Pontifícia Universidade Católica
do Rio de Janeiro

Felipe de Oliveira

A Characterization of Testable Graph Properties in the dense graph model

Dissertação de Mestrado

Thesis presented to the Programa de Pós–graduação em Matemática, do Departamento de Matemática da PUC-Rio in partial fulfillment of the requirements for the degree of Mestre em Matemática.

Advisor: Prof. Simon Griffiths

Rio de Janeiro
Abril 2023

Pontifícia Universidade Católica
do Rio de Janeiro

Felipe de Oliveira

A Characterization of Testable Graph Properties in the dense graph model

Thesis presented to the Programa de Pós–graduação em Matemática da PUC-Rio in partial fulfillment of the requirements for the degree of Mestre em Matemática. Approved by the Examination Committee:

Prof. Simon Griffiths
Advisor
Departamento de Matemática – PUC-Rio

Prof. Guilherme Oliveira Mota
Universidade de São Paulo – USP

Prof. Taísa Lopes Martins
Universidade Federal Fluminense – UFF

Prof. Rangel Baldasso
Departamento de Matemática – PUC-Rio

Rio de Janeiro, Abril 28th, 2023

Felipe de Oliveira

Majored in mathematics by Pontifícia Universidade Católica do Rio de Janeiro (Rio de Janeiro, Brazil).

Bibliographic data

de Oliveira, Felipe

A Characterization of Testable Graph Properties in the dense graph model / Felipe de Oliveira; advisor: Simon Griffiths. – 2023.

59 f: il. color. ; 30 cm

Dissertação (mestrado) - Pontifícia Universidade Católica do Rio de Janeiro, Departamento de Matemática, 2023.

Inclui bibliografia

1. Matemática – Teses. 2. Algoritmos de aproximação. 3. algoritmos aleatorizados. 4. algoritmos em grafos. 5. testagem de propriedades. 6. O lema de regularidade de Szemerédi. I. Griffiths, Simon. II. Pontifícia Universidade Católica do Rio de Janeiro. Departamento de Matemática. III. Título.

CDD: 510

To my wife, parents and relatives for their support and encouragement.

Acknowledgments

I would like to offer my special thanks to my wife Suemy Inagaki for believing in me when I decided to pursue a career in mathematics. Without her partnership and encouragement I would not have made it to this point.

I would also like to thank all my friends and family who in one way or another encouraged or helped me. In particular, I would like to thank my parents, Adriana and Glaucio, for encouraging me to study since I was a child.

I would like to thank Professor Simon Griffiths for encouragement and interesting conversations about the topics in this thesis. As well as trusting me to work with him and bringing back my passion for math.

I also would like to thank Professor Sinesio Pesco for all his patience and advices during the years we worked together during my undergraduate studies. When I thought about leaving a career in mathematics and following another path, his advice was essential to bring me back.

This study was financed in part by the Coordenação de Aperfeiçoamento de Pessoal de Nível Superior - Brasil (CAPES) - Finance Code 001

I also thank CNPq for their financial support.

Abstract

de Oliveira, Felipe; Griffiths, Simon (Advisor). **A Characterization of Testable Graph Properties in the dense graph model**. Rio de Janeiro, 2023. 59p. Dissertação de Mestrado – Departamento de Matemática, Pontifícia Universidade Católica do Rio de Janeiro.

We consider, in this thesis, the question of determining if a graph has a property $\mathcal{P}$ such as "G is triangle-free" or "G is 4-colorable". In particular, we consider for which properties $\mathcal{P}$ there exists a random algorithm with constant error probabilities that accept graphs that satisfy $\mathcal{P}$ and reject graphs that are ϵ-far from any graph that satisfies it. If, in addition, the algorithm has complexity independent of the size of the graph, the property is called testable. We will discuss the results of Alon, Fischer, Newman, and Shapira that obtained a combinatorial characterization of testable graph properties, solving an open problem raised in 1996. This characterization informally says that a graph property $\mathcal{P}$ is testable if and only if testing $\mathcal{P}$ can be reduced to testing the property of satisfying one of finitely many Szemerédi-partitions.

Keywords

Approximation Algorithms; Randomized Algorithm; Graph Algorithms; Property Testing; Szemerédi's regularity lemma.

Resumo

de Oliveira, Felipe; Griffiths, Simon. **Uma caracterização de propriedades testáveis no modelo de grafos densos**. Rio de Janeiro, 2023. 59p. Dissertação de Mestrado – Departamento de Matemática, Pontifícia Universidade Católica do Rio de Janeiro.

Consideramos, nesta dissertação, a questão de determinar se um grafo tem uma propriedade $\mathcal{P}$, tal como "G é livre de triângulos" ou "G é 4-colorível". Em particular, consideramos para quais propriedades $\mathcal{P}$ existe um algoritmo aleatório com probabilidades de erro constantes que aceita grafos que satisfazem $\mathcal{P}$ e rejeita grafos que são ϵ-longe de qualquer grafo que o satisfaça. Se, além disso, o algoritmo tiver complexidade independente do tamanho do grafo, a propriedade é dita testável. Discutiremos os resultados de Alon, Fischer, Newman e Shapira que obtiveram uma caracterização combinatória de propriedades testáveis de grafos, resolvendo um problema em aberto levantado em 1996. Essa caracterização diz informalmente que uma propriedade $\mathcal{P}$ de um grafo é testável se e somente se testar $\mathcal{P}$ pode ser reduzido a testar a propriedade de satisfazer uma das finitas partições Szemerédi.

Palavras-chave

Algoritmos de aproximação; algoritmos aleatorizados; algoritmos em grafos; testagem de propriedades; O lema de regularidade de Szemerédi.

Table of contents

List of algorithms

Hakuna Matata

Timon And Pumbaa, *The Lion King.*

1
Introduction

Imagine a poll where n voters must vote **Yes** or **No**. Let us encode the votes in an n-bit array A, where $A[i]$ is 1 if the ith voter voted **Yes** and 0 otherwise. The question we would like to address is: is it possible to decide whether **Yes** won the election by looking at just a small part of the array? If we call MAJ the set of arrays where **Yes** is the majority, we are wondering if it is possible to quickly decide whether $A \in \text{MAJ}$. If $A \in \text{MAJ}$ we say that A has the majority property.

0	**1**	**1**	**0**	**1**	**0**	**1**	**1**

Figure 1.1: Array in MAJ

0	**1**	**1**	**0**	**1**	**0**	**1**	**0**

Figure 1.2: Array not in MAJ

The natural answer is no, since in the worst case, the last voter may be able to decide the result.

However, in the age of Big Data, processing all the data can be a very costly task. It is natural then to try to relax the problem to enable a quick analysis. One way to do this is to note that the example in Figure 1.2 is not in MAJ, but it is very close to MAJ. We can turn it into a vector in MAJ, as the one in Figure 1.1, by editing just one entry. We'll change our question then to "Is it possible to quickly differentiate between arrays in MAJ and arrays that are far from MAJ?"

To answer this question, we first need to make it mathematically precise. Let us use the normalized Hamming distance to measure the distance between two arrays x and z:

$$\delta(x,z) = \begin{cases} \mid \{i \in [\mid x \mid] : x_i \neq z_i\} \mid / \mid x \mid & \text{if } \mid x \mid = \mid z \mid \\ \infty & \text{otherwise} \end{cases}, \tag{1-1}$$

where $\mid x \mid :=$ length of array x and $[\mid x \mid] := \{1, 2, 3 \ldots, \mid x \mid\}$.

Let C be a set, we define the distance from x to C as $\delta_C(x) := min_{z \in C}\{\delta(x,z)\}$. Clearly, $\delta_C(x) = 0 \iff x \in C$. We say that x is ϵ-far from C if $\delta_C(x) > \epsilon$, otherwise x is ϵ-close to C.

We will present a random algorithm in this section that will take x and ϵ as inputs and has three important properties: firstly, it returns the correct output with very high probability when given a $x \in MAJ$, secondly, it returns the correct ouput with very high probability when given a x ϵ-far from MAJ, and finally, it has complexity that only depends on ϵ. It is formally stated in Proposition 1.1.

Proposition 1.1 *Algorithm 1 satisfies the following 3 properties:*

1. *For every* $x \in MAJ$, $\boldsymbol{Pr}[Output = MAJ] \geq 99.99\%$*;*

2. *For each* x *such that* $\delta_{MAJ}(x) > \epsilon$, $\boldsymbol{Pr}[Output = FAR] \geq 99.99\%$*;*

3. *The complexity of the algorithm is independent of* $|x|$.

Also, the algorithm queries only $O(1/\epsilon^2)$ *entries.*

Notes:

– We are not making any guarantees when the input x is such that $0 < \delta_{MAJ}(x) < \epsilon$;

– The probabilities in the proposition are in relation to the randonmess of the algorithm. For example, the first property, $\Pr[Output = MAJ] \geq 99.99\%$ does not mean that it holds for 99.99% of the inputs. For every input $x \in$ MAJ the output is MAJ with probability at least 99.99%.

Algorithm 1: Random tester for majority property

```
Input : string x, proximity parameter ε
Output: MAJ if the algorithm says that x ∈ MAJ, FAR otherwise
Set m = 200/ε²
Let (i_1, ..., i_m) be a string of m uniformly and independently chosen
   indices in {1, ..., |x|}
if Σ_{j∈[m]} x_{i_j}/m > (1 − ε)/2 then
   Return MAJ
else
   Return FAR
```

To prove Proposition 1.1 we need the following auxiliary result.

Lemma 1.2

$$\boldsymbol{Pr}_{i_1,\ldots,i_m \in [|x|]}\left[\left|\frac{\sum_{j\in[m]} x_{i_j}}{m} - \frac{\sum_{i=1}^{|x|} x_i}{|\, x\,|}\right| > \epsilon/2\right] \geq 10^{-4}.$$

Where $i_1, \ldots, i_m \in [|\, x\,|]$ *means that* $i_1, \ldots, i_m$ *are independent and uniformly distributed over* $[|\, x\,|]$.

Proof Let $X_1, \dots, X_m$ be random variables associated with $x_{i_1}, \dots, x_{i_m}$. They are independent, identically distributed and have an image in $[0, 1]$. Furthermore, we have $\mathbf{E}[X_1] = \frac{\sum_{i=1}^{|x|} x_i}{|x|}$. Therefore, by a Chernoff inequality 10.2, we have:

$$\mathbf{Pr}\left[\left|\frac{\sum_{i\in[m]} X_i}{m} - \frac{\sum_{i=1}^{|x|} x_i}{\mid x \mid}\right| > \epsilon/2\right] < 2e^{-\epsilon^2 m/16} = 2e^{-200/16} < 10^{-4}.$$

■

With this lemma in mind, we may now analize the listed properties and prove Proposition 1.1.

Proof of Proposition 1.1

Property 1: If $x \in$ MAJ, then $\sum_{i\in[|x|]} x_i >\mid x \mid /2$. Applying Lemma 1.2, we have $\frac{\sum_{j\in[m]} x_{i_j}}{m} \in (1/2 - \epsilon/2, 1]$ with probability greater than 99.99%. In this case, we have that the probability of the algorithm returning MAJ is greater than 99.99%.

Property 2: If x is ϵ-far from MAJ, we have $\sum_{i\in[|x|]} x_i \leq (1/2 - \epsilon) \mid x \mid$. Applying Lemma 1.2, we have that $\frac{\sum_{j\in[m]} x_{i_j}}{m} \in [0, \frac{1}{2} - \frac{\epsilon}{2})$ with probability greater than 99.99%. In this case, we have that the probability of the algorithm returning FAR is greater than 99.99%.

Property 3: We perform m times the operation of randomly selecting an index between 1 and $\mid x \mid$, where we assume that each of these operations has a unit cost. All other operations have $O(m)$ complexity. The total complexity is therefore $O(m) = O(1/\epsilon^2)$, which is independent of $\mid x \mid$. ■

We are then able to differentiate between arrays in MAJ and arrays that are ϵ-far of MAJ by reading only $O(1/\epsilon^2)$ entries, which is indendepent of the size of the array! Would it be possible to do the same with a deterministic algorithm? The answer is no, as stated in Proposition 1.3 [1].

Proposition 1.3 *Any deterministic algorithm that distinguishes between inputs in* MAJ *and inputs that are 0.5-far from* MAJ *must make at least n/2 queries, where n is the length of the input.*

Algorithm 1 is called a tester for the property of being in MAJ. This kind of algorithm began to be systematically studied by Goldreich, Goldwasser and Ron [2].

For some graph problems, it is possible to obtain efficient testers. Graphs such that it is possible to color all of their vertices with 3 colors so that no two adjacent vertices connected with the same color are called 3-colorable. Given a graph, decide whether it is 3-colorable is an NP-complete problem, as discussed in the paper by Garey, Johnson, and Stockmeyer [3]. However, it is possible to test it in constant time using a randomized tester that differentiate between 3-colorable graphs and graphs that are ϵ-far from the 3-colorable graphs set.

It is not always possible, however, to develop a random algorithm to decide whether a graph has a certain property. Some examples of properties of this type can be found in Fischer [4]. Our main objective in this dissertation is to reproduce the result of Alon, Fischer, Newman and Shapira [5] which establishes a combinatorial characterization of the testable graph properties.

1.1 Overview of the dissertation

The dissertation is organized as follows: In Chapter 2 we introduce some basic notation about graphs and in Chapter 3 we make formal definitions about Graph Property Testing. This chapter states most of the definitions and notation used during the thesis and is essential, as definitions and notation vary in the literature.

In Chapter 4 we introduce the regularity lemma, a key result proved by Szemerédi [6], which is widely used in this thesis. In addition, we make the exact formulation of the main result using the definitions of previous chapters. In Chapter 5, we state some auxiliary results that will be necessary for the proof of the main result.

In Chapter 6 we state results that use the regularity lemma to estimate the number of induced subgraphs H in a graph G that satisfy certain regularity conditions. In Chapter 7 we prove the first direction of the main result. In Chapter 8 we state results about how regularity is passed to induced graphs. All these results appear in [5], which is an extension of the results of [7].

In Chapter 9 we prove the second part of the main theorem using the tools defined in the previous chapters.

2
Background on Graphs

In this section we define the notation and definitions about graphs used in this dissertation. We assume the reader is familiar with these concepts.

A simple graph $G = (V, E)$ consists of a finite set V of vertices and a finite set of edges E. Each edge is a set of 2 vertices u, v such that $u, v \in V$ and $u \neq v$. We call u and v endpoints of the edge. In this way, we are not considering parallel edges or self-loops.

In our case, the order of the n vertices is not important, so we can arbitrarily order them and think of V as $\{1, \ldots, n\} = [n]$.

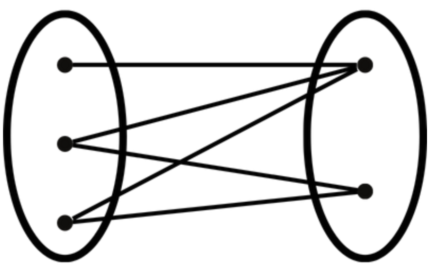
If there is an edge $\{u, v\} \in E$, we say that u and v are adjacent in the graph, or that u and v are neighbors.

For a vertex $v \in V$ of a graph $G = (V, E)$, $\Gamma_G(v)$ denotes the set of neighbors of v:

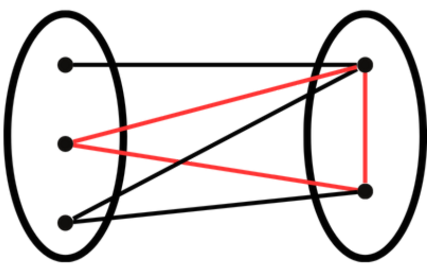
$$\Gamma_G(v) \coloneqq \{w \in V : \{v, w\} \in E\}. \tag{2-1}$$

A subgraph of the graph $G = (V, E)$ is any graph $G' = (V', E')$ such that $V' \subset V$ and $V' \subset E$. The subgraph induced by $V' \subset V$, denoted by $G[V']$ is the graph $G' = (V', E \cap \binom{V'}{2})$, where $\binom{V'}{2} \coloneqq \{\{u, v\}; u, v \in V'\}$. We say that V' spans $G[V']$.

Let H be a graph, we say that a graph G is H-free if it does not have H as a subgraph. For example, G is triangle-free if there is no subgraph of G that is a triangle. In Figure 2.1a we have an example of a triangle-free graph and in Figure 2.1b we have an example of a graph with one of its triangles highlighted.



(a) Example of Triangle-free graph



(b) Graph with a triangle in red

Figure 2.1

The adjacency matrix of graph G is a matrix that encondes the adjacencies between vertices of G. Given an arbitrary order of vertices, let u and v be the i-th and j-th vertices. We will denote by $A_G(u,v)$ the entry $A_{i,j}$ of the matrix. Using this notation, we have:

$$A_G(u,v) = \begin{cases} 1 & \text{if } \{u,v\} \in E, \\ 0 & \text{otherwise.} \end{cases}$$

We say $G = (V, E_G)$ and $H = (V, E_H)$ are *isomorphic* if there is a bijection $\pi : V \to V$ such that $\pi(E_G) = E_H$, where the image of a edge set under π is defined as:

$$\pi(E) \coloneqq \{\{\pi(u), \pi(v)\} : \{u,v\} \in E\} \ .$$

We define also $\pi(G) \coloneqq (\pi(V_G), \pi(E_G))$. The idea is that an isomorphism between graphs is a bijection between the vertices that preserves the neighborhood.

We define an automorphism of a graph G as being an isomorphism between G and G. We will denote the number of automorphisms of a graph G by $Aut(G)$.

In Figure 2.2 we can see an example of an automorphism π, where $\pi(v_1) = v_4, \pi(v_2) = v_1, \pi(v_3) = v_2, \pi(v_4) = v_3$. We can understand π as a 90° clockwise rotation.

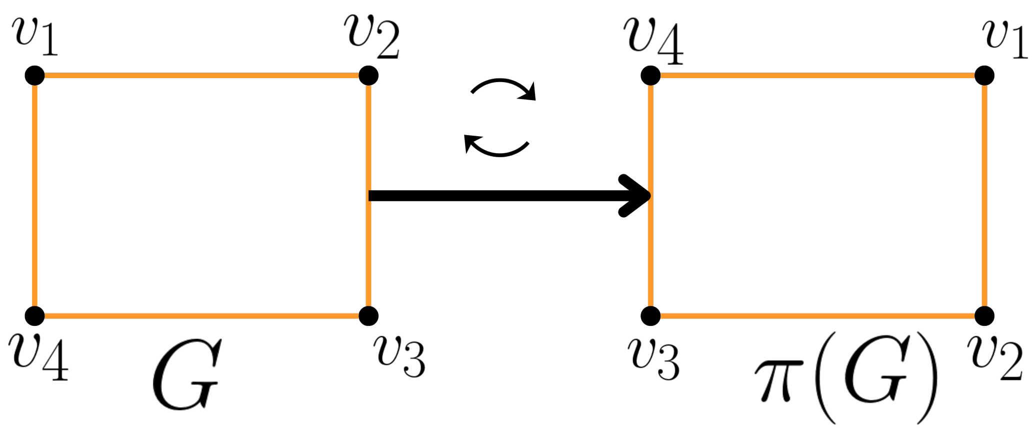


Figure 2.2: Example of one of the automorphisms of G

PUC-Rio - Certificação Digital Nº 2112392/CA

3
Background on Graph Property Testing

In this section we will introduce most of the notation and definitions used in this thesis.

As some propositions introduce constants, we will use the notation $\alpha_{i.j}$ to denote the constant α introduced in Proposition/Corollary/Lemma/Theorem $i.j$.

Let $G = (V, E_G)$ and $H = (V, E_H)$ be graphs. We will use the normalized symmetric difference to measure the distance between the graphs:

$$\delta(G, H) \coloneqq |E_G \Delta E_H|/n^2, \tag{3-1}$$

where $|V| = n$.

A graph property is defined as a set of graphs that is closed under isomorphism, as in Definition 3.1.

Definition 3.1 (Graph Property) *A* graph property *is a set $\mathcal{P}$ such that if $G \in \mathcal{P}$, then for every bijection $\pi : V_G \to V_G, \pi(G) \in \mathcal{P}$.*

The distance between a graph G on n vertices and a property $\mathcal{P}$ is defined as

$$\delta(G, \mathcal{P}) \coloneqq min_{H \in P_n} \delta(G, H), \tag{3-2}$$

where $P_n \coloneqq \{H = (V, E) \in \mathcal{P} : |V| = n\}$

We say that a graph G on n vertices is ϵ-far from $\mathcal{P}$ if $\delta(G, \mathcal{P}) > \epsilon$, or equivalently, if for every $G' \in \mathcal{P}_n, |\ E_G \Delta E_{G'}\ |> \epsilon n^2$. Otherwise, it is $\epsilon-$close from $\mathcal{P}$.

3.1
Complexity

To measure the complexity of an algorithm, it is common to calculate the time complexity or running time [8]. However, in Property Testing, the focus is on the *query complexity*, which is the number of times the input is accessed. In our case, the input is a graph and we need to somehow access it. The model that will be used by us is the dense graph model [2], where we have access to the adjacency matrix A_G of the graph. So, the query complexity measures the number of times we need to ask if a pair of nodes (i, j) is adjacent to each other.

3.2 Property Tester and Testable Properties

We can now define what is a Property Tester:

Definition 3.2 (Graph Property Tester) *Let $G = (V, E)$ be a graph on n vertices and $\mathcal{P} = \bigcup_{n\in\mathbb{N}} P_n$ a graph property. A tester for $\mathcal{P}$ is an algorithm that receives a parameter $\epsilon \in (0, 1)$, n and G as inputs and the following holds:*

1. *For any input G that satisfies $\mathcal{P}$, the tester accepts G with probability at least $2/3$;*

2. *For any input G that is ϵ-far from $\mathcal{P}$, the tester rejects G with probability at least $2/3$.*

Where we say that a tester for $\mathcal{P}$ accepts G if it returns that $G \in \mathcal{P}$ and rejects G if it returns that G is ϵ-far from $\mathcal{P}$.

In case the probability of the first condition is equal to 1, we say that the tester is one-sided error and otherwise it is two-sided error.

Note that we are dividing the possible inputs G into three types: The ones that satisfy $\mathcal{P}$, the ones that are ϵ-far from $\mathcal{P}$, and the ones that are ϵ-close from $\mathcal{P}$ but do not satisfy $\mathcal{P}$. We make no guarantees about the output in the third case.

Definition 3.3 (Testable Property) *A graph property $\mathcal{P}$ is testable if there is a tester that makes a number of edge queries which is bounded by some function $q(\epsilon)$ that is independent of the size of the input.*

Another definition that will be important in demonstrating the main result is the definition of estimable property.

Definition 3.4 ((ϵ, δ)-estimable Property) *Let $\mathcal{P} = \bigcup_{n\in\mathbb{N}} P_n$ be a graph property. We say $\mathcal{P}$ is (ϵ, δ)-estimable if there is a probabilistic algorithm $\mathcal{T}$ such that:*

1. *For every input G that is $(\epsilon - \delta)$-close to $\mathcal{P}$, the algorithm accepts G with probability at least $2/3$;*

2. *For every input G that is ϵ-far from $\mathcal{P}$, the algorithm rejects G with probability at least $2/3$;*

3. *The number of queries of the algorithm is constant (independent of n).*

Definition 3.5 (Estimable Property) *We call a property estimable if it is (ϵ, δ)-estimable for every fixed $\epsilon > 0$ and $\delta > 0$.*

The following result by Fischer and Newman [9] will be important to prove the main result.

Theorem 3.6 *Every testable graph property is estimable.*

3.3 Tester for Biclique

A graph $G = (V, E)$ is a *biclique*, or complete bipartite graph, if there exists a bipartition $V = V_1 \cup V_2$ such that $E = \{\{u, v\} : (u, v) \in V_1 \times V_2\}$. By this definition, a graph without edges is considered a biclique due to the bipartition $V = V \cup \emptyset$. We denote by $\mathcal{B}$ the property of being a biclique. In this section we show that the property of being a biclique is testable with an one-sided error tester.

The idea of the algorithm is simple. Select a vertex arbitrarily. Then repeat the following procedure $2/\epsilon$ times: select two more vertices at random. If the subgraph induced by them isn't a biclique, then return that the original graph doesn't satisfy the property of being a biclique. If in all procedures the induced subgraph is a biclique, return that the original graph satisfies the property of being a biclique. This algorithm is formally stated as Algorithm 2

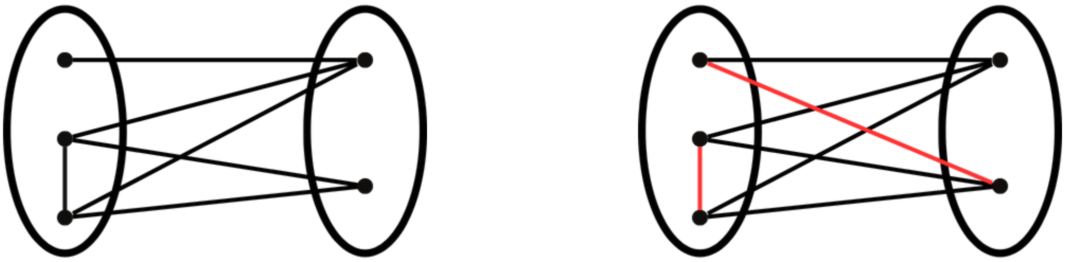
Proposition 3.7 *Algorithm 2 is a one sided error tester for $\mathcal{B}$ that makes $O(1/\epsilon)$ queries.*

Algorithm 2: Random tester for $\mathcal{B}$ property

Input : integer n, proximity parameter ϵ and access to G on n vertices
Output: YES if the algorithm says that $x \in \mathcal{B}$, FAR otherwise
Select the first vertice u of G
for $i = 1$ *to* $2/\epsilon$ **do**
    Select more 2 vertices v_i, w_i uniformly at random
    Query the 3 pairs of edges $\{u, v_i\}, \{u, w_i\}, \{v_i, w_i\}$
    if *$G[\{u, v_i, w_i\}]$ isn't a biclique* **then**
        Return FAR
Return YES

Note: As explained in Chapter 2, the order of the vertices is arbitrary. Thus, u is an arbitrary vertex of the graph.

Let (u, v) be a pair of vertices of G. We say that (u, v) is a *violating* pair with respect to a bipartition V_1, V_2 if the edge $\{u, v\}$ does not satisfy the property expected by an edge of a biclique. A violating pair can either be a pair of vertices u, v belonging to either V_1 or V_2 that have an inner edge or a pair of vertices u, v one belonging to V_1 and the other a V_2 that do not have an edge. In Figure 3.1a we have an example of a graph that has 2 violating pairs and in Figure 3.1b we have the same graph but with the edges related to the violating pairs painted in red.



(a) Example of Graph with 2 violating pairs

(b) Edges between the violating pairs in red

Figure 3.1

With this definition, we are ready to prove Proposition 3.7.

Proof of Proposition 3.7 To verify if $G[\{u, v_i, w_i\}]$ is a biclique, we make only 3 queries. The total number of queries is then bounded by $6/\epsilon = O(1/\epsilon)$.

Suppose $G = (V, E) \in \mathcal{B}$. In this case, because it is a biclique, any induced subgraph is a biclique, so the algorithm always returns YES.

Assume that $G = (V, E)$ is $\epsilon-$far from $\mathcal{B}$, this means that given any bipartition $V = V_1 \cup V_2$, there are more than ϵn^2 violating pairs with respect to this bipartition. We have that u defines a bipartition of V, where $V_1 = \Gamma(u)$ and $V_2 = V \setminus \Gamma(u)$. Note that $u \in V_2$. From the above observation, there are more than ϵn^2 violating pairs with respect to this bipartition. Thus, the probability that v_i and w_i are chosen such that (v_i, w_i) forms a violating pair is at least $\epsilon n^2/n^2 = \epsilon$.

If A_i is the event that none of the edges $(u, v_i), (v_i, w_i), (w_i, u)$ is violating for all i and B_i the event where the edge (v_i, w_i) is not violating for all i. Clearly $A_i \subset B_i$. We want to show that $\mathbf{Pr}[A_i] \leq 1/3$. However, by monotonicity, it is enough to show that $\mathbf{Pr}[B_i] \leq 1/3$. As discussed above, we have:

$$\mathbf{Pr}[B_i] \leq (1-\epsilon)^{\frac{2}{\epsilon}} \leq e^{-2} \leq 1/3.$$

Where the last inequality is stated in the Appendix (Lemma 10.3). Thus, the probability that some violating edge is chosen is at least 2/3.

■

Many graph property testers follow the same pattern: Given $G = (V, E)$, select a set V' of q vertices and accept G if $G[V']$ satisfies a given property, which motivates Definition 3.8.

Definition 3.8 (Canonical Tester) *A tester T for a property $\mathcal{P}$ is canonical if the following holds: There exists a function $s : \mathbb{Z}_+^* \times (0,1) \to \mathbb{Z}_+^*$ and a property $\mathcal{P}'$ such that given $n \in \mathbb{Z}_+^*$, $0 < \epsilon < 1$ and access to $G = (V, E)$ on* *n vertices, T uniformly selects a set of $s(n,\epsilon)$ vertices of G and accepts if and only if $G[s]$ has a property $\mathcal{P}'$.*

Note that $\mathcal{P}$ and $\mathcal{P}'$ do not have to be the same property. However, for many natural properties, such as *k-colorability*, $\mathcal{P} = \mathcal{P}'$ [10]. Also, note that the query complexity is $\binom{s(n,\epsilon)}{2}$.

The following proposition from Goldreich and Trevisan [10] guarantees that, at the cost of squaring the query complexity, we can assume without loss of generality that every tester works as a canonical tester.

Proposition 3.9 *Let $\mathcal{P}$ be any graph property with query complexity $q_{\mathcal{P}}$. Then $\mathcal{P}$ has a canonical tester of query complexity $q'_{\mathcal{P}}(n,\epsilon) = O(q_{\mathcal{P}}(n,\epsilon)^2)$. Furthermore, if $\mathcal{P}$ has a one-sided error tester of query complexity q, then $\mathcal{P}$ has a one sided-error canonical tester with query complexity $\binom{2q}{2}$.*

To prove that some properties are testable we will need a result known as Szemerédi regularity lemma, which will be introduced in Chapter 4.

4
Regularity Lemma and Characterization of Testability

In this chapter, we state the Szemerédi Regularity Lemma, an essential tool for this thesis along with some applications. At the end of the chapter, we enunciate the main result of this thesis, a combinatorial characterization of which properties are testable.

4.1
Regularity Lemma

Let A and B be subsets of vertices of a graph G. We define $e_G(A, B)$ as the number of edges of G between A and B. The edge density of the pair is defined by $d_G(A, B) = e_G(A, B)/|A||B|$. If the graph G is clear from the context, we omit the subscript G.

Definition 4.1 (γ-regular pair) *We say that the pair (A, B) is γ-regular if we have*

$$| \, d(A', B') - d(A, B) \, | \leq \gamma, \qquad (4\text{-}1)$$

for all $A' \subset A$ and $B' \subset B$ with

$$| \, A' \, | \geq \gamma \, | \, A \, | \;\; \textit{and} \;\; | \, B' \, | \geq \gamma \, | \, B \, | \, . \qquad (4\text{-}2)$$

Less formally, for large-enough subsets $A' \subset A$ and $B' \subset B$, the density between A' and B' should be close (within γ) to the density between A and B. This idea is shown in Figure 4.1.

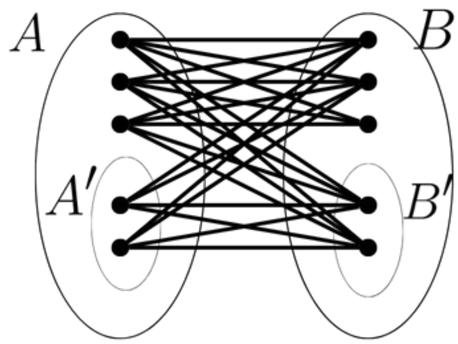


Figure 4.1: Density $d(A, B)$ and $d(A', B')$ should not differ by much in a γ-regular pair.

Note that we can change the inequalities of the condition 4-2 to equalities. This observation will be useful in many proofs. Also note that when $\gamma > 1$, any pair becomes trivially γ-regular, so it's interesting to analyze with $0 < \gamma < 1$.

In Figure 4.2 we have an example of 2/3-regular pair A, B that is not 1/5-regular, where A and B have 15 vertices each. Basically, the example consists of a bipartite graph where 14 of the 15 vertices of A connect with 14 of the 15 vertices of B.

To check this, let's call v_A and v_B the exceptional vertices of A and B. In this case, $d(A, B) = \frac{14 \cdot 14}{15 \cdot 15}$. By choosing $A' \subset A$ and $B' \subset B$ each with 10 vertices, we have 4 possibilities: $v_A \in A'$ and $v_B \in B$ or $v_A \notin A'$ and $v_B \in B$ or $v_A \in A'$ and $v_B \notin B$ or $v_A \notin A'$ and $v_B \notin B$. In any of the 4 cases it is verified that $|d(A', B') - d(A, B)| \leq 2/3$. Note however that (A, B) is not (1/5)-regular, since if we choose $A' \subset A$ and $B' \subset B$ each with 3 vertices such that $v_A \in A'$ and $v_B \in B'$, we have $d(A', B') = 4/9$, which makes $|d(A, B) - d(A', B')| > 1/5$.

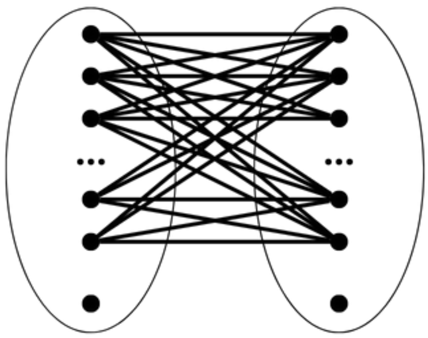


Figure 4.2: Example of a 2/3-regular pair that is not 1/5-regular pair

Let V be the vertices set of a graph G, a partition $V = V_1 \cup V_2 \cdots \cup V_k$ of V into k parwise disjoint sets is called an equipartition if $|| V_i | - | V_j || \leq 1$ for all $\leq i < j \leq k$. The number of partition classes, k, is called the *order* of the partition.

Definition 4.2 (γ-regular equipartition) *We say that an equipartition $V = V_1 \cup V_2 \cdots \cup V_k$ of the vertex set of a graph is γ-regular if at most $\gamma\binom{k}{2}$ pairs (V_i, V_j) are not γ-regular.*

We may now state the Szemerédi Regularity lemma [6]:

Proposition 4.3 (Regularity Lemma) *For any given $\gamma > 0$ and $k_0 \geq 1$, there is a constant $N = N_{4.3}(k_0, \gamma)$ such that any graph G on $n \geq N$ vertices admits a γ-regular equipartition of its vertex set with order k, for some $k_0 \leq k \leq N$.*

The Regularity Lemma says that every graph has an γ-regular partition of relatively small order. In many applications it is enough to know the densities of the bipartite graph that connects the parts V_i and V_j of the partition. If a graph G has a γ-regular partition of order k, we can define an auxiliary graph F on k vertices where each edge $\{i, j\}$ has weight $d(V_i, V_j)$ if (V_i, V_j) is γ-regular. Many times, we can prove properties of G by proving them for the graph F.

For example, if a graph $G = (V, E)$ admits a γ-regular equipartition such that the auxiliary graph F has a triangle with sufficiently large edge weights, we can guarantee that the graph G has many triangles. We formalize this statement in Proposition 4.4 below.

For the proof of this result and other applications of the Regularity Lemma, we recommend the reference [11].

Proposition 4.4 *Let $G = (V, E)$ be a graph. For any density $d > 0$, there exists a regularity parameter $\gamma(d) = d/2$ and a number of triangles $\delta(d) = (1 - d) - d^3/8$ such that if A, B, C are disjoiint subsets of V, each pair γ-regular with density at least d, then G has at least $\delta \cdot |A||B||C|$ distinct triangles with vertices from each of A, B and C.*

We will define a property that will be important for enunciating the main result [5].

Definition 4.5 (Regularity-instance) *A regularity-instance is a pair $R = (F, \gamma)$ where $0 < \gamma \leq 1$, and F is a weighted graph such that $e(F) \geq (1 - \gamma)\binom{k}{2}$ where $k := |V(F)|$ is called the order of the regularity-instance. If $ij \in E(F)$ then we write $d_{i,j}$ for the associated weight. We say that a graph $G = (V, E)$ satisfies the regularity-instance R if there is an equipartition $V = V_1 \cup V_2 \cup \cdots \cup V_k$ such that for all $(i, j) \in E(F)$, the pair (V_i, V_j) is γ-regular and there is $\lfloor d_{i,j}|V_i||V_j|\rfloor$ edges between V_i and V_j. The complexity of a regularity-instance is defined by* $\max(k, 1/\gamma)$.

Using this definition, the Regularity Lemma (Proposition 4.3) implies that for any $0 < \gamma \leq 1$, we have that every graph G satisfies some regularity-instance with parameter γ. Futhermore, the order, k of the regularity-instance is bounded as a function of γ.

Given a regularity-instance $R = (F, \gamma)$, in the next chapters it will be useful to think of F as a weighted complete graph, in this way we will denote by $\overline{F}$ the weighted complete graph obtained by adding edges not present in F with weight 0.

The following result, which we prove in Chapter 9, states that it is relatively easy to test whether a graph satisfies a regularity-instance.

Theorem 4.6 *For any regularity-instance R, the property of satisfying R is testable.*

As mentioned earlier, knowing that a graph G satisfies a certain regularity-instance R allows us to infer properties of G, such as the fact that it has many triangles. The theorem above says that we can make these inferences quickly.

4.2 Triangle-Free tester

The triangle removal lemma can be obtained as a consequence of the Regularity Lemma. This lemma was initially proposed by Ruzsa and Szemerédi [12] , later extended by Erdős, Frankl, and Rödl to the Graph removal lemma [13]. In Proposition 4.7 we state the triangle removal lemma in a different way from the original, using the definitions already introduced. Let us call $\mathcal{T}$ the triangle-free property.

Proposition 4.7 *For any positive $\epsilon < 1$, there exists $\delta_{4.7}(\epsilon)$ such that if $G = (V, E)$ is a graph on n vertices and ϵ-far from $\mathcal{T}$, then it contains at least $\delta_{4.7}(\epsilon)n^3$ triangles.*

We will show how to use it to verify that the property of being triangle-free is testable with one-sided error tester.

Proposition 4.8 *Algorithm 3 is a one-sided error tester for $\mathcal{T}$ that makes a number of queries that does not depend on n.*

Algorithm 3: Random tester for $\mathcal{T}$ property

Input : integer n, proximity parameter ϵ, access to G on n vertices
Output: YES if the algorithm says that $x \in \mathcal{T}$, FAR otherwise
Select randomly and independently $1/\delta_{4.7}(\epsilon)$ triples of vertices.
if *at least one of the triples is a triangle* **then**
  Return FAR
else
  Return YES

Proof To check if each triple is a triangle, we need to make 3 queries, so the total number of queries is at most $3/\delta_{4.7}(\epsilon)$, which only depends on ϵ.

Suppose $G = (V, E) \in \mathcal{T}$. In this case, no triple can form a triangle in the graph, causing the algorithm to always return YES.

Suppose G is ϵ-far from $\mathcal{T}$. By Proposition 4.7, the probability that a chosen triple is a triangle is greater than or equal to $\delta_{4.7}(\epsilon)n^3/\binom{n}{3}$, which in turn is greater than or equal to $6\delta_{4.7}(\epsilon)$. This way, the probability that the algorithm returns YES is bounded by $(1 - 6\delta_{4.7}(\epsilon))^{1/\delta_{4.7}(\epsilon)} \leq e^{-6} \leq 1/3$. ■

This proves that the Triangle-Free property is testable. We will prove this again at the end of the thesis using Corollary 9.4.

4.3 Characterization of Testable Graph Properties.

We can prove that the triangle-free property is testable following a different line of reasoning. This proof will be done rigorously at the end of the thesis, but here we will here sketch it.

Given $G = (V, E)$, by the Regularity Lemma, we can say that G has an equipartition $\mathcal{U} = \{U_1, U_2, \ldots, U_k\}$ γ-regular such that $k \geq 1/\gamma$. However, this does not give us any information about the inner edges of each set U_i nor about the edges between non-γ-regular pairs.

We define G' to be a cleaned-up version of G by making few modifications. As we have few pairs that are not γ-regular, we can make few changes so that such pairs do not have edges connecting them. Furthermore, as we have a large number of sets k, we will have few edges inside each U_i, so we can also make few modifications so that each U_i has no internal edges. Furthermore, we can delete the edges between of sparsely pairs. That is, we can make a few changes to ensure that the equipartition $\mathcal{U}$ has the following properties:

1. There are no internal edges inside each U_i
2. There are only edges between dense γ-regular pairs

After these changes, it follows that, if G' has a triangle, this triangle has vertices in distinct sets U_i, U_j, U_k, such that all pairs are γ-regular and dense. In this way, looking for a triangle in G' is the same thing as looking for a triangle in F, the auxiliary graph of the equipartition. Furthermore, if we find a triangle in F, by Proposition 4.4 we can say that G', and consequently G, has many triangles.

In this way, we can reduce our problem to testing whether, with a few changes, it is possible to transform G into a graph that satisfies a regularity instance R such that F has a triangle. Let us call $\mathcal{R}$ this family of regularity-instances. That is, we can reduce our problem to verifying whether G is close to satisfy some of the regularity-instances of $\mathcal{R}$. As, by Proposition 4.6, it is possible to do each test quickly, it is expected that the property of being triangle-free is testable. We will do this rigorous proof in Chapter 9.

It is possible to reduce other testable properties to the problem of testing a family of regularity-instances, such as the property of being k-colorable and the property of having a "large" cut [1]. The main result of [5] establishes that this is a characterization of which properties are testable. Now we present the formal definitions to state the result.

Definition 4.9 (Regular-Reducible) *A graph property $\mathcal{P}$ is regular-reducible if for any $\delta > 0$, there exists $r = r_{\mathcal{P}}(\delta)$ such that, for any n,* *there is a family $\mathcal{R}$ of at most r regularity-instances, each with complexity at most r, such that for every n-vertex graph G and for every $\epsilon > 0$:*

1. *If G satisfies $\mathcal{P}$, then for some $R \in \mathcal{R}$, G is δ-close to satisfying R*
2. *If G is ϵ-far from satisfying $\mathcal{P}$, then for all $R \in \mathcal{R}$, G is $(\epsilon - \delta)$-far from satisfying R.*

In the above definition ϵ is greater than zero, but can be arbitrarily small. If one takes $\epsilon = 0$, G satisfies $\mathcal{P}$ if and only if it satisfies one of the regularity-instances. That way, instead of testing $\mathcal{P}$, we can test finite regularity-instances. Theorem 4.10 is the main result of [5] that we prove in this thesis.

Theorem 4.10 (Main Result) *A graph property is testable if and only if it is regular-reducible.*

This means that we could have tested the property of being triangle-free in another way, just by showing that it is regular-reducible. An important observation is that this does not necessarily guarantee the tester with minimum query complexity, since the constants involving the Regularity Lemma are large [14]. Informally, a property $\mathcal{P}$ is testable if and only if knowing the auxiliary graph F of a regular partition of a graph G is enough to determine whether G is close or far from satisfying $\mathcal{P}$. We interpret this result by saying that the property of satisfying a certain instance of regularity R is the most difficult property to test, since any testable property can be reduced to testing finite instances of regularity. This result gives a purely combinatorial characterization of which properties are testable, solving a question asked in [2] that was left open for more than 10 years.

5
Enhancing Regularity

When defining a γ-regular pair of density d, many restrictions are imposed on the densities of the pairs of possible sets. The objective of this chapter is to show that if a bipartite graph almost satisfies all these restrictions, then it is close to be a γ-regular pair with density d. This idea is formalized in Proposition 5.1. Both propositions and lemmas were presented in the article [5]. We will use the notation $x = a \pm b$ to say that $x \in [a-b, a+b]$

The following abuse of notation will be common: $x = a \pm b = c \pm d$ means that $x \in [a-b, a+b] \implies x \in [c-d, c+d]$

Proposition 5.1 *The following holds for any* $0 < \delta \leq \gamma \leq 1$*: If* (A, B) *is a* $(\gamma+\delta)$*-regular pair with density* $d \pm \delta$*, where* $|A| = |B| = m \geq m_{5.1}(d, \delta)$*, then it is possible to make at most* $50\delta m^2/\gamma^2$ *modifications and turn* (A, B) *into a* γ*-regular pair with density* d*.*

To prove this proposition we need two lemmas. The first one will allow us to make modifications and adjust the pair density without changing the regularity too much.

Lemma 5.2 *If* (A, B) *is a* $(\gamma+\delta)$*-regular pair satisfying* $d(A, B) = d \pm \delta$*, where* $|A| = |B| = m \geq m_{5.2}(d, \delta)$*, so it is possible to make at most* $2\delta m^2$ *modifications and turn* (A, B) *into a* $(\gamma + 2\delta)$*-regular pair with density* d*.*

The second lemma takes a bipartite graph with density d and returns a bipartite graph with the same density but more regular.

Lemma 5.3 *The following holds for any* $0 < \delta \leq \gamma \leq 1$*. Let* A *and* B *be two sets of vertices of size* $m \geq m_{5.3}(\delta, \gamma)$ *with* $d(A, B) = d$*. Suppose further that for any pair of subsets* $A' \subset A$ *and* $B' \subset B$ *such that* $|A'| = |B'| = \gamma m$ *we have* $d(A', B') = d \pm (\gamma + \delta)$*, then it is possible to make at most* $3\delta m^2/\gamma$ *edge modifications and turn* (A, B) *into a* $\gamma-$*regular pair with density* d*.*

With these two lemmas in hand, we can prove Proposition 5.1.

Proof Proposition 5.1 By Lemma 5.2, we can make at most $2\delta m^2$ edge modifications and turn (A, B) into a $(\gamma+2\delta)$-regular pair with density d. Then, all pairs of subsets $A'' \subset A$ and $B'' \subset B$ of size γm have density at most

$$(d + \gamma + 2\delta)(\gamma + 2\delta)^2 m^2/(\gamma^2 m^2) \leq (d + \gamma + 2\delta)(1 + 8\delta/\gamma) \leq d + \gamma + 14\delta/\gamma.$$

Similarly, the density is at least $d - \delta - 14\delta/\gamma$. Thus, (A, B) has density d and each pair of subsets (A'', B'') of size γm has density $d \pm (\gamma + 14\delta/\gamma)$. Using

Lemma 5.3, we can make at most $42\delta m^2/\gamma^2$ additional modifications to the edges and then turn (A, B) into a γ-regular pair with density d. So we made a total of $42\delta m^2/\gamma^2 + 2\delta m^2 \leq 50\delta m^2/\gamma^2$ modifications and we got the desired result. ■

We can now state a result that will be useful to prove the main result.

Proposition 5.4 *Let $R = (F, \gamma)$ be a regularity-instance of order k. Suppose a graph G has an equipartition $V = V_1 \cup V_2 \cup \cdots \cup V_k$ of order k such that such that the following two conditions hold for all $(i, j) \in F$:*

– $d(V_i, V_j) = d_{i,j} \pm \gamma^2\epsilon/50$ *;*

– (V_i, V_j) *is* $(\gamma + \gamma^2\epsilon/50)$*-regular.*

Then G is ϵ-close to satisfaying R .

Proof We need to show that we can transform G into a graph satisfying R using at most ϵn^2 modifications. By Proposition 5.1, for each pair $(i, j) \in E(G_R)$, we can make at most $50\frac{\gamma^2\epsilon/50}{\gamma^2}(n/k)^2 \leq \epsilon n^2/k^2$ edge modifications to turn (V_i, V_j) into a ϵ-regular pair with density $d_{i,j}$. As we know that the number of pairs is less than or equal to $\binom{k}{2}$, we have that the total number of necessary modifications is less than or equal to ϵn^2, as desired. ■

Now, we prove the auxiliary lemmas.

Proof of Lemma 5.2

Let $G = (V, E)$ be a graph and $A, B \subset V$ subsets of vertices that satisfy the conditions of the statement. We have $d(A, B) = d \pm \delta$, so $d(A, B) = d + p$, where $p \in [-\delta, \delta]$. Let us analyze in cases. Note that when $\gamma + 2\delta \geq 1$, the condition of being $(\gamma + 2\delta)$-regular is satisfied for any pair, so in this case, just make the modifications in order to obtain the density d desired, which can be done with less than δm^2 modifications. In this way, we will assume that $\gamma + 2\delta < 1$.

Case 1: First, let us assume that $0 < p \leq \delta(\gamma + 2\delta)^2$. In that case, just remove any pm^2 edges. Let $H = (V, E_H)$ be the graph after removing the edges. We have, therefore, that $d_H(A, B) = d$. We will now show that the pair (A, B) is $(\gamma + 2\delta)$-regular in H.

Take a pair $(A', B'), A' \subset A, B' \subset B$ both with size $(\gamma + 2\delta)m$. By the observation of Definition 4.1, we just need to show that $d_H(A', B') = d_H(A, B) \pm (\gamma + 2\delta) = d \pm (\gamma + 2\delta)$.

Initially, as the pair (A, B) is $(\gamma \pm \delta)$-regular in G and $A' \subset A, B' \subset B$ with $|A'| = |B'| = (\gamma + 2\delta)m$, we have $d_G(A', B') = d_G(A, B) \pm (\gamma + \delta) = d + p \pm (\gamma + \delta) \implies d + p - \gamma - \delta \leq d_G(A', B') \leq d + p + \gamma + \delta$. However, as we remove $pm^2 \leq \delta(\gamma + 2\delta)m^2$ edges, we have $d - \gamma - 2\delta \leq d_H(A', B') \leq d + \gamma + 2\delta$.

Case 2: Assume now that $p \geq \delta(\gamma+2\delta)^2$. The strategy to show that there is a way to make less than $2\delta m^2$ modifications and obtain the desired conditions will be through the probabilistic method. That is, we will not explicitly find which changes to make, but will do it through a random process and show that the probability of the desired event is greater than 0, which is enough to demonstrate its existence. Our process will occur in two steps, first we remove some edges between A and B randomly. Then some deterministic adjustments are made. We will show that with probability greater than 3/4 we have that (A, B) is a $(\gamma+2\delta)$-regular pair, and also with probability greater than 3/4 we have that $d(A, B) = d$. Thus, by union bound, we obtain that with probability greater than 1/2 both conditions occur.

In the random process, we remove each edge between A and B randomly and independently with probability $p/(d+p)$. Let $H_1 = (V, E_{H_1})$ be the graph after these modifications. Thus, the expected number of edges removed will be:

$$\frac{p}{d+p}(d+p)|A||B| = p|A||B| \leq \delta|A||B|.$$

Thus, we have that the expected value of the density $d_{H_1}(A, B)$ is d. As we assume that $p \geq \delta(\gamma + 2\delta)^2$, we have $d_G(A, B) \geq p \geq \delta(\gamma + 2\delta)^2$. By Lemma 10.1, for m big enough, $m \geq m_{5.2}(\delta, \gamma)$, the probability that $d_{H_1}(A, B)$ deviates from d more than $1/m^{0.5}$ is at most 1/4. In particular, the number of edge modifications is at most $3\delta m^2/2$ with probability at least 3/4.

In the deterministic step, if we modify at most $m^{1.5}$ edges arbitrarily we will be sure that $d_{H_2}(A, B) = d$, where $H_2 = (V, E_{H_2})$ is the modified graph after these second step. The total number of modifications is then at most $3\delta m^2/2 + m^{1.5} \leq 2\delta m^2$ for m large enough. So far, with probability greater than or equal to 3/4, $d_{H_2}(A, B) = d$. It remains to show that with probability greater than 3/4, (A, B) is $(\gamma + 2\delta)$-regular in H_2.

As we initially assumed that (A, B) was a $(\gamma + \delta)$-regular pair in G, we have that $d_G(A', B') = d + p \pm (\delta + \gamma)$ for any pair of subsets $A' \subset A$ and $B' \subset B$ both with size $(\gamma + 2\delta)m$. As in the random process we remove each edge with probability $p/(p + d)$, the expected value of $d_{H_1}(A', B')$ is between:

$$(d + p + \delta + \gamma)\left(1 - \frac{p}{d+p}\right) \leq d + \gamma + \delta,$$

and

$$(d + p - \gamma - \delta)\left(1 - \frac{p}{d+p}\right) \geq d - \delta - \gamma.$$

That is, $\mathbb{E}[d_{H_1}(A', B')] = d \pm (\gamma + \delta)$ for A', B' of size $(\gamma + 2\delta)m$. So we want to show that with probability greater than 3/4, $d(A', B')$ does not deviate more than δ from its expected value. Assume first that $d_G(A', B') \leq \delta/2$ originally. We then have that $d_{H_1}(A', B')$ can change at most $\delta/2$ with relation to $d_G(A', B')$. Thus, in this case, $d_{H_1}(A', B')$ can deviate from its expectation by at most $\delta/2$. Furthermore, by adding or removing $m^{1.5}$ edges to (A, B) in the deterministic step, we can change $d_{H_2}(A', B')$ by at most $(\gamma+2\delta)^{-2}m^{-0.5} \leq \delta/2$ with relation to $d_{H_1}(A', B')$ for m big enough. So for these pairs, we have $d_{H_2}(A', B') = d \pm (\gamma + 2\delta)$.

Suppose now that $d_G(A', B') \geq \delta/2$. This way, we have at least $\delta(\gamma + 2\delta)^2 m^2/2$ edges available to be removed in the first process. Thus, by Lemma 10.1, the probability that $d_{H_1}(A', B')$ deviates from its expected value by more than $\delta/2$ is at most $2e^{-2(\frac{1}{2}\delta)^2\frac{1}{2}\delta(\gamma+2\delta)^2m^2}$. Since there are at most 2^{2m} pairs (A', B'), the probability that at least one of them is such that $d_{H_1}(A', B')$ deviates from its expected value by more than $\delta/2$ is, by union bound, at most $2^{2m}2e^{-2(\frac{1}{2}\delta)^2\frac{1}{2}\delta(\gamma+2\delta)^2m^2} \leq 1/4$ for m big enough. Thus, with probability at least $3/4$, all pairs (A', B') with $|A'| = |B'| = (\gamma + 2\delta)m$ satisfy $d_{H_1}(A', B') = d \pm (\gamma + 3\delta/2)$. By an argument similar to the previous paragraph, in the second step, adding or removing $m^{1.5}$ edges can change $d_{H_2}(A', B')$ by at most $\delta/2$ with relation to $d_{H_1}(A', B')$, where we get that $d_{H_2}(A', B') = d \pm (\gamma + 2\delta)$. This concludes the proof, as we can make an analogous argument for the case $-\delta \leq p \leq 0$, just adding edges instead of removing them. ■

Proof of Lemma 5.3

For each pair of vertices $a \in A$ and $b \in B$ we will either leave the pair unaltered or we will rerandomize whether there is an edge between a and b. We use the following two-step process.

In the first step, with probability $1 - \frac{2\delta}{(\delta+\gamma)}$ we declare that no change will be made between a and b, and with probability $\frac{2\delta}{(\delta+\gamma)}$ we move to the second step.

In the second step, we ignore whether or not there is an edge between a and b. With probability d we draw an edge between a and b and with probability $1 - d$ we do not draw it. This process is summarized in Figure 5.1.

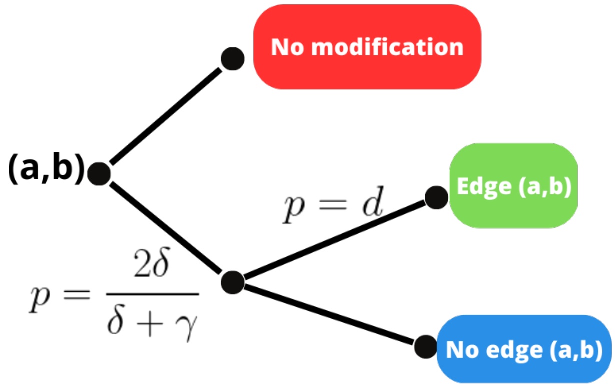


Figure 5.1: Random process to edit the graph.

Claim 5.5 *With probability greater than or equal to* $3/4$*, we make at most* $2.5\delta m^2/\gamma$ *edge modifications.*

Proof The number of edge modifications is less than or equal to the number of times we decided to move to the second step. Since we have m^2 possible pairs, we make this decision m^2 times. The distribution of the results of these m^2 decisions is given by the binomial distribution $B(m^2, \frac{2\delta}{\delta+\gamma})$ whose expected value is $\frac{2\delta}{\delta+\gamma}m^2$ and by Lemma 10.1, the probability of deviating from its expected value by more than $\frac{1}{2}\delta m^2$ is at most $2e^{-2(\delta/2)^2 m^2}$. For m large enough, $m \geq m_{5.3}(\delta, \gamma)$, we have $2e^{-2(\delta/2)^2 m^2} < 1/4$. This way, with probability greater than or equal to $3/4$, we make at most $\delta m^2/2 + \frac{2\delta}{\delta+\gamma}m^2 \leq 2.5\delta m^2/\gamma$ edge modifications. ■

Note that given a pair of adjacent vertices (a, b), for them to remain adjacent, either we decide not to do anything in the first step or we go to the second step and then decide to draw an edge between a and b. That is, the probability that the pair of vertices remains adjacent is $1 - \frac{2\delta}{\delta+\gamma} + \frac{2d\delta}{\delta+\gamma}$. If the vertices are non-adjacent, for them to become so, we need to go to the second step and then decide to draw an edge between (a, b), which happens with probability $\frac{2d\delta}{\delta+\gamma}$.

Claim 5.6 *With probability greater than or equal to 3/4, we have* $d(A, B) = d \pm 1/(m^{0.5})$.

Proof By hypothesis, initially we have dm^2 adjacent vertices and $(1-d)m^2$ non-adjacent. As adjacent vertices stay adjacent with probability $1 - \frac{2\delta}{\delta+\gamma} + \frac{2d\delta}{\delta+\gamma}$

and non-adjacent vertices become adjacent with probability $\frac{2d\delta}{\delta+\gamma}$, we have that the expected value of adjacent vertices is :

$$\left(1-\frac{2\delta}{\delta+\gamma}+\frac{2d\delta}{\delta+\gamma}\right)(dm^2)+\frac{2d\delta}{\delta+\gamma}(1-d)m^2=dm^2.$$

By Lemma 10.1, we have that, for m large enough, the probability that the number of adjacent vertices deviates from dm^2 by more than $m^{1.5}$ is at most $1/4$, from which the result follows. ■

Claim 5.7 *With probability at least 3/4, all sets $A' \subset A$ and $B' \subset B$ whose size is γm, satisfy $d(A', B') = d \pm (\gamma - \delta/2)$.*

Proof Let A' and B' be sets of the desired size. Let e be the number of edges between A' and B'. We then have that the number of non-adjacent pairs is $|A'||B'| - e$. In this way, we have that the expected value of the number of edges between A' and B' after the process is

$$\begin{aligned}&e\left(1-\frac{2\delta}{\delta+\gamma}+\frac{2d\delta}{\delta+\gamma}\right)+(|A'||B'|-e)\left(\frac{2d\delta}{\delta+\gamma}\right)\\&=e\left(1-\frac{2\delta}{\delta+\gamma}\right)+|A'||B'|\left(\frac{2d\delta}{\delta+\gamma}\right).\end{aligned}$$

We have $e = |A'||B'|(d \pm (\gamma+\delta))$ by assumption. Thus, the expected value of number of adjacent vertices is at most:

$$|A'||B'|(d+(\gamma+\delta))\left(1-\frac{2\delta}{\delta+\gamma}\right)+|A'||B'|\left(\frac{2d\delta}{\delta+\gamma}\right)=|A'||B'|(d+\delta-\gamma)$$

Analogously, substituting $e = |A'||B'|(d(\gamma+\delta))$, we obtain that the expected value of the number of edges between A' and B' is at least:

$$|A'||B'|(d-\gamma+\delta).$$

By Lemma 10.1, the probability that the number of edges between A' and B' deviates from its expected value by more than $\delta|A'||B'|/2$ is at most $2e^{-2(\frac{\delta}{2})^2(\gamma m)^2}$. Since the number of pairs (A', B') is at most 2^{2m}, by union bound, the probability that any of these pairs deviates from its expected value more than $\delta|A'||B'|/2$ is at most $2^{2m}2e^{-2(\frac{\delta}{2})^2(\gamma m)^2}$, which is less than or equal to $1/4$ for m big enough, $m \geq m_{5.3}$. Thus, with probability greater than or equal to $3/4$, all pairs (A', B') such that $|A'| = |B'| = \gamma m$ satisfy $d(A', B') = d \pm (\gamma - \delta/2)$. ■

With these three claims we can prove the lemma. We have that with positive probability, we make at most $2.5\delta m^2/\gamma$ modifications, $d(A, B) =$

$d \pm 1/m^{0.5}$ and in addition all pairs (A', B') of size γm satisfy $d(A', B') = d \pm (\gamma - \delta/2)$. Thus, as the probability is positive, there is a way to modify the graph satisfying these 3 restrictions. From this form, we can add or remove at most $m^{1.5}$ to ensure that $d(A, B) = d$. For every pair of sets (A', B') of size γm, this will change $d(A', B')$ by at most $\gamma^2/m^{0.5} \leq \delta/2$ for m large enough, which implies that $d(A', B') = d \pm \gamma$. This makes (A, B) γ-regular with density d, as desired.

■

6 More About Regular Pairs

According to Proposition 3.9, we can assume without loss of generality that every tester is canonical. This means that given a graph $G = (V, E)$ and a tester $\mathcal{T}$ for a graph property $\mathcal{P}_1$, there exists a graph property $\mathcal{P}_2$ such that $\mathcal{T}$ selects a subset $Q \subset V$ of q vertices and accepts or rejects G analyzing whether $G[Q]$ satisfies $\mathcal{P}_2$. Thus, given a graph G that satisfies a certain regularity-instance R, in this chapter we try to understand how the induced subgraph $G[Q]$ inherits this regularity from G. The results presented allow us to estimate the probability of G being accepted by a canonical tester given that G satisfies a certain regularity-instance. This will be deducted from Corollary 6.10. These results will then be used in Chapter 7 to prove that every testable property is regularly reducible.

According to the definition of ϵ-regular pairs, we realize that, intuitively, a pair (A, B) is "regular" if, for all subsets of non-negligible size of $X \subset A$ and $Y \subset B$, the density between such sets is very close to the density between A and B. This property ensures a good distribution of edges between A and B, which makes the pair (A, B) similar to a random bipartite graph in some sense.

Suppose that $V_1, V_2, \ldots, V_k$ are k sets of m vertices. Let us do a random process where we connect each vertex between V_i and V_j randomly and independently with probability $d_{i,j}$. Given a graph $H = (V', E_H)$ with k vertices and a permutation $\sigma : [k] \to [k]$, what is the probability that a specific k-tuple $v_1 \in V_1, v_2 \in V_2, \ldots, v_k \in V_k$ span H in the order defined by σ, that is, each v_i occupies the position defined by $\sigma(i)$? The answer is:

$$\prod_{(i,j)\in E_H} d_{\sigma(i),\sigma(j)} \prod_{(i,j\notin E_H)} (1 - d_{\sigma(i),\sigma(j)})$$

This motivates us to introduce the following definition.

Definition 6.1 *Let $H = (V, E_H)$ be a graph on k vertices, let W be weighted complete graph on k vertices, where the weight of an edge (i, j) is $d_{i,j}$. For a permutation $\sigma : [k] \to [k]$, define:*

$$IC(H, W, \sigma) := \prod_{(i,j)\in E_H} d_{\sigma(i),\sigma(j)} \prod_{(i,j\notin E_H)} (1 - d_{\sigma(i),\sigma(j)})$$

In the previous random process, the expected value of the number of k-tuples of vertices $v_1 \in V_1, v_2 \in V_2, \ldots, v_k \in V_k$ that span H in the order defined by σ is $IC(H, W, \sigma)m^k$, where W is the complete graph with weights $d_{i,j}$. Claim 6.2 shows that if we replace the random bipartite graphs by regular enough bipartite graphs the result is almost the same. This result was demonstrated in Fischer's article [4].

Claim 6.2 *For any δ and k, there is a $\gamma = \gamma_{6.2}(\delta, k)$ such that the following holds. Consider k disjoint sets $V_1, V_2, \ldots, V_k$ each with m vertices such that all pairs (V_i, V_j) are γ-regular and let W be a weighted complete graph of k vertices, with weights given by $d_{i,j} = d(V_i, V_j)$. Then for every graph H on k vertices and every permutation $\sigma : [k] \to [k]$, the number of k-tuples $v_1 \in V_1, v_2 \in V_2 \ldots v_k \in V_k$ that span an induced copy of H in the order defined by σ is*

$$(IC(H, W, \sigma) \pm \delta)m^k.$$

.

In the process defined before the Definition 6.1, how could we answer the same question about the probability without fixing a specific vertex order σ? The answer would be to sum over all permutations and divide by the number of times each copy of H is counted, that is, $\frac{1}{Aut(H)} \sum_\sigma IC(H, W, \sigma)$. This motivates Definition 6.3.

Definition 6.3 *Let H be a graph on k vertices, let W be a weighted complete graph on k vertices, where the weight of an edge (i, j) is $d_{i,j}$. We define:*

$$IC(H, W) := \frac{1}{Aut(H)} \sum_\sigma IC(H, W, \sigma).$$

Again, Claim 6.4 shows that if we replace the random bipartite graphs by regular enough bipartite graphs the result is almost the same.

Claim 6.4 *For any δ and k, there is a $\gamma = \gamma_{6.4}(\delta, k)$ such that: if $V_1, V_2, \ldots, V_k$ are k sets of m vertices, all pairs (V_i, V_j) are γ-regular and W is a complete graph of k weighted vertices, where the weights are $d_{i,j} = d(V_i, V_j)$, then for every graph H of k vertices the number of induced copies of H, which have precisely one vertex in each of the sets $V_1, \ldots, V_k$ is:*

$$(IC(H, W) \pm \delta)m^k.$$

Proof Let us prove that $\delta_{6.4}(\delta, k) = \delta_{6.2}(\delta/k!, k)$ is enough. Assume that $V_1, V_2, \dots, V_k$ are as stated and H is a graph on k vertices. By Claim 6.2, given a permutation σ, the number of induced copies of H that have precisely one vertex v_i in each V_i such that v_i occupies the role of the vertex $\sigma(i)$ is $(IC(H, W, \sigma) \pm \delta/k!)m^k$. Summing over all permutations will count $Aut(H)$ times each copy of H, so we need to divide the final result by $Aut(H)$:

$$\begin{aligned}
&\frac{1}{Aut(H)}\left(\sum_{\sigma}(IC(H, W, \sigma) \pm \delta/k!)\right) m^k \\
&= \left(\frac{1}{Aut(H)}\sum_{\sigma} IC(H, W, \sigma) \pm \delta\right) m^k \\
&= (IC(H, W) \pm \delta)m^k.
\end{aligned}$$

■

The previous results use as a hypothesis the fact that we have exactly k sets V_i. What if we have more sets?

Definition 6.5 *Let H be a graph with k vertices, let $R' = (V_{R'}, E_{R'})$ be a complete weighted graph with at least k vertices where the weight of an edge (i, j) is $d_{i,j}$ and let $\mathcal{W}$ be the set of all subsets of $V_{R'}$ with k vertices, we define:*

$$IC(H, R') := \sum_{W \in \mathcal{W}} IC(H, W).$$

Lemma 6.6 says that it is possible to estimate the number of induced copies of a given graph H in any graph G that satisfies a determined regularity-instance R.

Lemma 6.6 *For any δ and q, there are $k = k_{6.6}(\delta, q)$ and $\gamma = \gamma_{6.6}(\delta, q)$ such that: For any regularity-instance $R = (F, \gamma_R)$ of order at least k and with $\gamma_R \leq \gamma$, any G that satisfies R and for any graph H with $h \leq q$ vertices, the probability that a specific h-tuple of vertices of G spans an induced copy of H is $IC(H, \overline{F}) \pm \delta$.*

During the proof we will need the following probability result.

Claim 6.7 *Let E_1, E_2 and E_3 be events such that:*

1. $\boldsymbol{Pr}[E_1] \geq 1 - \delta/3$;

2. $\boldsymbol{Pr}[E_2] \geq 1 - \delta/3$;

3. $\boldsymbol{Pr}[E_3 | E_1 \cap E_2] = x \pm \delta/3$.

Then $\boldsymbol{Pr}[E_3] = x \pm \delta$.

Proof of Claim 6.7 We need to show that:

- $\mathbf{Pr}[E_3] \geq x - \delta$;
- $\mathbf{Pr}[E_3] \leq x + \delta$.

Let us start with the first inequality:

By the definition of conditional probability, we have:

$$\mathbf{Pr}[E_3|E_1 \cap E_2] = \mathbf{Pr}[E_1 \cap E_2 \cap E_3]/\mathbf{Pr}[E_1 \cap E_2] \leq \mathbf{Pr}[E_3]/\mathbf{Pr}[E_1 \cap E_2]. \quad (6\text{-}1)$$

That way:

$$\mathbf{Pr}[E_3] \geq \mathbf{Pr}[E_3|E_1 \cap E_2]\mathbf{Pr}[E_1 \cap E_2]. \quad (6\text{-}2)$$

However, we have that:

$$\mathbf{Pr}[E_1 \cap E_2] = -\mathbf{Pr}[E_1 \cup E_2] + \mathbf{Pr}[E_1] + \mathbf{Pr}[E_2] \geq 1 - \frac{2\delta}{3}. \quad (6\text{-}3)$$

So, using Inequalities 6-2 and 6-3 we conclude that:

$$\mathbf{Pr}[E_3] \geq \left(1 - \frac{2\delta}{3}\right)\mathbf{Pr}[E_3|E_1 \cap E_2] \geq \mathbf{Pr}[E_3|E_1 \cap E_2] - \frac{2\delta}{3} \geq x - \delta. \quad (6\text{-}4)$$

Now, we will show the second inequality. By the definition of conditional probability we know that:

$$\mathbf{Pr}[E_3|E_1 \cap E_2] = \mathbf{Pr}[E_1 \cap E_2 \cap E_3]/\mathbf{Pr}[E_1 \cap E_2] \geq \mathbf{Pr}[E_1 \cap E_2 \cap E_3]. \quad (6\text{-}5)$$

i.e:

$$\mathbf{Pr}[E_1 \cap E_2 \cap E_3] \leq \mathbf{Pr}[E_3|E_1 \cap E_2] \leq x + \delta/3. \quad (6\text{-}6)$$

But we know that$E_3 = (E_3 \cap E_1 \cap E_2) \cup (E_3 \cap E_1 \cap E_2^C) \cup (E_3 \cap E_1^C)$, that way:

$$\mathbf{Pr}[E_3] \leq \mathbf{Pr}[E_3 \cap E_1 \cap E_2] + \mathbf{Pr}[E_3 \cap E_1 \cap E_2^C] + \mathbf{Pr}[E_3 \cap E_1^C].$$

Rearranging and using conditions $\mathbf{Pr}[E_1] \geq 1 - \frac{\delta}{3}$ e $\mathbf{Pr}[E_2] \geq 1 - \frac{\delta}{3}$ we have that:

$$\mathbf{Pr}[E_3] \leq x + \delta/3 + \delta/3 + \delta/3 \leq x + \delta,$$

what concludes the proof. ■

Proof of Lemma 6.6

Let us show that choosing $k = \frac{10q^2}{\delta}$ and $\gamma = min\{\frac{\delta}{3q^2}\gamma_{6.4}(\frac{1}{3}\delta, q)\}$ is enough.

Let R be a regularity-instance as in the statement and G be a graph satisfying R. Let $V_1, V_2, \ldots, V_l$ be an equipartition of G satisfying R and H a graph on $h \leq q$ vertices. Choosing a h-tuple of G at random, what is the chance that at least 2 of these h vertices belong to the same set V_j of the equipartition? We will show in Claim 6.8 that this probability is at most $\frac{\delta}{3}$.

Claim 6.8 *Choosing a h-tuple of G at random, the chance that at least 2 of these h vertices belong to the same set V_j of the equipartition is at most $\delta/3$.*

Proof of Claim 6.8 Let A_i be the event where at least two of the h vertices are in the same V_i. We know from union bound that:

$$\mathbf{Pr}[A_i] \leq \binom{h}{2}\frac{1}{l^2} \leq \frac{h^2}{2l^2}.$$

We want to bound the probability of $\mathbf{Pr}[\bigcup_{i=1}^{l} A_i]$. By union bound we have:

$$\mathbf{Pr}[\bigcup_{i=1}^{l} A_i] \leq l\frac{h^2}{2l^2} = \frac{h^2}{2l}.$$

Since $h \leq q$ and $l \geq k$, we have

$$\mathbf{Pr}[\bigcup_{i=1}^{l} A_i] \leq \frac{q^2}{2k} = \frac{\delta}{5} \leq \frac{\delta}{3}.$$

■

Furthermore, as the equipartition is γ-regular and $\gamma \leq \delta$, we have the probability that this h-tuple contains a pair of vertices $v_i \in V_i$ and $v_j \in V_j$ such that (V_i, V_j) is not γ-regular is at most $\binom{h}{2}\gamma \leq \binom{q}{2}\gamma \leq \frac{1}{3}\delta$. Therefore, consider the following events:

1. $E_1 = \{$the h vertices $v_1, v_2, \ldots v_h$ belong to distinct sets $V_j\}$;
2. $E_2 = \{$if each vertex of the tuple $v_1, v_2, \ldots, v_h$ belongs to a distinct V_j, then (V_i, V_j) is γ-regular, where $v_i \in V_i$ without loss of generality $\}$;
3. $E_3 = \{$A specific tuple $v_1, v_2, \ldots, v_h$ spans $H\}$.

We want to show that $\mathbf{Pr}[E_3] = IC(H, \overline{F}) \pm \delta$. Since $\mathbf{Pr}[E_1] \geq 1 - \delta/3$ and $\mathbf{Pr}[E_2] \geq 1 - \delta/3$, by Claim 6.7 we only need to show that $\mathbf{Pr}[E_3 | E_1 \cap E_2] = IC(H, \overline{F}) \pm \delta/3$.

Assuming that E_1 and E_2 occur, let us calculate the probability that a specific tuple $v_1, v_2, \ldots, v_h$ spans H. Since E_1 and E_2 occur, we have that the tuple $v_1, v_2, \ldots, v_h$ is such that $v_i \in V_i$ and (V_i, V_j) is γ-regular for each $i < j \leq h$. In this way, as $\gamma \leq \gamma_{6.4}(\frac{1}{3}\delta, q)$, all conditions of Claim 6.4 are satisfied, and so for every possible set W of h sets V_i we get that the probability that they span an induced copy of H is $IC(H, W) \pm \delta/3$. This means that the probability that $\mathbf{Pr}[E_3] = IC(H, \overline{F}) \pm \delta/3$. ■

We could also have stated Lemma 6.6 as follows:

Lemma 6.9 *For any δ and q, there are $k = k_{6.6}(\delta, q)$ and $\gamma = \gamma_{6.6}(\delta, q)$ such that: For any regularity-instance $R = (F, \gamma_R)$ of order at least k and with $\gamma_R \leq \gamma$, any G that satisfies R and for any graph H on $h \leq q$ vertices, the number of induced copies of H in G is $\left(IC(H, \overline{F}) \pm \delta\right) \binom{n}{h}$.*

We can also extend this result to families of graphs, as in Corollary 6.10.

Corollary 6.10 *For any δ and q, there are $k = k_{6.10}(\delta, q)$ and $\gamma = \gamma_{6.10}(\delta, q)$ with the following properties: For any regularity-instance $R = (F, \gamma_R)$ of order at least k and with error parameter $\gamma_R \leq \gamma$, and for every family $\mathcal{A}$ of graphs on q vertices, the number of induced copies of graphs $H \in \mathcal{A}$ in any n-vertex graph satisfaying R is*

$$\left(\sum_{H \in \mathcal{A}} IC(H, \overline{F}) \pm \delta\right) \binom{n}{q}.$$

Proof Let us show that taking $k = k_{6.9}\left(2^{-\binom{q}{2}}\delta, q\right)$ and $\gamma = \gamma_{6.9}\left(2^{-\binom{q}{2}}\delta, q\right)$ is enough. The idea will be to use Lemma 6.9 for each $H \in \mathcal{A}$.

Let $H \in \mathcal{A}$, and R be as in the statement, by Lemma 6.9, we have that the number of induced copies of H in any graph satisfying R is $(IC(H, \overline{F}) \pm 2^{-\binom{q}{2}}\delta)\binom{n}{q}$. Since all graphs of $\mathcal{A}$ have q vertices, the number of possible graphs is at most $2^{\binom{q}{2}}$. Thus, summing over all graphs $H \in \mathcal{A}$, the total error is at most $\delta\binom{n}{q}$. ■

7
First Direction: Testable Implies Regular Reducible

In this chapter we will prove the first direction of the main result

Proposition 7.1 *If a graph property is testable then it is regular-reducible*

To prove this proposition, we will use Proposition 3.9 and assume that without loss of generality the tester is canonical. As explained in Chapter 3, canonical testers for a property $\mathcal{P}$ select a set of $q(\epsilon, n)$ vertices of a graph G and accept the graph if $G[q]$ satisfies a given property $\mathcal{P}'$. We will then use the technical results proved in Chapter 6.

Proof of Proposition 7.1

Assume that $\mathcal{P}$ is testable by a tester $\mathcal{T}$, which by Proposition 3.9 we can assume to be canonical without loss of generality.

Our intention is to build a family $\mathcal{R}$ of regularity-instances R as in Definition 4.9.

Let n and δ be given. We can assume, without loss of generality, that $\delta < 1/12$, otherwise just replace δ with $min\{\delta, 1/13\}$. Let $q' = q'(\delta, n)$ be the query complexity which is sufficient for $\mathcal{T}$ to distinguish between n-vertex graphs satisfying $\mathcal{P}$ and those that are δ-far from satisfying it with probability at least 2/3. Since $\mathcal{T}$ is testable, Definition 3.3, there exists $q(\delta)$ such that $q'(\delta, n) \leq q(\delta)$.

As the tester is canonical, given $G = (V, E)$ passed as input, it works by selecting a set of vertices $V' \subset V$, obtaining an induced graph $G[V']$ with q vertices and rejects or accepts G based on $G[V']$. So let us denote by $\mathcal{A}$ the family of graphs Q with q vertices that would be accepted by the algorithm if $G[V']$ were isomorphic to Q.

Define $k = k_{6.10}(\delta, q)$, $\gamma = \gamma_{6.10}(\delta, q)$ and $N = N_{4.3}(k, \gamma)$.

For all n such that $q \leq n \leq N$, consider all finite regularity-instances of order n, where densities $d_{i,j}$ belong to the finite set $S = \{0, \frac{\delta\gamma^2}{50q^2}, \frac{2\delta\gamma^2}{50q^2}, \ldots, 1\}$. We will define $\mathcal{I}$ as the union of all these finite regularity-instances. Since q depends only on δ, so far all constants depend exclusively on δ. Note that given $x \in (0, 1)$, this construction guarantees that there exists $s \in S$ such that $|x - s| \leq \frac{\delta\gamma^2}{50q^2}$.

We now prove that if we take $\mathcal{R} = \{R = (F, \gamma_R) \in \mathcal{I} : \sum_{H \in \mathcal{A}} IC(\overline{F}, H) \geq 1/2\}$, $\mathcal{P}$ is regular-reducible.

We need to show, by Definition 4.9, that for every G and $\epsilon > 0$:

1. If G satisfies $\mathcal{P}$, then for some $R \in \mathcal{R}$, G is δ-close to satisfying R;

2. If G is ϵ-far from satisfying $\mathcal{P}$, then for all $R \in \mathcal{R}$, G is $(\epsilon - \delta)$-far from satisfying R.

By construction, $\sum_{H \in \mathcal{A}} IC(\overline{F}, H)$ is an estimate of the fraction of induced copies of graphs $H \in \mathcal{A}$ in a graph G of n vertices satisfying R. By Corollary 6.10, the error of this estimation is at most δ.

First property: Let G be a graph satisfying $\mathcal{P}$. We have that $\mathcal{T}$ accepts G with probability at least $2/3$. Since $\mathcal{T}$ is canonical, this means that at least $2/3$ of the subsets V' of q vertices of G are such that $G[V']$ is isomorphic to a graph of $\mathcal{A}$, that is, there are at least $2\binom{n}{q}/3$ sets V'. Let K be that quantity.

By the Regularity Lemma (Proposition 4.3) there exists a γ-regular partition of G of order k', where $k \leq k' \leq N$. By the construction of $\mathcal{I}$, we have that there exists an regularity-instance R' with the same order k' such that $R' = (F', \gamma') \in \mathcal{I}$ and the densities of R' are close to the densities of the regular partition of G. More precisely, for each density d of the regular partition of G, there exists a density d' in R' such that $|d - d'| \leq \frac{\delta\gamma^2}{50q^2}$.

By Proposition 5.4, we can state that G is δ/q^2-close to satisfying such regularity-instance R', that is, we can transform G into a graph that satisfies R' with at most $\delta n^2/q^2$ changes. Given an edge, we have at most $\binom{n-2}{q-2}$ subgraphs of q vertices that use that edge, so changing an edge decreases K by at most $\binom{n-2}{q-2}$. By making the necessary changes to transform G into a graph that satisfies R', we then decrease K by at most $\delta n^2\binom{n-2}{q-2}/q^2 \leq \delta\binom{n}{q}$, making K to be at least $2\binom{n}{q}/3 - \delta\binom{n}{q} > (\frac{1}{2} + \delta)\binom{n}{q}$ since $\delta < 1/12$.

As $\sum_{H \in \mathcal{A}} IC(\overline{F'}, H)\binom{n}{q}$ is an estimative for K with an error of at most $\delta\binom{n}{q}$ and after changes $K > (\frac{1}{2} + \delta)\binom{n}{q}$, we have $\sum_{H \in \mathcal{A}} IC(\overline{F'}, H) \geq 1/2$. That is, with less than δn^2 changes, G can be transformed into a graph satisfying $R' \in \mathcal{R}$, or in other words, G is δ-close to satisfying $R' \in \mathcal{R}$, just like we wanted.

Second Property:

Suppose G is ϵ-far from satisfying $\mathcal{P}$. If $\delta \geq \epsilon$, we are done. Assume now that $\delta < \epsilon$.Suppose for a contradiction that G is $(\epsilon - \delta)$-close to satisfying a regularity-instance $R \in \mathcal{R}$. That is, there exists G' satisfying R such that G is $(\epsilon - \delta)$-close to G'. Claim 7.2 asserts that G' is δ-far from satisfying $\mathcal{P}$.

Claim 7.2 *G' is δ-far from satisfying $\mathcal{P}$.*

Proof Assume for a contradiction that G' is δ-close to satisfying $\mathcal{P}$. So it's possible to make less than δn^2 changes to G' to get a graph that satisfies $\mathcal{P}$. However, by hypothesis, it is possible to do less than $(\epsilon - \delta)n^2$ modifications in G to get G'. Making these changes in sequence, it is then possible to do less than $(\epsilon - \delta)n^2 + \delta n^2 = \epsilon n^2$ modifications to transform G into a graph satisfying $\mathcal{P}$, which is absurd given that G is ϵ-far from satisfying $\mathcal{P}$. ■

Thus, since G' is δ-far from satisfying $\mathcal{P}$, the probability that $\mathcal{T}$ accepts G' is less than $1/3$.

However, since G' satisfies R, this means that at least $(1/2 - \delta)\binom{n}{q}$ subsets of q vertices of G' span a graph of $\mathcal{A}$. As $\delta < 1/12$, we have $(1/2 - \delta)\binom{n}{q} > (1/3 + \delta)\binom{n}{q}$. So $\mathcal{T}$ accepts G' with probability at least $1/3 + \delta$, a contradiction. ■

8
Sampling Regular Partitions

So far, several times we start with a graph G and we need to get properties on some induced subgraph $G[S]$. In this chapter, we will prove technical results about how the regularity of G is inherited by $G[S]$ and vice versa. These results are important tools for demonstrating the main result.

Let us define what two close partitions are.

Definition 8.1 (δ-similar regular-partition) *An equipartition* $\mathcal{U} = \{U_1, U_2, \dots, U_k\}$ *is δ-similar to a γ-regular equipartition* $\mathcal{V} = \{V_1, V_2, \dots, V_k\}$ *if:*

1. $d(U_i, U_j) = d(V_i, V_j) \pm \delta, \qquad \forall i < j;$

2. *If* (V_i, V_j) *is γ-regular, then* (U_i, U_j) *is* $(\gamma + \delta)$*-regular.*

Note that when $\mathcal{U}$ and $\mathcal{V}$ may be equipartitions of different graphs. Furthermore, note that this definition is not symmetric due to the second condition imposed.

The main result of this chapter is to show that when S is large enough, G and $G[S]$ are close to satisfying the same regularity-instance. We now state this result formally.

Proposition 8.2 *For every k and δ there exists a* $q = q_{8.2}(k, \delta)$ *such that for every* $\gamma \geq \delta$ *and* $k' \leq k$ *the following holds. Let* $G = (V, E)$ *be a graph and* $S \subset V$ *a sample of q vertices, then each of the following has probability at least* $2/3$*:*

(i) For all γ-regular equipartition $\mathcal{V}$ of the vertices of G with order k' , it holds that $G[S]$ has an equipartition $\mathcal{U}$ with the same order which is δ-similar to $\mathcal{V}$;

(ii) For all γ-regular equipartition $\mathcal{U}$ of the vertices of $G[S]$ with order k', it holds that G has an equipartition $\mathcal{V}$ of the same order which is δ-similar to $\mathcal{U}$.

Our starting point will be the result of Fischer [7].

Lemma 8.3 *For every k and δ there is* $q = q_{8.3}(k, \delta)$ *such that the following holds for every* $\gamma \geq \delta$ *and* $k' \leq k$*: If a graph G has a γ-regular equipartiion* $\mathcal{V} = \{V_1, V_2, \dots, V_{k'}\}$*, then with probability at least* $2/3$*, a sample of q vertices will have an equipartition* $\mathcal{U} = \{U_1, U_2, \dots, U_{k'}\}$ *satisfying:*

(i) $d(U_i, U_j) = d(V_i, V_j) \pm \delta, \qquad \forall i < j;$

(ii) *if* (V_i, V_j) *is* γ*-regular,* (U_i, U_j) *is* $50\gamma^{1/5}$*-regular.*

This lemma guarantees to inherit a weaker regularity than the one guaranteed by Proposition 8.2. The strategy to strengthen Lemma 8.3 and thus obtain Proposition 8.2 will be to show that when two graphs share a γ-regular partition , then they share all γ'- regulars partitions where $\gamma' > \gamma$. We state this result formally in Lemma 8.5. To state it, we need the following definition.

Definition 8.4 ((δ, γ)-alike regular-partitions) *Two* γ*-regular equipartitions* $\mathcal{V} = \{V_1, V_2, \ldots, V_k\}$ *and* $\mathcal{U} = \{U_1, U_2, \ldots, U_k\}$ *are* (δ, γ)*-alike if:*

1. $d(U_i, U_j) = d(V_i, V_j) \pm \delta \qquad \forall i < j;$

2. *Both* (V_i, V_j) *and* (U_i, U_j) *are simultaneously* γ*-regular for at least* $(1 - \gamma)\binom{k}{2}$ *pairs* $i < j$.

This definition, unlike the Definition 8.1, is symmetric. In general, if we have two γ-regular equipartitions $\mathcal{U} = \{U_1, U_2, \ldots, U_k\}$ and $\mathcal{V} = \{V_1, V_2, \ldots, V_k\}$, (U_i, U_j) be γ-regular does not necessarily imply that (V_i, V_j) is γ-regular. If the partitions are (δ, γ)-alike regular, we have that implication in both directions and on at least $(1 - \gamma)\binom{k}{2}$ of the pairs.

Lemma 8.5 *For every* k *and* δ *there is* $\omega = \omega_{8.5}(k, \delta)$ *such that the following holds for every* $k' \leq k$ *and* $\gamma > 0$*. Suppose that two graphs* $G = (V, E)$ *and* $G' = (V', E')$ *have* (ω, ω)*-alike regular equipartitions* $\mathcal{V} = \{V_1, V_2, \ldots, V_l\}$ *and* $\mathcal{V}' = \{V_1', V_2', \ldots, V_l'\}$ *with* $l > 1/\omega$*. If* G' *has a* γ*-regular equipartition* $\mathcal{A}' = \{A_1', A_2' \ldots, A_{k'}'\}$ *then* G *has an equipartition* $\mathcal{A} = \{A_1, A_2, \ldots, A_{k'}\}$ *which is* δ*-similar to* $\mathcal{A}'$*.*

We will need the following simple inequality:

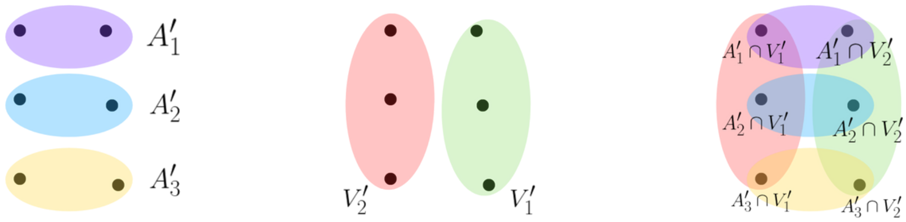
Claim 8.6 *Let* $a_1, \ldots, a_l$ *and* $b_1, \ldots, b_l$ *satisfy* $\sum_{1 \leq i \leq l} a_i = \sum_{1 \leq i \leq l} b_i = 1$ *and* $0 \leq a_i, b_i \leq c$*, then* $\sum_{1 \leq i \leq l} a_i b_i \leq c$

Proof $\sum_{1 \leq i \leq l} a_i b_i \leq \sum_{1 \leq i \leq l} a_i c \leq c \sum_{1 \leq i \leq l} a_i = c.$ ■

Proof of Lemma 8.5

Note that showing the case $k' = k$ is sufficient, as all $k' < k$ follow from it. Then let $k' = k$. Let us find ω that satisfies the statement. We have, by hypothesis, an equipartition $\mathcal{V}$ of G of order l, an equipartition $\mathcal{V}'$ of G' of order l and an equipartition $\mathcal{A}'$ γ-regular of G' of order k. Also, $\mathcal{V}$ and $\mathcal{V}'$ are (ω, ω)-alike regular partitions. We want to construct $\mathcal{A}$ partition of G δ-alike to $\mathcal{A}'$.

Let us start our construction with the equipartition $\mathcal{V}' = \{V'_1, V'_2, \dots, V'_l\}$ of G'. Let us think about the constitution of the other equipartition $\mathcal{A}' = \{A'_1, A'_2 \dots, A'_k\}$ of G'. The equipartition $\mathcal{A}'$ generates a partition $\mathcal{A}'_{V'_i}$ of the vertices of each V'_i, where $\mathcal{A}'_{V'_i} = \{A'_1 \cap V'_i, A'_2 \cap V'_i \dots A'_k \cap V'_i\}$. Let us call $p_{i,j}$ the proportion of the vertices of V'_i that are in A'_j, that is $p_{i,j} = |A'_i \cap V'_j|/|V'_j|$. In Figure 8.1 we have an example of a graph with two partitions $\mathcal{A}'$ and $\mathcal{V}'$ where $p_{i,j} = 1/3$ for all i, j.



(a) $\mathcal{A}'$ partition (b) $\mathcal{V}'$ partition (c)

Figure 8.1: In this example, each $p_{i,j} = 1/3$

We will construct $\mathcal{A}$ to have a similar effect in $\mathcal{V}$. More precisely, we want that for each (i, j), $|A_i \cap V_j|/|V_j| = p_{i,j}$. To do so, we will partition each V_j into k sets $AV_{1,j}, AV_{2,j} \dots AV_{k,j}$ where $|AV_{i,j}| = p_{i,j}|V_j|$ in an arbitrary way. We will define A_i to be $\bigcup_{j=1}^{l} AV_{i,j}$.

Let us show that this definition of $\mathcal{A} = \{A_1, A_2, \dots A_k\}$ is sufficient. For that, we have to show that $\mathcal{A}$ is δ-similar to $\mathcal{A}'$, which by Definition 8.1 means:

1. $d(A_i, A_j) = d(A'_i, A'_j) \pm \delta \qquad \forall i < j$;
2. if (A'_i, A'_j) is γ-regular, then (A_i, A_j) is $(\gamma + \delta)$-regular.

Let us assume that the first condition is true and show the second. Then we will show the first one.

Proof of Second Condition Without loss of generality, we will assume $(i, j) = (1, 2)$ and that the pair (A'_1, A'_2) is γ-regular. We will show that (A_1, A_2) is $(\gamma + 2\delta)$-regular to simplify the notation, and is clearly equivalent. We want to show then that $d(S_1, S_2) = d(A_1, A_2) \pm (\gamma + 2\delta)$ for each $S_1 \subset A_1, S_2 \subset A_2$ of sizes $(\gamma + 2\delta)|A_1|$ and $(\gamma + 2\delta)|A_2|$.

Let $d = d(A'_1, A'_2)$. As we are assuming the first condition, we have that $d(A_1, A_2) = d \pm \delta$. With this, to show that $d(S_1, S_2) = d(A_1, A_2) \pm (\gamma + 2\delta)$ we only need to prove that that $d(S_1, S_2) = d \pm (\gamma + \delta)$ for each $S_1 \subset A_1, S_2 \subset A_2$ of sizes $(\gamma + \delta)|A_1|$ and $(\gamma + \delta)|A_2|$. We will show the upper bound $d(S_1, S_2) \leq d + (\gamma + \delta)$ since the proof of the lower part is analogous. We want to show then that:

$$e(S_1, S_2) \leq (d + \gamma + \delta)|S_1||S_2| = (d + \gamma + \delta)(\gamma + \delta)^2|A_1||A_2| \quad (8\text{-}1)$$

Since $A_1 = \bigcup_{j=1}^{l} AV_{1,j}$, let us define $SV_{1,j} = S_1 \cap AV_{1,j}$ and $SV_{2,j} = S_2 \cap AV_{1,j}$. This way, we have a partition of the sets S_1 and S_2 and we can rewrite our goal as:

$$\sum_{1 \leq i,j \leq l} e(SV_{1,i}, SV_{2,j}) \leq (d + \gamma + \delta)(\gamma + \delta)^2|A_1||A_2| \quad (8\text{-}2)$$

Let n be the number of vertices of G. As $\mathcal{V}$ is an equipartition of order l, V_i has n/l vertices for each i and analogously each A_i has n/k vertices.

Since $\mathcal{V}$ and $\mathcal{V}'$ are (ω, ω)-alike regular equipartitions, then all but $\omega\binom{l}{2}$ of the pairs are such that both (V_i, V_j) and (V_i', V_j') are ω-regular. Let us call $\mathcal{M}$ the set of pairs i, j such that both (V_i, V_j) and (V_i', V_j') are ω-regular. To demonstrate 8-2, let us try to bound the contribution of each pair (i, j) in the sum $\sum_{1 \leq i,j \leq l} e(SV_{1,i}, SV_{2,j})$. For this, we will divide into cases:

Case 1 ($i = j$):

The number of edges connecting a pair $(SV_{1,i}, SV_{2,i})$ is at most:

$$\begin{aligned} |SV_{1,i}||SV_{2,i}| \leq |AV_{1,i}||AV_{2,i}| &= p_{1,i}p_{2,i}|V_i||V_i| \\ &= p_{1,i}p_{2,i}\left(\frac{n}{l}\right)^2 = p_{1,i}p_{2,i}\left(\frac{k}{l}\right)^2 |A_1||A_2| \,. \end{aligned}$$

In this way, when summing up to all i's, the contribution of these pairs is at most $|A_1||A_2| \sum_i \frac{k}{l}p_{1,i}\frac{k}{l}p_{2,i}$, which is less than $k|A_1||A_2|/l$ by Claim 8.6.

If we choose ω so that $l \geq 1/\omega \geq 6k/\delta^3 \geq 6k/(\delta)(\gamma + \delta)^2$, then the contribution to the summation of (8-2) is at most $\frac{1}{6}\delta(\gamma + \delta)^2|A_1||A_2|$.

Case 2 ($|SV_{1,i}| < \omega|V_i|$ ou $|SV_{2,j}| < \omega|V_j|$) :

In this case $|SV_{1,i}| < \omega|V_i| = \omega n/l$. The total number of vertices from G that belong to these sets is at most ωn, so the number of vertices that belong to A_1 is at most $\omega n = \omega kn/k = \omega k|A_1|$. Thus, the contribution of pairs (i, j) such that $|SV_{1,i}| < \omega|V_i| = \omega n/l$ is at most $\omega k|A_1||A_2|$. Analogously, the number of vertices of A_2 in sets $SV_{2,j}$ such that $|SV_{2,j}| < \omega|V_j|$ is at most $k\omega|A_2|$. Thus, the contribution of pairs (i, j) such that $|SV_{2,j}| < \omega|V_j| = \omega n/l$ is at most $\omega k|A_1||A_2|$.

This way, the contribution of both pairs to the summation of (8-2) is at most $2k\omega|A_1||A_2|$. Choosing ω so that $\omega \leq \frac{\delta^3}{12k} \leq \frac{\delta(\gamma+\delta)^2}{12k}$, the contribution is at most $\frac{1}{6}\delta(\gamma + \delta)^2|A_1||A_2|$.

Case 3 (Pairs (i, j) that do not belong to $\mathcal{M}$) : The number of pairs (V_i, V_j) that do not belong to $\mathcal{M}$ are at most $\omega\binom{l}{2}$, so the number of edges connecting sets V_i and V_j not ω-regulars are at most $\omega\binom{l}{2}|V_i||V_J| = \omega\binom{l}{2}n^2/l^2 \leq \omega n^2$. Using the fact that $|A_1| = |A_2| = n/k$, we have that the number of edges connecting A_1 and A_2 that belong to pairs (V_i, V_j) that are not

ω-regular are at most $\omega n^2 = \omega k^2(n/k)^2 = k^2\omega|A_1||A_2|$. Choosing ω such that $\omega \leq \frac{1}{6}\delta^3/k^2 \leq \frac{1}{6}\delta(\gamma+\delta)^2/k^2$, the contribution of these pairs to the summation of (8-2) is at most $\frac{1}{6}\delta(\gamma+\delta)^2|A_1||A_2|$.

Thus, combining the first three cases the total contribution to the summation of (8-2) is at most $\frac{1}{2}\delta(\gamma+\delta)^2|A_1||A_2|$.

Case 4 (other pairs) :

It remains to show that the other pairs contribute less than $(d+\gamma+\delta/2)(\gamma+\delta)^2|A_1||A_2|$. The proof will be extensive and we will use several implications. For readability, Figure 8.2 can be used to help understand which conditions were assumed and which implications were used to reach our goal.

Let B be the set of pairs that do not belong to any of the 3 cases. We have that the pairs (i,j) in B are such that:

1. $i \neq j$;
2. $|SV_{1,i}| \geq \omega|V_i|$ and $|SV_{2,j}| \geq \omega|V_j|$;
3. (V_i, V_j) and (V_i', V_j') are ω-regular.

We want to show then that

$$\sum_{(i,j)\in B} e(SV_{1,i}, SV_{2,j}) = \sum_{(i,j)\in B} d(SV_{1,i}, SV_{2,j})|SV_{1,i}||SV_{2,j}| \leq (d+\gamma+\delta/2)|S_1||S_2|. \tag{8-3}$$

By properties 2 and 3 above, we have:

$$d(SV_{1,i}, SV_{2,j}) = d(V_i, V_j) \pm \omega. \tag{8-4}$$

Furthermore, by the third property, if we take any $SV_{1,i}' \subset V_i'$ such that $|SV_{1,i}'| \geq \omega|V_i'|$ and $SV_{2,j}' \subset V_j'$ such that $|SV_{2,j}'| \geq \omega|V_j'|$ we have that:

$$d(SV_{1,i}', SV_{2,j}') = d(V_i', V_j') \pm \omega. \tag{8-5}$$

If we take $\omega \leq \frac{1}{6}\delta$ and use 8-4, to show 8-3 it is enough to show that :

$$\sum_{(i,j)\in B} d(V_i, V_j)|SV_{1,i}||SV_{2,j}| \leq (d+\gamma+\delta/3)|S_1||S_2|. \tag{8-6}$$

As we know that $\mathcal{V}$ and $\mathcal{V}'$ are (ω,ω) alike, we have that $d(V_i,V_j) = d(V_i',V_j') \pm \omega$. Taking $\omega \leq \frac{1}{6}\delta$, and using this information, it suffices to show that:

$$\sum_{(i,j)\in B} d(V_i', V_j')|SV_{1,i}||SV_{2,j}| \leq (d+\gamma+\delta/6)|S_1||S_2|. \tag{8-7}$$

Furthermore, by 8-5, taking the same $\omega \leq \frac{1}{6}\delta$, it suffices to show that:

$$\sum_{(i,j)\in B} d(SV_{1,i}', SV_{2,j}')|SV_{1,i}||SV_{2,j}| \leq (d+\gamma)|S_1||S_2|. \tag{8-8}$$

We will show that there are $SV'_{1,i} \subset V'_i$ and $SV'_{2,j} \subset V'_j$ satisfying 8-9, where $|SV'_{1,i}| \geq \omega|V'_i|$ and $|SV'_{2,j}| \geq \omega|V'_j|$. This implies 8-8 since it considers all pairs (i,j) and not just the ones in B.

$$\sum_{(i,j)\in 1\leq i,j\leq l} d(SV'_{1,i}, SV'_{2,j})|SV_{1,i}||SV_{2,j}| \leq (d+\gamma)|S_1||S_2|. \qquad (8\text{-}9)$$

To construct $SV'_{1,i} \subset V'_i$ and $SV'_{2,j} \subset V'_j$, let us define $b_{1,i} = |SV_{1,i}|/|S_1|$ and $b_{2,j} = |SV_{2,j}|/|S_2|$. In a similar way to the initial construction, let $(AV)'_{i,j} = A'_i \cap V'_j$, let $SV'_{1,i} \subset (AV)'_{1,i}$ of size $b_{1,i}|(AV)'_{1,i}|$. Similarly, let $SV'_{2,j} \subset (AV)'_{2,j}$ of size $b_{2,j}|(AV)'_{2,j}|$.

Let $S'_1 = \cup_{i=1}^{l} SV'_{1,i}$ and $S'_2 = \cup_{j=1}^{l} SV'_{2,j}$. As we have $|S_1| \geq \gamma|A_1|$ and $|S_2| \geq \gamma|A_2|$, this implies that $|S'_1| \geq \gamma|A'_1|$ and $|S'_2| \geq \gamma|A'_2|$. Dividing by $|S_1||S_2|$, 8-9 becomes:

$$\sum_{(i,j)\in 1\leq i,j\leq l} d(SV'_{1,i}, SV'_{2,j}) b_{1,i} b_{2,j} \leq (d+\gamma). \qquad (8\text{-}10)$$

However, note that $\sum_{(i,j)\in 1\leq i,j\leq l} d(SV'_{1,i}, SV'_{2,j}) b_{1,i} b_{2,j} = d(S'_1, S'_2)$. We want to show then that $d(S'_1, S'_2) \leq (d+\gamma)$. However, (A'_1, A'_2) is γ-regular, $|S'_1| \geq \gamma|A'_1|$, $|S'_2| \geq \gamma|A_2|$ and $d(A'_1, A'_2) = d$. Hence, $d(S'_1, S'_2) \leq (d+\gamma)$, which completes the proof of the second condition. ■

Proof of first condition

In the previous proof, we wanted to show that $d(S_1, S_2) = d \pm (\gamma + \delta)$. Now we want to show that $d(A_1, A_2) = d \pm \delta$. The proof will be very similar, so we will just explain which change results in this gain of a factor of γ in the estimate.

In the previous proof, we needed to bound the density of (S_1, S_2), where $S_1 \subset A_1$ and $S_2 \subset A_2$ As we now want to bound the density of (A_1, A_2), instead of working with S_1 and S_2, we will work directly with A_1 and A_2.

Previously, to obtain Equation 8-10 we had to bound the density of pairs of subsets of $S'_1 \subset A'_1$ and $S'_2 \subset A'_2$ through the γ- regularity of (A'_1, A'_2). The purpose was to use 8-10 to bound density of subsets of A_1 and A_2 through a series of implications, as best seen in Figure 8.2.

Now we want to bound the density of (A_1, A_2) and similarly we will do this by quoting the density $d(A'_1, A'_2)$. As $d(A'_1, A'_2) = d$, this equation plays the role of 8-10 in the previous proof, which results in a gain of a factor of γ in the density estimate of $d(A_1, A_2)$ ■

These two conditions prove Lemma 8.5. ■

With this lemma, we will be able to prove Proposition 8.2, which is our goal in this chapter.

Proof of Proposition 8.2 Let k and δ be given. Let $\omega' = \omega_{8.5}(k, \delta)$

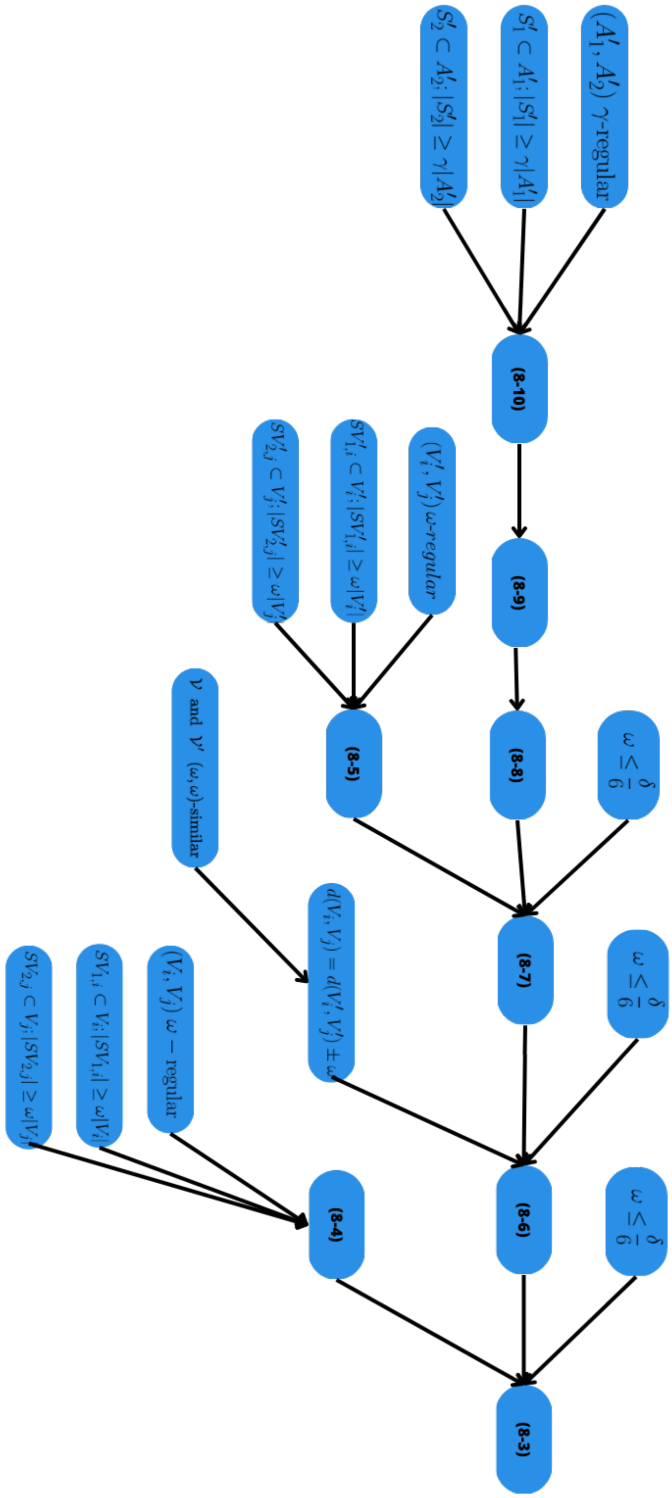


Figure 8.2: Flow on implications of the fourth case of the proof of Lemma 8.5.

and $\omega = (\omega'/50)^5$. We have $\omega, \omega' \leq \omega_{8.5}(k, \delta)$. By the Regularity Lemma (Proposition 4.3) there exists a ω-regular partition of G of order l such that $1/\omega \leq l \leq N_{4.3}(1/\omega, \omega)$. Let $\mathcal{V} = \{V_1, V_2, \dots, V_l\}$ be this partition. We will prove that taking $q = q_{8.2}(k, \delta) = q_{8.3}(l, \omega)$ is enough. Let Q be a set of at least q vertices of G. Let us initially proof Claim 8.7.

Claim 8.7 *With probability at least* $2/3$*,* $G[Q]$ *and* G *have partitions that are* (ω', ω')*-alike.*

Proof By Lemma 8.3, with probability at least $2/3$, $G[Q]$ has an equipartition $\mathcal{U} = \{U_1, U_2, \dots, U_l\}$ such that :

1. $d(V_i, V_j) = d(U_i, U_j) \pm \omega'$;
2. If (V_i, V_j) is ω-regular then (U_i, U_j) is ω'-regular.

Since $\omega' = \omega_{8.5}(k, \delta)$ and $\omega \leq \omega'$, the conditions imply that:

1. $d(V_i, V_j) = d(U_i, U_j) \pm \omega_{8.5}(k, \delta)$;
2. Both (V_i, V_j) and (U_i, U_j) are $\omega_{8.5}(k, \delta)$-regular for all but at most $\omega_{8.5}(k, \delta)\binom{k}{2}$ pairs $i < j$.

That is, with probability greater than $2/3$ the graphs G and $G[Q]$ have partitions that are $(\omega_{8.5}(k, \delta), \omega_{8.5}(k, \delta))$-alike. ∎

Now we will prove both assertions of Proposition 8.2.

Proof of assertion (i) By Claim 8.7, with probability at least $2/3$, $G[Q]$ and G have partitions that are (ω', ω')-alike, where $\omega' = \omega_{8.5}(k, \delta)$. Then, by Lemma 8.5, with probability at least $2/3$, for any γ-regular partition in G of order at most k, $G[Q]$ has an equipartition that is δ-similar to it. ∎

Proof of assertion (ii) By Claim 8.7, with probability at least $2/3$, $G[Q]$ and G have partitions that are (ω', ω')-alike, where $\omega' = \omega_{8.5}(k, \delta)$. Then, by Lemma 8.5, with probability at least $2/3$, for any γ-regular partition of $G[Q]$ with order at most k, G has a partition δ-similar to it. ∎

∎

9
Second Direction: Regular Reducibility Implies Testability

In this chapter, we apply the results demonstrated so far to prove the second direction of the main result (Theorem 4.10). Before that, we need to prove Theorem 4.6, which says that the property of satisfying a given regularity-instance is testable.

9.1
Proof of Theorem 4.6

Proof Let $R = (F, \gamma)$ be a regularity-instance of order k . Given $G = (V, E)$ and ϵ, let us construct a canonical tester that selects a subset Q of q vertices of E and accepts G if and only if $G[Q]$ is $\frac{\gamma^4\epsilon}{200k^2}$-close to satisfying R. Let us set q so that the tester works.

We need to show that there exists q such that:

1. If G satisfies R then $G[Q]$ is $\frac{\gamma^4\epsilon}{200k^2}$-close to satisfying R with probability at least $2/3$;

2. If G is ϵ-far from satisfying R then $G[Q]$ is $\frac{\gamma^4\epsilon}{200k^2}$-far from satisfying R with probability at least $2/3$.

Let us show that there exists $q = q(\epsilon, k, \gamma)$ such that both conditions are satisfied.

Claim 9.1 *If G satisfies R, and $q \geq q_1(\epsilon, k, \gamma)$, then $G[Q]$ is $\frac{\gamma^4\epsilon}{200k^2}$-close to satisfying R with probability at least 2/3.*

Proof of Claim 9.1 Assuming that G satisfies R, we have an equipartition $\mathcal{V} = \{V_1, V_2, \dots, V_k\}$ such that for all $(i, j) \in E(F)$, the pair (V_i, V_j) is γ-regular. By Proposition 8.2, taking $q_1(\epsilon, k, \gamma) = q_{8.2}(k, \frac{\gamma^6\epsilon}{10000k^2})$, we have that with probability at least $2/3$ the graph $G[Q]$ will have an equipartition $\mathcal{A} = \{A_1, A_2, \dots, A_k\}$ such that:

1. $d(A_i, A_j) = d_{i,j} \pm \frac{\gamma^6\epsilon}{10000k^2}$;

2. If (V_i, V_j) is γ-regular then (A_i, A_j) is $(\gamma + \frac{\gamma^6\epsilon}{10000k^2})$-regular.

By Proposition 5.4, this implies that $G[Q]$ is $\frac{\gamma^4\epsilon}{200k^2}$-close to satisfying R.

■

Claim 9.2 *If G is ϵ-far from satisfying R, and $q \geq q_2(\epsilon, k, \gamma)$, then $G[Q]$ is $\frac{\gamma^4\epsilon}{200k^2}$-far from satisfying R with probability at least $2/3$.*

Proof of Claim 9.2 Assuming that $G[Q]$ is $\frac{\gamma^4\epsilon}{200k^2}$-close to satisfying R, we will show that this, with probability at least 2/3, contradicts the hypothesis that G is ϵ- far from R. Let us take $q_2(\epsilon, k, \gamma) = q_{8.2}(\epsilon, k, \gamma)$.

$G[Q]$ be $\frac{\gamma^4\epsilon}{200k^2}$-close to satisfying R means we can do less than $q^2\frac{\gamma^4\epsilon}{200k^2}$ changes and make $G[Q]$ satisfy R. Let $\mathcal{U} = \{U_1, U_2, \ldots, U_k\}$ be the partition of $G[Q]$ that satisfies R after modifications. We will denote by $G' = (V, E')$ the graph G after making such changes. As after $q^2\frac{\gamma^4\epsilon}{200k^2}$ modifications each (U_i, U_j) satisfies $d_{G'}(U_i, U_j) = d_{i,j}$, we have $d_G(U_i, U_j) = d_{i,j} \pm \frac{\gamma^4\epsilon}{200}$. Let us take a pair $(i, j) \in E(F)$. This means that (U_i, U_j) is γ-regular in G'. This way, if we take subsets $SU_i \subset U_i$ and $SU_j \subset U_j$ such that $|SU_i| \geq \gamma|U_i|$ and $|SU_j| \geq \gamma|U_j|$, we have $d_{G'}(SU_i, SU_j) = d_{i,j} \pm \gamma$. So, before the modifications these pairs were such that $d_G(SU_i, SU_j) = d_{i,j} \pm (\gamma + \frac{\gamma^2\epsilon}{200}) = d(U_i, U_j) \pm (\gamma + \frac{\gamma^2\epsilon}{100})$. This then implies that each of these (U_i, U_j) pairs were $(\gamma + \frac{\gamma^2\epsilon}{100})$-regular in G. Thus, $G[Q]$ has a partition $\mathcal{U}$ which is $(\gamma + \frac{\gamma^2\epsilon}{100})$-regular, which by Proposition 8.2 implies that with probability at least 2/3 G has a partition $\mathcal{V}$ that is $\frac{\gamma^4\epsilon}{200k^2}$-similar to $\mathcal{U}$. Suppose that this event occurs.

This means that $\mathcal{V}$ is such that $d(V_i, V_j) = d \pm \frac{\gamma^2\epsilon}{50}$ for all $i < j$ and for all $(i, j) \in E(F)$, (V_i, V_j) is $(\gamma + \frac{\gamma^2\epsilon}{50})$-regular. By Proposition 5.4, G is ϵ-close to satisfying R, which is a contradiction. ■

To make both conditions true, just take q greater than or equal to q_1 and q_2. We can take for example $q = max((q_1(\epsilon, k, \gamma), q_2(\epsilon, k, \gamma))$. Furthermore, the complexity only depends on (ϵ, k, γ). As γ and k are fixed for each regularity instance, the complexity only depends on ϵ given a regularity instance R, which completes the proof. ■

9.2 Proof of Theorem 4.10

Now we will prove Theorem 4.10, the main result.

Proof of Theorem 4.10 The first direction has already been proved in Proposition 7.1. We will now prove the second direction. That is, we show that every regular-reducible graph property is testable.

Let $\mathcal{P}$ be a regular-reducible graph property. Let n and ϵ be given and set $\delta = \epsilon/4$. There is a family $\mathcal{R}$ of at most $r = r(\delta)$ regularity-instances each with complexity at most r such that for every G with n vertices:

1. If G satisfies $\mathcal{P}$, then for some $R \in \mathcal{R}$, G is δ-close to satisfying R;

2. If G is ϵ-far from satisfying $\mathcal{P}$, then for all $R \in \mathcal{R}$, G is $(\epsilon - \delta)$-far from satisfying R.

By Theorem 4.6, for any $R \in \mathcal{R}$, the property of satisfying R is testable. So, by Theorem 3.6, we can distinguish between graphs that are $\epsilon/4$-close to satisfying R from those that are $\frac{3}{4}\epsilon$-far from satisfying it, making a number of queries that is limited by a function of ϵ. Let T_R be the algorithm that makes this distinction.

We will improve the T_R algorithm through Claim 9.3.

Claim 9.3 *It is possible to build an algorithm T'_R that distinguishes between graphs that are $\epsilon/4$-close to satisfying R from those that are $\frac{3}{4}\epsilon$-far from satisfying it such that the following holds:*

1. *The error probability of T'_R is at most $\frac{1}{3r}$*
2. *The query complexity of T'_R is bounded by a function of ϵ*

Proof of Claim 9.3 The idea will be to repeat the algorithm T_R a number of times $m = m(\delta)$ and accept G if it returns the majority result. Thus, we need to prove that it is possible to choose m so that the error probability is less than or equal to $\frac{1}{3r}$. Let X_i be a random variable that assumes the value 1 if the ith algorithm is correct and 0 otherwise. We have that $\sum X_i$ is the number of algorithms that got it right. We want to show that:

$$\mathbf{Pr}\left[\sum X_i \leq \frac{m}{2}\right] \leq \frac{1}{3r}\,.$$

We know that $\mathbb{E}[X_i] = \sum_{i=1}^{m} \mathbf{Pr}[X_i = 1] \geq \frac{2m}{3}$. Let $E = \mathbb{E}[X_i]$. We will show Inequality 9-1, which implies the desired inequality:

$$\mathbf{Pr}[|\sum X_i - \mathbb{E}[X_i]| \geq m/6] \leq \frac{1}{3r}\,. \tag{9-1}$$

As $\frac{m}{6} \leq \frac{E}{4}$, it remains to show that:

$$\mathbf{Pr}\left[|\sum X_i - E| \geq \frac{E}{4}\right] \leq \frac{1}{3r}\,. \tag{9-2}$$

By Chernoff bound, Lemma 10.1, we have that:

$$\mathbf{Pr}\left[|\sum X_i - E| \geq \frac{E}{4}\right] \leq 2e^{-\frac{E}{48}}\,.$$

As $E \geq 2m/3$, we conclude that:

$$\mathbf{Pr}\left[|\sum X_i - E| \geq \frac{E}{4}\right] \leq 2e^{-\frac{m}{72}}\,.$$

Taking m such that $e^{-\frac{m}{72}} \leq \frac{1}{6r}$, we obtain the desired result. Since r depends only on δ, which in turn is a function of ϵ, this completes the proof.

■

We can now define our tester $T_{\mathcal{P}}$ for $\mathcal{P}$: Given $G = (V, E)$ with n vertices. We will accept G if for some $R \in \mathcal{R}$, T'_R says that G is $\epsilon/4$-close to satisfy R. Otherwise we will reject. We need to show that:

1. If G satisfies $\mathcal{P}$, $T_{\mathcal{P}}$ accepts G with probability at least $2/3$;

2. If G is ϵ-far from $\mathcal{P}$, $T_{\mathcal{P}}$ rejects G with probability at least $2/3$.

Proof of First condition Suppose G satisfies $\mathcal{P}$. As $\delta = \epsilon/4$ and $\mathcal{P}$ is regular-reducile to $\mathcal{R}$, G has to be $\epsilon/4$-close to satisfying some regularity-instance $R \in \mathcal{R}$. The probability that T'_R returns that G is $\epsilon/4$-close to R is at least $1 - \frac{1}{3r}$, which is greater than $2/3$. By probability monotonicity, the probability that $T_{\mathcal{P}}$ accepts G also is at least $2/3$.

■

Proof of Second condition Since $\mathcal{P}$ is regular-reducile to $\mathcal{R}$, G has to be $\frac{3\epsilon}{4}$-far from satisfying all of the regularity-instances $R \in \mathcal{R}$. Let A_R be the event where the algorithm T'_R says that G is $\frac{3}{4}$-far from R. We have $\mathbf{Pr}[A_R^C] \leq \frac{1}{3r}$ for every $R \in \mathcal{R}$.

We want to bound $\mathbf{Pr}[\bigcap_{R\in\mathcal{R}} A_R]$. However, we know that:

$$\mathbf{Pr}[\bigcap_{R\in\mathcal{R}} A_R] \geq 1 - \sum_{R\in\mathcal{R}} \mathbf{Pr}[A_R^C] \geq 2/3,$$

concluding the proof of the second condition. ■

These two conditions complete the proof of Theorem. ■

9.3 Triangle-Freeness Testability

In Section 4.2 we prove that the property of being Triangle-free is testable by presenting an algorithm. Now, we will show that this can also be deduced from the general result Theorem 4.10.

Corollary 9.4 *The property of being Triangle-Free is testable .*

Proof Since $\mathcal{P}$ is Triangle-Free, by Theorem 4.10 we need to show that $\mathcal{P}$ is regular-reducible. Let $\delta > 0$ and $\gamma' = \gamma_{6.4}(\gamma', \delta)$. Let $\gamma = min\{\gamma', \delta\}$. We define $\mathcal{R}$ to be all regularity-instances R satisfying the following:

1. All R have regularity parameter γ;

2. The order of each R is at least $\frac{1}{\gamma}$ and at most $T_{4.3}(\frac{1}{\gamma}, \gamma)$;

3. All densities $d_{i,j}$ of R belong to the set $C = \{0, \gamma, 2\gamma, \dots, 1\}$

4. For every R and every three V_i, V_j and V_k, at least one of the three densities $d_{i,j}, d_{i,k}, d_{j,k}$ is 0.

Let $G = (V, E)$ be a graph of n vertices, we need to show that:

- If G satisfies $\mathcal{P}$, then for some $R \in \mathcal{R}$, G is δ-close to satisfying $\mathcal{R}$;
- If G is ϵ-far from satisfying $\mathcal{P}$, then for any $R \in \mathcal{R}$, G is $(\epsilon - \delta)$-far from satisfaying R.

Proof of first condition We are assuming that G is triangle-free. By the Regularity Lemma (Proposition 4.3), G has a γ-regular equipartition $\mathcal{V} = \{V_1, V_2, \ldots, V_k\}$ of order $1/\gamma \leq k \leq T_{4.3}(1/\gamma, \gamma)$. Due to our choice of γ' as $\gamma \leq \gamma'$, we can use Claim 6.4 to say that there is no i, j, k such that $(V_i, V_j), (V_j, V_k), (V_i, V_k)$ are γ-regular and $d(V_i, V_j), d(V_j, V_k), d(Vi, V_k) \geq \delta$. otherwise, these pairs would necessarily generate triangles. By the third restriction that we imposed on $\mathcal{R}$, we can say that there is a regularity-instance R such that for each $d_{i,j}$ of $\mathcal{V}$, there is a density $d'_{i,j}$ of R such that $d'_{i,j} = d_{i,j} \pm \gamma$. Using Lemma 5.1, we can make less than $\gamma n^2 \leq \delta n^2$ and transform the densities of $\mathcal{V}$ into the same densities of R keeping the regularity. Thus, G is δ-close to satisfying $\mathcal{R}$. ■

Proof of second condition We are assuming that G is ϵ-far from satisfying $\mathcal{P}$. Assume that G is $(\epsilon - \delta)$-close to satisfying a regularity-instance $R \in \mathcal{R}$ and let us see that this generates a contradiction. We can then remove less than $(\epsilon - \delta)n^2$ edges from G and turn it into a graph satisfying R. Note that any graph satisfying R does not have triangles with a vertex in each set V due to the fourth condition imposed on $\mathcal{R}$. Make these modifications and additionally remove all edges from within the sets V_i. By the second restriction imposed on $\mathcal{R}$, each V_i has a maximum size of $\gamma n \leq \delta n$, which causes us to remove less than δn^2 edges in this additional modification. In total, we made less than ϵn^2 changes. These two changes ensure that the resulting graph is triangle-free. This is a contradiction since G is ϵ-far from $\mathcal{P}$.

■

■

10
Appendix

10.1
Useful Inequalities

Lemma 10.1 (A Chernoff bound [15]) *Suppose* $X_1, X_2, \dots, X_n$ *are* n *independent Boolean random variables, where* $\boldsymbol{Pr}[X_i = 1] = p_i$. *Let* $E = \sum_{i=1}^{n} p_i$ *and* $0 \leq \delta \leq 1$. *Then* $\boldsymbol{Pr}(|\sum_{i=1}^{n} X_i - E| \geq \delta E) \leq 2e^{-\delta^2 E/3}$.

Lemma 10.2 (Another Chernoff Bound [1]) *Let* $X_1, X_2, \dots, X_n$ *be identical independent random variables ranging in* $[0,1]$, *and let* $p = \mathbb{E}[X_1]$. *Then, for every* $\epsilon \in (0,1]$ *it holds that :*

$$\boldsymbol{Pr}\left[\left|\frac{1}{n}\sum_{i\in[n]} X_i - p\right| > \epsilon\right] < 2e^{-\epsilon^2 n/4}.$$

Lemma 10.3 *For all* $x \in \mathbb{R}$, $(1 + x) \leq e^x$.

Proof Note that $y = x + 1$ is the tangent line to the graph of e^x when $x = 0$. Since e^x is convex, the inequality follows. ■

11
Bibliography

1 GOLDREICH, O. **Introduction to Property Testing**. [S.l.]: Cambridge University Press, 2017.

2 GOLDREICH, O.; GOLDWASSER, S.; RON, D. Property testing and its connection to learning and approximation. **J. ACM**, Association for Computing Machinery, New York, NY, USA, v. 45, n. 4, p. 653–750, jul 1998. ISSN 0004-5411. Disponível em: <https://doi.org/10.1145/285055.285060>.

3 GAREY, M. R.; JOHNSON, D. S.; STOCKMEYER, L. Some simplified np-complete problems. In: **Proceedings of the Sixth Annual ACM Symposium on Theory of Computing**. New York, NY, USA: Association for Computing Machinery, 1974. (STOC '74), p. 47–63. ISBN 9781450374231. Disponível em: <https://doi.org/10.1145/800119.803884>.

4 FISCHER, E. The difficulty of testing for isomorphism against a graph that is given in advance. **SIAM Journal on Computing**, v. 34, n. 5, p. 1147–1158, 2005. Disponível em: <https://doi.org/10.1137/S009753970445603>.

5 ALON, N. et al. A combinatorial characterization of the testable graph properties: It's all about regularity. **SIAM Journal on Computing**, v. 39, n. 1, p. 143–167, 2009. Disponível em: <https://doi.org/10.1137/060667177>.

6 SZEMEREDI, E. **Regular Partitions of Graphs.** Stanford, CA, USA, 1975.

7 FISCHER, E. Testing graphs for colorable properties. In: **Proceedings of the Twelfth Annual ACM-SIAM Symposium on Discrete Algorithms**. USA: Society for Industrial and Applied Mathematics, 2001. (SODA '01), p. 873–882. ISBN 0898714907.

8 SIPSER, M. **Introduction to the Theory of Computation**. 1st. ed. [S.l.]: International Thomson Publishing, 1996. ISBN 053494728X.

9 FISCHER, E.; NEWMAN, I. Testing versus estimation of graph properties. In: **Proceedings of the Thirty-Seventh Annual ACM Symposium on Theory of Computing**. New York, NY, USA: Association for Computing Machinery, 2005. (STOC '05), p. 138–146. ISBN 1581139608. Disponível em: <https://doi.org/10.1145/1060590.1060612>.

10 GOLDREICH, O.; TREVISAN, L. Three theorems regarding testing graph properties. In: **Proceedings 42nd IEEE Symposium on Foundations of Computer Science**. [S.l.: s.n.], 2001. p. 460–469.

11 KOMLÓS, J. et al. The regularity lemma and its applications in graph theory. In: _____. **Theoretical Aspects of Computer Science: Advanced Lectures**. Berlin, Heidelberg: Springer Berlin Heidelberg, 2002. p. 84–112. ISBN 978-3-540-45878-4. Disponível em: <https://doi.org/10.1007/3-540-45878-6_3>.

12 RUZSA, I.; SZEMERÉDI, E. Triple systems with no six points carrying three triangles. **Combinatorica**, v. 18, 01 1976.

13 ERDŐS, P.; FRANKL, P.; RÖDL, V. The asymptotic number of graphs not containing a fixed subgraph and a problem for hypergraphs having no exponent. **Graph. Comb.**, Springer-Verlag, Berlin, Heidelberg, v. 2, n. 1, p. 113–121, dec 1986. ISSN 0911-0119. Disponível em: <https://doi.org/10.1007/BF01788085>.

14 GOWERS, W. T. Lower bounds of tower type for Szemerédi's uniformity lemma. **Geometric & Functional Analysis GAFA**, v. 7, p. 322–337, 1997.

15 GOEMANS, M. **Chernoff bounds, and some applications**. [S.l.]: MIT, 2015.