BACHELOR'S DEGREE THESIS

Degree in Mathematics

Title: Removal lemmas in sparse graphs

Author: Ander Lamaison Vidarte

Advisors: Oriol Serra Albo, Lluís Vena Cros

Department: Applied Mathematics IV

Academic year: 2014/2015

UNIVERSITAT POLITÈCNICA DE CATALUNYA
BARCELONATECH
Facultat de Matemàtiques i Estadística

Universitat Politècnica de Catalunya
Facultat de Matemàtiques i Estadística

Bachelor's Degree Thesis

Removal lemmas in sparse graphs

Ander Lamaison Vidarte

Advisors: Oriol Serra Albo
Lluís Vena Cros

Departament de Matemàtica Aplicada IV

This work is dedicated to my family, for providing me with unconditional support all these years, and to everyone who has helped me in the path through university.

Abstract

Key words: Pseudorandom, regularity lemma, removal lemma, sparse graphs

MSC2010: 05C35, 05C80

The aim of this work is to explain and prove the graph removal lemma, in both the dense and the sparse cases, and show how these can be applied to finite groups to obtain arithmetic removal lemmas. The graph removal lemma, in its most basic form, states that for any fixed graph H, if a graph G on n vertices contains $o(n^{v(H)})$ copies of H, then all copies can be deleted from G by deleting $o(n^2)$ edges. We will show how the concept of regularity, and the regularity and counting lemmas, are crucial in the proof of the removal lemmas. We will explain the motivation behind the development of the sparse case, and the role of pseudorandom graphs in sparse versions of the removal lemma. Finally, we will see how the removal lemma, both in its graph and its arithmetic versions, can be used to prove Roth's theorem, that is, the existence of non-trivial 3-term arithmetic progressions in any subset of the natural numbers with positive density.

Notation

$[n]$	$\{1, 2, ..., n\}$
$E(G)$	Set of edges of graph G
$V(G)$	Set of vertices of graph G
$e(G)$	Number of edges of graph G
$v(G)$	Number of vertices of graph G

Contents

Chapter 1
Introduction

The origins of the removal lemma can be traced back to a conjecture proposed by Erdős and Turán in 1936 [**ErdTur**]. This conjecture asked whether any subset of $\mathbb{N}$ in which the sum of the reciprocals of the elements is divergent necessarily contains a non-trivial k-term arithmetic progression for all positive integers k. An interesting particular case of this conjecture is whether this holds for subsets of $\mathbb{N}$ of positive density. This is a result known today as Szemerédi's Theorem on arithmetic progressions: for any density $\epsilon > 0$, subsets of $[n]$ with density at least ϵ, for n large enough, always contain k-term arithmetic progressions.

The first answer for the dense case came in 1953, when Roth [**Rot**] proved the case $k = 3$ using Fourier analysis. In the seventies, Szemerédi proved the result for general k using combinatorial methods [**Sze2**]. This proof introduced a tool that would be of great relevance in extremal combinatorics: regularity in graphs, and in particular the regularity lemma.

The concept of regularity is one of equidistribution of edges. We say that a graph is regular when the density of edges between any two large enough sets of vertices is approximately the same as the density of the entire graph. A partition of the vertex set of the graph is said to be regular if almost all of the pairs of parts are regular, and the parts are of the same size.

From the many results involving regularity that have been proven since it was introduced, the two that we will use are the regularity lemma and the counting lemma. The regularity lemma states that any graph admits a regular partition, and there is an upper bound on the number of parts required [**Sze3**]. Meanwhile, the counting lemma gives a lower bound on the number of embeddings of a graph H into a regular partition, under certain conditions [**KomSim**].

The combination of both lemmas produces the central result of this thesis: the graph removal lemma [**RuzSze**, **Fur**]:

Theorem 2.1 (Removal lemma). Let $\epsilon > 0$ be a constant and H be a graph on h vertices. Then there exists $\delta > 0$ for which the following property holds: any graph G on n vertices, which contains at most δn^h copies of H, can be made H-free (not containing any copies of H) by removing at most ϵn^2 edges.

This lemma has many applications, one of which is that it allows for a straightforward proof of Roth's theorem (the case $k = 3$ of Szemerédi's theorem) [**Rot**, **RuzSze**]. In fact, a generalization of the removal lemma which applies to hypergraphs instead of graphs allows to prove Szemerédi's theorem [**Gow**, **FraRod**]. The extension, however, is not simple because it requires a much more complicated definition for regular hypergraphs. Another important application is the arithmetic removal lemma [**Gre**], which deals with solutions to the equation $x_1 x_2 \cdots x_k = 1$ in a finite abelian group G, where $x_i \in S_i \subseteq G$. The lemma says that if the number of solutions to this equation is small, then a small number of elements can be removed from each set S_i in a way that removes all solutions of the equation.

When using the removal lemma, we run into problems when we try to apply it to sparse graphs, that is, graphs on n vertices in which the number of edges is $o(n^2)$ as n goes to infinity, because the statement of the lemma becomes trivial. Indeed, the statement of Theorem 2.1 allows us to remove up to ϵn^2 edges from G, but this means that we can remove all edges from G (and then trivially G becomes H-free). For this reason, a different version of this lemma is needed. However, this result can only be applied to a certain family of graphs, which is subgraphs of pseudorandom graphs. There are counterexamples to many statements that could reasonably be generalizations of the removal lemma if we do not impose conditions on the graph G.

Pseudorandom graphs are classes of graphs that satisfy certain properties that random graphs (especially Erdős-Rényi random graphs[1]) satisfy asymptotically almost surely, but that fail for most usual families of graphs that follow a certain structure. They are deterministic conditions that attempt to replicate certain behaviours of random graphs. Regular graphs, mentioned above, are one example of a pseudorandom class of graphs. Other examples include jumbled graphs and graphs satisfying discrepancy. For the sparse removal lemma, the pseudorandomness condition that we will impose is jumbledness.

The proof of the sparse version of the removal lemma [**ConFoxZha**] follows the same steps as the proof for the dense case, but the details get more complicated as additional calculations and considerations need to be made using the jumbledness condition. Here we will only give the proof for the case when H is a cycle of length at least 5, and we will mention the techniques used in other cases.

The lemma also leads to a sparse arithmetic removal lemma, but the proof now is analogous as that of the dense case. As in the graph removal lemma, pseudorandomness is required in the proof, so this result can only be applied to subsets of jumbled sets.

The aim of this work is to prove the different removal lemmas and some of the most important applications. We also want to understand the concept of pseudorandomness and why pseudorandom graphs are useful in adaptations of dense theorems to sparse graphs. In Chapter 2 we prove the dense case of the graph and arithmetic removal lemmas, and the applications resulting from them. In Chapter 3 we introduce the concept of pseudorandomness, and prove the sparse versions of the graph and arithmetic removal lemmas.

[1] We will denote by $G_{n,p}$ the graph on n vertices where the probability of each edge appearing is p and the edges are taken independently. These are called Erdős-Rényi graphs.

Chapter 2
Removal lemma in dense graphs

In this chapter we will introduce and prove the graph removal lemma in the dense case. First, we will introduce the concept of regularity and regular partitions, which is the main tool used in this proof. This will be followed by the proof of the regularity lemma. Then, we will prove the counting lemma. Later, we will use these two results to prove the removal lemma and a variant of it. Finally, we will give several applications of the graph removal lemma, including Roth's theorem and the arithmetic removal lemma.

The statement of the removal lemma is the following [**Fur**]:

Theorem 2.1 (Graph removal lemma). *Let $\epsilon > 0$ be a constant and H be a graph on h vertices. Then there exists $\delta > 0$ for which the following property holds: any graph G on n vertices, which contains at most δn^h copies of H, can be made H-free by removing at most ϵn^2 edges.*

As an observation, notice that the number of embedded copies of H in the complete graph on n vertices is $v(H)!\binom{n}{v(H)} = \Theta(n^{v(H)})$, while the number of edges grows with $\binom{n}{2} = \Theta(n^2)$. Thus the theorem says that if a graph G contains only a small proportion of all the possible copies of H, then a small proportion of the edges of G are responsible for the formation of all those copies.

The theorem, however, is not as trivial as this reformulation might make it seem, as for graphs H with at least three vertices the maximum number of copies allowed ($\delta n^{v(H)}$) grows much faster than the number of edges that we are allowed to remove (ϵn^2). If the latter number was bigger, then removing an edge from each copy of H would solve the problem.

Here is an example of how this theorem can be used: Consider the graph G from Figure 1, where each of the sets X_i contains t vertices, so $n = 5t$, and consider $H = K_3$. Every triangle in G must contain its three vertices in X_1, X_2 and X_5, and the vertices from X_2 and X_5 must be joined by an edge. The number of copies of H in G is $6te(X_2, X_5) = o(n^3)$, so the theorem tells us that we can take away $o(n^2)$ edges from G to make it triangle free. Sure enough, there are $o(n^2)$ red edges, and removing them makes G triangle free.

In this example, it is clear that there is a partition of the vertices of G, such that every copy of H in G has an edge between two parts that form a bipartite graph with very low density. In the

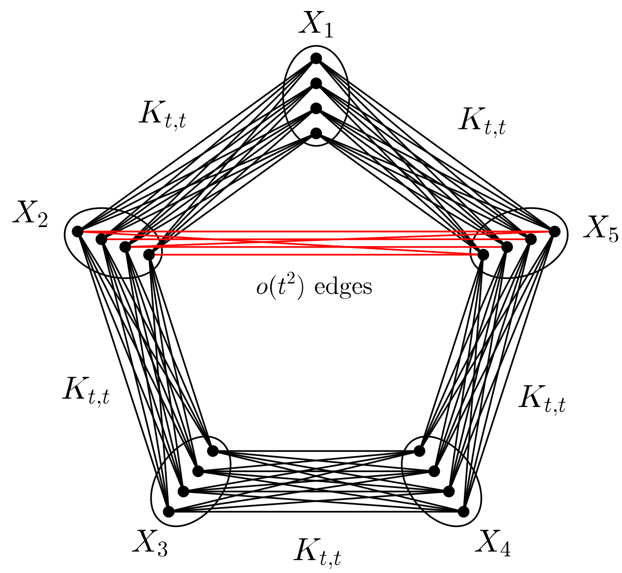

FIG. 1. Example of application of the removal lemma

example that we just picked the number of parts is very low and the partition is very clear, but this turns out to be the key step in the proof: if the graph G has $o(n^{v(H)})$ copies of H, then the set of vertices of G admits a partition in which, by removing all pairs with very small density and very few ones with big density, all copies of H can be removed.

To prove Theorem 2.1, we will need to introduce the concept of regularity, and prove two key results related to it: the regularity lemma and the counting lemma.

2.1. The regularity lemma

Before stating the regularity lemma we first need to understand the concept of regularity in a graph. For this purpose we need some definitions. For any two subsets $X, Y \subseteq V(G)$, we denote by $e_G(X, Y)$, or simply $e(X, Y)$, the number of ordered pairs (x, y) such that $x \in X$, $y \in Y$ and xy is an edge of G. For this definition, the vertex sets X and Y are not necessarily disjoint. We also define the density of the pair (X, Y) as the proportion of pairs $(x, y) \in X \times Y$ that are edges of G, that is, $d_G(X, Y) := \frac{e_G(X,Y)}{|X||Y|}$. This value is always in the interval $[0, 1]$. When the graph it refers to is clear, we will simply denote the density as $d(X, Y)$.

Definition 2.2 (ϵ-regular pair). Let $\epsilon > 0$. We say that a pair of vertex sets (X, Y) is ϵ-regular if, for every two subsets $X' \subseteq X$ and $Y' \subseteq Y'$ with $|X'| \geq \epsilon|X|$ and $|Y'| \geq \epsilon|Y|$ we have

$$|d(X', Y') - d(X, Y)| \leq \epsilon$$

This definition tells us that, in regular pairs, the edges do not form large clusters, nor are there sparse subsets of vertices: the density of the graph is uniform over all large enough subsets. The idea is, therefore, that the edges are equidistributed on large subsets. This is what one would expect, for example, from Erdős-Rényi graphs.

Definition 2.3 (ϵ-regular partition). Let $V = V_0 \cup V_1 \cup ... \cup V_k$ be a partition of the vertices of G. We say that this partition is ϵ-regular if the following three conditions are satisfied:

- $|V_1| = |V_2| = ... = |V_k|$ (we call partitions satisfying this propety equitable partitions)
- $|V_0| \leq \epsilon |V|$
- All except at most ϵk^2 of the pairs (V_i, V_j) with $1 \leq i, j \leq k$ are ϵ-regular.

In this case, the set V_0 is called the exceptional set.

In this definition, the exceptional set V_0 satisfies two roles. On the one hand, it makes the condition $|V_1| = |V_2| = ... = |V_k|$ satisfiable in cases where k does not divide $|V|$, as otherwise the number of vertices in each set would not be an integer. The other role is a matter of convenience, as when constructing an ϵ-regular partition we can place in V_0 all vertices that are not suitable to belong in any other set, provided that there are not too many of these vertices. However the inclusion of this set is not necessary, as there are alternative definitons of ϵ-regular partition which do not use this set[1], and also allow to prove a regularity lemma.

With this definition it is possible now to state the regularity lemma:

Lemma 2.4 (Regularity lemma). *For every $\epsilon > 0$ and every positive integer m there exists an integer M such that, for any graph G with at least m vertices, there exists an ϵ-regular partition $\{V_0, V_1, ..., V_k\}$ of the vertex set of G with $m \leq k \leq M$. Moreover, given any partition $\mathcal{P}$ of the vertex set of G into at most m parts, such an ϵ-regular partition can be found in a way that each set V_i with $1 \leq i \leq k$ is contained in one of the sets of $\mathcal{P}$.*

The proof that we will see follows the one from Diestel [**Die**].

In an ϵ-regular partition we will treat V_0 as a group of singletons, so the condition that V_i is contained in a set of $\mathcal{P}$ is similar to saying that the ϵ-regular partition refines $\mathcal{P}$. For a partition $\mathcal{P} = \{V_0, V_1, ..., V_k\}$ with exceptional set V_0, we define $\tilde{\mathcal{P}} = \{V_1, V_2, ..., V_k\} \cup (\bigcup_{v_0 \in V_0} \{\{v_0\}\})$ (where elements of V_0 belong to singletons) and, if we have two partitions with exceptional set $\mathcal{P}_1$ and $\mathcal{P}_2$, we say that the former refines the latter if $\tilde{\mathcal{P}}_1$ refines $\tilde{\mathcal{P}}_2$.

The following identity will be useful in the proof of the regularity lemma:

Lemma 2.5. *Let G be a bipartite graph on vertex sets X and Y, and let $\mathcal{X} = X_1 \cup X_2 \cup ... \cup X_a$ and $\mathcal{Y} = Y_1 \cup Y_2 \cup ... \cup Y_b$ be partitions of X and Y. Then*

$$\sum_{i=1}^{a} \sum_{j=1}^{b} |X_i||Y_j| d\left(X_i, Y_j\right) = \sum_{i=1}^{a} \sum_{j=1}^{b} |X_i||Y_j| d\left(X, Y\right) \tag{1}$$

[1] One option is to redefine the concept of equitable partition, so that $||V_i| - |V_j|| \leq 1$ for any $1 \leq i, j \leq k$.

Proof. The equality comes from the fact that each edge of G is contained in exactly one graph $G|_{X_iY_j}$, so by double counting,

$$\begin{aligned}\sum_{i=1}^{a}\sum_{j=1}^{b}|X_i||Y_j|d\left(X_i,Y_j\right) &= \sum_{i=1}^{a}\sum_{j=1}^{b}e(X_i,Y_j)\\ &= e(X,Y)\\ &= |X||Y|d(X,Y)\\ &= \left(\sum_{i=1}^{a}|X_i|\right)\left(\sum_{j=1}^{b}|Y_j|\right)d(X,Y)\\ &= \sum_{i=1}^{a}\sum_{j=1}^{b}|X_i||Y_j|d\left(X,Y\right)\end{aligned}$$

□

Let $\mathcal{X}$ and $\mathcal{Y}$ be partitions of X and Y. We consider the identity (1). The terms $|X_i||Y_j|$ are non-negative. If we treat them as weights, then $d(X_i,Y_j)$ and $d(X,Y)$ satisfy the conditions necessary to apply Jensen's inequality[2]. If $f:[0,1]\longrightarrow\mathbb{R}$ is a convex function, then

$$\sum_{i=1}^{a}\sum_{j=1}^{b}|X_i||Y_j|f(d\left(X_i,Y_j\right)) \geq \sum_{i=1}^{a}\sum_{j=1}^{b}|X_i||Y_j|f\left(d\left(X,Y\right)\right) \tag{2}$$

Definition 2.6 (Quadratic mean density). Let G be a graph on vertex set V, with $|V|=n$. Let $X,Y\subset V$. We define

$$q(X,Y) := \frac{|X||Y|}{n^2}d^2(X,Y)$$

Let $\mathcal{X}$ and $\mathcal{Y}$ be partitions of sets X,Y. Then

$$q(\mathcal{X},\mathcal{Y}) := \sum_{\substack{X_i\in\mathcal{X}\\ Y_j\in\mathcal{Y}}} q(X_i,Y_j)$$

If $\mathcal{P}$ is a partition of V without exceptional set, then

$$q(\mathcal{P}) := q(\mathcal{P},\mathcal{P})$$

If $\mathcal{P}=X_0\cup X_1\cup\ldots\cup X_k$ is a partition of V with exceptional set X_0, then define $\tilde{\mathcal{P}} := \{X_1,X_2,\ldots,X_k\}\cup\{\{v\}:v\in X_0\}$ (treating elements of the exceptional sets as singletons) and

$$q(\mathcal{P}) := q(\tilde{\mathcal{P}})$$

[2]Remember Jensen's inequality: if $c_1,c_2,\ldots,c_k$ are non-negative reals, $x_1,x_2,\ldots,x_k$ are reals, f is a convex function and x is a value such that $\sum_{i=1}^{k}c_ix_i=\sum_{i=1}^{k}c_ix$, then

$$\sum_{i=1}^{k}c_if(x_i)\geq\sum_{i=1}^{k}c_if(x)$$

Observation: If $\mathcal{P}$ is a partition of V then $0 \leq q(\mathcal{P}) \leq 1$, because

$$q(\mathcal{P}) = \sum_{X_i, X_j \in \mathcal{P}} \frac{|X_i||X_j|}{n^2} d^2(X_i, X_j)$$

(which is a sum of non-negative terms) and

$$\sum_{X_i, X_j \in \mathcal{P}} \frac{|X_i||X_j|}{n^2} d^2(X_i, X_j) \leq \sum_{X_i, X_j \in \mathcal{P}} \frac{|X_i||X_j|}{n^2} = \frac{|V||V|}{n^2} = 1$$

The same holds for partitions with exceptional set.

The quadratic mean density is the key tool in the proof of the regularity lemma. The proof consists of three steps:

- If $\mathcal{P}'$ is a refinement of $\mathcal{P}$, then $q(\mathcal{P}') \geq q(\mathcal{P})$
- If $\mathcal{P}$ is an equitable partition in k parts with small exceptional set and it is not ϵ-regular, then there is a refinement (not necessarily equitable) in at most $k4^k$ parts which does not increase the size of the exceptional set and with $q(\mathcal{P}') \geq q(\mathcal{P}) + \frac{\epsilon^5}{4}$.
- If $\mathcal{P}$ is a partition in k parts and $\delta > 0$, then there is a refinement of $\mathcal{P}$ in at most $\delta^{-1}k$ parts which is equitable and which increases the size of the exceptional set by at most δn.

Once those steps have been proven, a constructive proof of the regularity lemma goes as follows: start with any equitable partition into m parts. As long as the partition is equitable but not regular, use step 2 to find a partition that increases $q(\mathcal{P})$ by at least $\frac{\epsilon^5}{4}$, and then use steps 1 and 3 to find an equitable refinement that increases $|X_0|$ by little. Since applying these two steps increments $q(\mathcal{P})$ by at least $\frac{\epsilon^5}{4}$, and $q(\mathcal{P})$ is bounded by 1, the procedure must stop after applying step 2 at most $\lfloor 4\epsilon^{-5} \rfloor$ times, producing an ϵ-regular partition.

The first step is a direct application of inequality (2) for the function $f(x) = x^2$:

Lemma 2.7. *Let X and Y be sets of vertices of G, and let $\mathcal{A}$ and $\mathcal{A}'$ be two partitions of X and $\mathcal{B}$ and $\mathcal{B}'$ be two partitions of Y such that $\mathcal{A}'$ refines $\mathcal{A}$ and $\mathcal{B}'$ refines $\mathcal{B}$. Then $q(\mathcal{A}', \mathcal{B}') \geq q(\mathcal{A}, \mathcal{B})$. (Partitions may have exceptional sets)*

Proof. By expanding the formulas for $q(\mathcal{A},\mathcal{B})$ and $q(\mathcal{A}',\mathcal{B}')$, we obtain

$$\begin{aligned}
q(\mathcal{A}',\mathcal{B}') &= \sum_{A'\in\tilde{\mathcal{A}}'}\sum_{B'\in\tilde{\mathcal{B}}'} q(A',B') \\
&= \sum_{A\in\tilde{\mathcal{A}}}\sum_{B\in\tilde{\mathcal{B}}}\sum_{\substack{A'\in\tilde{\mathcal{A}}'\\A'\subset A}}\sum_{\substack{B'\in\tilde{\mathcal{B}}'\\B'\subset B}} q(A',B') \\
&= \sum_{A\in\tilde{\mathcal{A}}}\sum_{B\in\tilde{\mathcal{B}}}\sum_{\substack{A'\in\tilde{\mathcal{A}}'\\A'\subset A}}\sum_{\substack{B'\in\tilde{\mathcal{B}}'\\B'\subset B}} \frac{|A'|\,|B'|}{n^2}d^2(A',B') \\
&\overset{(2)}{\geq} \sum_{A\in\tilde{\mathcal{A}}}\sum_{B\in\tilde{\mathcal{B}}}\sum_{\substack{A'\in\tilde{\mathcal{A}}'\\A'\subset A}}\sum_{\substack{B'\in\tilde{\mathcal{B}}'\\B'\subset B}} \frac{|A'|\,|B'|}{n^2}d^2(A,B) \\
&= \sum_{A\in\tilde{\mathcal{A}}}\sum_{B\in\tilde{\mathcal{B}}} \frac{|A|\,|B|}{n^2}d^2(A,B) \\
&= \sum_{A\in\tilde{\mathcal{A}}}\sum_{B\in\tilde{\mathcal{B}}} q(A,B) \\
&= q(\mathcal{A},\mathcal{B})
\end{aligned}$$

□

Corollary 2.8. *If $\mathcal{P}$ and $\mathcal{P}'$ are two partitions of V such that $\mathcal{P}'$ refines $\mathcal{P}$, then $q(\mathcal{P}') \geq q(\mathcal{P})$*

Proof. $q(\mathcal{P}') = q(\mathcal{P}',\mathcal{P}') \geq q(\mathcal{P},\mathcal{P}) = q(\mathcal{P})$ □

This shows that, if we take a sequence of partitions, each of which refines the previous ones, then $q(\mathcal{P})$ is non-decreasing. The second step, which is the key step, consists of showing that, if a partition is equitable but not ϵ-regular, then we can increase $q(\mathcal{P})$ by a constant depending only on ϵ.

Lemma 2.9. *Let $0 < \epsilon < \frac{1}{2}$ and let $\mathcal{P} = \{X_i\}_{i=0}^k$ be an equitable partition of V with exceptional set V_0 and k non-exceptional sets. If $|X_0| < \epsilon|V|$ and the partition is not ϵ-regular, then there is another partition $\mathcal{P}'$, not necessarily equitable, with at most $k4^k$ non-exceptional parts, the same exceptional set X_0 and $q(\mathcal{P}') \geq q(\mathcal{P}) + \frac{\epsilon^5}{4}$.*

Proof. Let $S = \{(i,j) \in [k]^2 : (X_i,X_j) \text{ is not } \epsilon\text{-regular}\}$. If $\mathcal{P}$ is not ϵ-regular, then $\epsilon k^2 \leq |S| \leq k^2$. For every pair (i,j) that is not ϵ-regular, by definition of regularity, there are sets $X_i^j \subset X_i$ and $X_j^{[i]} \subset X_j$ such that $|X_i^j| \geq \epsilon|X_i|$, $|X_j^{[i]}| \geq \epsilon|X_j|$ and $|d(X_i^j, X_j^{[i]}) - d(X_i,X_j)| \geq \epsilon$.

Now take $\mathcal{P}'$ to be the coarsest partition that refines all the sets X_i^j and $X_j^{[i]}$. Within each set X_i there are at most k sets X_i^j and k sets $X_i^{[j]}$, which means that the coarsest partition of X_i that refines all those sets has at most $2^{2k} = 4^k$ sets, so the partition $\mathcal{P}'$ requires no more than $k4^k$ non-exceptional sets. Denote by $\tilde{\mathcal{P}}'(X)$ the partition of $X \in \tilde{\mathcal{P}}$ in $\tilde{\mathcal{P}}'$, and by $\mathcal{P}_i(X)$ the partition of X_i into two sets induced by $X \subset X_i$.

$$
\begin{aligned}
q(\mathcal{P}') - q(\mathcal{P}) &= \sum_{A' \in \tilde{\mathcal{P}}'} \sum_{B' \in \tilde{\mathcal{P}}'} q(A', B') - \sum_{A \in \tilde{\mathcal{P}}} \sum_{B \in \tilde{\mathcal{P}}} q(A, B) \\
&= \sum_{A \in \tilde{\mathcal{P}}} \sum_{B \in \tilde{\mathcal{P}}} q(\tilde{\mathcal{P}}'(A), \tilde{\mathcal{P}}'(B)) - \sum_{A \in \tilde{\mathcal{P}}} \sum_{B \in \tilde{\mathcal{P}}} q(A, B) \\
&\quad \text{Only taking the irregular pairs:} \\
&\overset{2.7}{\geq} \sum_{(i,j) \in S} \left(q(\tilde{\mathcal{P}}'(X_i), \tilde{\mathcal{P}}'(X_j)) - q(X_i, X_j) \right) \\
&\overset{2.7}{\geq} \sum_{(i,j) \in S} \left(q\left(\mathcal{P}_i(X_i^j), \mathcal{P}_j\left(X_j^{[i]}\right)\right) - q(X_i, X_j) \right) \\
&\overset{(*)}{\geq} \sum_{(i,j) \in S} \left(\frac{\epsilon}{2k}\right)^2 \epsilon^2 \\
&\geq (\epsilon k^2) \left(\frac{\epsilon}{2k}\right)^2 \epsilon^2 \\
&= \frac{\epsilon^5}{4}
\end{aligned}
$$

where inequality (*) is detailed here:

$$
\begin{aligned}
& q\left(\mathcal{P}_i(X_i^j), \mathcal{P}_j\left(X_j^{[i]}\right)\right) - q(X_i, X_j) \\
&= \sum_{A \in \mathcal{P}_i(X_i^j)} \sum_{B \in \mathcal{P}_j\left(X_j^{[i]}\right)} q(A, B) - q(X_i, X_j) \\
&= \sum_{A \in \mathcal{P}_i(X_i^j)} \sum_{B \in \mathcal{P}_j\left(X_j^{[i]}\right)} \frac{|A||B|}{n^2} d^2(A, B) - \frac{|X_i||X_j|}{n^2} d^2(X_i, X_j) \\
&= \sum_{A \in \mathcal{P}_i(X_i^j)} \sum_{B \in \mathcal{P}_j\left(X_j^{[i]}\right)} \frac{|A||B|}{n^2} \left(d^2(A, B) - d^2(X_i, X_j) \right) \\
&\overset{(1)}{=} \sum_{A \in \mathcal{P}_i(X_i^j)} \sum_{B \in \mathcal{P}_j\left(X_j^{[i]}\right)} \frac{|A||B|}{n^2} \left(d^2(A, B) - d^2(X_i, X_j) - 2d(A, B)d(X_i, X_j) + 2d^2(X_i, X_j) \right) \\
&= \sum_{A \in \mathcal{P}_i(X_i^j)} \sum_{B \in \mathcal{P}_j\left(X_j^{[i]}\right)} \frac{|A||B|}{n^2} \left(d(A, B) - d(X_i, X_j) \right)^2 \\
&\geq \frac{|X_i^j||X_j^{[i]}|}{n^2} \left(d(X_i^j, X_j^{[i]}) - d(X_i, X_j) \right)^2 \\
&\geq \left(\frac{\epsilon}{2k}\right)^2 \epsilon^2
\end{aligned}
$$

□

This completes the second step. The number of parts could be reduced to $k2^{k+1}$ because we can make $X_i^j = X_i^{[j]}$ whenever $i \neq j$, but here we are not trying to optimize our bounds, we just want to show that they exist. The third step concerns equitable refinements of partitions.

Lemma 2.10. *Let $\mathcal{P} = \{X_i\}_{i=0}^k$ be a (not necessarily equitable) partition of V with exceptional set X_0, and let $\delta > 0$. Then there exists an equitable partition $\mathcal{P}' = \{X_i'\}_{i=0}^{k'}$ with exceptional set X_0' which refines $\mathcal{P}$, with $k' \leq \delta^{-1}k$ and $|X_0'| \leq |X_0| + \delta|V|$.*

Proof. Let $m = \delta k^{-1}|V|$. To construct $\mathcal{P}'$, partition each set X_i with $1 \leq i \leq k$ into sets of size m, and if $|X_i|$ is not divisible by m, add the remaining vertices to the exceptional set X_0'. Once we have done this, every non-exceptional set has size m, so the partition is equitable. If $k' > \delta^{-1}k$, then $\left|\bigcup_{i=1}^{k'} X_i'\right| = mk' > m\delta^{-1}k = |V|$, which is impossible, hence $k' \leq \delta^{-1}k$. Finally, at most m elements from each X_i with $1 \leq i \leq k$ go to X_0', so $|X_0'| \leq |X_0| + km = |X_0| + \delta|V|$. □

We are now ready to prove the regularity lemma, using the previous three lemmas:

Proof of Lemma 2.4. Without loss of generality, assume that $\epsilon \leq \frac{1}{2}$ (indeed, if $\epsilon > \frac{1}{2}$, then any $\frac{1}{2}$-regular partition is also ϵ-regular, so finding a $\frac{1}{2}$-regular partition is enough). Take a partition $\mathcal{P}_0$ into exactly m parts, with empty exceptional set, that refines $\mathcal{P}$. Let $\delta = 4\epsilon^{-5}$. Now do the following:

- If $\mathcal{P}_i$ has k_i non-exceptional sets, then construct an equitable partition $\mathcal{Q}_i$ with at most $(\delta + 1)\epsilon^{-1}k_i$ non-exceptional parts in which the exceptional set increases by at most $(\delta+1)^{-1}\epsilon|V|$. The existence of such a partition is guaranteed by Lemma 2.10, by setting $\delta' = (\delta+1)^{-1}\epsilon$.
- If $\mathcal{Q}_i$ is equitable, has k_i' non-exceptional sets and its exceptional set has size at most $\epsilon|V|$, but it is not ϵ-regular, then construct $\mathcal{P}_{i+1}$ such that it has at most $k_i'4^{k_i'}$ non-exceptional parts, its exceptional set is the same as in $\mathcal{Q}_i$, refines $\mathcal{Q}_i$ and $q(\mathcal{P}_{i+1}) \geq q(\mathcal{Q}_i) + \delta^{-1}$. The existence of such a partition is guaranteed by Lemma 2.9.

We claim that the procedure produces a partition $\mathcal{Q}_i$ that is ϵ-regular for some $0 \leq i \leq \lfloor \delta \rfloor$. Assume the opposite, and we will reach a contradiction. First we will show that, if $\mathcal{Q}_i$ is not ϵ-regular for any of those values of i, then $\mathcal{Q}_i$ exists for $1 \leq i \leq \lfloor \delta \rfloor + 1$. If $\mathcal{Q}_i$ exists but $\mathcal{Q}_{i+1}$ does not, it is because $\mathcal{Q}_i$ is not equitable, or its exceptional set is bigger than $\epsilon|V|$. But $\mathcal{Q}_i$ is equitable by construction, so the first option is impossible.

The exceptional set of $\mathcal{Q}_i$ is the exceptional set of $\mathcal{P}_i$ with the addition of at most $(\delta+1)^{-1}\epsilon|V|$ vertices, and the exceptional set of $\mathcal{P}_i$ is the same as the one of $\mathcal{Q}_{i-1}$. Since $\mathcal{P}_0$ has an empty exceptional set, then $\mathcal{Q}_i$ has an exceptional set of size at most $(i+1)(\delta+1)^{-1}\epsilon|V|$, which for $i \leq \delta$ is less than or equal to $\epsilon|V|$. This implies that $\mathcal{Q}_i$ exists for $0 \leq i \leq \lfloor \delta \rfloor + 1$.

By Lemma 2.7, $q(\mathcal{Q}_i) \geq q(\mathcal{P}_i) \geq q(\mathcal{Q}_{i-1}) + \delta^{-1}$. Remember that $q(\mathcal{Q})$ is bounded between 0 and 1. Since $q(\mathcal{Q}_0) \geq 0$, by induction we obtain $q(\mathcal{Q}_i) \geq i\delta^{-1}$. But this means that $q\left(\mathcal{Q}_{\lfloor \delta \rfloor + 1}\right) \geq (\lfloor \delta \rfloor + 1)\,\delta^{-1} > 1$. This is a contradiction, so the partition $\mathcal{Q}_i$ is ϵ-regular for some $0 \leq i \leq \lfloor \delta \rfloor$.

$\mathcal{Q}_i$ refines $\mathcal{P}_i$, which in turn refines $\mathcal{Q}_{i-1}$. Since $\mathcal{P}_0$ refines $\mathcal{P}$, then the ϵ-regular partition constructed refines $\mathcal{P}$. Let $f(x) = (\delta+1)\epsilon^{-1}x$ and $g(x) = x4^x$. Then the partition $\mathcal{Q}_i$ has at most

$$M = f(g(f(g(f(\dots f(m)\dots)))))$$

non-exceptional parts, where f appears $\lfloor\delta\rfloor + 1$ times, and g appears $\lfloor\delta\rfloor$ times. M only depends on ϵ and m, so this value satisfies the conditions of the statement of lemma 2.4, and we are done. □

Observation: Ideally, once we fix m we would want M to grow as slowly as possible as ϵ tends to 0, but as we can see, this is not the case. The relation between M and ϵ is tower-like, that is, $M(\epsilon) = 4^{4^{\cdot^{\cdot^{4}}}}$, where the tower contains $O(\epsilon^{-5})$ layers. Using Knuth's arrow notation, this would be written as $M(\epsilon) = 4 \uparrow\uparrow O(\epsilon^{-5})$. This dependence is worse than what would be useful in most practical applications, so ϵ is usually treated as constant for this theorem. Conlon and Fox [**ConFox2**] showed, by finding a graph that whose smallest regular partition has that size, that it is impossible to obtain a lower bound less than tower type, in which the number of layers is at least $\Omega(\epsilon^{-1})$.

2.2. The counting lemma

The second part of the proof of the removal lemma consists of the proof of the counting lemma [**AloFisKriSze**]. The purpose of this lemma is to estimate the number of copies of a graph H in a graph G, which consists of several sets of vertices connected by fairly dense ϵ-regular bipartite graphs. The theorem says that, if each of the vertex sets of G contains m vertices, then the number of embedded copies of H in G grows like $m^{V(H)}$.

We will first start with some notation:

Definition 2.11. Let R be a graph and t be a positive integer. We denote by $R(t)$ the graph formed by replacing each vertex of R with an independent set of size t, and each edge with a complete bipartite graph $K_{t,t}$.

Definition 2.12. Let H and G be two graphs. Denote by $||H \to G||$ the number of embeddings[3] of H in G.

The proof will use this lemma:

Lemma 2.13. *Let G be a bipartite ϵ-regular graph on vertex sets X and Y. Let $d \leq d(X,Y)$. Let $Y' \subset Y$ with $|Y'| > \epsilon|Y|$. If $d > \epsilon$, then there are at most $\epsilon|X|$ vertices $x \in X$ such that $|N(x) \cap Y'| < (d-\epsilon)|Y'|$.*

Proof. Let X' be the set of vertices in X satisfying $|N(x) \cap Y'| < (d-\epsilon)|Y'|$. Each vertex of X' has less than $(d-\epsilon)|Y'|$ neighbours in Y', which means that $e(X',Y') < (d-\epsilon)|X'||Y'|$ and

[3] A morphism from H to G is an application $f : V(H) \to V(G)$ in which $f(v_i)f(v_j) \in E(G)$ for all $v_iv_j \in V(H)$. An embedding is an injective morphism.

$d(X',Y') < d - \epsilon$. If $|X'| \geq \epsilon|X|$, then by definition of regularity $\epsilon < |d(X',Y') - d| \leq |d(X',Y') - d(X,Y)| \leq \epsilon$, contradiction. This means that $|X'| \leq \epsilon|X|$. □

Now we are ready to state and prove the counting lemma. The proof follows the one found in **[KomSim]**:

Lemma 2.14 (Counting lemma). *Let $d > \epsilon > 0$ be two constants, let R be a graph and m be a positive integer. Let G be a graph produced by replacing each vertex of R with an independent set of size m, and each edge with an ϵ-regular bipartite graph with density at least d. Let H be a subgraph of $R(t)$ with h vertices and maximum degree $\Delta > 0$. Let $\delta = d - \epsilon$ and $\epsilon_0 = \frac{\delta^\Delta}{2+\Delta}$. If $\epsilon \leq \epsilon_0$ and $t - 1 \leq \epsilon_0 m$, then*

$$||H \to G|| \geq (\epsilon_0 m)^h$$

Observation: If we fix R, H, d and ϵ and let m grow, the number of vertices of G is $v(R)m$, and hence the maximum number of embeddings of H in G is $(v(R)m)^h$. This gives us the growth of $||H \to G||$ up to a constant: $||H \to G|| = \Theta(m^h)$.

Proof. The proof will be constructive. We begin by labelling the vertices of G as $u_1, u_2, ..., u_h$. To prove the result, we will embed the vertices u_i one by one in such a way that in every step there are at least $\epsilon_0 m$ choices for the embedding of the corresponding vertex, implying that the total number of embeddings is at least $(\epsilon_0 m)^h$. Actually, the bound on the number of choices that we obtain from the calculations is $(\delta^\Delta - \Delta\epsilon)m - (t-1)$, so the first step would be to prove that, under the hypothesis of the statement, this number is at least $\epsilon_0 m$. The following inequality implies this claim:

$$t - 1 + \epsilon_0 m \leq 2\epsilon_0 m = (\delta^\Delta - \Delta\epsilon_0)m \leq (\delta^\Delta - \Delta\epsilon)m$$

The procedure will go as follows: H is a subgraph of $R(t)$, so we denote by $v_{[i]}$ the vertex of R which produces u_i in $R(t)$, and by $V_{[i]}$ the set of vertices of G produced by $v_{[i]}$.

During our procedure, for $0 \leq j < i$, we will call $V_{i,j}$ the set of vertices of $V_{[i]}$ that are neighbours of v_k for all $k \leq j$ such that $u_k u_i \in E(H)$. The first sets are $V_{i,0} = V_{[i]}$. Note that $|V_{i,0}| = m$. The procedure goes like this: in the i-th step we choose as u_i any vertex from $V_{i,i-1}$ that has not been chosen before and which is adjacent to at least $\delta|V_{k,i-1}|$ vertices from every $V_{k,i-1}$ such that $u_i u_k \in E(H)$ and $k > i$.

We now want to show that the number of choices on each step is at least $\epsilon_0 m$. Consider vertex u_i. The value of $\frac{|V_{i,j}|}{|V_{i,j-1}|}$ is 1 if $v_i v_j \notin E(H)$ (because the set does not change, so $V_{i,j} = V_{i,j-1}$) and at least δ otherwise (by the choice of u_j). Since u_i has at most Δ neighbours, we obtain $|V_{i,i-1}| \geq \delta^\Delta |V_{i,0}| = \delta^\Delta m$.

Now let us see how many vertices from $V_{i,i-1}$ cannot be chosen as u_i. From the hypothesis of the statement, $\delta^\Delta m > \epsilon_0 m \geq \epsilon m$, which means that $|V_{k,i-1}| > \epsilon m$ for any $k > i$ such that $u_i u_k \in E(H)$. By Lemma 2.13, the number of vertices of $V_{i,i-1}$ that do not have at least $\delta|V_{k,i-1}|$ neighbours in $V_{k,i-1}$ is at most ϵm. Since u_i is adjacent to at most Δ vertices in H, then the number of discarded vertices for this reason is at most $\Delta\epsilon m$. In addition, at most $t-1$ vertices from $V_{[i]}$ have been chosen before, and hence at most $t-1$ from $V_{i,i-1}$. The number of choices for u_i is at least

$$|V_{i,i-1}| - \Delta\epsilon m - (t-1) \geq (\delta^{\Delta} - \epsilon m) - (t-1) \geq \epsilon_0 m$$

which brings the number of embeddings to at least $(\epsilon_0 m)^h$ □

2.3. The removal lemma

We now have all the necessary tools to prove the removal lemma. We remember the statement:

Theorem 2.1 (Removal lemma). Let $\epsilon > 0$ be a constant and H be a graph on h vertices. Then there exists $\delta(\epsilon, H) > 0$ for which the following property holds: any graph G on n vertices, with at most δn^h copies of H, can be made H-free by removing at most ϵn^2 edges.

A sketch of the proof, as given in [**ConFox1**], would go as follows: start by using the regularity lemma to find a μ-regular partition of the vertices of G, for an appropriate value of μ. Create a graph G^* by deleting from G the edges within pairs, the edges between non-regular pairs of parts and the edges in regular parts with density less than d. Observe that this only leaves edges between different and regular pairs of parts, each of which with density at least d. These are precisely the hypotheses to apply the counting lemma. For adequate values of μ and d, the number of removed edges is less than ϵn^2. Now, if G^* still has an embedding of H, we can use the counting lemma for m large enough (which is equivalent to n large enough once μ is fixed) to find a constant $\epsilon_0 = 2M\delta$ such that G^* has at least $(\delta n)^{v(H)}$ copies of H. Tweak the value of δ to account for small values of n and the result follows.

This is the proof with the details filled in:

Proof of theorem 2.1. Assume that $\epsilon \leq 1/2$, as otherwise the number of edges in G is less than ϵn^2. Also, assume that H contains at least one edge, and hence, two vertices. Let $\mu = \frac{(\epsilon/4)^{\Delta(H)}}{2+\Delta(H)} < \frac{\epsilon}{4}$. Then, on account of Lemma 2.4, there exists an integer M such that any graph G on at least $4\epsilon^{-1}$ vertices admits a μ- regular partition $\mathcal{P}$ on k non-exceptional parts, with $2\epsilon^{-1} \leq k \leq M$. Let m be the size of the non-exceptional parts, which satisfies $\frac{n}{2k} \leq \frac{n-|V_0|}{k} \leq m \leq \frac{n}{k}$.

Construct G^* from G as follows:

- Remove all edges having one or both of its ends in the exceptional set.
- Remove all edges with both endpoints in the same set
- Remove all edges between irregular pairs
- Remove all edges between regular pairs of density less than $d = \epsilon/2$.

We count the number of removed edges to see that the total is less than ϵn^2:

- The number of edges with at least one endpoint in the exceptional set is at most $|V_0||V| \leq \mu n^2 < \frac{\epsilon n^2}{4}$.

- The edges contained in the exceptional set were removed in the previous step. The number of edges contained inside the rest of the sets is at most $k\binom{m}{2} \le \frac{km^2}{2} \le \frac{n^2}{2k} \le \frac{\epsilon n^2}{4}$.
- There are at most $\mu k^2 \le \frac{\epsilon k^2}{4}$ irregular pairs, each one containing at most m^2 edges. The total number of edges is therefore at most $\frac{\epsilon k^2 m^2}{4} \le \frac{\epsilon n^2}{4}$
- We only consider pairs of different parts, as the pairs within the same part were already considered in the previous step. There are $\binom{k}{2} \le \frac{k^2}{2}$ pairs of different parts. A pair with density less than $\epsilon/2$ has at most $\frac{\epsilon m^2}{2}$ edges, so the number of edges that we remove in this step is at most $\frac{\epsilon m^2 k^2}{4} \le \frac{\epsilon n^2}{4}$

The number of edges removed in all four steps altogether is at most ϵn^2.

Now assume that H is a subgraph of G^*. Consider the graph R, where the vertices are the non-exceptional pairs of $\mathcal{P}$ and two vertices are connected if they are connected in G^* (in which case they are connected by a μ-regular bipartite graph of density at least d). Note that G^* is constructed from R following the procedure detailed in Lemma 2.14, and it is a subgraph of $R(m)$. If H is a subgraph of G^*, then it is also a subgraph of $R(v(H))$.

We will check that the hypotheses from the removal lemma are satisfied for m large enough. $d - \mu \ge \frac{\epsilon}{2} - \frac{\epsilon}{4} = \frac{\epsilon}{4}$. If $\mu_0 = \frac{(d-\mu)^\Delta}{2+\Delta}$, then $\mu_0 \ge \frac{(\epsilon/4)^{\Delta(H)}}{2+\Delta(H)} \ge \mu$ (this is the condition $\epsilon_0 \ge \epsilon$ from the counting lemma). If $m \ge \frac{v(H)-1}{\mu_0}$, then the other condition ($t-1 \le \epsilon_0 m$) is also satisfied. In this case, from the removal lemma,

$$||H \to G|| \ge ||H \to G^*|| \ge (\mu_0 m)^{v(H)} \ge \left(\frac{\mu_0 n}{2M}\right)^{v(H)}$$

To take care of the case $m < \frac{v(H)-1}{\mu_0}$, which means $n \le \frac{2M(v(H)-1)}{\mu_0}$, notice that in this case $\left(\frac{\mu_0 n}{2M(v(H)-1)}\right)^{v(H)} < 1$, so if $\delta \le \left(\frac{\mu_0 n}{2M(v(H)-1)}\right)^{v(H)}$, then any graph with $||H \to G|| \le \delta n^{v(H)}$ is H-free, so the removal lemma holds trivially in this case. Also, $\delta \le \left(\frac{\mu_0}{2M}\right)^{v(H)}$, so it also works for the case of large m.

To wrap up the whole proof, δ is a parameter that only depends on ϵ and H. If $m < \frac{2M(v(H)-1)}{\mu_0}$, then $\delta n^{v(H)} < 1$, so any graph with less than that many copies of H is H-free. If $m \ge \frac{2M(v(H)-1)}{\mu_0}$, then we find a μ-regular partition of G and construct G^* accordingly. G^* consists of removing at most ϵn^2 edges from G. If G, and therefore G^*, has less than $\delta n^{v(H)}$ copies of H, then it is H-free. This completes the proof of the removal lemma. □

The removal lemma admits many generalizations and variants. One possibility is the extension to sparse graphs (Lemma 3.34), which will be discussed in the next chapter, and requires a completely different approach. Another possible generalization is a removal lemma in which not any embedding of H in G counts, but only those embeddings satisfying some property. In this case, we can remove a bounded number of edges in such a way that it removes all the embeddings satisfying that property. For example, we can restrict the embedding of each vertex of H to a certain subset of $V(G)$:

Theorem 2.15 (Removal lemma on restricted sets). *Let H be a graph on h vertices, and $\epsilon > 0$. Let the vertices of H be $v_1, v_2, ..., v_h$. Then there exists $\delta > 0$ such that the following property holds: for any graph G on n vertices and any subsets $X_1, X_2, ..., X_h \subseteq V(G)$, denote by $||H \to G||_X$ the number of embeddings of H in G with $v_i \in X_i$ for all $1 \leq i \leq h$. If $||H \to G||_X \leq \delta n^h$, then it is possible to remove at most ϵn^2 edges from G to make $||H \to G||_X = 0$.*

The proof in this case is not too different to the proof in the previous case, it only requires a modification when taking $\mathcal{P}$:

Proof. It is enough to show this result for $\epsilon < 2^{2-h}$, as making ϵ smaller only makes the statement stronger. Take the coarsest partition $\mathcal{P}'$ that refines all X_i (as the sets X_i need not be disjoint). The number of parts of $\mathcal{P}'$ is at most $2^h < 4\epsilon^{-1}$. This means that we can find the μ-regular partition $\mathcal{P}^*$ refining $\mathcal{P}'$. Construct G^* in the same way as in the proof of 2.1. Now observe that if there is still a copy of H in G^* with its vertices in the corresponding X_i, then all copies generated by the counting lemma are also in the same sets (because $\mathcal{P}$ refines all sets X_i). Hence, the same bound $(\mu_0 m)^h$ applies in this case, and also the same δ. □

This theorem will be useful in the proof of the removal lemma for groups, in the next subsection. To illustrate a case in which theorem 2.15 can be applied but theorem 2.1 cannot, consider the following graph, for $H = C_4$.

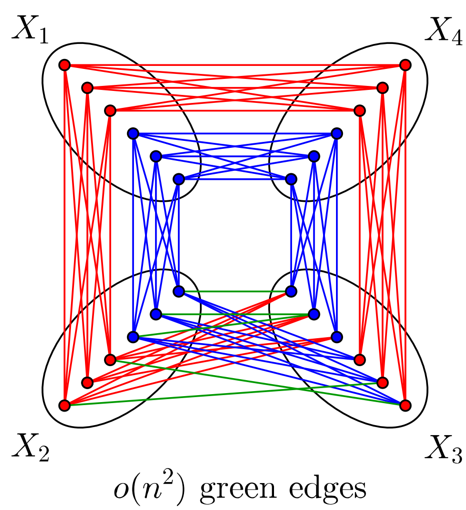

FIG. 2. Example of application of the removal lemma on restricted subsets

In the graph from Figure 2, only the cycles containing one vertex on each set are counted in $||H \to G||_X$. The number of copies of C_4 grows like $\Theta(n^4)$, but most of the copies have their vertices in two or three of the sets X_i. However, every cycle with each vertex in one set X_i must include a green edge, and since the number of green edges is $o(n^2)$, the number of cycles with vertices on all sets is $o(n^4)$. This means that the removal lemma can be applied here (the $o(n^2)$ edges that must be removed are the green edges).

2.4. Applications

We will now show some applications of the removal lemma. The two most important results from this section are Roth's theorem and the arithmetic removal lemma, but other results are included, either because they are used in the proof of those results or because they provide some insight into the possibilites that the removal lemma opens.

We begin with a result that can be obtained from the proof of the removal lemma:

Theorem 2.16. *Let H be a bipartite graph on h vertices, and $\epsilon > 0$. Then there exist N and δ such that the following holds: If G is a graph on n vertices, $n > N$, and $e(G) \geq \epsilon n^2$, then $||H \to G|| \geq \delta n^h$.*

Proof. Follow the proof of the remova lemma for $\epsilon' = \epsilon/2$. When we construct G^*, we remove at most $\frac{\epsilon}{2}n^2$ edges, so G^* has at least one edge. The graph H is a subgraph of $R(h)$, as this graph contains a copy of $K_{h,h}$ and $H \subset K_{h,h}$. For $m > \frac{h-1}{\mu_0}$, we obtain

$$||H \to G|| \geq ||H \to G^*|| \geq (\mu_0 m)^h \geq \left(\frac{\mu_0 n}{2M}\right)^h$$

Taking $N = \frac{2M(h-1)}{\mu_0}$ and $\delta = \left(\frac{\mu_0}{2M}\right)^h$ completes the proof. □

This result is an improvement over the Erdős-Stone thorem in the bipartite case [**ErdSto**], which asserts that, under the same hypotheses as in this theorem, $||H \to G|| > 0$. On the other hand, Sidorenko's conjecture claims that such an N exists for all $\delta < (2\epsilon)^{e(H)}$, which would be an improvement over this theorem. In a random Erdős-Rényi graph $G_{n,p}$ with constant probability $p = 2\epsilon$, the expected number of edges is $\epsilon n^2 + o(n^2)$, and the expected number of copies of H in G is $(2\epsilon)^{e(H)} n^h + o(n^h)$, so Sidorenko's conjecture says that the lowest possible asymptotic growth of $||H \to G||$ is precisely the expected value for random graphs. This conjecture has been proven for a wide family of bipartite graphs, including trees, hypercubes, grids, graphs with at most 4 vertices on one side of the partition [**ConFoxSud**] and graphs in which one vertex is adjacent to all the vertices in the other side of the partition [**Sze1**].

The next result, by Ruzsa ans Szemerédi [**Sol**, **RuzSze**], concerns induced matchings. In a graph G, a set of edges $\{e_1, ..., e_k\}$ forms an induced matching if the $2k$ endpoints of those edges are different, and the induced subgraph of G on those $2k$ vertices has only those k edges.

Lemma 2.17. *For any $\epsilon > 0$ there is $N > 0$ with the following property: if a graph G on $n > N$ vertices is the union of n induced matchings, then $e(G) \leq \epsilon n^2$.*

Proof. Let $v_1, v_2, ..., v_n$ be the vertices of G, and let M_1, M_2, ..., M_n be the matchings that form G. Suppose that each edge is contained in exactly one matching, otherwise remove it from every matching except one. Construct a graph G' as follows: take three sets of n vertices a_i, b_i and c_i. If $v_i v_j$ is an edge in M_k, then join a_i, b_j and c_k. The number of vertices is $n' = 3n$.

Now consider the triangles in this graph. There are some triangles of the form $a_i b_j c_k$, where $v_i v_j \in M_k$. In fact, there are exactly $2e(G)$ such triangles, two for each edge of G. Let us show that the edges of these triangles are disjoint. The edges $a_i b_j$ are all different because the edge $v_i v_j$

is in exactly one M_k. Also, the edges $a_i c_k$ are disjoint because, if one such edge appeared in two triangles $a_i b_{j_1} c_k$ and $a_i b_{j_2} c_k$, then $v_i v_{j_1}$ and $v_i v_{j_2}$ would both be in M_k, and then M_k would not be an induced matching. The same holds for the edges $b_j c_k$.

It is also true that those are the only triangles in G'. Indeed, any triangle in G' must contain one vertex from A, another from B and another from C, as otherwise it would be contained in a bipartite graph. If $a_i b_j c_k$ is a triangle in G' and $v_i v_j \notin M_k$, then M_k cannot be an induced matching, as M_k would contain an edge from v_i and an edge from v_j but not $v_i v_j$. The conclusion is that the only triangles are the ones described above. There are less than $n^2 = n'^2/9$ triangles.

Apply the removal lemma to $\epsilon' = 2\epsilon/9$ and $H = K_3$. This gives us a δ such that, if the number of triangles is less than $\delta n'^3$, then they can all be removed by removing $\epsilon' n'^2 = \frac{2\epsilon n'^2}{9} = 2\epsilon n^2$ edges. But since the triangles are edge disjoint, at least an edge must be removed from each triangle. For $n > \delta^{-1}$, we have $\delta n'^3 > \delta n^3 > n^2 > 2e(G)$, so the second condition must also apply, which means $2\epsilon n^2 \geq e(G)$ and $e(G) < \epsilon n^2$. $N = \delta^{-1}$ satisfies the condition from the statement. □

From this result, we can prove this other result, due to Ajtai and Szemerédi [**AjtSze**]. The proof presented here is due to Solymosi [**Sol**]:

Theorem 2.18 (Corners theorem). *For any $\epsilon > 0$ there exists N such that, for any $n > N$, any subset $S \in [n]^2$ of size at least ϵn^2 includes three elements of the form (a, b), $(a + d, b)$ and $(a, b + d)$ with $d \neq 0$.*

Proof. Suppose that S contains ϵn^2 elements from $[n]^2$ such that the configuration (a, b), $(a + d, b)$ and $(a, b + d)$ does not appear. We construct a graph G as follows: the set of vertices will be $v_1, v_2, ..., v_n, w_1, w_2, ..., w_n$. We join v_i and w_j iff $(i, j) \in S$. This graph has $2n$ vertices and $|S|$ edges.

Let M_k be the set of edges $v_i w_j$ such that $i + j = k$, for $1 \leq k \leq 2n$. This includes all edges of the graph. We claim that the sets M_k form an induced matching. Indeed, the endpoints of the edges of M_k are all different. Assume that $v_a w_y$ and $v_x w_b$ are two different edges from M_k and $v_a w_b \in E(G)$. Then, since $a + y = b + x$, we also have $x - a = y - b$. Setting this as d, we see that $d \neq 0$ and that (a, b), $(a + d, b)$ and $(a, b + d)$ are elements of S. This contradicts our initial hypothesis, so the sets M_k are induced matchings.

From lemma 2.17, there is N such that if $2n > N$, then $e(G) \leq \frac{\epsilon}{8}(2n)^2 = \frac{\epsilon}{2} n^2$. But since $e(G) = |S| \geq \epsilon n^2$, the only possibility is $2n \leq N$. □

The $d \neq 0$ from the statement can be replaced with a $d > 0$ using a symmetry argument. The proof goes as follows: consider the pairs of points $\{(p, q) | p, q \in S\}$. There are $|S|^2$ such pairs. The number of possible midpoints of the segment pq is $(2n - 1)^2$, as the coordinates of the midpoint are either an integer or half an integer from the interval $[1, n]$. By pigeonhole's principle, there are at least $\frac{|S|^2}{(2n-1)^2} \geq \frac{\epsilon^2 n^4}{4n^2} = \frac{\epsilon^2}{4} n^2$ pairs of points with the same midpoint m, and every point appears in at most twice. This means that there is a set $S_m \subseteq S$ with at least $\frac{\epsilon^2}{4} n^2$ points which is symmetric around m. By the corners theorem, there is an N such that for $n > N$ the set S_m contains a set of the form (a,b), $(a+d,b)$ and $(a,b+d)$. If $d<0$, then its symmetric about m is also of the same form and $d>0$.

This result can be used to prove Roth's theorem [**Rot**]:

Theorem 2.19 (Roth's theorem). *For any $\epsilon>0$ there is a positive integer N such that the following holds: for any integer $n>N$, any subset $S\subseteq\{1,2,...,n\}$ of size $|S|\geq\epsilon n$ contains a non-trivial 3-term arithmetic progression*[4]*.*

Proof. Let $S\subseteq\{1,2,...,n\}$ be a subset with size $|S|\geq\epsilon n$. Consider the square grid $G=\{-2n+1,-2n+2,...,n\}^2$. This grid contains $9n^2$ points. Now consider the set $T\subset G$ defined as $\{(x,y)\in G|x+2y\in S\}$. For every $1\leq y\leq n$ and every $s\in S$ there exists one x in $[-2n+1,n]$ such that $x+2y=s$. This means that $|T|\geq n|S|\geq\epsilon n^2$. According to theorem 2.18, for $n>N$ with N depending on ϵ only, the set T contains three elements of the form (a,b), $(a+d,b)$ and $(a,b+d)$. From the definition of T, the set S contains $a+2b$, $a+2b+d$ and $a+2b+2d$, which form a non-trivial arithmetic progression since $d\neq 0$. □

There is also another proof, this time coming directly from the removal lemma without going through the corners theorem. This proof can be found in [**KraSerVen**]:

Proof. Let $S\subseteq\{1,2,...,n\}$ be a subset with size $|S|\geq\epsilon n$. Assume that S does not contain a non-trivial 3-term arithmetic progression. Construct the following graph G: it will have three vertex sets, $\{a_i\}_{i=1}^{n}$, $\{b_j\}_{j=1}^{2n}$ and $\{c_k\}_{k=1}^{3n}$. The number of vertices in this graph is $6n$. Now, draw an edge a_ib_j if $j-i\in S$, an edge b_jc_k if $k-j\in S$ and an edge a_ic_k if $\frac{k-i}{2}\in S$.

Now, every triangle must contain a vertex from each set, otherwise it is contained in a bipartite graph. A triangle $a_ib_jc_k$ in G generates an arithmetic progression $j-i,\frac{k-i}{2},k-j$ in S. By our assumption, only triangles generating trivial arithmetic progressions (those with $j-i=\frac{k-i}{2}=k-j$) appear in G. These are of the form $i=a$, $j=a+s$ and $k=a+2s$ with $1\leq a\leq n$ and $s\in S$. There is a total of $n|S|$ triangles in G. Notice that the edges of those triangles are disjoint (any edge identifies a unique element s, and those triangles with the same s but different a do not share a vertex).

According to the removal lemma, there is a δ such that if G contains less than $\delta(6n)^3$ triangles then they can be removed by removing at most $\frac{\epsilon}{72}(6n)^2=\frac{\epsilon}{2}n^2$ edges from G. Since the number of triangles is $n|S|$, for $n>\delta^{-1}$ we have $n|S|\leq n^2<\delta(6n)^3$, so the lemma can be aplied. But the edges of the triangles are disjoint, so in order to remove all triangles one has to remove at least $n|S|\geq\epsilon n^2>\frac{\epsilon}{2}n^2$ edges. We conclude that, if there is no non-trivial 3-term arithmetic progression, then $n\leq\lfloor\delta^{-1}\rfloor=N$. □

One observation is that, because of the dependency of ϵ and δ in the regularity lemma, the bound that we prove for the N from Roth's theorem is $N(\epsilon)=4\uparrow\uparrow\Theta(\epsilon^{-5})$. This value is much greater than the value obtained by Roth in his origial proof, which was $N(\epsilon)=e^{e^{\Theta(\epsilon^{-1})}}$. Roth's method used Fourier analysis, rather than graph theory or other combinatorial methods.

There is an important generalization of Roth's theorem called Szemerédi's theorem:

[4] We say that an arithmetic progression is trivial if the difference between two consecutive terms is 0.

Theorem 2.20 (Szemerédi's theorem). *For any $\epsilon > 0$ and positive integer k there is a positive integer N such that the following holds: for any integer $n > N$, any subset $S \subseteq \{1, 2, ..., n\}$ of size $|S| \geq \epsilon n$ contains a non-trivial k-term arithmetic progression.*

Roth's theorem is the particular case $k = 3$. One could ask whether Szemerédi's theorem can be proved using the removal lemma, just as we did with Roth's theorem. The answer is that in order to prove Szemerédi's theorem a stronger removal lemma is required: the hypergraph removal lemma[5] [**ConFox1**], which was proven independently by Gowers [**Gow**], by Nagl, Rödl, Schacht and Skokan [**NagRodSch**, **RodSko**] and by Tao [**Tao**]:

Theorem 2.21 (Hypergraph removal lemma). *Let $\epsilon > 0$ be a real constant, k be a positive integer and H be a k-uniform hypergraph. Then there exists $\delta > 0$ such that the following property holds: if G is a k-uniform hypergraph on n vertices and $||H \to G|| \leq \delta n^h$, then G can be made H-free by removing at most ϵn^k hyperedges.*

The key in this theorem is how to generalize the concepts of ϵ-regularity and ϵ-regular partition. They must be weak enough that an ϵ-regular partition of bounded size always exists, yet it must also be strong enough that a counting lemma can be proven.

Finally, we present another application of the removal lemma, due to Green [**ConFox1**]. This time it considers subsets of a finite group The proof presented here is due to Král', Serra ans Vena.

Definition 2.22. Let G be a finite group, and let $S_1, S_2, ..., S_k$ be subsets of G. We denote by $C(S_1, S_2, ..., S_k)$ the number of solutions of $x_1 x_2 \cdots x_k = 1$ with $x_i \in S_i$ for all $1 \leq i \leq k$.

We observe that, if G contains n elements, then $C(G, G, ..., G) = n^{k-1}$. The result that we will introduce is called the group removal lemma, and it is similar to the graph removal lemma in the following sense: it states that when the number of solutions of $x_1 x_2 ... x_k$ is small, then they can all be removed by removing only a few elements from each S_i. Thus, the solutions would be analogous to copies of the graph H and the elements of the sets would be analogous to the edges of G.

Theorem 2.23 (Arithmetic removal lemma). *Let $\epsilon > 0$ be a constant and k be a positive integer. Then there exists a constant $\delta > 0$ for which the following holds: for any finite abelian group G with n elements and any subsets $S_1, S_2, ..., S_k \subseteq G$, if $C(S_1, S_2, ..., S_k) \leq \delta n^{k-1}$, then there are subsets $S_i' \subseteq S_i$ with $|S_i' \setminus S_i| \leq \epsilon n$ for which $C(S_1', S_2', ..., S_k') = 0$.*

The proof of this theorem [**KraSerVen**] uses the graph removal lemma in subsets (theorem 2.15) for $H = C_k$:

Proof. We will construct a graph K as follows: it will have k sets of vertices X_i, each of which containing n vertices. We will denote the vertices as $v_{i,g}$, where $1 \leq i \leq k$ denotes the set X_i it is in, and g is an element of G. We connect the vertex x_{i,g_1} to the vertex x_{i+1,g_2} if and only if $g_1^{-1} g_2 \in S_i$. We do the same for v_{k,g_1} and v_{1,g_2} (that is, we treat X_1 as X_{k+1}).

[5] A hypergraph consists of a vertex set V and a hyperedge set E, which is a set of nonempty subsets of V. In a k-uniform hypergraph, every hyperedge is a subset of V of size k.

We will now show that there is a correspondence between cycles $v_{1,g_1}v_{2,g_2}\cdots v_{k,g_k}$ and solutions of $x_1x_2\cdots x_k = 1$ with $x_i \in S_i$. If we take a cycle and make $x_i = g_i^{-1}g_{i+1}$, we see that $x_1x_2\cdots x_k = g_1^{-1}g_2g_2^{-1}g_3\cdots g_k^{-1}g_1 = 1$. There are n cycles that produce the same solution, as we can multiply all elements g_i by the same $g \in G$. Similarly, a solution $x_1x_2\cdots x_k$ produces n cycles in K of the form $v_{1,g_1}v_{2,g_2}\cdots v_{k,g_k}$. with $g_i = gx_1x_2\cdots x_{i-1}$ for any $g \in G$. We conclude that $||C_k \to K||_X = nC(S_1, S_2, ..., S_k) \leq \delta n^k$.

Using Theorem 2.15, we find a value of δ such all the cycles from K with the vertices in the respective X_i can be removed by removing $\frac{\epsilon}{k}n^2$ edges. Let E' be the set of these removed edges. To produce the sets S_i', we remove $x_i \in S_i$ if there are at least n/k edges of the form $v_{i,g_1}v_{i+1,g_2}$ with $g_1^{-1}g_2 = s_i$ in E'. The total number of removed elements is at most $\frac{\epsilon n^2/k}{n/k} = \epsilon n$.

Assume that $G(S_1', S_2', ..., S_k') \neq 0$. Then there is a solution $x_1x_2\cdots x_k = 1$ with $x_i \in S_i'$. Consider the n cycles generated by this solution (those with vertices v_{i,g_i} with $g_i = gx_1x_2\cdots x_{k-1}$). These cycles are vertex-disjoint and, therefore, edge-disjoint. By construction of E', there is an edge from each of these cycles in E'. By pigeonhole principle, there is an i for which there are at least n/k edges of the form $v_{i,g_i}v_{i+1,g_{i+1}}$, which satisfy $g_i^{-1}g_{i+1} = x_i$. But this means that x_i got removed when creating the set S_i', contradiction. This means that $G(S_1', G_2', ...G_k') = 0$, and this completes the proof. □

This result can lead to yet another proof of Roth's theorem, by taking $k = 3$, $G = \mathbb{Z}/(4n+1)\mathbb{Z}$, $S_1 = S_3 = S$ and $S_2 = \{-2s|s \in S\}$. It can be checked that, in this case, if $x_1 + x_2 + x_3 = 0$ in G, then $x_1 + x_2 + x_3 = 0$ in $\mathbb{Z}$, as $|x_1 + x_2 + x_3| \leq 4n$. This means that the elements of S that generate them form an arithmetic progression. But since the number of trivial arithmetic progressions is linear in n, and removing them requires removing a linear number of elements (all of them) from S, the hypotheses of theorem 2.23 cannot hold. This means that the number of solutions of $x_1 + x_2 + x_3 = 0$ must be quadratic, and for $n > N$ there is always a non-trivial solution.

Chapter 3
Sparse pseudorandom graphs

This section will deal with sparse graphs, that is, families of graphs with $e(G) = o(n^2)$. We will first examine, in Section 3.1, why the removal lemma from Chapter 2 is not adequate for use in sparse graphs, and why certain sparse versions only hold on certain families of sparse graphs. In Section 3.2, we will introduce the concept of pseudorandomness and some properties related to it, in particular jumbledness and discrepancy. Later, in Section 3.5 we will prove the removal lemma in the case where H is a cycle graph of length at least 5. Finally, in Section 3.6 we will provide a sparse pseudorandom version of the group removal lemma.

3.1. Motivation

The removal lemma (Theorem 2.1) holds for all graphs G, there is no restriction on the number of edges or its distribution. However, in some cases, specifically in sparse graphs, it only provides a trivial result. Let us see why. Indeed, the theorem claims that, under certain hypotheses, the graph G can be made H-free by removing at most ϵn^2 edges from G. But the number of edges in G is less than ϵn^2 for all but a finite number of graphs in any family of graphs with $o(n^2)$ edges, which means that the lemma allows us to remove all edges from the graph, therefore removing all copies of H.

Another option would be to treat δ in Theorem 2.1 as a parameter depending on ϵ, and make ϵ tend to 0. But then we run into another problem: the dependency between the two parameters is tower-like, which is too big for any practical applications (we could only apply the result if $||H \to G||$ grows slower than $\frac{cn^{v(H)}}{\exp(\exp(\dots\exp(1)\dots))}$, where there are $\Theta(\epsilon^{-5})$ exponentiations).

Part of the reason why this result does not translate well to sparse graphs has to do with regularity itself, the key concept from the proof. Let us look back at the definition of regularity. A pair (X, Y) is ϵ-regular if, for any $X' \subset X$ and $Y' \subset Y$ with $|X'| \geq \epsilon|X|$ and $|Y'| \geq \epsilon|Y|$, we have $|d(X', Y') - d(X, Y)| \leq \epsilon$. There are two problems with this: first, if $d(X, Y) \leq \epsilon^3$, then the pair (X, Y) is automatically ϵ-regular, regardless of the distribution of the edges, because

$$d(X', Y') = \frac{e(X', Y')}{|X'||Y'|} \leq \frac{e(X, Y)}{\epsilon^2|X||Y|} \leq \frac{\epsilon^3}{\epsilon^2} = \epsilon$$

Another problem is that, when $d(X,Y) < \epsilon$ we do not know anything about the ratios between densities: If $|X_1|, |X_2| \geq \epsilon|X|$ and $|Y_1|, |Y_2| \leq \epsilon|Y|$, then it is possible that, while $|d(X_1, Y_1) - d(X,Y)|$ and $|d(X_2, Y_2) - d(X,Y)|$ are both smaller than ϵ, the ratio $\frac{d(X_1,Y_1)}{d(X_2,Y_2)}$ can be arbitrarily large. Thus the concept of regularity is not as useful as when $d(X,Y) >> \epsilon$.

The conclusion is that another version of the removal lemma is required if we want to apply it to sparse graphs. Here is a proposed statement that, if true, would generalize Theorem 2.1:

Statement. Let $\epsilon > 0$ be a constant and H be a graph. Then there exists $\delta > 0$ for which the following property holds: for any two graphs G and Γ such that $G \subseteq \Gamma$, if $||H \to G|| \leq \delta ||H \to \Gamma||$, then G can be made H-free by removing at most $\epsilon e(\Gamma)$ edges.

The statement is a generalization of Theorem 2.1 because it is the particular case $\Gamma = K_n$. Indeed, any graph G on n vertices is a subgraph of K_n, the number of edges of K_n is $e(K_n) = \Theta(n^2)$ and the number of copies of H is $||H \to K_n|| = \Theta(n^{v(H)})$.

However, we can construct a counterexample to this statement. Consider the two graphs from Figure 1:

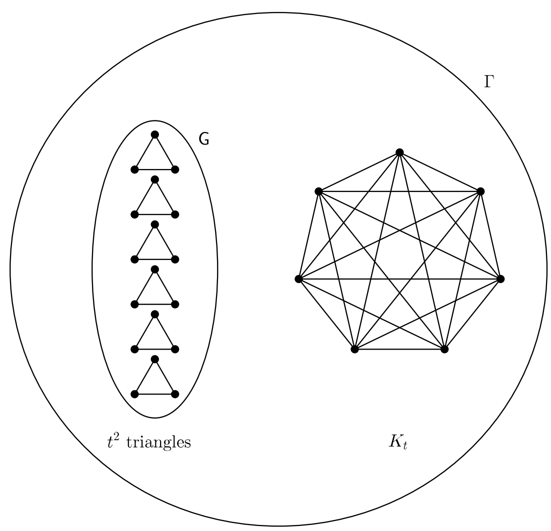

FIG. 1. Counterexample to the suggested generalization of the removal lemma

The graph Γ consists of t^2 disjoint triangles and a complete graph K_t, and the graph G consists only of the triangles from Γ. If we consider the case $H = K_3$, the relevant values from the statement are these:

$$e(\Gamma) < 4t^2 \qquad e(G) = 3t^2 \qquad ||H \to \Gamma|| > \binom{t}{3} = \Theta(t^3) \qquad ||H \to G|| = t^2$$

Now consider the case $\epsilon = 1/8$. For any fixed δ there is a value of t for which $||H \to G|| < \delta ||H \to \Gamma||$, which means that the triangles of G can be removed by removing at most $\epsilon e(\Gamma) \leq t^2/2$

edges from G. But, since the edges of the triangles of G are disjoint, we need to remove at least t^2 edges from G. This contradicts the statement.

We need more hypotheses in order to obtain a removal lemma similar to the statement that we proposed. One case in which we know that the generalization holds is the case where Γ is dense, that is, once we fix $\mu > 0$, we can apply it to any Γ with $e(\Gamma) \geq \mu n^2$. This can be deduced from Theorem 2.1, by noticing that $||H \to \Gamma|| = O(n^h)$ and setting $\epsilon' = \mu\epsilon$.

It seems reasonable to only impose restrictions on Γ rather than on G, to mimic the dense case. Let us look again at the example. We can see that the graph Γ has more than one eighth of its edges concentrated in a set S of vertices of size $o(|V|)$, and, as a consequence, most of the triangles in Γ are in this small set. This gives us an intuitive idea about what might be reasonable to impose on Γ: some condition that implies that the edges are well distributed among the vertices.

One option would be to make Γ a random graph $G_{n,p}$. In this case, there is a theorem by Conlon and Gowers [**ConGow**]. In this theorem, $m_2(H)$ is defined as the minimum of $\frac{e(H')-1}{v(H')-2}$, where H' is a proper induced subgraph of H. A graph is said to be strictly balanced if $m_2(H) > m_2(H')$ for any proper subgraph of H. All cycles and complete graphs, among others, are strictly balanced graphs.

Theorem 3.1. *For any strictly balanced graph H on h vertices and any $\epsilon > 0$, there exist positive constants δ and C such that, for $p \geq Cn^{-\frac{1}{m_2(H)}}$ the following holds a.a.s.: every subgraph G of $G_{n,p}$ with $||H \to G|| \leq \delta p^{e(H)} n^h$ can be made H-free by removing at most ϵpn^2 edges from G.*

In this theorem, $e(\Gamma)$ and $||H \to \Gamma||$ have been replaced by their expected values, since the distribution of those values is concentrated around the expected value (they have small standard deviations).

For any random graph Γ, however, we can only say that the theorem holds asymptotically, with a probability approaching 1. This does not allow us to determine whether or not we can apply the theorem to a particular graph. We would like to find a deterministic family of graphs Γ for which the generalization of the removal lemma always holds. Since we know that it holds for random graphs, we will consider a family of graphs that simulates the behaviour of random graphs: pseudorandom graphs or, more specifically, jumbled graphs.

3.2. Pseudorandom graphs

Pseudorandom graphs are graphs that attempt to replicate, deterministically, properties that random graphs satisfy a.a.s. There are several different measurements of pseudorandomness. In this section we will look at three: jumbledness, regularity and discrepancy. The idea behind all four is the same: restricting the density or the number of edges in certain subsets of vertices. The difference between them will be the relations between the limitations imposed and the size of the subsets.

We begin with the definition of jumbledness:

Definition 3.2 ((p, β)-jumbledness). Let G be a graph, and let p and β be positive constants. We say that G is (p, β)-jumbled if, for any two subsets $X', Y' \subseteq V(G)$, we have

$$\left|e(X', Y') - p|X'||Y'|\right| \leq \beta\sqrt{|X'||Y'|}$$

If G is a bipartite graph with stable vertex sets X and Y, we say that G is (p, β)-jumbled if the same condition holds for any subsets $X' \subseteq X$, $Y' \subseteq Y$.

For convenience, we say that a graph is $(p, \gamma = x)$-jumbled if it is (p, β)-jumbled for $\beta = x\sqrt{|X||Y|}$.

Let us see the meaning of each parameter. p in this definition is approximately equal to the density of the graph G: the number of edges $e(X', Y')$ is roughly the same as $p|X'||Y'|$, so $d(X, Y) \approx p$. This role is the same as in the random graph $G_{n,p}$. The parameter β is a measurement of how jumbled our graph is: a smaller β means that the error allowed in the number of edges is smaller, and consequently the edges are more evenly distributed. The parameter γ is similar to β, with the difference that, as we will see later, when we state our theorems the parameter γ will not depend on the size of the graph.

Every non-empty (p, β)-jumbled graph has $\beta > 0$. The complete graph on n vertices is $(1, 1)$-jumbled, because $|e(X', Y') - |X'||Y'|| = |X' \cap Y'| \leq \min\{|X'|, |Y'|\} \leq \sqrt{|X'||Y'|}$. For any fixed $\epsilon > 0$, any family of (p, β)-jumbled graphs with $p = d(V, V) \leq 1 - \epsilon$ satisfies $\beta = \Omega(\sqrt{pn})$. Let us see why. By double counting, pn is the average degree of the vertices of G, so there is a vertex $x \in V(G)$ with $|N(x)| \geq pn$. Then, by setting $X' = \{x\}$ and $Y' = N(x)$ we obtain

$$\beta \geq \frac{|e(X', Y') - p|X'||Y'||}{\sqrt{|X'||Y'|}} = \frac{||N(x)| - p|N(x)||}{\sqrt{|N(x)|}} = (1 - p)\sqrt{|N(x)|} \geq \epsilon\sqrt{np}$$

For a fixed p, the random graph $G_{n,p}$ is a.a.s. (p, β)-jumbled, with $\beta = O(\sqrt{pn})$, which means that it is optimally jumbled.

A similar concept is that of uniformity. This definition is similar to the definition of regularity, but in this chapter it will satisfy a very different role.

Definition 3.3 ((p, η)-uniformity). We say that a graph G is (p, η)-uniform if, for any two subsets $X', Y' \subseteq V$ satisfying $|X'|, |Y'| \geq \eta|V(G)|$, we have

$$|d(X', Y') - p| \leq \eta p$$

If G is a bipartite graph on vertex sets X and Y, we say that G is (p, η)-uniform if the same condition holds for any subsets $X' \subseteq X$, $Y' \subseteq Y$ with $|X'| \geq \eta|X|$ and $|Y'| \geq \eta|Y|$.

For $\beta = \Theta(pn)$ and $\eta = \Theta(1)$, the (p, β)-jumbledness condition is stronger than (p, η)-uniformity:

Lemma 3.4. *For every $\eta > 0$ there exists $c > 0$ such that the following holds: any (p, cpn)-jumbled graph is (p, η)-uniform.*

Proof. We consider $c = \eta^2$. Then, for any $X', Y' \subseteq V(G)$ satisfying $|X'|, |Y'| \geq \eta|V(G)|$ we have

$$|d(X', Y') - p| = \frac{|e(X', Y') - p|X'||Y'||}{|X'||Y'|} \leq \frac{\beta\sqrt{|X'||Y'|}}{|X'||Y'|} = \frac{\beta}{\sqrt{|X'||Y'|}} \leq \frac{\eta^2 pn}{\eta n} = \eta p$$

□

The same result holds for bipartite graphs, for $\beta = cp\sqrt{|X||Y|}$.

Finally, the last kind of pseudorandomness that we will introduce in this section is discrepancy:

Definition 3.5 (DISC(q, p, ϵ)). We say that a bipartite graph G on vertex sets X and Y satisfies $DISC(q, p, \epsilon)$ if, for any $X' \subseteq X$ and $Y' \subseteq Y$, we have

$$|e(X', Y') - q|X'||Y'|| \leq \epsilon p|X||Y|$$

We say that G satisfies DISC$_{\geq}(q, p, \epsilon)$ if, under the same conditions,

$$e(X', Y') - q|X'||Y'| \geq -\epsilon p|X||Y|$$

The role of discrepancy will be the same as regularity satisfied in the dense case. The proof of the removal lemma will consist on finding a partition in which most pairs of parts satisfy discrepancy, and show a counting lemma for graphs satisfying discrepancy. For the counting lemma, we will only use one-sided discrepancy (DISC$_{\geq}$): we will impose that the graph does not have subsets too sparse, and we will allow subsets too dense.

We notice that, from the discrepancy condition, if $\epsilon_1 \leq \epsilon_2$, then every graph satisfying (q, p, ϵ_1)-DISC also satisfies (q, p, ϵ_2)-DISC. The same happens with the parameter ϵ in DISC$_{\geq}$, with β and γ in jumbledness, and with η in uniformity.

Now we state the version of the removal lemma that we will prove. We consider $t_3 = 3$, $t_4 = 2$, $t_\ell = 1 + \frac{1}{\ell-3}$ for odd $\ell \geq 5$ and $t_\ell = 1 + \frac{1}{\ell-4}$ for even $\ell \geq 6$.

Theorem 3.34 (Removal lemma for pseudorandom graphs). For every integer $\ell \geq 5$, and every $\mu > 0$ there are $\delta > 0$ and $c > 0$ for which the following holds: let $X_1, X_2, ..., X_\ell$ be vertex sets, each with n vertices. Let Γ be a graph for which $(X_i, X_{i+1})_\Gamma$ is $(p, \gamma = cp^{t_\ell})$-jumbled for all $1 \leq i \leq \ell$, and let G be a subgraph of Γ. If $||C_\ell \to G||_X \leq \delta p^\ell n^\ell$, then it is possible to remove at most $\mu p n^2$ edges from G so that $||C_\ell \to G||_X = 0$.

3.3. The regularity lemma

This section will state and prove the sparse version of the regularity lemma. Most of the concepts and proofs are analogous to the ones from Section 2.1.

In this proof we will have a graph G which is subgraph of a graph of Γ. For this reason, we will need the notion of density within a graph:

Definition 3.6. Let G and Γ be two graphs such that G is a subgraph of Γ, and let X and Y be two subsets of $V(G)$. Then we define

$$d_{G,\Gamma}(X, Y) = \frac{e_G(X, Y)}{e_\Gamma(X, Y)} = \frac{\frac{e_G(X,Y)}{|X||Y|}}{\frac{e_\Gamma(X,Y)}{|X||Y|}} = \frac{d_G(X, Y)}{d_\Gamma(X, Y)}$$

If $e_\Gamma(X, Y) = 0$ (which implies $e_G(X, Y) = 0$) we define $d_{G,\Gamma}(X, Y) = 0$

With this definition it is easy to see that $0 \leq d_{G,\Gamma}(X,Y) \leq 1$. In this section, both for convenience and to highlight the difference between the two, we will denote $d(X,Y) = d_{G,\Gamma}(X,Y)$ and $r(X,Y) = d_\Gamma(X,Y)$.

We already introduced the concept of discrepancy in the previous section, so now we introduce ϵ-discrepant partitions:

Definition 3.7 (ϵ-DISC partition). Let G be a graph, $p \in [0,1]$ be a parameter and $\mathcal{P} = \{V_0, V_1, ..., V_k\}$ be a partition of $V(G)$, with exceptional set V_0. We say that $\mathcal{P}$ satisfies ϵ-DISC if the following holds:

- $|V_1| = |V_2| = ... = |V_k|$
- $|V_0| \leq \epsilon v(G)$
- All but at most ϵk^2 pairs of parts (V_i, V_j) with $1 \leq i,j \leq k$ satisfy $DISC(q_{i,j}, p, \epsilon)$ for some $q_{i,j}$.

We can see that now the condition that (V_i, V_j) needs to satisfy has a difference with respect to the one in regularity pairs: it depends on a parameter $q_{i,j}$. However, this parameter will generally be very close to $d_G(V_i, V_j)$, which makes it somewhat similar to the regular case. Moreover, a simple calculation shows that any DISC$(q_{i,j}, p, \epsilon)$ pair is also $(d_G(V_i,V_j), p, 2\epsilon)$-DISC.

In this version of the regularity lemma, Γ will be a jumbled (or uniform) graph, while G will be a graph in which we will want to find a partition satisfying the discrepancy condition. This is the reason why, when we defined DISC, we included three parameters (q,p,ϵ): q will measure the density of G, while p will measure the density of Γ. ϵ, as in the case of regularity, will be the parameter that measures how low the discrepancy is.

Now we state our sparse version of the regularity lemma:

Lemma 3.8. *For every $\epsilon > 0$ and every positive integer m there exist a constant $c > 0$ and a positive integer M such that, if G is a graph with at least m vertices which is a subgraph of graph Γ, if Γ is a (p,c)-uniform graph, then G admits an ϵ-DISC partition, with the same value of p, into k non-exceptional parts, with $m \leq k \leq M$. Moreover, if $\mathcal{P}$ is a fixed partition of $V(G)$ with at most m parts, then such an ϵ-DISC partition can be found in a way that each set X_i with $1 \leq i \leq k$ is contained in one of the sets of $\mathcal{P}$.*

Comparing this result to Lemma 2.4, we see that, other than the fact that we now have discrepancy instead of regularity, the biggest difference is that we now impose that the graph G is a subgraph of a graph satisfying a certain condition, which is uniformity.

The identity below will have a crucial role in the proof of this version of the regularity lemma:

Lemma 3.9. *Let Γ be a bipartite graph on stable vertex sets X and Y, let G be a subgraph of Γ, and let $\mathcal{X} = X_1 \cup X_2 \cup ... \cup X_a$ and $\mathcal{Y} = Y_1 \cup Y_2 \cup ... \cup Y_b$ be partitions of X and Y. Then*

$$\sum_{i=1}^{a}\sum_{j=1}^{b} e_\Gamma(X_i,Y_j)d\left(X_i,Y_j\right) = \sum_{i=1}^{a}\sum_{j=1}^{b} e_\Gamma(X_i,Y_j)d\left(X,Y\right) \quad (3)$$

Recall that $d(X,Y) = \frac{e_G(X,Y)}{e_\Gamma(X,Y)}$.

Proof. The equality comes from the fact that each edge of G is contained in exactly one graph $G|_{X_iY_j}$, so by double counting,

$$\begin{aligned}\sum_{i=1}^{a}\sum_{j=1}^{b} e_\Gamma(X_i,Y_j)d\left(X_i,Y_j\right) =& \sum_{i=1}^{a}\sum_{j=1}^{b} e_G(X_i,Y_j)\\ =& e_G(X,Y)\\ =& e_\Gamma(X,Y)d(X,Y)\\ =& \sum_{i=1}^{a}\sum_{j=1}^{b} e_\Gamma(X_i,Y_j)d\left(X,Y\right)\end{aligned}$$

□

As in the identity (1), the identity (3) comes from double counting, this time of $e_G(X,Y)$. The terms $e_\Gamma(X_i,Y_j)$ are non-negative, so applying Jensen's inequality to a convex function $f:[0,1]\to\mathbb{R}$ with those terms as weights yields

$$\sum_{i=1}^{a}\sum_{j=1}^{b} e_\Gamma(X_i,Y_j)f(d\left(X_i,Y_j\right)) \geq \sum_{i=1}^{a}\sum_{j=1}^{b} e_\Gamma(X_i,Y_j)f(d\left(X,Y\right)) \tag{4}$$

The analogous definition to quadratic mean density (Definition 2.6) is the following:

Definition 3.10. Let Γ be a graph on vertex set V, with $|V|=n$, and G be a subgraph of Γ. Let $X,Y\subset V$. We define

$$q(X,Y) := \frac{e_\Gamma(X,Y)}{n^2}d^2(X,Y)$$

Let $\mathcal{X}$ and $\mathcal{Y}$ be partitions of sets X,Y. Then

$$q(\mathcal{X},\mathcal{Y}) := \sum_{\substack{X_i\in\mathcal{X}\\ Y_j\in\mathcal{Y}}} q(X_i,Y_j)$$

If $\mathcal{P}$ is a partition of V without exceptional set, then

$$q(\mathcal{P}) := q(\mathcal{P},\mathcal{P})$$

If $\mathcal{P}=X_0\cup X_1\cup ...\cup X_k$ is a partition of V with exceptional set X_0, then

$$q(\mathcal{P}) := q(\tilde{\mathcal{P}})$$

(Remember the definition of $\tilde{\mathcal{P}}$ from Definition 2.6)

This function satisfies that, for any partition $\mathcal{P}$ of V, then $0\leq q(\mathcal{P})\leq\frac{e_G(V,V)}{n^2}$, since the partition in which we take each vertex individually refines $\mathcal{P}$ and, as we will see, refining a partition does not decrease $q(\mathcal{P})$ (we do not use the inequality in the proof of the property for refinements in Lemma 3.11, so we do not run into a circular reasoning). If Γ is (p,c)-uniform with $c<1$, then $q(\mathcal{P})\leq\frac{e_G(V,V)}{n^2}\leq\frac{e_\Gamma(V,V)}{n^2}\leq\frac{2p|V|^2}{n^2}=2p$.

The proof of the regularity lemma will be based on that of the sparse case from [**Die**], using ideas from [**Koh**]. It will consist of the same three steps as in Section 2.1, namely:

- If $\mathcal{P}'$ is a refinement of $\mathcal{P}$, then $q(\mathcal{P}') \geq q(\mathcal{P})$
- If $\mathcal{P}$ is an equitable partition in k parts with small exceptional set and it is not ϵ-DISC, then there is a refinement in at most $k4^k$ parts which does not increase the size of the exceptional set and with $q(\mathcal{P}') \geq q(\mathcal{P}) + \frac{\epsilon^3 p}{32}$.
- If $\mathcal{P}$ is a partition in k parts and $\delta > 0$, then there is a refinement of $\mathcal{P}$ in at most $\delta^{-1}k$ parts which is equitable and which increases the size of the exceptional set by at most δn.

Once we have these three steps we can complete the proof in a similar fashion as in the dense case. The only step with substantial differences is the second step. For the first step, a simple calculation is enough:

Lemma 3.11. *Let X and Y be sets of vertices of $G \subseteq \Gamma$, and let $\mathcal{A}$ and $\mathcal{A}'$ be two partitions of X and $\mathcal{B}$ and $\mathcal{B}'$ be two partitions of Y such that $\mathcal{A}'$ refines $\mathcal{A}$ and $\mathcal{B}'$ refines $\mathcal{B}$. Then $q(\mathcal{A}', \mathcal{B}') \geq q(\mathcal{A}, \mathcal{B})$. (Partitions may have exceptional sets)*

Proof. By expanding the formulas for $q(\mathcal{A}, \mathcal{B})$ and $q(\mathcal{A}', \mathcal{B}')$, we obtain

$$\begin{aligned}
q(\mathcal{A}', \mathcal{B}') &= \sum_{A' \in \tilde{\mathcal{A}}'} \sum_{B' \in \tilde{\mathcal{B}}'} q(A', B') \\
&= \sum_{A \in \tilde{\mathcal{A}}} \sum_{B \in \tilde{\mathcal{B}}} \sum_{\substack{A' \in \tilde{\mathcal{A}}' \\ A' \subset A}} \sum_{\substack{B' \in \tilde{\mathcal{B}}' \\ B' \subset B}} q(A', B') \\
&= \sum_{A \in \tilde{\mathcal{A}}} \sum_{B \in \tilde{\mathcal{B}}} \sum_{\substack{A' \in \tilde{\mathcal{A}}' \\ A' \subset A}} \sum_{\substack{B' \in \tilde{\mathcal{B}}' \\ B' \subset B}} \frac{e_\Gamma(A', B')}{n^2} d^2(A', B') \\
&\overset{(4)}{\geq} \sum_{A \in \tilde{\mathcal{A}}} \sum_{B \in \tilde{\mathcal{B}}} \sum_{\substack{A' \in \tilde{\mathcal{A}}' \\ A' \subset A}} \sum_{\substack{B' \in \tilde{\mathcal{B}}' \\ B' \subset B}} \frac{e_\Gamma(A', B')}{n^2} d^2(A, B) \\
&= \sum_{A \in \tilde{\mathcal{A}}} \sum_{B \in \tilde{\mathcal{B}}} \frac{e_\Gamma(A, B)}{n^2} d^2(A, B) \\
&= \sum_{A \in \tilde{\mathcal{A}}} \sum_{B \in \tilde{\mathcal{B}}} q(A, B) \\
&= q(\mathcal{A}, \mathcal{B})
\end{aligned}$$

□

The second step is where we require more work than in the previous case, and where uniformity of graph Γ will come into place:

Lemma 3.12. *For every $0 < \epsilon < \frac{1}{2}$ and integer k, there is $c > 0$ such that the following holds: let $\mathcal{P} = \{X_i\}_{i=0}^k$ be an equitable partition of $V(G)$ with exceptional set X_0 and k non-exceptional sets. If G is a subset of a (p, c)-uniform graph Γ, $\mathcal{P}$ satisfies $|X_0| < \epsilon |V|$ and it is not ϵ-DISC, with the same value of p, then there is another partition $\mathcal{P}'$ with at most $k4^k$ non-exceptional parts, the same exceptional set X_0 and $q(\mathcal{P}') \geq q(\mathcal{P}) + \frac{\epsilon^3 p}{32}$.*

Proof. Let $S = \{(i,j) \in [k]^2 : (X_i, X_j)$ is not (q,p,ϵ)-DISC for any value of $q\}$. If $\mathcal{P}$ is not ϵ-DISC, then $\epsilon k^2 \leq |S| \leq k^2$. For every pair (i,j) that is not DISC, by definition of discrepancy, there are sets $X_i^j \subset X_i$ and $X_j^{[i]} \subset X_j$ such that $\left|e_G(X_i^j, X_j^{[i]}) - \frac{e_G(X_i,X_j)}{|X_i||X_j|}|X_i^j||X_j^{[i]}|\right| \geq \epsilon p|X_i||X_j|$ (if the DISC condition does not hold for any $q_{i,j}$, in particular it does not hold for $q_{i,j} = \frac{e_G(X_i,X_j)}{|X_i||X_j|}$).

Now take $\mathcal{P}'$ to be the coarsest partition that refines all the sets X_i^j and $X_j^{[i]}$. Within each set X_i there are at most k sets X_i^j and k sets $X_i^{[j]}$, which means that the coarsest partition of X_i that refines all those sets has at most $2^{2k} = 4^k$ sets, so the partition $\mathcal{P}'$ requires no more than $k4^k$ non-exceptional sets. Denote by $\tilde{\mathcal{P}}'(X)$ the partition of $X \in \tilde{\mathcal{P}}$ in $\tilde{\mathcal{P}}'$, and by $\mathcal{P}_i(X)$ the partition of X_i into two sets induced by $X \subset X_i$.

We claim that, if Γ is (p,c)-uniform with $c \leq \frac{\epsilon}{8k}$ then $|X_i^j|, |X_j^{[i]}| \geq cn$. Indeed, if two sets $Y_i \subset X_i$ and $Y_j \subset X_j$ satisfy $\left|e_G(Y_i,Y_j) - q_{i,j}|Y_i||Y_j|\right| \geq \epsilon p|X_i||X_j|$ (that is, they are a counterexample to DISC), then either

(1) $e_G(Y_i,Y_j) \geq \epsilon p|X_i||X_j|$ or
(2) $q_{i,j}|Y_i||Y_j| \geq \epsilon p|X_i||X_j|$.

In the second case, $q_{i,j} = \frac{e_G(X_i,X_j)}{|X_i||X_j|} \leq \frac{e_\Gamma(X_i,X_j)}{|X_i||X_j|} \overset{\text{uniform}}{\leq} (1+c)p \leq 2p$, which implies

$$|Y_i| \overset{Y \subseteq X}{\geq} \frac{|Y_i||Y_j|}{|X_j|} \overset{(2)}{\geq} \frac{\epsilon p|X_i||X_j|}{q_{i,j}|X_j|} \geq \frac{\epsilon|X_i|}{2} \geq \frac{\epsilon n}{4k} \geq cn$$

The same works for Y_j. In the first case, assume that Y_i has size less than cn, so it is contained in a set Y_i' of size $cn \leq |Y_i'| \leq 2cn$. Then

$$e_G(Y_i,Y_j) \leq e_\Gamma(Y_i,Y_j) \leq e_\Gamma(Y_i',X_j) \leq (1+c)p|Y_i'||X_j| < 4cnp|X_j| \leq \frac{\epsilon}{2k}pn|X_j| \leq \epsilon p|X_i||X_j|$$

contradiction. Hence, $|Y_i| \geq cn$.

Suppose that Γ is $\left(p, \frac{\epsilon}{8k}\right)$-uniform. Assume for a moment the following inequality:

$$\left|d(X_i^j, X_j^{[i]}) - d(X_i,X_j)\right| \geq \frac{\epsilon p|X_i||X_j|}{2e_\Gamma(X_i^j, X_j^{[i]})} \tag{5}$$

We will prove this inequality later. Then:

$$
\begin{aligned}
q(\mathcal{P}') - q(\mathcal{P}) &\overset{\text{def.}}{=} \sum_{A' \in \tilde{\mathcal{P}}'} \sum_{B' \in \tilde{\mathcal{P}}'} q(A', B') - \sum_{A \in \tilde{\mathcal{P}}} \sum_{B \in \tilde{\mathcal{P}}} q(A, B) \\
&= \sum_{A \in \tilde{\mathcal{P}}} \sum_{B \in \tilde{\mathcal{P}}} q(\tilde{\mathcal{P}}'(A), \tilde{\mathcal{P}}'(B)) - \sum_{A \in \tilde{\mathcal{P}}} \sum_{B \in \tilde{\mathcal{P}}} q(A, B) \\
&\overset{\text{restriction, 3.12}}{\geq} \sum_{(i,j) \in S} \left(q(\tilde{\mathcal{P}}'(X_i), \tilde{\mathcal{P}}'(X_j)) - q(X_i, X_j) \right) \\
&\overset{3.12}{\geq} \sum_{(i,j) \in S} \left(q\left(\mathcal{P}_i(X_i^j), \mathcal{P}_j\left(X_j^{[i]} \right) \right) - q(X_i, X_j) \right) \\
&\overset{(*)}{\geq} \sum_{(i,j) \in S} \frac{\epsilon^2 p^2 |X_i|^2 |X_j|^2}{4n^2 e_\Gamma(X_i, X_j)} \\
&= \sum_{(i,j) \in S} \frac{\epsilon^2 p}{4} \frac{p|X_i||X_j|}{e_\Gamma(X_i, X_j)} \frac{|X_i||X_j|}{n^2} \\
&\overset{\Gamma\text{ unif.}}{\geq} \sum_{(i,j) \in S} \left(\frac{\epsilon^2 p}{4} \right) \left(\frac{1}{2} \right) \left(\frac{1}{4k^2} \right) \\
&\geq (\epsilon k^2) \left(\frac{\epsilon^2 p}{4} \right) \left(\frac{1}{2} \right) \left(\frac{1}{4k^2} \right) \\
&= \frac{\epsilon^3 p}{32}
\end{aligned}
$$

where inequality (*) is detailed here. For ease of notation, we will denote $X_i^j = Y_i$ and $X_j^{[i]} = Y_j$:

$$
\begin{aligned}
&q\left(\mathcal{P}_i(Y_i), \mathcal{P}_j\left(Y_j\right)\right) - q(X_i, X_j) \\
=&\sum_{A\in\mathcal{P}_i(Y_i)} \sum_{B\in\mathcal{P}_j\left(Y_j\right)} q(A,B) - q(X_i, X_j) \\
=&\sum_{A\in\mathcal{P}_i(Y_i)} \sum_{B\in\mathcal{P}_j\left(Y_j\right)} \frac{e_\Gamma(A,B)}{n^2} d^2(A,B) - \frac{e_\Gamma(X_i,X_j)}{n^2} d^2(X_i,X_j) \\
=&\sum_{A\in\mathcal{P}_i(Y_i)} \sum_{B\in\mathcal{P}_j\left(Y_j\right)} \frac{e_\Gamma(A,B)}{n^2}\left(d^2(A,B) - d^2(X_i,X_j)\right) \\
\overset{(3.9)}{=}&\sum_{A\in\mathcal{P}_i(Y_i)} \sum_{B\in\mathcal{P}_j\left(Y_j\right)} \frac{e_\Gamma(A,B)}{n^2}\left(d^2(A,B) - d^2(X_i,X_j) - 2d(A,B)d(X_i,X_j) + 2d^2(X_i,X_j)\right) \\
=&\sum_{A\in\mathcal{P}_i(Y_i)} \sum_{B\in\mathcal{P}_j\left(Y_j\right)} \frac{e_\Gamma(A,B)}{n^2}\left(d(A,B) - d(X_i,X_j)\right)^2 \\
\geq&\frac{e_\Gamma(Y_i,Y_j)}{n^2}\left(d(Y_i,Y_j) - d(X_i,X_j)\right)^2 \\
\geq&\frac{e_\Gamma(Y_i,Y_j)}{n^2}\left(\frac{\epsilon p|X||Y_i|}{2e_\Gamma(Y_i,Y_j)}\right)^2 \\
=&\frac{\epsilon^2 p^2 |X_i|^2 |X_j|^2}{4n^2 e_\Gamma(X_i,X_j)}
\end{aligned}
$$

Now we want to prove (5). Note that, from the uniformity of Γ, both $e_\Gamma(X_i, X_j)$ and $e_\Gamma(Y_i, Y_j)$ are nonzero. From the definition of Y_i and Y_j as sets that violate the $(q_{i,j}, p, \epsilon)$-DISC condition, we have

$$
\left|e_G(Y_i,Y_j) - \frac{e_G(X_i,X_j)}{|X_i||X_j|}|Y_i||Y_j|\right| \geq \epsilon p|X_i||X_j|
$$

$$
\left|\frac{e_G(Y_i,Y_j)}{e_\Gamma(Y_i,Y_j)} - \frac{e_G(X_i,X_j)}{e_\Gamma(Y_iY_j)}\frac{|Y_i||Y_j|}{|X_i||X_j|}\right| \geq \frac{\epsilon p|X_i||X_j|}{e_\Gamma(Y_i,Y_j)}
$$

$$
\left|\frac{e_G(Y_i,Y_j)}{e_\Gamma(Y_i,Y_j)} - \frac{e_G(X_i,X_j)}{e_\Gamma(X_i,X_j)}\frac{e_\Gamma(X_i,X_j)}{|X_i||X_j|}\frac{|Y_i||Y_j|}{e_\Gamma(Y_iY_j)}\right| \geq \frac{\epsilon p|X_i||X_j|}{e_\Gamma(Y_i,Y_j)}
$$

$$
\left|d(Y_i,Y_j) - d(X_i,X_j)\frac{r(X_i,X_j)}{r(Y_i,Y_j)}\right| \geq \frac{\epsilon p|X_i||X_j|}{e_\Gamma(Y_i,Y_j)} \tag{6}
$$

Now remember that by the (p,c)-uniform condition on Γ, we have that $|r(A,B) - p| \leq cp$ for $|A|,|B| \geq cn$. Since $|X_i|,|X_j|,|Y_i|,|Y_j| \geq cn$, we have that

$$|r(Y_i,Y_j)-r(X_i,X_j)| \leq |r(Y_i,Y_j)-p|+|p-r(X_i,X_j)| \leq 2cp \leq \frac{\epsilon p}{2} \leq \frac{\epsilon p|X_i||X_j|}{2|Y_i||Y_j|}$$

From this we can find

$$\left|\frac{r(Y_i,Y_j)-r(X_i,X_j)}{r(Y_i,Y_j)}\right| \leq \frac{\epsilon p|X_i||X_j|}{2|Y_i||Y_j|r(Y_i,Y_j)}$$

$$\left|d(X_i,X_j)\left(1-\frac{r(X_i,X_j)}{r(Y_i,Y_j)}\right)\right| \overset{(d\leq 1)}{\leq} \frac{\epsilon p|X_i||X_j|}{2e_\Gamma(Y_i,Y_j)}$$

$$\left|d(X_i,X_j)-d(X_i,X_j)\frac{r(X_i,X_j)}{r(Y_i,Y_j)}\right| \leq \frac{\epsilon p|X_i||X_j|}{2e_\Gamma(Y_i,Y_j)} \tag{7}$$

Finally, by the triangle inequality,

$$\begin{aligned} |d(Y_i,Y_j)-d(X_i,X_j)| &\geq \left|d(Y_i,Y_j)-d(X_i,X_j)\frac{r(X_i,X_j)}{r(Y_i,Y_j)}\right| - \left|d(X_i,X_j)-d(X_i,X_j)\frac{r(X_i,X_j)}{r(Y_i,Y_j)}\right| \\ &\overset{(6),(7)}{\geq} \frac{\epsilon p|X_i||X_j|}{e_\Gamma(Y_i,Y_j)} - \frac{\epsilon p|X_i||X_j|}{2e_\Gamma(Y_i,Y_j)} \\ &= \frac{\epsilon p|X_i||X_j|}{2e_\Gamma(Y_i,Y_j)} \end{aligned}$$

□

Finally, for the last step, we take a look at Lemma 2.10:

Lemma 2.10. Let $\mathcal{P}=\{X_i\}_{i=0}^k$ be a (not necessarily equitable) partition of V with exceptional set X_0, and let $\delta>0$. Then there exists an equitable partition $\mathcal{P}'=\{X_i'\}_{i=0}^{k'}$ with exceptional set X_0' which refines $\mathcal{P}$, with $k'\leq\delta^{-1}k$ and $|X_0'|\leq|X_0|+\delta|V|$.

This lemma only involves sets of vertices, and does not take in consideration the edges in between them (in fact, the lemma does not mention any graph at all). For this reason we can use the same lemma as in the dense case, without taking any special considerations.

With all of this we can prove the sparse version of the regularity lemma, using the same reasoning as in the dense case:

Proof of lemma 3.8. Let $\delta=64\epsilon^{-3}$, and suppose that $c<\frac{\epsilon}{8M}$, where we will define $M=M(\epsilon,m)$ later. Start with any partition $\mathcal{P}_0$ in m parts and without exceptional set. If a partition $\mathcal{P}$ into at most m parts is given, take $\mathcal{P}_0$ into exactly m parts such that it refines $\mathcal{P}$. Once we have that, do the following until we can not continue:

- Assume that $\epsilon\leq\frac{1}{2}$, as otherwise any $\frac{1}{2}$-DISC partition is ϵ-DISC (The ϵ-DISC condition is more restrictive for smaller values of ϵ). If $\mathcal{P}_i$ has k_i non-exceptional sets, then construct an equitable partition $\mathcal{Q}_i$ with at most $(\delta+1)\epsilon^{-1}k_i$ non-exceptional parts in which the exceptional set increases by at most $(\delta+1)^{-1}\epsilon|V|$ vertices. The existence of such a partition is guaranteed by Lemma 2.10, setting $\delta' = (\delta+1)^{-1}\epsilon$.

- If $\mathcal{Q}_i$ is equitable, has k'_i non-exceptional sets and its exceptional set has size at most $\epsilon|V|$, but it is not ϵ-DISC, then construct $\mathcal{P}_{i+1}$ such that it has at most $k'_i 4^{k'_i}$ non-exceptional parts, has the same exceptional set as $\mathcal{Q}_i$, refines $\mathcal{Q}_i$ and $q(\mathcal{P}_{i+1}) \geq q(\mathcal{Q}_i) + \frac{2p}{\delta}$. The existence of such a partition is guaranteed by Lemma 3.12 if $k'_i \leq M$.

We claim that the procedure produces a partition $\mathcal{Q}_i$ that is ϵ-DISC for some $0 \leq i \leq \lfloor\delta\rfloor$. Assume the opposite, and we will reach a contradiction. First we will show that, if $\mathcal{Q}_i$ is not ϵ-DISC for any of those values of i, then $\mathcal{Q}_i$ exists for $1 \leq i \leq \lfloor\delta\rfloor + 1$. If $\mathcal{Q}_i$ exists but $\mathcal{Q}_{i+1}$ does not, it is because $\mathcal{Q}_i$ is not equitable, or its exceptional set is bigger than $\epsilon|V|$, or its number of parts exceeds M. But $\mathcal{Q}_i$ is equitable by construction, so the first option is impossible.

Let $f(x) = (\delta+1)\epsilon^{-1}x4^x$. The numbers of parts k_i and k'_i satify $k'_i \leq (\delta+1)\epsilon^{-1}k_i \leq (\delta+1)\epsilon^{-1}k'_{i-1}4^{k'_{i-1}} = f(k'_{i-1})$. Since $k'_0 \leq ((\delta+1)\epsilon^{-1}m)$, then setting $M = f(f(...f((\delta+1)\epsilon^{-1}m)...))$, where f appears $\lfloor\delta\rfloor$ times guarantees that $k'_i \leq M$ for $0 \leq i \leq \lfloor\delta\rfloor$.

$\mathcal{Q}_i$ and $\mathcal{P}_{i+1}$ have the same exceptional set. By construction of $\mathcal{Q}_i$, if V_0^i is the exceptional set of $\mathcal{Q}_i$, then $|V_0^{i+1}| \geq |V_0^i| + (\delta+1)^{-1}\epsilon|V|$. By induction, this means that $|V_0^i| \leq (i+1)(\delta+1)^{-1}\epsilon|V|$. If $i \leq \lfloor\delta\rfloor$, then the size of the exceptional set of $\mathcal{Q}_i$ for $i \leq \lfloor\delta\rfloor$ is at most $(\lfloor\delta\rfloor+1)(\delta+1)^{-1}\epsilon|V| < \epsilon|V|$. This means that $\mathcal{Q}_i$ exists for $0 \leq i \leq \lfloor\delta\rfloor + 1$.

By Lemma 3.11 and Lemma 3.12, $q(\mathcal{Q}_i) \geq q(\mathcal{P}_i) \geq q(\mathcal{Q}_{i-1}) + \frac{2p}{\delta}$. Recall that if $c < 1$, then $q(\mathcal{P})$ is between 0 and $2p$. Since $q(\mathcal{Q}_0) \geq 0$ this means by induction that $q(\mathcal{Q}_i) \geq \frac{2pi}{\delta}$, and $q(\mathcal{Q}_{\lfloor\delta\rfloor+1}) \geq \frac{2p(\lfloor\delta\rfloor+1)}{\delta} > 2p$, contradiction. Hence $\mathcal{Q}_i$ must be ϵ-DISC for some $0 \leq i \leq \lfloor\delta\rfloor$, and the number of parts is at most M. Also, this partition refines $\mathcal{P}_0$, so it refines $\mathcal{P}$ too. □

Using this lemma and Lemma 3.4, we obtain the following version of the regularity lemma for sparse graphs:

Theorem 3.13 (Regularity lemma for jumbled graphs). *For every $\epsilon > 0$ and every positive integer m there exist a constant $c(\epsilon, m) > 0$ and a positive integer $M(\epsilon, m)$ such that, if G is a graph with at least m vertices which is a subgraph of graph Γ, and if Γ is a (p, cpn)-jumbled graph, then G admits an ϵ-DISC partition in k non-exceptional parts, with $m \leq k \leq M$. Moreover, if $\mathcal{P}$ is a fixed partition of $V(G)$ with at most m parts, then such an ϵ-DISC partition can be found in a way that each set X_i with $1 \leq i \leq k$ is contained in one of the sets of $\mathcal{P}$.*

Proof. By Lemma 3.8, there are constants M and $\delta > 0$ such that, if Γ is δ-uniform, then the result holds (here δ is the value of c returned by Lemma 3.8). Also, by Lemma 3.4, there is $c > 0$ such that, if Γ is (p, cpn)-jumbled then it is δ-uniform. The corollary follows trivially from these two results. □

The regularity that we required for this result is $\beta = cpn$, with p having exponent 1. We want to prove the removal lemma for an exponent as small as possible. The exponents in the statement of

Theorem 3.34 are all greater than 1, so the version of the regularity lemma that we proved is better than the one we need. This is because the most restrictive step in the proof is the counting lemma.

There is something to note about the proof of the regularity lemma. If we start with a partition $\mathcal{P}$, we do not look at whether X and Y satisfy the jumbledness condition when either X or Y is not contained in a pair of parts. In fact, the lemma still holds when we only impose jumbledness between parts of $\mathcal{P}$. For that reason, we can consider the following variation of the regularity lemma:

Lemma 3.14. *For every $\epsilon > 0$ and every positive integer m there exist a constant $c > 0$ and a positive integer M such that the following holds: let G be a graph with n vertices, $n \geq m$, which is a subgraph of graph Γ, and $\mathcal{P}$ be a partition of $V(\Gamma)$ into at most m parts, such that $(X_i, X_j)_\Gamma$ is a $(p_{i,j}, cp_{i,j}n)$-jumbled graph for any two parts X_i and X_j of $\mathcal{P}$. Then G admits a partition $\mathcal{P}'$ with exceptional set X_0' of size at most ϵn and k non-exceptional sets, with $m \leq k \leq M$, in which, for all but at most ϵk^2 pairs of parts X_a' and X_b', contained in X_i and X_j, respectively, the graph $(X_i', X_j')_G$ is $(q_{a,b}, p_{i,j}, \epsilon)$-DISC for some values of $q_{a,b}$.*

Sketch of the proof: If $p_{i,j} = 0$ for some X_i and X_j, the graph $(X_i, X_j)_\Gamma$ must be $(0,0)$-jumbled, which means that it is empty, and any two subsets of vertices between them form a $(0,0,\epsilon)$-DISC graph. So assume that $p_{i,j} \neq 0$. Then the proof is similar to the proof that we saw, except now the function $q(X_a', X_b')$ is replaced so that, if $X_a' \subseteq X_i$ and $X_b' \subseteq X_j$, then

$$q(X_a', X_b') = \frac{e_\Gamma(X_a', X_b')}{p_{i,j}n^2} d^2(X_a', X_b')$$

Now $q(\mathcal{P})$ is bounded between 0 and 2, and every time that we refine the parts not satisfying DISC q increases by at least $\frac{\epsilon^3}{32}$, so the number of steps is bounded. Conclude the proof as in the proof of Lemma 3.8.

The jumbledness condition is, in fact, not required at all in a regularity lemma like this. Scott proved a version of this lemma with no jumbledness condition [**Sco**]. While this result is a generalization of Lemma 3.8 and Theorem 3.14, it is not much more useful in practice, because most of the applications of the removal lemma have jumbledness as a hypothesis. If we do not impose the jumbledness condition, then one could have, for example, that all edges of G lie between pairs not satisfying $DISC$ (because all but at most ϵk^2 pairs of parts form empty graphs), so we cannot apply properties of $DISC$ in this case.

3.4. The counting lemma

Up until now, the proof of the sparse case has been simply an adaptation of the proof of the dense removal lemma. This has required some extra calculations, but the ideas were analogous as in the case where the graphs are dense. This will not be the case for the counting lemma. Remember that we proved the dense counting lemma by embedding the vertices one by one into different sets. Now we will not be able to do the same, and we will require other techniques.

The version of the counting lemma used in the proof of the sparse removal lemma from **[ConFox1]** is the following, where $d_2(H)$ denotes the 2-degeneracy[1] of H:

Lemma 3.15. *Let H be a graph with vertex set $V = \{v_1, ..., v_h\}$, and let $\alpha, \theta > 0$ be positive constants. Then there exist $c(H, \alpha, \delta) > 0$ and $\epsilon(H, \alpha, \delta) > 0$ such that the following holds: Let Γ be graph with vertex sets $X_1, X_2, ..., X_h$ such that the bipartite graph $(X_i, X_j)_\Gamma$ is $(p, \gamma = cp^{d_2(H)+3})$-jumbled for any $i < j$ such that $v_iv_j \in E(H)$. Let G be a subgraph of Γ such that $(X_i, X_j)_G$ is $(q_{i,j}, p, \epsilon)$-DISC with $\alpha p \leq q_{i,j} \leq p$ for any $i < j$ such that $v_iv_j \in E(H)$. Then*

$$||H \to G||_X \geq (1-\theta) \left(\prod_{\substack{i<j \\ v_iv_j \in E(H)}} q_{i,j} \right) \left(\prod_{i=1}^{h} |X_i| \right) \tag{8}$$

The exponent of p in γ is $d_2(H) + 3$. For cycles, the values of the 2-degeneracy are $d_2(C_3) = 1$ and $d_2(C_\ell) = \frac{1}{2}$ for $\ell \geq 4$. Remember that we will prove this result for $\gamma = cp^{t_\ell}$, with

$$t_\ell = \begin{cases} 3 & \text{if } \ell = 3 \\ 2 & \text{if } \ell = 4 \\ 1 + \frac{1}{\ell-3} & \text{if } \ell \geq 5 \text{ is odd} \\ 1 + \frac{1}{\ell-4} & \text{if } \ell \geq 6 \text{ is even} \end{cases}$$

This means that Lemma 3.20 is not enough for our purposes, because it requires a graph Γ more jumbled than the one in the hypotheses of our removal lemma. We will need a lemma that only focuses on cycles, but which requires less jumbledness.

Let us look again at equation (8). If G was made of bipartite random graphs, each of them with edge probability $q_{i,j}$, then the number of choices of h vertices, one from each subset X_i, would be $\prod_{i=1}^{h} |X_i|$. The probability of those vertices producing a copy of H in G is the product of the probabilities of each edge appearing in G, which is $\prod_{i<j, v_iv_j \in E(H)} q_{i,j}$, so equation (8) tells us that the number of embeddings of H in G, with the corresponding subset restrictions, is at the expected number of embeddings in a random graph, up to a factor of $1 - \theta$ for θ arbitrarily small.

In this section we will prove the case $H = C_\ell$ for $\ell \geq 5$. The techniques used in the proof for all other graphs, including $H = C_3$ and C_4, will be discussed, but we will not go into deep levels of detail on those.

The first concept that we need to introduce for the proof of the sparse counting lemma is that of weighted graphs. We usually consider that a graph G consists of a vertex set V and an edge set

[1]The 2-degeneracy of a graph H, denoted by $d_2(H)$, is the smallest value d for which there exists an ordering $v_1, v_2, ..., v_h$ of the vertices of H satisfying the following property: denote by $N_i(j)$ the number of t smaller than or equal to i such that $v_tv_j \in E(H)$. Then, for any $i < j$ such that $v_iv_j \in E(H)$, we have $N_{i-1}(i) + N_{i-1}(j) \leq 2d$. Note that d is not necesarily an integer, and it can be half an integer.

$E \subseteq \binom{V}{2}$. Another way of considering the same graph is as a function $f : V^2 \to \{0,1\}$ which satisfies $f(v,v) = 0$ and $f(u,v) = f(v,u)$ for all $u,v \in V$. In this case, we say that uv is an edge of the graph iff $f(u,v) = 1$.

A weighted graph is also a fuction, except the image of each pair of vertices does not need to be 0 or 1, but can be anywhere in the closed interval $[0,1]$. This gives an edge intermediate states between 'being in the graph' and 'not being in the graph':

Definition 3.16 (Weighted graph). A weighted graph G consists of a set of vertices V and a function $f : V^2 \to [0,1]$ satisfying that $f(v,v) = 0$ and $f(u,v) = f(v,u)$ for any $u,v \in V$.

A weighted bipartite graph consists of two vertex sets X and Y, and a function $f : X \times Y \to [0,1]$.

If G is a weighted graph, we will denote by $G(u,v)$ the value $f(u,v)$, where f is the function corresponding to the weighted edges of G.

This idea is related to random graphs: in random graphs, there are intermediate states between 'always appearing' and 'never appearing', which is appearing with a probability p. In weighted graphs, the intermediate states correspond to the edge having a certain weight $G(u,v)$, which can be regarded as analogous to p.

We will need to redefine concepts like number of embeddings, jumbledness, discrepancy, etc. The new definitions must be coherent with the old ones when the weight of every edge is 0 or 1. This is because we need to apply the results that we obtain to the case we want, which is unweighted graphs.

Let us start with $e_G(X,Y)$. The number of edges between X and Y was

$$e_G(X,Y) = |\{(x,y) : x \in X, y \in Y, (x,y) \in E(G)\}| = \sum_{x \in X} \sum_{y \in Y} \mathbb{1}_{E(G)}(x,y)$$

From this, a natural way of defining $e_G(X,Y)$ would be

$$e_G(X,Y) = \sum_{x \in X} \sum_{y \in Y} G(x,y)$$

which is the sum of the contributions of all possible edges between X and Y.

The idea for $||H \to G||_X$ is similar. We see that, in unweighted graphs,

$$\begin{aligned} ||H \to G||_X &= \left|\{(x_1,...,x_h) \in X_1 \times \cdots \times X_h : x_i x_j \in E(G) \;\; \forall i < j : v_i v_j \in E(H)\}\right| \\ &= \sum_{(x_1,...,x_h) \in X_1 \times \cdots \times X_h} \prod_{\substack{i<j \\ v_i v_j \in E(H)}} \mathbb{1}_{e(G)}(x_i, x_j) \end{aligned}$$

so a coherent definition would be

$$||H \to G||_X = \sum_{(x_1,...,x_h) \in X_1 \times \cdots \times X_h} \prod_{\substack{i<j \\ v_i v_j \in E(H)}} G(x_i, x_j)$$

If G was a random graph in which each edge appeared independently with probability $G(u,v)$, then this corresponds to the expected number of copies of H in G.

In the case where H is a cycle of length ℓ, the definition becomes

$$||C_\ell \to G||_X = \sum_{(x_1,\dots,x_\ell)\in X_1\times\cdots\times X_\ell} \prod_{i=1}^{\ell} G(x_i, x_{i+1})$$

where we are using cyclic notation, that is, we are considering $x_{\ell+1} = x_1$.

Let G and Γ be weighted graphs on the same set of vertices. We say that G is a subgraph of Γ if $G(u,v) \leq \Gamma(u,v)$ for each $u, v \in V(G)$.

Before finding the weighted equivalent of the pseudorandomness properties, we shall define some notation for sums, which will simplify the equations appearing in the proof. This notation is taken from [**ConFoxZha**]:

Definition 3.17. Let G be a weighted graph and S be a set of vertices of G. Let $f : S \to \mathbb{R}$ be a function. Then we define

$$\int_S f(s)ds = \frac{1}{|S|}\sum_{s\in S} f(s)$$

When it does not lead to confusion, we may omit the differential ds from the integral, or write simply f instead of $f(s)$.

The meaning of the integral sign is taking the average over the possible values of the variables. This integral sign satisfies several properties, some of which are in common with those of Riemann integrals:

Property 3.18.

- *If c is a constant, then*

$$\int_S c = c$$

- *If $f, g : S \to \mathbb{R}$ are two functions such that $f \geq g$, then*

$$\int_S f \geq \int_S g$$

In particular, if f is a non-negative function, then

$$\int_S f \geq \int_S 0 = 0$$

- *If $f, g : S \to \mathbb{R}$ are two functions, then*

$$\int_S (f+g) = \left(\int_S f\right) + \left(\int_S g\right)$$

- *If $S' \subseteq S$, then*

$$\int_S f(s)\mathbb{1}_{S'}(s) = \frac{|S'|}{|S|}\int_{S'} f(s) \tag{9}$$

In particular,

$$\int_S \mathbb{1}_{S'}(s) = \frac{|S'|}{|S|}$$

- *(Cauchy-Schwarz inequality) If* $f, g : S \to \mathbb{R}$ *are two functions, then*

$$\left(\int_S f(s)g(s)\right)^2 \le \left(\int_S f^2(s)\right)\left(\int_S g^2(s)\right)$$

Proof. Properties 1 through 3 are trivial from the expression of the integral sign as a sum. Let us prove the last two ones. First,

$$\int_S f(s)\mathbb{1}_{S'}(s) = \frac{1}{|S|}\sum_{s\in S} f(s)\mathbb{1}_{S'}(s) = \frac{1}{|S|}\sum_{s\in S'} f(s) = \frac{|S'|}{|S|}\int_{S'} f(s)$$

The special case is obtained by setting $f(s) = 1$. Finally, the last one is obtained by applying the Cauchy-Schwarz inequaliy for sums:

$$\left(\int_S f(s)g(s)\right)^2 = \left(\frac{1}{|S|}\sum_{s\in S} f(s)g(s)\right)^2 \le \frac{1}{|S|^2}\left(\sum_{s\in S} f^2(s)\right)\left(\sum_{s\in S} g^2(s)\right)$$

□

We can use this notation to rewrite the definitions of $e_G(X,Y)$ and $||C_\ell \to G||_X$:

$$e_G(X,Y) = |X||Y|\int_X\int_Y G(x,y)$$

$$||C_\ell \to G||_X = \prod_{i=1}^{\ell}|X_i|\int_{X_1}\cdots\int_{X_\ell} G(x_1,x_2)G(x_2,x_3)\cdots G(x_\ell,x_1)$$

The condition in Lemma 3.20 (equation (8)) can thus be rewriten as

$$\int_{X_1}\cdots\int_{X_\ell} G(x_1,x_2)G(x_2,x_3)\cdots G(x_\ell,x_1) \ge (1-\theta)\prod_{i=1}^{\ell} q_i$$

where $q_i = q_{i,i+1}$.

We will now find the weighted conditions for jumbledness and DISC. A bipartite graph on vertex sets X and Y is $(p, \beta = \gamma\sqrt{|X||Y|})$-jumbled if:

$$\left|e(X',Y') - p|X'||Y'|\right| \le \beta\sqrt{|X'||Y'|} \qquad \forall X' \subseteq X, Y' \subseteq Y$$

$$\Updownarrow$$

$$\left|\frac{e(X',Y')}{|X||Y|} - p\frac{|X'|}{|X|}\frac{|Y'|}{|Y|}\right| \le \gamma\sqrt{\frac{|X'|}{|X|}\frac{|Y'|}{|Y|}} \qquad \forall X' \subseteq X, Y' \subseteq Y$$

$$\Updownarrow$$

$$\left| \int_X \int_Y G(x,y) \mathbb{1}_{X'}(x) \mathbb{1}_{Y'}(y) - \int_X \int_Y p \mathbb{1}_{X'}(x) \mathbb{1}_{Y'}(y) \right| \leq \gamma \sqrt{\int_X \mathbb{1}_{X'}(x) \int_Y \mathbb{1}_{Y'}(y)} \qquad \forall X' \subseteq X, Y' \subseteq Y$$

$$\Updownarrow$$

$$\left| \int_X \int_Y (G(x,y) - p) f(x) g(y) \right| \leq \gamma \sqrt{\int_X f(x) \int_Y g(y)} \qquad \forall f : X \to \{0,1\}, g : Y \to \{0,1\}$$

We would like to know whether, at least in the case of unweighted graphs, we can extend this property to functions to the interval $[0,1]$. That is to say, whether every jumbled graph will satisfy this property for any $f : X \to [0,1]$ and $g : Y \to [0,1]$. We claim that this is indeed the case:

Lemma 3.19. *Let G be an unweighted bipartite graph on vertex sets X and Y. Then G is $(p, \beta = \gamma\sqrt{|X||Y|})$-jumbled if and only if, for any pair of functions $f : X \to [0,1]$ and $g : Y \to [0,1]$ we have*

$$\left| \int_X \int_Y (G(x,y) - p) f(x) g(y) \right| \leq \gamma \sqrt{\int_X f(x) \int_Y g(y)} \tag{10}$$

Proof. If G satisfies (10) for all funtions $f : X \to [0,1]$ and $g : Y \to [0,1]$, in particular it satisfies (10) for the functions $f : X \to \{0,1\}$ and $g : Y \to \{0,1\}$, which means that G is (p,β)-jumbled.

Assume that G does not satisfy (10) for some $f : X \to [0,1]$ and $g : Y \to [0,1]$, we claim that there are functions $f' : X \to \{0,1\}$ and $g' : Y \to \{0,1\}$ such that (10) does not hold. We proceed by induction on the number of elements from X and Y such that $f(x)$ or $g(y)$ are strictly between 0 and 1. If there are no such elements, then the functions f and g have image in $\{0,1\}$, and we can make $f = f'$ and $g = g'$.

Now, without loss of generality, assume that there is $\tilde{x} \in X$ such that $f(\tilde{x}) \in (0,1)$. We define $f_a(x)$ as $f_a(\tilde{x}) = a$ and $f_a(x) = f(x)$ for $x \neq \tilde{x}$. Consider the following inequality:

$$\left| \int_X \int_Y (G(x,y) - p) f_a(x) g(y) \right| \leq \gamma \sqrt{\int_X f_a(x) \int_Y g(y)} \tag{11}$$

By our hypothesis, inequality (11) does not hold for $a = f(\tilde{x})$. The term inside the absolute value of the left hand side is linear in a, so the LHS is convex in a. On the other hand, the term inside the square root of the RHS is linear in a, so the RHS is concave in a.

Therefore, if the inequality held for $a = 0$ and $a = 1$, then it would also hold for all intermediate values, including $a = f(\tilde{x})$. Since we assumed that this is not the case, the inequality is not satisfied for f_0 or f_1. This means that the pair (f_0, g) or (f_1, g) of functions does not satisfy (10) either, and the number of values different than 0 or 1 has decreased by one. We conclude that the existence of f and g implies that f' and g' also exist, and therefore any (p,β)-jumbled graph satisfies (10) for any $f : X \to [0,1]$ and $g : Y \to [0,1]$. □

For weighted graphs, the entire proof of Lemma 3.19 is still valid. That is to say, the condition $\left|\int_X\int_Y (G(x,y)-p)f(x)g(y)\right| \leq \gamma\sqrt{\int_X f(x)\int_Y g(y)}$ holds for all f and g if and only if the condition $|e(X',Y')-p|X'||Y'|| \leq \beta\sqrt{|X'||Y'|}$ holds for any subsets X' and Y'.

The same happens for (q,p,ϵ)-DISC graphs. The equivalent weighted condition is

$$\left|\int_X\int_Y (G(x,y)-p)f(x)g(y)\right| \leq \epsilon p \qquad \forall f: X\to[0,1], g: Y\to[0,1]$$

In the case of one-sided discrepancy ((q,p,ϵ)-DISC$_\geq$), the formula becomes

$$\int_X\int_Y (G(x,y)-p)f(x)g(y) \geq -\epsilon p \qquad \forall f: X\to[0,1], g: Y\to[0,1] \tag{12}$$

The proof in both cases is similar to the proof for jumbledness (Lemma 3.19).

We are ready to begin with the proof of the counting lemma. We will now state the version that we will prove:

Lemma 3.20 (Sparse counting lemma for cycles). *Let $\ell \geq 5$ be an integer, and let $\alpha,\theta > 0$ be positive constants. Then there exist $c>0$ and $\epsilon>0$ such that the following holds: Let Γ be graph with vertex sets $X_1, X_2, ..., X_\ell$ such that the bipartite graph $(X_i, X_{i+1})_\Gamma$ is $(p,\gamma = cp^{t_\ell})$-jumbled for all $1\leq i\leq \ell$. Let G be a subgraph of Γ such that $(X_i,X_{i+1})_G$ is (q_i,p,ϵ)-DISC$_\geq$ with $\alpha p \leq q_i \leq p$ for all $1\leq i\leq\ell$. Then*

$$\int_{X_1}\cdots\int_{X_\ell} G(x_1,x_2)G(x_2,x_3)\cdots G(x_\ell,x_1) \geq (1-\theta)\underbrace{\prod_{i=1}^{\ell} q_i}_{=q} \tag{13}$$

The proof of this lemma is taken from **[ConFoxZha]** and will use a key lemma, stated here as Lemma 3.22. For this key lemma we will use the following notation:

Definition 3.21. Let G be a graph, and let $X_1, X_2, ..., X_k$ be subsets of vertices of G. Fix some $x_1\in X_1$ and $x_k \in X_k$. We denote

$$G(x_1,X_2,...,X_{k-1},x_k) = \int_{X_2}\cdots\int_{X_{k-1}} G(x_1,x_2)G(x_2,x_3)\cdots G(x_{k-1},x_k)dx_{k-1}\cdots dx_2$$

$$G(x_1,X_2,...,X_{k-1},X_k) = \int_{X_2}\cdots\int_{X_k} G(x_1,x_2)G(x_2,x_3)\cdots G(x_{k-1},x_k)dx_k\cdots dx_2$$

$$G(X_1,X_2,...,X_{k-1},X_k) = \int_{X_1}\cdots\int_{X_k} G(x_1,x_2)G(x_2,x_3)\cdots G(x_{k-1},x_k)dx_k\cdots dx_1$$

Lemma 3.22 (Key lemma). *For any $\mu>0$ and $m\geq 2$ there are $c>0$ and $\epsilon>0$ such that the following holds: let Γ be a weighted graph and G be a weighted subgraph of Γ. Let X_0, X_1, ..., X_m be vertex sets, such that, for all $1\leq i\leq\ell$, $(X_{i-1},X_i)_\Gamma$ is $(p,\gamma = cp^{1+\frac{1}{2m-2}})$-jumbled and $(X_{i-1},X_i)_G$ is*

(q_i, p, ϵ)-DISC$_{\geq}$. Let $\tilde{G}(x_0, x_m) = G(x_0, X_1, ..., X_{m-1}, x_m)$ and $G' = \min\{\tilde{G}, 4p^m\}$. Then G' satisfies $(q_1 q_2 ... q_m, p^m, \mu)$-DISC$_{\geq}$

Intuitively, the meaning of this statement is the following: if we have several bipartite graphs forming a path, then we can make an average of them ($\tilde{G}$) and then bound the value of each edge (G'). If in the original G all the bipartite graphs satisfy DISC$_{\geq}$, and those bipartite graphs are subgraphs of jumbled graphs Γ, then after the average-and-bound process the graph still satisfies DISC$_{\geq}$, for some appropriate parameters.

We will use the lemma in the following way: for any $3 \leq a \leq \ell - 2$, we apply the lemma to construct $(X_1, X_a)_{G'}$ and $(X_a, X_\ell)_{G'}$. Note that $3 \leq \ell - 2$ implies $\ell \geq 5$, and for this reason we restrict ourselves to graphs of length at least 5. Then

$$\begin{aligned}\int_{X_1}\int_{X_a}\int_{X_\ell} G'(x_1, x_a)G'(x_a, x_\ell)G(x_\ell, x_1) &\leq \int_{X_1}\int_{X_a}\int_{X_\ell} \tilde{G}(x_1, x_a)\tilde{G}(x_a, x_\ell)G(x_\ell, x_1) \\ &= \int_{X_1}\int_{X_a}\int_{X_\ell} G(x_1, X_2, ..., X_{a-1}, x_a)G(x_a, X_{a+1}, ..., X_{\ell-1}, x_\ell)G(x_\ell, x_1) \\ &= \int_{X_1} ... \int_{X_\ell} G(x_1, x_2)G(x_2, x_3)...G(x_\ell, x_1)\end{aligned}$$

On the other hand, we can use that $G'(x_1, x_a) \leq 4p^{a-1}$ and $G'(x_a x_\ell) \leq 4p^{\ell-a}$ to define weights for the discrepancy condition on $(X_\ell, X_1)_G$ and obtain

$$\begin{aligned}\int_{X_1}\int_{X_a}\int_{X_\ell} G'(x_1, x_a)G'(x_a, x_\ell)(G(x_\ell, x_1) - q_\ell) =& 16p^{\ell-1}\int_{X_a}\left(\int_{X_1}\int_{X_\ell} \underbrace{\frac{G'(x_1, x_a)}{4p^{a-1}}}_{f(x_1)\in[0,1]}\underbrace{\frac{G'(x_a, x_\ell)}{4p^{\ell-a}}}_{g(x_\ell)\in[0,1]}(G(x_\ell, x_1) - q_\ell)dx_\ell dx_1\right)dx_a \\ \geq& 16p^{\ell-1}(-\epsilon p) \\ =& -16\epsilon p^\ell\end{aligned}$$

Combining the two inequalities, then the left hand side of (13) is greater than or equal to the integral of $G'(x_1, x_a)G'(x_a, x_\ell)q_\ell$ minus $16\epsilon p^\ell$. Notice that $16\epsilon p^\ell \leq 16\frac{\epsilon}{\alpha^\ell} q_1 q_2 \cdots q_\ell = 16\frac{\epsilon}{\alpha^\ell} q$, and the term $16\frac{\epsilon}{\alpha^\ell}$ goes to 0 as ϵ goes to 0. We can bound the integral of $G'(x_1, x_a)G'(x_a, x_\ell)q_\ell$ in a similar way, using DISC$_{\geq}$ twice more.

We can define $\tilde{\Gamma}$ and Γ' for Γ in the same way as $\tilde{G}$ and G' for G. The proof of this key lemma will consist of three steps:

- Show that $\tilde{G}$ satisfies DISC$_{\geq}$ (for some parameters).
- Show that, if the number of neighbours of every $x_i \in X_i$ in X_{i+1} is roughly the same, then capping off the edges (going from $\tilde{G}$ to G') does not have a big effect on discrepancy.
- Show that, under the hypotheses of the key lemma, the graph G has a big subgraph which satisfies the similar neighbourhoods condition.

We begin by showing the first step, which consists of Lemmas 3.23 and 3.24. These two will focus on the process of averaging.

The first lemma says that, by taking the average of two bipartite graphs, if each of them satisfies DISC$_{\geq}$, then the average also satisfies DISC$_{\geq}$, and the parameters q and p are the product of the equivalent parameters in each bipartite graph. The proof uses a 'few bad vertices' argument: we show that there are few vertices for which a certain value is far from the average, and show that the ones close to the average are enough that G' satisfies DISC$_{\geq}$ regardless of the behaviour of those bad vertices. We already used this argument in the proof of Lemma 2.13, and we will use it often in the proof of the key lemma.

Lemma 3.23. *Let G be a weighted graph on vertex sets X, Y and Z. Let $p_1, p_2, \epsilon \in (0,1]$ and $q_1 \in (0, p_1]$, $q_2 \in (0, p_2]$. If $(X,Y)_G$ satisfies (q_1, p_1, ϵ)-$DISC_{\geq}$ and $(Y,Z)_G$ satisfies (q_2, p_2, ϵ)-$DISC_{\geq}$, then the graph $\tilde{G}(x,z) = G(x,Y,z)$ satisfies $(q_1q_2, p_1p_2, 6\sqrt{\epsilon})$-$DISC_{\geq}$.*

Proof. Let $f : X \to [0,1]$ and $g : Z \to [0,1]$ be any two functions. Let

$$Y' = \left\{ y \in Y : \int_X (G(x,y) - q)f(x) \leq -\sqrt{\epsilon}p_1 \right\}$$

Then, applying the DISC$_{\geq}$ condition on $(X,Y)_G$ with weight functions f and $\mathbb{1}_{Y'}$ we obtain

$$\begin{aligned}
-\epsilon p_1 &\overset{DISC_{\geq}}{\leq} \int_X \int_Y (G(x,y) - q_1)f(x)\mathbb{1}_{Y'}(y) \\
&= \int_Y \left(\int_X (G(x,y) - q_1)f(x) \right) \mathbb{1}_{Y'}(y) \\
&\leq \int_Y -\sqrt{\epsilon}p_1\mathbb{1}_{Y'}(y) \\
&= -\sqrt{\epsilon}p_1\frac{|Y'|}{|Y|}
\end{aligned}$$

This means that $|Y'| \leq \sqrt{\epsilon}|Y|$. Similarly, we define Y'' as

$$Y'' = \left\{ y \in Y : \int_Z (G(y,z) - q_2)g(z) \leq -\sqrt{\epsilon}p_2 \right\}$$

which satisfies $|Y''| \leq \sqrt{\epsilon}|Y|$ for the same reason. The conslusion is that $|Y \setminus (Y' \cup Y'')| \geq (1 - 2\sqrt{\epsilon})|Y|$. We apply these to bound the integral:

$$
\begin{aligned}
\int_X \int_Z \tilde{G}(x,z)f(x)g(z)dzdx =& \int_X \int_Z \int_Y G(x,y)G(y,z)f(x)g(z)dydzdx \\
=& \int_Y \left(\int_X G(x,y)f(x)dx \right) \left(\int_Z G(y,z)g(z)dz \right) dy \\
\overset{G,f,g\geq 0}{\geq} & \int_Y \left(\int_X G(x,y)f(x)dx \right) \left(\int_Z G(y,z)g(z)dz \right) \mathbb{1}_{Y'\setminus(Y'\cup Y'')}(y)dy \\
=& \frac{|Y \setminus (Y' \cup Y'')|}{|Y|} \int_{Y'\setminus(Y'\cup Y'')} \left(\int_X G(x,y)f(x)dx \right) \left(\int_Z G(y,z)g(z)dz \right) dy \\
\geq & (1-2\sqrt{\epsilon}) \int_{Y'\setminus(Y'\cup Y'')} \left(\int_X G(x,y)f(x)dx \right) \left(\int_Z G(y,z)g(z)dz \right) dy \\
\overset{Y',Y''}{\geq} & (1-2\sqrt{\epsilon}) \left(q_1 \int_X f(x) - \sqrt{\epsilon}p_1 \right) \left(q_2 \int_Z g(z) - \sqrt{\epsilon}p_2 \right)
\end{aligned}
$$

Expanding and removing some positive terms

$$
\begin{aligned}
\geq & q_1q_2 \int_X f(x) \int_Z g(z) - 2\sqrt{\epsilon}q_1q_2 \int_X f(x) \int_Z g(z) - 2\epsilon\sqrt{\epsilon}p_1p_2 \\
& - \sqrt{\epsilon}p_1q_2 \int_Z g(z) - \sqrt{\epsilon}p_2q_1 \int_X f(x) \\
\overset{f,g\leq 1, q\leq p}{\geq} & \int_X \int_Z q_1q_2f(x)g(z) - 6\sqrt{\epsilon}p_1p_2
\end{aligned}
$$

Rearranging the terms we obtain $\int_X \int_Z (\tilde{G}(x,z) - q_1q_2)f(x)g(z) \geq -6\sqrt{\epsilon}p_1p_2$ □

This result applies for the case $m = 2$. If we apply induction on Lemma 3.23 we can prove that $\tilde{G}$ satisfies DISC$_{\geq}$ for a general $m \geq 2$, by merging the graphs two by two:

Lemma 3.24. *Let G be a weighted graph with vertex subsets $X_0, X_1, ..., X_m$, with $m \geq 2$. Let $0 < \epsilon < 1$. If $(X_{i-1}, X_i)_G$ satisfies (q_i, p_i, ϵ)-DISC$_{\geq}$ for all $1 \leq i \leq m$, then the graph $\tilde{G}(x_0, x_m) = G(x_0, X_1, ..., X_{m-1}, x_m)$ satisfies $(q_1q_2 \cdots q_m, p_1p_2 \cdots p_m, 36\epsilon^{\frac{1}{2m}})$-DISC$_{\geq}$.*

Proof. We can define $(X_i, X_j)_{\tilde{G}}$ for any $i < j$ as $\tilde{G}(x_i, x_j) = G(x_i, X_{i+1}, ..., X_{j-1}, x_j)$ (we are taking the average of the paths from one vertex sets to another). We will show that, if $0 < j - i \leq 2^k$ for a non-negative integer k, then $(X_i, X_j)_{\tilde{G}}$ satisfies $(q_{i+1} \cdots q_j, p_{i+1} \cdots p_j, 36\epsilon^{2^{-k}})$. We proceed by induction on k.

For $k = 0$ we have $j - i = 1$, so by hypothesis, (X_i, X_j) satisfies (q_j, p_j, ϵ)-DISC$_{\geq}$ and hence $(q_j, p_j, 36\epsilon)$-DISC$_{\geq}$. If $k > 0$ and $j - i \leq 2^{k-1}$, then by induction $(X_i, X_j)_{\tilde{G}}$ satisfies $(q_{i+1} \cdots q_j, p_{i+1} \cdots p_j, 36\epsilon^{2^{-(k-1)}})$-DISC$_{\geq}$ and therefore $(q_{i+1} \cdots q_j, p_{i+1} \cdots p_j, 36\epsilon^{2^{-k}})$-DISC$_{\geq}$ (because $36\epsilon^{2^{-(k-1)}} \leq 36\epsilon^{2^{-k}}$).

Assume that $k > 0$ and $j - i > 2^{k-1}$. Then by induction $(X_i, X_{i+2^{k-1}})_{\tilde{G}}$ satisfies $(q_{i+1} \cdots q_{i+2^{k-1}}, p_{i+1} \cdots p_{i+2^{k-1}}, 36\epsilon^{2^{-(k-1)}})$-DISC$_{\geq}$ and $(X_{i+2^{k-1}}, X_j)_{\tilde{G}}$ satisfies $(q_{i+2^{k-1}+1} \cdots q_j, p_{i+2^{k-1}+1} \cdots p_j, 36\epsilon^{2^{-(k-1)}})$-DISC$_{\geq}$. Applying Lemma 3.23 (with $\epsilon' = 36\epsilon^{2^{-(k-1)}}$) we obtain that $(X_i, X_j)_{\tilde{G}}$ satisfies $(q_{i+1} \cdots q_j, p_{i+1} \cdots p_j, 36\epsilon^{2^{-k}})$.

To finalize, there is an integer k such that $m \leq 2^k < 2m$. For this value of k, $(X_0, X_m)_{\tilde{G}'}$ satisfies $(q_1 q_2 \cdots q_m, p_1 p_2 \cdots p_m, 36\epsilon^{2^{-k}})$-DISC$_{\geq}$, and $36\epsilon^{2^{-k}} \leq 36\epsilon^{\frac{1}{2m}}$. We conclude that $(X_0, X_m)_{\tilde{G}'}$ satisfies $(q_1 q_2 \cdots q_m, p_1 p_2 \cdots p_m, 36\epsilon^{\frac{1}{2m}})$-DISC$_{\geq}$. □

This completes the first step of the proof, as we have shown that $\tilde{G}$ satisfies DISC$_{\geq}$. The second step is by far the most complex in the proof, and it will consist of Lemmas 3.26, 3.27 and 3.28. It will require the definition of bounded graphs:

Definition 3.25. Let Γ be a weighted bipartite graph on vertex sets X and Y. We say that $(X, Y)_\Gamma$ is (p, ξ, η)-bounded if the following two conditions hold: $\Gamma(x, y) \leq \eta$ for all $x \in X$ and $y \in Y$, and $|\Gamma(x, Y) - p| \leq \xi p$ for every $x \in X$.

The η condition is simple: it is a bound to the weight of the edges. The ξ condition is a bit more subtle, and says that every vertex from x has roughly the same number of neighbours in Y.

Two important things to note in this definition: first, this definition can only be applied to unweighted graphs if $\eta \geq 1$ (which is equivalent to $\eta = 1$), as the weight of any edge is either 0 or 1. Second, the definition is not symmetric: $(X, Y)_\Gamma$ satisfying boundedness does not imply that $(Y, X)_\Gamma$ satisfies boundedness, because we impose $|\Gamma(x, Y) - p| \leq \xi p$ for every $x \in X$ but we do not impose that $|\Gamma(X, y) - p| \leq \xi p$. For example, in Figure 2, the graph (X, Y) on the left is a good candidate to satisfy boundedness, but the graph (X, Y) on the right is not (because the ξ condition says that every vertex from X has roughly the same number of neighbours in Y).

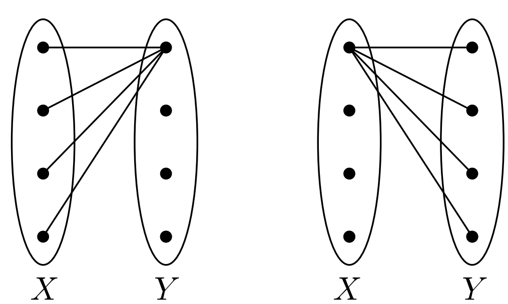

FIG. 2. Example of assymetry of boundedness

Now we can state the first lemma of step 2:

Lemma 3.26. *Let X, Y and Z be three vertex sets, and let $p_1, p_2, \xi_1, \xi_2, \xi_3 \in (0,1]$, and $\eta_1, \gamma_2 > 0$. Let Γ be a weighted graph such that $(X,Y)_\Gamma$ is (p_1, ξ_1, η_1)-bounded and $(Y,Z)_\Gamma$ is $(p_2, \xi_2, 1)$-bounded and $(p_2, \gamma = \gamma_2)$-jumbled. Let $\eta' = \max\{4\gamma_2^2 p_2^{-1} \xi_3^{-1} \eta_1, 4p_1p_2\}$ and $\xi' = \xi_1 + 2\xi_2 + 2\xi_3$. If $\tilde{\Gamma}(x,z) = \Gamma(x,Y,z)$ and $\Gamma' = \min\{\tilde{\Gamma}, \eta'\}$, then $(X,Z)_{\Gamma'}$ is (p_1p_2, ξ', η')-bounded.*

The statement says that, if $(X,Y)_\Gamma$ is bounded and $(Y,Z)_\Gamma$ is bounded and jumbled, then $(X,Z)_{\Gamma'}$ is also bounded, with certain parameters. The relations between the parameters are quite technical, but the role of each of them can clearly be seen in the statement, except for one: ξ_3 is a trade-off parameter. This means that, if we want (p, ξ', η')-boundedness in Γ', by increasing ξ_3 we can increase ξ' and decrease η', or the opposite by decreasing ξ_3. The proof will be very technical, and uses a 'few bad vertices' argument.

Proof. To see that Γ' is bounded, we must check that $\Gamma'(x,y) \leq \eta'$, $\Gamma'(x,Y) \leq (1+\xi')p_1p_2$ and $\Gamma'(x,Y) \geq (1-\xi')p_1p_2$. The first one is trivial from the definition of Γ'.

From the definition of Γ' we obtain $\Gamma'(x,y) \leq \tilde{\Gamma}(x,y)$ and $\Gamma'(x,Y) \leq \tilde{\Gamma}(x,Y)$. Now, we can use the boundedness of $(X,Y)_G$ and $(Y,Z)_G$ to obtain

$$\Gamma'(x,Z) \leq \tilde{\Gamma}(x,Z) = \Gamma(x,Y,Z) = \int\limits_Y \int\limits_Z \Gamma(x,y)\Gamma(y,z)dzdy \overset{\text{bound. YZ}}{\leq} \int\limits_Y \Gamma(x,y)(1+\xi_2)p_2dy \overset{\text{bound. XY}}{\leq} (1+\xi_1)p_1(1+\xi_2)p_2$$

This implies $\Gamma'(x,Z) \leq (1+\xi_1+2\xi_2)p_1p_2 \leq (1+\xi')p_1p_2$, which is the second condition of boundedness.

Finally, we need to prove $\Gamma'(x,Y) \geq (1-\xi')p_1p_2$. Fix some $x \in X$. We define $Z'_x = \{z \in Z : \Gamma(x,Y,z) > \eta'\}$ (this is the same 'bad vertex' idea as in other lemmas). Then we have that

$$\Gamma'(x,Z) \geq \int\limits_Y \int\limits_Z \Gamma(x,y)\Gamma(y,z)(1-\mathbb{1}_{Z'_x}(z)) = \Gamma(x,Y,Z) - \frac{|Z'_x|}{|Z|}\Gamma(x,Y,Z'_x)$$

Now we use the following chain of inequalities:

$$\begin{aligned}
\frac{1}{2}\frac{|Z'_x|}{|Z|}\frac{\eta'}{\eta_1} &\overset{(\eta'\geq 4p_1p_2)}{\leq} \eta_1^{-1}\left(\eta'\frac{|Z'_x|}{|Z|}-2p_1p_2\frac{|Z'_x|}{|Z|}\right)\\
&\leq \eta_1^{-1}\left(\eta'\frac{|Z'_x|}{|Z|}-(1+\xi_1)p_1p_2\frac{|Z'_x|}{|Z|}\right)\\
&\overset{(\text{def. } Z'_x,\text{ bound. } XY)}{\leq} \eta_1^{-1}\left(\frac{|Z'_x|}{|Z|}\Gamma(x,Y,Z'_x)-\Gamma(x,Y)p_2\frac{|Z'_x|}{|Z|}\right)\\
&\overset{(9)}{=} \int_Y\int_Z \eta_1^{-1}\Gamma(x,y)\Gamma(y,z)\mathbb{1}_{Z'_x}(z)-\int_Y\int_Z \eta_1^{-1}\Gamma(x,y)p_2\mathbb{1}_{Z'_x}(z)\\
&= \int_Y\int_Z \underbrace{\eta_1^{-1}\Gamma(x,y)}_{f(y)\in[0,1]}(\Gamma(y,z)-p_2)\underbrace{\mathbb{1}_{Z'_x}(z)}_{g(z)\in[0,1]}\\
&\overset{(\text{jumb. } YZ)}{\leq} \gamma_2\sqrt{\eta_1^{-1}\Gamma(x,Y)\frac{|Z'_x|}{|Z|}}\\
&\overset{(\text{bound. } XY)}{\leq} \gamma_2\sqrt{(1+\xi_1)p_1\eta_1^{-1}\frac{|Z'_X|}{|Z|}}
\end{aligned}$$

From this inequality we find a bound for $\frac{|Z'_x|}{|Z|}$, which is $\frac{|Z'_x|}{|Z|}\leq\frac{4\gamma_2^2(1+\xi_1)p_1\eta_1}{\eta'^2}$. Also, from the chain of inequalities we have $\eta_1^{-1}\left(\frac{|Z'_x|}{|Z|}\Gamma(x,Y,Z'_x)-\Gamma(x,Y)p_2\frac{|Z'_x|}{|Z|}\right)\leq\gamma_2\sqrt{(1+\xi_1)p_1\eta_1^{-1}\frac{|Z'_X|}{|Z|}}$, which can be rearranged as $\frac{|Z'_x|}{|Z|}\Gamma(x,Y,Z'_x)\leq\gamma_2\sqrt{(1+\xi_1)p_1\eta_1\frac{|Z'_X|}{|Z|}}+p_2\Gamma(x,Y)\frac{|Z'_x|}{|Z|}$. Plugging one into the other we obtain:

$$\begin{aligned}
\frac{|Z'_x|}{|Z|}\Gamma(x,Y,Z'_x) &\leq \gamma_2\sqrt{(1+\xi_1)p_1\eta_1\frac{|Z'_X|}{|Z|}}+p_2\Gamma(x,Y)\frac{|Z'_x|}{|Z|}\\
&\leq \frac{2\gamma_2^2(1+\xi_1)p_1\eta_1}{\eta'}+\frac{4\gamma_2^2(1+\xi_1)^2p_1^2p_2\eta_1}{\eta'^2}\\
&\overset{(\text{def. }\eta')}{\leq} \frac{2\gamma_2^2(1+\xi_1)p_1\eta_1}{4\gamma_2^2p_2^{-1}\xi_3^{-1}\eta_1}+\frac{4\gamma_2^2(1+\xi_1)^2p_1^2p_2\eta_1}{(4\gamma_2^2p_2^{-1}\xi_3^{-1}\eta_1)(4p_1p_2)}\\
&= \frac{1}{2}(1+\xi_1)\xi_3p_1p_2+\frac{1}{4}(1+\xi_1)^2\xi_3p_1p_2\\
&\leq 2\xi_3p_1p_2
\end{aligned}$$

Going back to what we wanted to prove,

$$\Gamma'(x,Z)\geq\Gamma(x,Y,Z)-\frac{|Z'_x|}{|Z|}\Gamma(x,Y,Z'_x)\geq(1-\xi_1)p_1(1-\xi_2)p_2-2\xi_3p_1p_1\geq(1-\xi')p_1p_2$$

This completes the proof of the third condition of boundedness, hence we conclude that $(X,Z)_{\Gamma'}$ is (p_1p_2,ξ',η')-bounded. □

Next, we will use Lemma 3.26 as an induction step to extend it to a path formed by m bipartite graphs, the biggest difference now is that there is no trade-off parameter:

Lemma 3.27. *Let $X_0, X_1, ..., X_m$ be vertex sets, with $m \geq 2$. Let c and ξ be such that $0 < 4c^2 < \xi < \frac{1}{4m}$, and $0 < p \leq 1$. Let Γ be a graph such that $(X_{i-1}, X_i)_\Gamma$ is $(p, \xi, 1)$-bounded and $(p, \gamma = cp^{1+\frac{1}{2m-2}})$-jumbled for all $1 \leq i \leq m$. Let $\tilde{\Gamma}(x_0, x_m) = \Gamma(x_0, X_1, ..., X_{m-1}, x_m)$ and $\Gamma' = \min\{\tilde{\Gamma}, 4p^m\}$. Then Γ' is $(p^m, 4m\xi, 4p^m)$-bounded.*

Again, there are three conditions that we need to prove to show boundedness. Like in the proof of Lemma 3.26, one is trivial, one requires few calculations, and the last one is the most complicated one. In this case, we apply Lemma 3.26 to sets of two graphs, applying the average-and-bound procedure to them and using induction to show the boundedness after i steps.

Proof. The condition $\Gamma'(x_0, x_m) \leq 4p^m$ comes from the definition of Γ'. To obtain $\Gamma'(x_0, X_m) \leq (1 + 4m\xi)p^m$ we expand and use boundedness on each graph:

$$\Gamma'(x_0, X_m) \leq \tilde{\Gamma}(x_0, X_m) = \int\limits_{X_1} \cdots \int\limits_{X_{m-1}} \int\limits_{X_m} \Gamma(x_0, x_1) \cdots \Gamma(x_{m-2}, x_{m-1})\Gamma(x_{m-1}, x_m)$$

$$\leq \int\limits_{X_1} \cdots \int\limits_{X_{m-1}} \Gamma(x_0, x_1) \cdots \Gamma(x_{m-2}, x_{m-1})(1 + \xi)p \leq ... \leq (1 + \xi)^m p^m$$

and $(1 + \xi)^m p^m \leq e^{m\xi} p^m \leq (1 + 4m\xi)p^m$ by the mean value theorem[2]. All we need to prove is $\Gamma'(x_0, X_m) \geq (1 - 4m\xi)p^m$

For this proof we will need to define some intermediate graphs. We will construct Γ_i for $1 \leq i \leq m$. Γ_i has vertex sets X_0, X_i and X_{i+1}, except for Γ_m which will only have X_0 and X_m.We construct Γ_1 as $(X_0, X_1)_{\Gamma_1} = (X_0, X_1)_\Gamma$ and $(X_1, X_2)_{\Gamma_1} = (X_1, X_2)_\Gamma$. For $2 \leq i < m$, we define $\Gamma_i(x_0, x_i) = \min\{\Gamma_{i-1}(x_0, X_{i-1}, x_i), \eta_i\}$ and $(X_i, X_{i+1})_{\Gamma_i} = (X_i, X_{i+1})_\Gamma$. Finally, $\Gamma_m(x_0, x_m) = \min\{\Gamma_{m-1}(x_0, X_{m-1}, x_m), \eta_m\}$. The value of η_i is

$$\eta_i = \max\{(4c^2\xi^{-1})^{i-1} p^{(i-1)(1+\frac{1}{m-1})}, 4p^i\}$$

First we see that $\eta_m = \max\{(4c^2\xi^{-1})^{m-1} p^m, 4p^m\} = 4p^m$, since $4c^2 < \xi$, and this means that $(4c^2\xi^{-1})^{m-1} p^m < p^m < 4p^m$. Moreover, we claim that, if $\eta_i = 4p^i$, then $\eta_{i+1} = 4p^{i+1}$. Indeed, for $i \geq 2$, we have $(4c^2\xi^{-1})^{i-1} p^{(i-1)(1+\frac{1}{m-1})} \geq 4p^i \Leftrightarrow 4c^2\xi^{-1} p^{(1+\frac{1}{m-1})} \geq 4p^{(1+\frac{1}{i-1})}$, and the right hand side of this last inequality is decreasing on i, while the left hand side does not depend on i. For $i = 1$, $\eta_1 = 4p \Rightarrow \max\{1, 4p\} = 4p \Rightarrow p \geq \frac{1}{4} \geq c^2\xi^{-1} \Rightarrow \eta_2 = \max\{4c^2\xi^{-1}p, 4p^2\} = 4p^2$. As a consequence, there is some t between 1 and m for which $\eta_i = (4c^2\xi^{-1})^{i-1} p^{(i-1)(1+\frac{1}{m-1})}$ for $i < t$ and $\eta_i = 4p^i$ for $i \geq t$. In addition, $\eta_1 = \max\{1, 4p\} \geq 1$.

We now claim that $(X_0, X_i)_{\Gamma_i}$ is $(p^i, 4i\xi, \eta_i)$-bounded. We proceed by induction on i. For $i = 1$, $(X_0, X_1)_{\Gamma_1} = (X_0, X_1)_\Gamma$ is $(p, \xi, 1)$-bounded, so it is also $(p, 4\xi, \eta_1)$-bounded. Now, for the

[2] For $f(x) = e^x$ and $0 < x \leq 1$, there is $c \in (0, 1)$ such that $e^x - 1 = f(x) - f(0) = xf'(c) = xf(c)$, where $1 = f(0) < f(c) < f(1) < 4$. Hence $x < e^x - 1 < 4x$, or equivalently, $1 + x < e^x < 1 + 4x$.

induction step, consider Lemma 3.26 with the following parameters: $p_1 = p^i$, $p_2 = p$, $\xi_1 = 4i\xi$, $\xi_2 = \xi_3 = \xi$, $\eta_1 = \eta_i$ and $\gamma_2 = cp^{1+\frac{1}{2m-2}}$. Then $\xi' = \xi_1 + 2\xi_2 + 2\xi_3 = 4(i+1)\xi$. We will see that $\eta_{i+1} = \max\{4\gamma_2^2 p_2^{-1}\xi_3^{-1}\eta_1, 4p_1p_2\}$.

If $i < t$, then $\max\{4\gamma_2^2 p_2^{-1}\xi_3^{-1}\eta_1, 4p_1p_2\} = \max\{4c^2\xi^{-1}p^{1+\frac{1}{m-1}}\eta_i, 4p^{i+1}\} = \max\{(4c^2\xi^{-1})^i p^{i(1+\frac{1}{m-1})}, 4p^{i+1}\} = \eta_{i+1}$. If $i \geq t$ then $\max\{4\gamma_2^2 p_2^{-1}\xi_3^{-1}\eta_1, 4p_1p_2\} = \max\{4c^2\xi^{-1}p^{1+\frac{1}{m-1}}(4p^i), 4p^{i+1}\} = 4p^{i+1} = \eta_{i+1}$.

By Lemma 3.26, if $(X_0, X_i)_{\Gamma_i}$ is $(p^i, 4i\xi, \eta_i)$-bounded, then $(X_0, X_{i+1})_{\Gamma_{i+1}}$ is $(p^{i+1}, 4(i+1)\xi, \eta_{i+1})$-bounded. By induction, $(X_0, X_m)_{\Gamma_m}$ is $(p^m, 4m\xi, 4p^m)$-bounded.

Finally, we notice that $\Gamma_i(x_0, x_i) \leq \Gamma(x_0, X_1, ..., X_{i-1}, x_i)$. Indeed, this is trivially true for $i = 1$, and if it holds for some i, then

$$\Gamma_{i+1}(x_0, x_{i+1}) \overset{\text{def. }\Gamma_{i+1}}{\leq} \Gamma_i(x_0, X_i, x_{i+1}) = \int_{X_i} \Gamma_i(x_0, x_i)\Gamma_i(x_i, x_{i+1})$$

$$\leq \int_{X_i} \Gamma(x_0, X_1, ..., X_{i-1}, x_i)\Gamma(x_i, x_{i+1}) = \Gamma(x_0, X_1, ..., X_i, x_{i+1})$$

We conclude that $\Gamma_m(x_0, x_m) \leq \min\{\Gamma(x_0, X_1, ..., X_m), \eta_m = 4p^m\} = \Gamma'(x_0, x_m)$ and

$$\Gamma'(x_0, X_m) \geq \Gamma_m(x_0, X_m) \overset{\text{bound.}}{\geq} (1 - 4m\xi)p^m$$

□

To finish the second step we need to extend this result from Γ to G. This is the first time that both Γ and G appear in the same lemma in the proof of the counting lemma. Lemma 3.28 says that if G is a path of bipartite graphs satisfying DISC$_\geq$, and they are subgraphs of bipartite graphs Γ which are jumbled and bounded, then G' (the result of the average-and-bound procedure) is also DISC$_\geq$, with some appropriate parameters, which is what the second step claims. The proof is based on the fact that we know that Γ' is bounded (Lemma 3.27) and $\tilde{G}$ satisfies DISC$_\geq$ (Lemma 3.24), and combining those two results using the inequality $\tilde{\Gamma} - \Gamma' \geq \tilde{G} - G'$.

Lemma 3.28. *Let $0 < 4c^2 < \xi$ and $0 < p \leq 1$. Let $X_0, X_1, ..., X_m$ be vertex sets, with $m \geq 2$. Let Γ be a graph and G be a subgraph of Γ such that $(X_{i-1}, X_i)_\Gamma$ is $(p, \xi, 1)$-bounded and $(p, \gamma = cp^{1+\frac{1}{2m-2}})$-jumbled, and $(X_{i-1}, X_i)_G$ satisfies (q_i, p, ϵ)-DISC$_\geq$, for all $1 \leq i \leq m$. Let $\tilde{G}(x_0, x_m) = G(x_0, X_1, ..., X_{m-1}, x_m)$ and $G' = \min\{\tilde{G}, 4p^m\}$. Then G' satisfies $(\underbrace{q_1q_2...q_m}_{=q}, p^m, 36\epsilon^{\frac{1}{2m}} + 8m\xi)$-DISC$_\geq$.*

Proof. We can suppose that $\xi < \frac{1}{4m}$, as otherwise $8m\xi \geq 2$ and any graph satisfies $(q, p, 2)$-DISC$_\geq$ (since $G \geq 0$, the integral $\int(G - q)uv$ is bounded by $-q$, and $-q \geq -p$).

Consider the following inequality: $\tilde{\Gamma} - \Gamma' \geq \tilde{G} - G'$, where $\tilde{\Gamma}$ and Γ' are defined analogously[3] as $\tilde{\Gamma}$ and Γ', respectively. Both the RHS and the LHS are non-negative. If the RHS is zero, then the

[3] $\tilde{\Gamma}(x_0, x_m) = \Gamma(x_0, X_1, ..., X_{m-1}, X_m)$ and $\Gamma' = \min\{\tilde{\Gamma}, 4p^m\}$

inequality holds. If the RHS is nonzero, then $G' = 4p^m$, which means that $\Gamma' = 4p^m = G'$ and the inequality becomes $\tilde{\Gamma} \geq \tilde{G}$, which is true. We conclude that $\tilde{\Gamma} - \Gamma' \geq \tilde{G} - G'$ holds in all cases.

We want to prove that, for any $f : X_0 \to [0,1]$ and $g : X_m \to [0,1]$,

$$\int_{X_0}\int_{X_m} (G'(x_0,x_m) - q)f(x_0)g(x_m) \geq -(36\epsilon^{\frac{1}{2m}} + 8m\xi)p^m$$

We split the integral into two:

$$\begin{aligned}
\int\int (G'-q)fg &= -\int\int(\tilde{G}-G')fg + \int\int(\tilde{G}-q)fg \\
&\overset{3.24}{\geq} -\int\int(\tilde{\Gamma}-\Gamma')fg - 36\epsilon^{\frac{1}{2m}}p^m \\
&\geq -\int\int(\tilde{\Gamma}-\Gamma') - 36\epsilon^{\frac{1}{2m}}p^m \\
&= -\int\int\tilde{\Gamma} + \int\int\Gamma' - 36\epsilon^{\frac{1}{2m}}p^m \\
&\overset{3.27}{\geq} -(1+\xi)^m p^m + (1-4m\xi)p^m - 36\epsilon^{\frac{1}{2m}}p^m \\
&\geq -(36\epsilon^{\frac{1}{2m}} + 8m\xi)p^m
\end{aligned}$$

□

where in the last inequality we used that $1 + x \leq e^x \leq 1 + 4x$ for $0 \leq x \leq 1$ to obtain $(1+\xi)^m \leq e^{\xi m} \leq 1 + 4m\xi$. This completes the proof of the lemmas forming the second step.

For the third step, we need to show that there is a large enough subgraph of G that satisfies boundedness. The proof will consist of applying a 'few bad vertices' argument on each set.

Lemma 3.29. *Let $0 < \delta, \tilde{\gamma}, \xi, p < 1$ satisfy $2\tilde{\gamma}^2 \leq \delta\xi^2p^2$. Let Γ be a graph with vertex subsets $X_0, X_1, ..., X_m$ such that $(X_{i-1}, X_i)_\Gamma$ is $(p, \gamma = (1-\delta)\tilde{\gamma})$-jumbled. Then we can find $\tilde{X}_0, \tilde{X}_1, ...\tilde{X}_m$, with $\tilde{X}_i \subseteq X_i$ and $|\tilde{X}_i| \geq (1-\delta)|X_i|$ such that $(\tilde{X}_{i-1}, \tilde{X}_i)_\Gamma$ is $(p, \xi, 1)$-bounded and $(p, \gamma = \tilde{\gamma})$-jumbled for all $1 \leq i \leq m$.*

Proof. Any choice of subsets $\tilde{X}_i$ with $|\tilde{X}_i| \geq (1-\delta)|X_i|$ will suffice for the jumbledness condition, because for any $\tilde{X}'_{i-1} \subseteq \tilde{X}_{i-1}$ and $\tilde{X}'_i \subseteq \tilde{X}_i$, using Lemma 3.19, we have

$$\left|e_\Gamma(\tilde{X}'_{i-1}, \tilde{X}'_i) - p|\tilde{X}'_{i-1}||\tilde{X}'_i|\right| \leq (1-\delta)\tilde{\gamma}\sqrt{|X_{i-1}||X_i||\tilde{X}'_{i-1}||\tilde{X}'_i|} \leq \tilde{\gamma}\sqrt{|\tilde{X}_{i-1}||\tilde{X}_i||\tilde{X}'_{i-1}||\tilde{X}'_i|}$$

The choice of subsets is only important for the boundedness condition. We will create the sets $\tilde{X}_i$ in decreasing order of i, from $\tilde{X}_m$ to $\tilde{X}_0$. We begin by making $\tilde{X}_m = X_m$. Now suppose that we have created $\tilde{X}_{i+1}$, and that $|\tilde{X}_{i+1}| \geq (1-\delta)|X_{i+1}|$. Let $X_{i,1}$ be the set of elements from X_i with $\Gamma(x_i, \tilde{X}_{i+1}) > (1+\xi)p$ this will play the role of the set of bad vertices. Now, using jumbledness with $f(x_i) = \mathbb{1}_{X_{i,1}}$ and $g(x_{i+1}) = \mathbb{1}_{\tilde{X}_{i+1}}$, we obtain

$$\frac{|X_{i,1}|}{|X_i|}\frac{|\tilde{X}_{i+1}|}{|X_{i+1}|}\xi p \overset{\text{def. } X_{i,1}}{\leq} \int_{X_i}\int_{X_{i+1}} (\Gamma(x_i, x_{i+1}) - p)\mathbb{1}_{X_{i,1}}(x_i)\mathbb{1}_{\tilde{X}_{i+1}}(x_{i+1}) \overset{\text{jumb.}}{\leq} (1-\delta)\tilde{\gamma}\sqrt{\frac{|X_{i,1}|}{|X_i|}\frac{|\tilde{X}_{i+1}|}{|X_{i+1}|}}$$

So $\frac{|X_{i,1}|}{|X_i|} \leq \frac{\tilde{(}1-\delta)^2\tilde{\gamma}^2|X_{i+1}|}{\xi^2 p^2 |\tilde{X}_{i+1}|} \leq \frac{\delta}{2}$. If we define $X_{i,2}$ as the elements from X_i with $\Gamma(x_i, \tilde{X}_{i+1}) \leq (1-\xi)p$, we also obtain $\frac{|X_{i,1}|}{|X_i|} \leq \frac{\delta}{2}$. We then can construct $\tilde{X}_i = X_i \setminus (X_{i,1} \cup X_{i,2})$. Then $\frac{|\tilde{X}_i|}{|X_i|} = \frac{|X_i|-|X_{i,1}|-|X_{i,2}|}{|X_i|} \geq 1 - \frac{2\delta}{2} = 1-\delta$. All the vertices from $\tilde{X}_i$ satisfy $|\Gamma(x_i, \tilde{X}_{i+1}) - p| \leq \xi p$, so $(\tilde{X}_i, \tilde{X}_{i+1})_\Gamma$ is $(p, \xi, 1)$-bounded. □

Finally, once we prove that G' satisfies $\text{DISC}_{\geq}$ for the subsets $\tilde{X}_i$, we need to extend it back to the complete sets X_i. For this purpose we use this lemma:

Lemma 3.30. *Let $0 \leq q \leq p \leq 1$ and let $\epsilon, \delta, \delta' > 0$. Let X and Y be vertex sets, and $\tilde{X}$ and $\tilde{Y}$ be subsets of X and Y, respectively, such that $|\tilde{X}| \geq (1-\delta)|X|$ and $|\tilde{Y}| \geq (1-\delta)|Y|$. Let G be a weighted bipartite graph on X and Y, and W be a weighted bipartite graph on $\tilde{X}$ and $\tilde{Y}$, such that $G(x,y) \geq (1-\delta')W(x,y)$ for every $x \in \tilde{X}$, $y \in \tilde{Y}$. If W satisfies (q,p,ϵ)-DISC$_{\geq}$ then G satisfies $(q,p,\epsilon+2\delta+2\delta')$-DISC$_{\geq}$.*

This means that, if we have two graphs G and W, with $V(W)$ being a large subset of $V(G)$, and G is not much smaller than W where W is defined, then $\text{DISC}_{\geq}$ in W implies $\text{DISC}_{\geq}$ in G. The proof consists on considering the contribution of the vertices of the sets $\tilde{X}_i$ to the integral from the definition of $\text{DISC}_{\geq}$.

Proof. Let $f : X \to [0,1]$ and $g : Y \to [0,1]$. Observe that, since $f \leq 1$, we have the following inequality:

$$\int_X f(x) - \frac{|\tilde{X}|}{|X|} \int_{\tilde{X}} f(x) = \int_X f(x)(1 - \mathbb{1}_{\tilde{X}}(x)) = \int_X f(x) \mathbb{1}_{X \setminus \tilde{X}}(x) = \frac{|X \setminus \tilde{X}|}{|X|} \int_{X \setminus \tilde{X}} f(x) \leq \frac{|X \setminus \tilde{X}|}{|X|} \leq \delta$$

This implies that $\frac{|\tilde{X}|}{|X|} \int_{\tilde{X}} f(x) \geq \int_X f(x) - \delta$ and, by analogy, $\frac{|\tilde{Y}|}{|Y|} \int_{\tilde{Y}} f(y) \geq \int_Y f(y) - \delta$.

Using that inequality we can show that

$$\begin{aligned}
\int_X \int_Y G(x,y)f(x)g(y) \geq& \int_X \int_Y G(x,y)f(x)g(y)\mathbb{1}_{\tilde{X}}(x)\mathbb{1}_{\tilde{Y}}(y) \\
=& \frac{|\tilde{X}||\tilde{Y}|}{|X||Y|}\int_{\tilde{X}}\int_{\tilde{Y}} G(x,y)f(x)g(y) \\
\geq& \frac{|\tilde{X}||\tilde{Y}|}{|X||Y|}\int_{\tilde{X}}\int_{\tilde{Y}} (1-\delta')W(x,y)f(x)g(y) \\
\overset{W\ \mathrm{DISC}_{\geq}}{\geq}& (1-\delta')\frac{|\tilde{X}||\tilde{Y}|}{|X||Y|}\left(q\left(\int_{\tilde{X}} f(x)\right)\left(\int_{\tilde{Y}} g(y)\right)-\epsilon p\right) \\
=& (1-\delta')\left(q\left(\frac{|\tilde{X}|}{|X|}\int_{\tilde{X}} f(x)\right)\left(\frac{|\tilde{Y}|}{|Y|}\int_{\tilde{Y}} g(y)\right)-\frac{|\tilde{X}||\tilde{Y}|}{|X||Y|}\epsilon p\right) \\
\geq& (1-\delta')\left(q\left(\int_X f(x)-\delta\right)\left(\int_Y g(y)-\delta\right)-\epsilon p\right) \\
&\text{Expanding and removing some positive terms} \\
\geq& q\int_X f(x)\int_Y g(y) - (\epsilon+2\delta+2\delta')p
\end{aligned}$$

□

We almost have everything in place to prove Lemma 3.22. Let us recapitulate what we have proved:

- If we have a jumbled graph Γ on vertex sets X_0, X_1, ..., X_m, then there are large subsets $\tilde{X}_i \subseteq X_i$ in which the graph Γ is both jumbled and bounded (Lemma 3.29).
- If Γ is a jumbled and bounded graph on vertex sets $\tilde{X}_0, \tilde{X}_1, ..., \tilde{X}_m$ and W is a subgraph of Γ satisfying $\mathrm{DISC}_{\geq}$, then the graph $W' = \min\{W(\tilde{x}_0, \tilde{X}_1, ..., \tilde{x}_m), 4p^m\}$ also satisfies $\mathrm{DISC}_{\geq}$ (Lemma 3.28).
- If G' is a graph on X_0, X_m and W' is a graph on $\tilde{X}_0, \tilde{X}_m$ satisfying $\mathrm{DISC}_{\geq}$, such that $G' \geq (1-\delta')W'$, then G' also satisfies $\mathrm{DISC}_{\geq}$ (Lemma 3.30).

Considering W as G restricted to the sets $\tilde{X}_i$, we see that the three steps above get us very close to the proof of the key lemma. We only need to take care of the details:

Proof of Lemma 3.22. Let $0 < c < 1$. Let $\delta = \frac{1}{4}c^{\frac{2}{3}}$, $\xi = 8c^{\frac{2}{3}}$ and $\gamma_1 = \frac{c}{1-\delta}p^{1+\frac{1}{2m-2}}$. These values satisfy $2\gamma_1^2 \leq \delta\xi^2p^2$. Indeed, $2\gamma_1^2 = 2\frac{c^2}{(1-\delta)^2}p^{2+\frac{1}{m-1}} \leq 8c^2p^2 \leq \delta\xi^2p^2$. This hypothesis will allow us to apply Lemma 3.29. In addition, we have $0 < 4c^2 < \xi$, the condition from Lemma 3.28.

Γ is $(p, \gamma = cp^{1+\frac{1}{2m-2}} = (1-\delta)\gamma_1)$-jumbled. By Lemma 3.29, we can find $\tilde{X}_i \subseteq X_i$ such that $|\tilde{X}_i| \geq (1-\delta)|X_i|$ and $(\tilde{X}_i, \tilde{X}_{i+1})_\Gamma$ is $(p, \xi, 1)$-bounded and $(p, \gamma = \gamma_1)$-jumbled for all $0 \leq i \leq$

$m-1$. Consider $W = G|_{\tilde{X}_0,\tilde{X}_1,\ldots,\tilde{X}_m}$, that is, the restriction of G to the vertex subsets constructed. Let $\tilde{W}(x_0,x_m) = W(x_0,\tilde{X}_1,\ldots,\tilde{X}_{m-1},x_m)$ and $W'(x_0,x_m) = \min\{\tilde{W}(x_0,x_m),4p^m\}$. The graph W is a subgraph of the jumbled bounded restricted graph Γ, so by Lemma 3.28 we obtain that W' is $(q,p^m,36\epsilon^{\frac{1}{2m}}+8m\xi)$-DISC$_\geq$.

We now want to find δ' such that $G' \geq (1-\delta')W'$.

$$\begin{aligned}
G'(x_0,x_m) =& \min\left\{\int\limits_{X_1}\cdots\int\limits_{X_{m-1}} G(x_0,x_1)\cdots G(x_{m-1},x_m),4p^m\right\}\\
\geq& \min\left\{\int\limits_{X_1}\cdots\int\limits_{X_{m-1}} G(x_0,x_1)\cdots G(x_{m-1},x_m)\mathbb{1}_{\tilde{X}_1}(x_1)\cdots\mathbb{1}_{\tilde{X}_{m-1}}(x_{m-1}),4p^m\right\}\\
=& \min\left\{\frac{|\tilde{X}_1|\cdots|\tilde{X}_{m-1}|}{|X_1|\cdots|X_{m-1}|}\int\limits_{\tilde{X}_1}\cdots\int\limits_{\tilde{X}_{m-1}} G(x_0,x_1)\cdots G(x_{m-1},x_m),4p^m\right\}\\
\geq& (1-\delta)^m\min\left\{\int\limits_{\tilde{X}_1}\cdots\int\limits_{\tilde{X}_{m-1}} W(x_0,x_1)\cdots W(x_{m-1},x_m),4p^m\right\}\\
\overset{\text{def. } W'}{=}& (1-\delta)^{m-1}W'(x_0,x_m)
\end{aligned}$$

This inequality shows that $G' \geq (1-\delta')W'$ is satisfied for $\delta' = 1-(1-\delta)^{m-1}$. Finally, Lemma 3.30 tells us that the graph G' satisfies $(q,p^m,\epsilon^{\frac{1}{2m}}+8m\xi+2\delta+2\delta')$-DISC$_\geq$, which is $(q,p^m,\epsilon^{\frac{1}{2m}}+64mc^{\frac{2}{3}}+\frac{1}{2}c^{\frac{2}{3}}+2-2(1-\frac{1}{4}c^{\frac{2}{3}}))$-DISC$_\geq$. Let $u(x) = x^{\frac{1}{2m}}$ and $v(x) = (64m+\frac{1}{2})x^{\frac{2}{3}}+2-2(1-x^{\frac{2}{3}})^{m-1}$. For a fixed m, we have $\lim_{x\to 0} u(x) = \lim_{x\to 0} v(x) = 0$.

Therefore, for every $\mu > 0$ there are $\epsilon > 0$ and $c > 0$ such that $u(\epsilon)+v(c) < \mu$. For those values, if $(X_{i-1},X_i)_\Gamma$ is $(p,\gamma = cp^{1+\frac{1}{2m-2}})$-jumbled and $(X_{i-1},X_i)_G$ is (q_i,p,ϵ)-DISC$_\geq$ for all $1\leq i\leq m$, then G' is (q,p^m,μ)-DISC$_\geq$. This completes the proof of Lemma 3.22. □

Before we go for the proof of the counting lemma, we will need the following lemma:

Lemma 3.31. *Let X, Y and Z be vertex sets. Let G be a graph on those vertex sets, such that $(X,Y)_G$ is $(q_{x,y},p_{x,y},\epsilon)$-DISC$_\geq$, $(Y,Z)_G$ is $(q_{y,z},p_{y,z},\epsilon)$-DISC$_\geq$ and $(Z,X)_G$ is $(q_{z,x},p_{z,x},\epsilon)$-DISC$_\geq$. If $G(y,z)\leq\eta_{y,z}$ for any $y\in Y$ and $z\in Z$, and $G(z,x)\leq\eta_{z,x}$ for any $z\in Z$ and $x\in X$, then*

$$\int\limits_X\int\limits_Y\int\limits_Z G(x,y)G(y,z)G(z,x) \geq q_{x,y}q_{y,z}q_{z,x}-\epsilon(p_{x,y}\eta_{y,z}\eta_{z,x}+q_{x,y}p_{y,z}\eta_{z,x}+q_{x,y}q_{y,z}q_{z,x})$$

No hypothesis is needed for the parameters q, p and η. We need this lemma beacuse we will use an integral in three vertex sets in the proof of the counting lemma, as we saw in page 41. The proof of this lemma consists of transforming the integral into a telescopic sum, and applying DISC$_\geq$

using $\frac{G}{\eta}$ as weights. The individual parameters are not important in the formula, all that matters is that, inside the parenthesis, each of the three products is of the form $s_{x,y}s_{y,z}s_{z,x}$, where s can be q, p or η. If $q_{y,z}$, $p_{y,z}$ and $\eta_{y,z}$ are all of the same order of magnitude, and so are the two terms $s_{x,y}$ and the three terms $s_{z,x}$, then the equation becomes

$$\int_X \int_Y \int_Z G(x,y)G(y,z)G(z,x) \geq q_{x,y}q_{y,z}q_{z,x} - O(\epsilon q_{x,y}q_{y,z}q_{z,x})$$

Proof. We can transform the integral into a telescopic sum:

$$\begin{aligned}\int_X \int_Y \int_Z G(x,y)G(y,z)G(z,x) = &\int_X \int_Y \int_Z (G(x,y) - q_{x,y})G(y,z)G(z,x) \\ &+ q_{x,y} \int_X \int_Y \int_Z (G(y,z) - q_{y,z})G(z,x) \\ &+ q_{x,y}q_{y,x} \int_X \int_Y \int_Z (G(z,x) - q_{z,x})\end{aligned}$$

Now we bound all integrals using DISC$_{\geq}$. The last one comes from using weights $f(z) = g(x) = 1$ in $(Z,X)_G$:

$$\int_X \int_Y \int_Z (G(z,x) - q_{z,x}) \geq -\epsilon p_{z,x}$$

The second integral can be bounded using $f(y) = 1$ and $g(z) = \frac{G(z,x)}{\eta_{z,x}}$ in $(Y,Z)_G$:

$$\int_X \int_Y \int_Z (G(y,z) - q_{y,z})G(z,x) = \eta_{z,x} \int_X \int_Y \int_Z (G(y,z) - q_{y,z}) \frac{G(z,x)}{\eta_{z,x}} \geq -\epsilon p_{y,x}\eta_{z,x}$$

Finally, for the first integral we bound using $f(x) = \frac{G(z,x)}{\eta_{z,x}}$ and $g(y) = \frac{G(y,z)}{\eta_{y,z}}$ in $(X,Y)_G$:

$$\begin{aligned}\int_X \int_Y \int_Z (G(x,y) - q_{x,y})G(y,z)G(z,x) =&\eta_{y,z}\eta_{z,x} \int_Z \int_X \int_Y (G(x,y) - q_{x,y}) \frac{G(y,z)}{\eta_{y,z}} \frac{G(z,x)}{\eta_{z,x}} \\ \geq&\epsilon p_{x,y}\eta_{y,z}\eta_{z,x}\end{aligned}$$

Substitution of these bounds into the original equation gives the inequality from the statement. □

Now we can finally prove the counting lemma:

Proof of Lemma 3.20. Let $a = \lfloor \frac{\ell+1}{2} \rfloor$. Consider first the case when ℓ is odd. If we want to apply Lemma 3.22 for graphs $X_1, X_2, ..., X_a$ and $X_a, X_{a+1}, ..., X_m$, we need the case $m = a - 1$. The exponent of p that we need in the jumbledness condition is $1 + \frac{1}{2m-2} = 1 + \frac{1}{\ell-3} = t_\ell$. If ℓ is even, then the cases are $m = a$ and $m = a - 1$, with exponents $1 + \frac{1}{2a-2}$ and $1 + \frac{1}{2a-4}$. The most restrictive exponent is $1 + \frac{1}{2a-4} = 1 + \frac{1}{\ell-4} = t_\ell$. This means that we can apply the lemma in both cases.

Let $X = X_1$, $Y = X_\ell$, and $Z = X_a$. We construct G' on these vertex sets as follows: $(X,Y)_{G'} = (X,Y)_G$, and $(Y,Z)_{G'}$ and $(Z,X)_{G'}$ following the construction from Lemma 3.22. Then $(X,Y)_{G'}$ satisfies (q_ℓ, p, ϵ)-DISC$_\geq$, and by Lemma 3.22, $(Z,X)_{G'}$ satisfies $(q_1q_2...q_{a-1}, p^{a-1}, \mu_1)$-DISC$_\geq$ and $(Y,Z)_{G'}$ satisfies $(q_aq_{a+1}...q_{\ell-1}, p^{\ell-a}, \mu_2)$-DISC$_\geq$. Also, $G'(z,x) \leq 4p^{a-1}$ and $G'(y,z) \leq 4p^{\ell-a}$.

If we take into account that, by hypothesis, $\alpha p \leq q_i \leq p$ for all i, then using Lemma 3.31 we obtain the following:

$$q_{x,y} = q_\ell \leq p \qquad q_{y,z} = q_aq_{a+1}\cdots q_{\ell-1} \leq p^{\ell-a} \qquad q_{z,x} = q_1q_2\cdots q_{a-1} \leq p^{a-1}$$
$$p_{x,y} = p \qquad p_{y,z} = p^{\ell-a} \qquad p_{z,x} = p^{a-1}$$
$$\eta_{y,z} = 4p^{\ell-a} \qquad \eta_{z,x} = 4p^{a-1}$$
$$\int_X\int_Y\int_Z G'(x,y)G'(y,z)G'(z,x) \geq q - \max\{\epsilon,\mu_1,\mu_2\}(21p^\ell)$$

Now, another consequence of lemma 3.22 is that there are values of $\epsilon > 0$ and $c > 0$ such that ϵ, μ_1 and μ_2 are all smaller than $\frac{\theta\alpha^\ell}{21}$, so for these values, $\max\{\epsilon,\mu_1,\mu_2\}(21p^\ell) \leq \theta(\alpha p)^\ell \leq q$.

On the other hand,

$$\begin{aligned}\int_X\int_Y\int_Z G'(x,y)G'(y,z)G'(z,x) &\leq \int_{X_0}\int_{X_a}\int_{X_\ell} G(x_0,X_1,...,x_a)G(x_a,X_{a+1},...,x_\ell)G(x_\ell,x_1)\\ &= \int_{X_1}\int_{X_2}...\int_{X_\ell} G(x_1,x_2)G(x_2,x_3)...G(x_\ell,x_1)\end{aligned}$$

Joining these inequalities we find that there are values of $\epsilon > 0$ and $c > 0$ for which

$$\begin{aligned}\int_{X_1}\int_{X_2}...\int_{X_\ell} G(x_1,x_2)G(x_2,x_3)...G(x_\ell,x_1) &\geq \int_X\int_Y\int_Z G'(x,y)G'(y,z)G'(z,x)\\ &\geq q - \max\{\epsilon,\mu_1,\mu_2\}(21p^\ell)\\ &\geq (1-\theta)q\end{aligned}$$

□

This completes the proof of the counting lemma for the case $H = C_\ell$ and $\ell \geq 5$.

Let us look back at one of the hypotheses of Lemma 3.20. We said that $(X_i, X_{i+1})_G$ is (q_i, p, ϵ)-DISC$_\geq$ with $\alpha p \leq q_i \leq p$. If we look back at the proof of the removal lemma in the dense case, we see that we first constructed a regular partition, and then remove, among others, the edges on pairs with small density. This explains why we are not consiering $q_i < \alpha p$. But what about $p < q_i$? We only used that condition because it was helpful with the calculations, but is was not specially important for the result, as it still holds if we allow $p < q_i$.

Indeed, to take care of the case $p < q_i$, however, note that we have only imposed one-sided discrepancy. We will use a property that DISC$_\geq$ satisfies, but $DISC$ does not:

Property 3.32. *Let $0 \leq q' \leq q \leq 1$ and $\epsilon, p \in (0,1]$. If G is a bipartite weighted graph on vertex sets X and Y, and is satisfies (q,p,ϵ)-DISC$_{\geq}$, then it also satisfies (q',p,ϵ)-DISC$_{\geq}$.*

Proof. Let $f : X \to [0,1]$ and $g : Y \to [0,1]$. Then

$$\int_X \int_Y (G(x,y) - q')f(x)g(y) \geq \int_X \int_Y (G(x,y) - q)f(x)g(y) \geq -\epsilon p$$

□

This means that, if G satisfies DISC$_{\geq}$ for $q_i > p$, then it also satisfies one-sided discrepancy for $q_i = p$. In particular, we can take all the parameters q_i from Lemma 3.20 to αp, and the value given by (13) becomes $(1-\theta)(\alpha p)^m$. This implies the following corollary, which is the one that we will use in the proof of the removal lemma:

Corollary 3.33. *Let $\ell \geq 5$ be a graph be an integer, and let $\alpha, \theta > 0$ be positive constants. Then there exist $c > 0$ and $\epsilon > 0$ such that the following holds: Let Γ be graph with vertex sets $X_1, X_2, ..., X_\ell$, each of them with n vertices, such that the bipartite graph $(X_i, X_{i+1})_\Gamma$ is $(p, \gamma = cp^{t_\ell})$-jumbled for all $1 \leq i \leq \ell$. Let G be a subgraph of Γ such that $(X_i, X_{i+1})_G$ is (q_i, p, ϵ)-DISC$_{\geq}$ with $\alpha p \leq q_i$ for all $1 \leq i \leq \ell$. Then*

$$||C_\ell \to G||_X \geq (1-\theta)\left(\alpha p n\right)^\ell \tag{14}$$

Before moving on to the next section, we will have a brief glimpse into the techniques used in the proof of the counting lemma for graphs other than a cycle of length at least 5. There are two main techniques: densification and doubling. Both of them are thoroughly discussed in [**ConFoxZha**].

The concept of densification is the same that we just saw in Lemma 3.22. If the graph H has a vertex v_3 of degree 2, with neighbours v_1 and v_2, and $v_1v_2 \notin E(H)$, then we construct a graph H' by deleting v_3 and its two edges, and join v_1 and v_2. For the graph G, we construct $G'(x_1,x_2) = \min\{G(x_1,X_3,x_2), 4p^2\}$. Then one can prove that, if the counting lemma holds for $||H' \to G'||_X$, then it also holds for $||H \to G||_X$.

Doubling is a more complicated technique, as it reduces counting of one graph to counting on three graphs. Also, unlike densification, doubling does not work well with one-sided counting, that is, to apply counting we need to prove both an upper bound and a lower bound for the number of embeddings, while in Corollary 3.33 we only give a lower bound.

If we want to apply doubling to a graph H, then choose a vertex a. Denote by H_{-a} the graph produced by removing a and the edges incident to it from H, and by H_a the graph consisting of a and its neighbours, with only the edges incident to a. Consider also the graph $H_{a,a\times 2}$, consisting of two copies of the vertex a, and one copy of each of its neighbours, forming a complete bipartite graph. One can show that, if the two-sided counting lemma (with an upper and a lower bound) holds for H_{-a}, for H_a and for $H_{a,a\times 2}$, then it also holds for H. This is illustrated in Figure 3.

Lemma 3.31 is enough to prove the counting lemma for triangles with at least two dense edges (a dense edge is one produced after the densification process). Proving the counting lemma for trees does not require special tecniques. The process to prove the counting lemma for $H = C_3$ is described in Figure 4. In that figure, red edges denote dense edges. We see that in the process we also prove the case $H = C_4$, and the proof consists of applying doubling and densification, twice each.

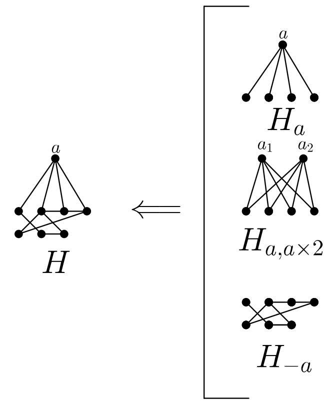

FIG. 3. Graphs resulting from the doubling process

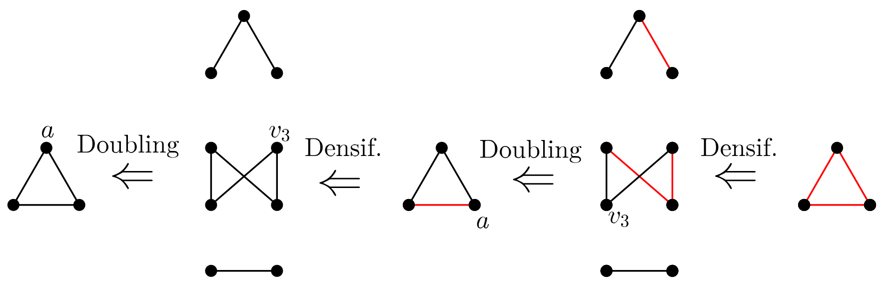

FIG. 4. Procedure to prove the counting lemma for triangles

3.5. The removal lemma

In this section we prove the removal lemma for cycles in sparse pseudorandom graphs. The proof will be analogous to the proof in the dense case, with some extra attention required for the pseudorandomness parameters involved.

For this proof, we will use the regularity lemma and the counting lemma. In particular, the versions of those lemmas that we will use will be Lemma 3.8 and Corollary 3.33, respectively. The version of the removal lemma that we prove is:

Theorem 3.34 (Removal lemma for pseudorandom graphs). *For every integer $\ell \geq 5$, and every $\mu > 0$ there are $\delta(\ell, \mu) > 0$ and $c(\ell, \mu) > 0$ for which the following holds: let $X_1, X_2, ..., X_\ell$ be vertex sets, each with n vertices. Let Γ be a graph for which $(X_i, X_{i+1})_\Gamma$ is $(p, \gamma = cp^{t_\ell})$-jumbled for all $1 \leq i \leq \ell$, and let G be a subgraph of Γ. If $||C_\ell \to G||_X \leq \delta p^\ell n^\ell$, then it is possible to remove at most $\mu p n^2$ edges from G so that $||C_\ell \to G||_X = 0$.*

Remember that $||C_\ell \to G||_X$ denotes the cycles in which the i-th vertex is in X_i for all $1 \leq i \leq \ell$.

Lemma 3.8 and Corollary 3.33, as well as this theorem, use the same notation (ϵ, c) for their parameters, but now we want to assign them different values while avoiding confusion. For this purpose, we will use the following functions:

- In Lemma 3.8, for every $\epsilon > 0$ and every positive integer m there exist $c = f_1(\epsilon, m)$ and $M = f_2(\epsilon, m)$, for which the statement holds.
- In Corollary 3.33, for every $\ell \geq 5$ and every $\alpha, \theta > 0$ there exist $c = f_3(\ell, \alpha, \theta)$ and $\epsilon = f_4(\ell, \alpha, \theta)$ for which the statement holds

Now we can prove the removal lemma for pseudorandom graphs:

Proof of Theorem 3.34. Once again, we consider $0 < \mu \leq \frac{1}{2}$, as otherwise it is enough to remove $\frac{1}{2}pn^2$ edges. We consider $c_1 = f_3(\ell, \frac{\mu}{2\ell}, \frac{1}{2})$ and $\epsilon_1 = f_4(\ell, \frac{\mu}{2\ell}, \frac{1}{2})$. We consider $\epsilon = \min\{\epsilon_1, \frac{\mu}{12\ell^2}\}$, $c_2 = f_1(\epsilon, \ell)$ and $M = f_2(\epsilon, \ell)$ Finally, let $\delta = \frac{1}{2}\left(\frac{\mu}{12\ell M}\right)^\ell$ and $c = \min\{c_1, \frac{c_2}{M}, \sqrt{\epsilon}, \frac{\ell}{M}\}$. We claim that these values satisfy the statement of the theorem.

If $(X_i, X_{i+1})_\Gamma$ is $(p, \gamma = cp^{t_\ell})$-jumbled, then it is also $(p, \gamma = c_2 p)$-jumbled. By Lemma 3.8, for this value of c there is an ϵ-regular partition $\mathcal{P}$ into k non-exceptional parts, $\ell \leq k \leq M$, which refines all sets X_i, by taking $\mathcal{P}_0 = \{X_1, X_2, ..., X_\ell\}$ as the initial partition.

Construct G^* by taking G and performing the following operations:

- Delete all edges having one of its ends in the exceptional set.
- Delete all edges between pairs not satisfying discrepancy.
- Delete all edges on pairs of parts satisfying $(q_{i,j}, p, \epsilon)$-DISC with $q_{i,j} < \frac{\mu}{6\ell^2}p$ (these are the pairs of parts with small density).

The number of edges that we remove is at most μpn^2. Let us see why:

- Let V_0 be the exceptional set from $\mathcal{P}$. Let $V_{0,i} = V_0 \cap X_i$. Then $|V_{0,i}| \leq |V_0| \leq \epsilon\ell n = \frac{\mu}{12\ell}n$. This implies

$$e_G(V_{0,i}, X_{i+1}) \leq e_\Gamma(V_{0,i}, X_{i+1}) \leq p|V_{0,i}||X_{i+1}| + \gamma n\sqrt{|V_{0,i}||X_{i+1}|} \leq p\frac{\mu}{12\ell}n^2 + cpn\sqrt{\epsilon n^2} \leq \frac{\mu}{6\ell}pn^2$$

  The same argument shows that $e_G(V_{0,i}, X_{i-1}) \leq \frac{\mu}{6\ell}pn^2$. The total number of edges removed in this step is at most

$$e_G(V_0, V) \leq \sum_{i=1}^{\ell} e_G(V_{0,i}, V) = \sum_{i=1}^{\ell}(e_G(V_{0,i}, X_{i-1}) + e_G(V_{0,i}, X_{i+1})) \leq \sum_{i=1}^{\ell}\frac{\mu}{3\ell}pn^2 = \frac{\mu}{3}pn^2$$

- There are at most ϵk^2 pairs not satisfying discrepancy. Each of them is between a pair of vertex sets V_i, V_j, with size $\frac{\ell n}{k} \geq |V_i| \geq \frac{\ell n}{2k}$. The number of edges between them is, by the jumbledness of Γ,

$$e_G(V_i, V_j) \leq e_\Gamma(V_i, V_j) \leq p|V_i||V_j| + cp^{t_\ell}n\sqrt{|V_i||V_j|} \leq pn^2\left(\left(\frac{\ell}{k}\right)^2 + c\frac{\ell}{k}\right) \leq \frac{2\ell^2}{k^2}pn^2$$

The number of edges removed in the second step is at most

$$\sum_{(V_i,V_j)\text{ irr.}} e_G(V_i,V_j) \le \epsilon k^2 \frac{2\ell^2}{k^2} pn^2 \le \frac{\mu}{12\ell^2} k^2 \frac{2\ell^2}{k^2} pn^2 < \frac{\mu}{3} pn^2$$

- There are at most k^2 pairs of parts satisfying $(q_{i,j}, p, \epsilon)$-DISC with $q_{i,j} \le \frac{\mu}{6\ell^2} p$. For those pairs,

$$e_G(V_i,V_j) \le q_{i,j}|V_i||V_j| + \epsilon p|V_i||V_j| \le \frac{\mu}{6\ell^2} p \left(\frac{\ell n}{k}\right)^2 + \frac{\mu}{12\ell^2} p \left(\frac{\ell n}{k}\right)^2 \le \frac{\mu p n^2}{3k^2}$$

The total number of edges removed in the third step is at most $k^2 \frac{\mu p n^2}{3k^2} \le \frac{\mu}{3} pn^2$.

Altogether, the number of edges that are removed in the construction of G^* is at most μpn^2.

Assume that $||C_\ell \to G^*||_X \ne 0$. Then each of the vertices of the cycle is contained in a non-exceptional part of $\mathcal{P}$, since in the construction of G^* we removed all edges incident to the exceptional set. Also, the edges of the cycle lie on different vertex sets X_i so, since $\mathcal{P}$ refines all the sets X_i, the vertices of the cycle lie on different parts of $\mathcal{P}$. We call those parts V_1, V_2, ..., V_ℓ, with $V_i \subseteq X_i$. The graph $(V_i, V_{i+1})_\Gamma$ is $(p, \gamma = cp^{t_\ell})$-jumbled, so it is also $(p, \gamma = c_2 p^{t_\ell})$-jumbled. The graph $(V_i, V_{i+1})_G$ is (q_i, p, ϵ)-DISC for some $q_i \ge \frac{\mu}{6\ell^2} p$, so it is also (q_i, p, ϵ_1)-DISC$_\ge$. By Corollary 3.33, the number of cycles with one vertex in each V_i is at least $\frac{1}{2}\left(\frac{\mu}{6\ell^2} p|V_i|\right)^\ell \ge \frac{1}{2}\left(\frac{\mu}{6\ell^2} p \frac{\ell n}{2M}\right)^\ell = \delta p^\ell n^\ell$.

This shows that, if $||C_\ell \to G||_X \le \delta p^\ell n^\ell$, then we can remove at most μpn^2 edges from G (by constructiong G^*) so that $||C_\ell \to G^*||_X = 0$, which is what the theorem states. □

3.6. Application: The sparse arithmetic removal lemma

As a conclusion to this thesis, we will prove a sparse version of Theorem 2.23, which can be found in [**ConFox1**], using the sparse removal lemma that we just proved. Like in the case of the graph removal lemma, when we move to a sparse environment we shall work in subsets of a pseudorandom set. For this reason, we will work with jumbled sets:

Definition 3.35 (Jumbled set). Let G be a finite group of order n. We say that a set S is (p, β)-jumbled if, for any $X, Y \subseteq G$, we have

$$||\{(x,y)|x \in X, y \in Y, xy \in S\}| - p|X||Y|| \le \beta\sqrt{|X||Y|}$$

This definition looks very similar to that of jumbled bipartite graphs. Indeed, this is what will allow us to go from one removal lemma to the other:

Lemma 3.36. *Let G be a finite group, let $p, \beta > 0$ and $S \subseteq G$. Let Γ be a bipartite graph defined as follows: it has two vertex sets X and Y, each with n vertices, which are labeled with the elements of G. We join $x_{g_1} \in X$ and $y_{g_2} \in Y$ if and only if $g_1^{-1} g_2 \in S$. Then the graph Γ is (p, β)-jumbled if and only if S is (p, β)-jumbled.*

Proof. Let A and B be subsets of G. We denote by A^{-1} the set of inverses of all the elements from A. Since inversion in a group is a bijection, then $|A^{-1}| = |A|$. Denote by $X_{A^{-1}}$ the set of vertices from X whose labels are in A^{-1}, and by Y_B the set of vertices from Y whose labels are in B. Then

$$e(X_{A^{-1}}, Y_B) = \{(x, y | x \in A^{-1}, y \in B, xy \in S\}$$

Also, since $|A| = |A^{-1}| = |X_{A^{-1}}|$ and $|B| = |Y_B|$ we have that

$$\big| |\{(x, y) | x \in A, y \in B, xy \in S\}| - p|A||B| \big| \leq \beta\sqrt{|A||B|}$$

$$\Updownarrow$$

$$\big| e(X_{A^{-1}}, Y_B) - p|X_{A^{-1}}||Y_B| \big| \leq \beta\sqrt{|X_{A^{-1}}||Y_B|}$$

This means that group jumbledness implies graph jumbledness. On the other hand, any subsets of X and Y can be written as $X_{A^{-1}}$ and Y_B for appropriate A and B, so the equivalence of both types of jumbledness follows. □

Using this equivalence, we can take the proof of Theorem 2.23 and extend it to the sparse jumbled case. Remember that $C(S_1, S_2, ..., S_k)$ is the number of solutions of $x_1 x_2 \cdots x_k = 1$ with $x_i \in S_i$:

Theorem 3.37 (Arithmetic sparse removal lemma). *For any integer $k \geq 3$ and any $\epsilon > 0$ there exist $\delta(k, \epsilon) > 0$ and $c(k, \epsilon) > 0$ for which the following holds: for any abelian group of order n, and any $(p, \beta = cp^{t_k} n)$-jumbled subset S, if S_1, S_2, ..., S_k are subsets of S for which $C(S_1, S_2, ..., S_k) \leq \delta p^k n^{k-1}$, then there are subsets $S_i' \subseteq S_i$ with $|S_i| \setminus |S_i'| \leq \epsilon pn$ and $C(S_1', S_2', ..., S_k') = 0$.*

Proof. We construct the graph K as follows: we consider k vertex sets X_i, each of which containing n vertices, each of which corresponds to an element of G. We denote by $v_{i,g}$ the vertex from X_i corresponding to $g \in G$. We join v_{i,g_1} and v_{i+1,g_2} if and only if $g_1^{-1} g_2 \in S_i$. We do the same for v_{k,g_1} and v_{1,g_2} (that is, we treat X_1 as X_{k+1}).

Let Γ be a graph on the same sets of vertices, where we join v_{i,g_1} and v_{i+1,g_2} if and only if $g_1^{-1} g_2 \in S$. Since $S_i \subseteq S$ for all i, every edge of K is also an edge of Γ, and K is a subgraph of Γ. In addition, due to the jumbledness condition on S, the graph $(X_i, X_{i+1})_\Gamma$ is (p, β)-jumbled.

As we showed in the proof of Theorem 2.23, $||C_k \to K||_X = nC(S_1, S_2, ..., S_k)$, since each solution of $x_1 x_2 \cdots x_k = 1$ generates n disjoint cycles. Consider the values of δ and c that result from Theorem 3.34 for $\ell = k$ and $\mu = \frac{\epsilon}{k}$. For those values, $(X_i, X_{i+1})_\Gamma$ is $(p, \beta = cp^{t_k} n)$-jumbled and $||C_k \to K||_X = nC(S_1, S_2, ..., S_k) \leq \delta p^k n^k$. This means that we can apply the sparse removal lemma.

Let E' be the set of at most μpn^2 edges from K such that removing them eliminates all cycles with one vertex in each X_i (the edges deleted in the removal lemma). To produce S_i' from S_i, we remove an element $s_i \in S_i$ if and only if there are at least $\frac{n}{k}$ edges of the form v_{i,g_1} and v_{i+1,g_2} with $g_1^{-1} g_2 = s_i$. Since every edge corresponds to exactly one element s_i, the number of removed elements from all sets is at most $\frac{|E'|}{n/k} \leq \frac{\mu pn^2}{n/k} \leq \epsilon n$.

Assume that $C(S'_1, S'_2, ..., S'_k) \neq 0$. Then there is a solution $x_1 x_2 \cdots x_k = 1$ with $x_i \in S'_i$. This solution generates n vertex-disjoint cycles in K, and in particular edge-disjoint. By construction of E', each of those cycles contains at least one edge of E', and that edge is of the form $v_{i,g_1} v_{i+1,g_2}$ with $g_1^{-1} g_2 = x_i$ for some $1 \leq i \leq k$. By pigeonhole principle, the value of i is the same for at least $\frac{n}{k}$ of those cycles, which means that there are at least $\frac{n}{k}$ different edges in E' of the form $v_{i,g_1} v_{i+1,g_2}$ with $g_1^{-1} g_2 = x_i$, and this implies that $x_i \notin S'_i$. This is a contradiction, so we must have $C(S'_1, S'_2, ..., S'_k) = 0$. □

This result can be used to prove a sparse version of Roth's theorem, but the proof is not as straightforward as in the dense case because we run into a small problem: it could be that S is contained in a jumbled set in G, but $2S$ is not. Fortunately, there is a workaround: if we consider $G = \mathbb{Z}/(4n+1)\mathbb{Z}$, then multiplying by 2 is an automorphism in G, so if a set is jumbled, then after multiplying each element by 2 it is still jumbled. This is because if f is an automorphism,

$$\begin{aligned} |\{(x,y) | x \in X, y \in Y, xy \in S\}| &= |\{(x,y) | x \in X, y \in Y, f(x)f(y) \in f(S)\}| \\ &= |\{(x,y) | x \in f(X), y \in f(Y), xy \in f(S)\}| \end{aligned}$$

This implies that, if S is contained in a jumbled set, then $2S$ is too.

3.7. Concluding remarks

We have seen that the regularity lemma opens a path to dealing with problems related to the structure of the graph. Cayley graphs or similar constructions allow us to extend to abelian finite groups this capability to analyze structures, which produces results such as Roth's theorem and the arithmetic removal lemma.

We have also seen that some results that are satisfied for dense graphs can be adapted to sparse graphs using pseudorandomness. This applies to the regularity lemma, the removal lemma and Roth's theorem, as seen here, but also to Turán's theorem [**ConFox1**], Erdős-Stone theorem [**ConFoxZha**] and results from Ramsey theory [**ConFoxZha**, **Koh**], among many others.

The adapted versions of those theorems that we saw used jumbledness, but this is not the only measure of pseudorandomness that can be used. Regularity and discrepancy, discussed here, and uniformity [**Sto**] are other commonly used measures of pseudorandomness which can serve the same purpose. Green and Tao [**GreTao**] used another pseudorandomness measure in which the set of prime numbers is a dense subset of a pseudorandom set in $\mathbb{N}$. This allowed them to extend Szemerédi's theorem to prime numbers:

Theorem 3.38 (Green-Tao). *The prime numbers contain an infinite number of non-trivial arithmetic k-term arithmeic progressions, for all positive integers k.*

References

[AjtSze] Ajtai, M. and Szemerédi, E., Sets of lattice points that form no squares, *Studia Scientiarum Mathematicarum Hungarica* **9** (1974), 9-11.

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

[ConFox1] Conlon, D. and Fox, J., Graph removal lemmas, *Surveys in Combinatorics* (2013), 1-50.

[ConFox2] Bounds for graph regularity and removal lemmas, *Geometric and Functional Analysis* **22** (2012), 1192-1256.

[ConFoxSud] Conlon, D., Fox, J. and Sudakov, B., Sidorenko's conjecture for a class of graphs: an exposition, *Geometric and Functional Analysis* **20** (2010), 1354-1366.

[ConFoxZha] Conlon, D., Fox, J. and Zhao, Y., Extremal results in pseudorandom graphs, *Advances in Mathematics* **256** (2014), 535-580.

[ConGow] Conlon, D. and Gowers, W.T., Combinatorial theorems in sparse random sets.

[Die] Diestel, R., "Graph theory", Electronic edition, Springer-Verlag Heidelberg, New York, 2005.

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

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

[FraRod] Frankl, P. and Rödl, V., Extremal problems on set systems, *Random Structures Algorithms* **20** (2002), 131-164.

[Fur] Füredi, Z., Extremal hypergraphs and combinatorial geometry, *Proceedings of the International Congress of Mathematics* **1** (1994), 1343-1352.

[Gow] Gowers, W.T., Hypergraph regularity and the multidimensional Szemerédi Theorem, *Annals of Mathematics* **166** (2007), 897-946.

[Gre] A Szemeréty-type regularity lemma in abelian groups, with applications, *Geometric and Functional Analysis* **15** (2005), 340-376.

[GreTao] Green, B. and Tao, T., The primes contain arbitrarily long arithmetic progressions, *Annals of Mathematics* **167** (2008), 481-547.

[Koh] Kohayakawa, Y., The Regularity Lemma of Szemerédi for sparse graphs (unpublished manuscript, 1993).

[KomSim] Komlós, J. and Simonovits, M., Szemerédi's Regularity Lemma and its applications in graph theory, *Combinatorics, Paul Erdős is eighty* **2** (1993), 295-352.

[KraSerVen] Král', D., Serra, O. and Vena, L., A combinatorial proof of the removal lemma for groups, *Journal of Combinatorial Theory, Series A* **116** (2009), 971-978.

[NagRodSch] Nagle, B., Rödl, V. and Schacht, M., The counting lemma for regular k-uniform hypergraphs, *Random Structures Algorithms* **28** (2006), 113-179.

[RodSko] Rödl, V. and Skokan, J., Regularity lemma for uniform hypergraphs, *Random Structures Algorithms* **25** (2004), 1-42.

[Rot] Roth, K.F., On certain sets of integers, *Journal of the London Mathematical Society* **28** (1953), 104-109

[RuzSze] Ruzsa, I.Z. and Szemerédi, E., Triple systems with no three points carrying three triangles, *Colloquia Mathematica Societatis János Bolyai* **18** (1978), 939-945.

[Sco] Scott, A., Szemerédi's Regularity lemma for matrices and sparse graphs, *Combinatorics, Probability and Computing* **20** (2011), 455-466.

[Sol] Solymosi, J., Note on a generalization of Roth's Theorem, *Algorithms and Combinatorics* **25** (2003), 825-827.

[Sze1] Szegedy, B., An information theoretic approach to Sidorenko's conjecture (2014).

[Sze2] Szemerédi, E., Integer sets containing no k elements in arithmetic progression, *Acta arithmetica* **27** (1975), 299-345.

[Sze3] Szemerédi, E., Regular partitions of graphs, *Colloques Internationaux C.N.R.S.* **260** (1976), 399-401.
[Tao] Tao, T., A variant of the hypergraph removal lemma, *Journal of Combinatorial Theory, Series A* **113** (2006), 1257-1280.