4-19-2017

The Regularity Lemma and its Applications

Elizabeth Sprangel
Western Michigan University, elizabethsprangel@gmail.com

WESTERN MICHIGAN UNIVERSITY
LEE HONORS COLLEGE

The Regularity Lemma and Its Applications

by

Elizabeth Sprangel

Committee:
Dr. Andrzej Dudek, Chair
Dr. Patrick Bennett, Member
Dr. Allen Schwenk, Member

May 2017

Abstract

The regularity lemma (also known as Szemerédi's Regularity Lemma) is one of the most powerful tools used in extremal graph theory. In general, the lemma states that every graph has some structure. That is, every graph can be partitioned into a finite number of classes in a way such that the number of edges between any two parts is "regular." This thesis is an introduction to the regularity lemma through its proof and applications. We demonstrate its applications to extremal graph theory, Ramsey theory, and number theory.

Contents

Abstract **i**

1 Introduction **1**

2 Szemerédi's Regularity Lemma **3**
2.1 Definitions, Notations, and Examples . . . 3
2.2 Szemerédi's Regularity Lemma . . . 6

3 Other Forms of the Regularity Lemma **14**
3.1 Degree Form . . . 14
3.2 Many Colors . . . 15

4 Applications **16**
4.1 Triangle Free Graphs . . . 16
4.2 Roth's Theorem for 3-Term Arithmetic Progressions . . . 18
4.3 Erdős and Stone's Theorem . . . 19
4.4 Chvatál, Rödl, Szemerédi, and Trotter's Theorem . . . 21
4.5 Ramsey-Turán for K_4 . . . 22

Bibliography **26**

Chapter 1

Introduction

Combinatorics is a branch of mathematics that can be loosely defined as the study of counting discrete structures. Many simple questions can be posed in combinatorics, including enumerating how many ways to arrange a set of objects, or counting the number of ways to choose a subset of a structure. Here are some introductory combinatorial problems:

1. Given a team of 12 basketball players, how many ways can we assign these players to 5 positions on the court?
2. How many ways can you put 6 balls in 14 boxes, where each box can have at most one ball?
3. If there are 30 colors of paint, what is the total number of ways one can distribute these colors to 9 rooms, ensuring that each room gets painted?

Although *"counting"* may sound easy, the level of difficulty rises quickly as the problems add more variables and conditions. It is often found that the ideas and techniques of combinatorics are being used not only in the traditional areas of mathematical application, namely computer science and physical sciences, but also in the social sciences, information theory and the biological sciences [4]. For example, determining the precise order of nucleotides within a DNA molecule, called DNA sequencing, is a typical problem where combinatorial methods are successfully applied.

A graph is a set of points and a set of lines, where a line connects two points. It is standard to call points vertices and lines edges. Extremal graph theory is a part of combinatorics. It was first considered a topic to be studied as its own subject by Paul Turán in 1940. Paul Erdős pioneered the subject through the problems, papers, and lectures he produced [1]. Extremal graph theory problems can take two forms.

First, given a certain quality of a graph, such as the number of vertices or edges, what properties does the graph have? Otherwise, given a property, how many edges or vertices can a graph have while still forbidding the property [6]? A classic example of the first form, proved by Mantel in 1907, is the following:

Mantel's Theorem (1907). Every graph with n vertices and at least $\frac{n^2}{4}$ edges contains a triangle.

This problem falls under the first form because given a certain number of vertices and edges, a graph is guaranteed to have the property that it contains a triangle. Mantel's theorem was generalized by Turán in 1941, bringing extremal graph theory to the forefront of mathematical research.

Turán's Theorem (1941). If G is a graph with n vertices that does not contain K_r, then G has at most $\frac{n^2(r-2)}{2(r-1)}$ edges.

Notice that this theorem was a solution to the second form of extremal graph problems. Many problems in extremal graph theory are concerned with finding the maximum number of edges a graph with n vertices can have while forbidding a certain type of subgraph. One of the most successful techniques used in solving these kind of problems is the so-called regularity method.

Chapter 2

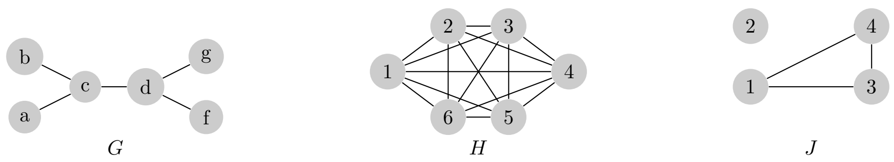
Szemerédi's Regularity Lemma

The regularity method, developed by Szemerédi, is a vital tool used in extremal graph theory. This lemma first appeared as a part of a larger result by Szemerédi [11]. The basic idea of the regularity lemma is that every graph has some structure, which can be approximated as some collection of random graphs with uniformly distributed edges. More precisely, all of the vertices of a graph G can be put into a finite number of classes. Most of these classes are "regular," meaning the number of edges between two subsets of the classes is about the same as the number of edges between the whole classes [2]. The regularity method is of great importance when solving extremal problems, since it provides structure to graphs that have seemingly little structure.

2.1 Definitions, Notations, and Examples

Before we can state and prove Szemerédi's Regularity Lemma, we must first define a few terms. First, we will define a few basic definitions in graph theory, and then what it means for a graph to have "regular" parts.

A **graph** is a collection two sets $G = (V, E)$ with $V := V(G)$ denoting a set of **vertices** and $E := E(G)$ denoting a set of **edges**. Vertices can be thought of as nodes or points and edges can be thought of as lines connecting the vertices. The elements of E are 2-element subsets of V, so $E \subseteq [V]^2$. An edge e is typically denoted xy (or yx), where $\{x, y\} \in e$. The following are a few examples of graphs:



A vertex v is **incident** with an edge e if $v \in e$, or the edge has an **endpoint** v. Two vertices x, y are **adjacent** if $xy \in E(G)$. The vertex x is called a **neighbor** of y if the two vertices are adjacent. For example, in graph H above, all vertices are adjacent to all other vertices, so every vertex has every other vertex as a neighbor. When all vertices are adjacent to one another, the graph is called **complete** and is denoted K_r, where $r = |V(G)|$. So, H is K_6.

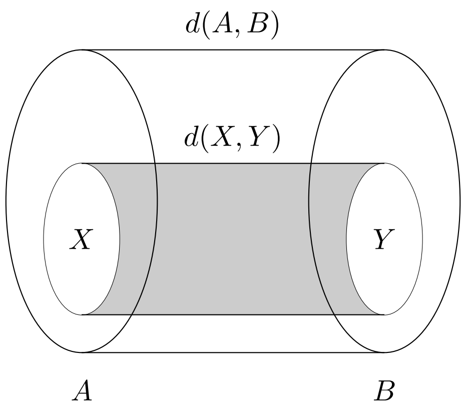
The **degree** of a vertex v, denoted $d(v)$, is the number of neighbors v has. The **minimum degree** of a graph G is $\delta(G) := min\{d(v) : v \in V\}$, and the **maximum degree** of G is $\Delta(G) := max\{d(v) : v \in V\}$.

Now, we will define what it means for a graph to be "regular." Let $G = (V, E)$ be a graph and X, Y be disjoint subsets of V. Let $e(X, Y)$ denote the number of edges between X and Y. Then,

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

is the **density** of the pair (X, Y). Notice that the density must always be a real number between 0 and 1, since $0 \leq e(X, Y) \leq |X| \cdot |Y|$.

A pair (A, B) of disjoint subsets of V is an **ε-regular pair** for some $\varepsilon > 0$ if for all $X \subseteq A$ with $|X| \geq \varepsilon|A|$ and $Y \subseteq B$ with $|Y| \geq \varepsilon|B|$, it follows that $|d(X, Y) - d(A, B)| < \varepsilon$.



Notice that $|X| \geq \varepsilon|A|$ and $|Y| \geq \varepsilon|B|$. This is because if we considered sets too small, say $|X| = 1 = |Y|$, then the density of the pair (X, Y) would be either 1 or 0, and thus $|d(X, Y) - d(A, B)| = d(A, B)$, which might not be less than ε.

Let $V = V_0 \cup V_1 \cup ... \cup V_k$ be a partition of the vertices of graph G. This parition is an **ε-regular partition** of G if the following are true:

(i) $|V_0| \leq \varepsilon|V|$

(ii) $|V_1| = |V_2| = ... = |V_k| = \frac{|V|-\varepsilon|V|}{k}$

(iii) all but at most εk^2 of the pairs (V_i, V_j) with $1 \leq i < j \leq k$ are ε-regular.

The set V_0 is called the **exceptional set**. This set exists so that the other partition sets can have the same number of vertices. These vertices are not considered when looking at the density of G, since they are not ε-regular and there are so few.

Example. We will show that any ε-regular pair in a graph is also ε-regular in the complement of the graph.

Consider the ε-regular pair (X, Y) in a graph G and let $A \subseteq X$ and $B \subseteq Y$ with $|A| \geq \varepsilon|X|$ and $|B| \geq \varepsilon|Y|$. Let $d(X, Y)$ be the density between X and Y in G, and let $d(A, B)$ be the density between A and B in G. Also, let $d_c(X, Y)$ be the density between X and Y in G^c, and let $d_c(A, B)$ be the density between A and B in G^c.

Notice, $d_c(X, Y) = 1 - d(X, Y)$ and $d_c(A, B) = 1 - d(A, B)$. Now,

$$|d_c(X, Y) - d_c(A, B)| = |1 - d(X, Y) - 1 + d(A, B)| = |d(X, Y) - d(A, B)| < \varepsilon.$$

So, (X, Y) is an ε-regular pair in G^c.

Now that we have an idea of what regular means in the context of graphs, we can prove the following proposition about ε-regular pairs. This proposition states that if a pair (A, B) is a ε-regular pair, then most of the vertices in A are neighbors with almost the density times the size of B number of vertices in B. We will use this proposition later in the first application we consider.

Proposition 2.1. *Let (A, B) be an ε-regular pair, of density d say, and let $Y \subseteq B$ have size $|Y| \geq \varepsilon|B|$. Then all but fewer than $\varepsilon|A|$ of the vertices in A have (each) at least $(d - \varepsilon)|Y|$ neighbors in Y.*

Proof. Let X be the set of vertices in A with at most $(d-\varepsilon)|Y|$ neighbors in Y. So, $e(X,Y) < |X|(d-\varepsilon)|Y|$. Now,

$$d(X,Y) < \frac{|X|(d-\varepsilon)|Y|}{|X||Y|} = d - \varepsilon = d(A,B) - \varepsilon.$$

Solving for ε yields

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

Since (A,B) is an ε-regular pair and $|Y| \geq \varepsilon|B|$, it must follow that $|X| < \varepsilon|A|$. Otherwise, (A,B) would not satisfy the definition of ε-regular. So, all but fewer than $\varepsilon|A|$ of the vertices in A have at least $(d-\varepsilon)|Y|$ neighbors in Y. □

2.2 Szemerédi's Regularity Lemma

Szemerédi's Regularity Lemma. For all $\varepsilon > 0$ and every integer $m \geq 1$, there exists an integer M such that for any graph G with $|V(G)| \geq m$, there exists an ε-regular partition $\mathcal{P} = \{C_0, C_1, \ldots, C_k\}$ with $m \leq k \leq M$.

This lemma states that for any $\varepsilon > 0$, every graph has a partition into a bounded number of ε-regular sets. If the number of sets was not bounded above by M, then any graph could be trivially partitioned into singletons. This lemma is powerful for larger graphs. We will now begin to prove the regularity lemma, using several preliminary lemmas.

Let G be a graph with $|V(G)| = n$. For disjoint sets $A, B \subseteq V$ we define

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

For partitions $\mathcal{A}$ of A and $\mathcal{B}$ of B we let

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

For a partition $\mathcal{P} = \{C_1, ..., C_k\}$ of $V(G)$ we let

$$q(\mathcal{P}) := \sum_{i<j} q(C_i, C_j).$$

Notice, $q(\mathcal{P})$ is bounded above by 1. This is because of the following:

$$\begin{aligned} q(\mathcal{P}) &= \sum_{i<j} q(C_i, C_j) \\ &= \sum_{i<j} \frac{|C_i||C_j|}{n^2} d^2(C_i, C_j) \\ &= \sum_{i<j} \frac{|C_i||C_j|}{n^2} \frac{e(C_i, C_j)^2}{(|C_i||C_j|)^2} \\ &\leq \sum_{i<j} \frac{1}{n^2} \frac{(|C_i||C_j|)^2}{|C_i||C_j|} \\ &= \frac{1}{n^2} \sum_{i<j} |C_i||C_j| \\ &\leq 1. \end{aligned}$$

Since every refinement of q increases by a constant and is bounded above by 1, there must be a finite number of refinements for the partition to be ε-regular.

If $\mathcal{P}$ has εk^2 or more irregular pairs (C_i, C_j) we may take the pairs (X, Y) of subsets of (C_i, C_j) which violate the regularity and make X and Y into partition sets of their own. Now, Lemma 2.2 will show that this refines $\mathcal{P}$ into a partition $\mathcal{P}'$, where $q(\mathcal{P}')$ is greater than $q(\mathcal{P})$. To prove this, we will use the Cauchy-Schwarz Inequality.

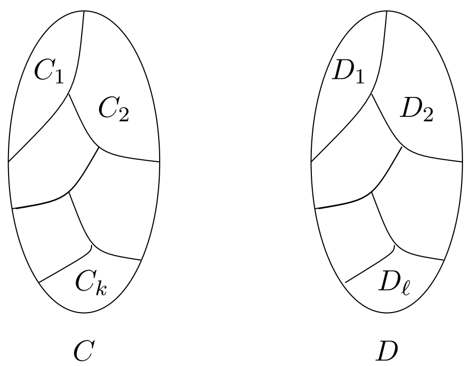
Cauchy-Schwarz Inequality.

$$\sum a_i^2 \sum b_i^2 \geq \left(\sum a_i b_i\right)^2.$$

Lemma 2.2. *(i) Let $C, D \subseteq V$ be disjoint. If $\mathcal{C}$ is a partition of C and $\mathcal{D}$ is a partition of D, then $q(\mathcal{C}, \mathcal{D}) \geq q(C, D)$.*

(ii) If $\mathcal{P}$, $\mathcal{P}'$ are partitions of V and $\mathcal{P}'$ refines $\mathcal{P}$, then $q(\mathcal{P}') \geq q(\mathcal{P})$.

Proof. (i) Let $\mathcal{C} = \{C_1, \ldots, C_k\}$ and $\mathcal{D} = \{D_1, \ldots, D_\ell\}$.



By the definition of $q(\mathcal{C}, \mathcal{D})$ and $q(C_i, D_j)$,

$$\begin{aligned} q(\mathcal{C}, \mathcal{D}) &= \sum_{i,j} q(C_i, D_j) \\ &= \frac{1}{n^2} \sum_{i,j} \frac{e(C_i, D_j)^2}{|C_i||D_j|}. \end{aligned}$$

Letting $a_i = e(C_i, D_j)/\sqrt{|C_i||D_j|}$ and $b_i = \sqrt{|C_i||D_j|}$, the Cauchy-Schwartz Inequality yields

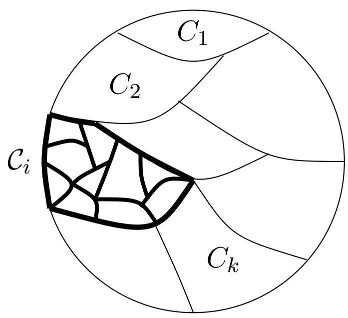
$$\begin{aligned} \sum_{i,j} \left(\frac{e(C_i, D_j)}{\sqrt{|C_i||D_j|}} \right)^2 \sum_{i,j} \left(\sqrt{|C_i||D_j|} \right)^2 &\geq \left(\sum_{i,j} e(C_i, D_j) \right)^2 \\ \sum_{i,j} \frac{e(C_i, D_j)}{|C_i||D_j|} &\geq \frac{\left(\sum_{i,j} e(C_i, D_j) \right)^2}{\sum_i |C_i| \sum_j |D_j|}. \end{aligned}$$

Notice that the sum of the edges between C_i and D_j for all i, j is the number of edges between C and D. Also, the sum over i, j of $|C_i|$ and $|D_j|$ is $|C|$ and $|D|$ respectively. Knowing this and the inequality we get from Cauchy-Schwartz, we have

$$\begin{aligned} \frac{1}{n^2} \sum_{i,j} \frac{e(C_i, D_j)^2}{|C_i||D_j|} &\geq \frac{1}{n^2} \cdot \frac{\left(\sum_{i,j} e(C_i, D_j) \right)^2}{\sum_i |C_i| \sum_j |D_j|} \\ &= \frac{1}{n^2} \cdot \frac{e(C, D)^2}{\sum_i |C_i| \sum_j |D_j|} \\ &= \frac{e(C, D)^2}{n^2|C||D|} \\ &= q(C, D). \end{aligned}$$

So, $q(\mathcal{C}, \mathcal{D}) \geq q(C, D)$, where $\mathcal{C}$ is a partition of C and $\mathcal{D}$ is a partition of D.

(ii) Let $\mathcal{P} := \{C_1, \ldots, C_k\}$ and let $\mathcal{C}_i$ be the partition of C_i induced by $\mathcal{P}'$, for $i = 1, \ldots, k$.



V

Using the definition of $q(\mathcal{P})$ and the inequality from part (i), we have

$$\begin{aligned} q(\mathcal{P}) &= \sum_{i<j} q(C_i, C_j) \\ &\leq \sum_{i<j} q(\mathcal{C}_i, \mathcal{C}_j) \\ &\leq q(\mathcal{P}') \end{aligned}$$

Notice that the last inequality holds because $q(\mathcal{P}')$ is the sum over i, j of the values of q within the part $\mathcal{C}_i$ and between parts $\mathcal{C}_i$ and $\mathcal{C}_j$, so $q(\mathcal{P}') = \sum_i q(\mathcal{C}_i) + \sum_{i,j} q(\mathcal{C}_i, \mathcal{C}_j)$. Therefore, $q(\mathcal{P}') \geq q(\mathcal{P})$.

□

Next, we will prove that the index $q(\mathcal{C}, \mathcal{D})$, where $\mathcal{C}$ and $\mathcal{D}$ are refinements of C and D, increases $q(C, D)$ by at most $\varepsilon^4 \frac{|C||D|}{n^2}$.

Lemma 2.3. *Let $\varepsilon > 0$ and let C, D be disjoint subsets of V. If (C, D) is not ε-regular, then there exists partitions $\mathcal{C} = \{C_1, C_2\}$ of C and $\mathcal{D} = \{D_1, D_2\}$ of D so that*

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

Proof. Suppose (C, D) is not an ε-regular pair. So, there exists sets $C_1 \subseteq C$ and $D_1 \subseteq D$ with $|C_1| > \varepsilon|C|$ and $|D_1| > \varepsilon|D|$ so that $\eta =: |d(C_1, D_1) - d(C, D)| \geq \varepsilon$. Let $C_2 =: C \backslash C_1$ and $D_2 =: D \backslash D_1$, and let $\mathcal{C} =: \{C_1, C_2\}$ and $\mathcal{D} =: \{D_1, D_2\}$.

Now, define the following notation:

$$\begin{aligned} c_i &:= |C_i| \qquad & d_j &:= |D_j| \qquad & e_{ij} &:= e(C_i, D_j) \\ c &:= |C| & d &:= |D| & e &:= e(C, D). \end{aligned}$$

We will now show that $q(\mathcal{C}, \mathcal{D}) \geq q(C, D) + \varepsilon^4 \frac{|C||D|}{n^2}$.

First, using the definition of $q(\mathcal{C}, \mathcal{D})$,

$$\begin{aligned} q(\mathcal{C}, \mathcal{D}) &= \sum_{i,j} q(C_i, D_j) \\ &= \frac{1}{n^2} \sum_{i,j} \frac{e_{ij}^2}{c_i d_j} \\ &= \frac{1}{n^2} \left(\frac{e_{11}^2}{c_1 d_1} + \sum_{i+j>2} \frac{e_{ij}^2}{c_i d_j} \right) \end{aligned}$$

Now, using the Cauchy-Schwartz Inequality with $a = e_{ij}/\sqrt{c_i d_j}$ and $b = \sqrt{c_i d_j}$, we have the following:

$$\frac{1}{n^2}\left(\frac{e_{11}^2}{c_1 d_1} + \sum_{i+j>2} \frac{e_{ij}^2}{c_i d_j}\right) \geq \frac{1}{n^2}\left(\frac{e_{11}^2}{c_1 d_1} + \frac{\sum_{i+j>2} e_{ij}^2}{\sum_{i+j>2} c_i d_j}\right)$$
$$= \frac{1}{n^2}\left(\frac{e_{11}^2}{c_1 d_1} + \frac{(e - e_{11})^2}{cd - c_1 d_1}\right).$$

Now, we will solve $\eta = d(C_1, D_1) - d(C, D)$ for e_{11}:

$$\eta = d(C_1, D_1) - d(C, D)$$
$$\eta = \frac{e_{11}}{c_1 d_1} - \frac{e}{cd}$$
$$e_{11} = \left(\eta + \frac{e}{cd}\right) c_1 d_1$$
$$e_{11} = \eta c_1 d_1 + \frac{e c_1 d_1}{cd}.$$

Substituting this for e_{11}, he have

$$\frac{1}{n^2}\left(\frac{e_{11}^2}{c_1 d_1} + \frac{(e - e_{11})^2}{cd - c_1 d_1}\right) = \frac{1}{n^2}\left[\frac{(\eta c_1 d_1 + (e c_1 d_1)/(cd))^2}{c_1 d_1} + \frac{(e - \eta c_1 d_1 - (e c_1 d_1)/(cd))^2}{cd - c_1 d_1}\right]$$
$$= \frac{1}{n^2}\left[\frac{\eta^2 c_1^1 d_1^2 + (2\eta e c_1^2 d_1^2)/(cd) + (e^2 c_1^2 d_1^2)/(c^2 d^2)}{c_1 d_1} + \frac{(e(cd - c_1 d_1)/(cd) - \eta c_1 d_1)^2}{cd - c_1 d_1}\right]$$
$$= \frac{1}{n^2}\left[\eta^2 c_1 d_1 + \frac{2\eta e c_1 d_1}{cd} + \frac{e^2 c_1 d_1}{c^2 d^2} + \frac{\frac{e^2 (cd - c_1 d_1)^2}{(c^2 d^2)} - \frac{2e\eta c_1 d_1 (cd - c_1 d_1)}{(cd)} + \eta^2 c_1^2 d_1^2}{cd - c_1 d_1}\right]$$
$$= \frac{1}{n^2}\left[\eta^2 c_1 d_1 + \frac{2\eta e c_1 d_1}{cd} + \frac{e^2 c_1 d_1}{c^2 d^2} + \frac{e^2 (cd - c_1 d_1)}{c^2 d^2} - \frac{2e\eta c_1 d_1}{cd} + \frac{\eta^2 c_1^2 d_1^2}{cd - c_1 d_1}\right]$$
$$= \frac{1}{n^2}\left[0 + \frac{e^2}{cd} + \eta^2 c_1 d_1 + \frac{2\eta^2 c_1^2 d_1^2}{cd - c_1 d_1}\right]$$
$$\geq \frac{1}{n^2}\left[\frac{e^2}{cd} + \eta^2 c_1 d_1\right].$$

Since $|C_1| > \varepsilon|C|$, $|D_1| > \varepsilon|D|$, and $|\eta| > \varepsilon$, we have

$$\begin{aligned}\left[\frac{e^2}{cd} + \eta^2 c_1 d_1\right] &\geq \frac{1}{n^2}\left[\frac{e^2}{cd} + \eta^2 \cdot \varepsilon|C| \cdot \varepsilon|D|\right] \\ &\geq \frac{1}{n^2}\left[\frac{e^2}{cd} + \varepsilon^4 cd\right] \\ &= q(C, D) + \varepsilon^4 \frac{cd}{n^2}.\end{aligned}$$

So, we have the desired conclusion to the lemma.

□

Now, in the Key Lemma, we will show that if $\mathcal{P}'$ is a refinement of $\mathcal{P}$, then the increase of the index is bounded below by some constant which only depends on ε.

Key Lemma. Let $0 < \varepsilon \leq 1/4$ and let $\mathcal{P} = \{C_0, C_1, \ldots, C_k\}$ be a partition of V with exceptional set C_0 with $|C_0| \leq \varepsilon n$ and $|C_1| = |C_2| = \cdots = |C_k| := c$. If $\mathcal{P}$ is not ε-regular, then there exists $\mathcal{P}' = \{C'_0, C'_1, C'_2, \ldots, C'_\ell\}$ where $k \leq \ell \leq k4^k$, so that $|C'_0| \leq |C_0| + n/2^k$, $|C'_1| = |C'_2| = \cdots = |C'_\ell|$, and $q(\mathcal{P}') \geq q(\mathcal{P}) + \varepsilon^5/2$.

Proof. Assume $\mathcal{P}$ is not ε-regular. Then, at least εk^2 pairs (C_i, C_j) are not ε-regular, so we refine these pairs as follows. For each pair (C_i, C_j) which is not ε-regular, Lemma 2 states that there exists partitions $\mathcal{C}_{ij}$ of C_i and $\mathcal{C}_{ji}$ of C_j so that $q(\mathcal{C}_{ij}, \mathcal{C}_{ji}) \geq q(C_i, C_j) + \varepsilon^4 c^2/n^2$. So, for each C_i, there exists a minimum parition, call it $\mathcal{C}_i$, with $|\mathcal{C}_i| \leq 2^{k-1}$. This is because for some C_i, each C_j such that (C_i, C_j) is not ε-regular, $\mathcal{C}_{ij}$ could parition C_i into at most twice as many parts.

Let $\mathcal{C}_0 =: \{\{v\} \mid v \in C_0\}$, and let $\mathcal{C} := \cup_{i=0}^{k} \mathcal{C}_i$. Notice,

$$|\mathcal{C}_0| + |\mathcal{C}_1| + \cdots + |\mathcal{C}_k| \leq |C_0| + k2^{k-1} \leq k2^k.$$

Also,

$$q(\mathcal{C}) = \sum_{1 \leq i < j} q(\mathcal{C}_i, \mathcal{C}_j) + \sum_{1 \leq i} q(\mathcal{C}_0, \mathcal{C}_i) + \sum_{0 \leq i} q(\mathcal{C}_i).$$

Note that

$$\begin{aligned}\sum_{0 \leq i} q(\mathcal{C}_i) &= q(\mathcal{C}_0) + \sum_{1 \leq i} q(\mathcal{C}_i) \\ &\geq q(\mathcal{C}_0).\end{aligned}$$

Since $\mathcal{C}_i$ is a partition of C_{ij}, $\mathcal{C}_j$ is a partition of C_{ji}, and $\mathcal{C}_i$ is a partition for $\{C_i\}$, Lemma 2(i) gives that

$$\begin{aligned}
\sum_{1\le i<j} q(\mathcal{C}_i,\mathcal{C}_j) + \sum_{1\le i} q(\mathcal{C}_0,\mathcal{C}_i) + \sum_{0\le i} q(\mathcal{C}_i) &\ge \sum_{1\le i<j} q(C_{ij},C_{ji}) + \sum_{1\le i} q(\mathcal{C}_0,\{C_i\}) + q(\mathcal{C}_0)\\
&\ge \sum_{1\le i<j}\left(q(\mathcal{C}_i,\mathcal{C}_j) + \frac{\varepsilon^4 c^2}{n^2}\right)\\
&\qquad + \sum_{1\le i} q(\mathcal{C}_0,\{C_i\}) + q(\mathcal{C}_0)\\
&\ge \sum_{1\le i<j} q(\mathcal{C}_i,\mathcal{C}_j) + \left(\varepsilon k^2\right)\frac{\varepsilon^4 c^2}{n^2}\\
&\qquad + \sum_{1\le i} q(\mathcal{C}_0,\{C_i\}) + q(\mathcal{C}_0)\\
&= q(\mathcal{P}) + \frac{\varepsilon^5 c^2 k^2}{n^2}\\
&\ge q(\mathcal{P}) + \frac{\varepsilon^5}{2}.
\end{aligned}$$

Now, we need to ensure that our refinement will have partitions of equal size. Let $C'_1, C'_2, \ldots, C'_\ell$ be a collection of disjoint sets of size $d := \lfloor c/4^k \rfloor$ so that every C'_i is contained in some $C \in \mathcal{C}\backslash\{C_0\}$ and $C'_0 := V\backslash \bigcup C'_i$. So, $\mathcal{P}' = \{C'_0, C'_1, \ldots, C'_\ell\}$ is a partition of V. Also, $\mathcal{P}'$ refines $\mathcal{C}$, so Lemma 2.2(ii) yields $q(\mathcal{P}') \ge q(\mathcal{C}) \ge q(\mathcal{P}) + \varepsilon^5/2$.

Since each C'_i is also contained in one of $C_1, \ldots, C_k$ and no more than 4^k sets C'_i can be contained in any one of the same C_j from the definition of d, it follows that $k \le \ell \le k4^k$. So $C'_1, C'_2, \ldots, C'_\ell$ use all but at most d vertices for each set $C_i \ne C_0$ in $\mathcal{C}$. Thus,

$$\begin{aligned}
|C'_0| &\le |C_0| + d|\mathcal{C}|\\
&= |C_0| + \frac{c}{4^k}|\mathcal{C}|\\
&\le |C_0| + \frac{c}{4^k}k2^k\\
&= |C_0| + \frac{ck}{2^k}\\
&\le |C_0| + \frac{n}{2^k}
\end{aligned}$$

The last inequality holds since ck is all of the vertices except those in the exceptional set. Therefore, $\mathcal{P}'$ satisfies the conditions in the lemma. □

We are now set up to choose a value for M, and then apply the Key Lemma to any graph until we obtain an ε-regular partition. The following is the proof of Szemerédi's Regularity Lemma.

Proof. Let G be a graph of orfer n with vertex set V. Let $0 < \varepsilon \leq 1/4$ and let m be a positive integer. Let $s := 2/\varepsilon^5$ be an upper bound on the number of applications of the Key Lemma, and recall that $q(\mathcal{P}) \leq 1$.

Now, we must choose M. Consider the function $f : x \to x4^x$. Take M to be the maximum of $f^5(k)$ and $2k/\varepsilon$. If $n \leq M$, then the partition of the graph into single vertices satisfies the conditions, so let $n > M$. Let $C_0 \subseteq V$ be minimal such that k divides $|V \backslash C_0|$, and consider any partition $\{C_1, \ldots, C_k\}$ of the vertices where $|C_i| = |V \backslash C_0|/k$.

We must now show $|C_0| \leq \varepsilon n$. Notice, $|C_0| < k$ and the exceptional set grows by at most $n/2^k$ each iteration of the Key Lemma. So, to obtain $k + sn/2^k \leq \varepsilon/n$, choose k large enough so $2^{k-1} \geq s/\varepsilon$. Then,

$$\begin{aligned} \frac{\varepsilon}{2} &\geq \frac{s}{2^k} \\ \varepsilon &\geq \frac{s}{2^k} + \frac{\varepsilon}{2} \\ \varepsilon n &\geq \frac{sn}{2^k} + \frac{\varepsilon n}{2}. \end{aligned}$$

Recall that $n > \frac{2k}{\varepsilon}$, since M cannot exceed this by definition and M is strictly larger than n. This implies that $k < \frac{n\varepsilon}{2}$, so it must follow that $\varepsilon n > \frac{sn}{2^k} + k$. This means that $|C_0|$ does not exceed εn. Now, we are set up to apply the Key Lemma until the partition is ε-regular. □

Chapter 3

Other Forms of the Regularity Lemma

Since the regularity lemma is such a powerful tool used in combinatorics, it has been adapted to many different forms to be better suited for different situtions. In this section, we will introduce two alternate forms of the lemma.

3.1 Degree Form

The degree form of the regularity lemma is frequently used in problems for which having more edges can only be beneficial. This form considers a subgraph G' of the graph G which we want to regularize. In G', there are no edges within any parts, all pairs are ε-regular, and all pairs have either a density of 0 or a sufficiently large density. Typically, the graph G' is easier to work with, which is why degree form can be preferred when it is applicable.

Degree Form of the Regularity Lemma. For every $\varepsilon > 0$, there exists M dependent only on ε such that if G is any graph and $d \in [0, 1]$, then there exists a partition $V = V_0 \cup V_1 \cup \cdots \cup V_k$ of the vertices of G and a subgraph $G' \subset G$ with the following properties:

(i) $k \leq M$,

(ii) $|V_0| \leq \varepsilon |V|$,

(iii) all parts V_i for $i \geq 1$ are of the same size $m \leq \lceil \varepsilon |V| \rceil$,

(iv) $deg_{G'}(v) > deg_G(v) - (d + \varepsilon)|V|$ for all $v \in V(G)$,

(v) $e(G'(V_i)) = 0$ for all $i \geq 1$,

(vi) all pairs $G'(V_i, V_j)$ for $1 \leq i < j \leq k$ are ε-regular with each having density either 0 or greater than d.

This form of the regularity lemma can be derived from Szemerédi's Regularity Lemma by using a "cleaning procedure."

3.2 Many Colors

In graph theory, it is common to consider coloring problems. In such problems, we will colors the vertices or edges of a graph and study what properties such graphs can have. Since it is common to work with colored graphs, there is a version of the regularity lemma which accounts for graphs with colored edges.

The many colors regularity lemma states that any graph colored with r colors can be partitioned in a way such that all but at most εk^2 pairs are epsilon regular for all r colors. That is, if we only consider the red edges, then all but at most εk^2 pairs are ε-regular with respect to the red edges. The following is a formal statement of this form of the lemma.

Many Colors Regularity Lemma. For all $\varepsilon \geq 0$ and integers r, m, there exists an M such that if the edges of a graph G with n vertices are r-colored, then the vertex set $V(G)$ can be partitioned into sets $V_0, V_1, ..., V_k$, for some $m \leq k \leq M$, so that $|V_i| = \frac{n}{k}$ for every $i \geq 1$, and for all but at most εk^2 pairs (V_i, V_j), $X \subset V_i$ and $Y \subset V_j$ of size $|X|, |Y| > \varepsilon n$ we have $|d_v(X, Y) - d_v(V_i, V_j)| < \varepsilon$ for $v = (1, \ldots, r)$.

The proof for the Many Colors Regularity Lemma is similar to the proof of the lemma without colors. To account for the different colors, we use a different index which sums over all the colors. If $\mathcal{P} = \{V_0, V_1, \ldots, V_k\}$ is a partition of G and $v \in (1, \ldots r)$, then

$$q(\mathcal{P}) = \frac{1}{k} \sum_{v} \sum_{i=1}^{k} \sum_{j=i+1}^{k} d_v^2(V_i, V_j)$$

is the index used to prove the Many Colors Regularity Lemma.

Chapter 4

Applications

In this chapter, we will show how the regularity lemma can be applied to problems in extremal combinatorics, Ramsey theory, and number theory. In the following applications, we use all three forms of the regularity lemma we have discussed thus far. First, we will consider the problem of triangle free graphs, which is a straight-forward application of the regularity lemma. Then, we illustrate how the regularity lemma can be used to prove Roth's theorem from number theory. Finally, we will look at a few results in Ramsey theory and the study of Turán numbers.

4.1 Triangle Free Graphs

The first application of the regularity method we are going to consider is about triangle free graphs. The theorem roughly states that if a graph contains at most some constant times n^3 triangles, then there exists at most a constant times n^2 edges which can be removed from the graph to make it triangle free.

Triangle Free Graphs Theorem. For every constant $c > 0$ there exists a constant $a > 0$ with the following property. If G is any graph with n vertices that contains at most an^3 triangles, then it is possible to remove at most cn^2 edges from G to make it triangle-free.

Proof. Apply the Regularity Lemma with $\varepsilon = c/4$ and $m = c$ obtaining a partition $V = \{V_0, V_1, ..., V_k\}$ of the vertices of G with all but at most $cn/4$ pairs (V_i, V_j), where $0 < i < j \leq k$, are $c/4$-regular. Now, we remove all edges between pairs which fail to be $c/4$-regular, which is at most

$$k^2 \frac{c}{4} \left(\frac{n - |V_0|}{k} \right)^2 \leq \frac{cn^2}{4}$$

edges. Now, we remove all of the edges between pairs which have density less than $c/2$, which is at most

$$\binom{k}{2}\frac{c}{2}\left(\frac{n-|V_0|}{k}\right)^2 \leq \frac{cn^2}{4}$$

edges. Finally, we remove all edges within each part. This is at most

$$\begin{aligned} k\binom{\frac{n-|V_0|}{k}}{2} &\leq k \cdot \frac{\left(\frac{n-|V_0|}{k}\right)^2}{2} \\ &\leq \frac{kn^2}{2k^2} \\ &\leq \frac{n^2}{2m} \\ &= \frac{cn^2}{2} \end{aligned}$$

edges. So, in total we removed at most cn^2 edges. Call this new subgraph with edges removed G'.

Now, we must show that G' is triangle free. Assume, to the contrary, that G' has a triangle. In particular, G' has a triangle with each vertex in a different part, say V_i, V_j, and V_ℓ. If there exists such a triangle, then (V_i, V_j), (V_j, V_ℓ), and (V_i, V_ℓ) must be $c/4$-regular and have density at least $c/2$. Since (V_i, V_j) is a $c/4$-regular pair, all but at most $(c/4)|V_i|$ vertices in V_i have at least $(c/2 - c/4)|V_j|$ neighbors in V_j, this set of vertices in V_j be denoted by V_j'. Similarly, all but at most $(c/4)|V_i|$ vertices in V_i have at least $(c/2 - c/4)|V_\ell|$ neighbors in V_ℓ, and call this set V_ℓ'. Now, since $c/2 - c/4 = c/4$, all but at most $(c/4)|V_i'|$ vertices in V_i have at least $(c/2 - c/4)|V_\ell'|$ neighbors in V_ℓ'.

So, $V_i \cup V_j \cup V_k$ has at least the following number of triangles:

$$\begin{aligned} \left(1-\frac{2c}{4}\right)\left(1-\frac{c}{4}\right)\left(\frac{c}{2}-\frac{c}{4}\right)^3\left(\frac{n-|V_0|}{k}\right)^3 &\geq \left(1-\frac{2c}{4}\right)\left(1-\frac{c}{4}\right)\left(\frac{c}{2}-\frac{c}{4}\right)^3\left(\frac{n-cn/4}{k}\right)^3 \\ &= \left(1-\frac{2c}{4}\right)\left(1-\frac{c}{4}\right)\left(\frac{c}{2}-\frac{c}{4}\right)^3\frac{(1-c/2)^3}{k^3}n^3 \end{aligned}$$

Now, since c and k do not depend on n, if

$$a = \left(1-\frac{2c}{4}\right)\left(1-\frac{c}{4}\right)\left(\frac{c}{2}-\frac{c}{4}\right)^3\frac{(1-c/2)^3}{k^3} - \delta$$

for some small δ, then G would contain less triangles than $V_i \cup V_j \cup V_k$, which is a contradiction. Note that a is positive, since $c < 1$. If $c \geq 1$, then it is trivial to remove at most n^2 edges to make G triangle free. So, G' must be triangle free and thus it is possible to remove at most cn^2 edges from G to make it triangle free.

□

The converse of this statement is known as the Triangle Removal Lemma, since it states that if it is necessary to remove at least cn^2 edges from G to make it triangle-free, then G contains at least an^3 triangles. We will use the Triangle Removal Lemma in the following application.

4.2 Roth's Theorem for 3-Term Arithmetic Progressions

A **k-term arithmetic progression**, or an AP_k, is a sequence of k numbers such that the difference of any successive numbers is constant. A subset A of the natural numbers has **positive upper density** if

$$\lim_{n\to\infty} \frac{|A|}{n} > 0.$$

In 1936, Erdős and Turán conjectured that every subset of the natural numbers with positive upper density contains arbitrarily long arithmetic progressions [8]. This conjecture was proven by Szemerédi in 1975 [11]. His proof was one of the first applications of the regularity lemma. The special case for AP_3 of this theorem was first proven by Klaus Roth in 1953 [9]. Roth's proof for this theorem utilized analytic number theory. In the following statement of the theorem, $[n]$ will be used to denote the set $\{1, 2, \ldots, n\}$.

Roth's Theorem for 3-Term Arithmetic Progressions (1953). For all $\varepsilon > 0$ there exists $N(\varepsilon)$ such that for all $n \geq N$ any set $S \subseteq [n]$ with $|S| \geq \varepsilon n$ contains an AP_3.

Proof. Let $S \subseteq [n]$ such that $|S| \geq \varepsilon n$, and assume that S contains no AP_3. We will begin by constructing a 3-partite graph $G = A \cup B \cup C$ from S. Let $V(A) = \{a : a \in [n]\}$, $V(B) = \{b : b \in [2n]\}$, and $V(C) = \{c : c \in [3n]\}$. Now, define the edges in the following way, where $a \in V(A)$, $b \in V(B)$, and $c \in V(C)$:

$$E(G) = \{ab : b - a \in S\} \cup \{bc : c - b \in S\} \cup \{ac : c - a \in 2S\}$$

Now, the vertices a, b, and c form a triangle in G if and only if $a + k = b$, $a + k + \ell = c$, and $k + \ell \in 2S$. For every vertex a and $k \in S$, there exists a triangle with vertices a, $a + k$, and $a + 2k$. Notice that these triangles are pairwise edge-disjoint. So, at least $c \in V(C)$ vertices in G must be removed to make the graph triangle free. Since $|S| \geq \varepsilon n$, $|S|n \geq \varepsilon n^2$. Hence, by the Triangle Removal Lemma, G contains at least γn^3 triangles, for some constant γ. Since there exists at least γn^3 triangles, there must be a nontrivial solution to $a + b = c$. This is a contradiction, so S must contain an AP_3. □

4.3 Erdős and Stone's Theorem

The following theorem published by Paul Erdős and Arthur Stone in 1946 [7] generalizes Turán's graph theorem to the extremal number of graphs containing no K_s^r, which is a complete r-partite graph where each part has s vertices. The Embedding Lemma, Turán's Theorem [12], and Lemma 4.1 will be used to prove the Erdős-Stone theorem.

Embedding Lemma. For all $d \in (0,1]$ and $\Delta \geq 1$, there exists $\varepsilon_0 > 0$ with the property that if G is any graph, H is a graph with $\Delta(H) \leq \Delta$, $s \in \mathbb{N}$, and R is any regularity graph of G with parameters $\varepsilon \leq \varepsilon_0$, $\ell \geq \frac{2s}{d^\Delta}$, and d, then $H \subseteq R_s$ implies $H \subseteq G$.

Turán's Theorem (1941). For all integers r, n with $r > 1$, every graph G which is K^r free with n vertices has at most $t_{r-1}(n) = \frac{r-2}{r-1} \cdot \frac{n^2}{2}$ edges.

Lemma 4.1.

$$\lim_{n\to\infty} t_{r-1}(n)\binom{n}{2}^{-1} = \frac{r-2}{r-1}.$$

In the following proof, we will use the regularity lemma to create a regularity graph. Then, using Turán's theorem, we will show that the regularized graph contains a complete graph on r vertices as a subgraph. If such a subgraph exists in the regularized graph, then the embedding lemma will give us that K_s^r is contained in the original graph, as desired.

Erdős-Stone Theorem (1946). For every $\varepsilon > 0$ and integers $r \geq 2$ and $s \geq 1$, there exists integer n_0 so that every graph G with $n \geq n_0$ vertices and at least $t_{r-1}(n) + \varepsilon n^2$ edges contains K_s^r as a subgraph.

Proof. Let $r \geq 2$ be given. If $s = 1$, then Turán's Theorem yeilds that $K^r \subseteq G$. So let $s \geq 2$ and set $\gamma > 0$. Let G be a graph of order n. Notice that if $|E(G)| \geq t_{r-1}(n) + \gamma n^2$ then $\gamma < 1$ because if $\gamma \geq 1$, then $|E(G)| > n^2$.

Let $d := \gamma$ and $\Delta := \Delta(K_s^r)$. The Embedding Lemma gives an $\varepsilon_0 > 0$. Now, let $m > \frac{1}{\gamma}$ and $\varepsilon > 0$ where $\varepsilon \leq \varepsilon_0$, $\varepsilon < \frac{\gamma}{2} < 1$, and

$$\delta := 2\gamma - \varepsilon^2 - 4\varepsilon - d - 1/m > 0.$$

This is possible because $m > \frac{1}{\gamma}$ implies that

$$2\gamma - d - \frac{1}{m} > 2\gamma - \gamma - \gamma = 0.$$

Given ε and m, Szemerdi's Regularity Lemma outputs an integer M. Let

$$n > \frac{2Ms}{d^\Delta(1-\varepsilon)} = \frac{2Ms}{\gamma^\Delta(1-\varepsilon)}.$$

Notice that this is at least m since $\delta, 1-\varepsilon < 1$. So Szemerdi's Regularity Lemma gives an ε-regular partition $\{V_0, V_1, ..., V_k\}$ of G where $m \leq k \leq M$ and $|V_1| = |V_2| = ... = |V_k| = \ell$ for some ℓ. Notice that $n \geq k\ell$, since k is the number of ε-regular partitions and ℓ is the number of verticies in each set V_i with $i \in (1, ..., k)$. Also, notice that

$$\begin{aligned}\ell &= \frac{n - |V_0|}{k} \\ &\geq \frac{n - \varepsilon n}{M} \\ &= n\left(\frac{1-\varepsilon}{M}\right) \\ &\geq \frac{2Ms}{d^{\Delta}(1-\varepsilon)}\left(\frac{1-\varepsilon}{M}\right) \\ &= \frac{2s}{d^{\Delta}}.\end{aligned}$$

Let R be the regularity graph of G with parameters ε, ℓ, and d. Given these parameters, the Embedding Lemma yields that if $K^r \subseteq R$, then $K^r_s \subseteq G$. To show that $K^r \subseteq R$, we will first show that R has more than $t_{r-1}(k)$ edges. Then, Turn's Theorem will give us $K^r \subseteq R$.

We will count the number of edges in G to show that there is a sufficient number. First, within V_0 there are at most $\binom{|V_0|}{2} \leq \frac{1}{2}(\varepsilon n)^2$ edges. Between V_0 and V_i, for $i \in (1, ..., k)$, there are at most $|V_0|k\ell \leq \varepsilon nk\ell$ edges. There are at most εk^2 pairs (V_i, V_j), for $i, j \in (1, ..., k)$, which are not ε-regular. Between such pairs there is at most ℓ^2 edges, so between all pairs which are not ε-regular, there are at most $\varepsilon k^2 \ell^2$ edges. The ε-regular pairs with density less than d have at most $d\ell^2$ edges between them, so altogether there are at most $\frac{1}{2}k^2 d\ell^2$ edges between pairs with insufficient density in G. Within a set V_i there are trivially at most $\binom{\ell}{2} \leq \frac{1}{2}\ell^2 k$ edges. The rest of the edges are between ε-regular pairs with density at least d, which are also edges in R_s. Every edge in R corresponds to at most ℓ^2 edges in G, so there are at most $|E(R)|\ell^2$ edges of this kind. By adding it all, we get

$$|E(G)| \leq \frac{1}{2}(\varepsilon n)^2 + \varepsilon nk\ell + \varepsilon k^2\ell^2 + \frac{1}{2}k^2 d\ell^2 + \frac{1}{2}\ell^2 k + |E(R)|\ell^2.$$

Solving for $|E(R)|$ gives

$$\begin{aligned}
|E(R)| &\geq \frac{|E(G)| - \frac{1}{2}(\varepsilon n)^2 + \varepsilon nk\ell + \varepsilon k^2\ell^2 + \frac{1}{2}k^2 d\ell^2 + \frac{1}{2}\ell^2 k}{\ell^2} \\
&= \frac{1}{2}k^2 \left(\frac{|E(G)| - \frac{1}{2}(\varepsilon n)^2 - \varepsilon nk\ell - \varepsilon k^2\ell^2 - \frac{1}{2}k^2 d\ell^2 - \frac{1}{2}\ell^2 k}{\frac{1}{2}k^2\ell^2} \right) \\
&\geq \frac{1}{2}k^2 \left(\frac{t_{r-1}(n) + \gamma n^2 - \frac{1}{2}(\varepsilon n)^2 - \varepsilon nk\ell}{\frac{1}{2}n^2} - 2\varepsilon - d - \frac{1}{k} \right) \\
&\geq \frac{1}{2}k^2 \left(\frac{t_{r-1}(n)}{\frac{1}{2}n^2} + 2\gamma - \varepsilon^2 - 4\varepsilon - d - \frac{1}{m} \right) \\
&= \frac{1}{2}k^2 \left(t_{r-1}(n)\binom{n}{2}^{-1}\left(1 - \frac{1}{n}\right) + \delta \right).
\end{aligned}$$

Lemma 1 can be applied to obtain

$$\begin{aligned}
\frac{1}{2}k^2 \left(t_{r-1}(n)\binom{n}{2}^{-1}\left(1 - \frac{1}{n}\right) + \delta \right) &> \frac{1}{2}k^2 \left(\frac{r-2}{r-1} \right) \\
&\geq t_{r-1}(k).
\end{aligned}$$

By Turán's Theorem, $K^r \subseteq R$, and so $K_s^r \subseteq R_s$. By the Embedding Lemma, $K_s^r \subseteq G$.

□

4.4 Chvatál, Rödl, Szemerédi, and Trotter's Theorem

The following theorem from Chvatál, Rödl, Szemerédi, and Trotter from 1983 [5] is a significant result in Ramsey theory. It states that for every graph with bounded maximum degree has a diagonal Ramsey number which is linear in the order of the graph. The diagonal Ramsey number $R_r(H)$ is the maximum number n such that there exists an r-edge-coloring of the complete graph K_n which contains no monochromatic copy of H as a subgraph. To prove this theorem, we will utilize Turán's Theorem and the Many Colors Form of the Regularity Lemma.

Chvatál, Rödl, Szemerédi, and Trotter (1983). For every positive integer Δ and $r \geq 2$, there is a $c(\Delta, r)$ such that $R_r(H) \leq c|H|$ for all graphs H with $\Delta(H) \leq \Delta$.

Proof. Let $\Delta \geq 1$ and $r \geq 2$ be given, and define $m := R_r(K_{\Delta+1})$. For $d := \frac{1}{r}$ and Δ, the Embedding Lemma gives ε_0. Choose $\varepsilon \leq \varepsilon_0$ be small enough so that $2\varepsilon < \frac{1}{m-1} - \frac{1}{m}$.

Notice that $\varepsilon < 1$. Now, let M be given by the Many Colors Regularity Lemma with inputs ε and m. Everything defined thus far depended only on Δ and r, so choose $c := \frac{2Mr^{\Delta}}{1-\varepsilon}$.

Let H be a graph with $\Delta(H) \leq \Delta$ and $s := |H|$, and let G be a graph with order $n \geq c|H|$. The Many Colors Regularity Lemma gives that G has an ε-regular partition $\{V_0, V_1, ..., V_k\}$ with exceptional set V_0 and $|V_1| = ... = |V_k| = l$ where $m \leq k \leq M$ and $|d_v(X,Y) - d_v(V_i,V_j)| < \varepsilon$ for $v = 1, ..., r$. Notice,

$$l = \frac{n - |V_0|}{k} \geq \frac{n(1-\varepsilon)}{M} \geq cs\frac{1-\varepsilon}{M} = \frac{2s}{(\frac{1}{r})^{\Delta}} = \frac{2s}{d^{\Delta}}.$$

Let R be the regularity graph of G with parameters ε, l, and 0. By definition, R has k vertices. Observe that

$$\begin{aligned}
|E(R)| &\geq \binom{k}{2} - \varepsilon k^2 \\
&= \frac{1}{2}k^2\left(1 - \frac{1}{k} - 2\varepsilon\right) \\
&> \frac{1}{2}k^2\left(1 - \frac{1}{k} - \frac{1}{1-m} + \frac{1}{m}\right) \\
&> \frac{1}{2}k^2\left(1 - \frac{1}{k} - \frac{1}{1-m} + \frac{1}{k}\right) \\
&= \frac{1}{2}k^2\left(\frac{m-2}{m-1}\right) \\
&\geq t(k, m-1).
\end{aligned}$$

By Turan's Theorem, R has K_m as a subgraph.

Now, color the edges of R by coloring an edge corresponding to a pair with color $v \in \{1, ..., r\}$ where $d_v(V_i, V_j) \geq \frac{1}{r}$. That is, the edge density in the v-th color is greater than or equal to density d. For each color, let R_v denote the subgraph of R induced by the v-th color (with parameters ε, l, and $\frac{1}{r}$). Since $m := R_r(K_{\Delta+1})$, K_m contains a monochromatic copy of $K_{\Delta+1}$, say the v-th color. Since $\chi(H) \leq \Delta + 1$, $H \subset (K_{\Delta+1})_s$, and thus $H \subset (R_v)_s$.

Since $\varepsilon \leq \varepsilon_0$, $l \geq \frac{2s}{d^{\Delta}}$, and $d \in (0, 1]$, $H \subset G$ by the Embedding Lemma. □

4.5 Ramsey-Turán for K_4

The study of Ramsey-Turán numbers is motivated by Ramsey theory and the study of Turán numbers. Given integers n and s, the Ramsey-Turán number $RT(n, H, s)$ is the maximum number of edges in a graph G with n vertices and independence number at most s so that G does not contain H as a subgraph. The following theorem was proven by Szemerédi in 1972 [10]. The basic idea of this theorem is if G is a graph with no K_4 as a subgraph and at most $o(n)$ independent vertices, then G has less than $1/8n^2+o(n^2)$ edges. In the proof, we will use the term cluster graph. Let G be a graph with partition $V = V_0 \cup V_1 \cup \cdots \cup V_k$, and let d be a number between 0 and 1. The **cluster graph** R of G has k vertices corresponding to the k parts of V, and there exists the edge ij if the corresponding parts V_i and V_j have density at least d.

Ramsey-Turán for K_4, Szemerédi (1972). Let G be a graph with n vertices with no K_4 as a subgraph and only $o(n)$ independent vertices. Then, $|E(G)| < \frac{1}{8}n^2 + o(n^2)$.

Proof. Let G be a graph with n vertices. Assume, to the contrary, that $|E(G)| > \frac{1}{8}n^2 + 4\varepsilon n^2$. Also, let $\alpha(G) \leq \frac{\varepsilon^2}{M(\varepsilon)}n - 1$ and $n \geq \frac{M(\varepsilon)}{\varepsilon}$, for $\varepsilon > 0$ and $M(\varepsilon)$ obtained from applying Degree Form to G with the parameter $d = 2\varepsilon$. Let G' denote the graph G in Degree Form, and let $G'' = G' - V_0$ be the usual pure graph. Also, let R be the cluster graph of G''. Notice,

$$E(G'') > E(G) - (d + 3\varepsilon)\frac{n^2}{2}.$$

Notice, by the assumption of the size of $E(G)$,

$$\begin{aligned} E(G'') &> \frac{1}{8}n^2 + 4\varepsilon n^2 - (d + 3\varepsilon)\frac{n^2}{2} \\ &= \frac{1}{8}n^2 + 4\varepsilon n^2 - (2\varepsilon + 3\varepsilon)\frac{n^2}{2} \\ &= \frac{1}{8}n^2 + \frac{3}{2}\varepsilon n^2 \\ &> \frac{1}{8}n^2 + \varepsilon n^2. \end{aligned}$$

Now, use the fact that $k \leq M(\varepsilon)$ to simplify $\alpha(G)$:

$$\begin{aligned} \alpha(G) &< \varepsilon^2 \left[\frac{n}{M(\varepsilon)} - 1\right] \\ &\leq \varepsilon^2 \left[\frac{n}{k} - 1\right] \\ &< \varepsilon^2 \left(\frac{n}{k}\right) \\ &\leq \varepsilon^2 m \end{aligned}$$

Proceed with two cases:

Case 1. Suppose there are more than $\frac{k^2}{4}$ edges in R. By Turán's Theorem, R contains a triangle. Let the vertices of said triangle in R correspond to the clusters V_i', V_j', and V_ℓ'

in G''. Let H be the graph consisting of the clusters A, B, and C in G, corresponding to V_i', V_j', and V_ℓ' respectively with the edges within the clusters put back.

By Proposition 2.1, all but at most $\varepsilon|A|$ vertices of A have at least $\varepsilon|B|$ neighbors in $|B|$ and all but at most $\varepsilon|A|$ vertices of A have at least $\varepsilon|C|$ neighbors in $|C|$. So, all but at most $2\varepsilon|A|$ vertices in A have at least $\varepsilon|B|$ neighbors in $|B|$ and $\varepsilon|C|$ neighbors in $|C|$. Choose a to be one of these vertices in A. Let the neighborhood of a in B and C be called B' and C' accordingly. Since (B, C) is an ε-regular pair, (B', C') is an ε-regular pair. So, all but at most $\varepsilon|B'| = \varepsilon^2|B|$ vertices in B' have at least $\varepsilon|C'| = \varepsilon^2|C|$ neighbors in C'. So, let b be such a vertex in B'. Let the neighborhood of b in C' be called C''. Since $\alpha(G) < \varepsilon^2 m = \varepsilon^2|C|$, there must be two adjacent vertices in C'', call them c_1 and c_2. Now, there must exist a K_4, consiting of vertices a, b, c_1, and c_2.
Case 2. Suppose there are at most $\frac{k^2}{4}$ edges in R. Notice,

$$\begin{aligned}\sum_{1\leq i<j\leq k} d(V_i, V_j) &= \frac{E(G'')}{m^2} \\ &\geq E(G'')\frac{k^2}{n^2} \\ &> \left(\frac{1}{8}+\varepsilon\right)k^2.\end{aligned}$$

Note that the last inequality follows from the observation that $E(G) > (\frac{1}{8}+\varepsilon)n^2$. So, the average of the densities which are not zero is at least:

$$d = \frac{(1/8+\varepsilon)k^2}{k^2/4} = \frac{1}{2} + 4\varepsilon.$$

Since the average is at least d, there is at least one pair (V_i, V_j) which has density greater than d. Let H be the graph consisting of this regular pair with the edges inside the two clusters put back, call this pair (A, B). Note that $|A| = |B| = m$. Notice that $\alpha(H) \leq \alpha(G) < \varepsilon m < 6\varepsilon m$. Let $\beta = d - \varepsilon = \frac{1}{2} + 3\varepsilon$, so $\alpha(H) < (2\beta - 1)m$. Notice that since $\alpha(H) < \varepsilon m$, there are at most εm independent vertices in A,so there exists at least $m - \varepsilon m$ adjacent vertices in A. Also notice that since $d \geq \frac{1}{2} + 4\varepsilon$, all but at most εm vertices in A are adjacent to at least $\frac{1}{2} + 4\varepsilon$ vertices in B. Choose two such vertices in A, and call them a_1 and a_2. Now, these vertices must have at least $4\varepsilon m$ of the same neighbors in B, call this set B'. Now, since $\alpha(H) < \varepsilon m$, at most εm vertices in B' are non-adjacent. So, there are at least two vertices in B' which are adjacent, call them b_1 and b_2. Now, there exists a K_4 as a subgraph of G, with vertices a_1, a_2, b_1, and b_2.

Since in both cases there exists a K_4 as a subgraph of G, there must be less than $\frac{1}{8}n^2 + o(n^2)$ edges in G. □

This result was proven to be optimal by Bollobás and Erdős in 1976 [3]. To prove this, they constructed a graph with $o(n)$ independent vertices and $\frac{1}{8}n^2 - o(n^2)$ edges.

Bibliography

[1] B. Bollobás, *Extremal Graph Theory*, Academic Press, London, 1978.

[2] ———, *Modern Graph Theory*, Springer, New York, 1998.

[3] B. Bollobás and P. Erdős, *On a Ramsey-Turán type problem*, J. Combin. Theory Ser. B **21** (1976), no. 2, 166–168.

[4] B. Chen, *What is combinatorics?*, available online: `http://www.math.ust.hk/~mabfchen/Math232/Introduction.pdf`.

[5] C. Chvatál, V. Rödl, E. Szemerédi, and W.T. Trotter, *The ramsey number of a graph with bounded maximum degree*, Journal of Combinatorial Theory **34** (1983), 239–243.

[6] R. Diestel, *Graph Theory*, third ed., Springer, New York, 2006.

[7] P. Erdős and A.H. Stone, *On the structure of linear graphs.*, Bulletin of the American Mathematical Society **52** (1946), 1087–1091.

[8] P. Erdős and P. Turán, *On some sequences of integers*, Journal of the London Mathematical Society **11** (1936), 261–264.

[9] K. Roth, *On certain sets of integers*, Journal of the London Mathematical Society **28** (1953), 101–109.

[10] E. Szemerédi, *On graphs containing no complete subgraph with* 4 *vertices*, Mat. Lapok **23** (1972), 113–116 (1973).

[11] ———, *On sets of integers containing no k elements in arithmetic progression*, Acta Arithmetica **27** (1975), 199–245, Collection of articles in memory of Juriĭ Vladimirovič Linnik.

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