

# DRP Notes: Graph Limits and Regularity Lemma

Youyou Wu

January 9, 2026

This is an expository writing broadly on the topic of graph homomorphisms, graph limits (graphons), regularity lemma and other applications of these tools. The note is largely based on "Large Networks and Graph Limits" by Laszlo Lovasz and the textbook of Yufei Zhao's course on "Graph Theory and Additive Combinatorics" at MIT. I have also consulted many other sources, but I have made sure to list them within the text when referenced.

These notes are prepared and organized by the writer during a Directed Reading Program (DRP) at UIUC during fall of 2025, with the help of Bowen Li. The program ran in a format where I deliver a presentation every week on the assigned contents, the mentor make some comments, and then we discuss the contents to read for next week.

The notes follows an intuition-first writing style. Although usually the textbook definitions are given, there are plenty of other sources that can also do that. I want to convey my personal intuition and understanding of these materials as much as possible, and try my best to make it self-contained.

# Contents

<b>1</b>	<b>Graph homomorphisms</b>	<b>4</b>
1.1	homomorphism number	4
1.2	homomorphism density	5
1.3	Relations between homomorphism numbers	6
<b>2</b>	<b>Kernels and graphons</b>	<b>8</b>
2.1	kernels, graphons, stepfunctions	8
2.2	generalized homomorphism	8
<b>3</b>	<b>The Cut Distance</b>	<b>9</b>
3.1	The Cut Distance of Graphs	9
3.1.1	Norms of a Matrix	9
3.1.2	Two Graphs on the Same Set of Nodes	9
3.1.3	Two Graphs with the Same Number of Nodes (Unlabeled)	10
3.1.4	Two Arbitrary Graphs	10
3.2	Cut Norm and Cut Distance of Kernels (Incomplete, but enough for our discussion)	11
<b>4</b>	<b>Szemerédi Partition and Regularity lemma</b>	<b>12</b>
4.1	Basic notation	12
4.2	$\epsilon$ -homogeneous bipartite pairs	12
4.3	Cut distance for unweighted graphs	13
4.4	Weak regularity lemma, almost original form	13
4.5	Weak regularity for graphs	13
4.6	Relation between the two version	14
4.7	A strong (Frieze–Kannan style) refinement statement	15
4.8	Proof of Weak Regularity Lemma, geometric form	16
4.8.1	Preliminaries and Notation	16
4.8.2	Lemma 9.7 (Matrix Version)	16
4.9	A Regularity Lemma in Hilbert Space	18
4.10	Compactness of the graphon space	20
<b>5</b>	<b>Graph Sampling</b>	<b>22</b>
5.1	Counting Lemma	22
<b>6</b>	<b>Convergence of dense graph sequences</b>	<b>24</b>
6.1	What it means to say a sequence of graphs is convergent	24
6.2	Random graphs as limit objects	24
6.2.1	Finite random graph models (frgm)	24
6.2.2	Countable random graph models (crgm)	25
<b>7</b>	<b>Graph Regularity Method: Some Applications</b>	<b>27</b>
7.1	Graph Regularity Lemma Revisit	27
7.2	Quasirandom Graphs	27
7.2.1	Multitype Quasirandom Graphs	29
7.3	Graph Counting Lemmas	29
7.3.1	Triangle Counting Lemma	29
7.3.2	$K_4$ Counting Lemma	30
7.3.3	General Graph Counting Lemma	32
7.4	Removal Lemmas	32

7.4.1	Triangle Removal Lemma	32
7.4.2	Extensions of the Removal Lemma	35
7.5	Applications	37
7.5.1	Proof of Roth's Theorem	37
7.5.2	Erdos-Stone-Simonovits Theorem	38
7.6	Statements of Other Removal Lemmas	39

# 1 Graph homomorphisms

## 1.1 homomorphism number

**Definition 0.1 (graph homomorphisms).** Let  $G$  and  $H$  be two simple graphs. An adjacency-preserving map from  $V(G)$  to  $V(H)$  is called a *homomorphism*.

We are interested in determining the existence of graph homomorphisms, but that problem will be NP-complete on most graphs.

We are also interested in determining the number of homomorphisms between graphs, which we call homomorphism numbers.

**Definition 0.2 (homomorphism Number for simple graphs).** For two simple graphs  $F$  and  $G$ ,

let  $Hom(F, G)$  denote the set of homomorphisms of  $F$  into  $G$  ( $F \Rightarrow G$ ), and let  $hom(F, G)$  be their number.

Let  $inj(F, G)$  be the number of *injective homomorphisms* of  $F$  into  $G$ , and  $ind(F, G)$  be the number of *embeddings* of  $F$  into  $G$  as induced subgraphs (i.e., also preserves non-adjacency between vertices).

The notion of  $hom(F, G)$  on multigraph is more complicated.

**Definition 0.3 (homomorphism number for multigraph).** A *node-and-edge homomorphism* preserves both edge and node adjacency, in other words, it is a pair of maps, the vertex map  $\phi$  and the edge map  $\delta$  such that if  $e \in E(G)$  connects  $i$  and  $j$ , then  $\delta(e)$  connects  $\phi(i)$  and  $\phi(j)$ .

This is usually the standard version of  $hom(F, G)$  that we will use for multigraph. Other versions include *node-homomorphisms*  $\phi$  such that the multiplicity of  $\phi(i)\phi(j)$  is at least as large as the multiplicity of  $ij$ . And *induced homomorphisms* that preserves edge multiplicity, and therefore preserving non-adjacency as well.

Next we extend this definition to weighted loop-simple graph (*loop-simple graphs* are simple graphs that are allowed loops).

**Definition 0.4 (graph homomorphism for weighted graphs).** For weighted loop-simple graph  $G$ , we denote  $\alpha_v(G)$  as the nodeweights and  $\beta_{uv}(G)$  as the edgeweights. For every map  $\phi : V(F) \rightarrow V(G)$ , we assign weights by

$$\alpha_\phi = \prod_{u \in V(F)} \alpha_{\phi(u)}(G),$$

and

$$hom_\phi(F, G) = \prod_{uv \in E(F)} \beta_{\phi(u)\phi(v)}(G).$$

We define

$$\text{hom}(F, G) = \sum_{\phi: V(F) \rightarrow V(G)} \alpha_{\phi} \text{hom}_{\phi}(F, G),$$

and

$$\text{inj}(F, G) = \sum_{\phi: V(F) \hookrightarrow V(G)} \alpha_{\phi} \text{hom}_{\phi}(F, G).$$

The above definition makes sense if  $F$  is a multigraph and  $G$  is a weighted graph. For  $G$  to be an unweighted multigraph, we can consider  $G'$  such that each edge is weighted by its multiplicity in  $G$ , where  $\text{hom}(F, G) = \text{hom}(F, G')$ .

Note that here we require  $F$  to be unweighted. One can try to define  $\text{hom}(F, G)$  when both are weighted graphs when the weights satisfy some reasonable conditions, but they may not be well defined.

Other definitions for signed graphs and partial labeled graphs exist. Refer to the books for details.

## 1.2 homomorphism density

We often normalize the homomorphism numbers to get *homomorphism densities*.

**Definition 0.5 (homomorphism density).** Let  $n = V(G)$ ,  $k = V(F)$ , we define

$$t(F, G) = \frac{\text{hom}(F, G)}{n^k}$$

which is the probability that a random map from  $F$  into  $G$  is a homomorphism.

Similarly, we define

$$t_{\text{inj}}(F, G) = \frac{\text{inj}(F, G)}{(n)_k}$$

and

$$t_{\text{ind}}(F, G) = \frac{\text{ind}(F, G)}{(n)_k}$$

which is the probability that a random injection from  $F$  into  $G$  preserves adjacency and non-adjacency.

For weighted graphs, we define

$$\alpha_H = \sum_{v \in V(H)} \alpha_v(H)$$

and

$$t(F, H) = \frac{\text{hom}(F, H)}{\alpha_H^{v(F)}}$$

If we normalize the weights of  $H$  by dividing it by the sum of the weights, the node weights form a probability distribution, which is the expectation that each vertex is chosen independently.

The above definition doesn't work that well for sparse graphs.

**Definition 0.6 (homomorphism frequency).** For bounded degree graphs  $G$  and connected graph  $F$ , we normalize them differently,

$$t^*(F, G) = \frac{\text{hom}(F, G)}{v(G)}$$

and we call this the *homomorphism frequency* of  $F$  in  $G$  to distinguish from the dense case.

### 1.3 Relations between homomorphism numbers

Below we summarize most important relationships between the homomorphism densities. We won't necessary use them in our discussion, but keeping them in mind, especially the relationship between  $t(F, G)$  and  $t_{\text{ind}}(F, G)$  will be reused some time later in our proof.

**Definition 0.7 (Injective/induced vs. hom).** For two simple graphs  $F$  and  $G$ ,

$$\text{inj}(F, G) = \sum_{F' \supseteq F} \text{ind}(F', G).$$

$$\text{hom}(F, G) = \sum_P \text{inj}(F/P, G).$$

**Definition 0.8 (Inclusion–exclusion and Möbius inversion).** For two simple graphs  $F$  and  $G$ ,

$$\text{ind}(F, G) = \sum_{\substack{F' \supseteq F \\ V(F')=V(F)}} (-1)^{e(F')-e(F)} \text{inj}(F', G).$$

$$\text{ind}(\cdot, G) = \text{inj}^{\uparrow}(\cdot, G).$$

$$\text{inj}(F, G) = \sum_P \mu_P \text{hom}(F/P, G).$$

$$\text{inj}(\cdot, G) = \text{hom}^{\downarrow}(\cdot, G).$$

**Definition 0.9 (Relations for homomorphism densities).** For two simple graphs  $F$  and  $G$ ,

$$t_{\text{inj}}(F, G) = \sum_{F' \supseteq F} t_{\text{ind}}(F', G).$$

$$t_{\text{ind}}(F, G) = \sum_{F' \supseteq F} (-1)^{e(F')-e(F)} t_{\text{inj}}(F', G) = t_{\text{inj}}^{\uparrow}(F, G).$$

$$|t_{\text{inj}}(F, G) - t(F, G)| \leq \frac{1}{v(G)} \binom{v(F)}{2}.$$

**Definition 0.10 (Complementation identities).** For two simple graphs  $F$  and  $G$ ,

$$\text{ind}(F, \bar{G}) = \text{ind}(\bar{F}, G).$$

$$\text{hom}(F, \bar{G}) = \sum_{\substack{F' \subseteq F \\ V(F')=V(F)}} (-1)^{e(F')} \text{hom}(F', G).$$

$$\text{inj}(F, \bar{G}) = \sum_{\substack{F' \subseteq F \\ V(F')=V(F)}} (-1)^{e(F')} \text{inj}(F', G).$$

**Definition 0.11 (Averaging identities).** Let  $v(F_0) = k \leq t \leq v(G) = n$ . Then

$$\text{ind}(F_0, G) = \frac{1}{\binom{n-k}{t-k}} \sum_{F: v(F)=t} \frac{\text{ind}(F_0, F) \text{ind}(F, G)}{\text{aut}(F)}.$$

If  $v(F) = k \leq t \leq v(G) = n$ , then

$$\text{inj}(F, G) = \frac{1}{\binom{n-k}{t-k}} \sum_{S \in \binom{V(G)}{t}} \text{inj}(F, G[S]).$$

$$t_{\text{inj}}(F, G) = \frac{1}{\binom{n}{t}} \sum_{S \in \binom{V(G)}{t}} t_{\text{inj}}(F, G[S]).$$

**Definition 0.12 (Graph operations).** If  $F_1$  and  $F_2$  are node-disjoint, then

$$\text{hom}(F_1 \cup F_2, G) = \text{hom}(F_1, G) \text{hom}(F_2, G).$$

If  $F$  is connected and  $G_1$  and  $G_2$  are node-disjoint, then

$$\text{hom}(F, G_1 \cup G_2) = \text{hom}(F, G_1) + \text{hom}(F, G_2).$$

$$\text{hom}(F, G_1 \times G_2) = \text{hom}(F, G_1) \text{hom}(F, G_2).$$

Chapter 5.3 on the book also has some nice examples on what homomorphism functions can express which are nice to look at.

## 2 Kernels and graphons

We introduce the notion of kernels and graphons, which will serve as the limit object when we talk about the limit of graph sequences. We first look at how we generalize graphs into more analytical objects that will simplify our discussion of limits.

### 2.1 kernels, graphons, stepfunctions

**Definition 0.13 (kernels and graphons).** Let  $W$  denote the space of all bounded symmetric measurable functions  $W: [0, 1]^2 \rightarrow \mathbb{R}$ . The elements of  $W$  will be called *kernels*.

Let  $W_0$  denote the set of all kernels such that  $0 \leq W \leq 1$ . The elements of  $W_0$  will be called *graphons*.

Intuitively, think of graphons as "adjacency matrix" of graphs. Graphons whose values are 0 and 1 can be considered graph on node set  $[0, 1]$ . And then we can talk about the usual notions of graphs like subgraphs, induced subgraphs and complements, etc.

Kernels generalize weighted graphs in the following sense:

**Definition 0.14 (stepfunction).** A kernel  $W$  is called a *stepfunction* if there's a partition  $S_1 \cup S_2 \cup \dots \cup S_k$  of  $[0, 1]$  into measurable sets such that  $W$  is constant on every product set of  $S_i \times S_j$ . The sets are called the *steps* of  $W$ .

For weighted graphs  $H$ , we define its stepfunction  $W_H$  as follows: split  $[0, 1]$  into  $n$  intervals  $J_i$ , each of length corresponding to the node weights  $\alpha_i/\alpha_H$ . And for  $x \in J_i$  and  $y \in J_j$ , let  $W_H(x, y) = \beta_{ij}(H)$ . Note that  $W_H$  depends on how the nodes of  $H$  are labeled.

A weighted graph uniquely determines a stepfunction, and vice versa.

If the edgeweights of  $H$  are from the interval  $[0, 1]$ , then  $W_H$  is a *graphon*. In particular, for every simple unweighted graph  $G$ ,  $W_G$  is a 0-1 valued graphon.

We can of course try to extend the properties of simple graphs to graphons. Most importantly, the *normalized degree function*:

### 2.2 generalized homomorphism

homomorphism densities in graphs extends to homomorphism densities in graphons and kernels. For every  $W$  and multigraph  $F$ , Define

View this as an infinite analogue of weighted homomorphism numbers.

As expected, we have  $t(F, G) = t(F, W_G)$ .

We can generalize it even further.

**Definition 0.15 (decorated graphs).** Let  $A$  be a set of kernels.

An *A-decorated graph* is a finite simple graph  $F = (V, E)$  in which every edge  $e \in E$  is labeled by function  $W_e \in A$ . For every  $W$ -decorated graph  $(F, w)$  we Define

The catch here is edges in a graph can be labeled from different kernels, and it can be used to express homomorphism densities in sums of kernels.

### 3 The Cut Distance

We have announced in the Introduction that we are going to define the distance of two arbitrary graphs, so that this distance will reflect structural similarity. The definition is quite involved, so we approach the problem in several steps: starting with two graphs on the same node set, then moving to graphs with the same number of nodes (but unrelated sets), then to the general case. Finally, we extend the definition to kernels. We consider dense graphs here; for sparse graphs (edge density  $o(1)$ ), these distances would be trivially small.

#### 3.1 The Cut Distance of Graphs

##### 3.1.1 Norms of a Matrix

Let  $A$  be an  $n \times n$  matrix. Several norms appear in various studies:

- The  $l_1$ -norm:  $\|A\|_1 = \frac{1}{n^2} \sum_{i,j=1}^n |A_{ij}|$ .
- The  $l_2$  (or Frobenius) norm:  $\|A\|_2 = \left( \frac{1}{n^2} \sum_{i,j=1}^n A_{ij}^2 \right)^{1/2}$ .
- The  $l_\infty$ -norm:  $\|A\|_\infty = \max_{i,j} |A_{ij}|$ .

Note the normalization by  $n^2$ , ensuring these norms are between 0 and 1 for adjacency matrices. Our main tool is the **cut norm**, introduced by Frieze and Kannan [1999]:

**Definition 0.16 (Cut Norm).**

$$\|A\|_{\square} = \frac{1}{n^2} \max_{S,T \subseteq [n]} \left| \sum_{i \in S, j \in T} A_{ij} \right|.$$

It is clear that the cut norm is bounded by the others:

$$\|A\|_{\square} \leq \|A\|_1 \leq \|A\|_2 \leq \|A\|_\infty.$$

**Example 0.1.** Let  $A$  be a random  $\pm 1$  matrix. Then  $\|A\|_1 = 1$ , but  $\|A\|_{\square} = \Theta(1/\sqrt{n})$ . This highlights that the cut norm measures "global" structure (or lack thereof), whereas  $l_1$  measures local absolute differences.

**Computability:** The cut norm is hard to compute exactly. However, Alon and Naor [2006] related it to the *Grothendieck norm*, which allows for a polynomial-time approximation within a constant factor (less than 2).

##### 3.1.2 Two Graphs on the Same Set of Nodes

Let  $G$  and  $G'$  be two graphs with the same node set  $[n]$ .

**Definition 0.17 ( $l_1$  and Cut Distance).** The  $l_1$  **distance** (or edit distance) is:

$$d_1(G, G') = \frac{|E(G) \Delta E(G')|}{n^2} = \|A_G - A_{G'}\|_1.$$

The **cut distance** (labeled) is:

$$d_{\square}(G, G') = \max_{S, T \subseteq V} \frac{|e_G(S, T) - e_{G'}(S, T)|}{n^2} = \|A_G - A_{G'}\|_{\square}.$$

Note that we normalize by  $n^2$  rather than  $|S| \times |T|$  to avoid over-emphasizing small sets.

**Comparison:**  $d_{\square}(G, G') \leq d_1(G, G')$ . For two independent random graphs  $G, G' \in \mathcal{G}(n, 1/2)$ , the edit distance  $d_1$  is large ( $\approx 1/2$ ), but the cut distance is small ( $O(1/\sqrt{n})$ ). This confirms that cut distance captures "structural" similarity (random graphs look the same globally) better than edit distance.

For weighted graphs with node weights  $\alpha_i$  and edge weights  $\beta_{ij}$ , these definitions generalize naturally by summing weighted differences.

### 3.1.3 Two Graphs with the Same Number of Nodes (Unlabeled)

If  $G$  and  $G'$  have the same number of nodes  $n$  but different (or unlabeled) node sets, we define their distance by minimizing over all possible labelings (bijections between vertex sets):

$$\hat{\delta}_{\square}(G, G') = \min_{\text{labelings}} d_{\square}(\hat{G}, \hat{G}'),$$

where  $\hat{G}, \hat{G}'$  are the graphs mapped to a common set  $[n]$ .

### 3.1.4 Two Arbitrary Graphs

To define the distance between graphs  $G$  and  $G'$  of different sizes ( $n$  and  $n'$ ), we have two equivalent approaches:

**1. Blow-up Method:** For a graph  $G$  and integer  $m$ , let  $G(m)$  be the graph obtained by replacing each node of  $G$  with  $m$  nodes (cloning). We can make the sizes compatible by blowing up  $G$  by  $n'$  and  $G'$  by  $n$  (or multiples thereof).

$$\delta_{\square}(G, G') = \lim_{k \rightarrow \infty} \hat{\delta}_{\square}(G(kn'), G'(kn)).$$

**2. Fractional Overlays:** A more direct definition uses *fractional overlays*. A fractional overlay of  $G$  and  $G'$  is a nonnegative  $n \times n'$  matrix  $X$  such that row sums are  $1/n$  and column sums are  $1/n'$  (a probability coupling with uniform marginals). The cut distance relative to an overlay  $X$  is:

$$d_{\square}(G, G', X) = \max_{Q, R \subseteq V \times V'} \left| \sum_{iu \in Q, jv \in R} X_{iu} X_{jv} (\mathbb{1}(ij \in E) - \mathbb{1}(uv \in E')) \right|.$$

Then the cut distance is the minimum over all valid overlays:

$$\delta_{\square}(G, G') = \min_{X \in \mathcal{X}(G, G')} d_{\square}(G, G', X).$$

**Properties:**

- $\delta_{\square}$  is a **pseudometric**:  $\delta_{\square}(G, G') = 0$  does not imply  $G \cong G'$  (e.g.,  $G$  and its blow-up  $G(k)$  have distance 0).
- $\delta_{\square}(G, G') \leq \hat{\delta}_{\square}(G, G')$ . The reverse inequality is weak (logarithmic), meaning convergence in one implies convergence in the other.

### 3.2 Cut Norm and Cut Distance of Kernels (Incomplete, but enough for our discussion)

The definitions extend cleanly to the space  $\mathcal{W}$  of kernels (graphons).

**Definition 0.18 (Cut Norm for Kernels).** For a kernel  $W \in \mathcal{W}$ , the cut norm is:

$$\|W\|_{\square} = \sup_{S, T \subseteq [0,1]} \left| \int_{S \times T} W(x, y) dx dy \right|.$$

This norm satisfies similar inequalities:  $\|W\|_{\square} \leq \|W\|_1 \leq \|W\|_2 \leq \|W\|_{\infty} \leq 1$ . The corresponding metric is  $d_{\square}(U, W) = \|U - W\|_{\square}$ .

**Relation to Step Functions:** For a step function  $U$  with  $k$  steps, we have  $\|U\|_1 \leq k^2 \|U\|_{\square}$  (refined to  $\sqrt{2k}$ ). This is crucial for approximation arguments.

We also define the  $\epsilon$ -neighborhood in the cut norm:

$$B_{\square}(\mathcal{R}, \epsilon) = \{W \in \mathcal{W}_0 : \exists U \in \mathcal{R} \text{ such that } d_{\square}(W, U) < \epsilon\}.$$

## 4 Szemerédi Partition and Regularity lemma

This is a very important tool in understanding large dense graphs. To state and understand it we will require some preliminary definitions and results.

### 4.1 Basic notation

**Definition 0.19.** Let  $G = (V, E)$  be a simple graph on  $n = |V|$  vertices.

- For  $X, Y \subseteq V$ , let  $e_G(X, Y)$  be the number of *cut edges* between  $X$  and  $Y$ . If  $X \cap Y \neq \emptyset$ , edges with both endpoints in  $X \cap Y$  are counted twice (once for each endpoint).
- The *edge density* between  $X$  and  $Y$  is

$$d_G(X, Y) := \frac{e_G(X, Y)}{|X||Y|}.$$

- If  $X, Y$  are disjoint,  $G[X, Y]$  is the bipartite graph obtained by only keeping the cut edges between  $X$  and  $Y$ .
- A *partition* of  $V$  is  $\mathcal{P} = \{V_1, \dots, V_k\}$ .

**Definition 0.20.** Quotient/averaged graph  $G_{\mathcal{P}}$  and its template on  $[k]$

Given a partition  $\mathcal{P} = \{V_1, \dots, V_k\}$ , define the *averaged* (weighted) graph  $G_{\mathcal{P}}$  on  $V$  by putting on each pair  $u \in V_i, v \in V_j$  the weight

$$(G_{\mathcal{P}})_{uv} := d_G(V_i, V_j).$$

Equivalently,  $G_{\mathcal{P}}$  is the “matrix averaging” of  $G$  over the blocks  $V_i \times V_j$ . One can also form the *template graph*  $G/\mathcal{P}$  on the vertex set  $[k]$ , where node  $i$  is weighted by  $|V_i|$  and the edge weight  $\{i, j\}$  is  $d_G(V_i, V_j)$ .

*Roughly speaking*, the regularity lemma says that the node set of every graph has an equitable partition  $\mathcal{P}$  into a relatively small number of classes such that  $G_{\mathcal{P}}$  is “close” to  $G$ . Depending what we meant for the notion of “close” here, we can derive various versions of this lemma.

To state this more formally, we quantify the notion of how random-like a graph is, or in a sense, very “homogeneous”.

### 4.2 $\epsilon$ -homogeneous bipartite pairs

**Definition 0.21.** Let  $U, W \subseteq V$  be disjoint. The bipartite graph  $G[U, W]$  is  $\epsilon$ -homogeneous if

$$|e_G(X, Y) - d_G(U, W)|X||Y|| \leq \epsilon |U||W|$$

(\*)

Equivalently a stronger condition with the restriction on the size of  $X, Y$  is called  $\epsilon$ -regular,

$$|d_G(X, Y) - d_G(U, W)| \leq \epsilon \quad \text{for all } X \subseteq U, Y \subseteq W, |X| \geq \epsilon|U|, |Y| \geq \epsilon|W|.$$

**Remark 0.1.** A remark on why we require a lower bound on the size.

If one tried to require the inequality for *all*  $X, Y$  without the size lower bounds, it would be too strong and would fail for very small  $X', Y'$ —hence the standard definition above.

The proof of this is non-trivial, a careful reader should try this. The hint is apply the remark in the book and prove that the choice of  $\epsilon$  is then made not arbitrary, thus a contradiction.

**Remark 0.2.** Another *nice-to-have* picture in mind:

Let  $H$  be the complete bigraph on  $(U, W)$  such that every edge has weight  $d_G(U, W)$ . Then  $e_H(X, Y) = d_G(U, W)|X||Y|$

Also recall the definition of cut distances.

### 4.3 Cut distance for unweighted graphs

Write  $A_G$  for the adjacency matrix of  $G$ , and set

$$d_{\square}(G_1, G_2) := \frac{1}{n^2} \|A_{G_1} - A_{G_2}\|_{\square} = \frac{1}{n^2} \max_{S, T \subseteq V} |e_{G_1}(S, T) - e_{G_2}(S, T)|.$$

This is the normalized cut distance. For a partition  $\mathcal{P}$ ,  $G_{\mathcal{P}}$  is the block-constant average of  $G$ .

### 4.4 Weak regularity lemma, almost original form

Now we are finally ready to state the Regularity Lemma.

**Theorem 0.1.** For every  $\epsilon > 0$ , there is an  $S(\epsilon) \in \mathbb{N}$  such that every graph  $G$  has an equitable partition into  $k$  classes, with  $\frac{1}{\epsilon} \leq k \leq S(\epsilon)$ , such that for all but  $\epsilon k^2$  pairs of indices, the bipartite graph between the pair of indices is  $\epsilon$ -homogeneous.

**Remark 0.3.** Something to notice about this theorem.

The problem of this version of the lemma is that the upper bound  $S$  is independent from the graph  $G$ .

The lower bound for number of classes bound from above the number of edges inside each class.

The only bad pairs are also bounded by same upper bound. But then this means all the edges inside these classes are all kind of error term, so the partitioned graph is very similar to the original graph in this sense.

### 4.5 Weak regularity for graphs

We want to have a more reasonable error bound. The following version addressed the problem.

**Theorem 0.2.** For every integer  $k > 1$  and every graph  $G = (V, E)$ , there exists a partition of  $V$  into  $k$  classes such that

$$d_{\square}(G, G_{\mathcal{P}}) \leq \frac{2}{\sqrt{\log k}}.$$

**Remark 0.4.** Notice how we didn't require equitable partition here. It's actually similar, however, and we just need to increase the error bound to  $\frac{4}{\sqrt{\log k}}$

**Remark 0.5.** Intuition:

the more classes you take, the closer  $G_{\mathcal{P}}$  is to  $G$  in cut distance.

## 4.6 Relation between the two version

Below I want to quickly show you how the two versions above relates to each other nicely. I will give a sketch of the proof and you can try to fill in the details.

*The explanation in the book is not that good in my opinion, and some sentences are too concise such that they are quite misleading.*

**Proof sketch** A first intuition we should have before dive into the proof is that, due to the nicer error bound, the second version should have a weaker restriction on the obtained partition, maybe a looser bound on the cut distance to the original graph.

Therefore consider the "difference" graph  $F = G - G_{\mathcal{P}}$ , where the edge weights are defined as  $e_F = e_G - e_{G_{\mathcal{P}}}$ . We want to bound the cut distance:

$$d_{\square}(G, G_{\mathcal{P}}) = \frac{1}{n^2} \max_{S, T \subseteq V} |e_F(S, T)|$$

Fix any two subsets  $S, T \subseteq V$ . Let  $\mathcal{P} = \{V_1, \dots, V_k\}$  be the  $\epsilon$ -regular partition. We can decompose the edges in  $F$  between  $S$  and  $T$  by summing over the pairs of partition classes:

$$e_F(S, T) = \sum_{i, j} e_F(S \cap V_i, T \cap V_j)$$

We split this sum into three parts: edges inside classes ( $i = j$ ), edges between "bad" pairs, and edges between "good" pairs.

1. **Inside classes ( $i = j$ ):** Assuming an equitable partition where  $|V_i| \leq \lceil n/k \rceil$ , the number of edges inside classes is bounded by the sum of squares of the sizes:

$$\sum_i |V_i|^2 \leq k \left( \frac{n}{k} + 1 \right)^2 \approx \frac{n^2}{k} \leq \epsilon n^2$$

(assuming  $k$  is large enough such that  $1/k \leq \epsilon$ ).

2. **Bad pairs:** By the definition of the Szemerédi partition, there are at most  $\epsilon k^2$  bad pairs. The contribution of these pairs is trivially bounded by the total number of vertex pairs between them:

$$\sum_{(i, j) \in \text{bad}} |e_F(S \cap V_i, T \cap V_j)| \leq \sum_{(i, j) \in \text{bad}} |V_i| |V_j| \leq \epsilon k^2 \left( \frac{n}{k} \right)^2 = \epsilon n^2$$

3. **Good pairs:** For a good pair  $(V_i, V_j)$ , the graph is  $\epsilon$ -regular. Let  $X = S \cap V_i$  and  $Y = T \cap V_j$ . By the definition of  $\epsilon$ -regularity:

$$|e_G(X, Y) - d(V_i, V_j)|X||Y|| \leq \epsilon |V_i| |V_j|$$

Note that  $e_{G_{\mathcal{P}}}(X, Y) = d(V_i, V_j)|X||Y|$  by construction. Thus,  $|e_F(X, Y)| \leq \epsilon |V_i| |V_j|$ . Summing over all good pairs:

$$\sum_{(i, j) \in \text{good}} |e_F(S \cap V_i, T \cap V_j)| \leq \sum_{(i, j) \in \text{good}} \epsilon |V_i| |V_j| \leq \epsilon n^2$$

Combining these three bounds:

$$|e_F(S, T)| \leq \epsilon n^2(\text{inside}) + \epsilon n^2(\text{bad}) + \epsilon n^2(\text{good}) \approx 3\epsilon n^2$$

Dividing by  $n^2$ , we obtain the bound on the cut distance:

$$d_{\square}(G, G_{\mathcal{P}}) \leq 3\epsilon$$

#### 4.7 A strong (Frieze–Kannan style) refinement statement

Another version, which is called the strong regularity lemma, strengthens the conclusion, however giving an even larger error bound.

A convenient strengthening that packages the above argument is:

**Lemma 0.3** (Strong refinement/approximation; “very strong”). For every  $\epsilon > 0$  and every partition  $\mathcal{P}$  of  $V$ , there exists a refinement  $\mathcal{Q}$  of  $\mathcal{P}$  with  $|\mathcal{Q}| \leq S(\epsilon)|\mathcal{P}|$  such that, for every pair of unions  $S, T$  of  $\mathcal{Q}$ -classes (except at most  $\epsilon|\mathcal{P}|^2$  exceptional pairs of  $\mathcal{P}$ -classes), we have

$$\left| \frac{e_G(S, T)}{|S||T|} - \frac{e_{G_{\mathcal{P}}}(S, T)}{|S||T|} \right| \leq \epsilon.$$

Consequently, taking  $\mathcal{P}$  equitable and combining with  $d_{\square}(G, G_{\mathcal{P}}) \leq \epsilon$  yields Szemerédi’s lemma with  $k \leq S(\epsilon)$  and at most  $\epsilon k^2$  exceptions.

#### Notes.

- The “refinement  $\mathcal{Q}$  is  $\epsilon$ -close to  $\mathcal{P}$ ” phrasing means: for all but  $\epsilon|\mathcal{P}|^2$  pairs of  $\mathcal{P}$ -classes, every union  $S$  of  $\mathcal{Q}$ -classes contained in one side and every union  $T$  on the other side preserve the average densities up to  $\epsilon$  after normalizing by  $|S||T|$ .
- Inside-edges and exceptional classes can be handled together by the crude bounds earlier (e.g.,  $\sum_i \binom{|V_i|}{2} \leq n^2/(2k) + n/2$  and choosing  $k$  large).

*Summary.* Weak regularity (via cut-norm approximation  $G \approx G_{\mathcal{P}}$ ) gives  $d_{\square}(G, G_{\mathcal{P}}) \lesssim 1/\sqrt{\log k}$ . For  $k$  large, this forces that, aside from at most  $\epsilon k^2$  pairs, the bipartite pieces  $G[V_i, V_j]$  are  $\epsilon$ -homogeneous. Choosing  $k \leq S(\epsilon)$  and taking  $\mathcal{P}$  equitable yields the common form of Szemerédi’s Regularity Lemma.

**Remark 0.6.** Notice the book also discusses how to extend this to kernels, and has proofs for it. But we’ll soon do a version of proof that’s very similar to it.

## 4.8 Proof of Weak Regularity Lemma, geometric form

Following the proofs in Lovász's *Graph and Geometry* book (Lemma 9.7), we will state and prove another version of the weak regularity lemma.

Start with some preparations.

### 4.8.1 Preliminaries and Notation

Let  $M \in \mathbb{R}^{n \times n}$ . For  $S, T \subseteq [n]$ , write  $1_{S \times T}$  for the indicator matrix,

$$(1_{S \times T})_{ij} = \begin{cases} 1, & i \in S, j \in T, \\ 0, & \text{otherwise.} \end{cases}$$

Define

$$M(S, T) = \sum_{i \in S, j \in T} M_{ij} = \langle M, 1_{S \times T} \rangle.$$

The Frobenius (Euclidean) norm is

$$\|M\|_F = \left( \sum_{i,j=1}^n M_{ij}^2 \right)^{1/2}.$$

The cut norm is

$$\|M\|_{\square} = \max_{S, T \subseteq [n]} |M(S, T)| = \max_{S, T} |\langle M, 1_{S \times T} \rangle|.$$

(When we minimize quadratic expressions below, it is convenient to think of the *normalized* version where the drop per step is  $\|M\|_{\square}^2$ ; the algebra below reflects this and is standard in weak regularity arguments.)

**Averaging operator.** For a partition  $\mathcal{P} = \{V_1, \dots, V_k\}$  of  $[n]$ , define for  $x \in \mathbb{R}^n$  and  $M \in \mathbb{R}^{n \times n}$ :

$$(x_{\mathcal{P}})_u = \frac{x(V_i)}{|V_i|} \quad (u \in V_i), \quad (M_{\mathcal{P}})_{uv} = \frac{M(V_i, V_j)}{|V_i||V_j|} \quad (u \in V_i, v \in V_j).$$

Then  $M \mapsto M_{\mathcal{P}}$  is the orthogonal projection (in Frobenius inner product) onto the linear subspace of  $\mathcal{P}$ -matrices (matrices constant on each block  $V_i \times V_j$ ). In particular,

$$\|M_{\mathcal{P}}\|_F \leq \|M\|_F, \quad \|M_{\mathcal{P}}\|_{\square} \leq \|M\|_{\square}.$$

### 4.8.2 Lemma 9.7 (Matrix Version)

**Lemma 0.4.** Let  $M \in \mathbb{R}^{n \times n}$ .

(a) For every  $r \geq 1$  there exist sets  $S_1, \dots, S_r, T_1, \dots, T_r \subseteq [n]$  and reals  $a_1, \dots, a_r$  such that

$$\left\| M - \sum_{i=1}^r a_i 1_{S_i \times T_i} \right\|_{\square} \leq \frac{1}{\sqrt{r}} \|M\|_F.$$

(b) For every  $k \geq 2$  there is a partition  $\mathcal{P}$  of  $[n]$  into  $k$  classes with

$$\|M - M_{\mathcal{P}}\|_{\square} \leq \frac{4}{\sqrt{\log k}} \|M\|_F.$$

(c) For every  $k \geq 2$  there is a partition  $\mathcal{P}$  of  $[n]$  into  $k$  classes and a  $\mathcal{P}$ -matrix  $B$  such that

$$\|M - B\|_{\square} \leq \frac{2}{\sqrt{\log k}} \|M\|_F.$$

*Proof of (a).* Fix nonempty  $S, T$  with  $|M(S, T)| = \|M\|_{\square}$ . For any real  $a$ ,

$$\begin{aligned} \|M - a \mathbf{1}_{S \times T}\|_F^2 &= \|M\|_F^2 + a^2 \|\mathbf{1}_{S \times T}\|_F^2 - 2a \langle M, \mathbf{1}_{S \times T} \rangle \\ &= \|M\|_F^2 + a^2 |S||T| - 2a M(S, T). \end{aligned}$$

This quadratic in  $a$  is minimized at  $a = M(S, T)/(|S||T|)$ , giving

$$\|M - a \mathbf{1}_{S \times T}\|_F^2 = \|M\|_F^2 - \frac{M(S, T)^2}{|S||T|} \leq \|M\|_F^2 - \|M\|_{\square}^2 \quad (\text{normalized view}). \quad (1)$$

Iterate greedily. Let  $M_0 := M$  and for  $j = 0, 1, \dots, r-1$  choose  $S_{j+1}, T_{j+1}$  attaining  $\|M_j\|_{\square}$  and set

$$a_{j+1} := \frac{M_j(S_{j+1}, T_{j+1})}{|S_{j+1}||T_{j+1}|}, \quad M_{j+1} := M_j - a_{j+1} \mathbf{1}_{S_{j+1} \times T_{j+1}}.$$

By (1),

$$\|M_{j+1}\|_F^2 \leq \|M_j\|_F^2 - \|M_j\|_{\square}^2.$$

Summing for  $j = 0$  to  $r-1$  yields

$$\sum_{j=0}^{r-1} \|M_j\|_{\square}^2 \leq \|M\|_F^2 - \|M_r\|_F^2.$$

We get

$$\|M_r\|_F^2 \leq \|M\|_F^2 - \sum_{j=0}^{r-1} \|M_j\|_{\square}^2.$$

Since the LHS is nonnegative, by averaging there exists  $j \in \{0, \dots, r-1\}$  with

$$\|M_j\|_{\square} \leq \frac{1}{\sqrt{r}} \|M\|_F.$$

Finally,

$$M - \sum_{i=1}^r a_i \mathbf{1}_{S_i \times T_i} = M_j$$

after setting  $a_{j+1} = \dots = a_r = 0$ , which proves (a).  $\square$

*Proof of (c).* Take

$$r = \left\lfloor \frac{\log k}{2} \right\rfloor, \quad B = \sum_{i=1}^r a_i \mathbf{1}_{S_i \times T_i}$$

from part (a). Construct a partition  $\mathcal{P}$  of  $[n]$  by grouping vertices that share the same membership pattern across the  $2r$  sets  $S_1, \dots, S_r, T_1, \dots, T_r$ :

$$u \sim v \iff (\forall i \leq r) [u \in S_i] = [v \in S_i] \text{ and } (\forall i \leq r) [u \in T_i] = [v \in T_i].$$

The equivalence classes form  $\mathcal{P}$ , and there are at most  $2^{2r} \leq k$  classes.

For any two parts  $A, A' \in \mathcal{P}$  and any  $(u, v) \in A \times A'$ , the value  $1_{S_i \times T_i}(u, v)$  depends only on whether  $A \subseteq S_i$  and  $A' \subseteq T_i$ , so each  $1_{S_i \times T_i}$  is constant on  $A \times A'$ . Therefore  $B$  is a  $\mathcal{P}$ -matrix. From part (a),

$$\|M - B\|_{\square} \leq \frac{1}{\sqrt{r}} \|M\|_F \leq \frac{2}{\sqrt{\log k}} \|M\|_F,$$

which proves (c). □

*Proof of (b).* This follows from part (c). Let  $\mathcal{P}$  be the partition and  $B$  be the  $\mathcal{P}$ -matrix obtained in part (c). Since  $B$  is a  $\mathcal{P}$ -matrix (constant on the blocks of  $\mathcal{P}$ ), applying the averaging operator to it leaves it unchanged, so  $B_{\mathcal{P}} = B$ .

We use the triangle inequality and the contractive property of the averaging operator with respect to the cut norm (i.e.,  $\|X_{\mathcal{P}}\|_{\square} \leq \|X\|_{\square}$ ).

$$\begin{aligned} \|M - M_{\mathcal{P}}\|_{\square} &= \|(M - B) + (B - M_{\mathcal{P}})\|_{\square} \\ &\leq \|M - B\|_{\square} + \|B - M_{\mathcal{P}}\|_{\square}. \end{aligned}$$

Substitute  $B = B_{\mathcal{P}}$  into the second term:

$$\|B - M_{\mathcal{P}}\|_{\square} = \|B_{\mathcal{P}} - M_{\mathcal{P}}\|_{\square} = \|(B - M)_{\mathcal{P}}\|_{\square}.$$

By the contractive property (applied to  $X = B - M$ ):

$$\|(B - M)_{\mathcal{P}}\|_{\square} \leq \|B - M\|_{\square} = \|M - B\|_{\square}.$$

Combining these inequalities:

$$\|M - M_{\mathcal{P}}\|_{\square} \leq \|M - B\|_{\square} + \|M - B\|_{\square} = 2\|M - B\|_{\square}.$$

Finally, substituting the bound from part (c),  $\|M - B\|_{\square} \leq \frac{2}{\sqrt{\log k}} \|M\|_F$ , we obtain:

$$\|M - M_{\mathcal{P}}\|_{\square} \leq 2 \left( \frac{2}{\sqrt{\log k}} \|M\|_F \right) = \frac{4}{\sqrt{\log k}} \|M\|_F.$$

□

Next we look at another form of weak regularity lemma in the hilbert space.

## 4.9 A Regularity Lemma in Hilbert Space

**Lemma 0.5** (Regularity Lemma in Hilbert Space). Let  $K_1, K_2, \dots$  be arbitrary nonempty subsets of a (real) Hilbert space  $\mathcal{H}$ . For every  $\varepsilon > 0$  and  $f \in \mathcal{H}$  there exists  $m \leq \lceil 1/\varepsilon^2 \rceil$ , elements  $f_i \in K_i$  ( $1 \leq i \leq m$ ), and scalars  $\gamma_1, \dots, \gamma_m \in \mathbb{R}$  such that for the residual  $r = f - (\gamma_1 f_1 + \dots + \gamma_m f_m)$  we have

$$|\langle g, r \rangle| \leq \varepsilon \|g\| \|f\| \quad \text{for every } g \in K_{m+1}.$$

*Proof.* We use only inner-product identities that also hold in any real inner-product space:

1.  $\langle x, y \rangle = \langle y, x \rangle$  and  $\|x\|^2 = \langle x, x \rangle$ .
2. Cauchy–Schwarz:  $|\langle x, y \rangle| \leq \|x\| \|y\|$ .

3. Line expansion: for all  $u, v, g$  and  $t \in \mathbb{R}$ ,

$$\|u - (v + tg)\|^2 = \|u - v\|^2 + t^2\|g\|^2 - 2t\langle g, u - v \rangle.$$

Let

$$\eta_k = \inf_{\gamma_i \in \mathbb{R}, f_i \in K_i} \left\| f - \sum_{i=1}^k \gamma_i f_i \right\|^2,$$

the best (squared) error after  $k$  terms. Then  $0 \leq \eta_{k+1} \leq \eta_k \leq \|f\|^2$  and hence there exists  $m \leq \lceil 1/\varepsilon^2 \rceil$  with

$$\eta_m < \eta_{m+1} + \varepsilon^2 \|f\|^2.$$

Choose coefficients  $\gamma_1, \dots, \gamma_m$  and  $f_i \in K_i$  so that, writing  $f^* = \sum_{i=1}^m \gamma_i f_i$ ,

$$\|f - f^*\|^2 \leq \eta_m + \varepsilon^2 \|f\|^2.$$

Let  $r = f - f^*$ . Fix  $g \in K_{m+1}$ . For any  $t \in \mathbb{R}$ ,

$$\|f - (f^* + tg)\|^2 \geq \eta_{m+1} > \eta_m - \varepsilon^2 \|f\|^2 \geq \|r\|^2 - \varepsilon^2 \|f\|^2.$$

Expanding along the line through  $f^* + tg$  gives

$$\|r - tg\|^2 = \|r\|^2 + t^2\|g\|^2 - 2t\langle g, r \rangle \geq \|r\|^2 - \varepsilon^2 \|f\|^2 \quad \forall t \in \mathbb{R}.$$

Thus the quadratic

$$q(t) = \|g\|^2 t^2 - 2\langle g, r \rangle t + \varepsilon^2 \|f\|^2$$

is nonnegative for all  $t$ , so its discriminant is  $\leq 0$ , i.e.

$$(2\langle g, r \rangle)^2 - 4\|g\|^2 \varepsilon^2 \|f\|^2 \leq 0 \implies |\langle g, r \rangle| \leq \varepsilon \|g\| \|f\|.$$

Since  $g \in K_{m+1}$  was arbitrary, the lemma follows.  $\square$

This finishes our discussion on the introduction of regularity lemma. We will see some interesting applications of it, but before that we want to go back to our topic of graphon, and prepare to define it formally as the limit object.

## 4.10 Compactness of the graphon space

In this section we prove a result that is equivalent to the regularity lemma, that *the graphon space is compact under cut distance*. One can verify this by deriving the theorem from the weak regularity lemma, and then derive the strong regularity lemma from it. The exact way to do this is tedious, and readers can refer to the book for the proof.

**Theorem 0.6.** Compactness of graphon space  
The space  $(W_0, \delta_\square)$  is compact.

The proof shown in the book is really too concise and requires a lot of prerequisite knowledge to actually follow, but it's worth the read in my opinion, since it provides some insight into the mechanism of the graphon space, and the method is very useful in a lot of similar proofs. Here I will try to give a detailed explanation of the proof steps, but may not be very precise on notations. *sorry*. Readers can try to get the intuition from my sketch first, and it will feel much easier to follow along the proof in the book.

*Proof.* A first intuition that may help is that *cut distance* categorizes equivalent classes of graphons, and compactness of graphon space implies that graph sets are "dense" in cut distance.

- Regularize

For every  $n > 1$ , apply weak regularity lemma to obtain a sequence of partitions  $P_{n,k}$  (This is equivalent to lemma 9.15 mentioned in the book) such that

- The partition  $P_{n,k+1}$  refines  $P_{n,k}$ .
- $|P_{n,k}| = m_k$  depends only on  $k$ .
- $\|W_n - W_{n,k}\|_\square \leq 1/k$ , where  $W_{n,k} = (W_n)_{P_{n,k}}$  (as  $k$  grows, they get closer)

- Then consider the following:

- For each  $k, i \in [m_k]$ , the measure of part  $i$  in Partition  $P_{n,k}$  converges to some value  $\alpha_{k,i}$  as  $n$  goes to infinity.
- For  $k, i, j \in [m_k]$ , the value of  $(W_n)_{P_{n,k}}$  on the product of the  $i^{th}$  and the  $j^{th}$  parts of  $P_{n,k}$  converges to value  $\beta_{k,i,j}$ , as  $n$  goes to infinity.
- Construct the above described limit object (or subsequence) for each  $k$ , as  $n$  goes to infinity along it. More precisely,

Let  $P_k = \{I_{k,1}, \dots, I_{k,m_k}\}$ , which is a partition of  $[0, 1]$  into intervals with length  $\lambda(I_{k,i}) = \alpha_{k,i}$  for  $i \in [m_k]$ .

Let  $U_k$  denote the stepfunction with steps  $P_k$ , values on  $I_{k,i} \times I_{k,j} = \beta_{k,i,j}$  for  $i, j \in [m_k]$ .

Then for each  $k$ , we have  $\delta_\square(W_{n,k}, U_k) \rightarrow 0$  as  $n$  goes to infinity. (\*).

And, for each  $k$ , since  $W_{n,k} = (W_{n,k+1})_{P_{n,k}}$  (Think of this as degeneration of partitions, like applying the original partition on the refined partiton will just give you the original partiton again).

- Consider  $U_k$ . We have

$$U_k = (U_{k+1})_{P_k}$$

By the definition of martingales,  $U_n$  sequence is a martingale, i.e.,

$$E[U_{k+1}(X, Y) | U_1, \dots, U_k] = U_k(X, Y)$$

By the *Martingale Convergence Theorem*,  $U_k$  is bounded and converges almost surely.

Therefore, there exists a limit graphon such that  $U_k \rightarrow U$  pointwise almost everywhere as  $k$  goes to infinity.

- We claim that the sequence of  $W_n$  converges to  $U$  in cut metric, and thus finish our proof. (Notice here that the  $W_n$  sequence is relabeled subsequence of the original sequence, see the book for a detail).

Pick  $\epsilon > 0$ . There exists some  $k > \frac{3}{\epsilon}$  such that  $\|U - U_k\|_{\square} < \frac{\epsilon}{3}$  for  $n$  big enough. (This from the property of the partition we obtained)

But since  $k > \frac{3}{\epsilon}$ , we have  $\delta_{\square}(W_{n,k}, W_n) < \frac{\epsilon}{3}$  for all  $n$ .

Thus  $\delta_{\square}(U, W_n) \leq \frac{\epsilon}{3} \cdot 3 = \epsilon$ .

Therefore, the subsequence  $W_n$  converges to  $U$  in cut metric.

□

This result is very important when we later consider graphon as a limit object more formally.

## 5 Graph Sampling

Recall the main purpose of this writing is to introduce the readers to the concept of graphon, and apply it to applications like regularity lemma to see the power of it.

To define formally graphon as the limit object of graphs we will need three tools: graph homomorphism (densities), notion of cut distance, and graph sampling, most importantly, *the counting lemma*.

The book has a chapter on  $W$  – random graph that’s good to have too.

### 5.1 Counting Lemma

Counting lemma relates homomorphism density and the cut distance together.

**Lemma 0.7.** Counting lemma for graphs

For any three simple graphs  $F, G, G'$ , we have

$$|t(F, G) - t(F, G')| \leq e(F)\delta_{\square}(G, G').$$

This extends to graphons.

**Lemma 0.8.** Counting lemma for graphons

Let  $F$  be a simple graph and let  $W, W'$  be graphons. Then we have,

$$|t(F, W) - t(F, W')| \leq e(F)\delta_{\square}(W, W').$$

Recall the definition of a  $W$  – decorated graph, which is a simple graph in which a graphon  $W_e$  is assigned to each edge  $e$ .

**Lemma 0.9.** Counting lemma for decorated graphs

Let  $(F, w)$  and  $(F, w')$  be two  $W$ -decorated graphs with the same underlying simple graph, where  $w = (W_e : e \in E)$  and  $w' = (W'_e : e \in E)$ . Then,

$$|t(F, w) - t(F, w')| \leq \sum_{e \in E(F)} \|W_e - W'_e\|_{\square}.$$

This setting actually makes the proof easier.

*Proof.* It suffices to prove the case when the generating graphon are equal for but one edge.

The reasoning is as follows:

- We change the decorations *one edge at a time*, slowly modifying  $w$  to  $w'$ .

Specifically, this creates a chain of intermediate decorated graphs:

$$w^0, w^1, w^2, \dots, w^m$$

Where  $w^0$  is  $w$ , and  $w^m$  is  $w'$ , and  $w^k$  means, for the first  $k$  edges, use  $W'_{e_j}$ , and for the remaining edges, use  $W_{e_j}$ .

In this way we slowly turns  $w$  into  $w'$ , one edge decoration at a time.

- Then for  $t(F, w) - t(F, w')$ , we can use a telescoping technique, like  $t(F, w) - t(F, w^1) + t(F, w^1) - t(F, w^2) + t(F, w^2) - \dots$
- Then we take the absolute value and use triangle inequality:

$$|t(F, w) - t(F, w')| \leq \sum_{k=1}^m |t(F, w^{k-1}) - t(F, w^k)| (*)$$

- Then we should compare (\*) with the original inequality, and it should be easy to see that if we can bound each absolute value with  $|t(F, w^{k-1}) - t(F, w^k)|$  with  $\|W_{e_k} - W'_{e_k}\|_{\square}$ , we can plug this back into (\*) and we get the required formula.

Now we proceed to prove the case where the two graphons differ by only one edge.

Assume that the two decorated graphs are the same on every edge except one edge, call it  $uv$ .

We have  $W_{uv} \neq W'_{uv}$

Then recall the definition of homomorphism density in decorated graphs, and plug in the definition, we get

$$t(F, w) - t(F, w') = \int_{[0,1]^V} \prod_{ij \in E(F) \setminus \{uv\}} W_{ij}(x_i, x_j) (W_{uv}(x_u, x_v) - W'_{uv}(x_u, x_v)) dx$$

The rest of the proof is rather easy to follow from the book, so I will omit the details here. □

There is a similar proof in lemma 4.1 of Lovasz's paper <https://arxiv.org/pdf/math/0408173>

## 6 Convergence of dense graph sequences

Finally we are ready to talk about the main problem: How to define convergent graph sequences, and what is their limit.

### 6.1 What it means to say a sequence of graphs is convergent

To actually define the notion of convergence, we need to fix a sampling method. Here we use subgraph sampling: uniformly random select  $k$  vertices from the graph, and consider the induced subgraph on them. Then we consider the probability a given graph  $F$  is the induced subgraph is the quantity  $t_{ind}(F, G)$ .

**Definition 0.22 (convergent graph sequence).** A sequence of graphs  $G_n$  is convergent if the induced subgraph densities  $t_{ind}(F, G_n)$  converge for every finite graph  $F$ .

Look at explanations on the book of why  $t(F, G)$  and  $t_{ind}(F, G)$  can be used interchangeably, and the two examples to gain some intuition.

**Theorem 0.10 (convergence in cut metric).** A sequence of simple graphs  $G_n$  with  $v(G_n) \rightarrow \infty$  is *convergent* if and only if it is a Cauchy sequence in the cut metric.

The proof is obvious, since we have already proved the counting lemma, and utilize a similar result called the *inverse counting lemma* we can prove the converse.

But if the reader really wants to write down all the steps of the proof, it is out of the scope of this note, as it requires some details in the graph sampling methods. Refer to the book for details.

### 6.2 Random graphs as limit objects

*What does the above mentioned graph/graphon sequence converge to?*

If you use the computer to draw out all the graphs with adjacency matrix, a "proof by picture" will tell you that they converge to graphons. But can we describe it more formally?

Let us try to define a limit object.

#### 6.2.1 Finite random graph models(frgm)

First we consider a construction called the *weak limit*. This notion is general in the sense that it can be constructed for convergence of any reasonable sequence of structures for which we have a reasonable sampling process.

Given a simple graph  $G$  and  $k \in [v(G)]$ , the random sample  $G(k, G)$  is a random graph on  $k$  labeled nodes. We denote its distribution by  $\sigma_{G,k}$ . By this construction we should have  $\sigma_{G,k}(F) = t_{ind}(F, G)$ . If  $G_n$  is a convergent sequence, then by the previous convergence theorem we have that the distributions  $\sigma_{G_n,k}$  tend to some distribution  $\sigma_k$  on  $k$  node labeled graphs. Conversely, if the distribution sequence tends to some limit for every  $k$ , then the graph sequence is convergent. *But notice trivially that some finitely many distribution may not be defined.*

Now we have successfully modeled the problem of the limit of the convergent graph sequence into the sequence of limit distribution. Which sequence is this?

**Definition 0.23 (random graph model).** A *random graph model* (i will refer to as *rgm* in the context) is a probability distribution  $\sigma_k$  on simple graphs on node set  $[k]$ , for every positive  $k$ , which is also *invariant under the reordering of the nodes*.

In other words, the *rgm* is a sequence of random variables, whose values are simple graphs on  $[k]$ , and isomorphic graphs have the same probability.

**Definition 0.24 (rgm cont.).** We continue our discussion on *rgm*.

- We say an *rgm* is *consistent* if deleting node  $k$  from  $G_k$ , the distribution of the resulting graph is the same as the distribution of  $G_{k-1}$ . Formally,

$$\sigma_{k-1}(H) = \sum_{F:F'=H} \sigma_k(F),$$

where  $F'$  denotes the graph obtained by deleting node  $k$  from  $F$ .

- We say an *rgm* is *local*, if for two disjoint subsets  $S, T \subseteq [k]$ , the subgraphs of  $G_k$  induced by  $S$  and  $T$  are independent as random variables.

**Remark 0.7.** Trivial to check that the *rgm*  $G_k = G(k, W)$  since both consistent and local. (hint: because  $S$  and  $T$  are disjoint, the sampling is independent; because the sampling method ensures that the edges depend only on previous latent variables, the construction should be the same.)

We formalize our discussion above rigorously.

**Theorem 0.11.** If a graph sequence  $G_n$  is convergent, then the distributions  $\sigma_k = \lim_{n \rightarrow \infty} \sigma_{k, G_n}$  forms a *consistent and local random graph model*. Conversely, every *consistent and local random graph model* should arise this way.

**Remark 0.8.** The proof of this theorem is beyond the scope due to the need of introducing specific sampling method. Careful readers should refer to the book.

We have know that, when two random graphs are very large, they should be very much alike. This hints us that a sequence of independently generated random graphs is convergent almost surely. Next, we state a lemma that shows all local and consistent random graph models have this similar property.

**Lemma 0.12 (generation method 1).** Let  $\sigma_n$  be a local and consistent random graph model, and generate a graph  $G_n$  from every  $\sigma_n$ , *independently from different values of  $n$* . Then the sequence  $G_n$  is convergent with probability 1.

Later we should see another generation method for convergent graph sequence, but notice the difference.

## 6.2.2 Countable random graph models (crgm)

We consider a setup that would allow us to extend the distribution from finite to countable.

Rearrange all labeled simple graphs in a locally finite rooted tree, where the empty graph is the root, and  $F'$  is the parent of  $F$ . If  $(\sigma_1, \sigma_2, \dots)$  is a consistent sequence of distributions, then  $\sigma_k$

is a probability distribution on the  $k$ -th level of the tree, and the probability of each node is the sum of probabilities of its children.

**Definition 0.25 (countable random graph model).** We have,

A *countable random graph model* is a probability distribution  $\sigma$  on  $(\Omega, \mathcal{A})$ , invariant under permutation.

Such a random graph can also be considered as a symmetric exchangeable array of 0-1 valued random variables.

The *crgm* is *local* if for any two finite disjoint subsets, the subgraphs induced by them are independent as random variables.

**Remark 0.9.** The discussion above shows obviously that every consistent random graph model defines a countable random graph model. More formally,

**Proposition 0.13 (bijection between crgm and consistent rgm).** There is a bijection between consistent random graph models and countable random graph models. This bijection also preserves locality.

There are some interesting examples that are good for intuition in the book. Readers should try to understand them themselves. However we will work through one example since it's to our interest.

**Example 0.2 (Infinite  $W$ -random graph).** We can extend the definition of  $W$ -random graphs to get a countable random graph.

Given a graphon  $W$ , we select a sequence of independent random points  $(X_1, X_2, \dots)$  from  $[0, 1]$ , and connect  $i$  and  $j$  with probability  $W(X_i, X_j)$ .

Notice how this construction generalizes the Rado graph.

It is shown independently by Lovasz and Szegedy and Diaconis and Janson later, that we can construct a convergent graph sequence from a countable random graph model, but without the assumption of locality. We conclude this section with this proposition.

**Proposition 0.14 (generation method 2).** Let  $G$  be a random graph on  $N^*$  drawn from a countable random graph model. Let  $G[n]$ , denote the subgraph induced by  $[n]$ . Then the sequence  $(G[1], G[2], \dots)$  is convergent almost surely.

**Remark 0.10.** The proof of this requires knowledge on martingale.

**Remark 0.11.** This second generation method may look similar to generation method 1, but they are actually very different. In this construction, locality is not needed.

Unlike in method 1, the subgraphs here are not *independently generated*.

The generation method 1 almost always generate sequences with the same limit for a fixed model. However, running the generation method 2 will possibly give us sequences that converge to different limits, since the sampling is not independent.

Compare these two methods to understand how random graph models work.

## 7 Graph Regularity Method: Some Applications

In this chapter we will first briefly revisit the regularity lemma, then we will move on to quasirandom graphs and the counting lemmas. Finally, we establish the removal lemmas and extend them to applications including Roth's theorem and the Erdős-Stone-Simonovits theorem.

### 7.1 Graph Regularity Lemma Revisit

Intuitively, the vertex set of every graph can be partitioned into a bounded number of parts so that the graph looks random-like between most pairs of parts.

The definition of  $\epsilon$ -regular pairs characterizes pairs of partitions that behave very much alike "inside" the parts, of course with restriction of "zooming in" to a certain degree (view it as a grayscale image, when you zoom in to some degree, the scale doesn't change much), and this is what we meant for random-like, that the edges are somewhat evenly distributed between parts.

**Definition 0.26 (Edge density).** Refer to previous chapter for precise definition.

**Definition 0.27 ( $\epsilon$ -regular pair).** Refer to previous chapter for precise definition.

For simplicity, I will refer to  $\epsilon$ -regular pairs as "regular pairs".

### 7.2 Quasirandom Graphs

Now we will look at some concepts related to regularity. First, we take a quick look at quasirandom graphs.

**Definition 0.28 (11.37 Quasirandom Graph).** Let  $(G_n)$  be a sequence of graphs with density  $p$ . We say  $(G_n)$  is a quasirandom sequence with density  $p$  if for every simple graph  $F$ ,

$$t(F, G_n) \rightarrow p^{e(F)}.$$

In terms of graphons, this corresponds to the sequence converging to the identically- $p$  graphon, i.e.,  $G_n \rightarrow p$  (where the limit graphon is the constant function  $W \equiv p$ ).

We mentioned in the introduction the surprising fact, due to Chung, Graham, and Wilson (1989), that it is enough to require this relation for just the edge  $F = K_2$  and the 4-cycle  $F = C_4$ :

**Theorem 0.15 (11.62 Chung, Graham, Wilson).** If  $(G_n)$  is a sequence of simple graphs such that  $v(G_n) \rightarrow \infty$ ,  $t(K_2, G_n) \rightarrow p$ , and  $t(C_4, G_n) \rightarrow p^4$ , then  $(G_n)$  is quasirandom with density  $p$ .

*Proof.* Suppose that  $(G_n)$  is not quasirandom, i.e., there is a simple graph  $F$  such that  $t(F, G_n) \not\rightarrow p^{e(F)}$ . We can select a subsequence for which  $t(F, G_n) \rightarrow c \neq p^{e(F)}$ , and then we can select a convergent subsequence. Let  $W$  be its limit graphon; then  $t(K_2, W) = p$ ,  $t(C_4, W) = p^4$ , and  $W \neq p$  (as  $t(F, W) = c \neq p^{e(F)}$ ). To get a contradiction, it suffices to prove:

**Claim 11.63** If  $W$  is a kernel such that  $t(K_2, W) = p$  and  $t(C_4, W) = p^4$  for some real number  $p$ , then  $W = pJ$ .

*Proof of Claim.* We start with two applications of the Cauchy–Schwarz inequality. Recall that  $t(C_4, W)$  is the norm squared of the operator  $T_W$  (or equivalently the inner product of the "cherry" density with itself). Let  $t_{xy}(P_3, W)$  denote the density of a 2-path rooted at endpoints  $x, y$ , which is exactly  $(W \circ W)(x, y)$ .

$$t(C_4, W) = \langle W \circ W, W \circ W \rangle \geq \langle W \circ W, J \rangle^2 = \langle W \circ W, J \rangle^2.$$

(Note: Assuming a normalized space where  $\langle J, J \rangle = 1$ ). The term  $\langle W \circ W, J \rangle$  can be expanded:

$$\langle W \circ W, J \rangle = \int \left( \int W(x, z)W(z, y) dz \right) dx dy = \int \left( \int W(x, z) dx \right) \left( \int W(z, y) dy \right) dz.$$

Let  $d(z) = \int W(x, z)dx$  be the marginal degree. Then  $\langle W \circ W, J \rangle = \langle d, d \rangle$ . Applying Cauchy–Schwarz again:

$$\langle d, d \rangle \geq \langle d, 1 \rangle^2 = t(K_2, W)^2.$$

Combining these, we get:

$$t(C_4, W) \geq t(K_2, W)^4.$$

Since we have  $t(C_4, W) = p^4$  and  $t(K_2, W) = p$ , we have equality throughout. Equality in the first step implies the function  $t_{xy}(P_3, W) = (W \circ W)(x, y)$  must be constant, and by integration we see its value is  $p^2$ . This means that  $W \circ W = p^2J$ . This means that the operator  $T_W \circ T_W = T_W^2$  has a single nonzero eigenvalue  $p^2$  with eigenfunction  $\equiv 1$ . But then trivially  $T_W$  has a single eigenvalue  $\pm p$  with the same eigenfunction, i.e.,  $W \equiv p$  or  $W \equiv -p$ . The condition  $t(K_2, W) = p$  rules out the second alternative. □

□

Next theorem is a little tangent to our topic, but I thought it would be fun and insightful to see the equivalent definitions of a graphon.

**Theorem 0.16** (11.52 Equivalence). The following are equivalent to a graphon:

1. graphons up to weak isomorphism, i.e.,  $t(F, U) = t(F, W)$ .
2. A multiplicative, normalized simple graph parameter that is non-negative on signed graphs.
3. consistent and local random graph models.
4. a point in the completion of the space of finite graphs with the cut distance.

Next we see the discussion of a specific type of graph.

**Example 0.3** (Paley Graph). Let  $q$  be a prime of the form  $4k + 1$ . The Paley graph  $P_q$  is defined as follows:

- $V(P_q) = \mathbb{F}_q$ .
- Connect vertices  $u, v$  if  $u - v$  is a quadratic residue modulo  $q$  (i.e., a square in  $\mathbb{F}_q^*$ ).
- **Density:** The edges have density  $1/2$  because exactly half the nonzero numbers in a field are squares.
- **Convergence:** This sequence converges to the constant graphon  $W \equiv 1/2$ .

## 7.2.1 Multitype Quasirandom Graphs

**Definition 0.29 (11.38 Multitype Quasirandomness).** Consider a sequence  $(G_n)$  and a template graph  $H$  with  $V(H) = [q]$ .

$V(G_n)$  has a partition  $(V_1, \dots, V_q)$  such that:

- The size of each part satisfies  $|V_i| = \alpha_i(H)n + o(n)$ .
- For each edge  $ij \in E(H)$ , the bipartite graphs  $G_n[V_i, V_j]$  form a quasirandom bipartite graph sequence with edge density  $\beta_{ij}$ .

This  $G_n$  tends to a step-function graphon  $W_H$  (and vice versa).

## 7.3 Graph Counting Lemmas

Now we look at some applications of the regularity lemma. Specifically, we want to make use of the regular partition obtained from it. The counting lemmas formalize the idea that a regular pair behaves like a random bipartite graph, allowing us to estimate the number of subgraphs.

### 7.3.1 Triangle Counting Lemma

The triangle counting lemma says that given three pairwise regular pairs, we can approximate them by a random tripartite graph.

**Lemma 0.17 (2.2 Triangle Counting Lemma).** Let  $G$  be a graph and  $X, Y, Z$  be subsets of the vertices of  $G$  such that  $(X, Y)$ ,  $(Y, Z)$ , and  $(Z, X)$  are all  $\epsilon$ -regular pairs for some  $\epsilon > 0$ . If  $d(X, Y), d(Y, Z), d(Z, X) \geq 2\epsilon$ , then the number of triangles is bounded below:

$$\begin{aligned} & |\{(x, y, z) \in X \times Y \times Z : xyz \text{ is a triangle in } G\}| \\ & \geq (1 - 2\epsilon)(d(X, Y) - \epsilon)(d(X, Z) - \epsilon)(d(Y, Z) - \epsilon)|X||Y||Z|. \end{aligned}$$

To prove this, we first need a helper lemma stating that in a regular pair, most vertices have the expected degree.

**Lemma 0.18 (2.2.3 Degree Property).** Let  $(X, Y)$  be an  $\epsilon$ -regular pair with density  $d = d(X, Y)$ . Then less than  $\epsilon|X|$  vertices in  $X$  have fewer than  $(d - \epsilon)|Y|$  neighbors in  $Y$ . *Vice versa*, less than  $\epsilon|Y|$  vertices in  $Y$  have fewer than  $(d - \epsilon)|X|$  neighbors in  $X$ .

*Proof.* Let  $A \subseteq X$  be the set of vertices in  $X$  that have fewer than  $(d - \epsilon)|Y|$  neighbors in  $Y$ . Then the density  $d(A, Y)$  satisfies:

$$d(A, Y) = \frac{e(A, Y)}{|A||Y|} < \frac{|A|(d - \epsilon)|Y|}{|A||Y|} = d - \epsilon.$$

So  $d(X, Y) - d(A, Y) > \epsilon$ . Since  $(X, Y)$  is  $\epsilon$ -regular, this implies  $|A| < \epsilon|X|$ .  $\square$

*Proof of Triangle Counting Lemma.* We want to find a subset  $X' \subseteq X$  such that every node in  $X'$  has "good" degree into both  $Y$  and  $Z$ . By Lemma 2.2.3 applied to  $(X, Y)$ , there are at most  $\epsilon|X|$  vertices in  $X$  with low degree into  $Y$ . Similarly, by Lemma 2.2.3 applied to  $(X, Z)$ , there are at most  $\epsilon|X|$  vertices in  $X$  with low degree into  $Z$ .

Let  $X'$  be the set of vertices in  $X$  excluding these bad sets. Then:

$$|X'| \geq |X| - \epsilon|X| - \epsilon|X| = (1 - 2\epsilon)|X|.$$

For every  $x \in X'$ , we have:

- Degree into  $Y$ :  $|N_Y(x)| \geq (d(X, Y) - \epsilon)|Y|$ .
- Degree into  $Z$ :  $|N_Z(x)| \geq (d(X, Z) - \epsilon)|Z|$ .

Since  $d(X, Y) \geq 2\epsilon$ , we have  $|N_Y(x)| \geq \epsilon|Y|$ . Similarly  $|N_Z(x)| \geq \epsilon|Z|$ . Because  $(Y, Z)$  is  $\epsilon$ -regular and these neighborhoods are large enough (at least an  $\epsilon$ -fraction), the density between them is close to  $d(Y, Z)$ . Specifically:

$$d(N_Y(x), N_Z(x)) \geq d(Y, Z) - \epsilon.$$

So the number of edges between  $N_Y(x)$  and  $N_Z(x)$  is:

$$e(N_Y(x), N_Z(x)) \geq (d(Y, Z) - \epsilon)|N_Y(x)||N_Z(x)|.$$

Each such edge  $yz$  (where  $y \in N_Y(x), z \in N_Z(x)$ ) forms a triangle  $xyz$  with  $x$ . Summing over all  $x \in X'$ :

$$\begin{aligned} \# \text{Triangles} &= \sum_{x \in X'} e(N_Y(x), N_Z(x)) \\ &\geq \sum_{x \in X'} (d(Y, Z) - \epsilon)(d(X, Y) - \epsilon)|Y|(d(X, Z) - \epsilon)|Z| \\ &= |X'| (d(Y, Z) - \epsilon)(d(X, Y) - \epsilon)(d(X, Z) - \epsilon)|Y||Z| \\ &\geq (1 - 2\epsilon)|X|(d(X, Y) - \epsilon)(d(X, Z) - \epsilon)(d(Y, Z) - \epsilon)|Y||Z|. \end{aligned}$$

□

### 7.3.2 $K_4$ Counting Lemma

This generalizes the triangle counting argument. We need a slightly stronger version of the degree lemma first.

**Lemma 0.19** (Generalized Degree Property). Let  $(X, Y)$  be  $\epsilon$ -regular. Let  $B \subseteq Y$  such that  $|B| \geq \epsilon|Y|$ . Then the number of vertices in  $X$  with fewer than  $(d(X, Y) - \epsilon)|B|$  neighbors in  $B$  is less than  $\epsilon|X|$ .

**Proposition 0.20** (2.6.1  $K_4$  Counting Lemma). Let  $0 < \epsilon < 1$ . Let  $X_1, X_2, X_3, X_4$  be vertex subsets of  $G$  such that every pair  $(X_i, X_j)$  is  $\epsilon$ -regular with edge density  $d_{ij} \geq 3\sqrt{\epsilon}$  (or sufficiently larger than  $\epsilon$ ). Then the number of quadruples  $(x_1, x_2, x_3, x_4)$  forming a  $K_4$  is at least:

$$(1 - 3\epsilon)(d_{12} - 3\epsilon)(d_{13} - \epsilon)(d_{14} - \epsilon)(d_{23} - \epsilon)(d_{24} - \epsilon)(d_{34} - \epsilon)|X_1||X_2||X_3||X_4|.$$

*Proof.* We proceed by picking "good" vertices one by one.

**Step 1: Pick good  $x_1 \in X_1$ .** We want  $x_1$  to have large neighborhoods in  $X_2, X_3, X_4$ . Let  $X'_1$  be the set of vertices in  $X_1$  that satisfy:

$$|N_{X_i}(x_1)| \geq (d_{1i} - \epsilon)|X_i| \quad \text{for } i = 2, 3, 4.$$

By the simple Degree Property, the number of bad vertices for each  $i$  is  $< \epsilon|X_i|$ . Thus:

$$|X'_1| \geq (1 - 3\epsilon)|X_1|.$$

Fix a choice of  $x_1 \in X'_1$ . Let  $Y_i = N_{X_i}(x_1)$  be the neighbors of  $x_1$  in  $X_i$ . Note that  $|Y_i| \geq (d_{1i} - \epsilon)|X_i|$ .

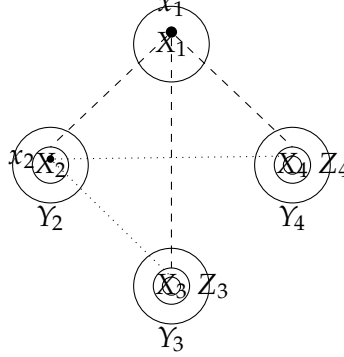


Figure 1: The slicing process: First fix  $x_1$  to get neighborhoods  $Y_i$ . Then inside  $Y_2$ , fix  $x_2$  to get neighborhoods  $Z_i$  inside  $Y_i$ .

**Step 2: Pick good  $x_2 \in Y_2$ .** Now we look inside  $Y_2$ . We want  $x_2 \in Y_2$  to have large neighborhoods in  $Y_3$  and  $Y_4$ . We use the Generalized Degree Property. The pair  $(X_2, X_3)$  is  $\epsilon$ -regular, and  $Y_3 \subseteq X_3$  is large enough (since  $|Y_3| \geq \epsilon|X_3|$ ). Thus, the number of vertices in  $X_2$  with few neighbors in  $Y_3$  is small. Specifically, the "bad" vertices in  $X_2$  are few. We focus on  $Y_2$ . Let  $X'_2$  be the subset of  $Y_2$  such that for every  $x_2 \in X'_2$ :

$$|N_{Y_i}(x_2)| \geq (d_{2i} - \epsilon)|Y_i| \quad \text{for } i = 3, 4.$$

The number of vertices in  $Y_2$  that fail this condition for  $Y_3$  is at most  $\epsilon|X_2|$ . Similarly for  $Y_4$ . Wait, we need the count relative to  $Y_2$ . The Generalized Degree Property says the number of bad vertices in the **whole** set  $X_2$  is  $< \epsilon|X_2|$ . So,

$$|X'_2| \geq |Y_2| - 2\epsilon|X_2|.$$

Since  $|Y_2| \approx d_{12}|X_2|$ , this is roughly  $|Y_2|(1 - \frac{2\epsilon}{d_{12}})$ . Using the crude bound from the image:

$$|X'_2| \geq (d_{12} - 3\epsilon)|X_2|.$$

(This accounts for the size of  $Y_2$  minus the bad vertices).

Fix  $x_2 \in X'_2$ . Let  $Z_i = N_{Y_i}(x_2)$  for  $i = 3, 4$ . We have  $|Z_i| \geq (d_{2i} - \epsilon)|Y_i|$ . Substituting the size of  $|Y_i|$ :

$$|Z_i| \geq (d_{2i} - \epsilon)(d_{1i} - \epsilon)|X_i| \geq \epsilon^2|X_i| \geq \epsilon|X_i|.$$

**Step 3: Count edges between  $Z_3$  and  $Z_4$ .** Finally, we have sets  $Z_3 \subseteq X_3$  and  $Z_4 \subseteq X_4$ . The pair  $(X_3, X_4)$  is  $\epsilon$ -regular. Since  $Z_3, Z_4$  are large enough, the density between them is roughly  $d_{34}$ .

$$e(Z_3, Z_4) \geq (d_{34} - \epsilon)|Z_3||Z_4|.$$

Any edge  $x_3x_4$  here completes a  $K_4$  with  $x_1, x_2$ .

**Total Count:** Combining everything:

- Choices for  $x_1$ :  $(1 - 3\epsilon)|X_1|$ .
- Choices for  $x_2$ :  $(d_{12} - 3\epsilon)|X_2|$ .
- Edges  $(x_3, x_4)$ :  $(d_{34} - \epsilon)(d_{23} - \epsilon)(d_{13} - \epsilon)|X_3| \cdot (d_{24} - \epsilon)(d_{14} - \epsilon)|X_4|$ .

Multiplying these terms gives the result. □

### 7.3.3 General Graph Counting Lemma

Luckily, using the same strategy works for counting any graphs. Compare the following statement of the general counting lemma to the above lemmas to understand what  $H$  characterizes.

**Theorem 0.21** (Graph Counting Lemma). Let  $H$  be a graph with vertices  $V(H) = \{1, \dots, h\}$  and let  $\delta > 0$ . There exists  $\epsilon > 0$  (dependent on  $H$  and  $\delta$ ) such that the following holds:

Let  $G$  be a graph and  $X_1, \dots, X_h$  be subsets of  $V(G)$ . Suppose that for every edge  $ij \in E(H)$ , the pair  $(X_i, X_j)$  is  $\epsilon$ -regular with edge density  $d_{ij} = d(X_i, X_j) \geq \delta$ . Then, the number of graph homomorphisms from  $H \rightarrow G$  where each vertex  $i \in V(H)$  is mapped to  $X_i$  is at least:

$$(1 - \delta) \left( \prod_{ij \in E(H)} (d_{ij} - \delta) \right) \prod_{i \in V(H)} |X_i|.$$

## 7.4 Removal Lemmas

This lemma says that a graph with few triangles can be made triangle-free by removing a few edges. We will also look at two proof methods, one via graphon, and one via regularity method.

### 7.4.1 Triangle Removal Lemma

**Theorem 0.22** (Triangle removal lemma). For every  $\epsilon > 0$ , there exists  $\delta > 0$  such that if a simple graph  $G$  with  $n$  nodes has at most  $\delta n^3$  triangles, then we can delete  $\epsilon n^2$  edges from  $G$  such that the remaining graph has no triangles.

*Proof 1: Via Graph Limits. Idea:* We proceed by contradiction.

1. **Construct a bad sequence.** Suppose the theorem is false. Then there exists an  $\epsilon > 0$  such that for any  $\delta > 0$ , there are graphs with triangle density  $t(K_3, G) < \delta$  but deleting any  $\epsilon n^2$  edges leaves at least one triangle. Let's construct a sequence of graphs  $(G_n)$  such that  $t(K_3, G_n) \rightarrow 0$  as  $n \rightarrow \infty$ , but for every  $n$ , one must remove at least  $\epsilon n^2$  edges to make  $G_n$  triangle-free.
2. **Compactness.** By the compactness of the graphon space, we can pick a subsequence such that  $G_n \rightarrow W$  in the cut distance. Since triangle density is continuous with respect to cut distance,
$$t(K_3, W) = \lim_{n \rightarrow \infty} t(K_3, G_n) = 0.$$
3. **Strategy.** We want to delete edges based on conditions derived from the limit graphon  $W$  and apply them back to  $G_n$ .
4. **Definitions.** Let  $S$  be the area where  $W$  is nonzero, i.e.,  $S = \{(x, y) : W(x, y) > 0\}$ . Let  $\mathbb{1}_S$  be the indicator function of  $S$ . This implies  $(1 - \mathbb{1}_S)W = 0$ . Let  $W_{G_n}$  be the step function of  $G_n$  that is 1 where there is an edge in  $G_n$ .
5. **Limit Analysis.** Lemma 8.22 gives: if  $W_{G_n} \rightarrow W$  in cut norm, then for any bounded measurable function  $f$ ,  $\int f W_{G_n} \rightarrow \int f W$ . Thus we have:

$$\int (1 - \mathbb{1}_S) W_{G_n} \rightarrow \int (1 - \mathbb{1}_S) W = 0.$$

The term  $(1 - \mathbb{1}_S)W_{G_n}$  is the indicator of edges of  $G_n$  that lie *outside* the region where  $W > 0$  (i.e., not in  $S$ ). Thus, the integral of it represents the total area of edges outside of  $S$ .

Now pick a large enough  $n$  (and let  $|V(G_n)| = N$ ) such that

$$\int (1 - \mathbb{1}_S)W_{G_n} < \frac{\epsilon}{4}.$$

6. **Edge Deletion Procedure.** Assume  $V(G_n) = [N]$ . The graphon  $W_{G_n}$  is exactly 1 on the rectangle  $R_{ij}$  corresponding to an edge  $ij$ . Each  $R_{ij}$  has area  $1/N^2$ . Let  $\lambda$  be the Lebesgue measure. We delete an edge  $ij$  if

$$\lambda(S \cap R_{ij}) < \frac{3}{4} \frac{1}{N^2},$$

meaning that less than 3/4 of the area of the rectangle  $R_{ij}$  lies in  $S$ . Let the resulting graph be  $G'_n$ .

7. **Verification (Triangle Free).** We try to find a triangle  $ijk$  inside  $G'_n$ . If such a triangle exists, the edges  $ij$ ,  $jk$ , and  $ki$  were *not* deleted. This means:

$$\lambda(S \cap R_{ij}) \geq \frac{3}{4} \frac{1}{N^2}, \quad \lambda(S \cap R_{jk}) \geq \frac{3}{4} \frac{1}{N^2}, \quad \lambda(S \cap R_{ki}) \geq \frac{3}{4} \frac{1}{N^2}.$$

Recall that  $t(K_3, W) = \int W(x, y)W(y, z)W(z, x) dx dy dz = 0$ . Since  $W$  is supported on  $S$ , this implies  $t(K_3, \mathbb{1}_S) = 0$ . But  $t(K_3, \mathbb{1}_S)$  is the measure of all triples lying inside  $S$ .

However, consider the contribution of the specific intervals  $J_i, J_j, J_k$  corresponding to the vertices  $i, j, k$ .

$$t(K_3, \mathbb{1}_S) \geq \int_{J_i \times J_j \times J_k} \mathbb{1}_S(x, y)\mathbb{1}_S(y, z)\mathbb{1}_S(z, x) dx dy dz \quad (*)$$

The total volume of the cube  $J_i \times J_j \times J_k$  is  $1/N^3$ . Thus  $(*) \geq \lambda(A)$ , where  $A$  is the set of triples inside the cube where the 3-tuples lie in  $S$ . The complement of  $A$  inside the cube corresponds to points where at least one pair lies outside  $S$ . The set of such triples lies inside:

$$(R_{ij} \setminus S) \times J_k \cup (R_{jk} \setminus S) \times J_i \cup (R_{ki} \setminus S) \times J_j.$$

Then, the measure of the area of edge  $ij$  except for  $S$  is  $\lambda(R_{ij} \setminus S)$ .

$$\lambda(R_{ij} \setminus S) \leq \frac{1}{N} \lambda(R_{ij} \setminus S) + \frac{1}{N} \lambda(R_{jk} \setminus S) + \frac{1}{N} \lambda(R_{ki} \setminus S).$$

For non-deleted edges, we know  $\lambda(R_{ij} \setminus S) \leq \frac{1}{4N^2}$ . Thus:

$$\lambda(A) \geq \frac{1}{N^3} - \left( \frac{1}{N} \frac{1}{4N^2} + \frac{1}{N} \frac{1}{4N^2} + \frac{1}{N} \frac{1}{4N^2} \right) = \frac{1}{N^3} - \frac{3}{4N^3} = \frac{1}{4N^3} > 0.$$

This implies  $t(K_3, \mathbb{1}_S) > 0$ , but we established earlier that  $t(K_3, \mathbb{1}_S) = 0$ , a contradiction. Thus, there are no triangles inside  $G'_n$ .

8. **Verification (Number of Deleted Edges).** It remains to show the number of edges deleted from  $G_n$  to obtain  $G'_n$  is less than  $\epsilon N^2$ . For deleted edges, they satisfy  $\lambda(S \cap R_{ij}) < \frac{3}{4N^2}$ , which implies  $\lambda(R_{ij} \setminus S) > \frac{1}{4N^2}$ .

We know that the areas of deleted edges must support area *outside* of  $S$ . Thus:

$$\int (1 - \mathbb{1}_S) W_{G_n} = \sum_{R_{ij}} \lambda(R_{ij} \setminus S).$$

Restricting the sum to only deleted edges  $E_{del}$ :

$$\sum_{ij \in E_{del}} \lambda(R_{ij} \setminus S) \geq \sum_{ij \in E_{del}} \frac{1}{4N^2} = |E_{del}| \frac{1}{4N^2}.$$

If we assumed we deleted at least  $\epsilon N^2$  edges, then:

$$\int (1 - \mathbb{1}_S) W_{G_n} \geq (\epsilon N^2) \frac{1}{4N^2} = \frac{\epsilon}{4}.$$

But this contradicts our choice of  $n$  in step 5, where we ensured  $\int (1 - \mathbb{1}_S) W_{G_n} < \frac{\epsilon}{4}$ .

Thus, we have found a set of edges of size  $< \epsilon N^2$ , the deletion of which results in a triangle-free graph. Thus the negation of the original lemma is wrong. □

Let's see another version of the proof using regularity lemma, which some readers may find more intuitive to read.

*Proof 2: Regularity Lemma Proof of Removal Lemma. Setup:* Let  $G$  be an  $n$ -vertex graph with fewer than  $\delta n^3$  triangles. Apply Szemerédi's Regularity Lemma with parameter  $\frac{\epsilon}{4}$  to obtain an  $\frac{\epsilon}{4}$ -regular partition of the vertex set  $V$  into parts  $V_1, \dots, V_m$ .

**Cleaning the Graph:** We remove edges from  $G$  that fall into "bad" categories to obtain a subgraph  $G'$ . Specifically, for each pair of parts  $(V_i, V_j)$ , we remove edges  $e \in E(G[V_i, V_j])$  if:

- (a) The pair  $(V_i, V_j)$  is **not**  $\frac{\epsilon}{4}$ -regular.
- (b) The density of the pair is low:  $d(V_i, V_j) < \frac{\epsilon}{2}$ .
- (c) The parts are small:  $\min\{|V_i|, |V_j|\} < \frac{\epsilon n}{4m}$  (This covers edges incident to the exceptional set  $V_0$  or inside parts).

**Bounding Removed Edges:** We calculate the number of edges removed in each step:

- **Bound (a):** By the definition of an  $\epsilon$ -regular partition, the number of edges in irregular pairs is at most the sum of sizes of such pairs:

$$\sum_{(V_i, V_j) \text{ not reg}} |V_i| |V_j| \leq \frac{\epsilon}{4} n^2.$$

- **Bound (b):** For low-density pairs, the number of edges is bounded by the density times the size:

$$\sum d(V_i, V_j) |V_i| |V_j| < \sum \frac{\epsilon}{2} |V_i| |V_j| \leq \frac{\epsilon}{2} n^2.$$

- **Bound (c):** Edges incident to small parts (or within parts/exceptional set) contribute:

$$< \frac{\epsilon n}{4m} \cdot m \cdot m = \frac{\epsilon}{4} n^2.$$

**Total edges removed:**

$$|E(G) \setminus E(G')| \leq \frac{\epsilon}{4}n^2 + \frac{\epsilon}{2}n^2 + \frac{\epsilon}{4}n^2 = \epsilon n^2.$$

**Contradiction (Finding a Triangle):** We choose  $\delta$  sufficiently small based on  $m$  and  $\epsilon$ . Suppose that after removing these edges, there still exists a triangle  $ijk$  in  $G'$  with vertices in  $V_i, V_j, V_k$ .

Since the edges  $ij, jk, ki$  survived the cleaning process:

1. The pairs  $(V_i, V_j), (V_j, V_k), (V_k, V_i)$  are all  $\frac{\epsilon}{4}$ -regular.
2. The densities satisfy  $d(V_i, V_j) \geq \frac{\epsilon}{2}, d(V_j, V_k) \geq \frac{\epsilon}{2},$  and  $d(V_k, V_i) \geq \frac{\epsilon}{2}.$

By the **Counting Lemma** (specifically the 3-graph version): If  $X, Y, Z$  are pairwise  $\epsilon$ -regular with densities at least  $2\epsilon$ , then the number of triangles is at least:

$$\#\Delta \geq (1 - \epsilon)d_{xy}d_{yz}d_{xz}|X||Y||Z|.$$

Applying this to our sets (using regularity parameter  $\epsilon/4$  and density lower bound  $\epsilon/2$ ):

$$\#\Delta(G) \geq \left(1 - \frac{\epsilon}{4}\right) \left(\frac{\epsilon}{2}\right)^3 |V_i||V_j||V_k|.$$

Since  $|V_i| \approx n/m$ , the number of triangles is roughly proportional to  $c(\epsilon)n^3$ .

By choosing  $\delta$  smaller than this constant factor (specifically  $\delta < (1 - \epsilon/4)(\epsilon/2)^3(1/m)^3$ ), the existence of even *one* triangle in  $G'$  implies that the original graph  $G$  must have contained at least  $\delta n^3$  triangles.

This contradicts the assumption that  $G$  had fewer than  $\delta n^3$  triangles. Therefore,  $G'$  must be triangle-free.  $\square$

#### 7.4.2 Extensions of the Removal Lemma

Before moving on, we should look at an equivalent statement of the triangle removal lemma, called the  $(6, 3)$ -theorem. This proof involves hypergraphs and is very interesting in how it connects the idea back to triangles.

But before it, we will need another lemma called the *diamond-free lemma*.

**Corollary 0.23** (Diamond-free lemma). If  $G$  is an  $n$ -vertex graph where every edge lies in a unique triangle, then  $G$  has  $o(n^2)$  edges.

*Proof.* The number of triangles is exactly  $|E(G)|/3$ . Let this count be  $N_\Delta$ . Since  $|E(G)| \leq \binom{n}{2}$ , we have  $N_\Delta = O(n^2) = o(n^3)$ .

By the Triangle Removal Lemma, we can remove a set of edges  $E_{rem}$  with size  $|E_{rem}| = o(n^2)$  to make the graph triangle-free.

Since every edge lies in a unique triangle, the triangles are edge-disjoint. To destroy a triangle, we must delete at least one of its edges. Because the triangles do not share edges, deleting an edge destroys exactly one triangle (and no others).

Therefore, to destroy all  $N_\Delta$  triangles, we must have deleted at least  $N_\Delta$  edges.

$$N_\Delta \leq |E_{rem}| = o(n^2).$$

Since  $|E(G)| = 3N_\Delta$ , it follows that  $|E(G)| = o(n^2)$ .  $\square$

**Theorem 0.24** ((6, 3)-theorem). Let  $H$  be an  $n$ -vertex 3-uniform hypergraph. If  $H$  contains no subgraph with 6 vertices and 3 edges (i.e., no induced matching of size 3, or any other configuration of 3 edges spanning  $\leq 6$  vertices), then  $|E(H)| = o(n^2)$ .

*Proof.* We partition the edges of  $H$  into "bad" and "good" edges based on their intersection properties.

**Step 1: Handle Bad Edges ( $E_{bad}$ )** Let  $E_{bad}$  be the set of edges that share 2 vertices with some other edge. Suppose  $e_1, e_2 \in E(H)$  share 2 vertices (say  $e_1 = \{a, b, c\}$  and  $e_2 = \{a, b, d\}$ ). The union of these two edges has 4 vertices  $\{a, b, c, d\}$ .

Consider any third edge  $e_3$  incident to any of these 4 vertices. If such an edge exists, then the union  $e_1 \cup e_2 \cup e_3$  would contain at most  $4 + 2 = 6$  vertices. This would form a set of 3 edges on  $\leq 6$  vertices, which is forbidden by the theorem statement.

Therefore, the vertices  $\{a, b, c, d\}$  must be isolated from the rest of the hypergraph; they cannot participate in any edges other than  $e_1$  and  $e_2$ .

This implies that "bad" pairs form disjoint blocks of 4 vertices, each contributing exactly 2 edges. The maximum number of such edges is:

$$|E_{bad}| \leq \frac{n}{4} \cdot 2 = O(n) = o(n^2).$$

We delete these edges. Let the remaining hypergraph be  $H'$ .

**Step 2: Construct Graph from Linear Hypergraph ( $H'$ )** The remaining hypergraph  $H'$  is linear, meaning no two edges share more than 1 vertex. We construct a simple graph  $G$  on the same vertex set  $V(H)$  as follows: For every hyperedge  $h = \{u, v, w\} \in E(H')$ , we add a triangle (edges  $uv, vw, wu$ ) to  $G$ .

**Step 3: Apply Diamond-free Lemma** We claim that every edge in  $G$  belongs to a unique triangle. Suppose an edge  $uv \in E(G)$  belongs to two triangles. This implies the pair  $\{u, v\}$  was contained in two distinct hyperedges in  $H'$ . However, this would mean those two hyperedges shared 2 vertices ( $u$  and  $v$ ), which contradicts the fact that we removed all such "bad" edges in Step 1.

Thus,  $G$  satisfies the condition of the Diamond-free Lemma.

$$|E(G)| = o(n^2).$$

Since every hyperedge in  $H'$  contributed exactly 3 edges to  $G$ :

$$|E(H')| = \frac{1}{3}|E(G)| = o(n^2).$$

Finally, the total edges in  $H$  are:

$$|E(H)| = |E_{bad}| + |E(H')| = O(n) + o(n^2) = o(n^2).$$

□

Finally we state general version of the Graph Removal lemma.

**Theorem 0.25** (Graph Removal Lemma). For every graph  $H$  and constant  $\epsilon > 0$ , there exists a constant  $\delta = \delta(H, \epsilon) > 0$  such that every  $n$ -vertex graph  $G$  with fewer than  $\delta n^{v(H)}$  copies of  $H$  can be made  $H$ -free by removing fewer than  $\epsilon n^2$  edges.

Below is a picture to visualize the proof of (6, 3)-theorem.

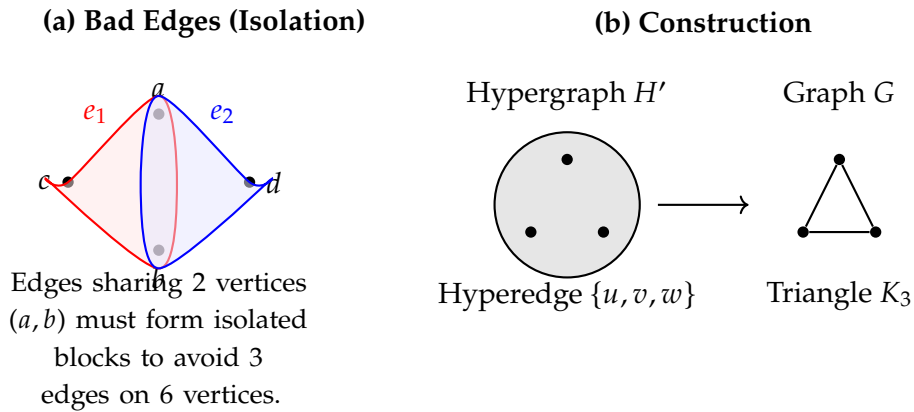


Figure 2: Visualization of the (6,3)-theorem proof. (a) Edges intersecting in 2 vertices must be isolated. (b) Transforming the linear hypergraph  $H'$  into a standard graph  $G$  by replacing hyperedges with triangles.

## 7.5 Applications

### 7.5.1 Proof of Roth's Theorem

With all the preparations we have made, this proof is rather subtle in my opinion, but it is a very important result in Additive Combinatorics, and the readers can see how graph theory and Additive Combinatorics are connected.

**Theorem 0.26 (Roth's theorem).** Let  $A \subseteq [N]$  be 3-AP-free (contains no arithmetic progression of length 3). Then  $|A| = o(N)$ .

**Remark 0.12 (Strategy).** Construct a tripartite graph where triangles correspond to solutions to  $x + z = 2y$  (arithmetic progressions). Use the Diamond-free Lemma.

*Proof.* Pick  $M = 2N + 1$ . Embed  $A$  into  $\mathbb{Z}/M\mathbb{Z}$  to avoid wraparound issues. Construct a tripartite graph with parts  $X, Y, Z$  each of size  $M$  (copies of  $\mathbb{Z}/M\mathbb{Z}$ ). Define edges:

- $(x, y) \in X \times Y$  if  $y - x \in A$ .
- $(y, z) \in Y \times Z$  if  $z - y \in A$ .
- $(x, z) \in X \times Z$  if  $(z - x)/2 \in A$ .

A triangle  $x, y, z$  exists iff  $y - x = a_1, z - y = a_2, (z - x)/2 = a_3$  for  $a_i \in A$ . This implies  $a_1 + a_2 = z - x = 2a_3$ , or  $a_1, a_3, a_2$  form an arithmetic progression. Since  $A$  is 3-AP-free, we must have  $a_1 = a_2 = a_3$  (trivial AP). This corresponds to a unique triangle for each edge. By the Diamond-free Lemma, the number of edges is  $o(M^2)$ . Total edges  $\approx 3M|A|$ . Thus  $3M|A| = o(M^2) \implies |A| = o(M) = o(N)$ .  $\square$

Higher dimension generalization like the Corner-free theorem exists.

## 7.5.2 Erdős-Stone-Simonovits Theorem

Another application is the ESS theorem. Notice how a general regularity method is used in the proof. This idea of isolating partitions and then process them respectively is very helpful.

**Theorem 0.27** (Erdős-Stone-Simonovits). Fix a graph  $H$  with at least one edge. Then:

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

*Proof.* Let  $r = \chi(H) - 1$ . The lower bound is given by the Turan graph  $T_n(r)$ , which is  $H$ -free because  $H$  cannot be embedded into an  $r$ -partite graph (since  $\chi(H) = r + 1$ ).

For the upper bound, fix  $\epsilon > 0$ . Let  $G$  be an  $n$ -vertex graph with at least

$$\left(1 - \frac{1}{r} + \epsilon\right) \frac{n^2}{2}$$

edges. We want to show that for sufficiently large  $n$ ,  $G$  contains  $H$  as a subgraph.

**Step 1: Regularity and Cleaning** Apply the Regularity Lemma to obtain an  $\eta$ -regular partition of  $V(G)$  into parts  $V_1, \dots, V_m$ , for a sufficiently small  $\eta > 0$  (depending on  $\epsilon, H$ ).

We construct a "cleaned" graph  $G'$  by removing edges from  $G$  that belong to "bad" pairs. Specifically, for any pair  $(V_i, V_j)$ , remove edges if:

- The pair  $(V_i, V_j)$  is not  $\eta$ -regular.
- The density is low:  $d(V_i, V_j) < \frac{\epsilon}{8}$ .
- The parts are small (or edges are internal):  $\min(|V_i|, |V_j|) < \frac{\epsilon n}{8m}$ .

We bound the number of edges removed:

- Condition (a): The number of edges in irregular pairs is  $\leq \eta n^2$ . We choose  $\eta \leq \frac{\epsilon}{8}$ , so this is  $\leq \frac{\epsilon}{8} n^2$ .
- Condition (b): The sum of densities is small. Total edges removed  $< \frac{\epsilon}{8} n^2$ .
- Condition (c): Total edges incident to small parts or inside parts is  $< m \frac{\epsilon n^2}{8m} = \frac{\epsilon}{8} n^2$ .

Total edges removed is  $\leq \frac{3}{8} \epsilon n^2$ . The remaining graph  $G'$  still has at least

$$\left(1 - \frac{1}{r} + \epsilon\right) \frac{n^2}{2} - \frac{3}{8} \epsilon n^2 > \left(1 - \frac{1}{r} + \frac{\epsilon}{4}\right) \frac{n^2}{2}$$

edges.

**Step 2: Finding a clique in the cluster graph** Let  $R$  be the *cluster graph* (or reduced graph) with vertices  $\{1, \dots, m\}$ . We place an edge  $ij$  in  $R$  if the pair  $(V_i, V_j)$  was *not* cleaned (i.e., it is  $\eta$ -regular and dense). The density of edges in  $G'$  is roughly the edge density of  $R$ . Since  $G'$  has density significantly higher than  $1 - \frac{1}{r}$ , the graph  $R$  must also have density  $> 1 - \frac{1}{r}$ .

By Turan's Theorem applied to  $R$ ,  $R$  contains a clique of size  $r + 1 = \chi(H)$ . Let the indices of this clique be  $i_1, \dots, i_{\chi(H)}$ . This means we have disjoint sets  $V_{i_1}, \dots, V_{i_{\chi(H)}}$  in  $G$  such that every pair is  $\eta$ -regular with density at least  $\frac{\epsilon}{8}$ .

**Step 3: Embedding  $H$**  Let the color classes of  $H$  be  $C_1, \dots, C_{\chi(H)}$ . We map the vertices of  $C_k$  to the cluster  $V_{i_k}$ .

Since each pair of clusters  $(V_{i_a}, V_{i_b})$  is regular and dense, we can apply the **Graph Counting Lemma**. The number of homomorphisms from  $H$  into  $G'$  respecting this partition is at least roughly  $c \cdot n^{v(H)}$ .

To count *injective* homomorphisms (embeddings), note that the number of non-injective homomorphisms (where two vertices map to the same node) involves choosing a pair  $u, v \in V(H)$  to collide. This effectively forces an embedding of a graph with  $v(H) - 1$  vertices. The number of such maps is  $O(n^{v(H)-1})$ . Since  $O(n^{v(H)-1})$  is an order of magnitude smaller than  $cn^{v(H)}$ , for large  $n$  there exists at least one injective embedding of  $H$  into  $G$ .  $\square$

## 7.6 Statements of Other Removal Lemmas

Beyond the standard Graph Removal Lemma, there are several powerful generalizations that extend the result to induced subgraphs, infinite families of forbidden graphs, and hypergraphs.

**Theorem 0.28** (Induced Graph Removal Lemma). For any graph  $H$  and  $\epsilon > 0$ , there exists  $\delta > 0$  such that if an  $n$ -vertex graph  $G$  has fewer than  $\delta n^{v(H)}$  induced copies of  $H$ , then it can be made induced  $H$ -free by adding and/or deleting fewer than  $\epsilon n^2$  edges.

**Remark.** Unlike the standard removal lemma, the induced version requires us to allow both **adding and deleting** edges. In terms of distance, this theorem states that every graph with few induced copies of  $H$  is close in *edit distance* to an induced  $H$ -free graph. If we were restricted to only deleting edges, the statement would be false (e.g., a clique with a few missing edges contains few induced empty graphs, but cannot be made empty by deleting edges).

**Theorem 0.29** (Infinite Graph Removal Lemma). For each (possibly infinite) set of graphs  $\mathcal{H}$  and  $\epsilon > 0$ , there exist integers  $h_0$  and a constant  $\delta > 0$  such that: If  $G$  is an  $n$ -vertex graph containing fewer than  $\delta n^{v(H)}$  induced copies of  $H$  for every  $H \in \mathcal{H}$  with at most  $h_0$  vertices, then  $G$  can be made induced  $\mathcal{H}$ -free by adding/removing fewer than  $\epsilon n^2$  edges.

**Remark.** This result is due to Alon and Shapira (2008). The parameter  $h_0$  essentially implies that "testing" for the property of being  $\mathcal{H}$ -free only requires checking forbidden subgraphs up to a certain size, which is a fundamental result in the field of Property Testing.

**Theorem 0.30** (Hypergraph Removal Lemma). For every  $r$ -uniform hypergraph  $H$  and  $\epsilon > 0$ , there exists  $\delta > 0$  such that every  $n$ -vertex  $r$ -graph with fewer than  $\delta n^{v(H)}$  copies of  $H$  can be made  $H$ -free by removing fewer than  $\epsilon n^r$  edges.

**Remark.** This theorem is substantially more difficult to prove than the graph case and was established by Gowers (2007) and Rödl et al. (2005). A major application of this lemma is providing a combinatorial proof of **Szemerédi's Theorem** on arithmetic progressions, where the existence of a  $k$ -term arithmetic progression corresponds to a specific configuration in a hypergraph.