Universitat Politècnica de Catalunya
Facultat de Matemàtiques i Estadística

Master in Advanced Mathematics and Mathematical Engineering

Master's thesis

# On the importance of details

**Severino Da Dalt**

Supervised by (Lluis Vena Cros)

August 29, 2025

Thanks to...

## Abstract

This should be an abstract in english, up to 1000 characters.

### Keywords

Stable Graphs, Graph Theory, Stability, VC-dimension, Szemerédi Regularity Lemma

# 1. Introduction

Szemerédi's Regularity Lemma [19] is a powerful tool in graph theory, stating that any sufficiently large graph can have its vertex set decomposed into an equitable partition such that most, but not all, pairs of parts are *regular*. A regular pair is one whose edge distribution resembles that of a random bipartite graph, a powerful property with many applications in extremal graph theory. On top of the presence of a small number of irregular pairs, a major drawback of the lemma is the immense bound on the required number of parts, which grows as a tower of exponentials whose height depends on the regularity parameter.

In the general setting, both limitations have been proven to be unavoidable. In [8] the author shows that there exist a family of graphs for which the lower bound on the number of parts is still a tower of exponentials[1]. On the other hand, it is folklore knowledge that large-enough half-graphs present irregular pairs in any regular partition ([1] gives a written proof of this fact). Having seen this unavoidability, it is natural to ask for the underlying reasons of those limitations and which additional conditions can be imposed or levied so that the parameters can be improved.

In this context, one of the first attempts was to make the regularity condition weaker so that the bound on the number of parts can be improved: this is now known as *the weak regularity lemma* [7] and, besides its own applications, allows to put the SzRL in the context of a spectra of regularity lemmas, with different strengths (the stronger the conclusion the larger the number of parts) [12, 11]; a notable example on the other direction (making the regularity lemma stronger) allows its use for Property Testing as it is suitable for the study of induced subgraphs [4].

Another option is to restrict the class of graphs where we want to find a regular partition. An example of this effort is the class of graphs with bounded VC-dimension: this concept was introduced in [20][2] and one can view it as a graph with "low complexity" (but not necessarily sparse), and the reader can find more details in Section 3. For these graphs the number of parts is highly reduced from a tower type to a polynomial in $1/\epsilon$, whose power depends on the bound on the dimension of the graph [11, 6, 2]. Even more, when the graphs have bounded VC-dimension, the density of edge in the regular pairs are either close to 1, or close to 0. However, the issue on the presence of irregular pairs remains, as any half-graph has bounded VC-dimension[3].

In this work we focus our attention on the result by Malliaris and Shelah [14, 13] which states that, if one cannot find a bi-induced copy of a half-graph, then a regular partition can be found, with not many pairs ($1/\epsilon$ to the power of an exponential on the size of the half-graph), and where no irregular pairs are found.

The graphs where no large half-graph can be found are called stable graphs, and their study stems from results in Model Theory and Logic. We shall stress that these stable graphs have, in fact, bounded VC-dimension (since we are forbidding a bi-induced copy of a fixed bipartite graph).

One of the many applications of the regularity lemma for which these bounds on the number of parts become relevant is in *property testing*. A property testing algorithm for a decision problem $P$ is a randomized algorithm that, by querying only a small portion of its input, can distinguish with high probability between

[1]To be more specific, the author shows that the number of parts is lower bounded by an exponenetial tower of 2's where the height of the tower is at least proportional to $\log(1/\epsilon)$. Meanwhile, in the usual proof of the theorem, the upper bound on the height of the tower is proportional to $\epsilon^{-5}$.

[2]See [21] for a translated version.

[3]Indeed, the fact that the neighbourhoods of the vertices on the same stable set can be ordered by inclusion, prevents the VC-dimension to grow beyond 2. Alternatively, in [11] it is shown that if a graph does not contain a bi-induced copy of a bipartite graph where the smaller size is k, then it has VC-dimension (strictly) bounded by k; in our case the half-graph has no bi-induced copy of $K_{3,3}$ minus a perfect matching.

objects that satisfy $P$ and those that are "far" from satisfying it. For instance, in [4] the authors use Szemerédi's Regularity Lemma to prove that it is possible to test the property of a graph $G$ being $H$-free (for a fixed graph $H$) using an algorithm which query complexity is independent on the size of the input graph $G$.

The query complexity of such testers, however, is intrinsically linked to the number of parts in the underlying regular partition. Consequently, the power-tower bounds of the standard regularity lemma lead to prohibitively large, although constant, query counts. This raises a natural question: can the superior bounds of the stable regularity lemma be exploited to create more efficient property testers for graphs in a half-graph-restricted setting?

In this thesis, we present an algorithm for testing $H$-freeness in stable graphs, thereby providing a concrete application that highlights the practical strength and utility of stable regularity partitions.

## 1.1 Main Contributions

The main contributions of this thesis are:

- We place a larger emphasis on the combinatorial part of the result in [14], making it self-contained and making some of the argument that previously used some Model Theory fully combinatorial. Further, we make some of the relations between the parameters explicit while correcting some of the typos that inevitably occur. In addition, we simplify some of the arguments, while making others more explicit and detailed. In particular, we make explicit that the excellence (see Section 5) depends on two parameters with opposite monotonic properties (see Definition 5.2 and Remark 5.5).
- The construction of an efficient property testing algorithm for H-freeness tailored to stable graphs. The algorithm's analysis leverages the stable regularity lemma to achieve a query complexity with significantly improved bounds compared to the general case.
- The development of a unified notational framework that cohesively integrates the concepts from extremal graph theory, stability, and property testing used throughout the thesis.

## 1.2 Summary

The remainder of this thesis is organized as follows. Section 2 reviews fundamental concepts from graph theory, culminating in a formal statement of Szemerédi's Regularity Lemma. Section 3 introduces the graph-theoretic notion of stability and proves some basic results in this context. Section 4 presents and analyzes some weaker variants of the stable regularity lemma, and illustrate both its strengths and its inherent limitations. Section 5 is dedicated to the proof of the main Stable Regularity Lemma, which forms the technical core of this work. Finally, Section 6 applies this previous results to prove our property testing algorithm for $H$-freeness in stable graphs works, providing explicit bounds on its query complexity.

# 2. Graphs and Regularity Lemma

Notation: - $\langle\cdot\rangle$ to represent tuples. - $\cdot^{<k}$. - $\frown$ and $\triangleleft$. - when we talk about equitable partitions, we mean +/-1.

Mention that during the thesis, a lot of results carry many conditions most of which seem almost trivial, but are necessary for the computations to work. In the final result of each section, the results are cleaned out and tried to be delivered in a more readable form.

Define neighbor?

## 2.1 Graphs and Basic Notation

In all this work we will consider only simple graphs, that is, unweighted, undirected graphs with no loops or multiple edges. The following definition accounts for this.

**Definition 2.1.** A (simple) *graph* is a pair $G = (V, E)$ where $V$ is a finite set whose elements are called *vertices* and $E \subseteq \{\{v_1, v_2\} \mid v_1, v_2 \in V \text{ and } v_1 \neq v_2\}$ is a set of unordered pairs of distinct vertices, whose elements are called *edges*. If $\{v_1, v_2\} \in E$, then $v_1$ and $v_2$ are said to be *the endpoints* of the edge.

By abuse of notation, we will often denote a graph $G = (V, E)$ simply by $G$ and write $v \in G$ to mean $v \in V$. Similarly, we will write $uv \in G$ to mean $\{u, v\} \in E$.

As most of this work is tightly related to model theory results, it is useful to note that vertices adjacency (two vertices being connected by an edge) is a symmetric and irreflexive binary relation on the vertex set. With this perspective, to denote vertex adjacency between twi vertices $v_1$ and $v_2$ we will often use the notation $v_1 R v_2$, where $R$ is the adjacency relation in $V$.

A class of graphs of particular relevance in this work is that of bipartite graphs.

**Definition 2.2.** A graph $G$ is *bipartite* if there exists a partition of its vertex set into two disjoint sets $L$ and $R$ such that every edge in $G$ connects a vertex in $L$ to a vertex in $R$. That is, no edge connects vertices within the same set of the partition.

Also, it is often useful to be able to restrict a graph to a subset of its vertices.

**Definition 2.3.** Let $G = (V, E)$ be a graph and let $S \subseteq V$ be a subset of its vertices. The *subgraph of $G$ induced by $S$*, denoted by $G[S]$, is the graph whose vertex set is $S$ and whose edge set consists of all edges in $E$ that have both endpoints in $S$. Formally, $G[S] = (S, E_S)$ where $E_S = \{\{v_1, v_2\} \in E \mid v_1, v_2 \in S\}$.

## 2.2 Szemerédi's Regularity Lemma

We now want to formalize the concept of regular pairs of vertex sets, which is central to Szemerédi's Regularity Lemma. The idea is that a pair of vertex sets is regular if the edges between them are "randomly" distributed, an idea that we can formalize using edge density.

**Definition 2.4.** Let $G$ be a graph and let $X, Y \subseteq G$ be two (not necessarily disjoint) subsets of its vertices. The *edge density* between $X$ and $Y$ is defined as

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

where $e(X, Y) = \{(x, y) \in E \mid x \in X, y \in Y\}$ is the set of edges with one endpoint in $X$ and the other in $Y$.

When $X$ and $Y$ are disjoint, the edge density $d(X, Y)$ measures the proportion of possible edges between $X$ and $Y$ that are actually present in the graph. If $X$ and $Y$ are not disjoint, this is not the case. On one hand, because simple graphs do not allow loops, and so edges between the same vertex are never present in

$e(X, Y)$, but they are counted in the denominator as "possible edges". On the other hand, edges between vertices in the intersection $X \cap Y$ are counted twice both in $e(X, Y)$ and $|X||Y|$. However, as this thesis deals with pairs of (non-necessarily different) parts of a partition of a vertex set, we only have to deal with two cases: either $X$ and $Y$ are disjoint, or $X = Y$. The first case has no problems, while for the second vase we note the following.

*Remark* 2.5. If $X$ is a subset of vertices of a graph $G$, then the proportion of possible edges between vertices in $X$ that are actually present in $G$ is larger than the density $d(X, X)$. That is,

$$\frac{|E_X|}{\binom{|X|}{2}} = \frac{|e(X,X)|/2}{(|X|-1)|X|/2} \geq \frac{|e(X,X)|}{|X|^2} = d(X,X)$$

Where first equality follows from the fact that $E_X$ counts each edge in $e(X, X)$ twice.

**Definition 2.6.** Given $\epsilon > 0$ and a graph $G$, a pair of (not necessarily disjoint) subsets of vertices $A, B \subseteq G$ is said to be *$\epsilon$-regular* if for all $A' \subseteq A$ and $B' \subseteq B$ such that $|A'| \geq \epsilon|A|$ and $|B'| \geq \epsilon|B|$, we have

$$|d(A', B') - d(A, B)| \leq \epsilon$$

Intuitively, this means that the edges of the pair are fairly uniformly distributed, and the pair behaves similarly to a random bipartite graph with edge density $d(A, B)$.

Now, this notion of regularity can be used in the context od a partition of a graph's vertex set. This partition allows a small number of pairs to be irregular. Also, a small *remainder set* is allowed, which is a set of vertices that are not included in any part of the partition.

**Definition 2.7.** Let $G$ be a graph and let $\epsilon > 0$. An *$\epsilon$-regular partition* of $G$ is a partition of its vertex set into $k$ parts $\{A_1, \dots, A_k\}$ with remainder set $B$ such that:

- $|B| \leq \epsilon|G|$, and may be empty.
- All but at most $\epsilon k^2$ of the pairs $(A_i, A_j)$ with $1 \leq i < j \leq k$ are $\epsilon$-regular.

Also, we say that the partition is *equitable* if $|A_1| = |A_2| = \cdots = |A_k|$.

The following is the celebrated Szemerédi's Regularity Lemma. The statement and proof we provide in this thesis follows the one given in [5], with minor notation modifications.

**Theorem 2.8** (Szemerédi's Regularity Lemma, [19])**.** *For every $\epsilon > 0$ and every positive integer $m$, there exists a positive integer $M = M(\epsilon, m)$ such that every graph with at least $m$ vertices admits an equitable $\epsilon$-regular partition $\{A_1, \dots, A_k\}$ and reminder $B$ with $m \leq k \leq M$.*

The principal strength of this lemma lies in the fact that it guarantees the existence of a regular partition which number of parts is independent of the size of the graph, and only depends on the regularity parameter $\epsilon$ and the minimum number of parts (and thus vertices) $m$.

Note that the lower bound on the number of parts $m$ can be used to increase the number of edges that connect different parts of the partition, over edges in the parts themselves. This is useful in some applications of the lemma.

Before proving the theorem, it is useful to introduce some additional notation and definitions. First, we note that the following inequality holds for any $\mu_1, \dots, \mu_k > 0$ and for all $e_1, \dots, e_k \geq 0$:

$$\sum_{i=1}^{k} \frac{e_i^2}{\mu_i} \geq \frac{(\sum_{i=1}^{k} e_i)^2}{\sum_{i=1}^{k} \mu_i} \tag{1}$$

This is a direct consequence of applying the Cauchy-Schwarz inequality $\sum a_i^2 \sum b_i^2 \geq (\sum a_i b_i)^2$ with the sequences $a_i = \sqrt{\mu_i}$ and $b_i = e_i/\sqrt{\mu_i}$.
A crucial concept in the proof of the Regularity Lemma is that of the *energy* of a partition.

**Definition 2.9.** Let $G$ be a graph with $n$ vertices and let $A_1, A_2$ be two disjoint subset of its vertex set. Then, we define

$$q(A_1, A_2) = \frac{|A_1||A_2|}{n^2} d(A_1, A_2)^2 = \frac{e(A_1, A_2)^2}{n^2|A_1||A_2|}$$

For a partition $\overline{A_1}$ of $A_1$ and $\overline{A_2}$ of $A_2$, we define

$$q(\overline{A_1}, \overline{A_2}) = \sum_{A_1' \in \overline{A_1}, A_2' \in \overline{A_2}} q(A_1', A_2')$$

Finally, we define the *energy* of a partition $\overline{A} = \{A_1, \dots, A_k\}$ of the vertex set of $G$ as

$$q(\overline{A}) = \sum_{1 \leq i < j \leq k} q(A_i, A_j)$$

Let $\overline{A}$ be a partition with reminder set $B$, we define $\widetilde{A} := \overline{A} \cup \overline{B}$, and we use $\overline{B}$ to denote the set of singletons of the remainder set, $\overline{B} := \{\{b\} \mid b \in B\}$. Then, $q(\widetilde{A}) = q(\overline{A} \cup \overline{B})$
The energy function $q$ i the main tool to prove the Regularity Lemma. We will prove that, if a given partition has a large enough number of irregular pairs, taking the associated parts and make them a partition of their own, results in a refinement of the original partition with a large increase in energy, which only depends on the regularity parameter $\epsilon$. But the energy of a partition is upper bounded:

$$\begin{aligned} q(\widetilde{A}) &= \sum_{\substack{C_1, C_2 \in \widetilde{A} \\ C_1 \neq C_2}} q(C_1, C_2) \\ &= \sum_{\substack{C_1, C_2 \in \widetilde{A} \\ C_1 \neq C_2}} \frac{|C_1||C_2|}{n^2} d(C_1, C_2)^2 \\ &\leq \frac{\sum |C_1||C_2|}{n^2} \leq 1 \end{aligned}$$

Hence, the process of refinement must stop after a finite number of steps.
We first prove that refining a pair of parts or a whole partition does not decrease its energy.

**Lemma 2.10.** *Let $G$ be a graph.*

1. *Let $A_1, A_2 \subseteq G$ be disjoint. If $\overline{A}_1$ is a partition of $A_1$ and $\overline{A}_2$ is a partition of $A_2$, then $q(\overline{A}_1, \overline{A}_2) \geq q(A_1, A_2)$.*

2. *If $\overline{A}, \overline{A}'$ are partitions of $G$ and $\overline{A}'$ is a refinement of $\overline{A}$, then $q(\overline{A}') \geq q(\overline{A})$.*

*Proof.* 1. Let $\overline{A}_1 = \{A_{1,1}, \dots, A_{1,k}\}$ and $\overline{A}_2 = \{A_{2,1}, \dots, A_{2,\ell}\}$. Then

$$\begin{aligned} q(\overline{A}_1, \overline{A}_2) &= \sum_{i=1}^{k} \sum_{j=1}^{\ell} q(A_{1,i}, A_{2,j}) \\ &= \frac{1}{n^2} \sum_{i=1}^{k} \sum_{j=1}^{\ell} \frac{e(A_{1,i}, A_{2,j})^2}{|A_{1,i}||A_{2,j}|} \\ &\overset{(1)}{\geq} \frac{1}{n^2} \frac{\left(\sum_{i=1}^{k} \sum_{j=1}^{\ell} e(A_{1,i}, A_{2,j})\right)^2}{\sum_{i=1}^{k} \sum_{j=1}^{\ell} |A_{1,i}||A_{2,j}|} \\ &= \frac{1}{n^2} \frac{e(A_1, A_2)^2}{(\sum_{i=1}^{k} |A_{1,i}|)(\sum_{j=1}^{\ell} |A_{2,j}|)} \\ &= q(A_1, A_2) \end{aligned}$$

2. Let $\overline{A} = \{A_1, \dots, A_k\}$, and for all $i \in \{1, \dots, k\}$ let $\overline{A}_i$ be the partition of $A_i$ induced by $\overline{A}'$. Then,

$$\begin{aligned} q(\overline{A}) &= \sum_{1 \leq i < j \leq k} q(A_i, A_j) \\ &\overset{1.}{\leq} \sum_{1 \leq i < j \leq k} q(\overline{A}_i, \overline{A}_j) \\ &\leq q(\overline{A}') \end{aligned}$$

where last inequality follows from the fact that $q(\overline{A}') = \sum_{1 \leq i \leq k} q(\overline{A}_i) + \sum_{1 \leq i < j \leq k} q(\overline{A}_i, \overline{A})$.

□

Next, we show that refining an irregular pair results in a significant increase in energy. This amount, does not yet depend only on $\epsilon$, but it will when applied to all irregular pairs at the same time.

**Lemma 2.11.** *Let $G$ be a graph with $n$ vertices, $A_1, A_2 \subseteq G$ be disjoint subsets and $\epsilon > 0$. If the pair $(A_1, A_2)$ is not $\epsilon$-regular, then there exist partitions $\overline{A}_1 = \{A_{1,1}, A_{1,2}\}$ of $A_1$ and $\overline{A}_2 = \{A_{2,1}, A_{2,2}\}$ of $A_2$ such that*

$$q(\overline{A}_1, \overline{A}_2) \geq q(A_1, A_2) + \epsilon^4 \frac{|A_1||A_2|}{n^2}$$

*Proof.* Suppose that $(A_1, A_2)$ is not $\epsilon$-regular. Then there are subsets $A_{1,1} \subseteq A_1$ and $A_{2,1} \subseteq A_2$ with $|A_{1,1}| \geq \epsilon |A_1|$ and $|A_{2,1}| \geq \epsilon |A_2|$ such that

$$|\eta| > \epsilon \tag{2}$$

where $\eta = d(A_{1,1}, A_{2,1}) - d(A_1, A_2)$. We now show that $\overline{A}_1 = \{A_{1,1}, A_{1,2}\}$ and $\overline{A}_2 = \{A_{2,1}, A_{2,2}\}$, where $A_{1,2} \coloneqq A_1 \setminus A_{1,1}$ and $A_{2,2} \coloneqq A_2 \setminus A_{2,1}$, satisfy the statement.

For ease of notation, we write $c_i \coloneqq |A_{1,i}|$, $d_i \coloneqq |A_{2,i}|$, $e_{ij} \coloneqq e(A_{1,i}, A_{2,j})$, $c \coloneqq |A_1|$, $d \coloneqq |A_2|$ and

$e = e(A_1, A_2)$. Then, we have

$$\begin{aligned} q(\overline{A}_1, \overline{A}_2) &= \frac{1}{n^2} \sum_{i=1}^{2} \sum_{j=1}^{2} \frac{e_{ij}^2}{c_i d_j} \\ &= \frac{1}{n^2} \left( \frac{e_{11}^2}{c_1 d_1} + \sum_{i+j>2} \frac{e_{ij}^2}{c_i d_j} \right) \\ &\overset{(1)}{\geq} \frac{1}{n^2} \left( \frac{e_{11}^2}{c_1 d_1} + \frac{(e - e_{11})^2}{cd - c_1 d_1} \right) \end{aligned}$$

By definition of $\eta$, in the new notation we have that $e_{11} = \frac{c_1 d_1 e}{cd} + \eta c_1 d_1$, and so

$$\begin{aligned} n^2 q(\overline{A}_1, \overline{A}_2) &\geq \frac{1}{c_1 d_1} \left( \frac{c_1 d_1 e}{cd} + \eta c_1 d_1 \right)^2 + \frac{1}{cd - c_1 d_1} \left( e - \frac{c_1 d_1 e}{cd} - \eta c_1 d_1 \right)^2 \\ &\geq \frac{1}{c_1 d_1} \left( \frac{c_1 d_1 e}{cd} + \eta c_1 d_1 \right)^2 + \frac{1}{cd - c_1 d_1} \left( \frac{cd - c_1 d_1}{cd} e - \eta c_1 d_1 \right)^2 \\ &= \frac{c_1 d_1 e^2}{c^2 d^2} + \frac{2e\eta c_1 d_1}{cd} + \eta^2 c_1 d_1 + \frac{(cd - c_1 d_1) e^2}{c^2 d^2} - \frac{2e\eta c_1 d_1}{cd} + \frac{\eta^2 c_1^2 d_1^2}{cd - c_1 d_1} \\ &\geq \frac{e^2}{cd} + \eta^2 c_1 d_1 \\ &\overset{(2)}{\geq} \frac{e^2}{cd} + \epsilon^4 cd = n^2 q(A_1, A_2) + \epsilon^4 cd \end{aligned}$$

and we obtain the inequality from the statement by simply dividing by $n^2$ at each side of the inequality. □

The next lemma show that applying the previous lemma to all irregular pairs of a partition achieves the desired constant increase in energy.

**Lemma 2.12.** *Let $0 < \epsilon \leq \frac{1}{4}$, let $G$ be a graph with $n$ vertices, and let $\overline{A} = \{A_1, \dots, A_k\}$ be a partition of its vertex set with remainder set $B$ such that $|B| \leq \epsilon n$ and $|A_1| = \dots = |A_k| =: c$. If the partition $\overline{A}$ is not $\epsilon$-regular, then there is a refinement $\overline{A}' = \{A'_1, \dots, A'_\ell\}$ of $\overline{A}$ with remainder set $B'$ such that $k \leq \ell \leq k4^{k+1}$, $|A'_0| \leq |A_0| + \frac{n}{2^k}$, $|A'_1| = \dots = |A'_\ell|$ and either $\overline{A}'$ is $\epsilon$-regular, or*

$$q(\widetilde{A}') \geq q(\widetilde{A}) + \frac{\epsilon^5}{2}$$

*Proof.* For all $1 \leq i < j \leq k$, let $\overline{A}_{ij}$ be a partition of $A_i$ and $\overline{A}_{ji}$ a partition of $A_j$ as follows. If the pair $(A_i, A_j)$ is $\epsilon$-regular, then $\overline{A}_{ij} := \{A_i\}$ and $\overline{A}_{ji} := \{A_j\}$. Otherwise, we can apply Lemma 2.11 to obtain a partition $\overline{A}_{ij}$ of $A_i$ and a partition $\overline{A}_{ji}$ of $A_j$ with $|\overline{A}_{ij}| = |\overline{A}_{ji}| = 2$ such that

$$q(\overline{A}_{ij}, \overline{A}_{ji}) \geq q(A_i, A_j) + \epsilon^4 \frac{c^2}{n^2} \tag{3}$$

Now, consider two vertices $u, v \in A_i$ to be equivalent if for every $j \neq i$ they belong to the same set of the partition $\overline{A}_{ij}$. We can define $\overline{A}_i$ to be the set of such equivalence classes. Then, each partition $\overline{C}_{ij}$ may at most double the number of parts of $\overline{A}_i$, we have that $|\overline{A}_i| \leq 2^{k-1}$. Putting all of this together, we have a new partition

$$\overline{A}'' := \bigcup_{i=1}^{k} \overline{A}_i$$

of $G$ with reminder set still $B$. Note that $\overline{A}''$ refines $\overline{A}$, and that

$$k \leq |\overline{A}''| \leq k2^{k-1} \leq k2^k \tag{4}$$

By hypothesis, we now that $\overline{A}$ is not $\epsilon$-regular, and so there are at least $\epsilon k^2$ pairs $(A_i, A_j)$ with $1 \leq i < j \leq k$, the partition $\overline{A}_{ij}$ is non-trivial. Thus,

$$\begin{aligned}
q(\widetilde{A}'') &= \sum_{1\leq i<j\leq k} q(\overline{A}_i, \overline{A}_j) + \sum_{1\leq i\leq k} q(\overline{A}_i, \overline{B}) + \sum_{1\leq i\leq k} q(\overline{A}_i) + q(\overline{B}) \\
&\geq \sum_{1\leq i<j\leq k} q(\overline{A}_{ij}, \overline{A}_{ji}) + \sum_{1\leq i\leq k} q(\{A_i\}, \overline{B}) + q(\overline{B}) \\
&\overset{(3)}{\geq} \sum_{1\leq i<j\leq k} q(A_i, A_j) + \epsilon k^2 \epsilon^4 \frac{c^2}{n^2} + \sum_{1\leq i\leq k} q(\{A_i\}, \overline{B}) + q(\overline{B}) \\
&= q(\widetilde{A}) + \epsilon^5 (\frac{ck}{n})^2 \\
&\geq q(\widetilde{A}) + \frac{\epsilon^5}{2}
\end{aligned}$$

First equality follows from the definition of energy, first inequality uses 1. from Lemma 2.10 and last inequality follows from the fact that $|B| \leq \epsilon n \leq \frac{1}{4}$, so $kc$ is necessarily at least $\frac{3}{4}n$.

Finally, we need to turn $\overline{A}''$ into an equitable partition. In order to achieve this, we will split refine each part into pieces of equal size, and move the remaining vertices to the reminder set. We need to separate two cases, as we may not have enough vertices to make substantially sized parts.

If $c < 4^k$, we just consider all the parts to be singletons, and keep the reminder set $B$ as it is. Since there are at most $k$ parts in $\overline{A}$, we have that the resulting partition $\overline{A}'$ of size $\ell$ satisfies $k \leq \ell = kc < k4^k$.

Otherwise, if $c \geq 4^k$, consider $A'_1, \dots, A'_\ell$ to be a maximal collection of disjoint sets of size $d := \lfloor \frac{c}{4^k} \rfloor \geq 1$ such that each $A'_i$ is contained in some part of $\overline{A}''$. Then, the remainder set $B'$ is obtained by adding to $B$ all the remaining vertices from all the parts of $\overline{A}''$, or simply $B' = G \setminus \bigcup_{i=1}^{\ell} A'_i$.

The resulting partition $\overline{A}' = \{A'_1, \dots, A'_\ell\}$ is a refinement of $\overline{A}''$ and, following 2. from Lemma 2.10, satisfies

$$q(\widetilde{A}') \geq q(\widetilde{A}'') \geq q(\widetilde{A}) + \frac{\epsilon^5}{2}$$

Now, no more than $\frac{c}{d} \leq 4^{k+1}$ sets $A'_i$ can lie within the same part of $\overline{A}''$, so the condition $k \leq \ell k 4^{k+1}$ is satisfied. Also, no more than $d$ vertices are left out from each part of $\overline{A}''$, and so

$$\begin{aligned}
|B'| &\leq |B| + d|\overline{A}''| \\
&\overset{(4)}{\leq} |B| + \frac{c}{4^k} k2^k \\
&= |B| + \frac{kc}{2^k} \\
&\leq |B| + \frac{n}{2^k}
\end{aligned}$$

Thus, the partition $\overline{A}'$ with remainder set $B'$ satisfies all the conditions in the statement, and we are done. □

We now have all the tools required to prove Szemerédi's Regularity Lemma. The idea will be to start with an arbitrary equitable partition, with a large enough number of parts and small enough reminder set, and then keep refining it until we reach a regular partition. Then, reaching regularity is inevitable, as the previous result guarantees a constant increase in energy which we previously proved to be upper bounded.

*Proof of Theorem 2.8.* Let $\epsilon > 0$, $m \geq 1$ and assume without loss of generality that $\epsilon \leq \frac{1}{4}$. This is possible by monotonicity of the regularity condition. Also, set $s := \frac{2}{\epsilon^5}$.
While refining repeatedly the partition using Lemma 2.12, ($s$ times) we need to make sure that the remainder set does not grow too large, as the lemma requires it to be at most $\epsilon n$. At each refinement, the size of the reminder set increases by at most $\frac{n}{2^k}$, where $k$ is the number of parts of the partition before refining. Since at each iteration the number of parts can only increase, each time at most $\frac{n}{2^m}$ vertices are added to the reminder set. By choosing $k$ and $n$ large enough, we can ensure that the initial size of the remainder set and the total growth of it over all the $s$ steps is at most $\frac{\epsilon n}{2}$ each.
With this in mind, we choose $k$ large enough to satisfy $\frac{s}{2^k} \leq \frac{\epsilon}{2}$, and $n$ large enough so that $k \leq \frac{\epsilon n}{2}$. Then,

$$k + \frac{sn}{2^k} \leq \frac{\epsilon n}{2} + \frac{\epsilon n}{2} = \epsilon n \tag{5}$$

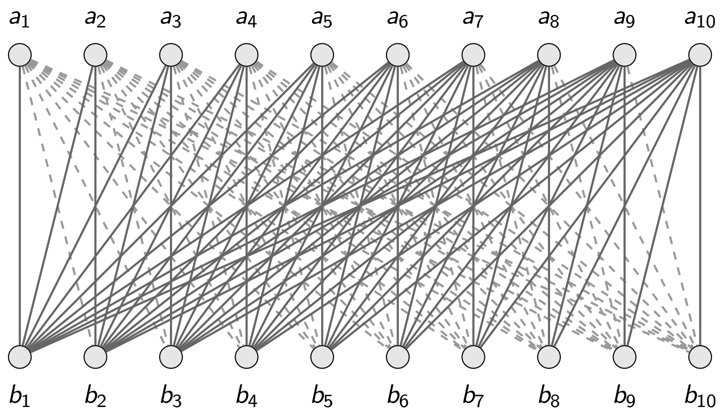
Now, let's bound the number of parts of the partition at end of the process. Since at each step the number of parts goes from $r$ up to at most $r4^{r+1}$, starting with $k$ parts, we can simply set $M := \max(f^s(k), 2\frac{k}{\epsilon})$, where $f(r) = r4^{r+1}$. The second term ensures that $n \geq M$ is sufficiently large for (5) to hold.
Now, given a graph $G$ with $n \geq m$ vertices, we can build a partition into $k'$, with $m \leq\leq M$ parts with remainder $B$ as follows. If $n \leq M$, simply take the partition to be all the vertices as singletons, and the remainder set to be empty. The resulting partition is trivially $\epsilon$-regular, as pairs of singletons are always either complete or empty. Suppose now that $n > M$. We randomly partition the vertex set of $G$ into parts of size $k$, and put the remaining vertices in the remainder set. This remainder set has size at most $k - 1 < \epsilon n$ by (5). We now can apply Lemma 2.12 repeatedly, as the choice of $k$ and $n \geq M$ in (5) ensures that the reminder is at most $\epsilon n$ during $s$ steps. But this process must stop in at most $s$ steps, as the energy of the partition increases by at least $\frac{\epsilon^5}{2}$ at each step, so after $s$ steps the energy would be at least 1, which is the theoretical maximum as shown earlier. □

# 3. Stable Graphs

In this section we introduce the class of *stable* graphs. A graph is considered stable, if it does not contain bi-induced (as defined in [15]) large half-graphs, a particularly non-quasi-random structure in graphs. See Figure 1 for an example of such a graph.

Maybe move cite to section 2, and mention there other references such as "induced subbigraph"

Explain somewhere what this means.




| | $b_1$ | $b_2$ | $b_3$ | $b_4$ | $b_5$ | $b_6$ | $b_7$ | $b_8$ | $b_9$ | $b_{10}$ |
|---|---|---|---|---|---|---|---|---|---|---|
| $a_1$ | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| $a_2$ | 1 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| $a_3$ | 1 | 1 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| $a_4$ | 1 | 1 | 1 | 1 | 0 | 0 | 0 | 0 | 0 | 0 |
| $a_5$ | 1 | 1 | 1 | 1 | 1 | 0 | 0 | 0 | 0 | 0 |
| $a_6$ | 1 | 1 | 1 | 1 | 1 | 1 | 0 | 0 | 0 | 0 |
| $a_7$ | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 0 | 0 | 0 |
| $a_8$ | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 0 | 0 |
| $a_9$ | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 0 |
| $a_{10}$ | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |

Figure 1: A half-graph with $2 \times 10$ vertices. *On the left*, solid lines show adjacent vertices, and dashed lines show non-adjacent vertices. Pairs of vertices without a line may or may not be connected. *On the right* is the corresponding adjacency matrix.

First, stability implies a bounded *Vapnik-Chervonenkis (VC) dimension*, which limits the variety of neighborhoods of vertices within the graph. While stability implies a bounded VC-dimension for the entire graph (See [11]), our work primarily focuses on bounding the VC-dimension restricted to a subset of vertices. This is formalized in Lemma 3.10.

Second, stability implies a finite *tree bound*. This property is the foundational tool we use to prove the existence of parts that are quasi-random with respect to the rest of the graph. We use this to establish the existence of indivisible parts in Section 4 (Lemma 4.12) and excellent parts in Section 5 (Lemma 5.6).

## 3.1 $k$-order Property

First, we formally define stability as the non-$k$-order property, where $k$ determines the size of the excluded half-graphs.

**Definition 3.1.** Let $G$ be a graph. We say that $G$ has the *$k$-order property* if there exist two sequences of vertices $\langle a_i \mid i \in \{1, \dots, k\} \rangle$ and $\langle b_i \mid i \in \{1, \dots, k\} \rangle$ such that for all $i, j \leq k$, $a_i R b_j$ if and only if $i \geq j$. Otherwise, we say that $G$ has the *non-$k$-order property* or that $G$ is *$k$-stable*.

*Remark* 3.2. It is important to note what is left unspecified in Definition 3.1. First, the vertices within each sequence must be distinct, as their neighborhoods within the other sequence differ. However, the sequences themselves need not be disjoint. One may have $a_i = b_j$, provided $i < j$ (so that $\neg(a_i R b_j)$). Furthermore, the definition does not specify the presence or absence of edges within the same sequence. Consequently, the non-$k$-order property requires the containment of a subgraph from a broad class of structures, not merely a k-half-graph.

Possibly add visual example of this too.

*Remark* 3.3. $G$ having the $k$-order property implies that $G$ has the $k'$-order property for all $k' \leq k$. Conversely, $G$ having the non-$k$-order property implies that $G$ has the non-$k'$-order property for all $k' \geq k$.

An important concept used all over the thesis is that of *exceptional edges* and *exceptional vertices*. That is, edges and vertices that, in the context of a pair of sets of vertices, do not "behave" as the rest. In order to classify what is the expected behaviour in a graph, or more specifically, in a pair of sets of vertices, we define the *truth value*.

Echarle un ojo.

**Definition 3.4** (Truth value)**.** Let $G$ be a graph. For any (not necessarily disjoint) $A, B \subseteq G$, we say that

$$t(A,B) = \begin{cases} 0 & \text{if } |\{(a,b) \in A \times B \mid aRb, a \neq b\}| < |\{(a,b) \in A \times B \mid \neg aRb, a \neq b\}| \\ 1 & \text{otherwise} \end{cases}$$

is the *truth value* of the pair $(A, B)$. That is, $t(A, B) = 0$ if $A$ and $B$ are mostly disconnected, and $t(A, B) = 1$ if they are mostly connected. When $B = \{b\}$, we write $t(A, b)$ instead of $t(A, \{b\})$, and we say that it is the truth value of $A$ with respect to $b$.

In this context, we say that a vertex $a \in A$ is *exceptional* with respect to $B \subseteq G$ if $t(a, B) \not\equiv t(A, B)$, or that it is exceptional with respect to $b \in G$ if $aRb \not\equiv t(A, b)$. On the other hand, we say that an edge $ab$ with $a \in A$ and $b \in B$ is exceptional in $(A, B)$ if $aRb \not\equiv t(A, B)$. Also, it is useful to define the following set of vertices.

- $B_{A,b} = \{a \in A \mid aRb \equiv t(A, b)\}$, i.e. the set of non-exceptional vertices of $A$ with respect to $B$.
- $\overline{B}_{A,b} = \{a \in A \mid aRb \not\equiv t(A, b)\}$, the set of exceptional vertices of $A$ with respect to $B$.
- $B^+_{A,b} = \{a \in A \mid aRb\}$, the vertices of $A$ connected to $b$.
- $B^-_{A,b} = \{a \in A \mid \neg aRb\}$, the vertices of $A$ that are not connected to $b$.

With this notation, notice that either $t(A, b) = 1$ and thus $B_{A,b} = B^+_{A,b}$, or $t(A, b) = 0$ and $B_{A,b} = B^-_{A,b}$. Sets of vertices $A$ with a large number of large $\overline{B}_{A,b}$ are a great obstacle towards creating a quasi-random, and more specifically homogeneous partition, as the number of exceptional edges with respect to the entire graph is large and concentrated. A useful tool to deal with them is Lemma 3.10, which gives a bound on the number of such sets under the non-$k$-order property. In order to prove it, we first need to introduce the *VC dimension* of a family of sets, and relate it to the $k$-order property. This, together with Lemma 3.7, will give us the desired result.

Define homogeneous

**Definition 3.5.** Let $G$ be a set and $S = \{S_i \subseteq G \mid i \in I\}$ be a family of sets. A set $A \subseteq G$ is said to be *shattered* by $S$ (and $S$ is said to *shatter* $A$) if for every $B \subseteq A$, there exists $S_i \in S$ such that $S_i \cap A = B$.

**Definition 3.6.** Let $G$ be a set and $S = \{S_i \subseteq G \mid i \in I\}$ be a family of sets. The *VC dimension* of $S$ is the size of the largest set $A \subseteq G$ that is shattered by $S$.

**Lemma 3.7** (Sauer-Shelah (-Perles -Vapnik-Chervonenkis) Lemma, [17], [18])**.** *Let $G$ be a set and $S = \{S_i \subseteq G \mid i \in I\}$ be a family of sets. If the VC dimension of $S$ is at most $k$, and the union of all the sets in $S$ has $n$ elements, then $S$ consists of at most $\sum_{i=0}^{k} \binom{n}{i} \leq n^k$ sets.*

We'll begin by proving a stronger version of this lemma from Pajor, for which Sauer-Shelah will be a straightforward consequence.

**Lemma 3.8** (Pajor's variant, [16]). *Let $G$ be a set and $S$ be a finite family of sets in $G$. Then $S$ shatters at least $|S|$ sets.*

*Proof.* We will prove this by induction on the cardinality of $S$. If $|S| = 1$, then $S$ consists of a single set, which only shatters the empty set. If $|S| > 1$, we may choose an element $x \in S$ such that some sets of $S$ contain $x$ and some do not. Let $S^+ = \{s \in S \mid x \in S\}$ and $S^- = \{s \in S \mid x \notin S\}$. Then $S = S^+ \sqcup S^-$, and both $S^+$ and $S^-$ are non-empty. By induction hypothesis, we know that $S^+ \subsetneq S$ shatters at least $|S^+|$ sets, and $S^- \subsetneq S$ shatters at least $|S^-|$ sets. Let $T$, $T^+$, $T^-$ be the families of sets shattered by $S$, $S^+$ and $S^-$ respectively. To conclude the proof, we just need to show that for each element in $T^+$ and $T^-$, there is a corresponding one in $T$. If a set is shattered by only one of the two families $S^+$ and $S^-$, then it only contributes by one unit to $|T^+| + |T^-|$ and one unit to $|T|$. Notice that no set shattered by $S^+$ or $S^-$ may contain $x$, otherwise all or none of the intersections will contain this element. Thus, if a set $s$ is shattered by both $S^+$ and $S^-$, it will contribute by two units to $|T^+| + |T^-|$ and one unit to $|T|$. But then, for each such set, we can consider $s \cup \{x\}$ which is not in $T^+$ or $T^-$, but it is in $T$. Indeed, for each subset of $s$, if it does not contain $x$ it is the intersection with some set in $S^- \subsetneq S$, and if it does contain $x$ it is the intersection with some set in $S^+ \subsetneq S$. All in all, we conclude that

$$|T| \geq |T^+| + |T^-| \geq |S^+| + |S^-| \geq |S|$$

□

*Proof of Lemma 3.7.* Suppose that $\bigcup S$ has $n$ elements. By Lemma 3.8, $S$ shatters at least $|S|$ subsets, and since there are at most $\sum_{i=0}^{k} \binom{n}{i}$ subsets of $S$ of size at most $k$, if $|S| > \sum_{i=0}^{k} \binom{n}{i}$, at least one of the shattered sets has cardinality larger than $k$, and hence the VC dimension of $S$ is larger than $k$. □

Next, we want to prove that if $G$ has the non-$k$-order property, then the size of the family of exceptional sets of $A$, relative to each vertex $b \in G$, is bounded by $|A|^k$. Instead, we prove a stronger result, that is we prove this same bound with only the condition that $G$ has the "disjoint" non-$k$-order property, in which the two sequences of vertices in the Definition 3.1 are in fact disjoint. This stronger version (Lemma 3.10) is neither more useful nor easier to prove, but remarks that the non-disjointness of the sequences, and thus the broadening of the excluded structures, is not needed to obtain the bound, but later on.

**Lemma 3.9.** *Let $G$ be a graph and $A \subseteq G$. Let $S = \{B^+_{A,b} \mid b \in G \setminus A\}$. If $S$ has VC dimension (at least) $k$, then $G$ has the (disjoint) $k$-order property.*

Lluis: faig una definició separada o s'enten pel context que ja he posat?

*Proof.* If $S$ has VC dimension $k$, then it shatters a set $A' \subseteq A$ of size $k$. Now, choose any order of the vertices of $A' = \langle a_1, \dots, a_k \rangle$. Then, consider the increasing sequence of subsets $A_1 \subseteq A_2 \subseteq \cdots \subseteq A_k = A'$, where $A_i = \{a_j \mid j \in \{1, \dots, i\}\}$. Since $A'$ is shattered by $S$, for each $i \in \{1, \dots, k\}$ there exists a $b_i \in G$ such that $b_i R a$ if and only if $a \in A_i$. In particular, the two sequences $\langle a_i \mid i \in \{1, \dots, k\} \rangle$ and $\langle b_i \mid i \in \{1, \dots, k\} \rangle$ satisfy

$$a_i R b_j \Leftrightarrow i \leq j$$

and thus $G$ has the $k$-order property. □

**Lemma 3.10** (Claim 2.6 in [14]). *Let $G$ be a graph with the (disjoint) non-$k$-order property. Then, for any finite non-trivial $A \subseteq G$,*

$$|\{B^+_{A,b} \mid b \in G\}| \leq |A|^k$$

*Proof.* By Lemma 3.9, if $G$ has the non-$k$-order property, then the family $\{B^+_{A,b} \mid b \in G \setminus A\}$ has VC dimension at most $k-1$, so by the Sauer-Shelah Lemma 3.7 we have $\{B^+_{A,b} \mid b \in G \setminus A\} \leq \sum_{i=0}^{k-1} \binom{|A|}{i}$. Since $\{B^+_{A,b} \mid b \in A\} \leq |A|$, we conclude that

$$|S| = |\{B^+_{A,b} \mid b \in G\}| \leq \sum_{i=0}^{k-1} \binom{|A|}{i} + |A|$$

This conditions should be set at some point of the tfm. Specify that if they are not met, the problem becomes trivial.

Finally, when $|A| = n, k > 1$:

- if $n \leq k$, then $|S| \leq 2^n \leq 2^k \leq n^k$.
- if $n > k$, then $|S| \leq \sum_{i=0}^{k-1} \binom{n}{i} + n \leq n^{k-1} + n \leq 2n^{k-1} \leq n^k$.

We conclude that $|S| \leq n^k$. □

*Remark* 3.11. The condition $n, k > 1$ is trivial. If $n = 1$ then $A$ is the trivial graph with a single vertex. If $k = 1$ we are not allowing even a single edge, so $G$ is the empty graph.

We now prove the following equivalent versions of the lemma, which will be useful in the different sections of the thesis. The idea is that any choice of either the exceptional or the non-exceptional vertices set of $A$ with respect to each vertex $b \in G$, have the same bound.

**Corollary 3.12** (Claim 2.6.1). *Let $G$ be a graph with the non-$k$-order property. Then:*

1. *For any finite $A \subseteq G$*
$$|\{B^-_{A,b} \mid b \in G\}| \leq |A|^k$$
2. *For any finite $A \subseteq G$*
$$|\{\overline{B}_{A,b} \mid b \in G\}| \leq |A|^k$$

*Proof.* 1. First of all, notice that $B^+_{A,b} = A - B^-_{A,b}$, since by definition they are complementary. Thus, for any $b, b' \in G$, $B^+_{A,b} = B^+_{A,b'} \Leftrightarrow B^-_{A,b} = B^-_{A,b'}$. It follows that

$$|\{B^-_{A,b} \mid b \in G\}| = |\{B^+_{A,b} \mid b \in G\}| \leq |A|^k$$

where the last inequality follows from Lemma 3.10.

2. Consider the following map:

$$\begin{aligned} \pi : \{B^+_{A,b} \mid b \in G\} &\longrightarrow \{\overline{B}_{A,b} \mid b \in G\} \\ B^+_{A,b} &\longmapsto \overline{B}_{A,b} \end{aligned}$$

We first prove that the map $\pi$ is well-defined. If $B^+_{A,b}$ and $B^+_{A,b'}$ are equal, then they have the same size, and thus the same truth value. Then,

- if $t(A,b) = t(A,b') = 1$, we have that $\overline{B}_{A,b} = B^+_{A,b} = B^+_{A,b'} = \overline{B}_{A,b'}$.
- if $t(A,b) = t(A,b') = 0$, we have that $\overline{B}_{A,b} = B^-_{A,b} = A \setminus B^+_{A,b} = A \setminus B^+_{A,b'} = B^-_{A,b'} = \overline{B}_{A,b'}$.

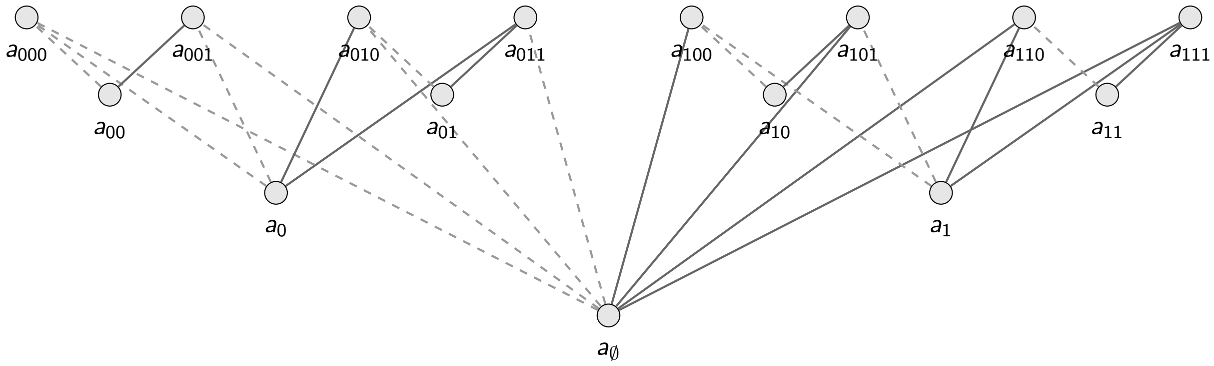



Figure 2: Example of a 3-tree. Notice that connections between disjoint sub-trees are not defined, and may be edges or non-edges in any combination.

which proves that the map is well-defined. The map $\pi$ is also surjective, since for each $b \in G$, and thus for each $\overline{B}_{A,b}$, the set $B^+_{A,b}$ is mapped to $\overline{B}_{A,b}$ by construction. Hence,

$$|\{\overline{B}_{A,b} \mid b \in G\}| \leq |\{B^+_{A,b} \mid b \in G\}| \leq \sum_{i \leq k} \binom{|A|}{i} \leq |A|^k$$

This concludes the proof. Notice that, actually, the map $\pi$ is a not necessarily a bijection, since (at most) two $b$'s with different truth value with respect to $A$ may induce the same set $\overline{B}_{A,b}$.

□

## 3.2 Tree Bound

During the next sections, it will be a key point proving that some sort of "regular" subgraphs (*independent* in Section 4 and *excellent* in Section 5) exist in a given stable graph. In order to do so, a useful structure strongly related to the $k$-order property is the $k$-*tree*.

**Definition 3.13.** A $k$-*tree* in $G$ is an ordered pair $H = (\overline{c}, \overline{b})$ comprising:

- $\overline{c} = \{c_\eta \in G \mid \eta \in \{0,1\}^{<k_{**}}\}$, the set of *nodes*.
- $\overline{b} = \{b_\rho \in G \mid \rho \in \{0,1\}^{k_{**}}\}$, the set of *branches*.

satisfying that, for all $\eta \in \{0,1\}^{<k_{**}}$ and $\rho \in \{0,1\}^{k_{**}}$, if given $\ell \in \{0,1\}$ we have $\eta \frown \langle \ell \rangle \triangleleft \rho$, then $(b_\rho R c_\eta) \equiv (\ell = 1)$. The two sequences are not necessarily disjoint. Lluis: millor una remark?

See Figure 2 for an example of such a structure.
Similarly to stability, we can define the *tree bound* of a graph to measure the level of freeness from $k$-trees of graph.

**Definition 3.14** (Definition 2.11)**.** Suppose $G$ is a finite graph. We denote the *tree bound* $k_{**} = k_{**}(G)$ as the minimal positive integer such that there is no $k_{**}$-tree $H = (\overline{c}, \overline{b})$ in $G$.

As mentioned earlier, the tree bound is closely related to the $k$-order property. The following theorem states that if a graph has a sufficiently large tree bound, then it has the $k$-order property and vice versa.

**Theorem 3.15** (Lemma 6.7.9 in [9]). *If a graph $G$ has the $2^{k_{**}}$-order property, then the tree bound of $G$ is at least $k_{**}+1$. On the other hand, if a graph $G$ has tree bound at least $k_{**}=2^{k_*+1}-3$, then it has the $k_*$-order property.*

*Proof.* For the first implication, just consider $\langle a_i \mid i \in \{1, \dots, 2^{k_{**}}-1\}\rangle$ and $\langle b_i \mid i \in \{0, \dots, 2^{k_{**}}-1\}\rangle$ to be the two sequences of vertices witnessing the $2^{k_{**}}$-order property in $G$, and thus for all $i, j \leq k$, $a_i R b_j$ if and only if $i \geq j$. It is straightforward to build a $k_{**}$-tree using these vertices. Take $\langle b_i \mid i \in \{0, \dots, 2^{k_{**}}-1\}\rangle$ to be the branches of the tree, indexing them by the binary decomposition of their index, and run the following construction for the nodes:

- Initiate $C_{=} \langle a_i \mid i \in \{0, \dots, 2^{k_{**}}-2\}\rangle$.
- At each step $k \in \{0, k_{**}-1\}$, for each $\eta \in \{0,1\}^k$, take the middle element of the sequence $C_\eta$ and set it to be the node $c_\eta$. Then, the remaining first half of $C_\eta$ becomes the sequence $C_{\eta \frown \langle 0 \rangle}$ and the second half is $C_{\eta \frown \langle 1 \rangle}$.

Notice that at each step, the sequence $C_\eta$ has an odd number of elements. The resulting two sequences of nodes and branches form a $k_{**}$-tree. See **??** for a visual example of this construction.

Add visual example of order implies tree.

During the proof of the second implication, we say that a set of nodes $N$ of a $k$-tree $H = (\bar{c}, \bar{b})$ *contains* a $k'$-tree, if there exists a map $f\colon \{0,1\}^{<k'} \longrightarrow \{0,1\}^{<k}$ such that for all $\eta, \eta' \in \{0,1\}^{<k'}$, $c_{f(\eta)}$ and $c_{f(\eta')}$ are in $N$, and if $\eta \frown \langle i \rangle = \eta'$ then $f(\eta) \frown \langle i \rangle \triangleleft f(\eta')$, for all $i \in \{0,1\}$. This clearly implies that there is a $k'$-tree $H'$ with nodes in $N$ and branches in $\bar{b}$. Simply, for each $\eta \in \{0,1\}^{k'-1}$, pick exactly two branches $b_{\rho_0}$ and $b_{\rho_1}$ such that $f(\eta) \frown \langle i \rangle \triangleleft \rho_i$ for $i \in \{0,1\}$.

Also, we will use $H'_i$ to denote the subtree of $H'$ consisting of the nodes $c_{f(\eta)}$ and branches $b_{f(\rho)}$ such that $\langle i \rangle \triangleleft \eta$ and $\langle i \rangle \triangleleft \rho$, with $\eta \in \{0,1\}^{<k'}$ and $\rho \in \{0,1\}^{k'}$. Notice that, if $H$ is an $h$-tree, $H_0$ and $H_1$ are $(h-1)$-trees, and together with the root node $c_{f(\emptyset)}$, they partition $H$.

Next, we prove the following claim, which shows that we can always find a tree in one of the parts of a bipartition of the nodes of a larger tree.

*Claim* 3.16. For all $n, k \geq 0$, if $H$ is a $(n+k)$-tree and the nodes of $H$ are partitioned into two sets $N$ and $P$, then either $N$ contains an $n$-tree or $P$ contains a $k$-tree.

*Proof of Claim 3.16.* We prove this by induction on $n+k$. Clearly, the statement is true for the trivial case $n = k = 0$. Suppose $n + k > 0$. Without loss of generality, we may assume that the root node $c_\emptyset$ is in $N$. Let $Z_i$ be the set of nodes of $H_i$, which is an $(n+k-1)$-tree. By I.H., for each $i \in \{0,1\}$, either $N \cap Z_i$ contains an $(n-1)$-tree or $P \cap Z_i$ contains a $k$-tree. If either $P \cap Z_0$ or $P \cap Z_1$ contains a $k$-tree, then $P$ contains a $k$-tree, and we are done. Otherwise, both $N \cap Z_0$ and $N \cap Z_1$ contain an $(n-1)$-tree. Since $c_\emptyset$ is in $N$, the root with the two $(k-1)$-tree are in $N$ and make an $n$-tree. Thus, $N$ contains an $n$-tree. $\square$

Suppose that $G$ has a tree bound of at least $2^{k_*+1}-3$, and thus contains a $(2^{k_*+1}-2)$-tree. We show by induction on $k_* - r$, with $1 \leq r \leq k_*$, that the following scenario $S_r$ holds. There are

$$b_0, c_0, \dots, b_{q-1}, c_{q-1}, H, b_q, c_q, \dots, b_{k_*-r-1}, c_{k_*-r-1} \tag{6}$$

such that:

1. for all $i \in \{0, \dots, k_* - r - 1\}$, $b_i$ and $c_i$ are vertices in $G$, and $H$ is a $(2^{r+1}-2)$-tree in $G$.
2. for all $i, j \in \{0, \dots, k_* - r - 1\}$, $b_i R c_j \Leftrightarrow i \geq j$.

3. if $c$ is a node of $H$, $b_i R c \Leftrightarrow i \geq q$.

4. if $b$ is a branch of $H$, $b R c_i \Leftrightarrow i < q$.

The initial case $S_{k_*}$ only requires the existence of a $(2^{k_*+1} - 2)$-tree in $G$, which is the premise. If the final case $S_1$ is true, then we are done: this case assumes that $H$ is a 2-tree, in which case there is a node $c_*$ and branch $b_*$ in $H$ which are connected. These vertices satisfy conditions 3. and 4., so the sequence resulting from replacing $H$ in (6) by $b_*$, $c_*$ implies that $G$ has the $k_*$-order property. specify k-order

To conclude the proof it remains to show that if $S_r$ holds, then so does $S_{r-1}$ for $r > 1$. Assume $S_r$. Fixing $h = 2^r - 2$, by 1. we have that $H$ is a $(2h+2)$-tree. For each branch $b$ of $H$ we denote $Z(b)$ the set of nodes $c$ of $H$ such that $bRc$.

We have two cases:

- *Case 1.* There is a branch $b_*$ such that $Z(b_*)$ contains an $(h+1)$-tree $H'$. In that case, we can take $c_*$ to be the top node of the $(h+1)$-tree, and $H_*$ to be the $h$-subtree $H'_0$. Replacing $H$ in (6) with $H_*$, $b_*$, $c_*$ in this order, the conditions for $S_{r-1}$ are satisfied.

- *Case 2.* There is no branch $b$ such that $Z(b)$ contains an $(h+1)$-tree. Now, let $c_*$ be the top node of $H$, $Z_1$ the set of nodes of $H_1$, and $b_*$ any branch of $H_1$. By the case assumption, $Z(b) \cap Z_1$ contains no $(h+1)$-tree, so by the claim and the fact that $Z_1$ is the set of nodes of a $(2h+1)$-tree, $Z_1 \setminus Z(b)$ contains an $h$-tree $H_*$. Finally, replacing $H$ in (6) by $b_*$, $c_*$, $H_*$ in this order, the conditions for $S_{r-1}$ are satisfied.

In any case, $S_{r-1}$ is satisfied, and the proof is complete. □

*Remark* 3.17. The key point of the proof of the second implication of Theorem 3.15 is that the found $k$-order does not only utilize edges and non-edges of the $k$-tree structure itself. Instead, it relies on the fact that, for a tall enough tree, a $k$-order must appear in some way, leveraging some "unknown" edges, independently on the choice of those.

The second implication of this theorem is of special interest in the next sections, as it proves that in the context of a $k$-stable graph no $2^{k+1} - 2$-trees can be found.

Given that the stability of the studied graphs is fixed for all proofs in the next sections, from now on we will use $k_*$ as the value of the non-$k$-property of the studied graphs, and $k_{**}$ for the associated tree bound.

# 4. Unbounded Stable Regularity Lemmas

This section works around the concept of $\epsilon$-*indivisible* sets, a strong condition on the quasi-randomness of a subset respect to all the vertices of the graph. This condition results in pairs of sufficiently large subsets of vertices satisfying the *average condition*, which (asymmetrically) strictly bounds the number of exceptional edges in the pair. Using these tools we obtain the first result in Lemma 4.14, which proves the existence of a partition of highly quasi-random pairs with no exceptions, at the cost of a non-homogeneous partition. Next, we improve the results obtaining an equitable partition in Theorem 4.21, but this time with a small number of exceptional pairs, and a tradeoff between a non-negligible remainder set and even smaller parts. The final result, Theorem 4.27, achieves removing non-quasi-random pairs and reduce the size of the remainder set. All in all, even though the partitions of this section present a very strong form of quasi-randomness, they all share the same drawback: a large number of parts that grows with the size of the graph, something that we will be dealing with in the next section.

## 4.1 Indivisibility and Average Condition

First step is defining *indivisibility*. The general definition is for any function $f$, but for the rest of the section we are mostly interested in the case of $f(n) = n^\epsilon$, which we call $\epsilon$-indivisible, and at the end in the constant case $f(n) = c$.

**Definition 4.1** (Definition 4.2(b))**.** Let $f : \mathbb{R} \longrightarrow \mathbb{R}$ be a function. We say that $A \subseteq G$ is $f$-*indivisible* if for every $b \in G$,

$$|\overline{B}_{A,b}| < f(|A|)$$

**Definition 4.2** (Definition 4.2(a))**.** Let $\epsilon \in (0,1)$. We say that $A \subseteq G$ is $\epsilon$-*indivisible* if for every $b \in G$,

$$|\overline{B}_{A,b}| < |A|^\epsilon$$

Redundant.

*Remark* 4.3. An $\epsilon$-indivisible set is $f$-indivisible for $f(n) = n^\epsilon$.

A natural follow-up question, is how strongly bounded are exceptions in the context of two indivisible sets. The following lemma measures precisely that, although doing so in asymmetrically.

**Lemma 4.4** (Claim 4.6))**.** *Let $G$ be a finite graph. Suppose $A, B \subseteq G$ such that $A$ is $f$-indivisible, $B$ is $g$-indivisible, and $f(|A|)g(|B|) < \frac{1}{2}|B|$. Then, the truth value $t = t(A, B)$ satisfies that for all but $< f(|A|)$ of the $a \in A$ for all but $< g(|B|)$ of the $b \in B$ we have that $aRb \equiv t$.*

*Proof.* Since $B$ is $g$-indivisible, for each $a \in A$ we have that $|\overline{B}_{B,a}| < g(|B|)$. Let $U_i = \{a \in A \mid t(a, B) \equiv i\}$ for $i \in \{0, 1\}$. If either $U_i$ satisfies $|U_i| < f(|A|)$ then the statement is true. Suppose not. Then, there are $W_i \subseteq U_i$ with $|W_i| = f(|A|)$ for $i \in \{0, 1\}$. Now, let $V = \{b \in B \mid (\exists a \in W_0 \mid aRb) \vee (\exists a \in W_1 \mid \neg aRb)\}$, i.e. the $b$'s which are an exception for some $a \in W_0 \cup W_1$. Then, $|V| < |W_0|g(|B|) + |W_1|g(|B|) = 2f(|A|)g(|B|) < |B|$, where the first inequality follows the $g$-indivisibility of $B$. Finally, there is a $b_* \in B \setminus V$ such that $\forall a \in W_0\ \neg aRb_*$ and $\forall a \in W_1\ aRb_*$ with $|W_0| = |W_1| = f(|A|)$, which contradicts the $f$-indivisibility of $A$. $\square$

**Definition 4.5.** We say that the pair $(A, B)$ with $A$ $f$-indivisible and $B$ $g$-indivisible satisfies the *average condition* if $f(|A|)g(|B|) < \frac{1}{2}|B|$ and thus the statement of Lemma 4.4 is true for the pair $(A, B)$.

*Remark* 4.6. The condition $f(|A|)g(|B|) < \frac{1}{2}|B|$ makes ordering of the pair $(A, B)$ matter, that is,

$$(A, B) \text{ has the average condition } \not\Rightarrow (B, A) \text{ has the average condition}$$

*Remark* 4.7 (Remark 4.7). When $f(n) = n^{\epsilon}$ and $g(n) = n^{\zeta}$, the average condition is $|A|^{\epsilon}|B|^{\zeta} < \frac{1}{2}|B|$.

I don't think this is useful in any way.

Next, we are interested in studying how the average condition of an indivisible pair controls the homogeneity of large enough subpairs, in the sense of bounding exceptional edges. We study the $f$ and $\epsilon$ case separately, as the specific case of $\epsilon$ gives a slightly better condition on the range of the size of the subpair.

Define homogeneity.

**Lemma 4.8** (Claim 4.8)**.** *Let $A$ be $\epsilon$-indivisible, $B$ $\zeta$-indivisible and let the pair $(A, B)$ satisfy the average condition. Then, for all $\epsilon_1 \in (0, 1-\epsilon)$, $\zeta_1 \in (0, 1-\zeta)$, $A' \subseteq A$ and $B' \subseteq B$ such that $|A'| \geq |A|^{\epsilon+\epsilon_1}$ and $|B'| \geq |B|^{\zeta+\zeta_1}$, we have that:*

$$\frac{|\{(a, b) \in (A', B') \mid aRb \equiv \neg t(A, B)\}|}{|A' \times B'|} \leq \frac{1}{|A|^{\epsilon_1}} + \frac{1}{|B|^{\zeta_1}}$$

*Proof.* Notice:

- There are at most $|A|^{\epsilon}$ vertices of $A$ (hence in $A' \subseteq A$) which are exceptional (in the sense of the average condition).
- For each $a \in A$ (hence in $A' \subseteq A$) not exceptional, there are at most $|B|^{\zeta}$ elements $b \in B$ such that $(a, b)$ does not satisfy the truth value $t(A, B)$, i.e. that are exceptional.

This is not the same use of the word exceptional as defined in Section 3.

Putting it all together:

This makes my eyes bleed.

$$\begin{aligned}
\frac{|\{(a, b) \in (A', B') \mid aRb \equiv \neg t(A, B)\}|}{|A' \times B'|} &\leq \frac{|A|^{\epsilon}|B'| + (|A'| - |A|^{\epsilon})|B|^{\zeta}}{|A'||B'|} \\
&= \frac{|A|^{\epsilon}}{|A'|} + \frac{|A'| - |A|^{\epsilon}}{|A'|}\frac{|B|^{\zeta}}{|B'|} \\
&\leq \frac{|A|^{\epsilon}}{|A'|} + \frac{|B|^{\zeta}}{|B'|} \\
&\leq \frac{|A|^{\epsilon}}{|A|^{\epsilon+\epsilon_1}} + \frac{|B|^{\zeta}}{|B|^{\zeta+\zeta_1}} \\
&= \frac{1}{|A|^{\epsilon_1}} + \frac{1}{|B|^{\zeta_1}}
\end{aligned}$$

□

This next generalization may seem to make previous lemma redundant. But they actually prove different results. But is it worth to keep both?

**Lemma 4.9** ($f$-indivisible version)**.** *Let $A$ be $f$-indivisible, $B$ $g$-indivisible and let the pair $(A, B)$ satisfy the average condition. Then, for all $\epsilon_1 \in (0, 1 - \frac{f(|A|)}{|A|})$, $\zeta_1 \in (0, 1 - \frac{g(|B|)}{|B|})$, $A' \subseteq A$ and $B' \subseteq B$ such that $|A'| \geq f(|A|)|A|^{\epsilon_1}$ and $|B'| \geq g(|B|)|B|^{\zeta_1}$, we have that:*

$$\frac{|\{(a, b) \in (A', B') \mid aRb \equiv \neg t(A, B)\}|}{|A' \times B'|} \leq \frac{1}{|A|^{\epsilon_1}} + \frac{1}{|B|^{\zeta_1}}$$

*Proof.* Notice:

- There are at most $f(|A|)$ elements of $A$ (hence in $A' \subseteq A$) which are exceptional (in the sense of the average condition).
- For each $a \in A$ (hence in $A' \subseteq A$) not exceptional, there are at most $g(|B|)$ elements $b \in B$ such that $(a, b)$ does not satisfy the truth value $t(A, B)$, i.e. that are exceptional.

Putting it all together:

$$\begin{aligned}
\frac{|\{(a,b) \in (A', B') \mid aRb \equiv \neg t(A,B)\}|}{|A' \times B'|} &\leq \frac{f(|A|)|B'| + (|A'| - f(|A|))g(|B|)}{|A'||B'|} \\
&= \frac{f(|A|)}{|A'|} + \frac{|A'| - f(|A|)}{|A'|}\frac{g(|B|)}{|B'|} \\
&\leq \frac{f(|A|)}{|A'|} + \frac{g(|B|)}{|B'|} \\
&\leq \frac{f(|A|)}{f(|A|)|A|^{\epsilon_1}} + \frac{g(|B|)}{g(|B|)|B|^{\zeta_1}} \\
&= \frac{1}{|A|^{\epsilon_1}} + \frac{1}{|B|^{\zeta_1}}
\end{aligned}$$

□

This may be skipped, and be directly commented in the appropiate remark following the theorem.

For later use, we are particularly interested in the case when $f(n) = c$.

**Corollary 4.10** (Corollary 4.9)**.** *Let $A$ and $B$ be $f$-indivisible with $f(n) = c$ and $(A, B)$ satisfy the average condition. Then, for all $\epsilon_1 \in (0, 1 - \frac{c}{|A|})$, $\zeta_1 \in (0, 1 - \frac{c}{|B|})$, $A' \subseteq A$ and $B' \subseteq B$ with $|A'| \geq c|A|^{\epsilon_1}$ and $|B'| \geq c|B|^{\zeta_1}$, we have:*

$$\frac{|\{(a,b) \in (A', B') \mid aRb \equiv \neg t(A,B)\}|}{|A' \times B'|} \leq \frac{1}{|A|^{\epsilon_1}} + \frac{1}{|B|^{\zeta_1}}$$

*Proof.* Use Lemma 4.9 with $f(n) = c$. □

*Remark* 4.11. Notice that the average condition is easily satisfied if the pair satisfies a condition on the size of its sets. If $f(n) = n^{\epsilon}$, $A$ and $B$ are $f$-indivisible, and $|B| \geq |A| \geq m$, then $m^{1-2\epsilon} > 2$ is sufficient for the average condition to hold for the pair $(A, B)$:

$$\frac{|A|^{\epsilon}|B|^{\epsilon}}{|B|} \leq \frac{|B|^{2\epsilon}}{|B|} = \frac{1}{|B|^{1-2\epsilon}} = \frac{1}{m^{1-2\epsilon}} < \frac{1}{2}$$

We will be using this fact in the context of a sequence of non-zero natural numbers $\{m_\ell \mid \ell \in \{0, \dots, k_{**}\}\}$ where $\lfloor m_\ell^{\epsilon} \rfloor = m_{\ell+1}$ for some $\epsilon \in (0, \frac{1}{2})$ and for all $\ell \in \{0, \dots, k_{**} - 1\}$. Here, $2 < (m_{k_{**}-1})^{1-2\epsilon}$ is sufficient for any $f$-indivisible $A$ and $B$, with $|A|, |B| \in \{m_0, \dots, m_{k_{**}-1}\}$, to satisfy the average condition.

Now that we have proven some properties of indivisible sets, we are actually interested in whether they can be found in a graph. It turns out that the non-$k$-order property, or more specifically the associated tree bound, is sufficient for proving it. The proof resumes in assuming that there is no indivisible set to recursively refine a "semi-partition" which by construction contains a $k_{**}$-tree.

Lluis: hi ha alguna manera de dir una partició que no cubreix tots els vertex amb una paraula?

**Lemma 4.12** (Claim 4.3)**.** *Let $G$ be a finite graph with the non-$k_*$-property and $f : \mathbb{R} \longrightarrow \mathbb{R}$ a function such that $x \geq f(x)$. Let $\{m_\ell \mid \ell \in \{0, \dots, k_{**}\}\}$ be a sequence of non-zero natural numbers such that for all $\ell \in \{0, \dots, k_{**} - 1\}$, $f(m_\ell) \geq m_{\ell+1}$. If $A \subseteq G$, $|A| \geq m_0$, then for some $\ell \in \{0, \dots, k_{**} - 1\}$ there is a subset $B \subseteq A$ of size $m_\ell$ which is $f$-indivisible.*

*Proof.* Suppose not. Then we can construct the sequences $\langle b_\eta \mid \eta \in \{0,1\}^{<k} \rangle$ and $\langle A_\eta \mid \eta \in \{0,1\}^{\leq k} \rangle$ on induction over $k = |\eta|$, satisfying:

1. $A_{\eta ^\frown \langle i \rangle} \subseteq A_\eta$, $\forall i \in \{0, 1\}$, $\forall k \in \{0, \dots, k_{**} - 1\}$
2. $A_{\eta ^\frown \langle 0 \rangle} \cap A_{\eta ^\frown \langle 1 \rangle} = \emptyset$, $\forall k \in \{0, \dots, k_{**} - 1\}$
3. $|A_\eta| = m_k$, $\forall k \in \{0, \dots, k_{**}\}$
4. $b_\eta \in G$ witnessing that $A_\eta$ is not $f$-indivisible, $\forall k \in \{0, \dots, k_{**} - 1\}$
5. $A_{\eta ^\frown \langle i \rangle} \subseteq A_\eta^{(i)} = \{a \in A_\eta \mid aRb_\eta \equiv (i = 1)\}$, $\forall \in \{0, 1\}$, $\forall k \in \{0, \dots, k_{**} - 1\}$

Let's prove the induction. For $k = 0$, consider any $A_{\langle \cdot \rangle} \subseteq A$, satisfying $|A_{\langle \cdot \rangle}| = m_0$, and any $b_{\langle \cdot \rangle}$ witnessing the non-$f$-indivisibility of $A_{\langle \cdot \rangle}$. For $k > 0$ we can assume by hypothesis that $A_\eta$, with $|A_\eta| = m_k$, is not $f$-indivisible. Thus, there exists $b_\eta$ such that $A_\eta^{(i)} \geq f(m_k) \geq m_{k+1}$ (4.), and we can choose $A_{\eta ^\frown \langle i \rangle} \subseteq A_\eta^{(i)}$ (5.), such that $|A_{\eta ^\frown \langle i \rangle}| = m_{k+1}$ $\forall i \in \{0, 1\}$ (3.). 1. and 2. are satisfied by the definition of $A_\eta^{(i)}$. Now, for all $\eta$ such that $|\eta| = k_{**}$, consider some element $a_\eta \in A_\eta$, which exists since $m_\ell > 0$ for all $\ell$. Then, we have two sequences $\langle b_\eta \mid \eta \in \{0,1\}^{<k_{**}} \rangle$ and $\langle a_\eta \mid \eta \in \{0,1\}^{k_{**}} \rangle$ satisfying the $k_{**}$-tree property: for all $\rho \in \{0,1\}^{<k_{**}}$ and $\eta \in \{0,1\}^{k_{**}}$ if given $\ell \in \{0, 1\}$ we have $\rho ^\frown \langle \ell \rangle \trianglelefteq \eta$ then $(a_\eta R b_\rho) \equiv (\ell = 1)$ since $a_\eta \in A_\eta \subseteq A_{\rho ^\frown \langle i \rangle}$. This contradicts the $k_{**}$ tree bound. □

The previous proof can be iteratively used to partition the graph into indivisible parts, with a small reminder. As the average condition cares about the ordering, we define the partition as a tuple instead of a family of sets, and fix an ascending order on the size of the parts.

**Lemma 4.13** (Claim 4.4 + 4.5)**.** *Let $G$ be a finite graph with the non-$k_*$-order property and $f : \mathbb{R} \longrightarrow \mathbb{R}$ a function such that $x \geq f(x)$. Let $\{m_\ell \mid \ell \in \{0, \dots, k_{**}\}\}$ be a sequence of non-zero natural numbers such that for all $\ell \in \{0, \dots, k_{**} - 1\}$, $f(m_\ell) \geq m_{\ell+1}$. If $A \subseteq G$ with $|A| = n$, then we can find a sequence $\overline{A} = \langle A_j \mid j \in \{1, \dots, j(*)\} \rangle$ and reminder $B = A \setminus \bigcup \overline{A}$ such that:*

1. *For each $j \in \{1, \dots, j(*)\}$, $A_j$ is $f$-indivisible.*
2. *For each $j \in \{1, \dots, j(*)\}$, $|A_j| \in \{m_0, \dots, m_{k_{**}-1}\}$.*
3. *$A_j \subseteq A \setminus \bigcup \{A_i \mid i < j\}$, in particular $A_i \cap A_j = \emptyset$ $\forall i \neq j$.*
4. *$|B| < m_0$.*
5. *$\overline{A}$ is $\leq$-increasing.*

*Proof.* Iteratively, apply Lemma 4.12 to the remainder $A \setminus \bigcup \{A_i \mid i < j\}$ (3.) to get an $f$-indivisible $A_j$ (1.) of size $m_\ell$, $\ell \in \{0, \dots, k_{**} - 1\}$ (2.) until less then $m_0$ vertices are available (4.). To conclude, reorder the indices of the $A_j$'s in ascending size order (5.). □

Finally, we ensure the pairs satisfy the average condition by simply requiring a minimal size of the parts, a condition that can be easily integrated in the definition of the sequence of integers.

Last condition can be changed (and probably should) for a condition on $m_{k_{**}}$

**Lemma 4.14** (Claim 4.10)**.** *Let $G$ be a finite graph with the non-$k_*$-order property. Let $\{m_\ell \mid \ell \in \{0, \dots, k_{**}\}\}$ be a sequence of non-zero natural numbers such that $n \geq m_0$ and for all $\ell \in \{0, \dots, k_{**} - 1\}$, $\lfloor m_\ell^\epsilon \rfloor = m_{\ell+1}$, for some $\epsilon \in (0, \frac{1}{2})$ such that $2 < (m_{k_{**}-1})^{1-2\epsilon}$. If $A \subseteq G$ with $|A| = n$, then we can find a sequence $\overline{A} = \langle A_i \mid i \in \{1, \dots, i(*)\}\rangle$ and reminder $B = A \setminus \bigcup \overline{A}$ satisfying:*

1. *For each $i \in \{1, \dots, i(*)\}$, $A_i$ is $\epsilon$-indivisible.*
2. *For each $i \in \{1, \dots, i(*)\}$, $|A_i| \in \{m_0, \dots, m_{k_{**}-1}\}$.*
3. *$A_i \cap A_j = \emptyset$ for all $i \neq j$.*
4. *$|B| < m_0$.*
5. *$\overline{A}$ is $\leq$-increasing.*
6. *If $\zeta \in (0, \epsilon^{k_{**}})$ then for every $i, j \in \{1, \dots, i(*)\}$ with $i < j$, $A \subseteq A_i$ ad $B \subseteq A_j$ such that $|A| \geq |A_i|^{\epsilon+\zeta}$ and $|B| \geq |A_j|^{\epsilon+\zeta}$ we have that:*

$$\begin{aligned}\frac{|\{(a,b) \in (A,B) \mid aRb \equiv \neg t(A_i, A_j)\}|}{|A \times B|} &\leq \frac{1}{|A_i|^\zeta} + \frac{1}{|A_j|^\zeta} \\ &\leq \frac{1}{|A|^\zeta} + \frac{1}{|B|^\zeta}\end{aligned}$$

*Proof.* The five points are direct consequence of Lemma 4.13, setting $f(x) = x^\epsilon$. Now, by 2., for any $A_i, A_j \in \overline{A}$ with $i < j$ there is some $\ell \in \{0, \dots, k_{**} - 1\}$ such that $|A_i| \leq |A_j| = m_\ell$. Also, it follows the condition $2 < (m_{k_{**}-1})^{1-2\epsilon}$ and Remark 4.11 that the pair $(A_i, A_j)$ satisfies the average condition. Finally, notice that $\epsilon^{k_{**}} < \epsilon < 1 - \epsilon$ since $\epsilon \in (0, \frac{1}{2})$, so that $\zeta \in (0, \epsilon^{k_{**}}) \subseteq (0, 1 - \epsilon)$ and the condition for Lemma 4.8 is satisfied. This gives us 6. and concludes the proof of the statement. ☐

Maybe merge the last two lemmas?

*Remark* 4.15. For sufficiently small $\epsilon$, the condition $2 < (m_{k_{**}-1})^{1-2\epsilon}$ is almost, trivial. For example, if $\epsilon < \frac{1}{4}$, then we are just requiring that $m_{k_{**}-1} \geq 4$.

## 4.2 $\epsilon$-indivisible Equitable Partition

Hehe, "partition indivisible sets", sona contradictiu.

As stated earlier, the principal drawback of the previous result is that the obtained partition is not equitable. To deal with this, we study the event of randomly partitioning a pair of indivisible sets into subparts of equal size. We prove that the event of a pair of subparts of the refinement being either fully connected or completely empty, is satisfied with very high probability.

**Definition 4.16.** Let $A, B$ be $f$-indivisible sets with $f(A)f(B) < \frac{1}{2}|B|$. Let $\langle A_i \mid i \in \{1, \dots, i_A\}\rangle$ be a partition of $A$ with $|A_i| = m$ for all $i \in \{1, \dots, i_A\}$ and $\langle B_i \mid i \in \{1, \dots, i_B\}\rangle$ be a partition of $B$ with $|B_i| = m$ for all $i \in \{1, \dots, i_B\}$. We define $\varepsilon^+_{A_i, A_j, m}$ as the event:

$$\forall a \in A_i \; \forall b \in B_i, aRb = t(A, B)$$

Should be mentioned that this is a strong limitation which is not mentioned in the original paper, but required for the calculations (sin hacer trampa).

Here we have enforced the equality. Should be commented that this is easily achievable as we do in the next results.

**Lemma 4.17** (Claim 4.13). *Let $G$ be a finite graph with the non-$k_*$-order property. Let $\langle m_\ell \mid \ell \in \{0, \dots, k_{**}\}\rangle$ be a sequence of non-zero natural numbers such that $n \geq m_0 \geq n^\epsilon$ and for all $\ell \in \{0, \dots, k_{**}-1\}$, $m_\ell^\epsilon = m_{\ell+1}$, for some $\epsilon \in (0, \frac{1}{2})$ such that $2 < (m_{k_{**}-1})^{1-2\epsilon}$. Let $A_1, A_2 \subseteq G$ be two $\epsilon$-indivisible subsets such that $|A_1| = m_{\ell_1}$ and $|A_2| = m_{\ell_2}$ for some $\ell_1, \ell_2 \in \{0, \dots, k_{**}-1\}$ and $|A_1| \leq |A_2|$. Let $c \in (0, 1-\epsilon)$ and $\zeta \leq \frac{1-\epsilon-c}{3}\epsilon^{k_{**}}$ such that $m := n^\zeta$ divides $|A_1|$ and $|A_2|$. Then, let $\left\langle A_{1,s} \mid s \in \left\{1, \dots, \frac{|A_1|}{m}\right\}\right\rangle$ and $\left\langle A_{2,t} \mid t \in \left\{1, \dots, \frac{|A_2|}{m}\right\}\right\rangle$ be random partitions of $A_1$ and $A_2$ respectively, with pieces of size $m$. We have that*

$$P(\varepsilon^+_{A_{1,s},A_{2,t},m}) \geq 1 - \frac{2}{n^{c\epsilon^{k_{**}}}}$$

*Proof.* Fix $s \in \frac{|A_1|}{m}$, $t \in \frac{|A_2|}{m}$. It follows from the condition $2 < (m_{k_{**}-1})^{1-2\epsilon}$ and Remark 4.11 that the pair $(A_1, A_2)$ satisfies the average condition. Let $U_1 = \left\{a \in A_1 \mid \left|\{b \in A_2 \mid aRb \equiv \neg t(A_1, A_2)\}\right| \geq |A_2|^\epsilon\right\}$ and for each $a \in A_1 \setminus U_1$ let $U_{2,a} = \{b \in A_j \mid aRb \equiv \neg t(A_1, A_2)\}$. By Lemma 4.4, $|U_1| \leq |A_1|^\epsilon$ and $\forall a \in A_1 \setminus U_1$, $|U_{2,a}| \leq |A_2|^\epsilon$. Now, given $\left\{1, \dots, \frac{|A_1|}{m}\right\}$, we can bound the probability $P_1$ that $A_{1,s} \cap U_1 \neq \emptyset$ as follows:

$$\begin{aligned}
P_1 &\leq \frac{|U_1|}{|A_1|} + \cdots + \frac{|U_1|}{|A_1| - m + 1} < \frac{m|U_1|}{|A_1| - m} \leq \frac{m|A_1|^\epsilon}{|A_1| - m} \\
&\leq \frac{m^2|A_1|^\epsilon}{|A_1|} = \frac{m^2}{|A_1|^{1-\epsilon}} = \frac{m^2}{m_0^{(1-\epsilon)\epsilon^{\ell_1}}} \leq \frac{n^{2\zeta}}{n^{(1-\epsilon)\epsilon^{\ell_1+1}}} \\
&\leq \frac{n^{2\frac{1-\epsilon-c}{3}\epsilon^{k_{**}}}}{n^{(1-\epsilon)\epsilon^{k_{**}}}} \leq \frac{n^{(1-\epsilon-c)\epsilon^{k_{**}}}}{n^{(1-\epsilon)\epsilon^{k_{**}}}} = \frac{1}{n^{c\epsilon^{k_{**}}}}
\end{aligned}$$

The forth inequality comes from the fact that $\frac{(|A_i|-m)m}{|A_i|} \geq 1$. Then, if $A_{1,s} \cap U_1 = \emptyset$, we have that $|\bigcup_{a \in A_{1,s}} U_{2,a}| \leq |A_{1,s}||A_2|^\epsilon$. So, given $\left\{1, \dots, \frac{|A_2|}{m}\right\}$, we can bound $P_2$, the probability that $A_{2,t} \cap \bigcup_{a \in A_{1,s}} U_{2,a} \neq \emptyset$, by:

$$\begin{aligned}
P_2 &\leq \frac{|\bigcup_{a \in A_{1,s}} U_{2,a}|}{|A_2|} + \cdots + \frac{|\bigcup_{a \in A_{1,s}} U_{2,a}|}{|A_2| - m + 1} < \frac{m|\bigcup_{a \in A_{1,s}} U_{2,a}|}{|A_2| - m} \leq \frac{mm|A_2|^\epsilon}{|A_2| - m} \\
&\leq \frac{m^3|A_2|^\epsilon}{|A_2|} = \frac{m^3}{|A_2|^{1-\epsilon}} = \frac{m^3}{m_0^{(1-\epsilon)\epsilon^{\ell_2}}} \leq \frac{n^{3\zeta}}{n^{(1-\epsilon)\epsilon^{\ell_2+1}}} \\
&\leq \frac{n^{3\frac{1-\epsilon-c}{3}\epsilon^{k_{**}}}}{n^{(1-\epsilon)\epsilon^{k_{**}}}} \leq \frac{n^{(1-\epsilon-c)\epsilon^{k_{**}}}}{n^{(1-\epsilon)\epsilon^{k_{**}}}} = \frac{1}{n^{c\epsilon^{k_{**}}}}
\end{aligned}$$

Putting it all together:

$$P(\varepsilon^+_{A_{1,s},A_{2,t},m}) \geq (1 - P_1)(1 - P_2) \geq \left(1 - \frac{1}{n^{c\epsilon^{k_{**}}}}\right)^2 \geq 1 - \frac{2}{n^{c\epsilon^{k_{**}}}}$$

□

*Remark* 4.18. The condition on the size of $m_0$, which is both an upper and lower bound, is very strong and will be carried over up to Theorem 4.21. The greater limitations of this resides in the fact that the size of the parts of the resulting partition $m_{**}$ is set by the size of $m_0$, and thus inherits the same limitations.

Now, since the event of a given subpair not satisfying the desired property is very unlikely, it can be easily proven that a random refinement of the partition given by Lemma 4.13 only has a small number of exceptional pairs.

**Lemma 4.19** (Claim 4.14)**.** *Let $G$ be a finite graph with the non-$k_*$-order property. Let $\langle m_\ell \mid \ell \in \{0, \dots, k_{**}\}\rangle$ be a sequence of non-zero natural numbers such that for all $\ell \in \{0, \dots, k_{**}-1\}$, $m_\ell^\epsilon = m_{\ell+1}$, for some $\epsilon \in (0, \frac{1}{2})$ such that $2 < (m_{k_{**}-1})^{1-2\epsilon}$. Also, suppose $m_0$ satisfies $n^\epsilon \leq m_0 < \min(\frac{\sqrt{2}-1}{\sqrt{2}}n, \frac{n}{n^{c\epsilon^{k_{**}}}})$, with $c \in (0, 1-\epsilon)$. Finally, let $m_{**}$ be a divisor of $m_\ell$ for all $\ell \in \{0, \dots, k_{**}-1\}$ and $m_{**} \leq n^{\frac{1-\epsilon-c}{3}\epsilon^{k_{**}}}$. If $A \subseteq G$ with $|A| = n$, then we can find a partition $\overline{A} = \langle A_i \mid i \in \{1, \dots, r\}\rangle$ with reminder $B = A \setminus \bigcup_{i\in\{1,\dots,r\}} A_i$ such that:*

1. *$|A_i| = m_{**}$ for all $i \in \{1, \dots, r\}$.*
2. *For all but $\frac{2}{n^{c\epsilon^{k_{**}}}}r^2$ of the pairs $(A_i, A_j)$ with $i < j$ there are no exceptional edges, i.e.*
$$\{(a, b) \in A_i \times A_j \mid aRb \not\equiv t(A_i, A_j)\} = \emptyset$$
3. *$|B| < m_0$*

*Proof.* We can use Lemma 4.13 to get a partition $\overline{A'} = \langle A_i' \mid i \in \{1, \dots, i(*)\}\rangle$ and remainder $B' = A \setminus \bigcup A'$. We can refine the partition by randomly splitting each $A_i'$ into pieces of size $m_{**}$ (1.). Consider the resulting partition $\overline{A} = \langle A_i \mid i \in \{1, \dots, r\}\rangle$ with remainder $B = B'$ (3.). First of all, notice that for each pair $(A_i, A_j)$ such that $A_i \subseteq A_{i_1}'$ and $A_j \subseteq A_{j_1}'$ with $i_1 \neq j_1$, the probability of the pair having exceptional edges is upper bounded by $\frac{2}{n^{c\epsilon^{k_{**}}}}$. This follows Lemma 4.17. Thus, given $X$ the random variable counting the number of exceptional pairs of this kind, we have

$$\mathbb{E}(X) = \sum_{\substack{A_i, A_j \text{ s.t.} \\ A_i \subseteq A_{i_1}', A_j \subseteq A_{j_1}' \\ i_1 \neq j_1}} \mathbb{E}(X_{A_i, A_j}) = \sum_{\substack{A_i, A_j \text{ s.t.} \\ A_i \subseteq A_{i_1}', A_j \subseteq A_{j_1}' \\ i_1 \neq j_1}} P(\varepsilon_{A_i, A_j, m_{**}}) \leq \frac{r^2}{2} \frac{2}{n^{c\epsilon^{k_{**}}}}$$

where $X_{A_i,A_j}$ is the random variable giving 1 if $(A_i, A_j)$ is exceptional, and 0 otherwise. Since the expectation is an average, for some refinement $\overline{A}$ of $\overline{A'}$ we have that the number of exceptional pairs when $i_1 \neq j_1$ is at most $\frac{r^2}{n^{c\epsilon^{k_{**}}}}$. Now, we have no control if $i_1 = j_1$, so let's bound how many of these we have:

$$\begin{aligned}
\left|\{\text{Exceptional } (A_i, A_j) \mid A_i, A_j \subseteq A_{i_1}', i_1 \in \{1, \dots, i(*)\}\}\right| &\leq \binom{\frac{m_0}{m_{**}}}{2} \frac{n}{m_0} \\
&\leq \frac{(\frac{m_0}{m_{**}})^2}{2} \frac{n}{m_0} = \frac{m_0 n}{2m_{**}^2} = \frac{m_0}{n}\left(\frac{n}{\sqrt{2}m_{**}}\right)^2 \\
&\leq \frac{m_0}{n}\left(\frac{n - m_0}{m_{**}}\right)^2 \leq \frac{m_0}{n} r^2 < \frac{r^2}{n^{c\epsilon^{k_{**}}}}
\end{aligned}$$

Notice that the third inequality comes after the condition $m_0 \leq \frac{\sqrt{2}-1}{\sqrt{2}}n$. Putting it all together, we see that the number of exceptional pairs is upper bounded by $\frac{2r^2}{n^{c\epsilon^{k_{**}}}}$ satisfying 2.. □

*Remark* 4.20 (Remark 4.15). In the previous proof, the condition $m_0 < \frac{n}{n^{c\epsilon^{k_{**}}}}$ can be weakened at the cost of increasing the number of exceptional pairs. More specifically, since this condition is only used to bound the exceptional sub-pairs in the same pair (the second part of the proof), the number of exceptional pairs can be generally bounded by

$$|\{\text{Exceptional pairs}\}| \leq \left(\frac{m_0}{n} + \frac{2}{n^{c\epsilon^{k_{**}}}}\right) r^2$$

Notation here is confusing. $r$ is another thing, and $m$ becomes the number of parts.

We now resume the previous results in a theorem with minimal conditions.

**Theorem 4.21** (Theorem 4.16)**.** *Let $\epsilon = \frac{1}{r} \in (0, \frac{1}{2})$ with $r \in \mathbb{N}$ (this avoids rounding errors), $c \in (0, 1-\epsilon)$ and $k_*$ be given. Let $G$ be a finite graph with the non-$k_*$-order property. Let $A \subseteq G$ with $|A| = n$, and $n > 2^{\frac{r^{k_{**}}}{1-2\epsilon}}$. Then, for any $m_{**} \in \left[n^{\frac{(1-\epsilon-c)}{3}\epsilon^{k_{**}+1}}, (\frac{\sqrt{2}-1}{\sqrt{2}})^{\frac{1-\epsilon-c}{3}\epsilon^{k_{**}}} n^{\frac{(1-\epsilon-c)}{3}\epsilon^{k_{**}} - \frac{(1-\epsilon-c)c}{3}\epsilon^{2k_{**}}}\right]$, there is a partition $\overline{A} = \langle A_i \mid i \in \{1, \dots, m\}\rangle$ of $A$ with remainder $B = A \setminus \bigcup \overline{A}$ such that:*

1. *$|A_i| = m_{**}$ for all $i \in \{1, \dots, m\}$.*
2. *$|B| < m_{**}^{\frac{3}{(1-\epsilon-c)} r^{k_{**}}}$.*
3. *$\left|\{(i,j) \mid i,j \in \{1, \dots, m\}, i < j \text{ and } \{(a,b) \in A_i \times A_j \mid aRb\} \notin \{A_i \times A_j, \emptyset\}\}\right| \leq \frac{2}{n^{c\epsilon^{k_{**}}}} m^2$*

*Proof.* Let $m_{k_{**}} = m_{**}^{\frac{3}{1-\epsilon-c}}$, and consider the sequence

$$m_{**} \leq m_{k_{**}} < \cdots < m_0$$

such that for all $\ell \in \{1, \dots, k_{**}\}$ we have that $m_{\ell-1} = m_\ell^r$. Notice that:

1. $m_{**}$ divides $m_\ell$ for all $\ell \in \{0, \dots, k_{**}\}$ since the $m_\ell$'s are powers of $m_{k_{**}}$ and $m_{**}$ divides $m_{k_{**}}$ by construction.

Probably it is not needed that $m_{**}$ divides $m_{k_{**}}$, with $m_{k_{**}-1}$ is enough, but it comes for free.

2. $(m_{\ell-1})^\epsilon = m_\ell$ for all $\ell \in \{1, \dots, k_{**}\}$, by construction.
3. $m_{**} \leq n^{\frac{1-\epsilon-c}{3}\epsilon^{k_{**}}}$, by choice of $m_{**}$.
4. $m_0 = m_{**}^{\frac{3}{1-\epsilon-c} r^{k_{**}}}$, so on one hand

$$m_0 = m_{**}^{\frac{3}{1-\epsilon-c} r^{k_{**}}} \geq n^{\frac{1-\epsilon-c}{3}\epsilon^{k_{**}+1} \frac{3}{1-\epsilon-c} r^{k_{**}}} \geq n^\epsilon$$

   and on the other hand,

$$m_0 = m_{**}^{\frac{3}{1-\epsilon-c} r^{k_{**}}} \leq (\frac{\sqrt{2}-1}{\sqrt{2}}) n^{1-c\epsilon^{k_{**}}}$$

   and thus $n$ is both smaller than $(\frac{\sqrt{2}-1}{\sqrt{2}})n$ and smaller than $n^{1-c\epsilon^{k_{**}}}$.

5. $m_{k_{**}-1} = m_{**}^{\frac{3}{1-\epsilon-c} r} \geq n^{\epsilon^{k_{**}}} > 2^{\frac{1}{1-2\epsilon}}$.

So, all the conditions of Lemma 4.19 are satisfied, and we can use it to get a partition $\overline{A}$ with remainder $B$ satisfying the statement. Notice that 2. is satisfied by the fact that $|B| < m_0 \leq m_{**}^{\frac{3}{(1-\epsilon-c)} r^{k_{**}}}$. □

*Remark* 4.22. Some notes on the partition obtained in the previous theorem:

- With any choice of $c$ and $m_{**}$, the fraction of exceptional pairs is asymptotically small, but we obtain very small parts, that is, $m_{**} \approx n^{\epsilon^{k_{**}}}$.
- A smaller value of $c$ results in larger parts and smaller reminder, at the cost of a larger fraction of exceptional pairs.
- The window of choice of $m_{**}$ is very small, and taking a larger value (in the given interval), results in a strongly larger reminder. The edge case of choosing $m_{**}$ as the larger value, results in the bound on the size of the reminder becoming $|B| < \frac{\sqrt{2}-1}{\sqrt{2}} n^{1-\epsilon^{k_{**}}}$.

## 4.3 $f_c$-indivisible Equitable Partition

Next, we will follow another approach to obtain an equitable partition. That is, we prove a result similar to that of Lemma 4.12, but this time the size of the resulting quasi-random set can be chosen in advance. The resulting Lemma 4.26 has also the advantage that the associated quasi-random property is $f_c$-indivisibility, where $f_c$ is the constant function $f_c(x) = c$, which is much stronger then $\epsilon$-indivisibility as the bound on the number of exceptions is constant.

To prove this result, we use a probabilistic argument, and show that the event of there existing a subset which has intersection smaller than $c$ with every $\overline{B}_{A,b}$ (Definition 4.23) is highly probable under some very specific conditions (Lemma 4.24).

In what follows, $c$ should be another letter, it collides with previous definition. Also, what about renaming $c$-indivisible to $f_c$-indivisible or something like that?

**Definition 4.23** (Definition 4.18)**.** For $n, c \in \mathbb{N}$ and $\epsilon, \zeta, \xi \in \mathbb{R}$, let $\oplus[n, \epsilon, \zeta, \xi, c]$ be the statement: For any set $A$ and family of subsets $P \subseteq \mathcal{P}(A)$ such that $|A| = n$ and $|P| \le n^{\frac{1}{\zeta}}$, and for all $B \in P$ with $|B| \le n^{\epsilon}$, there exists $U \subseteq A$ with $|U| = \lfloor n^{\xi} \rfloor$ such that for all $B \in P$, $|U \cap B| \le c$.

**Lemma 4.24** (Lemma 4.19)**.** *If the reals $\epsilon, \zeta, \xi$ satisfy $\epsilon \in (0, 1)$, $\zeta > 0$ and $0 < \xi < \frac{1}{2}$, the natural number $n$ is sufficiently large ($n > N(\epsilon, \zeta, \xi, c)$) to satisfy the equation*

$$\frac{1}{2n^{1-2\xi}} + \frac{1}{n^{(1-\xi-\epsilon)c-\frac{1}{\zeta}}} < 1 \tag{7}$$

*and $c > \frac{1}{\zeta(1-\xi-\epsilon)}$, then $\oplus[n, \epsilon, \zeta, \xi, c]$ holds.*

*Proof.* First of all, notice that the condition on $c$ implies that $(1 - \xi - \epsilon) > 0$, and thus $\xi < 1 - \epsilon$. Let $m = \lfloor n^{\xi} \rfloor$ be the size of the set $U$ we want to build, and let $\mathcal{F}_* = [A]^m$ the set of sequences of elements of $A$ with length $m$. Let $\mu$ be a probability distribution on $\mathcal{F}_*$ such that for all $F \in \mathcal{F}_*$, $\mu(F) = \frac{|F|}{|\mathcal{F}_*|}$. We want to prove that the probability that a random $U$ satisfies:

1. All elements of $U$ are distinct.
2. For all $B \in P$, $|U \cap B| < c$.

is non-zero. First of all let's bound the converse of 1., i.e. the probability that there are two equal elements in $U$:

$$P_1 = P(\exists s < t \in [m] \mid U_s = U_t) \le \binom{m}{2} \frac{n}{n^2} \le \frac{m^2}{2n} \le \frac{n^{2\xi}}{2n} < \frac{1}{2n^{1-2\xi}}$$

Now, in order to bound 2., let's first bound the probability that at least $c$ elements of $U$ are in a given $B \in P$:

$$P_B = P(\exists^{\ge c} t \in [m] \mid U_t \in B) \le \binom{m}{c} \left(\frac{|B|}{n}\right)^c \le \frac{m^c |B|^c}{n^c} \le \frac{n^{\xi c} n^{\epsilon c}}{n^c} = \frac{1}{n^{c(1-\xi-\epsilon)}}$$

Then, we can bound the converse of 2., i.e. the probability that this happens for some $B \in P$, by:

$$P_2 = P(\exists B \in P \mid \exists^{\ge c} t \in [m], U_t \in B) \le \sum_{B \in P} P_B = \frac{|P|}{n^{c(1-\xi-\epsilon)}} \le \frac{1}{n^{c(1-\xi-\epsilon)-\frac{1}{\zeta}}}$$

Putting it all together, we have that

$$P((1.) \cup (2.)) \le P_1 + P_2 < \frac{1}{2n^{1-2\xi}} + \frac{1}{n^{c(1-\xi-\epsilon)-\frac{1}{\zeta}}}$$

Notice that

- Since $\xi < \frac{1}{2}$ we have that $1 - 2\xi > 0$.
- $c(1 - \xi - \epsilon) - \frac{1}{\zeta} > 0$.

so, the $n$-large enough condition (7) is well defined and

$$P((1.) \cup (2.)) < \frac{1}{2n^{1-2\xi}} + \frac{1}{n^{c(1-\xi-\epsilon)-\frac{1}{\zeta}}} < 1$$

holds. We conclude that the probability that there exists a $U \subseteq A$ satisfying the condition is non-trivial, and $\oplus[n, \epsilon, \zeta, \xi, c]$ holds. □

*Remark* 4.25. In the context of the condition $c > \frac{1}{\zeta(1-\xi-\epsilon)}$ from the previous lemma, we note that the lower bound on $c$ increases as $\xi$ and $\epsilon$ grow, and as $\zeta$ decreases.

A similar pattern is also followed by the large enough condition of $n$ given by Equation (7). For the condition to be met, $n$ needs to grow as the exponents $1 - 2\xi$ and $(1 - \xi - \epsilon)c - \frac{1}{\zeta}$ become smaller. That is, the lower bound on $n$ becomes larger as $\xi$ and $\epsilon$ grow, and as $\zeta$ and $c$ decrease.

**Lemma 4.26** (Claim 4.21)**.** *Let $k_*, c \in \mathbb{N}$ and $\epsilon, \xi \in \mathbb{R}$ such that:*

1. *$G$ is a graph with the non-$k_*$-order property.*
2. *$\epsilon \in (0, \frac{1}{2}]$.*
3. *$\xi \in (0, \frac{\epsilon^{k_{**}}}{2})$.*
4. *$c$ satisfies*

$$c > \frac{1}{\frac{1}{k_*}(1 - \frac{\xi}{\epsilon^{k_{**}}} - \epsilon)}$$

*Then, for every sufficiently large $n \in \mathbb{N}$ (it suffices that $n > g_\epsilon^{k_{**}}(N_{4.24}(\epsilon, \frac{1}{k_*}, \frac{\xi}{\epsilon^{k_{**}}}, c))$, where $g_\epsilon(x) = (x+1)^{\frac{1}{\epsilon}}$), if $A \subseteq G$ with $|A| = n$, there is $Z \subseteq A$ such that*

*(a) $|Z| = \lfloor n^\xi \rfloor$.*

*(b) $Z$ is $f_c$-indivisible in $G$.*

*Proof.* Let $n = m_0 > m_1 > \cdots > m_{k_{**}}$ with $m_\ell = \lfloor m_{\ell-1}^\epsilon \rfloor \geq g_\epsilon^{-1}(m_{\ell-1}) \geq g_\epsilon^{-\ell}(n)$. Then, $m_\ell \geq m_{\ell+1}$ and we can use Lemma 4.12 to get an $\epsilon$-indivisible subset $A_1 \subseteq A$, with $|A_1| = m_\ell$ for some $\ell \in \{0, \dots, k_{**} - 1\}$. Notice that:

- $\epsilon \in (0, 1)$ by 2..
- We can set $\zeta := \frac{1}{k_*} > 0$.
- By 3., $0 < \frac{\xi}{\epsilon^\ell} \leq \frac{\xi}{\epsilon^{k_{**}}} < \frac{1}{2}$.
- For all $\ell \in \{0, \dots, k_{**}\}$, $m_\ell$ is sufficiently large:

$$m_\ell \geq g_\epsilon^{-\ell}(n) \geq g_\epsilon^{-k_{**}}(n) > N_{4.24}(\epsilon, \frac{1}{k_*}, \frac{\xi}{\epsilon^{k_{**}}}, c) > N_{4.24}(\epsilon, \zeta, \frac{\xi}{\epsilon^\ell}, c)$$

- $c > \frac{1}{\frac{1}{k_*}(1-\frac{\xi}{\epsilon^{k_{**}}}-\epsilon)} = \frac{1}{\zeta(1-\frac{\xi}{\epsilon^{k_{**}}}-\epsilon)}$, by 4..

Conditions of Lemma 4.24 are met, so $\oplus[m_\ell, \epsilon, \zeta, \frac{\xi}{\epsilon^\ell}]$ (as defined in Definition 4.23) holds. We can take $A_{(4.23)}$ and $P_{(4.23)}$ to be $A_1$ and $P := \{\overline{B}_{A_1,b} \mid b \in G\}$ respectively, which satisfy:

- $|A_1| = m_\ell$.
- $|P| \leq m_\ell^{k_*} = m_\ell^{\frac{1}{\zeta}}$, where first inequality follows 2. of Corollary 3.12.
- $\forall B \in P$, $|B| \leq |A_1|^\epsilon$ by $\epsilon$-indivisibility of $A_1$.

Thus, by Definition 4.23 we have that there exists $Z \subseteq A_1$ such that:

- $|Z| = \lfloor m_\ell^{\frac{\xi}{\epsilon^\ell}} \rfloor = \lfloor n^{\epsilon^\ell \frac{\xi}{\epsilon^\ell}} \rfloor = \lfloor n^\xi \rfloor$ satisfying a..
- $Z$ is $f_c$-indivisible since $|B \cap Z| \leq c$ for all $B \in P$, satisfying b..

This proves the statement. □

We now use the previous result to build an equitable partition. Similarly to Lemma 4.13, we will iteratively extract an $f_c$-indivisible set from the reminder using Lemma 4.26, until the sufficiently large condition holds.

**Theorem 4.27** (Theorem 4.23)**.** *Let $G$ be a graph with the non-$k_*$-property. For any $\epsilon \in (0, \frac{1}{2}]$, $\xi \in (0, \frac{\epsilon^{k_{**}}}{2})$ and $c > \frac{k_*}{1-\frac{\xi}{\epsilon^{k_{**}}}-\epsilon}$, any $A \subseteq G$ with $|A| = n$ has a partition $\overline{A} = \langle A_i \mid i \in \{1, \dots, i(*)\}\rangle$ of $A$ with remainder $B = A \setminus \bigcup_{i\in\{1,\dots,i(*)\}} A_i$ satisfying:*

- *$|A_i| = \lfloor n^\xi \rfloor$ for all $i \in \{1, \dots, i(*)\}$.*
- *$A_i$ is $f_c$-indivisible for all $i \in \{1, \dots, i(*)\}$, where $f_c(x) = c$ is a constant function.*
- *$|B| \leq N := g_\epsilon^{k_{**}}(N_{4.24}(\epsilon, \frac{1}{k_*}, \frac{\xi}{\epsilon^{k_{**}}}, c))$ where $g_\epsilon(x) = (x+1)^{\frac{1}{\epsilon}}$.*

*Proof.* We will build a sequence of disjoint $f_c$-indivisible subsets $A_i$ by induction on $i$ as follows. Let $R_i = A \setminus \bigcup_{j<i} A_j$ (so $R_1 = A$). At each step, if $|R_i| > N$, we can apply Lemma 4.26 to $R_i$ with the values $f_c$, $\epsilon$ and $\xi$ of the statement of this theorem, to obtain a $f_c$-indivisible subset $A_i$ of $R_i$ of size $\lfloor n^\xi \rfloor$ which will be disjoint with all $A_j$ with $j < i$. Otherwise, we stop and let $\overline{A} = \langle A_j \mid j < i = i(*)\rangle$ and $B = R_i$. By the case hypothesis, $|B| = |R_i| \leq N$, and we are done. □

*Remark* 4.28. Some notes on the partition obtained in the previous theorem:

- The partition is exceptionally quasi-random, and the number of exceptional edges in each pair of parts and subparts is strongly bounded as shown by Corollary 4.10.
- As the upper bound on the size of the remainder is constant with respect to the size of the graph $n$, the remainder as a fraction of the total graph can be made as small as possible (but not completely avoided). If we want the remainder to be at most $\frac{1}{x}$ of the total graph, we can simply impose $n \geq x \cdot N$, and we are done.
- The parts are exponentially smaller then the size of the graph. Hence, the number of parts grows with the size of the graph, which is actually the principal drawback of this theorem. This will be solved in the partition studied in Section 5.

# 5. The Stable Regularity Lemma

This section focuses in leveraging the stability of a graph to create a stable partition which maximum number of parts does not grow with the size of the graph. In order to do so, we first prove the existence of a partition which parts satisfy a property which we prove stronger then regularity: *excellence*.

## 5.1 Goodness and Excellence

We proceed to formalize this concept.

**Definition 5.1** (Definition 5.2(a))**.** Let $G$ be a finite graph with the non-$k_*$-property. We say that $A \subseteq G$ is $\epsilon$-*good* when for every $b \in G$ the truth value $t = t(b, A) \in \{0, 1\}$ satisfies $|\overline{B}_{A,b}| = |\{a \in A \mid aRb \not\equiv t\}| < \epsilon|A|$.

**Definition 5.2** (Definition 5.2(b))**.** Let $G$ be a finite graph with the non-$k_*$-property. We say that $A \subseteq G$ is $(\epsilon, \zeta)$-*excellent* when $A$ is $\epsilon$-good and, if $B$ is $\zeta$-good, then the truth value $t = t(B, A)$ satisfies $|\{a \in A \mid t(a, B) \neq t(A, B)\}| < \epsilon|A|$. In particular, we say $A$ is $\epsilon$-*excellent* if $A$ is $(\epsilon, \epsilon)$-excellent.

We now make some observations about these two properties.

*Remark* 5.3. For comparison with the properties studied in the previous section, a set being $\epsilon$-good is equivalent to the set being $f$-indivisible with $f(n) = \epsilon n$, while $\epsilon$-indivisibility is a much stronger condition then $\epsilon$-goodness, as for large enough $n$, we have that $n^\epsilon < \epsilon n$.

On the other hand, $\epsilon$-excellence carries some kind of reciprocity with other good (and in particular, excellent) sets, which makes it particularly suitable for studying quasi-randomness between pairs of sets. While independence and goodness only bound the number of exceptions with each vertex of the graph independently, excellence of a set $A$ also ensures that the truth values of each of its vertex with respect to each good set in the graph remains mostly the same. **??** shows an example of an $\epsilon$-good set that, as it does not satisfy this reciprocity condition with another good set, it is not $\epsilon$-excellent.

*Remark* 5.4. If $A, B \subseteq G$ are two (not necessarily disjoint) subsets of vertices with $A$ $(\epsilon, \epsilon')$-excellent and $B$ $\epsilon'$-good set, then the number of exceptional edges between $A$ and $B$, i.e. these vertex pairs that do not follow $t(A, B)$, is relatively small:

$$|\{\text{Exceptional edges between } A \text{ and } B\}| < \epsilon|A||B| + (1-\epsilon)|A|\epsilon'|B| = (\epsilon + (1-\epsilon)\epsilon')|A||B|$$

A relevant example is that of two disjoint $\epsilon$-excellent sets, in which case we have that the fraction of exceptional edges between them is less than $2\epsilon$. If they are not disjoint, we can still use the same reasoning to conclude that the fraction of exceptional edges is less than $2\epsilon\frac{|A||B|}{e(A,B)} < 8\epsilon$, since $e(A, B) > \frac{|A||B|}{4}$.

Discuss with Luis, this may be reduced but I am not sure.

*Remark* 5.5. A final important remark, is the fact that differently then most quasi-random properties, $\epsilon$-excellence is not monotonic. That is, in general, for $\epsilon < \epsilon'$, a set being $\epsilon$-excellent does not imply it being also $\epsilon'$-excellent (and trivially neither the converse). See **??** for a counter example to the monotonicity of this property.

Lluis: això està ben dit?

On the other hand, the non-symmetric $(\epsilon, \epsilon')$-excellence satisfies some sort of monotonicity. That is, if a given set is $(\epsilon_1, \epsilon_1')$-excellent, then it is also $(\epsilon_2, \epsilon_2')$-excellent for all $\epsilon_1 \leq \epsilon_2$ and $\epsilon_1' \geq \epsilon_2'$, since restricting the condition on the goodness of the relevant good sets ($\epsilon_1'$ to $\epsilon_2'$) takes less of such sets into account, and relaxing the condition on the "exceptional truth values" ($\epsilon_1$ to $\epsilon_2$) only enlarges the error accepted.

## 5.2 Excellent Partitions

The first step towards constructing a partition of sets with such property, is to prove their existence under the stability condition. Similar to Lemma 4.12 in Section 4, we will prove this by assuming the converse and getting to contradiction with the tree bound.

We actually show two versions of the same lemma on existence of excellent sets. Lemma 5.6 is slightly more readable, while Lemma 5.8 is the one we will be using for further proofs, as it fixes the possibles sizes of the resulting set. For that reason, in this section we only prove the first one, and leave the proof of the other in Appendix A.

**Lemma 5.6** (Claim 5.4). *Let $G$ be a finite graph with the non-$k_*$-order property. Let $\zeta \leq \frac{1}{2^{k_{**}}}$, $\epsilon \in (0, \frac{1}{2})$. Then, for every $A \subseteq G$ with $|A| \geq \frac{1}{\epsilon^{k_{**}}}$ there exists an $(\epsilon, \zeta)$-excellent subset $A' \subseteq A$ such that $|A'| \geq \epsilon^{k_{**}-1}|A|$.*

*Proof.* Suppose the converse. We use this fact to build sets $\{b_\eta \mid \eta \in \{0,1\}^{<k_{**}}\}$ and $\{A_\eta \mid \eta \in \{0,1\}^{\leq k_{**}}\}$ on induction over $k < k_{**}$, where $k = |\eta|$, satisfying:

1. $A_{\langle \cdot \rangle} = A$.
2. $B_\eta$ is a $\zeta$-good set witnessing that $A_\eta$ is not $(\epsilon, \zeta)$-excellent, for $k < k_{**}$.
3. $A_{\eta \frown \langle i \rangle} = \{a \in A_\eta \mid t(a, B_\eta) \equiv i\}$ for all $i \in \{0,1\}$ and $k < k_{**}$.
4. $|A_{\eta \frown \langle i \rangle}| \geq \epsilon |A_\eta|$ for all $i \in \{0,1\}$ and $k < k_{**}$.
5. $|A_\eta| \geq \epsilon^k |A|$, for $k \leq k_{**}$.
6. $A_\eta = A_{\eta \frown \langle 0 \rangle} \sqcup A_{\eta \frown \langle 1 \rangle}$, for $k < k_{**}$.
7. $\overline{A_k} = \{A_\eta \mid \eta \in \{0,1\}^k\}$ is a partition of $A$, for $k \leq k_{**}$.

First of all, notice that at each step, the non-$(\epsilon, \zeta)$-excellence of $A_\eta$ comes by IH from 1. or 5., and thus allows the existence of $B_\eta$ in 2.. 4. follows the definition of $A_{\eta \frown \langle i \rangle}$ in 3. and the fact $B_\eta$ is witnessing that $A_\eta$ is not $(\epsilon, \zeta)$-excellent. Applying recursively this last point we obtain 5.. Finally, by definition 3., we have the disjoint union 6. which ensures the partition 7..

Now, our goal is to build two sequences $\{b_\eta \mid \eta \in \{0,1\}^{<k_{**}}\}$ and $\{a_\eta \mid \eta \in \{0,1\}^{k_{**}}\}$ to contradict the tree bound $k_{**}$. First of all, notice that, for $\eta \in \{0,1\}^{k_{**}}$

$$|A_\eta| \geq \epsilon^{k_{**}} |A| \geq \epsilon^{k_{**}} \frac{1}{\epsilon^{k_{**}}} = 1$$

So, for each $\eta \in \{0,1\}^{k_{**}}$, $A_\eta \neq \emptyset$ and we may choose an $a_\eta \in A_\eta$. Now, for $\nu \in \{0,1\}^{<k_{**}}$ and $\eta \in \{0,1\}^{k_{**}}$ such that $\nu \lhd \eta$, let

$$U_{\nu,\eta} = \{b \in B_\nu \mid a_\eta R b \not\equiv t(a_\eta, B_\nu)\}$$

be the subset of elements of $B_\nu$ that do not relate with $a_\eta$ in the expected way. By $\zeta$-goodness of $B_\nu$, $|U_{\nu,\eta}| < \zeta |B_\nu|$, and thus for every $\nu \in \{0,1\}^{<k_{**}}$,

$$|\bigcup \{U_{\nu,\eta} \mid \nu \lhd \eta \in \{0,1\}^{k_{**}}\}| < 2^{k_{**}} \zeta |B_\nu| \leq |B_\nu|$$

We may choose $b_\nu \in B_\nu \setminus \bigcup\{U_{\nu,\eta} \mid \nu \triangleleft \eta \in \{0,1\}^{k_{**}}\}$, for all $\nu \in \{0,1\}^{<k_{**}}$. Finally, the sequences $\langle a_\eta \mid \eta \in \{0,1\}^{k_{**}}\rangle$ and $\langle b_\nu \mid \nu \in \{0,1\}^{<k_{**}}\rangle$ satisfy that $\forall \eta, \nu$ such that $\nu \frown \langle i \rangle \triangleleft \eta$, $(a_\eta R b_\nu)^i$ by 3. and 6.. This contradicts Definition 3.14 of tree bound $k_{**}$. $\square$

*Remark* 5.7. The two sequences $\langle a_\eta \mid \eta \in \{0,1\}^{k_{**}}\rangle$ and $\langle b_\nu \mid \nu \in \{0,1\}^{<k_{**}}\rangle$ are not necessarily disjoint. This is the reason why, for this to work, the Definition 3.13, and consequently Definition 3.1, do not take this condition. Although it makes the non-$k$-order assumption on the graph stricter, this also allows the definition of excellence to work with respect to the set itself (as it is good by definition). Thus, the resulting partition will not only satisfy quasi-randomness between different parts, but actually ensures that the parts are quasi-random within themsleves.

**Lemma 5.8** (Claim 5.4.1)**.** *Let $G$ be a finite graph with the non-$k_*$-order property. Let $\zeta < \frac{1}{2^{k_{**}}}$, $\epsilon \in (0, \frac{1}{2})$. Let $\langle m_\ell \mid \ell \in \{0, \dots, k_{**}\}\rangle$ be a decreasing sequence of natural numbers such that $\epsilon m_\ell \geq m_{\ell+1}$ for all $\ell \in \{0, \dots, k_{**}-1\}$ and $m_{k_{**}} \geq 1$. Then, for every $A \subseteq G$ with $|A| \geq m_0$ there exists $(\frac{m_{\ell+1}}{m_\ell}, \zeta)$-excellent subset $A' \subseteq A$ such that $|A'| = m_\ell$ for some $\ell \in \{0, \dots, k_{**}-1\}$.*

Now, we can get the first version of a partition by applying the previous lemma recursively, until the remainder is too small for the condition on the size of the graph to be satisfied.

**Lemma 5.9** (Claim 5.14.1)**.** *Let $G$ be a finite graph with the non-$k_*$-order property. Let $\epsilon \in (0, \frac{1}{2})$ and $\epsilon' \leq \frac{1}{2^{k_{**}}}$. Let $A \subseteq G$ such that $|A| = n$. Let $\langle m_\ell \mid \ell \in \{0, \dots, k_{**}\}\rangle$ be a decreasing sequence of natural numbers such that $\epsilon m_\ell \geq m_{\ell+1}$ for all $\ell \in \{0, \dots, k_{**}-1\}$ and $m_{k_{**}} \geq 1$. Denote $m_* := m_0$ and $m_{**} := m_{k_{**}}$. Then, there is a partition $\overline{A} = \langle A_j \mid j \in \{1, \dots, j(*)\}\rangle$ with remainder $B = A \setminus \bigcup_{j<j(*)} A_j$ such that:*

*(a) For all $j \in \{1, \dots, j(*)\}$, $|A_j| \in \langle m_\ell \mid \ell \in \{0, \dots, k_{**}-1\}\rangle$.*

*(b) For all $i \neq j \in \{1, \dots, j(*)\}$, $A_i \cap A_j = \emptyset$.*

*(c) For all $j \in \{1, \dots, j(*)\}$, $A_j$ is $(\epsilon, \epsilon')$-excellent.*

*(d) $|B| < m_*$.*

*Proof.* Apply Lemma 5.8 recursively to the remainder $A \setminus \bigcup_{i<j} A_i$, to obtain $A_j$ at each step. The process stops at $j(*)$ when the remainder is smaller than $m_0$, and thus the lemma cannot be applied. Notice that, since $\frac{m_\ell}{m_{\ell-1}} \leq \epsilon$, $(\frac{m_\ell}{m_{\ell-1}}, \epsilon')$-excellence implies $(\epsilon, \epsilon')$-excellence. $\square$

The next step is refining this partition to obtain an equitable partition. In order to do so, we first show that any random sample of a given size from an excellent set is still excellent with high probability, at the cost of a slightly reduced excellence (c. of Lemma 5.11). Then, we use this result in a union-bound argument to show that we can actually fully partition the excellent set into pieces of equal size (d. of Lemma 5.11), which are still excellent. Finally, Lemma 5.15 applies this result to the partition from Lemma 5.9 to get an equitable excellent partition.

Say that if $A$ is smaller than $m_0$, then the partition is empty and $B = A$.

Before getting to it, we prove the following calculus result, which will be required in the subsequent proof. The statement comes from [**no_me_acuerdo**] and, for completeness, we provide here a short proof.

To do.

**Lemma 5.10.** *For $k > 1$, $\zeta, \eta \in (0,1)$ the function $f(m) = m^k \cdot e^{-2\zeta^2 m}$ satisfies $f(m) \leq \eta$ for all $m \geq \frac{1}{\zeta^2}(k \log \frac{1}{\zeta^2} k - \log \eta)$.*

*Proof.* First of all, notice that for $m = \frac{1}{\zeta^2}(k \log \frac{1}{\zeta^2} k - \log \eta)$,

$$f(m) = \frac{m^k}{e^{2\zeta^2 m}} = \frac{(\frac{1}{\zeta^2}(k \log \frac{1}{\zeta^2} k - \log \eta))^k}{(\frac{k}{\zeta^2})^{2k} \eta^{-2}} \leq \frac{k^k (\log \frac{k}{\zeta^2} (\frac{1}{\eta})^{\frac{1}{k}})^k}{k^k (\frac{k}{\zeta^2} (\frac{1}{\eta})^{\frac{1}{k}})^k} \eta < \eta$$

To conclude, we prove that $f$ is decreasing for larger values of $m$:

$$f'(m) = \frac{k m^{k-1} e^{2\zeta^2 m} - 2\zeta^2 m^k e^{2\zeta^2 m}}{(e^{2\zeta^2 m})^2} = (k - 2m\zeta^2) \frac{m^{k-1}}{e^{2\zeta^2 m}}$$

The second factor is always positive, and $m > \frac{k}{\zeta^2} > \frac{k}{2\zeta^2}$, proving that $f'(m) < 0$ and thus $f$ is decreasing. □

**Lemma 5.11** (Claim 5.13)**.** *Let $G$ be a finite graph with the non-$k_*$-order property. Then:*

*(a) For every $\epsilon \in (0, \frac{1}{2})$, $\zeta \in (0, \frac{1}{2} - \epsilon)$, $\xi \in (0, 1)$ and $m \geq \frac{1}{\zeta^2}(k_* \log \frac{1}{\zeta^2} k_* - \log \xi)$, if $A \subseteq G$ is an $\epsilon$-good subset of size $n \geq m$, then a random subset $A' \subseteq A$ of size $m$ is $(\epsilon + \zeta)$-good with probability $1 - \xi$.*

*(b) Moreover, such $A'$ satisfies $t(b, A') = t(b, A)$ for all $b \in G$.*

*(c) For every $\zeta \in \{0, \frac{1}{2}\}$ and $\zeta' < \zeta$, there is $\epsilon_1 = \epsilon_1(\zeta, \zeta')$ such that for every $\epsilon < \epsilon' \leq \epsilon_1$, if*

- *$A \subseteq G$ is $\{\epsilon, \epsilon'\}$-excellent.*
- *$A' \subseteq A$ is $\{\epsilon + \zeta'\}$-good.*

*then, $A'$ is $(\epsilon + \zeta, \epsilon')$-excellent.*

*(d) For all $\zeta \in (0, \frac{1}{2})$, $\zeta' < \zeta$, $r \geq 1$ and for all $\epsilon < \epsilon'$ small enough (in the sense of the previous point) there exists $N = N(k_*, \zeta', r)$ such that, if $|A| = n > N$, $r$ divides $n$ and $A$ is $(\epsilon, \epsilon')$-excellent, there exists a partition into $r$ disjoint pieces of equal size, each of which is $(\epsilon + \zeta, \epsilon')$-excellent.*

*Proof.* (a) For each $b \in G$, we say that $B_{A,b}$ is *bad* if $|B_{A,b}| \geq \epsilon |A'|$. For each bad $B_{A,b}$, let $X_{A,b}$ be the event that $|B_{A,b}| \geq (\epsilon + \zeta)|A'|$ for a random subset $A' \subseteq A$ of size $m$. Notice that $X_{A,b}$ is modelled by a hypergeometric distribution, and so the probability of upperly deviating from the mean by $\zeta$, can be modeled by

$$P(X_{A,b} = 1) \leq e^{-2\zeta^2 m}$$

Now we want to study the random variable $X$ counting the number of events $X_{A,b}$ that are satisfied. That is, $X = \sum_{\text{bad } B_{A,b}} X_{A,b}$. We compute the expectation

$$\mathbb{E}[X] = \sum_{\text{bad } B_{A,b}} \mathbb{E}[X_{A,b}] = \sum_{\text{bad } B_{A,b}} P(X_{A,b} = 1) \leq \sum_{\text{bad } B_{A,b}} e^{-2\zeta^2 m}$$

Following 2., the number of intersections of bad $B_{A,b}$'s with $A'$, can be bounded by $m^{k_*}$. Thus, using the First Moment Method, we have that:

$$P(X \geq 1) \leq \mathbb{E}[X] \leq m^{k_*} \cdot e^{-2\zeta^2 m} \leq \xi$$

Last inequality follows Lemma 5.10 using the lower bound on $m$. Thus, with probability at least $1 - \xi$, we have that $A'$ is $(\epsilon + \zeta)$-good.

(b) Suppose that $A'$ is the subset described in a.. We proved that, such set satisfies

$$|A' \cap B_{A,b}| < (\epsilon + \zeta)|A'|$$

for all $b \in G$ such that $|B_{A,b}| \geq \epsilon m$. Thus, we have that:

- If $|B_{A,b}| < \epsilon m$, then $|\{a \in A' \mid aRb \not\equiv t(b, A)\}| \leq |B_{A,b}| < \epsilon m < (\epsilon + \zeta)m$.
- If $|B_{A,b}| \geq \epsilon m$, then $|\{a \in A' \mid aRb \not\equiv t(b, A)\}| = |A' \cap B_{A,b}| < (\epsilon + \zeta)m$.

We conclude that $t(b, A) = t(b, A')$ for all $b \in G$.

(c) Let $B \subseteq G$ be an $\epsilon'$-good set. We first upperbound the number of exceptional vertices of $B$ with respect to $A'$:

$$\begin{aligned}
|\{b \in B \mid t(b, A') \not\equiv t(B, A)\}| &= |\{b \in B \mid t(b, A) \not\equiv t(B, A)\}| \\
&\leq \frac{(\epsilon + (1-\epsilon)\epsilon')|A||B|}{(1-\epsilon)|A|} \\
&= (\epsilon' + \frac{\epsilon}{1-\epsilon})|B|
\end{aligned}$$

The first equality follows b., and the first inequality follows from Remark 5.4 for the numerator, and taking the worst case of only $(1-\epsilon)|A|$ exceptional edges per exceptional $b \in B$ (considering that $A$ is $\epsilon$-good).

Now, let $Q$ be the set of exceptional vertices of $A'$ with respect to $B$, i.e.:

$$Q = \{a \in A' \mid t(a, B) \not\equiv t(A, B)\}$$

We want to double-count the number of exceptional edges between $Q$ and $B$. On one hand, we have that:

$$|\{(a, b) \in Q \times B \mid t(a, b) \not\equiv t(A, B)\}| < (\epsilon' + \frac{\epsilon}{1-\epsilon})|B||Q| + (1 - \epsilon' - \frac{\epsilon}{1-\epsilon})|B|(\epsilon + \zeta')|A'|$$

The first term is the maximum number of exceptional edges associated to exceptional $b \in B$ (considering all edges exceptional), while the second term bounds the number of exceptional edges of non-exceptional $b \in B$, using the fact that $A'$ is $(\epsilon + \zeta')$-good.

On the other hand, we have that:

$$|\{(a, b) \in Q \times B \mid t(a, b) \not\equiv t(A, B)\}| \geq |Q|(1 - \epsilon')|B|$$

which follows $B$ being $\epsilon'$-good.

Putting it all together:

$$(1 - \epsilon' - \epsilon' - \frac{\epsilon}{1-\epsilon})|B||Q| < (1 - \epsilon' + \frac{\epsilon}{1-\epsilon})(\epsilon + \zeta')|B||A'|$$

So, we have that:

$$\begin{aligned}
|Q| &< \frac{(1 - \epsilon' - \frac{\epsilon}{1-\epsilon})}{(1 - \epsilon' - \frac{\epsilon}{1-\epsilon}) - \epsilon'}(\epsilon + \zeta')|A'| \\
&= (1 + \frac{\epsilon'}{1 - 2\epsilon' - \frac{\epsilon}{1-\epsilon}})(\epsilon + \zeta')|A'|
\end{aligned}$$

Notice that $f(\epsilon, \epsilon') := \frac{\epsilon'}{1-2\epsilon'-\frac{\epsilon}{1-\epsilon}}$ decreases with $\epsilon$ and $\epsilon'$. In particular,

$$f(\epsilon, \epsilon') \xrightarrow{\epsilon' \to 0} 0$$

and $\epsilon' > \epsilon$. Then,

$$|Q| < (\epsilon + (\underbrace{\epsilon f(\epsilon, \epsilon')}_{\to 0} + \underbrace{(1 + f(\epsilon, \epsilon'))}_{\to 1})\zeta')|A'| \xrightarrow{\epsilon' \to 0} (\epsilon + \zeta')|A'|$$

So, there exists an $\epsilon_1 = \epsilon_1(\zeta, \zeta')$ small enough such that for all $(\epsilon <)\ \epsilon' \leq \epsilon_1$, we have that $|Q| < (\epsilon + \zeta)|A'|$, and since $A'$ is $(\epsilon + \zeta')$-good, and thus $(\epsilon + \zeta)$-good, we conclude that $A'$ is $(\epsilon + \zeta, \epsilon')$-excellent.

Mention that in the next claim we show valid values for this.

(d) Let $\zeta, \zeta', \epsilon, \epsilon'$ and $r$ be given satisfying the conditions of the statement. Set $\xi = \frac{1}{r+1}$. We will see that the condition $n > N = N(k_*, \zeta', r) := r\frac{1}{\zeta'^2}(k_* \log \frac{1}{\zeta'^2} k_* - \log \frac{1}{r+1})$ is sufficient. First of all, randomly choose a function $h : A \longrightarrow \{1, \dots, r-1\}$ such that for all $s < n$ we have that $|\{a \in A \mid h(a) = s\}| = \frac{n}{r}$. Since $h$ is random, each $A' \in [A]^{\frac{n}{r}}$ has the same probability of being part of the partition induced by $h$, i.e. to satisfy $A' = h^{-1}(s)$ for some $s \in \{1, \dots, r-1\}$. Since each element of the partition $A'$ has size $\frac{n}{r} > \frac{N}{r} = \frac{1}{\zeta'^2}(k_* \log \frac{1}{\zeta'^2} k_* - \log \xi)$, we can apply a. to get that

$$P(A' \text{ is not } (\epsilon + \zeta')\text{-good}) < \xi$$

In particular, since $A$ is $(\epsilon, \epsilon')$-excellent, it follows c. that if $A'$ is $(\epsilon + \zeta')$-good then it is also $(\epsilon + \zeta, \epsilon')$-excellent, so:

$$P(A' \text{ is not } (\epsilon + \zeta, \epsilon')\text{-excellent}) < \xi$$

To conclude, by the union bound, we have that:

$$\begin{aligned} P(\bigcup_{s<r} h^{-1}(s) \text{ is not } (\epsilon + \zeta, \epsilon')\text{-excellent}) &\leq \sum_{s<r} P(h^{-1}(s) \text{ is not } (\epsilon + \zeta, \epsilon')\text{-excellent}) \\ &< r\xi = \frac{r}{r+1} < 1 \end{aligned}$$

All in all, there is a non-zero chance that the partition satisfies the statement, i.e. there exists at least one.

□

*Remark* 5.12 (Remark 5.13.1). For following applications, we would like to use d. from Lemma 5.11 with $\epsilon' > k(\epsilon + \zeta)$, for an arbitrarily large $k \in \mathbb{N}$. Notice that if $\epsilon, \zeta' \leq \frac{1}{t}, \epsilon' \leq \frac{1}{t'}$ and $t > t' \geq 5$, then:

(a) $\frac{\epsilon}{1-\epsilon} \leq \frac{\frac{1}{t}}{1-\frac{1}{t}} = \frac{\frac{1}{t}}{\frac{t-1}{t}} = \frac{1}{t-1}$

(b) $1 - 2\epsilon' - \frac{\epsilon}{1-\epsilon} \geq 1 - \frac{2}{t'} - \frac{1}{t-1} > 1 - \frac{3}{t'-1} = \frac{t'-4}{t'-1}$

(c) $(1 + \frac{\epsilon'}{1-2\epsilon'-\frac{\epsilon}{1-\epsilon}}) < 1 + \frac{\epsilon'}{1-\frac{3}{t'-1}} = (1 + \frac{t'-1}{t'-4}\epsilon')(\epsilon + \zeta')$

Then, by requiring $\frac{1}{t} \leq \frac{1}{4k}\epsilon'$ we have that

$$\begin{aligned}
\epsilon + \zeta' &\leq \frac{2}{t} \leq 2(\frac{1}{4k}\epsilon') = \frac{1}{2}(\frac{1}{k}\epsilon') \\
&< \frac{t'-4}{t'-3}\frac{1}{k}\epsilon' = \frac{1}{k}\frac{\epsilon'}{1+\frac{1}{t'-4}} \\
&< \frac{1}{k}\frac{\epsilon'}{1+\frac{t'-1}{t'}\frac{1}{t'-4}} = \frac{1}{k}\frac{\epsilon'}{1+\frac{t'-1}{t'-4}\frac{1}{t'}} \\
&\leq \frac{1}{k}\frac{\epsilon'}{1+\frac{t'-1}{t'-4}\epsilon'}
\end{aligned}$$

i.e., we have:

$$(1+\frac{t'-1}{t'-4}\epsilon')(\epsilon+\zeta') < \frac{1}{k}\epsilon'$$

which by c. gives us:

$$(1+\frac{\epsilon'}{1-2\epsilon'-\frac{\epsilon}{1-\epsilon}}) < \frac{1}{k}\epsilon'$$

All in all, a sufficient condition, for the lemma to hold under the constraint $\epsilon' \geq k(\epsilon+\zeta)$, is:

$$\epsilon, \zeta' \leq \frac{1}{4k}\epsilon' \quad \text{and} \quad \epsilon' \leq \frac{1}{5}$$

We use this fact to reformulate point d. of Lemma 5.11 as:

**Lemma 5.13** (Claim 5.13.2(3))**.** *Let $G$ be a finite graph with the non-$k_*$-property. For all $k, r \geq 1$, $\epsilon' \leq \frac{1}{5}$ and $\epsilon \leq \frac{1}{4k}\epsilon'$, there exists $N = N(k, k_*, \epsilon', r)$ large enough such that, for all $n > N$ and $r$ dividing $n$, if $A \subseteq G$ is $(\epsilon, \epsilon')$-excellent, with $|A| = n$, then there exists a partition into $r$ disjoint pieces of equal size, each of which is $(\frac{\epsilon'}{k}, \epsilon')$-excellent.*

*Proof.* Choose any $\zeta' \leq \frac{1}{4k}\epsilon'$ and set $N := N_{5.11}(k_*, \zeta', r)$. Remark 5.12 sufficiency condition is satisfied, d. from Lemma 5.11 holds and we are done. □

*Remark* 5.14. A sufficient condition for $N_{5.13}$ to be large enough is to choose $\zeta' = \frac{1}{4k}\epsilon'$ in which case $N_{5.13}(k, k_*, \epsilon', r) := N_{5.11}(k_*, \frac{1}{4k}\epsilon', r)$

Now we proceed to refine the partition from Lemma 5.9 into an equitable one.

**Lemma 5.15** (Claim 5.14.1a)**.** *Let $G$ be a finite graph with the non-$k_*$-order property. Let $\epsilon'$ and $\epsilon$ be two real numbers such that $\epsilon' \leq \min(\frac{1}{5}, \frac{1}{2^{k_{**}}})$ and $\epsilon \leq \frac{1}{4k}\epsilon'$ for some $k > 1$. Also, let $m_*$, $m_{**}$ and $q$ be natural numbers such that $q \geq \lceil \frac{1}{\epsilon} \rceil$, $m_{**} > \frac{N_{5.13}(k, k_*, \epsilon', \frac{m_*}{m_{**}})}{q}$ and $m_* := q^{k_{**}} m_{**}$. Then, for any $A \subseteq G$ with $|A| = n \geq m_*$ there exists a partition $\overline{A} = \langle A_i \mid i \in \{1, \dots, i(*)\} \rangle$ with remainder $B = A \setminus \bigcup \overline{A}$ such that:*

(a) *$i(*) \leq \frac{n}{m_{**}}$.*

(b) *For all $i \in \{1, \dots, i(*)\}$, $|A_i| = m_{**}$.*

(c) *For all $i \in \{1, \dots, i(*)\}$, $A_i$ is $(\frac{\epsilon'}{k}, \epsilon')$-excellent.*

(d) *$|B| < m_*$.*

*Proof.* Consider the decreasing sequence of natural numbers

$$m_0 \geq m_1 \geq \cdots \geq m_{k_{**}} = m_{**}$$

defined by $m_\ell = qm_{\ell+1}$, so that it satisfies $m_\ell \geq \frac{m_{\ell+1}}{\epsilon}$ for all $\ell \in \{0, \dots, k_{**}-1\}$. Then $m_0 = q^{k_{**}} m_{**} = m_* \leq n$, and $m_{k_{**}-1} = qm_{**} > N_{5.13}(k, k_*, \epsilon', \frac{m_*}{m_{**}})$. With such a sequence, we can apply Lemma 5.9 to $A$ to obtain a partition $\overline{A}' = \left\langle A'_j \mid j \in \{1, \dots, j(*)\} \right\rangle$ and remainder $B$ with $|B| < m_*$. Then, we can apply Lemma 5.13 to each of the parts $A'_j$ with $r = \frac{m_*}{m_{**}}$, as $m_{**} \mid m_\ell$ for all $\ell \in \{0, \dots, k_{**}-1\}$. Putting together all the new subparts, we obtain a new partition $\overline{A} = \langle A_i \mid i \in \{1, \dots, i(*)\} \rangle$ with remainder $B$, satisfying all the conditions of the statement. □

Notice that our partition still has reminder, which is unwanted and, as the next lemma proves, it is avoidable at the cost of another slight increase of the excellence parameter.

**Lemma 5.16** (Claim 5.14.2)**.** *Under the same condition of Lemma 5.15, we can get a partition $\overline{A} = \langle A_i \mid i \in \{1, \dots, i(*)\} \rangle$ with no remainder, such that:*

*(a) For all $i, j \in \{1, \dots, i(*)\}$, $||A_i| - |A_j|| \leq 1$.*

*(b) For all $i, j \in \{1, \dots, i(*)\}$, $A_i \cap A_j = \emptyset$.*

*(c) For all $i \in \{1, \dots, i(*)\}$, $A_i$ is $(\epsilon'', \epsilon')$-excellent, where*

$$\epsilon'' \leq \frac{\frac{\epsilon'}{k} m_{**} + \left\lceil \frac{m_*}{i(*)} \right\rceil}{m_{**} + \left\lceil \frac{m_*}{i(*)} \right\rceil}$$

*(d) $A = \bigcup \overline{A}$.*

*Proof.* Let $\overline{A}' = \langle A'_i \mid i \in \{1, \dots, i(*)\} \rangle$ and $B$ from Lemma 5.15. We can partition $B$ into $\overline{B} = \langle B_i \mid i \in \{1, \dots, i(*)\} \rangle$ in such a way that for all $i \in \{1, \dots, i(*)\}$,

$$|B_i| \in \{ \left\lfloor \frac{|B|}{i(*)} \right\rfloor, \left\lceil \frac{|B|}{i(*)} \right\rceil \}$$

Notice that we are allowing $B_i = \emptyset$. Then, the new partition $\overline{A} = \langle A'_i \cup B_i \mid i \in \{1, \dots, i(*)\} \rangle$ satisfies a., b. and d. by construction. To conclude, notice that for each $\epsilon'$-good set $B$, the number of exceptions is bounded by

$$\begin{aligned} |\{a \in A_i \mid t(a, B) \not\equiv t(A_i, B)\}| &\leq \frac{\epsilon'}{k} |A'_i| + |B_i| \\ &= \frac{\frac{\epsilon'}{k}|A'_i| + |B_i|}{|A'_i| + |B_i|} (|A'_i| + |B_i|) \\ &\leq \frac{\frac{\epsilon'}{k} m_{**} + \left\lceil \frac{m_*}{i(*)} \right\rceil}{m_{**} + \left\lceil \frac{m_*}{i(*)} \right\rceil} |A_i| \end{aligned}$$

which proves that c. can be satisfied. □

We now have an $(\epsilon'', \epsilon')$-excellent equitable partition with no remainder. Also $\epsilon''$ is bounded by something very close to $\frac{\epsilon'}{k}$, where $k$ is a settable parameter which only affects the large-enough condition on the size of the graph. It is reasonable to assume that, under some conditions of $m_*$ and $m_{**}$, and under an appropriate choice of $k$, we can upper bound $\epsilon''$ by $\epsilon'$, thus ensuring that the partition is $\epsilon'$-excellent.

*Remark* 5.17 (Remark 5.14.3). In the context of Lemma 5.16, if:

(a) $m_{**} \geq \frac{1}{\frac{\epsilon'}{k}}$

(b) $m_* \leq \frac{\frac{\epsilon'}{k}n+1}{\frac{\epsilon'}{k}+1}$

then $\epsilon'' \leq \frac{3\epsilon'}{k}$.

*Proof.* Notice that, if $|B_i| \leq 2\frac{\epsilon'}{k}|A_i|$ for all $i \in \{1, \dots, i(*)\}$, then $\epsilon''$ can be bounded by:

$$\epsilon'' \leq \frac{\frac{\epsilon'}{k}|A_i| + |B_i|}{|A_i| + |B_i|} \leq \frac{\frac{\epsilon'}{k}|A_i| + 2\frac{\epsilon'}{k}|A_i|}{|A_i|} = \frac{3\epsilon'}{k}$$

Let's now prove that $|B_i| \leq \frac{2\epsilon'}{k}|A_i|$ is satisfied. Notice that, by construction:

$$|B_i| \leq \left\lceil \frac{|B|}{i(*)} \right\rceil \leq \left\lceil \frac{m_* - 1}{i(*)} \right\rceil \leq \frac{m_* - 1}{i(*)} + 1$$

Also we can bound $i(*)$ by:

$$\frac{n}{m_{**}} \geq i(*) \geq \frac{n - |B|}{m_{**}} \geq \frac{n - m_* + 1}{m_{**}} > \frac{n - m_*}{m_{**}}$$

Thus, $|B_i| - 1 \leq \frac{m_*-1}{i(*)} \leq \frac{(m_*-1)m_{**}}{n-m_*}$, then $\frac{|B_i|-1}{m_{**}} \leq \frac{m_*-1}{n-m_*}$, and since $|A_i| = m_{**}$ we get: Is the lower bound needed?

$$\frac{|B_i|}{|A_i|} \leq \frac{m_* - 1}{n - m_*} + \frac{1}{m_{**}}$$

Finally, notice that condition a. implies:

$$\frac{\epsilon'}{k} \geq \frac{1}{m_{**}}$$

and condition b. implies:

$$\frac{\epsilon'}{k} \geq \frac{m_* - 1}{n - m_*}$$

We conclude:

$$\frac{|B_i|}{|A_i|} \leq \frac{m_* - 1}{n - m_*} + \frac{1}{m_{**}} \leq 2\frac{\epsilon'}{k}$$

completing the proof. □

We now resume all the conditions necessaries for the previous result to hold in the context of the values $m_*$ and $m_{**}$ given by the previous remark.

**Lemma 5.18** (Corollary 5.15). *Let $G$ be a graph with the non-$k_*$-order property. Suppose that we are given:*

1. *A real value* $\epsilon \leq \min(\frac{1}{5}, \frac{1}{2^{k_{**}}})$.
2. *Three natural numbers* $m_*$, $m_{**}$ *and* $q$ *such that:*
   (a) $q \geq \lceil \frac{1}{\epsilon} \rceil$.
   (b) $m_{**} > \frac{N_{5.13}(3, k_*, \epsilon, \frac{m_*}{m_{**}})}{q}$
   (c) $m_* := q^{k_{**}} m_{**}$.
3. $A \subseteq G$ *such that* $|A| = n$, *where* $n$ *is large enough to satisfy* $m_* \leq \frac{1+\frac{\epsilon}{3}n}{1+\frac{\epsilon}{3}}$.

*Then, there exists* $i(*) \leq \frac{n}{m_{**}}$ *and a partition of* $A$ *into disjoint pieces* $\overline{A} = \langle A_i \mid i \in \{1, \dots, i(*)\}\rangle$ *such that:*

(i) *For all* $i, j \in \{1, \dots, i(*)\}$, $||A_i| - |A_j|| \leq 1$.

(ii) *For all* $i \in \{1, \dots, i(*)\}$, $A_i$ *is* $\epsilon$*-excellent,*

(iii) *For all* $i, j \in \{1, \dots, i(*)\}$, $(A_i, A_j)$ *is* $\epsilon$*-uniform.*

*Proof.* First of all, notice that condition 2.b. is a tighter bound then $m_{**} \geq \frac{3}{\epsilon}$. To prove the statement, we simply apply Lemma 5.16 in the context of Remark 5.17 with $k = 3$, $\epsilon'_{5.16} = \epsilon$ and $\epsilon_{5.16} \leq \frac{1}{12}\epsilon$. This results in a partition of $A$ into disjoint pieces that satisfy i. and that are $(\epsilon''_{5.16}, \epsilon'_{5.16})$-excellent, with $\epsilon''_{5.16} \leq \frac{3\epsilon'_{5.16}}{k}$. But since $k \geq 3$, $\epsilon''_{5.16} \leq \epsilon'_{5.16}$, they are also $\epsilon'_{5.16}$-excellent, satisfying ii. and iii.. □

To conclude, we prove that the conditions of the previous lemma can be satisfied, under some minimal conditions of the two parameters $\epsilon$ (the excellence parameter) and $m$ (the minimum number of parts in the resulting partition), and rewrite the statement accordingly.

**Theorem 5.19** (Theorem 5.18)**.** *Let* $k_*$ *and therefore* $k_{**}$ *be given. Then, for all* $\epsilon \leq \min(\frac{1}{5}, \frac{1}{2^{k_{**}}})$ *and* $m > 1$, *there is* $M = M(\epsilon, m, k_*)$ *and* $N = N(\epsilon, m, k_*)$ *such that, for every finite graph* $G$ *with the non-*$k_*$*-order property, and every* $A \subseteq G$ *with* $|A| \geq N$, *there exists a partition* $\overline{A} = \langle A_i \mid i \in \{1, \dots, i(*)\}\rangle$ *of* $A$, *such that:*

Move the bound on $M$ to another point?

1. *The number of parts is bounded by* $m \leq i(*) \leq M := \max(\lceil \frac{12}{\epsilon} \rceil^{k_{**}+1}, 4m)$.
2. *For all* $i, j \in \{1, \dots, i(*)\}$, $||A_i| - |A_j|| \leq 1$.
3. *For all* $i \in \{1, \dots, i(*)\}$, $A_i$ *is* $\epsilon$*-excellent.*

Redundant?

4. *For all* $i, j \in \{1, \dots, i(*)\}$, $(A_i, A_j)$ *is* $\epsilon$*-uniform.*

*Proof.* Our goal is to apply Lemma 5.18. Let $q = \lceil \frac{12}{\epsilon} \rceil$. For $N(\epsilon, m, k_*)$, and thus $n$, large enough, we can then choose the smallest $m_{**}$ satisfying:

(a) $m_{**} \in [\delta n - 1, \delta n]$, where $\delta = \min(\frac{\epsilon}{(3+\epsilon)q^{k_{**}}}, \frac{1}{m+q^{k_{**}}})$

(b) $m_{**} > \frac{3}{\epsilon}$.

(c) $m_{**} > \frac{N_{5.13}(3, k_*, \epsilon, q^{k_{**}})}{q}$.

By a. we have that $m_* \leq \frac{\epsilon n}{3+\epsilon}$. This sequence satisfies all the conditions of Lemma 5.18:

2.a. $q \geq \lceil \frac{1}{\epsilon} \rceil$, and in particular defined it to be equal.

2.b. $m_{**} > \frac{N_{5.13}(3, k_*, \epsilon, \frac{m_*}{m_{**}})}{q}$ by choice of $m_{**}$.

2.c. $m_* := q^{k_{**}} m_{**}$.

3. $m_{k_{**}-1} = qm_{**} > q\frac{N_{5.13}(3, k_*, \epsilon, q^{k_{**}})}{q} = N_{5.13}(3, k_*, \epsilon, \frac{m_*}{m_{**}})$.

We can apply Lemma 5.18 to obtain a partition satisfying 2., 3. and 4..
We proceed to bound the number of parst $i(*)$. First, the upper bound follows from the fact that $m_{**} \geq \frac{1}{2}\min(\frac{\epsilon}{3+\epsilon}, \frac{1}{m+q^{k_{**}}})n$:

$$i(*) \leq \frac{n}{m_{**}} \leq \frac{2\max(\frac{3+\epsilon}{\epsilon}q^{k_{**}}, m+q^{k_{**}})n}{n} < 2\max(\frac{3+\epsilon}{\epsilon}q^{k_{**}}, 2m) \leq \max(\left\lceil \frac{12}{\epsilon} \right\rceil^{k_{**}+1}, 4m)$$

In the last inequality, we used that if $m < q^{k_{**}}$, then $m + q^{k_{**}} \leq 2q^{k_{**}} < \frac{3+\epsilon}{\epsilon}q^{k_{**}}$, which is dealt in the first argument of the maximum, so we may assume that $m \geq q^{k_{**}}$. We also show that the lower bound is satisfied:

$$i(*) \geq \frac{n - m_*}{m_{**}} \geq \frac{n - m_{**}q^{k_{**}}}{m_{**}} = \frac{n}{m_{**}} - q^{k_{**}} \geq \frac{m + q^{k_{**}}}{n}n - q^{k_{**}} = m$$

□

*Remark* 5.20. We now see how large $N$, and thus $n$, actually needs to be. First of all, we see that:

$$\begin{aligned}\frac{1}{q}N_{5.13}(4, k_*, \epsilon, q^{k_{**}}) &= \frac{1}{q}N_{5.11}(k_*, \frac{1}{4\cdot 3}\epsilon, q^{k_{**}}) \\ &= \frac{1}{q}q^{k_{**}}(\frac{12}{\epsilon})^2(k_* \log(\frac{12}{\epsilon})^2 k_* - \log\frac{1}{q^{k_{**}}+1}) \\ &< k_*^2 q^{2k_{**}+3}\end{aligned}$$

Also, $\frac{3}{\epsilon}$ is clearly smaller than this value. Then, since $m_{**}$ is the smallest integer larger than both values, we conclude:

$$\begin{aligned}\frac{m_{**}}{\delta} &\leq \frac{k_*^2 q^{2k_{**}+3}}{\min(\frac{\epsilon}{(3+\epsilon)q^{k_{**}}}, \frac{1}{m+q^{k_{**}}})} \\ &= k_*^2 q^{2k_{**}+3}\max(\frac{3+\epsilon}{\epsilon}q^{k_{**}}, m+q^{k_{**}}) \\ &\leq \max(q^{k_{**}+1}, 4m)k_*^2 q^{2k_{**}+3}\end{aligned}$$

Define or remove uniformity.

## 5.3 Stable Regularity Lemma

Lluis: is it ok to call a subsection as the section?

As mentioned in the beginning of the section, it can be proven that excellence is a stronger condition then regularity. In fact, as shown in the following lemma, excellence of a pair not only implies some level of regularity, but also it ensures that the pair is mostly full or empty of edges.

Mention homogeneity?

**Lemma 5.21** (Lemma 5.17)**.** *Suppose that $\epsilon_1, \epsilon_2, \epsilon_3 \in (0, \frac{1}{2})$ with $\frac{\epsilon_1+\epsilon_2}{\epsilon_3} < \frac{1}{2}$ and the pair $(A, B)$ is $(\epsilon_1, \epsilon_2)$-uniform. Let $A' \subseteq A$ with $|A'| \geq \epsilon_3|A|$, $B' \subseteq B$ with $|B'| \geq \epsilon_3|B|$ and denote $Z = \{(a, b) \in (A \times B) \mid aRb \not\equiv t(A, B)\}$ and $Z' = \{(a, b) \in (A' \times B') \mid aRb \not\equiv t(A, B)\}$. Then, we have:*

1. $\frac{|Z|}{|A||B|} < \epsilon_1 + \epsilon_2$.
2. $\frac{|Z'|}{|A||B|} < \frac{\epsilon_1+\epsilon_2}{\epsilon_3}$.

*In particular, if for some $\epsilon_0, \epsilon \in (0, \frac{1}{2})$, the pair $(A, B)$ is $\epsilon_0$-uniform, for $\epsilon_0 \leq \frac{\epsilon^2}{2}$, then:*

a. *$(A, B)$ is $\epsilon$-regular.*

b. *If $A' \in [A]^{\geq \epsilon|A|}$and $B' \in [B]^{\geq \epsilon|B|}$, then $d(A', B') < \epsilon$ or $d(A', B') \geq 1 - \epsilon$.*

*Proof.* Let $U = \{a \in A \mid |\overline{B}_{B,a}| > \epsilon_1|A|\}$, i.e. the set of exceptional vertices $a \in A$. Then,

$$Z \subseteq U \times B \cup \bigcup_{a \in A \setminus U} \{a\} \times \overline{B}_{B,a}$$

and

$$Z' \subseteq U \times B' \cup \bigcup_{a \in A' \setminus U} \{a\} \times \overline{B}_{B,a}$$

Notice that, if $a \in A \setminus U$, then $|\overline{B}_{B,a}| < \epsilon_2|B|$, so

$$|Z| < \epsilon_1|A||B| + |A|\epsilon_2|B|$$

which can be written as

$$\frac{|Z|}{|A||B|} < \epsilon_1 + \epsilon_2$$

which proves 1.. Similarly,

$$\begin{aligned} |Z'| &\leq |U||B'| + |A'| \max\{|\overline{B}_{B,a}| \mid a \notin U\} \\ &< \epsilon_1|A||B'| + |A'|\epsilon_2|B| \end{aligned}$$

By dividing both sides by $|A'||B'|$ we conclude

$$\frac{|Z'|}{|A'||B'|} < \epsilon_1\frac{|A|}{|A'|} + \epsilon_2\frac{|B|}{|B'|} \leq \frac{\epsilon_1|A|}{\epsilon_3|A|} + \frac{\epsilon_2|B|}{\epsilon_3|B|} = \frac{\epsilon_1 + \epsilon_2}{\epsilon_3}$$

proving 2.. Let's now prove a. and b.. First of all, notice that:

This only works so nice when $A$ and $B$ are disjoint. Check what happens when they are not. Something more on the line of $d(A, B) > 1 - 4(\epsilon_1 + \epsilon_2)$

- if $t(A, B) = 1$, then $d(A, B) > 1 - (\epsilon_1 + \epsilon_2)$ and $d(A', B') > 1 - \frac{\epsilon_1+\epsilon_2}{\epsilon_3}$, which follows 1. and 2. respectively. Thus,

$$\begin{aligned} |d(A, B) - d(A', B')| &\leq \max\{d(A, B) - d(A', B'), d(A', B') - d(A, B)\} \\ &< \max\{1 - (1 - \frac{\epsilon_1 + \epsilon_2}{\epsilon_3}), 1 - (1 - \epsilon_1 - \epsilon_2)\} \\ &= \frac{\epsilon_1 + \epsilon_2}{\epsilon_3} \end{aligned}$$

- if $\underline{t(A,B) = 0}$, similarly $d(A,B) < (\epsilon_1 + \epsilon_2)$ and $d(A',B') < \frac{\epsilon_1+\epsilon_2}{\epsilon_3}$. Thus,

$$\begin{aligned}|d(A,B) - d(A',B')| &\leq \max\{d(A,B) - d(A',B'), d(A',B) - d(A,B)\} \\ &< \max\{(\epsilon_1+\epsilon_2), \frac{\epsilon_1+\epsilon_2}{\epsilon_3}\} \\ &= \frac{\epsilon_1+\epsilon_2}{\epsilon_3}\end{aligned}$$

In both cases, we have that $|d(A,B) - d(A',B')|$ is bounded by $\frac{\epsilon_1+\epsilon_2}{\epsilon_3} < \frac{1}{2}$. Also, $d(A',B')$ may only differ by $\frac{\epsilon_1+\epsilon_2}{\epsilon_3}$ with either 0 or 1. In particular, we may choose $\epsilon_3 = \epsilon$ and $\epsilon_1 = \epsilon_2 = \epsilon_0 \leq \frac{\epsilon^2}{2}$. This way, the condition $\frac{\epsilon_1+\epsilon_2}{\epsilon_3} \leq \epsilon < \frac{1}{2}$ is satisfied. We conclude that $(A,B)$ is $\epsilon$-regular (a.) and that $d(A',B')$ is either $< \epsilon$ or $\geq 1-\epsilon$ (b.). □

We finally prove the Stable Regularity Lemma using the previous lemma to reformulate Theorem 5.19 in the context of regularity.

**Theorem 5.22** (Theorem 5.19). *For every $k_* \in \mathbb{N}$ and $\epsilon \in (0, \frac{1}{2})$ and $m > 1$, there exist $N = N(\epsilon, m, k_*)$ and $M = M(\epsilon, m, k_*)$ such that, for every finite graph $G$ with the non-$k_*$-order property, and every $A \subseteq G$ with $|A| \geq N$, there is $m < \ell < M$ and a partition $\overline{A} = \langle A_i \mid i \in \{1, \dots, \ell\}\rangle$ of $A$ such that each $A_i$ is $\frac{\epsilon^2}{2}$-excellent, and for every $i, j \in \{1, \dots, \ell\}$,*

1. *$||A_i| - |A_j|| \leq 1$.*
2. *$(A_i, A_j)$ is $\epsilon$-regular, and moreover if $B_i \in [A_i]^{\geq \epsilon |A_i|}$ and $B_j \in [A_j]^{\geq \epsilon |A_j|}$, then either $d(B_i, B_j) < \epsilon$ or $d(B_i, B_j) \geq 1 - \epsilon$.*
3. *If $\epsilon \leq \min(\frac{1}{5}, \frac{1}{2^{k_{**}}})$, then $M \leq \max(\lceil \frac{12}{\epsilon} \rceil^{k_{**}+1}, 4m)$.*

*Proof.* If $\epsilon \leq \min(\frac{1}{5}, \frac{1}{2^{k_{**}}})$, then we can apply Theorem 5.19 to $A$ with $\frac{\epsilon^2}{2}$, and then use Lemma 5.21 to replace the $\frac{\epsilon^2}{2}$-uniformity of pairs by $\epsilon$-regularity. Otherwise, to get 1. and 2., just do the same process for some $\epsilon' = \min(\frac{1}{5}, \frac{1}{2^{k_{**}}}) \leq \epsilon$. Then, since regularity is monotone, we get the wanted $\epsilon$-regularity from the resulting $\epsilon'$-regularity. In this last case, the bound on $M$ is $M \leq \max(\lceil \frac{12}{\epsilon'} \rceil^{k_{**}+1}, 4m)$. □

*Remark* 5.23. By Theorem 3.15, we have that $k_{**} \leq 2^{k_*+1} - 2$ in the context of the non-$k_*$-order property. Thus, the bound on the number of parts $M$ can clearly be reformulated as a function of only $k_*$, $\epsilon$ and $m$:

$$M \leq \max(\left\lceil \frac{12}{\epsilon} \right\rceil^{2^{k_*+1}-1}, 4m)$$

# 6. Property Testing

Property testing is a field of theoretical computer science, concerned about finding low-complexity algorithms for testing (approximate) properties in large objects, such as graphs. These algorithms need to be successful with high probability, and are only required to distinguish between objects that do not satisfy the property, and those which are "far" from satisfying it. For the purposes of this thesis, it is useful to formalize these concepts in the context of graphs.

**Definition 6.1.** We say that a graph $G$ is $\epsilon$*-far* from satisfying a graph property $\mathcal{P}$ if no adding or removing of up to $\epsilon\binom{|G|}{2}$ edges in $G$ results in the graph satisfying the property.

**Definition 6.2.** An $\epsilon$*-test* $\mathcal{A}$ deciding a graphs property $\mathcal{P}$ with query complexity $q(n)$ is a randomized algorithm that, on input graph $G$ of size $n$, satisfies:

1. If $G \in \mathcal{P}$, then $P(\mathcal{A} \text{ accepts } G) \geq \frac{2}{3}$.
2. If $G$ is $\epsilon$-far from satisfying $\mathcal{P}$, then $P(\mathcal{A} \text{ rejects } G) \geq \frac{2}{3}$.

The query complexity $q(n)$ is the maximum number of queries the algorithm can make, and (in our case) a query discerns whether a desired pair of vertices in the input graph $G$ are adjacent or not.

Of course, the most desirable testers are those with lower query complexity. A class of particular interest is that of testers which complexity does not grow with the size of the graph.

**Definition 6.3.** We say that a property $\mathcal{P}$ is *testable* if there exists an $\epsilon$-test deciding $\mathcal{P}$ with a constant query-complexity with respect to the size of the input graph, that is, it only depends on the parameter $\epsilon$.

In [3], Alon and Shapira showed that a large class of properties, a subclass of which will be the center of our attention, are testable.

**Theorem 6.4** (Alon & Shapira Theorem in [3])**.** *Every hereditary graph property is testable (with one-sided error).*

A property is said to be *hereditary* if it is preserved under taking induced subgraphs. A property is testable *with one-sided error* if the first condition in Definition 6.2 is strengthened to $P(\mathcal{A} \text{ accepts } G) = 1$, and thus the associated algorithm does not give false negatives.

Although constant with respect to the size of the input graph, the query complexity of the resulting $\epsilon$-test from Alon & Shapira Theorem is very large. This is due to the tower function bound of Szemerédi's Regularity Lemma, which is unavoidable in the general setting [8]. Another problem caused by the use of the Regularity Lemma, although less concerning, is generated by the presence of irregular pairs. Due to this, a subsequent refinement of the resulting partition may be required (as in [14]), further increasing the complexity of the tester.

Now, by moving to the context of stable graphs, both these difficulties are easily avoided by using the Stable Regularity Lemma instead. The partition size is only exponential with respect to the error parameter $\epsilon$, and irregular pairs are completely avoided.

Mention homogeneity?

The remaining of this section will be dedicated to the construction of an $\epsilon$-test for a known case of hereditary property, $H$*-freeness* in stable graphs. A graph $G$ is said to be $H$-free, where $H$ is another graph, if no copy of $H$ appears as an induced subgraph in $G$. Thus, the given $\epsilon$-test needs to be able to distinguish between graphs that are $H$-free and graphs that are $\epsilon$-far from it, with some error. In fact, our $\epsilon$-test will only have one-sided error, as if the input graph is $H$-free the tester will report so with probability 1.

The first step towards constructing such tester is proving Theorem 6.10. This theorem uses the Stable Regularity Lemma to prove that a graph being $\epsilon$-far from being $H$-free implies it containing many (as a fixed fraction of all induced subgraphs of size $|H|$) induced copies of $H$. This point is central for the construction, and once proved we can simply let the tester ask for all the edges in a sample of vertices of fixed size. The algorithm then simply checks whether a copy of $H$ can be found in the subgraph induced by the sample, and report accordingly.

Adapt if multiple samples is still the strategy.

## 6.1 Unavoidable is Abundant

We now briefly formalize the concepts of being far from $H$-freeness, and containing many copies of $H$ using the notation from [4].

**Definition 6.5.** A graph $H$ is $\gamma$-*unavoidable* in a graph $G$ if no adding or removing of up to $\gamma\binom{|G|}{2}$ edges in $G$ results in $H$ not appearing as an induced subgraph of $G$.

**Definition 6.6.** A graph $H$ is $\eta$-*abundant* in a graph $G$ if $G$ contains at least $\eta|G|^{|H|}$ induced copies of $H$.

An important property of regularity, which is needed for the proof of the theorem, is that the regularity is partially maintained when moving to subsets. Not only that, but it also ensures that the density of the pair does not change too much.

**Lemma 6.7** (Lemma 3.1 in [4])**.** *Let $\epsilon \leq \epsilon' < \frac{1}{2}$ and $\delta \in (0, 1)$. If $(A, B)$ is an $\epsilon$-regular pair with density $\delta$, $A' \subseteq A$ with $A \geq \epsilon'|A|$, and $B' \subseteq B$ with $|B'| \geq \epsilon'|B|$, then $(A', B')$ is an $(\frac{\epsilon}{\epsilon'})$-regular pair with density at least $\delta - \epsilon$ and at most $\delta + \epsilon$.*

*Proof.* Let $A'' \subseteq A' \subseteq A$, $B'' \subseteq B' \subseteq B$ be such that

$$|A''| \geq \frac{\epsilon}{\epsilon'}|A'| \geq \frac{\epsilon}{\epsilon'}\epsilon'|A| = \epsilon|A| \text{ and}$$
$$|B''| \geq \frac{\epsilon}{\epsilon'}|B'| \geq \frac{\epsilon}{\epsilon'}\epsilon'|B| = \epsilon|B|$$

By $\epsilon$-regularity of $(A, B)$, $|d(A, B) - d(A'', B'')| < \epsilon$. Thus,

$$\begin{aligned}|d(A', B') - d(A'', B'')| &= |d(A', B') - d(A, B) + d(A, B) - d(A'', B'')| \\ &\leq |d(A', B') - d(A, B)| + |d(A, B) - d(A'', B'')| \\ &< 2\epsilon \leq \frac{\epsilon}{\epsilon'}\end{aligned}$$

This proves the $(\frac{\epsilon}{\epsilon'})$-regularity of $(A', B')$.
Also, since $(A, B)$ is $\epsilon$-regular, $|d(A, B) - d(A', B')| < \epsilon$, and thus,

$$\delta - \epsilon < d(A', B') < \delta + \epsilon$$

□

The pivotal point in the proof of Theorem 6.10 is the fact that, if the reduced graph from a regular partition contains an induced structure resembling $H$, i.e. where pairs of parts are mostly connected if the corresponding vertices in $H$ are connected, and mostly not connected otherwise, then the original graph contains many induced copies of $H$ (this is a version of the so called *Counting Lemma* from [10]). The following lemma formalizes this idea.

Define reduced subgraph as a remark of the stable regularity lemma.

**Lemma 6.8** (Lemma 3.2 in [4]). *For every $\delta \in (0,1)$ and $\ell > 0$ there exist $\epsilon = \epsilon(\delta, \ell)$ and $\eta = \eta(\delta, \ell)$ satisfying the following property:*

*Let $H$ be a graph with vertices $v_1, \dots, v_\ell$ and let $V_1, \dots, V_\ell$ be an $\ell$-tuple of disjoint sets of vertices of a graph $G$ such that for every $1 \le i < i' \le \ell$, the pair $(V_i, V_{i'})$ is $\epsilon$-regular, with density at least $\delta$ if $v_i v_{i'}$ is an edge of $H$, and at most $1-\delta$ if $v_i v_{i'}$ is not an edge of $H$. Then, at least $\eta \prod_{i=1}^{\ell} |V_i|$ of $\ell$-tuples $w_1 \in V_1, \dots, w_\ell \in V_\ell$ span induced copies of $H$ where $w_i$ plays the role of $v_i$.*

*Proof.* Without loss of generality, we assume that $H$ is the complete graph, since we can simply replace each non-edge $v_i v_{i'}$ of $H$ with an edge by exchanging all edges and non-edges between $V_i$ and $V_{i'}$.

We prove the lemma by induction on $\ell$. The case $k = 1$ is trivial, and the number of induced copies of $H$ is $|V_1|$, so $\eta(\delta, 1) = 1$ and $\epsilon(\delta, 1) = 1$ (No regularity needed if no pairs). The I.H. is that the values $\eta(\delta, \ell-1)$ and $\epsilon(\delta, \ell-1)$ exist and are known for all $\ell$. We proceed to prove that the following values $\eta$ and $\epsilon$ hold:

$$\epsilon = \epsilon(\delta, \ell) = \min\left(\frac{1}{2\ell-2}, \frac{1}{2}\delta\epsilon\left(\frac{1}{2}\delta, \ell-1\right)\right)$$
$$\eta = \eta(\delta, \ell) = \frac{1}{2}(\delta-\epsilon)^{\ell-1}\eta\left(\frac{1}{2}\delta, \ell-1\right)$$

For each $1 < i \le \ell$, the number of vertices of $V_1$ which have less then $(\delta-\epsilon)|V_i|$ neighbors in $V_i$ is less then $\epsilon|V_i|$. Otherwise, the set of such vertices, say $U \in [V_1]^{\ge \epsilon |V_1|}$ together with $V_i$ would form a subpair $(U, V_i)$ with density $< \delta - \epsilon$ which, by Lemma 6.7 contradicts the $\epsilon$-regularity of the pair $(V_1, V_i)$.

Therefore, at least $(1-(\ell-1)\epsilon)|V_1|$ of the vertices of $V_1$ have at least $(\delta-\epsilon)|V_i|$ neighbors in $V_i$ for all $1 < i \le \ell$. In particular, since $\epsilon \le \frac{1}{2\ell-2}$ we have that $(\ell-1)\epsilon \le \frac{1}{2}$ and then $1-(\ell-1)\epsilon \ge \frac{1}{2}$, so at least half of the vertices of $V_1$ satisfy the above condition.

For each such vertex $w_1 \in V_1$, let $V_i'$ denote the subset of vertices of $V_i$ which are neighbors of $w_1$. Since *epsilon* $\le \frac{1}{2}\delta$, Lemma 6.7 implies that for all $1 < i < i' \le \ell$, the pair $(V_i', V_{i'}')$ is $(\frac{\epsilon}{\delta-\epsilon})$-regular, and given that $(\frac{\epsilon}{\delta-\epsilon}) \le (\frac{2\epsilon}{\delta}) \le \epsilon(\frac{1}{2}\delta, \ell-1)$, it is $\epsilon(\frac{1}{2}\delta, \ell-1)$-regular. Also, it has density at least $\delta - \epsilon \ge \frac{1}{2}\delta$. By the induction hypothesis, we have at least

$$\eta\left(\frac{1}{2}\delta, \ell-1\right)\prod_{i=2}^{\ell}|V_i'| \ge \eta\left(\frac{1}{2}\delta, \ell-1\right)\prod_{i=2}^{\ell}(\delta-\epsilon)|V_i|$$

possible choices of $w_2 \in V_2, \dots, w_\ell \in V_\ell$ such that the induced subgraph spanned by $w_1, \dots, w_\ell$ is complete. Since there are at least $\frac{1}{2}|V_1|$ vertices $w_1$ which satisfy the above condition, the chosen values of $\eta$ satisfies the lemma, and we are done. □

*Remark* 6.9. The non-recursive form of $\epsilon$ and $\eta$ for $\ell > 1$ is:

$$\epsilon(\delta, \ell) = 2\left(\frac{\delta}{4}\right)^{\ell-1}$$
$$\eta(\delta, \ell) \ge \frac{1}{2^{\frac{(\ell+2)(\ell+1)}{2}-4}}\delta^{\frac{\ell(\ell-1)}{2}}$$

We are now ready to prove the main theorem of this section. The proof is similar to that of [4, Theorem 5.1], but with some major simplification and optimization allowed by using the Stable Regularity Lemma. The main difference is the fact that we do not need to refine the partition to get rid of irregular pairs. To resume, we first apply Theorem 5.22 to get a regular partition, then, we create a copy of the graph where pairs become either complete or empty, by adding or subtracting, overall, less than $\gamma\binom{|G|}{2}$ edges. By the $\gamma$-unavoidability of $H$, this new graph still contains a copy of $H$. This fact ensures the existence of an induced structure in the partition of the original graph which allows us to apply Lemma 6.8 and conclude that $H$ is abundant in $G$. Such conclusion is formalized in the following theorem.

**Theorem 6.10.** *For every $k_*, \gamma, \ell$ there is a $\eta(k_*, \gamma, \ell)$ such that if $H$ is a graph with $\ell$ vertices, $G$ has the non-$k_*$-order property and $H$ is $\gamma$-unavoidable in $G$, then $H$ is $\eta$-abundant in $G$.*

*Proof.* Apply Theorem 5.22 to $G$ with $\epsilon = \min(\frac{\sqrt{\gamma}}{2}, \frac{\epsilon_{6.8}(1-\frac{\sqrt{\gamma}}{2},\ell)}{\ell})$, $k_*$ and $m = 0$. We have a partition $\overline{A} = \{A_i \mid i \in \{1, \dots, m_+\}\}$ into $m_* \leq M$ disjoint parts with,

$$M \leq \left\lceil 12\max(\frac{2}{\sqrt{\gamma}}, \frac{\ell}{\epsilon_{6.8}(1-\frac{\sqrt{\gamma}}{2},\ell)})\right\rceil^{2^{k_*+1}-1}$$

such that all pairs of parts are $\epsilon$-regular, and self-pairs are $4\epsilon$-regular. Also, by Remark 5.4 and $\frac{\epsilon^2}{2}$-excellence of the parts, pairs have density at most $\epsilon^2$ or at least $1-\epsilon^2$.

Maybe make a remark in Theorem 5.19

Now, we randomly partition each part $A_i$ into $\ell$ equitable subparts $A_{i,j}$. By Lemma 6.7, each pair of such subparts is $\ell\epsilon$-regular. On the other hand, Theorem 5.22 guarantees that such pairs have density at most $\epsilon$ or at least $1-\epsilon$.

Next, we modify the graph $G$ into $G'$ by only adding and removing no more than $\gamma\binom{|G|}{2}$ edges:

- For each pair of parts $(A_{i_1}, A_{i_2})$ with $i_1 \neq i_2$, if the pair's density is at most $\epsilon^2$, we remove all edges between $A_{i_1}$ and $A_{i_2}$. Otherwise, the pair's density is at least $1-\epsilon^2$, and we add all remaining edges. This changes at most a fraction $\epsilon^2$ of the edges between (disjoint) parts.

- For each self-pair $(A_i, A_i)$, if the pair's density is at most $4\epsilon^2$ again we remove all edges in $A_i$. Otherwise, the pair's density is at least $1-4\epsilon^2$, and we add all remaining edges. This changes at most a fraction $4\epsilon^2$ of the edges in self-pairs.

The resulting graph $G'$ differs from $G$ in at most $4\epsilon^2\binom{|G|}{2} \leq \gamma\binom{|G|}{2}$ edges. Thus, the $\gamma$-unavoidability of $H$ in $G$ ensures that there is still a copy of $H$ in $G'$. Denote its vertices $v_{i_1}, \dots, v_{i_\ell}$, choosing $i_1, \dots, i_\ell$ such that $v_{i_1} \in A_{i_1,1}, \dots, v_{i_\ell} \in A_{i_\ell,\ell}$. Notice that $A_{i_1,1}, \dots, A_{i_\ell,\ell}$ satisfy the conditions of Lemma 6.8 with $\delta_{6.8} = 1 - \frac{\sqrt{\gamma}}{2}$:

- Each subpair $(A_{i_j,j}, A_{i_{j'},j'})$ with $j \neq j'$ is $\ell\epsilon$-regular, and since $\epsilon \leq \frac{\epsilon_{6.8}(1-\frac{\sqrt{\gamma}}{2},\ell)}{\ell}$, in particular is $\epsilon_{6.8}(1-\frac{\sqrt{\gamma}}{2},\ell)$-regular.

- For each $i_j \neq i_{j'}$, if $v_{i_j}v_{i_{j'}}$ is an edge of $G$ then, by construction of $G'$, the subpair $(A_{i_j,j}, A_{i_{j'},j'})$ has density at least $1-\epsilon \leq 1-\frac{\sqrt{\gamma}}{2}$, and if $v_{i_j}v_{i_{j'}}$ is not an edge of $G$, the subpair $(A_{i_j,j}, A_{i_{j'},j'})$ has density at most $\epsilon \geq 1-(1-\frac{\sqrt{\gamma}}{2})$

Hence, the lemma guarantees that there are at least $\eta_{6.8}(1-\frac{\sqrt{\gamma}}{2},\ell)\prod_{j=1}^{\ell}\{A_{i_j},j\}$ copies of $H$ in $G$. The fraction of induced copies of $H$ in $G$ is at least

$$\frac{\eta_{6.8}(1-\frac{\sqrt{\gamma}}{2},\ell)\prod_{j=1}^{\ell}\{A_{i_j},j\}}{n^{\ell}} \geq \eta_{6.8}(1-\frac{\sqrt{\gamma}}{2},\ell)(\frac{\frac{n}{M\cdot\ell}}{n})^{\ell} = \eta_{6.8}(1-\frac{\sqrt{\gamma}}{2},\ell)(M\cdot\ell)^{-\ell} =: \eta$$

and $H$ is at least $\eta$-abundant in $G$. □

Notice that this same result can be proved in the general context instead of only for stable graphs as the original Theorem 5.1 from [4] proves. The difference is that the resulting $\eta$ is much larger (although not given explicitly).

Explain where is the improvement (irregular pairs, partition size, etc.)

*Remark* 6.11. A more explicit lower bound for $\eta$ only depending on $\gamma$, $k_*$ and $\ell$ is:

$$\eta \geq \frac{1}{2^{\frac{(\ell+2)(\ell+1)}{2}-4}}\left(1-\frac{\sqrt{\gamma}}{2}\right)^{\frac{\ell(\ell-1)}{2}}\left(\frac{1}{24}\min\left\{\frac{\sqrt{\gamma}}{2},\frac{2}{\ell}\left(\frac{2-\sqrt{\gamma}}{8}\right)^{\ell-1}\right\}\right)^{\ell\left(2^{k_*+1}-1\right)}\left(\frac{1}{\ell}\right)^{\ell}$$

## 6.2 The Algorithm

Now we have all the tools needed to build our $\epsilon$-test $\mathcal{A}$ for deciding $H$-freeness for a given graph $H$ of size $\ell$. See Algorithm 1 for pseudo-code of the steps that $\mathcal{A}$ follows to make its decision.

**Algorithm 1** $\epsilon$-test $\mathcal{A}$ for deciding $H$-freeness for a given graph $H$ of size $\ell$

**Require:** a graph $G$ of size $n$ with non-$k_*$-order property

1: $t \leftarrow \frac{\ell \log(\frac{2}{3})}{\log(1-\eta_{6.10}(k_*,\epsilon,\ell))}$
2: **if** $n < \ell$ **then**
3:     return 0
4: **else if** $n < t$ **then**
5:     query all edges in $G$
6:     **if** $\exists v_{i_1}, \dots, v_{i_\ell} \in G$ such that $\{v_{i_1}, \dots, v_{i_\ell}\}$ induces a copy of $H$ in $G$ **then**
7:         return 1
8:     **else**
9:         return 0
10:     **end if**
11: **else**
12:     $S \leftarrow \emptyset$
13:     **while** $i \leq t$ **do**
14:         $s_i \sim G$
15:         **while** $s_i \in S$ **do**
16:             $s_i \sim G$
17:         **end while**
18:         $S \leftarrow S \cup \{s_i\}$
19:     **end while**
20:     query all edges induced by the vertex set $S$
21:     **if** $\exists v_1, \dots, v_\ell \in S$ such that $\{v_1, \dots, v_\ell\}$ induces a copy of $H$ in $G$ **then**
22:         return 1
23:     **else**
24:         return 0
25:     **end if**
26: **end if**

Indeed, $\mathcal{A}$ is an $\epsilon$-test. If the input graph $G$ is $H$-free, then the algorithm returns 0, either because the graph $G$ is too small to contain $H$ (line 3) or because all attempts of finding $H$ as an induced subgraph of $G$ failed (either line 9 or line 24). On the other hand, if $G$ is $\epsilon$-far from being $H$-free, Theorem 6.10 ensures that $H$ is $\eta_{6.10}(k_*, \epsilon, \ell)$-abundant in $G$. Thus, checking $t_*$ times whether a random sample of $\ell$ vertices contains an induced copy of $H$, the probability of not finding any copy of $H$ is at most $(1-\eta_{6.10}(k_*, \epsilon, \ell))^{t_*}$. By letting $t_* = \frac{\log(\frac{2}{3})}{\log(1-\eta_{6.10}(k_*,\epsilon,\ell))}$ the probability of finding at least one copy of $H$ is at least $\frac{2}{3}$. The total number of vertices included in the samples is at most (as there may be repetitions) $t_* := t \cdot \ell$, and this probability is at most as high as simply querying all the edges within a sample of vertices of size $t_*$, and checking whether $H$ appears as an induced subgraph of $G$. For completeness, we also need to ensure that $n \geq t_*$. If $n < t_*$, then the algorithm simply queries all edges of $G$, checks whether $H$ appears as an induced subgraph of $G$ and reports accordingly (either line 7 or line 9).

The resulting query complexity of the algorithm $\mathcal{A}$ can bounded by

$$q \leq \binom{t_*}{2} \leq \left(\frac{\log(\frac{2}{3})}{\log(1-\eta_{6.10}(k_*, \epsilon, \ell))}\right)^2$$

Comment on optimization such as checking if copies of $H$ are found as soon as the sample is large enough and stopping early if so.

# References


[1] Ayush Basu (https://mathoverflow.net/users/389789/ayush-basu). *irregular pairs in half graphs - Szemeredi regularity*. MathOverflow. URL:https://mathoverflow.net/q/404954 (version: 2021-09-27). eprint: https://mathoverflow.net/q/404954. URL: https://mathoverflow.net/q/404954.

[2] Noga Alon, Eldar Fischer, and Ilan Newman. "Efficient Testing of Bipartite Graphs for Forbidden Induced Subgraphs". en. In: *SIAM Journal on Computing* 37.3 (Jan. 2007), pp. 959–976. ISSN: 0097-5397, 1095-7111. DOI: 10.1137/050627915. URL: http://epubs.siam.org/doi/10.1137/050627915 (visited on 08/16/2025).

[3] Noga Alon and Asaf Shapira. "A characterization of the (natural) graph properties testable with one-sided error". In: *SIAM Journal on Computing* 37.6 (2008), pp. 1703–1727.

[4] Noga Alon et al. "Efficient testing of large graphs". In: *Combinatorica* 20.4 (2000), pp. 451–476.

[5] Reinhard Diestel. "Extremal graph theory". In: *Graph theory*. Springer, 2024, pp. 179–226.

[6] Jacob Fox, János Pach, and Andrew Suk. "Erdős–Hajnal Conjecture for Graphs with Bounded VC-Dimension". en. In: *Discrete & Computational Geometry* 61.4 (June 2019), pp. 809–829. ISSN: 0179-5376, 1432-0444. DOI: 10.1007/s00454-018-0046-5. URL: http://link.springer.com/10.1007/s00454-018-0046-5 (visited on 06/28/2025).

[7] Alan Frieze and Ravi Kannan. "Quick Approximation to Matrices and Applications". en. In: *Combinatorica* 19.2 (Feb. 1999), pp. 175–220. ISSN: 0209-9683, 1439-6912. DOI: 10.1007/s004930050052. URL: http://link.springer.com/10.1007/s004930050052 (visited on 12/11/2018).

[8] W. T. Gowers. "Lower bounds of tower type for Szemerédi's uniformity lemma". In: *Geom. Funct. Anal.* 7.2 (1997), pp. 322–337. ISSN: 1016-443X,1420-8970. DOI: 10.1007/PL00001621. URL: https://doi.org/10.1007/PL00001621.

[9] Wilfrid Hodges. *Model theory*. Cambridge university press, 1993.

[10] János Komlós et al. "The regularity lemma and its applications in graph theory". English. In: *Theoretical aspects of computer science. Advanced lectures*. Berlin: Springer, 2002, pp. 84–112. ISBN: 3-540-43328-7. DOI: 10.1007/3-540-45878-6_3.

[11] László Lovász and Balázs Szegedy. "Regularity partitions and the topology of graphons". In: *An irregular mind*. Vol. 21. Bolyai Soc. Math. Stud. János Bolyai Math. Soc., Budapest, 2010, pp. 415–446. ISBN: 978-963-9453-14-2; 978-3-642-14443-1. DOI: 10.1007/978-3-642-14444-8_12. URL: https://doi.org/10.1007/978-3-642-14444-8_12.

[12] László Lovász and Balázs Szegedy. "Szemerédi's Lemma for the Analyst". en. In: *GAFA Geometric And Functional Analysis* 17.1 (Apr. 2007), pp. 252–270. ISSN: 1016-443X, 1420-8970. DOI: 10.1007/s00039-007-0599-6. URL: http://link.springer.com/10.1007/s00039-007-0599-6 (visited on 12/11/2018).

[13] M. Malliaris and S. Shelah. "Notes on the stable regularity lemma". en. In: *The Bulletin of Symbolic Logic* 27.4 (Dec. 2021), pp. 415–425. ISSN: 1079-8986, 1943-5894. DOI: 10.1017/bsl.2021.69. URL: https://www.cambridge.org/core/product/identifier/S107989862100069X/type/journal_article (visited on 10/18/2022).

[14] Maryanthe Malliaris and Saharon Shelah. "Regularity lemmas for stable graphs". In: *Transactions of the American Mathematical Society* 366.3 (2014), pp. 1551–1585.

[15] Tung Nguyen, Alex Scott, and Paul Seymour. "Induced subgraph density. VI. Bounded VC-dimension". In: *arXiv preprint arXiv:2312.15572* (2023).

[16] A Pajor. "Sous-Spaces 1: Des Espaces de Banach Travaux en Cours". In: *Hermann, Paris* (1985).

[17] N Sauer. "On the density of families of sets". In: *Journal of Combinatorial Theory, Series A* 13.1 (1972), pp. 145–147. ISSN: 0097-3165. DOI: https://doi.org/10.1016/0097-3165(72)90019-2. URL: https://www.sciencedirect.com/science/article/pii/0097316572900192.

[18] Saharon Shelah. "A combinatorial problem; stability and order for models and theories in infinitary languages". In: *Pacific Journal of Mathematics* 41.1 (1972), pp. 247–261.

[19] Endre Szemerédi. "Regular partitions of graphs". In: *Problèmes combinatoires et théorie des graphes (Colloq. Internat. CNRS, Univ. Orsay, Orsay, 1976)*. Vol. 260. Colloq. Internat. CNRS. CNRS, Paris, 1978, pp. 399–401. ISBN: 2-222-02070-0.

[20] V. N. Vapnik and A. Ja. Červonenkis. "The uniform convergence of frequencies of the appearance of events to their probabilities". In: *Teor. Verojatnost. i Primenen.* 16 (1971), pp. 264–279. ISSN: 0040-361x.

[21] V. N. Vapnik and A. Ya. Chervonenkis. "On the uniform convergence of relative frequencies of events to their probabilities". In: *Measures of complexity*. Reprint of Theor. Probability Appl. **16** (1971), 264–280. Springer, Cham, 2015, pp. 11–30. ISBN: 978-3-319-21851-9; 978-3-319-21852-6.

# A. Other proofs

For completeness, here we leave the proof of Lemma 5.8.

*Proof.* Suppose the converse. We use this fact to build sets $\{b_\eta \mid \eta \in \{0,1\}^{<k_{**}}\}$ and $\{A_\eta \mid \eta \in \{0,1\}^{\leq k_{**}}\}$ on induction over $k < k_{**}$, where $k = |\eta|$, satisfying:

1. $A_{\langle\cdot\rangle} \subseteq A$, with $|A|_{\langle\cdot\rangle} = m_0$.
2. $B_\eta$ is an $\zeta$-good set witnessing that $A_\eta$ is not $(\frac{m_{k+1}}{m_k}, \zeta)$-excellent, for all $k < k_{**}$.
3. $A_{\eta\frown\langle i\rangle} = \{a \in A_\eta \mid t(a, B_\eta) \equiv i\}$ for all $i \in \{0,1\}$ and $k < k_{**}$.
4. $|A_\eta| = m_k$, for all $k \leq k_{**}$.
5. $A_{\eta\frown\langle 0\rangle} \sqcup A_{\eta\frown\langle 1\rangle} \subseteq A_\eta$, for all $k < k_{**}$.
6. $\overline{A_k} = \{A_\eta \mid \eta \in \{0,1\}^k\}$ is a partition of a subset of $A$, for all $k \leq k_{**}$.

Notice that, by 1. and 4., the size of $A_\eta$ is $m_k$, so by IH none of the sets $A_\eta$ is $(\frac{m_{k+1}}{m_k}, \zeta)$-excellent. Then, $B_\eta$ in 2. is well-defined. Also, by $\zeta$-goodness of $B_\eta$, $t(a, B_\eta)$ in 3. is well-defined. Then, since $B_\eta$ is witnessing the non-$(\frac{m_{k+1}}{m_k}, \zeta)$-excellence of $A_\eta$, we have that $|A_{\eta\frown\langle i\rangle}| \geq \frac{m_{k+1}}{m_k} m_k = m_{k+1}$ for all $i \in \{0,1\}$, satisfying 4.. Finally, by definition 3., we have the disjoint union 5. which by itself ensures 6..
Now, our goal is to build two sequences $\{b_\eta \mid \eta \in \{0,1\}^{<k_{**}}\}$ and $\{a_\eta \mid \eta \in \{0,1\}^{k_{**}}\}$ to contradict the tree bound $k_{**}$. First of all, notice that, for $\eta \in \{0,1\}^{k_{**}}$

$$|A_\eta| = m_k \geq m_{k_{**}} \geq 1$$

So, for each $\eta \in \{0,1\}^{k_{**}}$, $A_\eta \neq \emptyset$ and we may choose an $a_\eta \in A_\eta$. Now, for $\nu \in \{0,1\}^{<k_{**}}$ and $\eta \in \{0,1\}^{k_{**}}$ such that $\nu \triangleleft \eta$, let

$$U_{\nu,\eta} = \{b \in B_\nu \mid (a_\eta R b) \not\equiv t(a_\eta, B_\nu)\}$$

be the subset of elements of $B_\nu$ that do not relate with $a_\eta$ in the expected way. By $\zeta$-goodness of $B_\nu$, $|U_{\nu,\eta}| < \zeta|B_\nu|$, and thus for every $\nu \in \{0,1\}^{<k_{**}}$,

$$|\bigcup\{U_{\nu,\eta} \mid \nu \triangleleft \eta \in \{0,1\}^{k_{**}}\}| < 2^{k_{**}}\zeta|B_\nu| \leq |B_\nu|$$

We may choose $b_\nu \in B_\nu \setminus \bigcup\{U_{\nu,\eta} \mid \nu \triangleleft \eta \in \{0,1\}^{k_{**}}\}$, for all $\nu \in \{0,1\}^{<k_{**}}$. Finally, the sequences $\langle a_\eta \mid \eta \in \{0,1\}^{k_{**}}\rangle$ and $\langle b_\nu \mid \nu \in \{0,1\}^{<k_{**}}\rangle$ satisfy that $\forall \eta, \nu$ such that $\nu \frown \langle i\rangle \triangleleft \eta$, $(a_\eta R b_\nu)^i$, which follows 3.. This contradicts Definition 3.14 of tree bound $k_{**}$. $\square$

# B. 1000 razones para querer morirme...