{2} \leq kp^2/2 \leq \frac{n^2}{2k}$.
- Remove the edges that are between two non-ϵ-regular pairs: $\leq \epsilon k^2\binom{p}{2} \leq \epsilon(kp)^2/2 \leq (\epsilon/2)n^2$.
- Remove the edges that, although they are between an ϵ-regular pair, they compute a density less than d: suppose all the pairs are like so this makes $\leq \binom{k}{2}dp^2 \leq dn^2/2$ total edges.

Lets name this graph G'.

We have that if $\mathcal{P}$ is an ϵ-regular partition we can produce a reduced graph by deleting at most:

$$\epsilon n^2 + \frac{n^2}{2k} + \frac{\epsilon n^2}{2} + \frac{dn^2}{2} = \left(\epsilon + \frac{\epsilon}{2} + \frac{d}{2} + \frac{1}{2k}\right) n^2$$

So if we choose the minimum number of clusters as $m = 1/\epsilon$, also take $d = \beta$ and choose ϵ with $\beta/3 > \epsilon$ and such that $\epsilon \leq (\beta - \epsilon)^h/(2+h) = \epsilon_0$, we will be able to apply the Counting Lemma. By letting $\epsilon = (\beta/4)^h$ is enough: with this we have removed less than βn^2 edges. We can find this number by: if $\epsilon_0 \geq \epsilon$ then:

$$\frac{(d-\epsilon)^h}{h+2} \geq \epsilon \Longrightarrow h\ln(\beta - \epsilon) - \ln(\epsilon) > \ln(h+2) \Longrightarrow h\ln\left(\frac{\beta - \epsilon}{\sqrt[h]{\epsilon}}\right) \geq \ln(h+2)$$

So if we let $\epsilon = (\beta/4)^h$ we have:

$$h\ln\left(\frac{\beta - \left(\frac{\beta}{4}\right)^h}{\frac{\beta}{4}}\right) \geq \ln(h+2) \Longrightarrow h\ln\left(\frac{\frac{3}{4}\beta}{\frac{\beta}{4}}\right) \geq \ln(h+2) \Longrightarrow h\ln(3) \geq \ln(h+2)$$

as we want, because we can reverse the implications.

Now we have two possible configurations: the graph that remains after deleting those edges can be H-free or not.

Suppose that G' is not H-free: this means that we could transform G' into the Reduced Graph R (by continuing the process: collide the vertices in the same V_i to v_i, etc.) and find a monomorphism $incr_H$ from H to R so that $H \to R$. We can construct this map by sending each vertex u_i from H into the vertex in R following the copy of H in G': each vertex u_j of H will go to a cluster, say V_i: decide to send $incr_H(u_j) = v_i$, where v_i is the representant of V_i in the reduced graph. The application is well defined because we are just using edges that are from ϵ-regular pairs and with densities more than d. In this case we have found not just one copy of H but many. Applying the Lemma 3.1, using the remark that follows the lemma and the chose of ϵ we have:

$$|H \subset G| > \left(\frac{\epsilon_0 p}{\sqrt[h]{h!}}\right)^h \geq \left(\frac{\epsilon p}{\sqrt[h]{h!}}\right)^h \geq \left(\frac{\epsilon n}{M\sqrt[h]{h!}}\right)^h = \left(\frac{(\beta/4)^h n}{M\sqrt[h]{h!}}\right)^h = \left(\frac{(\beta/4)^h}{M\sqrt[h]{h!}}\right)^h n^h.$$

So if we let $\gamma = \left(\frac{(\beta/4)^h}{M\sqrt[h]{h!}}\right)^h$ the proportion of copies of H in G, we are sure that in G' there would be no copy of H: so G' is H-free (if there where some copy, we would have a contradiction).

We have found a sure-threshold for the number of copies of H in G that can be deleted by removing a maximum of βn^2 edges in G. We have proved the claim because M is just dependent on ϵ and m, and therefore on β. □

We can reformulate the Removal Lemma: if G_n is a graph with $o(n^h)$ copies of H, we can, by deleting $o(n^2)$ edges of G_n let G'_n be H-free.

The speed as we can let $\beta \to_{n\to\infty} 0$ when we let $\gamma \to_{n\to\infty} 0$ is remarkably slow as every time that we slightly change the ϵ in the Regularity Lemma, we would get large variations on M.

Now we make explicit the reciprocal statement of the Removal Lemma as, for some applications, it will be useful.

Corollary 3.4 (Reciprocal: Removal Lemma for Graphs). *Let G_n be a graph with n vertices. If G has at least $O(n^2)$ edge-disjoint copies of H, then the total number of copies is $O(n^h)$.*

Proof. If G have $O(n^2)$ edge-disjoint copies of H we need to remove, at least $O(n^2)$ edges of G (at least one per edge-disjoint copy) to make G H-free. Therefore the total number of copies should be $O(n^h)$: because if there were asymptotically-less copies of H, this is $o(n^h)$, using the Corollary 3.3 we would be able to find a set of edges with size $o(n^2)$ such that, by deleting this set we would make G H-free, but this is not case as we need $O(n^2)$ edges to be removed. □

3. The Removal Lemma: directed and colored cases

Once proved the Removal Lemma for the undirected graph case, the other two cases are proved similarly but by adjusting the constants and the argument: we will proof the directed and colored case and the monocolored direct case will follow. First we state both lemmas:

Corollary 3.5 (Removal Lemma for Directed Graphs, [**1**]). *For every $\beta > 0$. Let $G = G_n$ be a digraph of order n and a subgraph $H \subset G$, if there is a $\gamma = \gamma(\beta, H) > 0$ such that G_n is a graph with at most γn^h copies of H, then by deleting at most βn^2 edges one can make G_n H-free.*

Note : The relation between β and γ is explicit and if we have a collection of graphs with $o(n^h)$ H-subgraphs, then we can let $\beta \to_{n\to\infty} 0$, so we can delete $o(n^2)$ edges.

And the colored version:

Corollary 3.6 (Removal Lemma for Directed and Colored Graphs). *For every $\beta > 0$ and set of colors $\mathcal{L}$. Let $G = G_n$ be a digraph with n vertices, with colors in $\mathcal{L}$. Let H be a colored digraph with h vertices. If there is a $\gamma = \gamma(\beta, H, |\mathcal{L}|) > 0$ such that G_n is a graph with at most γn^h copies of H, then by deleting at most βn^2 edges one can make G_n H-free.*

Note : The relation between β and γ is explicit and if we have a collection of graphs with $o(n^h)$ H-subgraphs, then we can let $\beta \to_{n\to\infty} 0$, so we can delete $o(n^2)$ edges.

Proof of the colored case. Let $\beta > 0$ be a constant: the proportion of edges we want to remove at most. Let d be a density, let $\epsilon > 0$ be a constant, let G be the graph an H the subgraph, let $\mathcal{L}$ be the color set. Let ϵ_0 be the one defined in the Counting Lemma 3.1 and such that fulfills the hypothesis for ϵ_0 in the Counting Lemma. Let $l = 3|\mathcal{L}|$. Let m be the minimum number of sets in the regular partition.

Apply the Regularity Lemma for directed and edge–colored graphs (Lemma 2.4) to G and let $\mathcal{P} = \{V_0, V_1, \ldots, V_k\}$ be the ϵ-regular partition that outputs the lemma, also let DMM be the upper bound for k. With this we can find the Reduced Graph: let $R = R(G, d, \epsilon, \mathcal{L})$ be the reduced graph of G with the partition $\mathcal{P}$. Let $p = |V_1|$ be the size of the clusters.

Remark : As $m \leq k \leq DMM$ we have that $kp \leq n$ and $p \geq n/DMM$

If we let $\gamma = \left(\frac{\epsilon}{DMM\sqrt[h]{h!}}\right)^h$ we can be sure that there would be no “monomorphism” $incr_H$ from H to R since, if there were one, we would get more than γn^h copies of H inside G, contradicting the hypothesis. That is because if $H \to R$ then we can apply the Counting Lemma and have:

$$|H \subset G| > \left(\frac{\epsilon_0}{\sqrt[h]{h!}}p\right)^h \geq \left(\frac{\epsilon}{\sqrt[h]{h!}}p\right)^h \geq \left(\frac{\epsilon}{\sqrt[h]{h!}}\frac{n}{DMM}\right)^h = \gamma n^h.$$

Now we just have to find that we can, by deleting at most βn^2, find a subgraph of G such that: its reduced graph R' will be a subgraph of R, hence, we would not get that $H \to R'$.

Take G and the partition $\mathcal{P}$.

- Remove the colored edges that touch V_0: if we suppose the vertices in V_0 have maximum degree then $\leq 2|V_0|n \leq 2\epsilon n^2$ edges.
- Remove the colored edges inside each V_i, $i \neq 0$: if we suppose we have the complete graph $K_{|V_i|}$ then all those edges are $\leq kp^2 \leq \frac{n^2}{k}$.

- Remove the colored edges that are between two non-ϵ-regular pairs: if we suppose there are all the edges in that pair $\leq \epsilon k^2 p^2 \leq \epsilon (kp)^2 \leq \epsilon n^2$.
- Remove the colored edges that, although they are between an ϵ-regular pair, they compute a density less than d (with the given color α^*): we can suppose all the colors have density less than d, so the total number of edges $\leq ldp^2\binom{k}{2} \leq ldn^2/2$.

Lets name this graph G'. Lets continue to build the Reduced Graph doing the quotient and place an edge labeled α^* whenever we have an ϵ-regular pair and a density of edges colored with α^* larger than d. We could find that some pairs that where ϵ-regular now they are ϵ-irregular because we have sets of edges with non-empty intersections: when we have deleted an $\overline{\alpha}$ color because we have not reach the density d we have, maybe, deleted some edges from $\overleftarrow{\alpha}$ necessary for the ϵ-regularity of that pair of clusters. But we can sure that the resultant reduced graph R' is a subgraph of R, hence we could not have $H \to R'$.

Lets omit the pairs that are not ϵ-regular after deleting some edges because of the lack of density. If we place edges in the Reduced Graph R' between pairs of vertices whenever the color still have more density than d and the pair was originally ϵ-regular we would get, another time, a subgraph of R, say R'', $R'' \subset R$.

In both ways we cannot use the remaining edges to build copies of H in G', since we cannot use edges with colors α^* with more density than d and between ϵ-regular pairs to build copies of H inside G, hence we have deleted all the γn^h copies of H.

Now we should fulfill all the hypothesis of the Counting Lemma 3.1 and delete less than βn^2 edges.

So we have deleted at most:

$$2\epsilon n^2 + \frac{n^2}{k} + \epsilon n^2 + \frac{ldn^2}{2} = \left(3\epsilon + \frac{1}{k} + \frac{ld}{2}\right) n^2$$

If we choose in the Regularity Lemma $m = 1/\epsilon$ (the minimum number of clusters), $d = \beta/l$ and choose ϵ with $\beta/5 \geq \epsilon$ and such that $\epsilon \leq (d-\epsilon)^h/(2+lh) = \epsilon_0$ we will be able to apply the Counting Lemma. By letting $\epsilon = \left(\frac{\beta}{5l^2}\right)^h$ is enough. With this we have removed less than βn^2 edges. To find this $\epsilon = \left(\frac{\beta}{5l^2}\right)^h$ we can proceed with the same strategy used in the undirected graph case (see the proof of the Lemma 3.3).

Thus we have $\gamma = \left(\frac{\left(\frac{\beta}{5l^2}\right)}{DMM\sqrt[h]{h!}}\right)^h$.

We have proved the claim because DMM is just dependent on ϵ and m, and therefore, by the election of ϵ, on β ($m = 1/\epsilon$ in our case). □

If we have that our number of copies of H is $o(n^h)$, as $n \to \infty$ (this is $\gamma \to_{n\to\infty} 0$), we are allowed to reduce our threshold γ, and, therefore our β; in fact, we can let $\beta \to_{n\to\infty} 0$.

Remarks :

- By letting $l = 3$ we have also proved the Collorary 3.5: the Removal Lemma for digraphs.
- The same arguments for the undirected graph case can be applied here, so we can reformulate both Removal Lemmas by saying that: if G_n is a colored digraph of order n and we have $o(n^h)$ copies of a colored subdigraph H in G_n, then we can remove $o(n^2)$ edges so to make G_n H-free.
- Although we will not use them in this work, the reciprocals of both lemmas are also worth a mention. They say the same as in the undirected graph version: if we have $O(n^2)$ colored diedge-disjoint copies of H then we should have a total of $O(n^h)$ copies of H inside G.

We will use those corollaries, both the undirected an the directed cases, extensively in the following chapters.

CHAPTER 4

Classical Applications

The early applications of the Regularity Method where in Combinatorial Number Theory and Extremal Graph Theory. In this chapter we discuss two of the main achievements which historically motivated the interest in the method. The first one is a proof of the Erdős–Stone Theorem. Although the Erdős–Stone Theorem was obtained two decades before the SzRL, it is generally acknowledged that the proof using the SzRL is far more transparent and illustrative than the other proofs known.

The second classical example is the so–called $(6,3)$–problem, which was connected with the original motivation of the SzRL. The $(6,3)$ Theorem of Ruzsa and Szemerédi provides a quite simple combinatorial proof of Roth's Theorem which states that a set of integers with positive density contains 3–term arithmetic progressions. We include this simple proof here as well. It was shown by Varnavides [**27**] that in fact, the number of 3–term arithmetic progressions is, in order of magnitude, as large as it can be. Namely, for a set A of positive density and N large enough, the set $A \cap [1,\dots,N]$ contains $O(N^2)$ three–term arithmetic progressions. This last result can be also derived from the Removal Lemma.

1. The Erdős-Stone Theorem

Extremal Graph Theory is generally concerned with evaluating the maximum edge density of an H–free graph. The first result of this type is the Theorem of Turán which considers the case $H = K_p$. The Erdős–Stone Theorem is a far reaching statement which shows that the maximal number of edges in an H–free graph depends essentially on the chromatic number of H.

1.1. The Theorem of Turán. In this section we present the well–known Turán's Theorem from 1941. This theorem was the starting point of Extremal Graph Theory: Turán asked about the maximum number of edges a graph with n vertices can have without containing a complete graph with p vertices, K_p, as a subgraph.

This theorem, along with the SzRL and its applications, will allow us to give a proof for the Erdös–Stone Theorem in a way that illustrates how the SzRL is used in extremal problems.

THEOREM 4.1 (Turán, 1941, [**25**]). *Let G be a graph with n vertices without K_p as a subgraph. Then*

$$E(G) \leq \left(1 - \frac{1}{p-1}\right) \frac{n^2}{2}$$

Which implies that, if a graph has more than $\left(1 - \frac{1}{p-1}\right) \frac{n^2}{2}$ edges, then it should have $K_p \subset G$.

The bound given by Theorem 4.1 is tight. The extremal examples are the so–called Turán graphs which are complete multipartite graphs with $p - 1$ clusters of equal size (or almost equal size).

1.2. The Erdös-Stone with the Regularity Method: an example of use. The proof of the Theorem of Erdős-Stone follows one of the most typical strategies involving the use of the SzRL. First we use the Regularity Lemma in order to have information about the graph: this usually is done by computing the reduced graph. Apply a classical theorem for graphs on R, in this case the Turán Theorem, and then try to get back to the original graph with the new information.

We introduce some notation. Let $\mathcal{H}$ be a family of graphs: $ex(n, \mathcal{H})$ denotes the upper bound for the number of edges a graph of order n can have without containing none of the graphs $H \in \mathcal{H}$ as subgraphs. If the number of edges in an $\mathcal{H}$–free graph attains the bound $ex(n, \mathcal{H})$, we say that it is extremal for $\mathcal{H}$. The notation $K_p(t, \ldots, t)$ stands for the graph where there is a cluster with t vertices for every vertex in K_p and we put the edges of the subgraph $K_{t,t}$ instead of a simple edge between each pair of clusters. The original form of the theorem is as follows.

THEOREM 4.2 (Erdős-Stone, 1946, [**9**]). *For every $p \geq 2$ and $t \geq 1$,*

$$ex(n, K_p(t, \ldots, t) = \left(1 - \frac{1}{p-1}\right) \binom{n}{2} + o(n^2).$$

In fact, we will prove that, for every $p \geq 2$ and $t \geq 1$, every sufficiently large graph G of order n such that

$$|E(G)| \geq \left(1 - \frac{1}{p-1}\right) \binom{n}{2} + \gamma n^2.$$

for some fixed $\gamma > 0$, contains $K_p(t, \ldots, t)$ as a subgraph.

PROOF. The idea of the proof is the following: if we can have, in the reduced graph R of some regular partition of G, enough edges to be sure that $K_p \subset R$ then, by using the Counting Lemma, we will be sure to find $K_p(t, \ldots, t)$ in G.

We first select a density d an $\epsilon > 0$ and the minimal number m of clusters. We apply the SzRL and find an ϵ-regular partition $\mathcal{P} = \{V_0, V_1, \ldots, V_k\}$ with V_0 the exceptional set. We should find the reduced graph and count how many edges it has. Let $l = |V_1|$ be the common size of each part of the partition. Then do the following: find the reduced graph R and count the edges.

- Delete the edges inside V_0. There are at most $\frac{1}{2}\epsilon^2 n^2$ of them.
- Delete the edges that connect V_0 with the other vertices, at most ϵnkl.
- Delete the edges inside each cluster V_i. There are at most $\binom{l}{2}k \leq \frac{1}{2}kl^2$ of them.
- Delete the edges between ϵ-irregular pairs. As there are no more than ϵk^2 irregular pairs we delete at most $\leq \epsilon k^2 l^2$ of them.
- Delete the edges between pairs with edge-density lower than d. This results in at most $\leq \binom{k}{2} dl^2 \leq \frac{1}{2}k^2 dl^2$ deletions.

We thus obtain the reduced graph. As we want to count the number of edges in R, we should notice that every edge in R corresponds to at most l^2 edges in G since we are switching an entire pair for a single edge. Hence

$$|E(G)| \leq \frac{1}{2}\epsilon^2 n^2 + \epsilon nkl + \frac{1}{2}kl^2 + \epsilon k^2 l^2 + \frac{1}{2}k^2 dl^2 + |E(R)| \cdot l^2$$

Note that, since $m \leq k$ and there is an exceptional set, we have $kl \leq n$.

As we want to estimate a lower bound for $|E(R)|$:

$$\begin{aligned}
|E(R)| &\geq \frac{1}{l^2}\left(|E(G)| - \frac{1}{2}\epsilon^2 n^2 - \epsilon nkl - \frac{1}{2}kl^2 - \epsilon k^2 l^2 - \frac{1}{2}k^2 dl^2\right) \\
&\geq \frac{1}{2}k^2\left(\frac{|E(G)| - \frac{1}{2}\epsilon^2 n^2 - \epsilon nkl - \frac{1}{2}kl^2 - \epsilon k^2 l^2 - \frac{1}{2}k^2 dl^2}{\frac{1}{2}k^2 l^2}\right) \\
&\geq \frac{1}{2}k^2\left(\frac{|E(G)| - \frac{1}{2}\epsilon^2 n^2 - \epsilon nkl}{\frac{1}{2}k^2 l^2} - \frac{1}{k} - 2\epsilon - d\right) \\
&\geq \frac{1}{2}k^2\left(\frac{|E(G)| - \frac{1}{2}\epsilon^2 n^2 - \epsilon nkl}{\frac{1}{2}n^2} - \frac{1}{k} - 2\epsilon - d\right) \\
&\geq \frac{1}{2}k^2\left(\frac{|E(G)|}{\frac{1}{2}n^2} - \epsilon^2 - 2\epsilon\frac{nkl}{n^2} - \frac{1}{k} - 2\epsilon - d\right) \\
&\geq \frac{1}{2}k^2\left(\frac{|E(G)|}{\frac{1}{2}n^2} - \epsilon^2 - 2\epsilon - \frac{1}{k} - 2\epsilon - d\right) \\
&\geq \frac{1}{2}k^2\left(\frac{\left(1 - \frac{1}{p-1}\right)\binom{n}{2} + \gamma n^2}{\frac{1}{2}n^2} - \epsilon^2 - 4\epsilon - \frac{1}{m} - d\right)
\end{aligned}$$

$$\geq \frac{1}{2}k^2 \left(\frac{\left(1-\frac{1}{p-1}\right) n^2/2}{\frac{1}{2}n^2} - \left(\frac{p-2}{p-1}\right)\frac{1}{n} + 2\gamma - \epsilon^2 - 4\epsilon - \frac{1}{m} - d \right)$$
$$\geq \frac{1}{2}k^2 \left(\left(1-\frac{1}{p-1}\right) - \frac{1}{n} + 2\gamma - \epsilon^2 - 4\epsilon - \frac{1}{m} - d \right)$$

Therefore, if we have

$$|E(R)| > \frac{1}{2}k^2\left(1-\frac{1}{p-1}\right),$$

the Turán bound for K_p, then R contains K_p as a subgraph. In order to blow–up this copy to $K_p(t,\dots,t)$ we have to check that the conditions of the Counting Lemma hold. We should manage to get:

$$-\frac{1}{n} + 2\gamma - \epsilon^2 - 4\epsilon - \frac{1}{m} - d > 0$$

By letting $d = \gamma/3$, $m > 3/\gamma$ and $\epsilon \leq (\gamma/16)^{pt}$, for example, for an n large enough we will be able to meet the requirements. Also, with this choice we are allowed to use the Counting Lemma because $\epsilon_0 = (d-\epsilon)^{pt}/(2+pt)$ is greater than ϵ and for an n large enough we will be able to find enough room for an ϵ portion of $l = |V_1|$ to be greater than t, since we know there will be no more than t vertices per V_i.

Once we have a copy of K_p inside R we proceed into building the homomorphism $incr_{K_p(t,\dots,t)}$: $K_p(t,\dots,t) \to R$ in a natural way: send the first cluster of t vertices where the first vertex of K_p goes, and so on for the rest of the clusters of $K_p(t,\dots,t)$. By the preceding comments, if n is large enough we would be able to apply the Counting Lemma and, hence, assure a copy of $K_p(t,\dots,t)$ in G_n. □

One of the important points in the above proof was to be sure that the vast majority of the edges where outside the clusters: this is done by increasing m. Once this is made we can be sure to delete few edges by choosing a small enough d and ϵ but ensuring that the difference is large enough so the ϵ-error do not damages the relatively many edges that d gives to us (at this stage the ϵ_0 appears). Finally, we just need to choose n large enough so we can be sure to have some vertices inside the clusters to be susceptible to be chosen as vertices of the $K_p(t,\dots,t)$ as a subgraph of G_n.

Now we know that $ex(n, K_p(t,\dots,t))$ should be such that

$$ex(n, K_p(t,\dots,t)) \leq \left(1-\frac{1}{p-1}\right)\binom{n}{2} + o(n^2)$$

as the (small but fixed) γ cannot exist.

On the other hand, the above inequality cannot be strict since we know, by Turán's Theorem that $ex(n, K_p(1, \ldots, 1)) > \left(1 - \frac{1}{p-1}\right)\binom{n}{2}$. Hence equality must hold, proving the theorem.

An important consequence of the Erdős–Stone Theorem pointed out by Erdős and Simonovits in [**8**] is the following. Since every p–colorable graph is a subgraph of $K_p(t, \ldots, t)$ for large enough t, if we have a finite family of graphs $\mathcal{L}$ with a minimal chromatic number $\chi(\mathcal{L}) = \min\{\chi(L) : L \in \mathcal{L}\}$ then

$$ex(n, \mathcal{L}) = \left(1 - \frac{1}{\chi(\mathcal{L}) - 1}\right)\binom{n}{2} + o(n^2),$$

as we should not have $K_p(t, \ldots, t)$ as a subgraph, and we need to exclude $K_p(t, \ldots, t)$ because excluding $K_{p-1}(t, \ldots, t)$ is not enough. This is the usual modern formulation of the Erdős–Stone Theorem.

2. The (6,3)–Theorem of Ruzsa and Szemerédi

The so-called (6,3)–Theorem was proved in 1976 by Ruzsa and Szemerédi. They used it to prove Roth's Theorem on the existence of 3–term arithmetic progressions in sets of integers with positive density. Their proof uses the original version of the Regularity Lemma just for bipartite graphs (see [**21**]). The proof we give here is slightly different, but uses the same background tool: the Regularity Lemma and the Removal Lemma for triangles.

A *3-uniform hypergraph* is a hypergraph all the hyperedges of size three. With this definition we can state the $(6, 3)$–Theorem.

THEOREM 4.3 (The (6,3)-Theorem, Ruzsa-Szemerédi 1976, [**21**]). *If H_n is a 3-uniform hypergraph on n vertices such that no set of six points contains three or more edges, then $e(H_n) = o(n^2)$.*

PROOF. First we will translate the problem from the 3-uniform hypergraph to a problem on a graph. We will change each hyperedge of H_n by a K_3, a triangle: but we should check and analyze the output graph, finding out how are the triangles we get in the new graph. Let G_n be a graph with the same set of vertices V as the hypergraph H_n. For every 3-hyperedge $\{v, u, w\}$ in H_n we will put the edges vu, uw and vw in G_n, hence forming a K_3.

Claim 1: If two hyperedges in H_n share two vertices, then they are "alone", that is, the vertices in these two hyperdges are not incident with a further hyperedge of H_n .

PROOF. Suppose that h_1 and h_2 are two hyperedges from H_n which share two vertices: in that case, we have already two triangles in four vertices say $\{v_1, v_2, v_3, v_4\}$. If we have some other 3–hyperedge, say h_3 with some v_i with $1 \leq i \leq 4$, say v_1, v_5, v_6, we would have three triangles (or three hyperedges) in six points contradicting our assumption. Hence, the vertices $\{v_1, v_2, v_3, v_4\}$ can only see 2 hyperedges of H_n. □

If we exclude the first case, where two hyperedges share two vertices, all the other triangles that come from an hyperedge are edge-disjoint. Call G' the graph obtained from considering only 3-hyperedges that turned out to be edge–disjoint triangles . Let us suppose now that we can find some H_n with no $(6, 3)$–configuration such that $e(H_n) = \Omega(n^2)$, in this case, as the maximum number of 3-hyperedges that share two vertices with some other hyperedge (an hence, forms two non-edge-disjoint triangles) should be $n/2$ or less, we can be sure there will be still $\Omega(n^2)$ edge–disjoint triangles. As the 3-hyperedged triangles we have excluded are "alone" we can focus on the set of vertices with $\Omega(n^2)$ edge-disjoint triangles.

Now we can apply the Removal Lemma to the present context to deduce that the total amount of triangles in $G' \subset G$ will be $\Omega(n^3)$. Indeed, if G' contains $o(n^3)$ triangles then we could remove them by deleting $o(n^2)$ edges. However, since G' contains $\Omega(n^2)$ edge–disjoint triangles, this cannot be the case. Thus we should find some new triangles in the graph, different from the ones that come from 3-hyperedges.

Let T be one of these new triangles. As G and G' have only edges that come from a 3-hyperedge, T only have edges of that type. Also T must have edges from three different triangles, since if T receives two edges from the same edge-disjoint triangle, then it should receive the third one. We will call t_1, t_2 and t_3 the three T-edges and T_1, T_2 and T_3 the respective edge-disjoint triangles they come from.

We claim that this T, more precisely the three edge-disjoint triangles that hold each one of the edges of T, form a configuration where there are six points with three triangles. This is so because if T is built from t_1, t_2 and t_3 then:

- if $t_1 = uv$ then $t_2 = uw$ (or $t_2 = vw$), because they are form T.
- if $t_1 = uv$ and $t_2 = uw$ then $t_3 = vw$ because they should form the T.

So we have that $T_1 = \{u, v, h_1\}$, $T_2 = \{u, w, h_2\}$ and $T_3 = \{v, w, h_3\}$. But now we see that T_1, T_2 and T_3 are three triangles on six points: $\{u, v, w, h_1, h_2, h_3\}$; this is a contradiction with the fact that H_n has no such configuration, because G' has only edges that comes from 3-hyperedges in H_n. So the number of hyperedges in H_n should be no more than $o(n^2)$. □

3. The Roth's Theorem

Roth [**20**] proved in 1953 that a set of integers in $[N]$ which contains no 3–term arithmetic progressions must have cardinality $o(N)$. The original proof uses harmonic analysis. Although the proof using the SzRL through the Removal Lemma give us a worse bound than the original one (see [**20**] or [**24**] and the better bound found by Bourgain in 1999, see [**6**]), it reflects the wide ranges where the Regularity Lemma can be applied and the simple combinatorial arguments that can be derived from it.

Let $r_3(N)$ denote the size of the maximum subset of $[N]$ that does not contain any three elements in non-trivial arithmetic progression.

Theorem 4.4 (Roth, 1953, [**20**]). *The function $r_3(N)$ is $o(N)$.*

Proof. Let $A \subset [N]$ be a set with size $r_3(N)$ such that there is no arithmetic progression in A, so a maximal one. We want to see that $|A|$ must be $o(N)$. For this purpose we will build a graph G. Take three sets of disjoint vertices: $V_1 = [N]$, $V_2 = [2N]$ and $V_3 = [3N]$. Let the set of vertices V of G be $V(G) = V_1 \cup V_2 \cup V_3$. To construct the edges in G we will use the set A.

For every element g_1 with $g_1 \in V_1$ we will connect it to $g_1 + A \subset V_2$. That is, if there exists a $g_2 \in V_2$ such that $g_1 + a_i = g_2$ for some $a_i \in A$ then connect g_1 and g_2 by an edge. Do the same between V_2 and V_3: connect every element $g_2 \in V_2$ to $g_2 + A \subset V_3$. Let $2 \cdot A$ denote the set of integers such that every element is twice an element of A. Finally connect V_1 with V_3 using $2 \cdot A$: that is, connect every element $g_1 \in V_1$ with $g_1 + 2A \subset V_3$. We would have a triangle in G if and only if we have three numbers in arithmetic progression, since if x, y and z are in arithmetic progression, with $x \leq y \leq z$, then $x + z = 2y$. For any given 3–term arithmetic progression we have N copies of that arithmetic progression in the graph: one for each $g_1 \in V_1$. For every vertex $g_1 \in V_1$ we will get $|A|$ edge-disjoint triangles with g_1 as a vertex since we have $|A|$ trivial 3–term arithmetic progressions (the ones with zero difference), that gives a total of $|A|N$ edge-disjoint triangles. Also we have that every edge is from one, and only one, of this edge-disjoint triangles. Notice that the total amount of vertices is $6N = O(N)$.

Now we can use the $(6, 3)$–Theorem, or the Removal Lemma, to know that we can only have a maximum of $o(N^2)$ edge-disjoint triangles if we don't want them to generate another triangle: the non-trivial arithmetic progression. Hence the size of A should be no more than $o(N)$. □

3.1. Lower bounds on the $(6, 3)$–problem. We have seen that if we have $O(N^2)$ edge-disjoint triangles or, similarly, $O(N)$ integers in $[N]$, we would get either a new triangle or a non-trivial arithmetic progression. This is: we have found upper bound for the $r_3(N)$ function or, for the maximum number of edge-disjoint triangles we can have without three of them building a new one. But we can ask about lower bounds on those functions; this is: how many edge-disjoint triangles can we have in a graph without creating a new one?.

In 1946 Behrend (see [**3**]) found a construction, for every N, of a quite dense set without 3–term arithmetic progressions. As we can build the graph with a similar construction as the one done in the proof of Roth's theorem, there we will find no more triangles than the edge-disjoint original ones that come from the trivial 3–arithmetic progressions. So we will find a lower bound on the number of edge-disjoint triangles, as Ruzsa and Szemerédi did in [**21**] for the $(6,3)$–problem.

Behrend's construction assures that we can find sets A without 3–term arithmetic progressions of size, asymptotically:

$$|A| > N^{1-\frac{2\sqrt{2\log 2}+\epsilon}{\sqrt{\log N}}} = b(N)$$

for every $\epsilon > 0$.

This means that we can found $b(N)N$ edge-disjoint triangles in $6N$ vertices without the need of building a new non-edge-disjoint triangle.

3.2. Varnavides' Theorem. A few years after Roth published his result (see [**20**]), Varnavides proved that, in a set of integers with positive density, there should be not only one 3–term arithmetic progression but many: in fact, $O(N^2)$.

Theorem 4.5 (Varnavides, 1959 [**27**]). *Let δ be a number satisfying $0 < \delta < 1$, and let $a_1, a_2, \ldots, a_m$ be any set of distinct positive integers not exceeding x. Suppose that*

$$m > \delta x \quad \text{and } x > x_0(\delta)$$

where $x_0(\delta)$ depends only on δ. Then the number of solutions of

$$a_i + a_j = 2a_h \quad (i \neq j)$$

is at least $C(\delta)x^2$, where $C(\delta)$ is a positive number depending only on δ.

Proof. We can prove this theorem by using the same proof as for the Roth's Theorem but, once in the last step, use the Removal Lemma to say that one would get $O(N^3)$ triangles in total, hence $O(N^3)$ triangles more than the original edge–disjoint ones. This means that there should be $O(N^2)$ new 3-arithmetic progressions as the Varnavides Theorem says.

This also means that there should be some distance d with at least $O(N)$ 3-arithmetic progressions with this common difference d. ☐

CHAPTER 5

The Removal Lemma for groups

Green presents in [**12**] an algebraic version of the Regularity method for abelian groups. The main feature of this algebraic version is the fact that the clusters of ϵ–regular partitions are sets close to subgroups.

One of the highlights of the SzRL for groups is a Removal Lemma which can be stated in terms of the number of solutions of a linear equation in the group: if the equation $a_1 + \cdots + a_k = 0$ has $o(|G|^{k-1})$ solutions in a subset $A \subset G$, then we can remove $o(|G|)$ elements in A such that all solutions are eliminated. Green derives this version of the Removal Lemma from the more general SzRL for groups. This in turn is proved by heavy use of the machinery of Fourier Analysis, and as such, it is limited to abelian groups. In Section 1 we give a general statement of the Removal Lemma for groups which is valid in an arbitrary finite group. The proof relies on the directed version of the SzRL, and, from that lemma, is considerably simpler than the derivation from the SzRL for groups. Moreover it essentially requires only a finite algebraic structure with a cancelation law, although we state it just for finite groups. The result is the object of a preprint in collaboration with Daniel Kral.

The more general edge colored digraph version of the SzRL allows us to extend the Removal Lemma to a class of linear systems. This result is presented in Section 2.

In Section 3 we show an application of this Removal Lemma for groups that can also can be found in [**12**] which concerns sum-free sets: sets in which no element in the set that can be written as the sum of other two elements in the set.

We close the chapter with a discussion of future work and open problems which arose during the preparation of this work.

1. The Removal Lemma for groups

In 2004 Ben Green proved a Removal Lemma-like theorem referring to the abelian groups. More precisely:

Let G be a finite abelian group with cardinal N. Let A be a subset of G. A triple $(x, y, z) \in A^3$ will be called a triangle if $x + y + z = 0$.

THEOREM 5.1 (Green 2004, [**12**]). *Suppose that $A \subseteq G$ is a set with $o(N^2)$ triangles. Then we may remove $o(N)$ elements from A to leave a set which is triangle-free.*

In fact, he proved the more general one:

THEOREM 5.2 (Green 2004, [**12**]). *Let $k \geq 3$ be a fixed integer. Let G be a finite abelian group with cardinality N and suppose that $A_1, \ldots, A_k$ are subsets of G such that there are $o(N^{k-1})$ solutions to the equation $a_1 + \ldots + a_k = 0$ with $a_i \in A_i$ for all i. Then we may remove $o(N)$ elements from each A_i so as to leave sets A'_i, such that there are no solutions to $a'_1 + \ldots + a'_k = 0$ with $a'_i \in A'_i$ for all i.*

Green proved this theorem as an application of a SzRL theorem for groups. The proof relies heavily on Fourier Analysis (Harmonic Analysis). This technics allow the author to show that the clusters in the Regularity Lemma can be found to structures with subgroup reminiscence. As it use extensively Fourier Analysis the proof should be restricted to abelian groups.

We present another proof that allows us to extend Theorem 5.2 to arbitrary finite groups, not just abelian ones, as stated in Theorem 5.3 below. In Section 2 we will discuss extending this proof to other structures.

THEOREM 5.3. *Let k be an integer with $k \geq 3$. Let G be a finite group of order N. let $A_1, \ldots, A_k$ be subsets of G and $g \in G$. Suppose that the equation $x_1 x_2 \cdots x_k = g$ has $o(N^{k-1})$ solutions with $x_i \in A_i$, $1 \leq i \leq k$. Then there are subsets $A'_1 \subset A_1, \ldots, A'_k \subset A_k$ verifying $|A_i \setminus A'_i| = o(n)$ such that there is no solution of the equation $x_1 x_2 \cdots x_k = g$ with $x_i \in A'_i$ for all i.*

Remark : We will prove Theorem 5.3 for the homogeneous case $g = 1$. The general case can be easily handled just by letting $A'_k := A_k g^{-1}$: thus the new equation equals 1 and we can get back to the new set A_k by multiplying by g to the corresponding set.

PROOF. We will build a digraph and then work with it, as our intention is to use the Removal Lemma for digraphs (Corollary 3.5). Take k copies of the group G, say $G_1, G_2, \ldots, G_k$ and consider the graph $\mathcal{G}$ with the set of vertices $V = \{G_1, G_2, \ldots, G_k\}$. There will be no edges inside G_i; just between them.

For all $i \neq k$ connect each element (vertex) of $g_i \in G_i$ to $g_i a_l \in G_{i+1}$, where $a_l \in A_i$, with the directed edge $(g_i, g_i a_l)$. In case we have $i = k$ connect $g_k \in G_k$ to $g_k a_l \in G_1$ where $a_l \in A_k$ with the edge $(g_k, g_k a_l)$. All those $(g_j, g_j a_l)$, for j running over the set of clusters and l over the elements of the corresponding A_j will be the set of edges, $E(\mathcal{G})$. Label each arc with the $a_{m,l}$ that build it.

Claim : Whenever we have a k-directed-cycle and with all the arc with the "same way" (from here on a k–cycle), it is because we have found a solution to the equation $a_1 a_2 \dots a_k = 1$, also it is right in the other way: whenever we have a k-tuple $(a_{1,0}, a_{2,0}, \dots a_{k,0})$ in $\prod_{i\in[k]} A_i$ such that $a_{1,0}a_{2,0}\dots a_{k,0} = 1$, it forms a k–cycle in $\mathcal{G}$. In fact, for every k-tuple in G, such equation outputs N k–cycles in $\mathcal{G}$, one for every vertex in G_1 (this is so because we are considering k-tuples, and no k-sets, but in the abelian case the difference is just a $\leq k$-times factor).

To prove the claim: suppose we have a k–cycle in the graph $\mathcal{G}$ with vertex g_1. This would mean that there exists a g_2 and an edge from some $a_1 \in A_1$ such that $g_1 a_1 = g_2$; the same is true for $g_i a_i = g_{i+1}$, with $g_i \in G_i$, $g_{i+1} \in G_{i+1}$ and $a_i \in A_i$ for all $i < k$. If it should be a k–cycle, there exists a $a_k \in A_k$ such that $g_k a_k = g_1$, the initial element. We have then that:

$$g_k a_k = g_1 \;\Rightarrow\; g_{k-1}a_{k-1}a_k = g_1 \;\Rightarrow\; \dots \;\Rightarrow\; g_1 a_1 \dots a_{k-1}a_k = g_1 \;\Rightarrow\; a_1 a_2 \dots a_{k-1}a_k = 1.$$

The other way is also true because if we fix $g_1 \in G_1$ (or any $g_i \in G_i$) and if we choose to "travel" between G_j always by a k–tuple of elements whose product is 1 it means that we will end up in the same g_1 using k edges, and so a k–cycle is formed in the graph $\mathcal{G}$.

Now paying attention to the fact that, for every fixed vertex, there are as many k-cycles that go through that vertex as the number of k–tuples whose product is 1 we can conclude that there are $o(N^k)$ in total (using the hypothesis of the theorem): fixing the cluster G_1, for every element $g_i \in G_1$ there are $o(N^{k-1})$ k–cycles, and so this makes a total of $No(N^{k-1}) = o(N^k)$ k–cycles in the graph (not much and also no less k–cycles).

Remark : If $k \equiv 0 \mod 2$ we can find other k-directed-cycles but they will be cycles with 2 exceptional vertices: from one there will *leave* 2 di-edges (out-degree 2) and another will *receive* other 2 (in-degree 2), and those are not the k–cycles we are considering: we are considering the k–cycles that have all the arcs in the same way.

So we have a digraph, $\mathcal{G}$, with $o(N^k)$ k–cycles in it: therefore we can apply the Removal Lemma (Corollary 3.5) over digraphs to be sure that, by deleting at most $o(N^2)$ edges, we have made $\mathcal{G}$ k–cycle-free. Let us call E_k this set of edges.

With this information we will analyze a bit more the graph. If we pay attention to a single k–tuple whose product is 1 it forms N edge-disjoint k–cycles (in-between the N k-cycles), but with edges that have the same label between a fixed pair of clusters (G_j, G_{j+1}). So we know for sure that, if we want to delete all those k–cycles we need to pick, at least, N edges (one for every edge-disjoint-k–cycle) from the set of edges E that will be in E_k to erase all the k-cycles in the graph.

By the pigeonhole argument we can be sure that, if we look at the labels of the edges (for a single k–tuple), there is one label of the k–tuple elements that gets more than N/k edges (what is important, $O(N)$ edges). We will choose to delete this element from the correspondent set.

By doing so we have deleted all the edges with that label, and also we have thrown those N k–cycles away.

Now we continue to do this: check for a k–tuple that still (we may have deleted many already) is a solution of the considered equation: that corresponds to N k–cycles in $\mathcal{G}$; look in E_k the edges that are in the N k–cycles associated to the k–tuple to be deleted; check which one is the most popular label (we know we should have, at least, N/k) and choose to delete this in the corresponding set A_i.

By doing so we can delete all the k–tuples which are solutions of our equation in G and also be sure that at most $o(N)$ elements in A_i are deleted. The reason is because as we choose to delete the most popular label (in the k–tuple and from E_k) we can also get rid of those edges from this set. So each time we choose one element in A_i to be deleted also we can delete $O(N)$ edges in the k-cycle-removing deleting set E_k: this process should be done no more than $o(N)$ times, because E_k is $o(N^2)$ and we delete $O(N)$ different edges (with the same label but different *parallel* edges in $\mathcal{G}$) each time. Thus we are done.

The important fact is that we can associate at least N different edges that needed to be deleted to erase the k-tuple with a constant number of elements to delete (namely k): hence with a constant number of steps we can delete $O(N)$ edges. ☐

It is important to notice that the freedom in choosing A_i in Theorem 5.3 allows us to deal with quite a general family of equations in the group. The important hypothesis is contained in the $o(n^{k-1})$ number of solutions. The following corollary is an example of the possible range of applications.

Corollary 5.4. *Let $k \geq 3$ be an integer. Let G be a finite group of order N and $A \subset G$. Let $\pi_1, \ldots, \pi_k$ be arbitrary permutations of the elements of G. Suppose that the equation*

$$x_1^{\pi_1} x_2^{\pi_2} \cdots x_k^{\pi_k} = g$$

has $o(N^{k-1})$ solutions in A. Then there is a subset $A' \subset A$ with $A' = o(N)$ such that there are no solutions of the equation in $A \setminus A'$.

The statement of the above corollary applies, for instance, to any linear equation in a finite Abelian group whose coefficients are relatively prime with the exponent of the group.

In the proof of Theorem 5.3 we have essentially used only the cancelation property of the group. Thus we could state a similar statement for quasigroups, or latin squares (and even slightly more relaxed structures where the cancelation property holds for all but $o(n)$ elements of G, in the sense that all but a negligible number of elements $g \in G$ verify $xg \neq yg$ whenever $x \neq y$). We have nevertheless restricted ourselves to the case of groups, where the applications coming from Additive Combinatorics are more apparent.

2. Extensions to systems of equations

The directed edge–colored version of the Removal Lemma allows us to extend Theorem 5.3 to a class of systems of equations. In order to make the exposition clearer we will restrict ourselves to the case of linear systems in abelian groups.

Let $A = (a_{ij})$ be a $(0,1)$–matrix of order $k \times m$. We say that A is *nice* if, up to rearranging rows and columns, the following conditions hold:

- There is $k' \leq k$ such that the first row has k' ones in the first positions and $k - k'$ zeros, that is, $a_{1i} = 1$ for $1 \leq i \leq k'$ and $a_{1j} = 0$ for $k' + 1 \leq j \leq k$.
- For each row, if $a_{it} = 1$ for some $t \leq k'$, then $a_{ij} = 1$ for each $j \leq t$.
- For each $j > k'$, the j-th column has exactly one nonzero entry.
- For each $i \geq 2$ there is a nonzero entry a_{ij} in the i-th row for some $j > k'$.

Thus a nice matrix (in canonical form) has two parts, one of them in triangular form, and the second one with vectors of disjoint supports. If A is a nice matrix in canonical form, we call $k'(A)$ the integer for which the above conditions hold. For example, the matrix

$$\begin{pmatrix} 1 & 1 & 1 & 0 & 0 & 0 \\ 1 & 1 & 0 & 1 & 0 & 0 \\ 1 & 1 & 1 & 0 & 1 & 1 \end{pmatrix}$$

is nice, and

$$\begin{pmatrix} 1 & 1 & 1 & 0 & 0 & 0 \\ 1 & 1 & 0 & 1 & 0 & 0 \\ 1 & 0 & 1 & 1 & 1 & 1 \end{pmatrix}$$

is not.

THEOREM 5.5. *Let G be a finite abelian group of order N. Let A be a $k \times m$ nice matrix. Let $B_1, \ldots, B_k \subset G$ If the number of solutions of the linear system $Ax = 0$ with $x = B_1 \times \cdots \times B_k$*

is $o(N^{k-m})$ *then there are subsets* $B'_1 \subset B_1, \dots B'_k \subset B_k$ *such that the equation has no solutions in* $B'_1 \times \cdots \times B'_k$ *and and* $|B_i \setminus B'_i| = o(N)$, $1 \leq i \leq k$.

Proof. Up to rearranging rows and columns we may assume that A is in canonical form. We construct an edge colored digraph H as follows. Take $k-m+1$ copies of G, $G_1, \dots, G_{k-m+1}$. For each $i = 1, \dots, m$, let $J(i) = \{j_1, \dots, j_{r_i}\}$ denote the support of the vector in row i of the matrix A. We connect every element $g \in G_{j_i}$ to the elements $gb \in G_{j_{i+1}}$, by an directed edge colored j_i and labelled b for each $b \in B_{j_i}$ except for $i = r_i$, in which case the terminal vertex is in G_1. We thus construct an edge colored directed cycle for each equation in the system, all pairwise edge disjoint except for the initial edges shared with the cycle corresponding to the first equation. We then identify parallel arcs to obtain the simple edge colored digraph H. Note that all edges incident to some vertex not in G_1 have the same in-color.

Let us fix a solution to the linear system, namely $(b_1, \dots, b_k)$. If we fix $g_1 \in G_1$ and use the solution as a way to travel through the graph, we will get a subgraph with edges having all the possible colors and no two edges with the same color.

Thus, a solution of the equation corresponds to an edge–disjoint union of N such edge colored subgraphs.

Reciprocally, if we have an edge–colored subgraph S of H with exactly one cycle with edges colored j, $j \in J(i)$, and label b_j, for each $i = 1, \dots, m$, then we get a solution $(b_1, \dots, b_k)$ of our system. Note here that we need to consider colored edges, since other isomorphic (uncolored) subgraphs to S in H may not correspond to solutions of the system.

Now, if the system has $o(N^{k-m})$ solutions we can be sure that we have $o(N^{k-m+1})$ subgraphs we want to remove. Since the number of vertices of each subgraph corresponding to a solution is $k-m+1$, we can apply the Removal Lemma in the directed and colored case (Corollary 3.6): we can choose a set E' with $o(N^2)$ colored arcs such that, if we delete them we make G free of this solution-related subgraphs.

To travel from the edge-set to the group we can simply use the same argument as in the one equation case (see proof of Theorem 5.3): as we have N copies of the same solution and they are edge-disjoint, we should remove at least one edge from each subgraph to erase it, hence N edges. As the graph has less than km edges we can select to remove the most popular one: it should get, at least, N/mk hits. We can continue to do this till we have no solutions: since we have $o(N^2)$ edges and we have removed, at each step, at least N/mk edges from the E' set, so we should do this process no more than $o(N)$ times. So, we have deleted $o(N)$ elements in total. At this point we are sure to find subsets in every B_j, with size $o(N)$ so that, once removed, we will have no solutions of $Ax = 0$, proving the theorem. □

Remark : Notice that in this case the edge coloring is important since we can find configurations where, if we do not use colors, we could run into difficulties when applying the Removal Lemma. We will have $o(N^{k-m+1})$ copies of the subgraph we *really* want to delete, but we can have, in the digraph, many unwanted copies of the same graph: even as much as $O(N^{k-m+1})$ and, hence, the Removal Lemma for digraphs does not assures a set of removable edges with size $o(N^2)$ since it will compute also those unwanted subgraphs as ones to be removed. Let us illustrate this by an example. Consider the system of equations:

$$\begin{cases} x_1 + x_2 + x_3 + x_4 \qquad\qquad = 0 \\ x_1 + x_2 \qquad\qquad + x_5 + x_6 = 0 \end{cases}$$

and let $B_1, B_2, \ldots, B_6$ be subsets of linear size $O(N)$. Consider the edge colored directed graph build in the proof of Theorem 5.5. The subgraphs which represent solutions of the system are isomorphic to two directed 4–cycles with two edges in common. However we could find many such subgraphs with edges colored only $\{1, 2, 3, 4\}$ (as solutions of $b_3 + b_4 = b_3' + b_4'$ with $b_3, b_3' \in B_3$ and $b_4, b_4' \in B_4$). Therefore, if we remove the colors we could find $O(N^{k-m+1})$ such subgraphs and we would not be in the conditions of applying the Removal Lemma.

Theorem 5.5 has also the same general feature as Theorem 5.3 which lies in the freedom of choosing the sets B_i. For instance, it can be applied to linear systems with coefficients relatively prime with the exponent of the group and with a matrix whose support is a nice matrix. Also we can state a more general form for nonabelian groups, except that the formulation of the corresponding notion of nice matrix is more involved. In the next section we illustrate some applications of the above results.

3. Applications of the Removal Lemma for groups

In this section we present two applications of the Removal Lemma for groups. The first uses the Theorem 5.3 for one linear equation and the second one uses the Theorem 5.5 for systems of equations.

3.1. Applications to sum-free sets.

3.1. Applications to sum-free sets. Using the Theorem 5.3 (or the Theorem 5.2) we can, by choosing the sets properly, get different theorems. For example:

THEOREM 5.6 (Sum-free sets, [**12**]). *Suppose that $A \subseteq [N-1]$ is a set containing $o(N^2)$ triples with $x + y = z$. Then $A = B \cup C$ where B is sum-free and $|C| = o(N)$.*

So, if the number of collisions is “small” we can delete a few elements in the set so to make no collision at all in the final set.

PROOF. Choose $G = \mathbb{Z}_{2N}$ and choose $A_1 = A$, $A_2 = A$ and $A_3 = -A$ considered as being modulo $2N$. In the way we have chosen G we know we have not generated new Schur triples as if $x \in A$ and $y \in A$ then $x + y \leq 2N$ so both A modulo $2N$ and $A + A$ modulo $2N$ can be mapped to the same set of representatives in the natural way. Thus, if there exists a $z \in A$ such that $x + y \equiv z$ the same equations holds on the integers.

As we are under the hypothesis of $o(N^2)$ triples, we can remove a set with size $o(N)$ (the union of three $o(N)$ sets) such that, once removed, there are no solutions to the equation $x+y-z = 0$ and, hence, we have found a set B that is sum-free. □

Let us mention that there are sets of integers in the interval $[0, N]$ that are sum-free and that have size $O(N)$. For example the elements larger than $N/2$ or any subset of the odd numbers.

3.2. Example of application for systems of equations. Now we will see an example of how Theorem 5.5 can be used.

Let A and B be two subsets of $[1, \ldots, N-1]$. Let S be the system of equations with:

$$\begin{cases} x + y + z = t \\ x + y = 2r \end{cases}$$

where $x, y, z, t \in A$ and $r \in B$.

So we are asking about elements from $2B+A = A$ such that: we add elements from $2B$ whenever the point in B is the middle one from two other points in A. The set A adds also multiplicity to that counting.

We use Theorem 5.5 and the group $G = \mathbb{Z}_{3N}$, instead of the $G = \mathbb{Z}_{2N}$ used in the application to sum-free sets, with the sets $B_1 = A$, $B_2 = A$, $B_3 = A$, $B_4 = -A$ and $B_5 = -2B$. As we are not generating new solutions to the system, we know that if this system of equations has $o(N^{5-2})$ solutions, we will be able to delete $o(N)$ elements from A and $o(N)$ elements of B so that the system S has no solution.

4. Open problems and future work

During the preparation of this work two questions arose which are a natural continuation. One of them is to obtain more precise estimates in the Removal Lemma, at least in the algebraic setting considered here. The second one is a possible application of such refinement in estimating the number of arithmetic progressions with a common difference in a dense set of integers. We next discuss both of them in more detail.

4.1. Combinatorial proof for the Removal Lemma. The proof of the Removal Lemma presented in this work involves the Regularity Lemma, but the proof of the Regularity Lemma is not constructive and gives very large bounds on the number of sets in relation to the ϵ (tower type bounds on $1/\epsilon^5$). By a result of Gowers [**10**] this is, at some point, unavoidable.

One open question is whether a combinatorial proof of the Removal Lemma can be found: the motivation behind this is the large bounds that the Regularity Lemma outputs. If a more direct, combinatorial, proof could be found it would be useful for two reason: the first one is that is plausible to think that this possible proof would give better bounds on the Removal Lemma, and, also, it might give better understanding on why the Removal Lemma works. This question can be found in [**24**] where Tao and Vu quote a paper of Gowers [**11**].

Now we present one approach that we have followed to try to answer this question.

Instead of proving the triangle Removal Lemma itself our intention is to try to find how many edge-disjoint triangles can have a graph G without three of them forming a new one: the formulation of the $(6,3)$–Theorem of Ruzsa and Szemerédi. More precisely, for which orders and number of triangles we can be sure this will happen. We will suppose that every edge in G comes from an edge-disjoint triangle and we have a total of αn^2 edge-disjoint triangles for a fixed α. Also we will restrict, in order to simplify the computations, to the case where every vertex has the same number of edge-disjoint triangles on it: this case is relevant since it corresponds to the case with $k=3$ in the Removal Lemma for groups which allows us to proof the Roth's Theorem (using a similar strategy to the one used to proof the theorem for sum-free sets combined with the proof of the Roth's Theorem proved in this work). Also, although this is not proved, it seems to be the worst or nearly the worst case, see the example of the Turán graph in [**7**].

The main idea is, try to find a big enough $n = |V(G)|$ such that, if a graph has αn^2 edge-disjoint triangles then one should get some additional triangle, not an edge-disjoint one. To do this we will suppose that, for every n, we have no new triangles besides the edge-disjoint ones and try to reach a contradiction. We will also suppose that we have no other edges besides the ones from the edge-disjoint triangles and that the graph G is regular.

In this case we have $3\alpha n$ triangles per vertex, so a degree of $6\alpha n$ per vertex. Let v be a vertex of G. Let $N(v)$ be the neighbourhood of v. Let u_1 and u_2 be two vertices from $N(v)$. As we are supposing there is no new triangle, if there is the edge u_1u_2 then u_1u_2v must be an edge-disjoint triangle, hence v just "see" the edges of edge-disjoint triangles that has v as a vertex: those are $3\alpha n$ edges. Also the set of edges between neighbours of v form a perfect matching within the vertices of $N(v)$.

If we find two vertices v_1, v_2 with $|N(v_1) \cap N(v_2)| > 3\alpha n$ we would have found a new triangle, since, as we can have only $3\alpha n$ triangles per vertex, there would be an edge u_1u_2 from one of these edge-disjoint triangles pivoting over, say, v_2 that will connect two neighbours of v_1 (u_1 and u_2), but without $v_1u_1u_2$ being an edge-disjoint triangle (since $v_2u_1u_2$ was already a triangle!) and, therefore, making $v_1u_1u_2$ a new triangle.

The main idea is to find a needed configuration in the graph where there are two vertices v_1, v_2 with $|N(v_1) \cap N(v_2)| > 3\alpha n$. We will try to build a "small" set of vertices V' where some other vertices, W', would have V' as their neighbourhood. Hence there would be a pair of vertices of W' which share a large portion of their neighbourhood in V'.

We can also view it as there are some edges that are forbidden, since if we had those edges then we would have a triangle. Any edge between two vertices u_1, u_2 at distance 2 produce a forbidden edge, u_1u_2, the same edge can be forbidden many times, but, since the number of paths of distance two between u_1 and u_2 are the same as $|N(u_1) \cap N(u_2)|$ we should have no more than $3\alpha n$ per forbidden edge. Any vertex v forbids $(6\alpha n)(6\alpha n - 2)/2$ edges, since we can have only $3\alpha n$ edges between neighbours of v, this is the same for each vertex so we forbid a total of $n(6\alpha n)(6\alpha n - 2)/2$. Since the total number of edges we do not have is $\binom{n}{2} - 3\alpha n^2$ we can define:

$$\mu := \frac{n(6\alpha n)(3\alpha n - 1)}{\binom{n}{2} - 3\alpha n^2}$$

as the average number of times we forbid an edge. Obviously for small α's we have that this amount is less than $3\alpha n$.

Let $v_1, \ldots, v_k$ be vertices of G. Let $\Delta = N(v_1) \cap \ldots \cap N(v_k)$ be the intersection of the neighbourhoods. Lets suppose $k \geq 2$. If we have $w \in \Delta$ then w can just be connected to $v_1, \ldots, v_k$, the other vertex of the triangles $(\cdot)v_iw$ and to other vertices that cannot be from $\bigcup_{i=1}^{k} N(v_1)$, otherwise we should have some other triangle.

Our main idea is to assure that, for some $v_1, \ldots, v_k$ we have $|\Delta|$ large and also $|\bigcup_{i=1}^{k} N(v_1)|$ large, without forming any new triangle. So we would be able to assure that $N(\Delta)$ has to share not many vertices allowing us to find a new triangle.

We can compute all the $v_1, \ldots, v_k$ neighbourhood intersections: we pick one vertex, say v, and choose k vertices within its neighbours. They will have v (at least) as a common intersection. So we have:

$$n\binom{6\alpha n}{k}$$

k-vertices intersections.

So a first approximation for $\Delta(v_1, \ldots, v_k)$ for a k-set of vertices will be:

$$\frac{n\binom{6\alpha n}{k}}{\binom{n}{k}}.$$

But in $\binom{n}{k}$ we have many sets for whom we know their intersection. We define three classes of vertices, named 1,2 and 3.

- Class 1: the k-sets of vertices for which we know that $|\bigcap_{i=1}^k N(v_1)| = 1$
- Class 2: the k-sets of vertices for which we know that $|\bigcap_{i=1}^k N(v_1)| = 0$
- Class 3: the rest, maybe some will have $\cap 1$, others $\cap 0$ and others $\cap j$.

Lets call the *associates* of Δ to the set of vertices u such that wv_iu is an edge-disjoint triangle of G with $w \in \Delta(v_1, \ldots, v_k)$.

Let $v_1, \ldots, v_k$ be vertices of G, suppose they are from class 2: then for every v_{k+1} we have that $v_1, \ldots, v_k, v_{k+1}$ will be from class 2.

Let $v_1, \ldots, v_k$ be vertices of G, suppose they are from class 1: suppose that $w = \bigcap_{i=1}^k N(v_1)$, then if $v_{k+1} \in N(w) \setminus \{v_1, \ldots, v_k\}$ we have that $v_1, \ldots, v_k, v_{k+1}$ will be from class 1. Otherwise $v_1, \ldots, v_k, v_{k+1}$ is from class 2.

Let $v_1, \ldots, v_k$ be vertices of G, suppose they are from class 3. If v_{k+1} is an associate of $\Delta(v_1, \ldots, v_k)$ then $v_1, \ldots, v_k, v_{k+1}$ is from class 1. If v_{k+1} is from the rest of $\bigcup_{i=1}^k N(v_1)$ then $v_1, \ldots, v_k, v_{k+1}$ is from class 2. Otherwise $v_1, \ldots, v_k, v_{k+1}$ remains to class 3.

We can count, for every k, the size of the class 1: they are sets with at least one edge and with all the vertices are neighbours of a fixed v (their intersection). So:

$$|\text{Class } 1_k| = \sum_{1 \leq i \leq \lfloor \frac{k}{2} \rfloor} n \cdot \binom{3\alpha n}{i} \cdot \binom{3\alpha n - i}{k - 2i} \cdot 2^{k-2i}.$$

The class 2 can be counted "exactly" (a sure lower bound) for $k = 3$ because for $k = 2$ is empty or hard to compute. The 3-sets will be formed from elements that has one edge union one vertex such that has not the third vertex of the triangle as a neighbour plus the 3-sets that forms edge-disjoint triangles (the only ones we are supposing) plus the pairs of edges that forms 2-paths and no triangles. Those make the class 2 to be in size:

$$|\text{Class } 2_3| = 3\alpha n^2 \left(n - (18\alpha n - 3)\right) + \alpha n^2 + \frac{n\left(6\alpha n\left(6\alpha n - 2\right)\right)}{2}.$$

The following k-classes are difficult to count but we know they will be a 3-set from class 2 joined with any other vertex of G. Thus we can estimate them using the Kruskal-Katona Theorem (see [**2**], [**17**], [**13**]) that gives general lower bounds for shadows of collections of k-sets but can be easily reformulate to compute lower bounds on shades of collections of k-sets (the things we want to count).

Thus we have less k-sets for which we should look at their k-intersection. Also we can try to use the relations between classes to get better knowledge about the union of the k-neighbourhoods (its size) or the size of the vertex set that $N(\delta)$ should share. At this point some questions arises:

Questions:

- Can we know that the proportion of sets in k-th class 2 versus the whole $\binom{n}{k}$ will growth with n quick enough so we can be sure to have large intersections of k-sets and large unions of theirs neighbourhoods as n grows?
- Can we be sure to have some k-set with, although not big Δ, large k-neighbourhood union?

4.2. Other problems and future work. The original argument used by Varnavides to prove the Varnavides Theorem (see [**27**]) can be used along with the Szemerédi Theorem to say that, in a set $A \subset [N]$ of positive density ($|A| = \delta N$ for some δ) we will have $O(N^2)$ k-arithmetic progression. In particular, so we should have some distance d for which there are $O(N)$ k-arithmetic progressions with common difference d, as N goes to infinity.

One can ask about how is this constant: how many k-arithmetic progressions should share some common difference d in a set with δN elements. In [**4**] Bergelson, Host and Kra asked if in the cases where $k = 3, 4$ one can found, for every $\epsilon > 0$, an N big enough such that one can found $(1-\epsilon)\delta^3 N$ or $(1-\epsilon)\delta^4 N$ 3 and 4-arithmetic progressions with common difference d respectively. They also bounded from above the $k \geq 5$ case, proving that a similar statement is false for $k \geq 5$ based on an example by Ruzsa.

In [**12**] Ben Green answers affirmatively to the 3-arithmetic progression case using the Szemerédi Regularity Lemma for groups with its Counting Lemma. The question for the 4-arithmetic progression case remains open. One open issue is to find an alternative proof for the case of 3–tern arithmetic progression which makes no use of the Regularity Lemma for groups.

The second question is to deal with the case of 4–term arithmetic progressions for which the question remains open. One possibility is to use the Removal Lemma for systems of equations. We have not been able to obtain a version of this Removal Lemma for the systems of equations which describe k–term arithmetic progresions for $k \geq 4$, although this was one of the motivations of Section 2. One way to try to solve this problem is to state a colored version of the Regularity

Lemma for groups. This theorem could be useful to extend the Counting Lemma for groups to other more complex structures.

Bibliography

[1] N. Alon, A. Shapira, Testing subgraphs in directed Graphs, *Proc. of the 35 ACM STOC, ACM Press* (2003), 700–709. Also: *JCSS* **69** (2004), 354–382.

[2] I. Anderson, Combinatorics of Finite Sets, Dover Publications, New York 2002.

[3] F.A. Behrend, On sets of integers wich contain no three terms in arithmetical progression, *Proc. Nat. Acad. Sci. USA*, **32** (1946), 331–332.

[4] V. Bergelson, B. Host, B. Kra, Multiple recurrence and nilsequences, *Inventiones Math.* **160,2** (2005) 261–303.

[5] B. Bollobás, Modern Graph Theory, Springer-Verlag, New York, (1998).

[6] J. Bourgain, On triples in arithmetic progression, *GAFA* **9** (1999), no.5, 968–984.

[7] R. Diestel, Graph Theory, Third Edition, Springer-Verlag, New York, (2005).

[8] P. Erdös, M. Simonovits, A limit theorem in graph theory, *Studia Sci. Math. Hung.* **1** (1966), 51–57.

[9] P. Erdös, A.H. Stone, On the structure of linear graphs, *Bull. Amer. Math. Soc.* **52** (1946), 1089–1091.

[10] W. T. Gowers, Lower bounds of tower type for Szemerédi Uniformity Lemma, *GAFA* **7** (1997), 322–337.

[11] W. T. Gowers, Quasirandomness, counting, and regularity for 3-uniform hypergraphs, *Comb. Probab. Comput.* **15** (1-2). (2006), pp. 143–184.

[12] B. Green, A Szemerédi-type regularity lemma in abelian groups, with applications, *GAFA* **15** (2005) no.2, 340–376.

[13] G. O. H. Katona, A theorem on finite sets, *Theory of graphs. Proc. Colloq. Tihany*, (1966), 187–207, Akademiai Kiado, Academic Press, New York.

[14] J. Komlós, A. Shokoufandeh, M. Simonovits, E. Szemerédi, The regularity lemma and its applications in graph theory, *Theoretical aspects of computer science* (Tehran, 2000), 84–112, Lecture Notes in Comput. Sci., **2292**, Springer, Berlin, (2002).

[15] J. Komlós, M. Simonovits, Szemerédi's regularity lemma and its applications in graph theory. *Combinatorics, Paul Erdös is eighty, Vol.2 (Keszthely, 1993)*, 295–352, Bolyai Soc. Math. Stud., **2**, János Bolyai Math. Soc., Budapest, (1996), 295–352.

[16] D. Kral, L. Vena, Note on a Ben Geen theorem: extension to the non-abelian case, *preprint.*

[17] J. B. Kruskal, The number of simplices in a comlex, *Mathematical optimization techniques* (ed. R. Bellman) (1963), 251–278, University of California Press, Berkeley.

[18] B.M. Landman, A. Robertson, Ramsey Theory on the Integers, American Mathematical Society, 2004.

[19] J. Nešetřil, A. Raspaud, Colored homomorphisms of colored mixed graphs, *J. Combin. Theory Ser. B* **80** (2000), no. 1, 147–155.

[20] K.F. Roth, On certain sets of integers,*J. London Math. Soc.* **28** (1953), 245–252.

[21] I. Ruzsa, E. Szemerédi, Triple systems with no six points carrying three triangles, *Colloq. Math. Soc. J. Bolyai Combinatorics (Keszthely, 1976)*, **18** (1978), Vol. II., 939–945. North-Holland, Amsterdam-New York.

[22] E. Szemerédi, On sets of integers containing no k elements in arithmetic progression, Acta. Arith. **27** (1975),299–345.

[23] E. Szemerédi, Regular partitions of graphs, *Colloques Internationaux C.N.R.S. N°* 260 - *Problèmes Combinatoires et Théorie des Graphes*, Orsay (1976), 399–401.

[24] T. Tao, V. Vu, Additive Combinatorics, Cambridge University Press, New York, 2006.

[25] P. Turán, On an extremal problem in graph theory (in Hungarian), *Matematikai és Fizikai Lapok* **48** (1941), 436–452.

[26] B.L.Van der Waerden, Beweis einer baudetshen Vermutung, *Nieuw Archief voor Wiskunde* **15** (1927), 212–216.

[27] P. Varnavides, On certain sets of positive density, *J. London Math. Soc.* **34** (1959) 358–360.

[28] Spring School on Combinatorics: http://kam.mff.cuni.cz/ spring/2007/index.html