MASTER'S THESIS

Master of Science in Advanced Mathematics and Mathematical Engineering

Title: The Regularity Lemma for Stable Graphs and its applications in Property Testing

Author: Severino Da Dalt

Advisor: Lluís Vena Cros

Department: Department of Mathematics

Academic year: 2024/2025

UNIVERSITAT POLITÈCNICA DE CATALUNYA
BARCELONATECH
Facultat de Matemàtiques i Estadística

Universitat Politècnica de Catalunya
Facultat de Matemàtiques i Estadística

Master in Advanced Mathematics and Mathematical Engineering

Master's thesis

The Regularity Lemma for Stable Graphs and its applications in Property Testing

Severino Da Dalt

Supervised by (Lluís Vena Cros)

September 21, 2025

Thanks to...

Abstract

Szemerédi's Regularity Lemma is a cornerstone of modern graph theory, asserting that any graph can be partitioned into a bounded number of vertex sets, where the connections between most pairs of sets behave quasi-randomly. Despite its wide-ranging applications in areas like number theory, combinatorics and computer science, the lemma suffers from two major limitations: a partition size bounded by a tower of exponentials, and the presence of irregular pairs, both unavoidable in the general case.

This work focuses on a specific subclass of graphs, the *stable graphs*, where these limitations can be overcome. By avoiding a bipartite substructure known as the half-graph, stable graphs admit a much stronger regularity lemma. This specialized lemma, originally developed by Malliaris and Shelah, guarantees a partition where all pairs are regular and the number of parts is bounded by a polynomial, a significant improvement over the general tower-type bound.

This thesis first presents a self-contained, combinatorial, and complete presentation of the proof of the stable regularity lemma, developing a unified notational framework to bridge concepts from extremal graph theory, stability, and property testing. Building on this theoretical foundation, we then construct an efficient algorithm for testing *H-freeness* (the property of not containing an induced copy of a fixed graph H) for stable graphs. This application leverages the lemma's superior properties to achieve a query complexity with significantly improved bounds compared to testers for general graphs.

Keywords

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

Contents

1. Introduction

1.1 Szemerédi's Regularity Lemma

Szemerédi's Regularity Lemma (SzRL) [43] is a powerful tool in graph theory, stating that every 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 (in the sense of satisfying its expected properties). The strength of the quasi-random properties is measured with the regularity parameter ϵ.

This theorem has seen many applications in wide variety areas of mathematics (see an early survey from the mid 1990's [22]), such as graph theory [16, 17, 18, 19, 21, 23, 32, 39][1], limits of dense graphs [6, 7, 25, 26] (see the book [24]), number theory [42, Szemerédi's Theorem], and Ramsey Theory [15, 33, 40], to name a few. The regularity method has also been generalized to hypergraphs [13, 29, 34, 35, 44]. Also, it has seen important applications to computer science, such as in property testing (see Section 1.3). This brief summary is by no means an exhaustive list of the many applications that has seen the Regularity Lemma as a key component.

Since this result is applicable to *any* graph and the number of parts of this partition with good properties is constant, it should not be surprising that some limitations arise. Firstly, not necessarily all pairs are regular, but most crucially, the required upper bound on the number of parts is, although constant, very large. More specifically, it is a tower of exponentials (it has the form $2^{2^{2^{\cdots}}}$) whose height depends on the regularity parameter ϵ.

In the general setting, both limitations have been proven to be unavoidable. In [14], Gowers shows that there exists a family of graphs for which the lower bound on the number of parts is still a tower of exponentials[2]. On the other hand, it is folklore knowledge that large-enough half-graphs present irregular pairs in any regular partition ([5] 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.

1.2 Versions of SzRL

In order to reduce the bound on the number of parts, one can relax the property required on the partition. One of the first successful implementations of such approach was given by [12, Theorem 12] which is now known as *the weak regularity lemma*. Towards the other direction one can strengthen the property on the partition [1, Lemma 4.1], at the cost of increasing the number of parts. This is known as the *strong regularity lemma* and it is particularly useful when working with induced subgraphs (more on this in Section 1.3). The previous results can be thought as three instances of a family of regularity lemmas, with varying strength [25, 26]. This is hinted in [26, Lemma 4.1 and its discussion], and for example, an even stronger instance in this family is given explicitly in [25, Section 5.1 - pg. 439], where is referred to as an *ultra-strong* regularity lemma.

[1]Aside from the results that directly apply the SzRL, there are many that either use suitable variants of SzRL or are deeply inspired by its ideas.

[2]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 ϵ^{-5}.

Now, another way of tackling the limitations of the SzRL is to reduce the scope to an appropriate subclass of graphs. A relevant example of this approach is the class of graphs with bounded VC-dimension; the notion of VC-dimension was firstly introduced by Vapnik & Chervonenkis in [45][3] and one can view it as a graph with "low complexity" (but not necessarily sparse). The reader can find more details in Section 3. For this class of graphs the bound on the number of parts can be greatly reduced. Indeed, if a given graph has VC-dimension bounded by k, we can obtain a regular partition with only $(1/\epsilon)^{f(k)}$ parts, where ϵ is the regularity parameter. Furthermore, the low complexity of graphs with bounded VC-dimension, translates into the additional property that the regular pairs are either almost fully connected or almost empty.

The first regularity lemma for graphs with bounded VC-dimension appears in the context of matrices [2], which gives a regularity lemma for bipartite graphs. In [25], the authors prove a similar result for (not necessarily bipartite) graphs with bounded VC-dimension. Fox, Pach, & Suk give an alternative proof with better bounds of the previous result, which we state below.

Theorem A (Theorem 1.3 in [11] for graphs)**.** *Each graph G with VC-dimension bounded by k admits an equitable partition of its vertex set with at most $c(k) \cdot (1/\epsilon)^{2k+1}$ parts such that all but at most an ϵ-fraction of the pairs of parts are ϵ-regular and have density either less than ϵ or greater than $1 - \epsilon$.*

Another class of graphs that has been considered to alleviate the limitations of SzRL is that of the *stable* graphs. The concept of stability originates in Model Theory (see [38]). A graph is k-stable if it avoids any *bi-induced* (see Definition 2.4) copy of the *half-graph* on $2k$ vertices, which is a bipartite graph that behaves in a very *non-quasi-random* way (see Figure 1). In fact, Malliaris & Shelah showed in [28, Theorem 5.19] that restricting to this class of graphs not only achieves a partition with a bound on the number of parts which is only exponential on $1/\epsilon$, but also completely avoids irregular pairs; the exponent depends on the size of the avoided half-graph. Again, (all) pairs are either almost fully connected or almost empty. This result is the pivotal point of this work, and Section 5 is devoted to its proof, culminating in Theorem 5.23, which we informally give below.

Theorem B. *Each k-stable graph G admits an equitable partition of its vertex set with at most $c(k) \cdot (1/\epsilon)^{2^{k+1}}$ parts such that ALL pairs of parts are ϵ-regular and have density either less than ϵ or greater than $1 - \epsilon$.*

Note that the stable graphs is a subclass of graphs with bounded VC-dimension. Indeed, if a graph does not contain a bi-induced copy of a bipartite graph with stable sets of size $\geq k$, then it has VC-dimension strictly bounded by k [25]. Hence, k-stable graphs have VC-dimension strictly bounded by k. Additionally, any half-graph has VC-dimension 1[4], so the k-stable graphs is a proper subclass of the graphs with VC-dimension strictly bounded by k.

Notice that, in Theorem B there are no irregular pairs. This fact, shows that the presence of the half-graph plays a key role in requiring irregular pairs in the partition. On the other hand, the exponent in the bound on the number of parts is exponential on k while, in Theorem A it is only linear. It is an open question whether the exponential exponent in the bound of Theorem B is needed [47].

[3]See [46] for a translated version.

[4]Indeed, the fact that the neighbourhoods of the vertices on the same stable set of a half-graph can be ordered by inclusion, and it is a bipartite graph, results in a VC-dimension of 1. Alternatively, in [25] 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.

1.3 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, called *tests*, 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. The *complexity* of a test is measured by the number of queries it needs to perform in order to decide whether a given input, each query returning whether two given vertices of the input graph 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 whose complexity does not grow with the size of the input graph. Properties for which such testers exist are called *testable*. An important result in this context is given by Alon and Shapira in [4], where they showed that a large class of properties, a subclass of which will be the center of our attention, are testable.

Theorem C (Alon & Shapira Theorem in [4])**.** *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.

A stepping stone towards Alon & Shapira Theorem was the work of Alon-Fischer-Krivelevich-Szegedy [1] where they show, among other things, that *H-freeness* is testable. 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, and H-freeness is clearly an hereditary property.

Both Alon-Fischer-Krivelevich-Szegedy and Alon & Shapira results use the strong regularity lemma [1, Lemma 4.1], which we mentioned earlier. Furthermore, the query complexity of the testers associated to the previous results, is intrinsically linked to the number of parts in the underlying regular partition. Not only that, but even though the standard SzRL is good to understand most of the structure of the graph, it has no control over the irregular pairs, which becomes a problem when looking for induced subgraphs. For this reason, the stronger version is required. This worsens the already enormous power-tower bounds of the standard regularity lemma, leading to prohibitively large, although constant, query counts.

This raises a natural question on whether the superior bounds and the lack of irregular pairs of the stable regularity lemma can be exploited to create more efficient property testers for graphs in a half-graph-restricted setting. The final of this work is dedicated to this question.

1.4 Main contributions

The main contributions of this thesis are:

- We place a larger emphasis on the combinatorial part of the result in [28], 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. In addition, we simplify some of the arguments, while making others more explicit and detailed. For example, 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). A more details list of changes is provided in Appendix B.

- 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.

1.5 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 the SzRL

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 inspired by model theory and logic results (see the use of k-trees in Section 3.3), it is useful to note that vertices adjacency (whether two vertices are the endpoints of an edge) is a symmetric and irreflexive binary relation on the vertex set. With this perspective, to denote vertex adjacency between two 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. Also, in order to simplify future notation, we will assume that a logical true statement and the value 1 are equivalent, and similarly a false statement and the value 0. As an example, if two vertices v_1 and v_2 are not adjacent, we say that $\neg v_1 R v_2 \equiv \neg 1 \equiv 0$.

Now, a class of graphs of particular relevance in this work is that of bipartite graphs, which we define as follows.

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 $E_S = \{\{v_1, v_2\} \in E \mid v_1, v_2 \in S\}$.

A similar restriction can defined for bipartite graphs, but only controlling edges between the two disjoint sets.

Definition 2.4. We say that a bipartite graph H with disjoint sets L and R is *bi-induced* in a graph G if there exist two injective homomorphisms $\phi_L : L \longrightarrow G$ and $\phi_R : R \longrightarrow G$ such that, for all $u \in L$ and $v \in G$, $uv \in H \Leftrightarrow uv \in G$.

Notice that this definition does not require the two sets $\phi_L(L)$ and $\phi_R(R)$ to be disjoint (as defined in [25, pg. 417] and [3, pg. 2]). This is important for the arguments used in this thesis, and needs to be noted that other works define such condition without this relaxation [30, pg. 3].

2.2 Regular pairs and partitions

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 2.5. Let G be a graph and let $X, Y \subseteq G$ be two (not necessarily disjoint) non-empty 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, we will only be interested in knowing the exact proportion of edges in a pair in two specific cases: either when X and Y are disjoint, or when they are equal. The first case has no problems, while for the second case we note the following.

Remark 2.6. If X is a subset of vertices of a graph G such that $|X| \geq 2$, then the proportion of possible edges between vertices in X that are actually present in G is at most twice the density $d(X, X)$. That is,

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

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

$$d(X, X) \leq \frac{|E_X|}{\binom{|X|}{2}} \leq 2d(X, X).$$

This also implies that the proportion of possible edges between vertices in X that are actually not present in G lies between $1 - 2d(X, X)$ and $1 - d(X, X)$.

Definition 2.7. Given $\epsilon > 0$ and a graph G, a pair of (not necessarily disjoint) subsets of vertices $A, B \subseteq G$ is said to be *ϵ-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 of a partition of a graph's vertex set.

Definition 2.8. Given a graph G, we say that $\{A_1, \dots, A_k\}$ is a partition of the vertex set of G with *remainder* set B, if $G = A_1 \cup \dots \cup A_k \cup B$, and $A_1, \dots, A_k$ are non-empty sets. Implicitly, we allow the remainder to be empty.

The partition we want to study needs to satisfy that most pairs of parts are regular, but we allow a small number of such pairs to be irregular.

Definition 2.9. Let G be a graph and let $\epsilon > 0$. An *ϵ-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 ϵk^2 of the pairs (A_i, A_j) with $1 \leq i < j \leq k$ are ϵ-regular.

Also, we want the partition's sets to be roughly of the same size, which can be formalized in two different ways.

Definition 2.10. A partition $\{A_1, \dots, A_k\}$ of the vertex set of a graph G is said to be *equitable* if for all $1 \leq i \leq j \leq k$, we have that $||A_i| - |A_j|| \leq 1$. On the other hand, a partition $\{A_1, \dots, A_k\}$ with remainder B of the vertex set of a graph G is said to be *even* if $|A_1| = |A_2| = \cdots = |A_k|$.

Remark 2.11. The two previous definitions, although very close in concept, have a key difference that needs to be noted. As most of the results requires the partition property (such as regularity) to be satisfied only by parts in the partition, and not necessarily by the reminder, in even partitions the behaviour of a (not necessarily trivial) fraction of vertices is unknown. Thus, results with equitable partitions are generally preferable over those with even partitions, but require some extra arguments. For example, in the context of regular partitions, one can make an even partition into an equitable one by distributing the remainder (which by definition is small) evenly between all the parts (with some extra arguments). The resulting partition is equitable with a 1-vertex difference between parts[5], and with a small increase in the regularity error. In other cases, such as the results of Section 4, the remainder is much larger, and such a strategy does not work. These (secondary) results will be presented with even partitions. The more relevant Stable Regularity Lemma in Section 5 presents an equitable one.

2.3 Szemerédi's Regularity Lemma

The following is the celebrated Szemerédi's Regularity Lemma. The statement and proof we provide in this thesis follows the one given in [10], with minor notation modifications.

Theorem 2.12 (Szemerédi's Regularity Lemma, [43])**.** *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 even ϵ-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 whose number of parts is independent of the size of the graph, and only depends on the regularity parameter ϵ and the minimum number of parts (and thus vertices) m.[6]

The proof of the regularity lemma uses a density-increment argument. There is a quantity that we shall call *energy* of the partition (Definition 2.13) that is upper bounded by a constant (2) and which is non-decreasing by partition refinement (Lemma 2.14). Also, we prove that if an even partition is not ϵ-regular, then one could refine the partition in such a way that the energy increases by a constant depending only on ϵ, and the number of parts in the new partition only depends on the size of the previous partition (Lemma 2.16). Thus, one can iteratively refine until reaching a regular partition, a process that must culminate in finitely many steps (Theorem 2.12).

[5]This ± 1 size difference is a simple consequence of the number of vertices possibly not being divisible by the number of parts. It has no major consequences, since it becomes proportionally more trivial as the size of the parts gets larger.

[6]The dependency of M on m has more to do with practical and applicability purposes (in this form of the result we do not control de edges within each part) than conceptual ones. Since we want to be able to choose a minimal number of parts m, the upper bound on the number of pairs will also depend on such value.

The following inequality will be useful during the proof. 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 $b_i = e_i / \sqrt{\mu_i}$.

We now formalize the concept of the *energy* of a partition.

Definition 2.13. 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, $q(\widetilde{A}) = q(\overline{A} \cup \overline{B})$

As promised, we see that the energy of a partition is upper bounded by a constant:

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

We now prove that refining a pair of parts or a whole partition does not decrease its energy.

Lemma 2.14. *Let G be a graph.*

I. 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 $q(\overline{A}_1, \overline{A}_2) \geq q(A_1, A_2)$.

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

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

II. 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,

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

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 ϵ, but it will when applied to all irregular pairs at the same time.

Lemma 2.15. *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 ϵ-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 ϵ-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

$$|\eta| > \epsilon, \tag{3}$$

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 $A_{1,2} := A_1 \setminus A_{1,1}$ and $A_{2,2} := A_2 \setminus A_{2,1}$, satisfy the statement.

For ease of notation, we write $c_i := |A_{1,i}|$, $d_i := |A_{2,i}|$, $e_{ij} := e(A_{1,i}, A_{2,j})$, $c := |A_1|$, $d := |A_2|$ 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 η, in the new notation we have that $e_{11} = \frac{c_1 d_1 e}{cd} + \eta c_1 d_1$, and so

$$\begin{aligned} n^2 q(\overline{A}_1, \overline{A}_2) &\geq \frac{1}{c_1 d_1} \left(\frac{c_1 d_1 e}{cd} + \eta c_1 d_1 \right)^2 + \frac{1}{cd - c_1 d_1} \left(e - \frac{c_1 d_1 e}{cd} - \eta c_1 d_1 \right)^2 \\ &\geq \frac{1}{c_1 d_1} \left(\frac{c_1 d_1 e}{cd} + \eta c_1 d_1 \right)^2 + \frac{1}{cd - c_1 d_1} \left(\frac{cd - c_1 d_1}{cd} e - \eta c_1 d_1 \right)^2 \\ &= \frac{c_1 d_1 e^2}{c^2 d^2} + \frac{2e\eta c_1 d_1}{cd} + \eta^2 c_1 d_1 + \frac{(cd - c_1 d_1) e^2}{c^2 d^2} - \frac{2e\eta c_1 d_1}{cd} + \frac{\eta^2 c_1^2 d_1^2}{cd - c_1 d_1} \\ &\geq \frac{e^2}{cd} + \eta^2 c_1 d_1 \\ &\overset{(3)}{\geq} \frac{e^2}{cd} + \epsilon^4 cd = n^2 q(A_1, A_2) + \epsilon^4 cd \end{aligned}$$

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

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

Lemma 2.16. *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 an even partition of its vertex set with remainder set B such that $|B| \leq \epsilon n$ and $|A_1| = \cdots = |A_k| =: c$. If the partition $\overline{A}$ is not ϵ-regular, then there is an even refinement $\overline{A}' = \{A_1', \dots, A_\ell'\}$ of $\overline{A}$ with remainder set B' such that $k \leq \ell \leq k4^{k+1}$, $|A_0'| \leq |A_0| + \frac{n}{2^k}$, and*

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

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 ϵ-regular, then $\overline{A}_{ij} := \{A_i\}$ and $\overline{A}_{ji} := \{A_j\}$. Otherwise, we can apply Lemma 2.15 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

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

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, since each partition $\overline{A}_{ij}$

may at most double the number of parts that end up in $\overline{A}_i$, we have that $|\overline{A}_i| \leq 2^{k-1}$. Putting all of this together, we have a new (not necessarily even) 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

$$k \leq |\overline{A}''| \leq k2^{k-1} \leq k2^k. \tag{5}$$

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

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

First equality follows from the definition of energy, first inequality uses I from Lemma 2.14, 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 even partition. In order to achieve this, we split 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 ℓ satisfies $k \leq \ell = kc < k4^k$.

Otherwise, if $c \geq 4^k$, consider $A'_1, \ldots, A'_\ell$ 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 $\overline{A}' = \{A'_1, \ldots, A'_\ell\}$ is a refinement of $\overline{A}''$ and, following II from Lemma 2.14, 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 $k \leq \ell \leq k4^{k+1}$

is satisfied. Also, no more than d vertices are left out from each part of $\overline{A}''$, and so

$$\begin{aligned}|B'| &\leq |B| + d|\overline{A}''| \\ &\overset{(5)}{\leq} |B| + \frac{c}{4^k}k2^k \\ &= |B| + \frac{kc}{2^k} \\ &\leq |B| + \frac{n}{2^k}.\end{aligned}$$

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 even 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 2.12. 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[7]. Also, set $s := \frac{2}{\epsilon^5}$.
While refining repeatedly the partition using Lemma 2.16, (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 ϵn. 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, at most $\frac{n}{2^k}$ 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 are 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,

$$k + \frac{sn}{2^k} \leq \frac{\epsilon n}{2} + \frac{\epsilon n}{2} = \epsilon n. \tag{6}$$

Now, let's bound the number of parts of the partition at the 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 if n is sufficiently large (in particular when $n \geq M$) then (6) holds.
Now, given a graph G with $n \geq m$ vertices, we can build a partition into k', with $m \leq k' \leq M$ parts, and 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 ϵ-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 $k := m$ maximal parts of equal size, and put the remaining vertices in the remainder set. This remainder set has size at most $k - 1 < \epsilon n$ by (6). We now can apply Lemma 2.16 repeatedly, as the choice of k and $n \geq M$ in (6) ensures that the reminder is at most ϵn 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. □

[7] By monotonicity of the regularity property we mean that, if a partition is ϵ-regular, than it is also ϵ'-regular for any $\epsilon' \geq \epsilon$. This follows the Definition 2.9, as both the allowed error in regular pairs and the number of irregular ones permitted increase with the regularity parameter.

For the matters of this thesis, it is important to note that it is actually known that:

- The remainder set can be avoided in the resulting partition of Szemerédi's Regularity Lemma, moving from an even partition to an equitable one (Definition 2.10). This is done by evenly distributing the leftover vertices evenly throughout the large clusters of the part, and overserving that energy lost in this operation is smaller than the gains from the former.
- It can be ensured that not only (most) pairs of different parts are regular, but also (most) parts with themselves (self-pairs) satisfy this property.

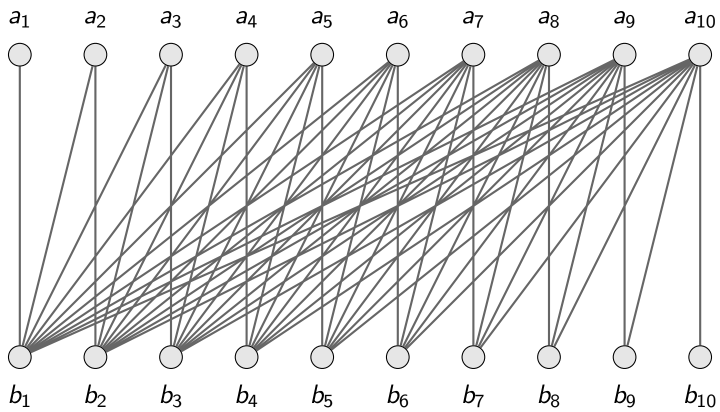
In this work we have focused our attention to the case of the Stable Regularity Lemma, but we have opted to include a proof of a (less technically involved but conceptually complete) version of the SzRL for completeness.

The interested reader is redirected to [9, 48][8] for more detailed proofs on how to obtain such partitions.

[8] In [48], authors show how to obtain a regular parition that includes regularity within pairs themselves, but omit the details on how to get an equitable partition. [9] proves the existence of a partition has regular self-pairs and no reminder, but the proof of the critical lemma to refine a partition into an equitable one (at the loss of a small amount of energy) is hinted at but omitted.

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 (see Definition 2.4) large half-graphs, a particularly non-quasi-random structure in graphs. See Figure 1 for an example of such a graph.



	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: *On the left*, a half-graph with 2×10 vertices. *On the right*, the corresponding bi-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 [25]), 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 4.10) and excellent parts in Section 5 (Lemma 5.6).

3.1 The 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*[9] 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*. Additionally, we say that G has the *disjoint k-order property* if $\langle a_i \rangle_i \cap \langle b_i \rangle_i = \emptyset$.

Remark 3.2. Notice that the vertices within each sequence $\langle a_i \rangle_i$, $\langle b_i \rangle_i$ must be distinct, as their neighborhoods within the other sequence differ, which makes this definition equivalent to "the graph not containing a bi-induced copy of a k-half-graph", as defined in Definition 2.4. 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

[9]Note that the vertex tuples $\langle a_i \rangle_i$ and $\langle b_i \rangle_i$ are "ordered" according to their neighbourhood, and thus the name *k-order* comes very naturally.

property requires avoiding not only the k-half-graph, but a whole family of induced subgraphs (the ones resulting by adding edges in the independent sets $\langle a_i \rangle_i$, $\langle b_i \rangle_i$, and possibly identifying some pairs of vertices (a_i, b_j)).

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*.

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 $B_{A,b} = B^-_{A,b}$. Naturally, sets of vertices A with a large number of large $\overline{B}_{A,b}$ are a great obstacle towards creating (close to) full or empty pairs.

3.2 VC-dimension and the Sauer-Shelah Lemma

Recall from the introduction, that graphs with the non-k-order property have bounded VC-dimension. We proceed to formally define the concept of VC-dimension in Definition 3.6, and some of its properties (Lemma 3.7) to bound the number of $\overline{B}_{A,b}$ under the non-k-order property (Lemma 3.10).

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, [36], [37])**.** *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, [31])**.** *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

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

which proves the statement. $\square$

Proof of Lemma 3.7. Suppose that $\bigcup_{s \in S} s$ 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. $\square$

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 (recall that then the two sequences of vertices in the Definition 3.1 are disjoint). Even though results in this thesis use the weaker non-disjoint version of Lemma 3.10, we prove it in this form to highlight that the non-disjointness of the sequences (and thus the broadening of the excluded structures in stable graphs) is not crucial to obtain this value of the bound, but later on[10].

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.*

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

[10] Specifically, the non-disjointness of the sequences beceomes relevant in Lemma 4.10 of Section 4 and in Lemma 5.6 of Section 5, as it allows to prove the existence of quasi-randompairs.

$b_i \in G$ 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 [28]). *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

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

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.

The following equivalent versions of Lemma 3.10 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 $b \in G$, has the same bound. The proof is given in Appendix A.

Corollary 3.12. *Let G be a graph with the non-k-order property. Then:*

I. For any finite $A \subseteq G$

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

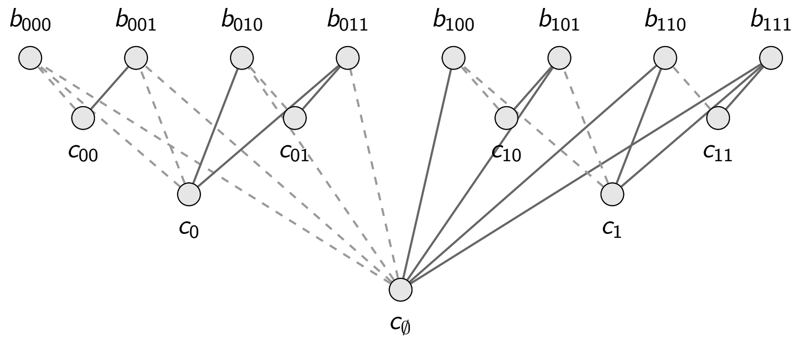
II. For any finite $A \subseteq G$

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

3.3 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. A useful structure strongly related to the k-order property is the k-*tree*.

Defining such concept requires us to introduce some tuple notation. First of all, we use $\langle a_1, \dots, a_n\rangle$ to denote an n-tuple which is an ordered list of objects (in this work, such objects will be integers). When using such tuples as a subscript of a variable and the tuples are sequences of 0's and 1's, we may skip



	c_{00}	c_0	c_{01}	$c_\emptyset$	c_{10}	c_1	c_{11}
b_{000}	0	0	-	0	-	-	-
b_{001}	1	0	-	0	-	-	-
b_{010}	-	1	0	0	-	-	-
b_{011}	-	1	1	0	-	-	-
b_{100}	-	-	-	1	0	0	-
b_{101}	-	-	-	1	1	0	-
b_{110}	-	-	-	1	-	1	0
b_{111}	-	-	-	1	-	1	1

Figure 2: *On the left*, example of a 3-tree. Solid lines show adjacent vertices, and dashed lines show non-adjacent vertices. Pairs of vertices without a line may or may not be connected. In particular, notice that connections between disjoint sub-trees are not defined, and may be edges or non-edges in any combination (e.g. the pair (c_1, c_{01})). *On the right*, the corresponding bi-adjacency matrix.

the $\langle\rangle$ and commas for ease of read (for example, $\langle 0, 0, 1\rangle$ would be written as 001). The empty tuple is denoted as $\langle\cdot\rangle$ and occasionally in subscripts as $\emptyset$. A useful operation is the concatenation of tuples, which we denote with the symbol $\frown$. Finally, we say that $\eta_1 \triangleleft \eta_2$ if for some tuple η_3 we have that $\eta_1 \frown \eta_3 = \eta_2$. We now have all the notation to formally define the concept of k-tree.

Definition 3.13. A *k-tree* in G is an ordered pair $H = (\overline{c}, \overline{b})$ comprising two (not necessarily disjoint) sequences:

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

satisfying that for all $\eta \in \{0,1\}^{<k}$ and $\rho \in \{0,1\}^k$, if for some $\ell \in \{0,1\}$ we have $\eta \frown \langle\ell\rangle \triangleleft \rho$, then $b_\rho R c_\eta \equiv \ell$. Nothing else is said about the adjacency of the rest of pairs of vertices. 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 in [28])**.** 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 [20])**.** *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 $k_{**} = 2^{k_*+1} - 1$, 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 $C_\emptyset = \langle a_i \mid i \in \{0, \dots, 2^{k_{**}} - 2\} \rangle$.
- At each step $k \in \{0, k_{**} - 1\}$, for each $\eta \in \{0,1\}^k$, take the middle element of the sequence C_η and set it to be the node c_η. Then, the remaining first half of C_η 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_η has an odd number of elements. The resulting two sequences of nodes and branches form a k_{**}-tree. See Figure 3 for a visual example of this construction. This finishes the argument for the first part of the argument.

During the proof of the second implication, we say that a set of nodes N of a k-tree $H = (\overline{c}, \overline{b})$ *contains* a k'-tree H', 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 $\overline{b}$. Simply, for each $\eta \in \{0,1\}^{k'-1}$, pick exactly two branches b_{ρ_0} and b_{ρ_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 singleton $\{c_{f(\emptyset)}\}$, 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. ☐

Suppose that G has a tree bound of at least $2^{k_*+1} - 1$, 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

$$b_0, c_0, \dots, b_{q-1}, c_{q-1}, H, b_q, c_q, \dots, b_{k_*-r-1}, c_{k_*-r-1} \tag{7}$$

such that:

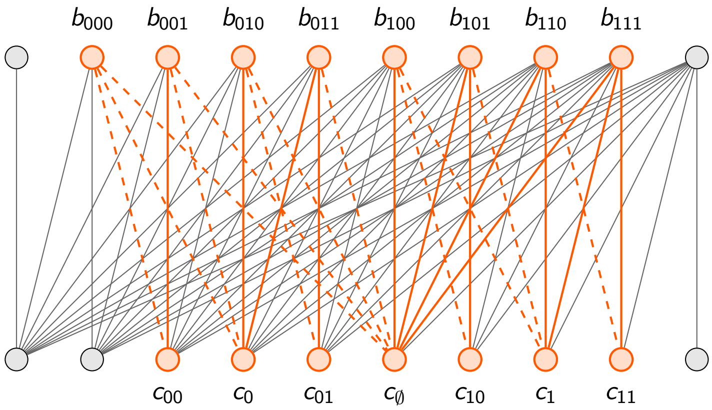
I. for all $i \in \{0, \dots, k_* - r - 1\}$, b_i and c_i are vertices in G, and H is a $(2^{r+1} - 2)$-tree in G.

II. for all $i, j \in \{0, \dots, k_* - r - 1\}$, $b_i R c_j \Leftrightarrow i \geq j$.

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

IV. 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, hence there is a node c_* and



			c_{00}	c_0	c_{01}	$c_\emptyset$	c_{10}	c_1	c_{11}	
	1	0	0	0	0	0	0	0	0	0
b_{000}	1	1	0	0	0	0	0	0	0	0
b_{001}	1	1	1	0	0	0	0	0	0	0
b_{010}	1	1	1	1	0	0	0	0	0	0
b_{011}	1	1	1	1	1	0	0	0	0	0
b_{100}	1	1	1	1	1	1	0	0	0	0
b_{101}	1	1	1	1	1	1	1	0	0	0
b_{110}	1	1	1	1	1	1	1	1	0	0
b_{111}	1	1	1	1	1	1	1	1	1	0
	1	1	1	1	1	1	1	1	1	1

Figure 3: *On the left*, example of a 3-tree in a half-graph with 2×10 vertices. Orange lines and nodes highlight the 3-tree structure, with dashed orange lines remarking the relevant non-edges. *On the right* is the corresponding bi-adjacency matrix. Again, orange cells highlight edges relative to the 3-tree structure.

branch b_* in H that are adjacent. These vertices satisfy conditions III and IV, so the sequence resulting from replacing H in (7) by b_* and c_*,

$$b_0, c_0, \dots, b_{q-1}, c_{q-1}, b_*, c_*, b_q, c_q, \dots, b_{k_*-2}, c_{k_*-2},$$

implies that G has the k_*-order property.

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 I we have that H is a $(2h+2)$-tree. For each branch b of H we denote with $Z(b)$ 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 (7) 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 (7) 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 bi-induced half-graph copy 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 of disposition of the edges.

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 is centered around the concept of ϵ-*indivisible* sets, a strong condition on bounding the non-homogeneous behaviour 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 4.12, which proves the existence of a partition where *all* pairs are highly close to full or empty, at the cost of a uneven partition. Next, we improve the results obtaining an even partition in Theorem 4.19, but this time with a small number of exceptional pairs, and a tradeoff between a non-negligible remainder set and parts of even smaller size. The final result, Theorem 4.26, achieves removing exceptional pairs and reduces 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 Section 5.

4.1 Indivisibility and the average condition

Let's first define *indivisibility*. The general definition is dependent on a certain function f, but we shall be mostly interested in the case of $f(n) = n^{\epsilon}$, which we call ϵ-indivisible, and in the constant case $f(n) = c$.

Definition 4.1 (Definition 4.2 (2) in [28])**.** 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 (1) in [28])**.** Let $\epsilon \in (0,1)$. We say that $A \subseteq G$ is ϵ-*indivisible* if for every $b \in G$,

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

As mentioned before, an ϵ-indivisible set is f-indivisible for $f(n) = n^{\epsilon}$.

Thus, indivisibility upper bounds the number of exceptional edges with respect to each vertex. A natural follow-up question is how two indivisible sets interact between each other. The following lemma tackles precisely that, although doing so asymmetrically.

Lemma 4.3 (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 $\lceil f(|A|)\rceil\, g(|B|) < \frac{1}{2}|B|$. Then, the truth value $t = t(A,B)$ satisfies the following. For all but $< f(|A|)$ of the $a \in A$ we have that for all but $< g(|B|)$ of the $b \in B$ satisfy $aRb \equiv t$.*

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, $|U_i| \geq f(|A|)$ and we can take $W_i \subseteq U_i$ with $|W_i| = \lceil f(|A|)\rceil$ 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, $|V| < |W_0| g(|B|) + |W_1| g(|B|) = 2\lceil f(|A|)\rceil\, g(|B|) < |B|$, where the first inequality follows the g-indivisibility of B. Since $|V| < |B|$, 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 $|W_0| = |W_1| = \lceil f(|A|)\rceil$, which contradicts the f-indivisibility of A. □

Definition 4.4. We say that the pair (A,B) with A f-indivisible and B g-indivisible satisfies the *average condition* if $\lceil f(|A|)\rceil\, g(|B|) < \frac{1}{2}|B|$ and thus the statement of Lemma 4.3 is true for the pair (A,B).

Remark 4.5. The condition $\lceil f(|A|)\rceil\, g(|B|) < \frac{1}{2}|B|$ makes ordering of the pair (A, B) matter, that is,

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

Next, we are interested in studying how the average condition of an indivisible pair bounds the number of exceptional edges of large subpairs. We study the f-indivisible and ϵ-indivisible case separately, as the specific case of ϵ-indivisibility gives a slightly better condition on the range of the size of the subpair. In any case, the two proofs are very similar, so we prove the ϵ-indivisible case here and the f-indivisible case in Appendix A.

Lemma 4.6 (Claim 4.8 in [28])**.** *Let A be ϵ-indivisible, B ζ-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 that, by the average condition of the pair (A, B):

- there are at most $|A|^\epsilon$ vertices of A (hence in $A' \subseteq A$), say S, which are exceptional with respect to B, so the number of edges $(a, b) \in S \times B'$ which are exceptional is at most $|S| \cdot |B'|$, and
- for each $a \in A$ (hence in $A' \subseteq A$) not in S, 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. Thus, we have at most $(a, b) \in (A' \setminus S) \times B'$ is at most $(|A'| - |S|)|B|^\zeta$.

The overall worse case in this scenario is when S is maximum ($|S| = |A|^\epsilon$), and thus we have at most $|A|^\epsilon|B'| + (|A'| - |A|^\epsilon)|B|^\zeta$ exceptional edges in $A' \times B'$, as $|B'| \geq |B|^\zeta$. Putting it all together:

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

This finishes the proof. ◻

Lemma 4.7 (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 in Appendix A. □

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

Corollary 4.8 (Corollary 4.9 in [28]). *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 4.7 with $f(n) = c$. □

Remark 4.9. Notice that the average condition is easily satisfied if the pair satisfies a condition on the size of its sets. Namely, if $f(n) = n^{\epsilon}$, A and B are f-indivisible, and $|B| \geq |A| \geq m$, then $m^{1-2\epsilon} > 4$ is sufficient for the average condition to hold for the pair (A, B):

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

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, $4 < (m_{k_{**}-1})^{1-2\epsilon}$ 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 partition of a subset, which by construction must contain a k_{**}-tree, and gives a contradiction with the non-k_*-property.

Lemma 4.10 (Claim 4.3 in [28]). *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_ℓ which is f-indivisible.*

Proof. Suppose that such subset B does not exist. Then we shall construct by induction on $k = |\eta|$ the sequences $\langle b_\eta \mid \eta \in \{0, 1\}^{<k} \rangle$ and $\langle A_\eta \mid \eta \in \{0, 1\}^{\leq k} \rangle$, satisfying the following properties:

I. $|A_\eta| = m_k$, $\forall k \in \{0, \dots, k_{**}\}$.

II. $b_\eta \in G$ witnessing that A_η is not f-indivisible, $\forall k \in \{0, \dots, k_{**} - 1\}$.

III. $A_{\eta^\frown \langle i \rangle} \subseteq A_\eta^{(i)} = \{a \in A_\eta \mid aRb_\eta \equiv i\}$, $\forall \in \{0, 1\}$, $\forall k \in \{0, \dots, k_{**} - 1\}$.

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_η, with $|A_\eta| = m_k$, is not f-indivisible. Thus, there exists a b_η such that $A_\eta^{(i)} \geq f(m_k) \geq m_{k+1}$ (II), and we can choose $A_{\eta^\frown \langle i \rangle} \subseteq A_\eta^{(i)}$ (III), such that $|A_{\eta^\frown \langle i \rangle}| = m_{k+1}$ $\forall i \in \{0, 1\}$ (I). Now, for all η such that $|\eta| = k_{**}$, consider some element $a_\eta \in A_\eta$, which exists since $m_\ell > 0$ for all ℓ. Then, we have two sequences $\langle b_\eta \mid \eta \in \{0, 1\}^{<k_{**}} \rangle$ and $\langle a_\eta \mid \eta \in \{0, 1\}^{k_{**}} \rangle$ are a k_{**}-tree in G: 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 $a_\eta R b_\rho \equiv \ell$ since $a_\eta \in A_\eta \subseteq A_{\rho^\frown \langle \ell \rangle}$ by III. This contradicts the tree bound k_{**} (see Theorem 3.15). □

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 of the elements of the pair, 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 4.11 (Claim 4.4 and 4.5 in [28])**.** *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 $B = A \setminus \bigcup_{A_i \in \overline{A}} A_i$ such that:*

I. For each $j \in \{1, \dots, j(*)\}$, A_j is f-indivisible.

II. For each $j \in \{1, \dots, j(*)\}$, $|A_j| \in \{m_0, \dots, m_{k_{**}-1}\}$.

III. $A_j \subseteq A \setminus \bigcup\{A_i \mid i < j\}$, in particular $A_i \cap A_j = \emptyset \ \forall i \neq j$.

IV. $|B| < m_0$.

V. $|A_i| \leq |A_j| \Leftrightarrow i \leq j$.

Proof. Iteratively, apply Lemma 4.10 to the remainder $A \setminus \bigcup\{A_i \mid i < j\}$ (III) to get an f-indivisible A_j (I) of size m_ℓ, $\ell \in \{0, \dots, k_{**} - 1\}$ (II) until less than m_0 vertices are available (IV). To conclude, reorder the indices of the A_j's in ascending size order (V). □

Finally, we ensure the pairs satisfy the average condition by simply requiring a minimal size of the parts using Remark 4.9, which can be easily integrated in the definition of the sequence of integers.

Lemma 4.12 (Claim 4.10 in [28])**.** *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 $4 < (m_{k_{**}-1})^{1-2\epsilon}$. 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 $B = A \setminus \bigcup_{A_i \in \overline{A}} A_i$ satisfying:*

I. For each $i \in \{1, \dots, i(*)\}$, A_i is ϵ-indivisible.

II. For each $i \in \{1, \dots, i(*)\}$, $|A_i| \in \{m_0, \dots, m_{k_{**}-1}\}$.

III. $A_i \cap A_j = \emptyset$ for all $i \neq j$.

IV. $|B| < m_0$.

V. $\overline{A}$ is $\leq$-increasing.

VI. 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 4.11, setting $f(x) = x^{\epsilon}$. Now, by II, for any $A_i, A_j \in \overline{A}$ with $i < j$ there is some $\ell \in \{0, \dots, k_{**} - 1\}$ such that $|A_i| \le |A_j| = m_\ell$. Also, it follows the condition $4 < (m_{k_{**}-1})^{1-2\epsilon}$ and Remark 4.9 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 4.6 is satisfied. This gives us VI, where in last inequality we used that $A_i \subseteq A, A_j \subseteq B$. □

Remark 4.13. For sufficiently small ϵ, the condition $4 < (m_{k_{**}-1})^{1-2\epsilon}$ is rather mild. For example, if $\epsilon < \frac{1}{4}$, then we are just requiring that $m_{k_{**}-1} \ge 16$.

4.2 ϵ-indivisible even partition

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

Definition 4.14. Let A, B be f-indivisible sets with $\lceil f(A) \rceil f(B) < \frac{1}{2}|B|$. 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 $\langle B_j \mid j \in \{1, \dots, j_B\} \rangle$ be a partition of B with $|B_j| = m$ for all $j \in \{1, \dots, j_B\}$. We define $\varepsilon^{+}_{A_i, B_j, m}$ as the event:

$$\forall a \in A_i \ \forall b \in B_j, aRb = t(A, B).$$

Notice that, in the previous definition, we are implicitly requiring that $|A| = i_A \cdot m$ and $|B| = j_B \cdot m$.

Lemma 4.15 (Claim 4.13 in [28]). *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 $n \ge m_0 \ge n^{\epsilon}$ and for all $\ell \in \{0, \dots, k_{**} - 1\}$, $m_\ell^{\epsilon} = m_{\ell+1}$,*[11] *for some $\epsilon \in (0, \frac{1}{2})$ such that $4 < (m_{k_{**}-1})^{1-2\epsilon}$. Let $A_1, A_2 \subseteq G$ be two ϵ-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 $|A_1| \le |A_2|$. Let $c \in (0, 1-\epsilon)$ and $m \le n^{\frac{1-\epsilon-c}{3}\epsilon^{k_{**}}}$ such that m 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*

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

Proof. Fix $s \in \frac{|A_1|}{m}$, $t \in \frac{|A_2|}{m}$. It follows from the condition $4 < (m_{k_{**}-1})^{1-2\epsilon}$ and Remark 4.9 that the pair (A_1, A_2) satisfies the average condition. Let $U_1 = \left\{a \in A_1 \mid \left|\{b \in A_2 \mid aRb \equiv \neg t(A_1, A_2)\}\right| \ge |A_2|^{\epsilon}\right\}$ 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 4.3, $|U_1| \le |A_1|^{\epsilon}$ and $\forall a \in A_1 \setminus U_1$, $|U_{2,a}| \le |A_2|^{\epsilon}$. Now, given $s \in \left\{1, \dots, \frac{|A_1|}{m}\right\}$, we can bound the probability P_1 that

[11] Implicitly, we are requiring each m_ℓ^{ϵ} to be a natural number. This can be easily achieved by fixing ϵ to be a fraction $\frac{1}{r}$ for some natural number r, and constructing the sequence starting from $m_{k_{**}}$ as $m_\ell = m_{\ell+1}^{r}$ for all $\ell \in \{0, \dots, k_{**} - 1\}$. This is precisely the strategy we will use in following results.

$A_{1,s} \cap U_1 \neq \emptyset$ as follows:

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

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 $t \in \{1, \dots, \frac{|A_2|}{m}\}$, we can similarly bound P_2, the probability that $A_{2,t} \cap \bigcup_{a \in A_{1,s}} U_{2,a} \neq \emptyset$, by:

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

Putting it all together:

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

□

Remark 4.16. The condition $n \geq m_0 \geq n^{\epsilon}$, which is both an upper and lower bound of m_0, is very strong and will be carried over up to Theorem 4.19. 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 4.11 only has a small number of exceptional pairs.

Lemma 4.17 (Claim 4.14 in [28]). *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 $4 < (m_{k_{**}-1})^{1-2\epsilon}$. 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_ℓ 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:*

I. $|A_i| = m_{**}$ for all $i \in \{1, \dots, r\}$.

II. 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.

$$\{(a, b) \in A_i \times A_j \mid aRb \not\equiv t(A_i, A_j)\} = \emptyset.$$

III. $|B| < m_0$.

Proof. We can use Lemma 4.11 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_{**} (I). Consider the resulting partition $\overline{A} = \langle A_i \mid i \in \{1, \dots, r\}\rangle$ with remainder $B = B'$ (III). 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 4.15. Thus, given X the random variable counting the number of exceptional pairs of this kind, we have

$$\mathbb{E}(X) = \sum_{\substack{A_i, A_j \text{ s.t.} \\ A_i \subseteq A'_{i_1}, A_j \subseteq A'_{j_1} \\ i_1 \neq j_1}} \mathbb{E}(X_{A_i, A_j}) = \sum_{\substack{A_i, A_j \text{ s.t.} \\ A_i \subseteq A'_{i_1}, A_j \subseteq A'_{j_1} \\ i_1 \neq j_1}} 1 - \mathbb{P}(\varepsilon^{+}_{A_i, A_j, m_{**}}) \leq \frac{r^2}{2} \frac{2}{n^{c\epsilon^{k_{**}}}},$$

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:

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

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 II. □

Remark 4.18 (Remark 4.15 in [28]). 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

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

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

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

Theorem 4.19 (Theorem 4.16 in [28]). *Let $r, r' > 1$ be two natural numbers, and let $\epsilon = \frac{1}{r}$ and $c = \frac{r'-1}{r'} - \epsilon$.*[12] *Let G be a finite graph with the non-k_*-order property. Let $A \subseteq G$ with $|A| = n$, and $n > 4^{\frac{r^{k_{**}}}{1-2\epsilon}}$. Then, for any $m_{**} \in \left[n^{\frac{(1-\epsilon-c)}{3}\epsilon^{k_{**}+1}}, \left(\frac{\sqrt{2}-1}{\sqrt{2}}\right)^{\frac{1-\epsilon-c}{3}\epsilon^{k_{**}}} n^{\frac{(1-\epsilon-c)}{3}\epsilon^{k_{**}} - \frac{(1-\epsilon-c)c}{3}\epsilon^{2k_{**}}}\right]$, 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:*

I. $|A_i| = m_{**}$ *for all* $i \in \{1, \dots, m\}$.

[12] Choosing ϵ and c this way ensures avoiding aproximation errors, and makes the proof more readable

II. $|B| < m_{**}^{\frac{3}{(1-\epsilon-c)}r^{k_{**}}}$.

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

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:

i. $c \in (0, 1-\epsilon)$, since $0 < \frac{1}{2} - \epsilon \leq \frac{r'-1}{r'} - \epsilon < 1 - \epsilon$.

ii. All m_ℓ's are powers of $m_{k_{**}}$, and so they are divisible by it. Also, since $\frac{3}{1-\epsilon-c} = 3r'$ is an integer by choice of c, m_{**} divides $m_{k_{**}}$. Thus, m_{**} divides m_ℓ for all $\ell \in \{0,\dots,k_{**}\}$.

iii. $(m_{\ell-1})^\epsilon = m_\ell$ for all $\ell \in \{1,\dots,k_{**}\}$, by construction.

iv. $m_{**} \leq n^{\frac{1-\epsilon-c}{3}\epsilon^{k_{**}}}$, by choice of m_{**}.

v. $m_0 = m_{**}^{\frac{3}{1-\epsilon-c}r^{k_{**}}}$, 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 $n^{1-c\epsilon^{k_{**}}}$.

vi. $m_{k_{**}-1} = m_{**}^{\frac{3}{1-\epsilon-c}r} \geq n^{\epsilon^{k_{**}}} > 4^{\frac{1}{1-2\epsilon}}$.

So, all the conditions of Lemma 4.17 are satisfied, and we can use it to get a partition $\overline{A}$ with remainder B satisfying the statement. Notice that II is satisfied by the fact that $|B| < m_0 \leq m_{**}^{\frac{3}{(1-\epsilon-c)}r^{k_{**}}}$. □

Remark 4.20. 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 even partition

Next, we will follow another approach to obtain an even partition. That is, we prove a result similar to that of Lemma 4.10, but this time the size of the resulting quasi-random set can be chosen in advance. The resulting Lemma 4.25 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 ϵ-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 4.21) is highly probable under some very specific conditions (Lemma 4.22).

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 4.21 (Definition 4.18 in [28]). 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 such that $|A| = n$, and for any family of subsets $P \subseteq \mathcal{P}(A)$ such that $|P| \leq n^{\frac{1}{\zeta}}$ and each $B \in P$ satisfies $|B| \leq n^{\epsilon}$, then there exists $U \subseteq A$ with $|U| = \lfloor n^{\xi} \rfloor$ such that for all $B \in P$, $|U \cap B| \leq c$.

Lemma 4.22 (Lemma 4.19 in [28]). *If the reals ϵ, ζ, ξ satisfy $\epsilon \in (0,1)$, $\zeta > 0$ and $0 < \xi < \min(\frac{1}{2}, 1-\epsilon)$, the natural number n is sufficiently large ($n > N(\epsilon, \zeta, \xi, c)$) to satisfy the equation*

$$\frac{1}{2n^{1-2\xi}} + \frac{1}{n^{(1-\xi-\epsilon)c-\frac{1}{\zeta}}} < 1, \tag{8}$$

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

Proof. Let $m = \lfloor n^{\xi} \rfloor$ be the size of the set U we want to build, and let $\mathcal{F}_* = [A]^m$ be the set of sequences of length m of (not necessarily distinct) elements of A. Let μ be the uniform distribution on $\mathcal{F}_*$, i.e. for all $F \subseteq \mathcal{F}_*$, $\mu(F) = \frac{|F|}{|\mathcal{F}_*|}$. Let $P \subseteq \mathcal{F}_*$ be a family of sequences where the elements of the sequences are pairwise distinct. We want to prove that the probability that a random $U \in \mathcal{F}_*$ satisfies:

I. All elements of U are distinct.

II. For all $B \in P$, $|U \cap B| < c$ (where the intersection counts repetitions).

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

Change P to $\mathbb{P}$

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

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

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

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

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

Putting it all together, we have that

$$\mathbb{P}(\text{U does not satisfy I or does not satisfy II}) \leq P_1 + P_2 < \frac{1}{2n^{1-2\xi}} + \frac{1}{n^{c(1-\xi-\epsilon)-\frac{1}{\zeta}}}.$$

Notice that

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

so, the n-large enough condition (8) is well defined and

$$\mathbb{P}(\text{U does not satisfy I or does not satisfy II}) < \frac{1}{2n^{1-2\xi}} + \frac{1}{n^{c(1-\xi-\epsilon)-\frac{1}{\zeta}}} < 1$$

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 4.23. 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 ξ and ϵ grow, and as ζ decreases. A similar pattern is also followed by the large enough condition of n given by (8). For the condition to be met, n needs to grow as the exponents $1-2\xi$ and $(1-\xi-\epsilon)c - \frac{1}{\zeta}$ becomes smaller. That is, the lower bound $N_{4.22}(\epsilon, \zeta, \xi, c)$ on n becomes larger as ξ and ϵ grow, and as ζ and c decrease.

The following claim will be useful in the proof of Lemma 4.25.

Claim 4.24. *Let $f_\epsilon(x)$ be the map such that $f_\epsilon(x) = \lfloor x^\epsilon \rfloor$. Then, for any $x \geq 1$ and $\epsilon_1, \dots, \epsilon_k \in (0,1)$, the following inequality is satisfied:*

$$f_{\epsilon_1\epsilon_2\dots\epsilon_k}(x) \leq f_{\epsilon_1} \circ f_{\epsilon_2} \circ \cdots \circ f_{\epsilon_k}(x) + k - 1$$

Proof in Appendix A. □

Lemma 4.25 (Claim 4.21 in [28]). *Let $k_*, c \in \mathbb{N}$ and $\epsilon, \xi \in \mathbb{R}$ such that:*

I. G is a graph with the non-k_*-order property.

II. $\epsilon \in (0, \frac{1}{2}]$.

III. $\xi \in (0, \frac{\epsilon^{k_{**}}}{2})$.

IV. 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 $n > g_\epsilon^{k_{**}}(N_{4.22}(\epsilon, \frac{1}{k_*}, \frac{\xi}{\epsilon^{k_{**}}}, c))$, where $g_\epsilon(x) = (x+1)^{\frac{1}{\epsilon}}$), if $A \subseteq G$ with $|A| = n$, there is $Z \subseteq A$ such that

i. $|Z| \geq \lfloor n^\xi \rfloor - k_{**}$.

ii. 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 4.10 to get an ϵ-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 II.

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

$$m_\ell \geq g_\epsilon^{-\ell}(n) \geq g_\epsilon^{-k_{**}}(n) > N_{4.22}(\epsilon, \frac{1}{k_*}, \frac{\xi}{\epsilon^{k_{**}}}, c) > N_{4.22}(\epsilon, \zeta, \frac{\xi}{\epsilon^\ell}, c).$$

  Second inequality follows from the fact that function g_ϵ^{-1} satisfies that $g_\epsilon^{-1}(x) \leq x$ for $x \geq 1$. For third inequality we use that g_ϵ^{-1} is decreasing. Forth inequality uses Remark 4.23 and by definition $\zeta = \frac{1}{k_{**}}$.
- $c > \frac{1}{\frac{1}{k_*}(1-\frac{\xi}{\epsilon^{k_{**}}}-\epsilon)} = \frac{1}{\zeta(1-\frac{\xi}{\epsilon^{k_{**}}}-\epsilon)}$, by IV.

Conditions of Lemma 4.22 are met, so $\oplus[m_\ell, \epsilon, \zeta, \frac{\xi}{\epsilon^\ell}, c]$ (as defined in Definition 4.21) holds. We can take $A_{(4.21)}$ and $P_{(4.21)}$ to be A_1 and $P := \left\{\overline{B}_{A_1,b} \mid b \in G\right\}$ respectively, which satisfy:

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

Thus, by $\oplus[m_\ell, \epsilon, \zeta, \frac{\xi}{\epsilon^\ell}, c]$ we have that there exists $Z \subseteq A_1$ such that:

- Condition i is satisfied:

$$|Z| = \left\lfloor m_\ell^{\xi\epsilon^{-\ell}} \right\rfloor = f_{\xi\epsilon^{-\ell}} \circ f_\epsilon^\ell(n) \overset{4.24}{\geq} f_{\xi\epsilon^{-\ell}\epsilon^\ell}(n) - \ell \geq f_\xi(n) - k_{**} = \lfloor n^\xi \rfloor - k_{**},$$

  where $f_\epsilon(x)$ is defined in Claim 4.24.
- Z is f_c-indivisible since $|B \cap Z| \leq c$ for all $B \in P$, satisfying ii.

This proves the statement. □

We now use the previous result to build an even partition. Similarly to Lemma 4.11, we will iteratively extract an f_c-indivisible set from the reminder using Lemma 4.25, while the sufficiently large condition holds.

Theorem 4.26 (Theorem 4.23 in [28]). *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:*

- *$|A_i| = \lfloor n^\xi \rfloor - k_{**}$ 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.*
- *$|B| \leq N := g_\epsilon^{k_{**}}(N_{4.22}(\epsilon, \frac{1}{k_*}, \frac{\xi}{\epsilon^{k_{**}}}, c))$ where $g_\epsilon(x) = (x+1)^{\frac{1}{\epsilon}}$.*

Proof. We 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 4.25 to R_i with the values f_c, ϵ and ξ of the statement of this theorem, to obtain a f_c-indivisible subset A_i of R_i of size (at least[13])

[13]Since f_c-indivisibility is preserved under taking subsets, we can take A_i to have exactly size $\lfloor n^\xi \rfloor - k_{**}$.

$\lfloor n^{\xi} \rfloor - k_{**}$, 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. $\square$

Remark 4.27. 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 4.8.
- 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 desired (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 $-k_{**}$ factor in the size of the parts is just a by-product of using Claim 4.24 to deal with floor functions, and not a crucial dependence of the size of the parts on this parameter. Careful computations may reduce this factor to a constant independent of any parameter.
- The parts are exponentially smaller than 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 whose maximum number of parts only depends on the error and stability parameters; hence it does not grow with the size of the graph. In order to do so, we first prove the existence of a partition whose parts satisfy a property which we prove stronger than regularity: *excellence*.

5.1 Goodness and excellence

We proceed to formalize this concept.

Definition 5.1 (Definition 5.2 (1) in [28])**.** Let G be a finite graph with the non-k_*-property. We say that $A \subseteq G$ is ϵ-*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 (2) in [28])**.** Let G be a finite graph with the non-k_*-property. We say that $A \subseteq G$ is (ϵ, ζ)-*excellent* when A is ϵ-good and, if B is ζ-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 ϵ-*excellent* if A is (ϵ, ϵ)-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 ϵ-good is equivalent to the set being f-indivisible with $f(n) = \epsilon n$, while ϵ-indivisibility is a much stronger condition then ϵ-goodness, as for large enough n, we have that $n^\epsilon < \epsilon n$.

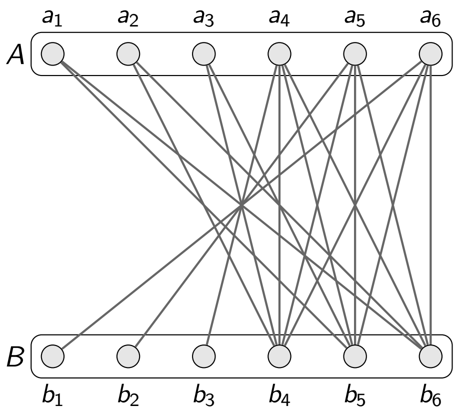
On the other hand, ϵ-excellence carries some kind of reciprocity with other good (and in particular, excellent) sets, which makes it particularly suitable for studying quasi-randomness between pairs of sets. While independence and goodness only bound the number of exceptions with each vertex of the graph 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 ϵ-good set that, as it does not satisfy this reciprocity condition with another good set, it is not ϵ-excellent.

Remark 5.4. If $A, B \subseteq G$ are two (not necessarily disjoint) subsets of vertices with A (ϵ, ϵ')-excellent and B ϵ'-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 (not necessarily disjoint) ϵ-excellent sets, in which case we have that the density (as defined in Definition 2.5) of exceptional edges between them is less than 2ϵ. See Remark 2.6 to see the relation of this density and the real fraction of exceptional edges of the pair.

Remark 5.5. A final important remark, is the fact that differently then most quasi-random properties, ϵ-excellence is not "monotonic". That is, in general, for $\epsilon < \epsilon'$, a set being ϵ-excellent does not imply it being also ϵ'-excellent (and trivially neither the converse). See Figure 4 for a counterexample to the monotonicity of this property. More precisely, each of the two variables in the (ϵ, ϵ')-excellence are oppositely monotonic. 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 (ϵ_1' to ϵ_2') takes less of such sets into account, and relaxing the condition on the "exceptional truth values" (ϵ_1 to ϵ_2) only enlarges the error accepted.



	b_1	b_2	b_3	b_4	b_5	b_6
a_1	0	0	0	0	1	1
a_2	0	0	0	1	0	1
a_3	0	0	0	1	1	0
a_4	0	0	1	1	1	1
a_5	0	1	0	1	1	1
a_6	1	0	0	1	1	1

Figure 4: Example of the ϵ-excellence property not being monotonic. *On the left*, a bipartite graph with two independent sets A and B. A simple exhaustive check shows that A is $\frac{1}{5}$-excellent. On the other hand, raising the value of ϵ up to $\frac{2}{5}$ introduces a new $\frac{2}{5}$-good set B witnessing that A is not $\frac{2}{5}$-excellent, as half of the vertices of A have one truth value, and half the other. *On the right* is the corresponding bi-adjacency matrix.

5.2 Excellent partitions

The first step towards constructing a partition of excellent sets is to prove the existence of such sets under the stability condition. Similar to Lemma 4.10 in Section 4, we prove their existence 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 in 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 latter in Appendix A.

Lemma 5.6 (Claim 5.4 (I) in [28])**.** *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 (ϵ, ζ)-excellent subset $A' \subseteq A$ such that $|A'| \geq \epsilon^{k_{**}-1}|A|$.*

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

I. $A_{\langle\cdot\rangle} = A$.

II. B_η is a ζ-good set witnessing that A_η is not (ϵ, ζ)-excellent, for $k < k_{**}$.

III. $A_{\eta\frown\langle i\rangle} = \{a \in A_\eta \mid t(a, B_\eta) \equiv i\}$ for all $i \in \{0,1\}$ and $k < k_{**}$.

IV. $|A_{\eta\frown\langle i\rangle}| \geq \epsilon|A_\eta|$ for all $i \in \{0,1\}$ and $k < k_{**}$.

V. $|A_\eta| \geq \epsilon^k|A|$, for $k \leq k_{**}$.

VI. $A_\eta = A_{\eta\frown\langle 0\rangle} \sqcup A_{\eta\frown\langle 1\rangle}$, for $k < k_{**}$.

VII. $\overline{A_k} = \{A_\eta \mid \eta \in \{0,1\}^k\}$ is a partition of A, for $k \leq k_{**}$.

First of all, notice that at each step, the non-(ϵ,ζ)-excellence of A_η comes by the initial supposition (negation of the statement's thesis) and by IH from I or V. This allows the existence of the B_η claimed in II. Condition IV follows from the definition of $A_{\eta\frown\langle i\rangle}$ in III and the fact that B_η is witnessing that A_η is not (ϵ,ζ)-excellent. Applying recursively this last point we obtain V. Finally, by definition III, we have the disjoint union VI which ensures the partition VII.

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

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

Therefore, for all $\nu \in \{0,1\}^{<k_{**}}$, we may choose $b_\nu \in B_\nu \setminus \bigcup\{U_{\nu,\eta} \mid \nu \lhd \eta \in \{0,1\}^{k_{**}}\}$. Finally, by III and VI the sequences $\langle a_\eta \mid \eta \in \{0,1\}^{k_{**}}\rangle$ and $\langle b_\nu \mid \nu \in \{0,1\}^{<k_{**}}\rangle$ satisfy:

$$\forall \eta, \nu \text{ such that } \nu \frown \langle i\rangle \lhd \eta,\ a_\eta R b_\nu \equiv i,$$

which forms a k_{**}-tree. This contradicts the tree bound k_{**} (see Definition 3.14). □

Remark 5.7. The two sequences $\langle a_\eta \mid \eta \in \{0,1\}^{k_{**}}\rangle$ and $\langle b_\nu \mid \nu \in \{0,1\}^{<k_{**}}\rangle$, constructed in the proof of Lemma 5.6, 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 *excellent* sets are *good* by definition). Thus, the resulting partition will not only satisfy a quasi-random property between different parts, but actually ensures that the parts are quasi-random within themsleves.

Lemma 5.8 (Claim 5.4 (II) in [28]). *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 $\ell \in \{0, \dots, k_{**}-1\}$.*

Proof in Appendix A. □

Now, we can get a first (not necessarily even) partition by applying the previous lemma recursively, until the remainder is too small to further apply the previous statement.

Lemma 5.9 (Claim 5.14 (1) in [28]). *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 $|A| = n \geq m_0$. 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:*

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

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

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

IV. $|B| < m_*$.

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 $m_0 = m_*$, and thus the lemma cannot be further applied. Notice that, since $\frac{m_\ell}{m_{\ell-1}} \leq \epsilon$, then by Remark 5.5 the pair being $(\frac{m_\ell}{m_{\ell-1}}, \epsilon')$-excellence implies it is also (ϵ, ϵ')-excellence. □

Note that, in Lemma 5.9, if $n < m_0$ then the statement holds for an empty partition where the reminder is the whole set A.

The next step is refining this partition to obtain an even 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 (condition III in 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 (condition IV in Lemma 5.11), which still are excellent. Finally, Lemma 5.15 applies this result to the partition from Lemma 5.9 to get an even excellent partition.

Before getting to it, we prove the following calculus result, which we use in the proof of Claim 5.10. The statement is inspired by [46, page 272] and, for completeness, we provide here a short proof.

Claim 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:

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

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 in [28]). *Let* G *be a finite graph with the non-*k_**-order property. Then:*

I. 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* ϵ*-good subset of size* $n \geq m$, *then a subset* $A' \subseteq A$ *of size* m, *sampled uniformly at random, is* $(\epsilon + \zeta)$*-good with probability* $1 - \xi$.

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

III. 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 (ϵ, ϵ')-excellent and*
- *$A' \subseteq A$ is $(\epsilon + \zeta')$-good,*

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

IV. For all $\zeta \in (0, \frac{1}{2})$, $\zeta' < \zeta$, $r \geq 1$ and for all $\epsilon < \epsilon'$ small enough ($\leq \epsilon_1(\zeta, \zeta')$ from the previous point) there exists $N = N(k_*, \zeta', r)$ such that: if $|A| = n > N$, r divides n and A is (ϵ, ϵ')-excellent, then there exists a partition into r disjoint pieces of equal size, each of which is $(\epsilon + \zeta, \epsilon')$-excellent.

Proof. I. The proof of this point uses the hypergeometric distribution. This distribution works on a set of size N in which $M \leq N$ elements are *fail* elements and the rest are *pass* elements. The hypergeometric describes the probability of having $m \leq n$ fail elements out of n draws *without replacement*. The probability of upperly deviating from the mean M/N by more than φ can be bounded (see [8, 41]) by

$$\mathbb{P}\left(k \geq \left(\frac{M}{N} + \varphi\right)n\right) \leq e^{-2\varphi^2 n}. \tag{9}$$

Now, moving to the context of the statement, we are interested in studying $\overline{B}_{A,b}$ sets, for $b \in G$, such that $|\overline{B}_{A,b}| \geq \epsilon m$, and how they intersect with the randomly sampled A'. We call such sets *bad* sets.

For each $b \in G$ such that $\overline{B}_{A,b}$ is bad, the elements of $\overline{B}_{A,b}$ are our fail elements when drawing from A, and A' is our sample set (where elements are uniformly drawn at random without replacement). This clearly describes a hypergeometric distribution.[14] We define $X_{A,b,A'}$ to be the event of $|\overline{B}_{A,b} \cap A'|$ upperly deviating from ϵm by ζm, i.e. $|\overline{B}_{A,b} \cap A'| \geq (\epsilon + \zeta)m$. Then, since $\epsilon \geq |\overline{B}_{A,b}|/|A|$ by A goodness, the probability of upperly deviating more than ζ from ϵ is at most the probability of deviating more than ζ from the real mean. Thus,

$$\mathbb{P}(X_{A,b,A'}) \leq \mathbb{P}\left(|\overline{B}_{A,b} \cap A'| \geq \left(\frac{|\overline{B}_{A,b}|}{|A|} + \zeta\right)m\right) \leq e^{-2\zeta^2 m},$$

where the second inequality follows (9).

Now we want to study the random variable X counting the number of events in $X_{A,b,A'}$ that are satisfied. That is, $X = \sum_{\text{bad } \overline{B}_{A,b}} \mathbb{1}_{X_{A,b,A'}}$, where $\mathbb{1}_Y$ is the indicator random variable of the event Y. We compute the expectation

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

Following II in Corollary 3.12, the number of intersections of bad $\overline{B}_{A,b}$'s with A', can be bounded by m^{k_*}. Thus, using the First Moment Method, we have that:

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

Last inequality follows Claim 5.10 using the lower bound $m \geq \frac{1}{\zeta^2}(k_* \log \frac{1}{\zeta^2} k_* - \log \xi)$. Thus, with probability at least $1 - \xi$, we have that A' is $(\epsilon + \zeta)$-good.

[14] To avoid confusions, in the notation used to describe the hypergeometric distribution, $N = |A|$, $M = |\overline{B}_{A,b}|$, $n = |A'|$ and $m = |\overline{B}_{A,b} \cap A'|$.

II. Suppose that A' is the subset described in I. We proved that such set satisfies

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

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

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

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

III. Let $B \subseteq G$ be an ϵ'-good set. It follows Remark 5.4 that number of exceptional edges in the pair (A, B) is at most $(\epsilon + (1 - \epsilon)\epsilon')|A||B|$. Also, since A is ϵ-good, each exceptional vertex of B has at least $(1 - \epsilon)|A|$ exceptional edges. Thus,

$$\begin{aligned} |\{b \in B \mid t(b, A) \not\equiv t(B, A)\}| \cdot (1 - \epsilon)|A| &\leq \sum_{\substack{b \in B \\ t(b,A) \not\equiv t(B,A)}} |\{\text{Exceptional edges of } (b, A)\}| \\ &\leq (\epsilon + (1 - \epsilon)\epsilon')|A||B|. \end{aligned} \tag{10}$$

This allows us to upperbound the number of exceptional vertices of B with respect to A':

$$\begin{aligned} |\{b \in B \mid t(b, A') \not\equiv t(B, A)\}| &= |\{b \in B \mid t(b, A) \not\equiv t(B, A)\}| \\ &\leq \frac{(\epsilon + (1 - \epsilon)\epsilon')|A||B|}{(1 - \epsilon)|A|} \\ &= (\epsilon' + \frac{\epsilon}{1 - \epsilon})|B|. \end{aligned} \tag{11}$$

The first equality follows II. and the first inequality uses (10).

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:

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

The first term is the maximum number of exceptional edges associated to exceptional $b \in B$ assuming the worse case of all edges being exceptional in each vertex. The second term bounds the number of exceptional edges incident to non-exceptional vertices $b \in B$, using the fact that A' is $(\epsilon + \zeta')$-good. We are assuming the worse case scenario in which the number of exceptional vertices $b \in B$ is maximum with the bound given by (11).

On the other hand, we have that:

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

which follows from B being ϵ'-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:

$$|Q| < \frac{(1-\epsilon'-\frac{\epsilon}{1-\epsilon})}{(1-\epsilon'-\frac{\epsilon}{1-\epsilon})-\epsilon'}(\epsilon+\zeta')|A'|$$
$$= (1+\frac{\epsilon'}{1-2\epsilon'-\frac{\epsilon}{1-\epsilon}})(\epsilon+\zeta')|A'|.$$

Notice that $f(\epsilon,\epsilon') := \frac{\epsilon'}{1-2\epsilon'-\frac{\epsilon}{1-\epsilon}}$ decreases when ϵ or ϵ' decrease. In particular,

$$f(\epsilon,\epsilon') \overset{\epsilon'\to 0}{\longrightarrow} 0.$$

Then,

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

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.

IV. Let $\zeta,\zeta',\epsilon,\epsilon'$ and r be given satisfying the conditions of the statement. Set $\xi = \frac{1}{r+1}$. We 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. Consider all functions $h : A \longrightarrow \{1,\dots,r\}$ such that for all $s < n$ we have that $|\{a \in A \mid h(a) = s\}| = \frac{n}{r}$, and draw one h uniformly at random (this can be seen as sampling a partition of A into r parts uniformly at random). Then, 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 $s \in \{1,\dots,r\}$. 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 statement I to get that

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

Observe that the hypotheses of statement III are satisfied, crucially A is (ϵ,ϵ')-excellent. It follows that if A' is $(\epsilon+\zeta')$-good, then it is also $(\epsilon+\zeta,\epsilon')$-excellent. Therefore:

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

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

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

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

Remark 5.12. In some of the following arguments, we would like to use statement IV from Lemma 5.11 with $\epsilon' > k\cdot(\epsilon+\zeta)$, 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:

I. $\frac{\epsilon}{1-\epsilon} \leq \frac{\frac{1}{t}}{1-\frac{1}{t}} = \frac{\frac{1}{t}}{\frac{t-1}{t}} = \frac{1}{t-1}$.

II. $1 - 2\epsilon' - \frac{\epsilon}{1-\epsilon} \geq 1 - \frac{2}{t'} - \frac{1}{t-1} > 1 - \frac{3}{t'-1} = \frac{t'-4}{t'-1}$.

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

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 III gives us:

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

All in all, we conclude that, if

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

then statement IV from Lemma 5.11 holds with $\epsilon' \geq k \cdot (\epsilon + \zeta)$.

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

Lemma 5.13. *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 (ϵ, ϵ')-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, so statement IV from Lemma 5.11 holds together with $\frac{\epsilon'}{k} \geq \epsilon + \zeta$ and, given Remark 5.5, we are done. □

Remark 5.14. It is sufficient to choose $\zeta' = \frac{1}{4k}\epsilon'$, for $N_{5.13}$ to be large enough, 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 even one.

Lemma 5.15 (Claim 5.14 (1A) in [28])**.** *Let G be a finite graph with the non-k_*-order property. Let ϵ' and ϵ 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:*

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

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

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

IV. $|B| < m_*$.

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 $\ell \in \{0, \ldots, k_{**} - 1\}$. 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 $\overline{A}' = \left\langle A'_j \mid j \in \{1, \ldots, j(*)\} \right\rangle$ 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 $\ell \in \{0, \ldots, k_{**} - 1\}$. Putting together all the new subparts, we obtain a new partition $\overline{A} = \langle A_i \mid i \in \{1, \ldots, i(*)\}\rangle$ with remainder B, satisfying all the conditions of the statement. □

Notice that our partition is even with a small reminder. We can turn it into an equitable one, as the next lemma proves, at the cost of another slight increase of the excellence parameter.

Lemma 5.16 (Claim 5.14 (2) in [28]). *Under the same condition of Lemma 5.15, we can get a partition* $\overline{A} = \langle A_i \mid i \in \{1, \ldots, i(*)\}\rangle$ *with no remainder, such that:*

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

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

III. *For all* $i \in \{1, \ldots, i(*)\}$, A_i *is* (ϵ'', ϵ')*-excellent, for some* ϵ'' *satisfying*

$$0 < \epsilon'' \leq \frac{\frac{\epsilon'}{k} m_{**} + \left\lceil \frac{m_*}{i(*)} \right\rceil}{m_{**} + \left\lceil \frac{m_*}{i(*)} \right\rceil}.$$

IV. $A = \bigcup \overline{A}$.

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

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

Notice that we are allowing $B_i = \emptyset$. Then, the new partition $\overline{A} = \langle A'_i \cup B_i \mid i \in \{1, \ldots, i(*)\}\rangle$ satisfies the conditions I, II and IV by construction. To conclude, notice that, since A'_i is $(\epsilon'/k, \epsilon)$-excellent, we have that for each ϵ'-good set C, the number of exceptions is bounded by

$$\begin{aligned} |\{a \in A_i \mid t(a, C) \not\equiv t(A_i, C)\}| &\leq \frac{\epsilon'}{k}|A'_i| + |B_i| \\ &= \frac{\frac{\epsilon'}{k}|A'_i| + |B_i|}{|A'_i| + |B_i|}(|A'_i| + |B_i|) \\ &\leq \frac{\frac{\epsilon'}{k} m_{**} + \left\lceil \frac{m_*}{i(*)} \right\rceil}{m_{**} + \left\lceil \frac{m_*}{i(*)} \right\rceil}|A_i|, \end{aligned} \tag{12}$$

where first inequality assumes the worse case scenario of all added vertices B_i being exceptional. This proves that III can be satisfied. ☐

We now have an (ϵ'', ϵ')-excellent equitable partition. Also ϵ'' 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 ϵ'' by ϵ', thus ensuring that the partition is ϵ'-excellent.

Remark 5.17 (Remark 5.14 (3) in [28]). Under the hypotheses of Lemma 5.16, if:

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

II. $m_* \leq \frac{\frac{\epsilon'}{k}n+1}{\frac{\epsilon'}{k}+1}$,

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

Proof. Using the notation of the proof of Lemma 5.16, we notice that, if $|B_i| \leq 2\frac{\epsilon'}{k}|A_i|$ for all $i \in \{1, \dots, i(*)\}$, then ϵ'' can be bounded by (starting at the more precise (12)):

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

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:

$$\frac{|B_i|}{|A_i|} \leq \frac{m_* - 1}{n - m_*} + \frac{1}{m_{**}}. \tag{13}$$

Finally, notice that condition I implies:

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

and condition II implies:

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

We conclude:

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

where the first inequality follows from (13). This completes the proof. □

We now summarize the conditions and arguments from previous results into the following statement.

Lemma 5.18 (Corollary 5.15 in [28]). *Let G be a graph with the non-k_*-order property. Suppose that we are given:*

I. A real value $\epsilon \leq \min(\frac{1}{5}, \frac{1}{2^{k_{**}}})$.

II. 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. $m_* := q^{k_{**}} m_{**}$.

III. $A \subseteq G$ 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 ϵ-excellent.

Proof. First of all, notice that if condition 2.b is satisfied, then we also have $m_{**} \geq \frac{3}{\epsilon}$. To prove the statement, we use Lemma 5.16 in the context of Remark 5.17 with the following parameters: $k = 3$, $\epsilon'_{5.16} = \epsilon$ 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 $k \geq 3$, then $\epsilon''_{5.16} \leq \epsilon'_{5.16}$, and thus by Remark 5.5 they are also $\epsilon'_{5.16}$-excellent, satisfying ii. □

To conclude, we prove how, under some minimal conditions of the excellence parameter ϵ and the minimum size of a partition m, the conditions of the previous lemma can be satisfied. We rewrite the statement accordingly as follows.

Theorem 5.19 (Theorem 5.18 in [28]). *Let k_* be given (which by Theorem 3.15 gives a bound on k_{**}). 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:*

I. The number of parts is bounded by $m \leq i(*) \leq M := \max(\lceil \frac{12}{\epsilon} \rceil^{k_{**}+1}, 4m)$.

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

III. For all $i \in \{1, \dots, i(*)\}$, A_i is ϵ-excellent.

Proof. We aim at applying 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:

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

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

By I 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 $m_* := q^{k_{**}} m_{**}$.

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

We can apply Lemma 5.18 to obtain a partition satisfying II, III.
We proceed to bound the number of parts of the partition $i(*)$. 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 in the previous equation, 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 $N_{5.19}(\epsilon, m, k_*)$, and thus n, actually needs to be. First of all, we see that:

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

Then, since m_{**} is the smallest integer larger than $\frac{1}{q}N_{5.13}(4, k_*, \epsilon, q^{k_{**}})$, we conclude:

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

5.3 The Stable Regularity Lemma

As mentioned in the beginning of this section, we prove that excellence is a stronger condition then regularity (see Definition 2.7). 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.

Lemma 5.21 (Lemma 5.17 in [28]). *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 (not necessarily disjoint) pair (A, B) satisfies that A is ϵ_1-excellent and B is ϵ_2-good. 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:*

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

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

In particular, for any $\epsilon_0, \epsilon \in (0, \frac{1}{2})$ such that $\epsilon_0 \leq \frac{\epsilon^2}{2}$, and for any two ϵ_0-excellent sets A, B, we have that:

i. *(A, B) is ϵ-regular.*

ii. *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 $U = \{a \in A \mid t(a, B) \not\equiv t(A, B)\}$, i.e. the set of exceptional vertices $a \in A$. Then,

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

and

$$Z' \subseteq (U \times B') \cup \left(\bigcup_{a \in A' \setminus U} \{a\} \times \overline{B}_{B,a} \right).$$

Notice that, by the ϵ_1-excellence of A, $|U| < \epsilon_1|A|$. Furthermore, by the ϵ_2-goodness of B, 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

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

proving I. 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 II. Let's now prove i and ii. First of all, notice that:

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

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

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

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

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 $\frac{\epsilon_1+\epsilon_2}{\epsilon_3}$ from 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. Thus, we conclude that (A, B) is ϵ-regular (statement i) and that $d(A', B')$ is either $< \epsilon$ or $\geq 1 - \epsilon$ (statement ii) for each $A' \subseteq A$ and $B' \subseteq B$ such that $|A'| \geq \epsilon|A|$ and $|B'| \geq \epsilon|B|$. □

Remark 5.22. The fact that the pairs are almost empty or almost full is to be expected. Indeed, in a regular pair of density far away from 0 and 1 we can (asymptotically) find any bi-induced bipartite graph, in particular the half-graph. It can be derived from a similar argument to that of Lemma 6.7.

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

Theorem 5.23 (Theorem 5.19 in [28]). *For every $k_* \in \mathbb{N}$ (which by Theorem 3.15 gives a bound on k_{**}) 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 $|A| \geq N$, the following holds. There exists $m < \ell < M$ 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 (not necessarily distinct) $i, j \in \{1, \dots, \ell\}$,*

I. $||A_i| - |A_j|| \leq 1$, i.e. the partition is equitable.

II. (A_i, A_j) is ϵ-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 $d(B_i, B_j) \geq 1 - \epsilon$.

III. 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 obtain I, II and III applying Theorem 5.19 to A with $\epsilon_{5.19} = \frac{\epsilon^2}{2}$, and then use Lemma 5.21 to get the ϵ-regularity of pairs from the $\frac{\epsilon^2}{2}$-excellence of the parts.

Otherwise, to get I and II, we use the same argument for $\epsilon' = \min(\frac{1}{5}, \frac{1}{2^{k_{**}}}) \leq \epsilon$. Then, since regularity is monotone, we get the wanted ϵ-regularity from the resulting ϵ'-regularity. In this last case, the bound on M is $M \leq \max(\lceil \frac{12}{\epsilon'} \rceil^{k_{**}+1}, 4m)$. □

Remark 5.24. 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_*, ϵ and m:

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

6. Property testing

This section is dedicated at showcasing the benefits of the stable regularity lemma regarding its partition size and lack of irregular pairs, by giving a concrete application in property testing. More specifically we focus at studying H-freeness in stable graphs. We now formalize some key concepts that were loosely defined in the introduction.

Definition 6.1. We say that a graph G is ϵ-*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 ϵ-*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 the following.

I. If $G \in \mathcal{P}$, then $\mathbb{P}(\mathcal{A} \text{ accepts } G) \geq \frac{2}{3}$.

II. If G is ϵ-far from satisfying $\mathcal{P}$, then $\mathbb{P}(\mathcal{A} \text{ rejects } G) \geq \frac{2}{3}$.

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 is adjacent or not.

Definition 6.3. We say that a property $\mathcal{P}$ is *testable* if there exists an ϵ-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 ϵ.

The remaining of this section is dedicated to the construction of an ϵ-test for H-freeness in stable graphs. Such an ϵ-test needs to be able to distinguish between graphs that are H-free and graphs that are ϵ-far from it, with some error. In fact, our ϵ-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 6.9. This theorem uses the Stable Regularity Lemma to prove that if a graph is ϵ-far from being H-free then it contains many induced copies of H (at least a fraction of all induced subgraphs of size $|H|$). 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.

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 [1].

Definition 6.4. A graph H is γ-*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 6.5. A graph H is η-*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 6.6 (Lemma 3.1 in [1]). *Let $\epsilon \leq \epsilon' < \frac{1}{2}$ and $\delta \in (0,1)$. If (A, B) is an (not necessarily disjoint) ϵ-regular pair with density δ, $A' \subseteq A$ 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

$$\begin{aligned} |A''| &\geq \frac{\epsilon}{\epsilon'}|A'| \geq \frac{\epsilon}{\epsilon'}\epsilon'|A| = \epsilon|A| \text{ and} \\ |B''| &\geq \frac{\epsilon}{\epsilon'}|B'| \geq \frac{\epsilon}{\epsilon'}\epsilon'|B| = \epsilon|B|. \end{aligned}$$

By ϵ-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 ϵ-regular, $|d(A, B) - d(A', B')| < \epsilon$, and thus,

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

□

In the context of the regularity method, the *reduced graph* H associated to a partition $\{A_1, \dots, A_m\}$ of a graph A, is a graph (possibly with loops) with m vertices $v_1, \dots, v_m$ such that for all (not necessarily distinct) $i, j \in \{1, \dots, m\}$, it satisfies that $v_i R v_j$ if and only if $d(A_i, A_j) > T$, for some threshold value $T \in [0, 1)$. Since pairs of parts in a stable regular partition are almost empty or almost full, any threshold value in $(\epsilon, 1 - \epsilon)$ associates *almost full* pairs with edges, and *almost empty* pairs with non-edges. In particular, one may consider $T = 1/2$.

The pivotal point in the proof of Theorem 6.9 is the fact that, if the reduced graph from a regular partition contains a copy of H as an induced subgraph, and all edges lie in regular pairs, then the original graph contains many induced copies of H (this is a version of the so called *Counting Lemma* from [22]). The following lemma formalizes this idea.

Lemma 6.7 (Lemma 3.2 in [1]). *For every real number $\delta \in (0, 1)$ and integer $\ell > 0$ 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 ℓ-tuple of (not necessarily disjoint) sets of vertices of a graph G such that for every $1 \leq i < i' \leq \ell$, the pair $(V_i, V_{i'})$ is ϵ-regular, with density at least δ 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 ℓ-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 ℓ. The case $\ell = 1$ 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 ℓ. We proceed to prove that the following values η and ϵ hold.

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

For each $1 < i \leq \ell$, 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 $U \in [V_1]^{\geq \epsilon|V_1|}$ together with V_i would form a subpair (U, V_i) with density $< \delta - \epsilon$ which, by Lemma 6.6 contradicts the ϵ-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 $1 < i \leq \ell$. In particular, since $\epsilon \leq \frac{1}{2\ell-2}$ we have that $(\ell-1)\epsilon \leq \frac{1}{2}$ and then $1-(\ell-1)\epsilon \geq \frac{1}{2}$, 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 $\epsilon \leq \frac{1}{2}\delta$, Lemma 6.6 implies that for all $1 < i < i' \leq \ell$, the pair $(V_i', V_{i'}')$ is $(\frac{\epsilon}{\delta-\epsilon})$-regular, and given that $(\frac{\epsilon}{\delta-\epsilon}) \leq (\frac{2\epsilon}{\delta}) \leq \epsilon(\frac{1}{2}\delta, \ell-1)$, it is $\epsilon(\frac{1}{2}\delta, \ell-1)$-regular. Also, it has density at least $\delta-\epsilon \geq \frac{1}{2}\delta$. By the induction hypothesis, we have at least

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

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 η satisfies the lemma, and we are done. □

Remark 6.8. We now give a non-recursive form of ϵ and a lower bound of η, for all $\ell \geq 1$.

$$\epsilon(\delta, \ell) = \left(\frac{1}{2}\right)^{\frac{\ell(\ell-1)}{2}} \cdot \delta^{\ell-1}.$$
$$\eta(\delta, \ell) \geq \left(\frac{1}{2}\right)^{\frac{\ell^3+5\ell-6}{6}} \cdot \delta^{\frac{\ell(\ell-1)}{2}}.$$

We are now ready to prove the main theorem of this section. The proof is similar to that of [1, 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 5.23 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 γ-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 6.7 and conclude that H is abundant in G. Such conclusion is formalized in the following theorem.

Theorem 6.9. *For every k_*, γ, ℓ there is a $\eta(k_*, \gamma, \ell)$ such that if H is a graph with ℓ vertices, G has the non-k_*-order property and H is γ-unavoidable in G, then H is η-abundant in G.*

Proof. Apply Theorem 5.23 to G with $\epsilon = \min\big(\sqrt{\gamma}, \epsilon_{6.7}(1-\gamma, \ell)\big)$, 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

$$M \leq \left\lceil 12\max\left(\frac{1}{\sqrt{\gamma}}, \frac{1}{\epsilon_{6.7}(1-\gamma, \ell)}\right)\right\rceil^{2^{k_*+1}-1}$$

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

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 ϵ^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 ϵ^2 of the edges between (disjoint) parts.

- For each self-pair (A_i, A_i), if the pair's density is at most ϵ^2, again we remove all edges in A_i. Otherwise, the pair's density is at least $1-\epsilon^2$, and we add all remaining edges. Notice that, in self-pairs, the density (1 minus the density respectively) is at most the fraction of possible edges in the pair that actually are edges (non-edges), as noted in Remark 2.6. Thus, the fraction of changed edges in all self-pairs is at most ϵ^2.

The resulting graph G' differs from G in at most $\epsilon^2\binom{|G|}{2} \leq \gamma\binom{|G|}{2}$ edges, and satisfies that each pair of parts (disjoint or not) is either complete or empty. Then, the γ-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 $v_{i_1} \in A_{i_1}, \dots, v_{i_\ell} \in A_{i_\ell}$. Notice that $A_{i_1}, \dots, A_{i_\ell}$ satisfy the conditions of Lemma 6.7 with $\delta_{6.7} = 1-\gamma$: each pair $(A_{i_j}, A_{i_{j'}})$ with $j \neq j'$ is ϵ-regular, and since $\epsilon \leq \epsilon_{6.7}(1-\gamma,\ell)$, in particular is $\epsilon_{6.7}(1-\gamma,\ell)$-regular. Hence, the lemma guarantees that there are at least $\eta_{6.7}(1-\gamma,\ell)\prod_{j=1}^{\ell}\{A_{i_j}, j\}$ copies of H in G.

The fraction of induced copies of H in G is at least

$$\frac{\eta_{6.7}(1-\gamma,\ell)\prod_{j=1}^{\ell}\{A_{i_j}\}}{n^\ell} \geq \eta_{6.7}(1-\gamma,\ell)\left(\frac{n/M}{n}\right)^\ell = \eta_{6.7}(1-\gamma,\ell)(M)^{-\ell} =: \eta,$$

and H is at least η-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 [1] proves. The difference is that the resulting η is much larger (although not given explicitly). The main reasons of such an improvement, as mentioned earlier, is that neither a double partition is required to elude irregular pairs, neither the bound on the number of parts is a tower of powers. This allows the resulting set of parts $A_{i_1}, \dots, A_{i_\ell}$, the ones used as a restricted copy of H in G in the previous theorem, to be a larger proportion of the whole graph, and thus inducing more copies of H.

Remark 6.10. We now provide a more explicit lower bound for η only depending on γ, k_* and ℓ is:

$$\eta \geq \left(\frac{1}{2}\right)^{\frac{\ell^3+5\ell-6}{6}} \cdot (1-\gamma)^{\frac{\ell(\ell-1)}{2}} \cdot \left(\frac{1}{24}\min\left\{\sqrt{\gamma}, \left(\frac{1}{2}\right)^{\frac{\ell(\ell-1)}{2}}(1-\gamma)^{\ell-1}\right\}\right)^{\ell\left(2^{k_*+1}-1\right)}.$$

6.2 The algorithm

Now we have all the tools needed to build an ϵ-test $\mathcal{A}$, which decides H-freeness for a given graph H of size ℓ, in the context of graphs with the non-k_*-order property.

$\mathcal{A} = \mathcal{A}(H, \epsilon, k_*)$ works as follows. Given a graph H, with all edges known, a natural number k_*, a real number ϵ, and a graph G with the non-k_*-order property, whose edges are unknown, the algorithm computes $\ell = |H|$, and the value $t = \frac{\ell\log(\frac{2}{3})}{\log(1-\eta_{6.9}(k_*,\epsilon,\ell))}$. It then samples t different vertices from G uniformly at random. $\mathcal{A}$ queries all edges from the sampled set, and checks whether a copy of H as an induced subgraph can be found in it. If a copy of H is found, then $\mathcal{A}$ accepts G. Otherwise, $\mathcal{A}$ rejects it. See Algorithm 1 for a more detailed step to step description of $\mathcal{A}$.

Algorithm 1 ϵ-test $\mathcal{A}$ for deciding H-freeness

Require: a graph H of size ℓ, a natural number k_* and a real number $\epsilon > 0$.
Require: an oracle $\mathcal{O}$ accepting queries of whether two vertices of a graph G are adjacent. The graph G has the non-k_*-order property, and only its size n is known.

1: $t \leftarrow \ell \log(\frac{2}{3}) / \log(1 - \eta_{6.9}(k_*, \epsilon, \ell))$ ▷ Compute sample size t
2: **if** $n < \ell$ **then** ▷ Check if G is large enough to contain H
3:     *$\mathcal{A}$ rejects G*
4: **else if** $n < t$ **then** ▷ Check if G is small enough to query all edges
5:     query all pairs of vertices of G to $\mathcal{O}$
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:         *$\mathcal{A}$ accepts G*
8:     **else**
9:         *$\mathcal{A}$ rejects G*
10:     **end if**
11: **else** ▷ Sample t vertices uniformly at random, without repetitions
12:     $S \leftarrow \emptyset$
13:     **while** $i \leq t$ **do**
14:         $s_i \sim G$
15:         **while** $s_i \in S$ **do** ▷ Repeat until a new vertex is sampled
16:             $s_i \sim G$
17:         **end while**
18:         $S \leftarrow S \cup \{s_i\}$
19:     **end while**
20:     query all pairs of vertices of S to $\mathcal{O}$
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:         *$\mathcal{A}$ accepts G*
23:     **else**
24:         *$\mathcal{A}$ rejects G*
25:     **end if**
26: **end if**

We now proceed to prove that, indeed, $\mathcal{A}$ is an ϵ-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 ϵ-far from being H-free, Theorem 6.9 ensures that H is $\eta_{6.9}(k_*, \epsilon, \ell)$-abundant in G. Thus, checking t_* times whether a random sample of ℓ vertices contains an induced copy of H, the probability of not finding any copy of H is at most $(1 - \eta_{6.9}(k_*, \epsilon, \ell))^{t_*}$. By letting $t_* = \frac{\log(\frac{2}{3})}{\log(1 - \eta_{6.9}(k_*, \epsilon, \ell))}$ 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) $t := t_* \cdot \ell$. Hence, querying for all pairs of vertices in a random sample of t vertices has strictly more probabilities of finding a copy of H appearing 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

$$q \leq \binom{t}{2} \leq \left(\frac{\ell \log(\frac{2}{3})}{\log(1 - \eta_{6.9}(k_*, \epsilon, \ell))} \right)^2 .$$

7. Conclusion

In this thesis, we delved into the powerful framework of Szemerédi's Regularity Lemma, focusing on a restricted version for the class of stable graphs. Our primary contributions have been threefold: we have provided a detailed, self-contained combinatorial proof of the Stable Regularity Lemma from [28], clarifying parameters and simplifying arguments; we have developed a unified notational framework to bridge concepts from extremal graph theory, stability, and property testing; and, most significantly, we have designed an efficient property testing algorithm for H-freeness specifically for stable graphs.

Our motivation for this final contribution stemmed from a desire to exploit the most striking feature of the Stable Regularity Lemma: the complete absence of irregular pairs in its partitions. We hypothesized that this structural guarantee could be a powerful tool in property testing, and we found a direct application in testing for forbidden induced subgraphs (H-freeness).

During the design of our tester, we observed that a similar argument could be constructed using the (Ultra-) Strong Regularity Lemma for graphs with bounded VC-dimension, as presented in [25]. The stronger regularity conditions of this lemma allows one to effectively "avoid" the issue of irregular pairs when dealing with induced subgraphs. This led to a crucial comparison of the bounds on the number of parts in the partitions. The (Ultra-) Strong lemma provides a bound of $(1/\epsilon)^{c \cdot d^2}$, where d is the bound on the VC-dimension, while the Stable Regularity Lemma's bound is $(1/\epsilon)^{c \cdot 2^k}$, where k is the stability parameter. As established in Section 3, stability is a stronger condition than bounded VC-dimension, and the parameters k and d are closely related (when comparing the two in this situation, we can assume $k = d$).

This comparison suggests that the bound for the Stable Regularity Lemma might not be optimal. The fact that a broader class of graphs (bounded VC-dimension) admits a partition with a polynomially better exponent raises the possibility that the bound for stable graphs could be improved to something akin to $(1/\epsilon)^{c \cdot k^2}$, a question also posed by [47]. Despite the potentially suboptimal bound, the Stable Regularity Lemma remains a valuable tool. Its guarantee of having no irregular pairs whatsoever allows for a uniquely clean and straightforward proof of the testability of H-freeness in the context of stable graphs.

Finally, this work opens several avenues for future research, which could not be included in this thesis due to time constraints. One promising direction is to leverage the lack of irregular pairs to test for H_n-freeness, where the forbidden subgraph H_n grows in size with the input graph G. Another intriguing possibility is to move beyond simple freeness testing towards subgraph counting, using the clean structure of the stable regular partition to develop a tester that provides an interval estimate for the number of induced copies of a graph H.

References

[1] N. Alon, E. Fischer, M. Krivelevich, and M. Szegedy. Efficient testing of large graphs. *Combinatorica*, 20(4):451–476, 2000.

[2] N. Alon, E. Fischer, and I. Newman. Efficient Testing of Bipartite Graphs for Forbidden Induced Subgraphs. en. *SIAM Journal on Computing*, 37(3):959–976, Jan. 2007. ISSN: 0097-5397, 1095-7111. DOI: 10.1137/050627915. URL: http://epubs.siam.org/doi/10.1137/050627915 (visited on 08/16/2025).

[3] N. Alon, J. Fox, and Y. Zhao. Efficient arithmetic regularity and removal lemmas for induced bipartite patterns. *arXiv preprint arXiv:1801.04675*, 2018.

[4] N. Alon and A. Shapira. A characterization of the (natural) graph properties testable with one-sided error. *SIAM Journal on Computing*, 37(6):1703–1727, 2008.

[5] A. Basu. Irregular pairs in half graphs - szemeredi regularity. MathOverflow. eprint: https://mathoverflow.net/q/404954. URL: https://mathoverflow.net/q/404954. (version: 2021-09-27).

[6] C. Borgs, J. T. Chayes, L. Lovász, V. T. Sós, and K. Vesztergombi. Convergent sequences of dense graphs II. Multiway cuts and statistical physics. *Ann. of Math. (2)*, 176(1):151–219, 2012. ISSN: 0003-486X,1939-8980. DOI: 10.4007/annals.2012.176.1.2. URL: https://doi.org/10.4007/annals.2012.176.1.2.

[7] C. Borgs, J. T. Chayes, L. Lovász, V. T. Sós, and K. Vesztergombi. Convergent sequences of dense graphs. I. Subgraph frequencies, metric properties and testing. *Adv. Math.*, 219(6):1801–1851, 2008. ISSN: 0001-8708,1090-2082. DOI: 10.1016/j.aim.2008.07.008. URL: https://doi.org/10.1016/j.aim.2008.07.008.

[8] V. Chvátal. The tail of the hypergeometric distribution. *Discrete Mathematics*, 25(3):285–287, 1979. ISSN: 0012-365X. DOI: https://doi.org/10.1016/0012-365X(79)90084-0. URL: https://www.sciencedirect.com/science/article/pii/0012365X79900840.

[9] D. Conlon and J. Fox. Graph removal lemmas. *Surveys in combinatorics*, 409:1–49, 2013.

[10] R. Diestel. Extremal graph theory. In *Graph theory*, pages 179–226. Springer, 2024.

[11] J. Fox, J. Pach, and A. Suk. Erdős–Hajnal Conjecture for Graphs with Bounded VC-Dimension. en. *Discrete & Computational Geometry*, 61(4):809–829, June 2019. 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).

[12] A. Frieze and R. Kannan. Quick Approximation to Matrices and Applications. en. *Combinatorica*, 19(2):175–220, Feb. 1999. ISSN: 0209-9683, 1439-6912. DOI: 10.1007/s004930050052. URL: http://link.springer.com/10.1007/s004930050052 (visited on 12/11/2018).

[13] W. T. Gowers. Hypergraph regularity and the multidimensional Szemerédi theorem. *Ann. of Math. (2)*, 166(3):897–946, 2007. ISSN: 0003-486X,1939-8980. DOI: 10.4007/annals.2007.166.897. URL: https://doi.org/10.4007/annals.2007.166.897.

[14] W. T. Gowers. Lower bounds of tower type for Szemerédi's uniformity lemma. *Geom. Funct. Anal.*, 7(2):322–337, 1997. ISSN: 1016-443X,1420-8970. DOI: 10.1007/PL00001621. URL: https://doi.org/10.1007/PL00001621.

[15] A. Gyárfás, M. Ruszinkó, G. N. Sárközy, and E. Szemerédi. Three-color Ramsey numbers for paths. *Combinatorica*, 27(1):35–69, 2007. ISSN: 0209-9683,1439-6912. DOI: 10.1007/s00493-007-0043-4. URL: https://doi.org/10.1007/s00493-007-0043-4.

[16] J. Hladký, J. Komlós, D. Piguet, M. Simonovits, M. Stein, and E. Szemerédi. The approximate Loebl-Komlós-Sós conjecture I: The sparse decomposition. *SIAM J. Discrete Math.*, 31(2):945–982, 2017. ISSN: 0895-4801,1095-7146. DOI: 10.1137/140982842. URL: https://doi.org/10.1137/140982842.

[17] J. Hladký, J. Komlós, D. Piguet, M. Simonovits, M. Stein, and E. Szemerédi. The approximate Loebl-Komlós-Sós conjecture II: The rough structure of LKS graphs. *SIAM J. Discrete Math.*, 31(2):983–1016, 2017. ISSN: 0895-4801,1095-7146. DOI: 10.1137/140982854. URL: https://doi.org/10.1137/140982854.

[18] J. Hladký, J. Komlós, D. Piguet, M. Simonovits, M. Stein, and E. Szemerédi. The approximate Loebl-Komlós-Sós conjecture III: The finer structure of LKS graphs. *SIAM J. Discrete Math.*, 31(2):1017–1071, 2017. ISSN: 0895-4801,1095-7146. DOI: 10.1137/140982866. URL: https://doi.org/10.1137/140982866.

[19] J. Hladký, J. Komlós, D. Piguet, M. Simonovits, M. Stein, and E. Szemerédi. The approximate Loebl-Komlós-Sós conjecture IV: Embedding techniques and the proof of the main result. *SIAM J. Discrete Math.*, 31(2):1072–1148, 2017. ISSN: 0895-4801,1095-7146. DOI: 10.1137/140982878. URL: https://doi.org/10.1137/140982878.

[20] W. Hodges. *Model theory*. Cambridge university press, 1993.

[21] J. Komlós, G. Sárközy, and E. Szemerédi. Proof of the alon–yuster conjecture. *Discrete Mathematics*, 235(1):255–269, 2001. ISSN: 0012-365X. DOI: https://doi.org/10.1016/S0012-365X(00)00279-X. URL: https://www.sciencedirect.com/science/article/pii/S0012365X0000279X. Chech and Slovak 3.

[22] J. Komlós, A. Shokoufandeh, M. Simonovits, and E. Szemerédi. The regularity lemma and its applications in graph theory. English. In *Theoretical aspects of computer science. Advanced lectures*, pages 84–112. Berlin: Springer, 2002. ISBN: 3-540-43328-7. DOI: 10.1007/3-540-45878-6_3.

[23] D. Kühn and D. Osthus. Hamilton decompositions of regular expanders: applications. *J. Combin. Theory Ser. B*, 104:1–27, 2014. ISSN: 0095-8956,1096-0902. DOI: 10.1016/j.jctb.2013.10.006. URL: https://doi.org/10.1016/j.jctb.2013.10.006.

[24] L. Lovász. *Large networks and graph limits*, volume 60. American Mathematical Soc., 2012.

[25] L. Lovász and B. Szegedy. Regularity partitions and the topology of graphons. In *An irregular mind*. Volume 21, Bolyai Soc. Math. Stud. Pages 415–446. János Bolyai Math. Soc., Budapest, 2010. 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.

[26] L. Lovász and B. Szegedy. Szemerédi's Lemma for the Analyst. en. *GAFA Geometric And Functional Analysis*, 17(1):252–270, Apr. 2007. 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).

[27] M. Malliaris and S. Shelah. Notes on the stable regularity lemma. en. *The Bulletin of Symbolic Logic*, 27(4):415–425, Dec. 2021. 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).

[28] M. Malliaris and S. Shelah. Regularity lemmas for stable graphs. *Transactions of the American Mathematical Society*, 366(3):1551–1585, 2014.

[29] B. Nagle, V. Rödl, and M. Schacht. The counting lemma for regular k-uniform hypergraphs. *Random Structures Algorithms*, 28(2):113–179, 2006. ISSN: 1042-9832,1098-2418. DOI: 10.1002/rsa.20117. URL: https://doi.org/10.1002/rsa.20117.

[30] T. Nguyen, A. Scott, and P. Seymour. Induced subgraph density. vi. bounded vc-dimension. *arXiv preprint arXiv:2312.15572*, 2023.

[31] A. Pajor. Sous-spaces 1: des espaces de banach travaux en cours. *Hermann, Paris*, 1985.

[32] V. Rödl. On universality of graphs with uniformly distributed edges. *Discrete Mathematics*, 59(1):125–134, 1986. ISSN: 0012-365X. DOI: https://doi.org/10.1016/0012-365X(86)90076-2. URL: https://www.sciencedirect.com/science/article/pii/0012365X86900762.

[33] V. Rödl and A. Ruciński. Threshold functions for Ramsey properties. *J. Amer. Math. Soc.*, 8(4):917–942, 1995. ISSN: 0894-0347,1088-6834. DOI: 10.2307/2152833. URL: https://doi.org/10.2307/2152833.

[34] V. Rödl and M. Schacht. Regular partitions of hypergraphs: regularity lemmas. *Combin. Probab. Comput.*, 16(6):833–885, 2007. ISSN: 0963-5483,1469-2163. DOI: 10.1017/s0963548307008553. URL: https://doi.org/10.1017/s0963548307008553.

[35] V. Rödl and J. Skokan. Regularity lemma for k-uniform hypergraphs. *Random Structures Algorithms*, 25(1):1–42, 2004. ISSN: 1042-9832,1098-2418. DOI: 10.1002/rsa.20017. URL: https://doi.org/10.1002/rsa.20017.

[36] N. Sauer. On the density of families of sets. *Journal of Combinatorial Theory, Series A*, 13(1):145–147, 1972. ISSN: 0097-3165. DOI: https://doi.org/10.1016/0097-3165(72)90019-2. URL: https://www.sciencedirect.com/science/article/pii/0097316572900192.

[37] S. Shelah. A combinatorial problem; stability and order for models and theories in infinitary languages. *Pacific Journal of Mathematics*, 41(1):247–261, 1972.

[38] S. Shelah. *Classification theory: and the number of non-isomorphic models*, volume 92. Elsevier, 1990.

[39] A. Shokoufandeh and Y. Zhao. Proof of a tiling conjecture of Komlós. *Random Structures Algorithms*, 23(2):180–205, 2003. ISSN: 1042-9832,1098-2418. DOI: 10.1002/rsa.10091. URL: https://doi.org/10.1002/rsa.10091.

[40] M. Simonovits and V. T. Sós. Ramsey-Turán theory. In volume 229, number 1-3, pages 293–340. 2001. DOI: 10.1016/S0012-365X(00)00214-4. URL: https://doi.org/10.1016/S0012-365X(00)00214-4. Combinatorics, graph theory, algorithms and applications.

[41] M. Skala. Hypergeometric tail inequalities: ending the insanity. *arXiv preprint arXiv:1311.5939*, 2013.

[42] E. Szemerédi. On sets of integers containing k elements in arithmetic progression. eng. *Acta Arithmetica*, 27(1):199–245, 1975. URL: http://eudml.org/doc/205339.

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

[44] T. Tao. A variant of the hypergraph removal lemma. *J. Combin. Theory Ser. A*, 113(7):1257–1280, 2006. ISSN: 0097-3165,1096-0899. DOI: 10.1016/j.jcta.2005.11.006. URL: https://doi.org/10.1016/j.jcta.2005.11.006.

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

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

[47] J. Wolf. Private communication.

[48] Y. Zhao. *Graph Theory and Additive Combinatorics: Exploring Structure and Randomness*. Cambridge University Press, 2023.

A. Other proofs

For completeness, here we leave secondary proofs we skipped in the thesis.

Proof of Corollary 3.12. 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 π 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 $\overline{B}_{A,b} = B^-_{A,b} = A \setminus B^+_{A,b} = A \setminus B^+_{A,b'} = B^-_{A,b'} = \overline{B}_{A,b'}$.

which proves that the map is well-defined. The map π 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 π 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 $\overline{B}_{A,b}$. $\square$

Proof of Lemma 4.7. Notice that, by the average condition of the pair (A, B):

- there are at most $f(|A|)$ vertices of A (hence in $A' \subseteq A$), say S, which are exceptional with respect to B, so the number of edges $(a,b) \in S \times B'$ which are exceptional is at most $|S| \cdot |B'|$, and
- for each $a \in A$ (hence in $A' \subseteq A$) not in S, 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. Thus, we have at most $(a,b) \in (A' \setminus S) \times B'$ is at most $(|A'| - |S|)g(|B|)$.

The overall worse case in this scenario is when S is maximum ($|S| = f(|A|)$), and thus we have at most

$f(|A|)|B'| + (|A'| - f(|A|))g(|B|)$ exceptional edges in $A' \times B'$, as $|B'| \geq g(|B|)$. 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 finishes the proof. □

Proof of Claim 4.24. We first note that for all natural number $n \geq 1$, and real values $x \geq 1$ and $\epsilon < 1$, we have that:

$$(x+n)^{\epsilon} \leq x^{\epsilon} + n \tag{14}$$

and

$$\lfloor x+n \rfloor \leq \lfloor x \rfloor + n. \tag{15}$$

We now prove the statement by induction on k. If $k = 2$,

$$f_{\epsilon_1\epsilon_2} = \lfloor x^{\epsilon_1\epsilon_2} \rfloor \leq \lfloor \lfloor x^{\epsilon_1} + 1 \rfloor^{\epsilon_2} \rfloor \leq \lfloor \lfloor x^{\epsilon_1} \rfloor^{\epsilon_2} \rfloor + 1,$$

where the second inequality uses (14) and (15). If $k > 2$,

$$\begin{aligned} f_{\epsilon_1\epsilon_2\ldots\epsilon_k}(x) &\leq f_{\epsilon_1\epsilon_2} \circ f_{\epsilon_3} \circ \cdots \circ f_{\epsilon_k}(x) + k - 2 \\ &= f_{\epsilon_1\epsilon_2}(f_{\epsilon_3} \circ \cdots \circ f_{\epsilon_k}(x)) + k - 2 \\ &\leq f_{\epsilon_1} \circ f_{\epsilon_2}(f_{\epsilon_3} \circ \cdots \circ f_{\epsilon_k}(x)) + k - 1, \end{aligned}$$

where the first inequality uses I.H. for $k-1$, and the second inequality uses I.H. for 2. This proves the statement. □

Proof of Lemma 5.8. 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_η is an ζ-good set witnessing that A_η 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_η is m_k, so by IH none of the sets A_η is $(\frac{m_{k+1}}{m_k}, \zeta)$-excellent. Then, B_η in 2 is well-defined. Also, by ζ-goodness of B_η, $t(a, B_\eta)$ in 3 is well-defined. Then, since B_η is witnessing the non-$(\frac{m_{k+1}}{m_k}, \zeta)$-excellence of A_η, 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_ν that do not relate with a_η in the expected way. By ζ-goodness of B_ν, $|U_{\nu,\eta}| < \zeta |B_\nu|$, 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 $\nu \frown \langle i \rangle \triangleleft \eta$, $a_\eta R b_\nu \equiv i$, which follows 3. This contradicts Definition 3.14 of tree bound k_{**}. $\square$

B. Main changes

This section of the appendix is dedicated at showing the main changes this thesis applies to the original results of [28] and [27].

- Definition 2.3 of the k-order property in [28] does not specify adjacency (or not) of vertices with the same index.
- In order for the arguments of Section 4 to work, most results require that the function f (of the f-indivisibility) satisfies $x \geq f(x)$, instead of the *non-decreasing* condition given in Definition 4.2 in [28], which is redundant.
- In order for the average condition to be satisfied, and thus being able to apply Claim 4.8 in the proof of Claim 4.10 in [28], something like the extra condition provided by Remark 4.9 needs to be added to the claim statement.
- Second to last inequality in the equation of P_1 at page 1569 of [28] is actually opposite (the $<$ should be a $>$). The same occurs, with last inequality of P_2 equation at the same page. This breaks the proof's argument, requiring extra conditions and some (non-trivial) changes in the argument. The most important change in the result is the extra condition $m_0 \geq n^{\epsilon}$ in Lemma 4.15, which strongly reduces the interval of possible choices of parts size in the result, and needs to be carried until the end of the subsection.
- Condition $m_{**} > k_{**}$, which is persistent in results of Section 4 in [28] can be relaxed into $m_{**} \geq 1$.
- Proof of Theorem 4.16 in [28] is unclear, even more when previous points are noted. Theorem 4.19 provides a complete proof of a weaker (but coherent) version of the same result.
- Theorem 4.23 proof construction first finds an ϵ-indivisible set, and then applies Claim 4.21 to find a c-indivisible set. But Claim 4.21 itself does not require an ϵ-indivisible set as input, as it is constructed in its own proof. Noticing this allows to fully rewrite the theorem for a stronger (and more interpretable) result (Theorem 4.26).

To do...

- Non-monotonicity

Section 3 argument does not work because...

Also, we note that...

C. Excellence is not monotonic.

Here give more details on the counterexample to the monotonicity of the excellence property given in Figure 4. We see that this example is in fact the smallest bipartite graph of a family of counterexamples. Each element of the family can be described by the following adjacency matrix, defined by blocks:

$$G_r = \left[\begin{array}{c|c} 0 & H_r \\ \hline H_r^T & 0 \end{array}\right], \text{ with } \quad H_r = \left[\begin{array}{c|c} 0 & \mathbb{1}_r - \mathbb{I}_r \\ \hline \mathbb{J}_r & \mathbb{I}_r \end{array}\right]$$

where $\mathbb{1}_r$ is the $r \times r$ matrix of all 1's, $\mathbb{I}_r$ is the $r \times r$ diagonal matrix, and $\mathbb{J}_r$ is the $r \times r$ anti-diagonal matrix. Also, we use H^T to refer to the transpose of H.

By calling A the set of the first (as indices) $2r$ vertices of G_r, and B the last $2r$, we have the desired counterexample: $A \subseteq G$ is $\frac{1}{2r-1}$-excellent, but B witnesses that A is not $\frac{1}{2r-1}$-excellent. The example in Figure 4 shows G_r for $r = 3$.

A sufficient proof of this is a simple exhaustive check, and code for this precise purpose is provided with all the material of this thesis in a GitHub repository[15]. There are two main relevant scripts in the repository. One allows to check whether G_r for a given value of r is in fact $\frac{1}{2r-1}$-excellent and not $\frac{1}{2r-1}$-excellent. The other allows for an exhaustive search of *possible* counterexamples under some given parameters, which is how this counterexamples were found in the first place. Read the documentation for more information on how to run the code.

[15]See `https://github.com/SeverinoDaDalt/tfm_severino_da_dalt/`