The induced removal lemma in sparse graphs

Shachar Sapir * Asaf Shapira †

Abstract

The induced removal lemma of Alon, Fischer, Krivelevich and Szegedy states that if an n-vertex graph G is ϵ-far from being induced H-free then G contains $\delta_H(\epsilon) \cdot n^h$ induced copies of H. Improving upon the original proof, Conlon and Fox proved that $1/\delta_H(\epsilon)$ is at most a tower of height $\text{poly}(1/\epsilon)$, and asked if this bound can be further improved to a tower of height $\log(1/\epsilon)$. In this paper we obtain such a bound for graphs G of density $O(\epsilon)$. We actually prove a more general result, which, as a special case, also gives a new proof of Fox's bound for the (non-induced) removal lemma.

1 Introduction

The graph removal lemma [1, 6, 10] is undoubtedly one of the cornerstone results of modern combinatorics. It states that if one must remove from an n-vertex graph G at least ϵn^2 edges in order to make it H-free then G contains at least $n^h/\text{Rem}_H(\epsilon)$ copies of H for some function $\text{Rem}_H(\epsilon)$ (we use $h = |V(H)|$). The original proof of the removal lemma relied on Szemerédi's regularity lemma [11] and thus only gave the very weak bound $\text{Rem}_H(\epsilon) \leq \text{twr}(\text{poly}(1/\epsilon))$, where $\text{twr}(x)$ is a tower of exponents of height x (so $\text{twr}(3) = 2^{2^2}$). Thanks to Gowers' lower bound for the regularity lemma [8], we know that any proof applying this lemma cannot yield better bounds. In a major breakthrough, Fox [7] (see also [4]) found a new proof of the removal lemma which avoided the regularity lemma and was thus able to show that $\text{Rem}_H(\epsilon) \leq \text{twr}(O(\log(1/\epsilon)))$. A different proof of Fox's result, more similar in nature to the proof of the removal lemma based on the regularity lemma, was recently given in [9].

The removal lemma was extended to the setting of induced subgraphs by Alon, Fischer, Krivelevich and Szegedy [2]. Let us say that an n-vertex graph G is ϵ-far from satisfying a property $\mathcal{P}$ if one must add/delete at least ϵn^2 edges to make it satisfy $\mathcal{P}$. The induced graph removal lemma of [2] then states that if G is ϵ-far from being induced H-free then G contains at least $n^h/\text{IRem}_H(\epsilon)$ induced copies of H. The original proof of this lemma in [2] introduced and used the so called strong regularity lemma and thus supplied only wowzer-type bounds for $\text{IRem}_H(\epsilon)$, where the wowzer function is the iterated version of the twr function. Conlon and Fox [3] gave a new proof of the induced removal lemma which avoided the strong regularity lemma, and were thus able to prove that $\text{IRem}_H(\epsilon) \leq \text{twr}(O(1/\epsilon^4))$. They later asked [4] if one can further improve this

*School of Mathematics, Tel Aviv University, Tel Aviv 69978, Israel.

†School of Mathematics, Tel Aviv University, Tel Aviv 69978, Israel. Email: asafico@tau.ac.il. Supported in part by ISF Grant 1028/16 and ERC Starting Grant 633509.

and show that $\mathrm{IRem}_H(\epsilon) \leq \mathrm{twr}(O(\log(1/\epsilon)))$ (and more generally prove such a bound for linear hypergraphs), that is, if one can extend Fox's bound for the removal lemma to the setting of the induced removal lemma. For more background and information on the removal lemma and its many variants, we refer the reader to the excellent survey [4].

1.1 New bounds for the induced removal lemma

Fox's new proof of the removal lemma [7], and Conlon and Fox's new proof of the induced removal lemma [3] used completely different methods. Our main result in this paper, stated as Theorem 1 below, is a new variant of the induced removal lemma, which contains both results as a special case, and furthermore answers the problem of [4] for sparse graphs. Before stating this theorem let us slightly extend the notion of induced H-freeness to finite families of graphs $\mathcal{H}$ by saying that G is induced $\mathcal{H}$-free if it is induced H-free for every $H \in \mathcal{H}$. It is easy to see that one can assume without loss of generality that all the graphs in a finite family $\mathcal{H}$ have the same size h. Let us also say that a graph has density p if $p = 2|E(G)|/n^2$. We are now ready to state the main result of this paper.

Theorem 1. *Fix a real $0 < p \leq \frac{1}{2}$ and a family of graphs $\mathcal{H}$ on h vertices. If an n-vertex graph G is ϵ-far from being induced $\mathcal{H}$-free and G has density p then G contains at least $n^h / \mathrm{twr}(O(\frac{p^2}{\epsilon^2} \log \frac{1}{p}))$ induced[1] copies of some $H \in \mathcal{H}$.*

First, note that since in the setting of induced subgraphs we can either work with a graph or its complement, the assumption $p \leq \frac{1}{2}$ actually covers all the range of possible edge densities. Hence, setting $p = \frac{1}{2}$ in the above theorem we get the following bound for the induced removal lemma which (slightly) improves the one obtained by Conlon and Fox [3].

Corollary 2. *For every H, $\mathrm{IRem}_H(\epsilon) \leq \mathrm{twr}(O(1/\epsilon^2))$.*

With regards to the problem of [4] whether $\mathrm{IRem}_H(\epsilon) \leq \mathrm{twr}(O(\log 1/\epsilon))$, setting $p = \epsilon$ (or more generally $p = O(\epsilon)$) in Theorem 1 we obtain the following corollary showing that such a bound holds for graphs of density $O(\epsilon)$.

Corollary 3. *If an n-vertex graph G is ϵ-far from being induced H-free and G has density $O(\epsilon)$ then G contains at least $n^h / \mathrm{twr}(O(\log 1/\epsilon))$ induced copies of H.*

Recalling the above mentioned observation that we can always switch between a graph and its complement, we immediately see that the above corollary also holds for graphs G of density $1 - O(\epsilon)$.

1.2 Reducing the general case to the sparse case

We believe that it should be possible to reduce the general case of the induced removal lemma to the sparse case, handled in Corollary 3, and thus answer positively the problem of [4] whether $\mathrm{IRem}_H(\epsilon) \leq \mathrm{twr}(O(\log 1/\epsilon))$. A special case in which we can achieve this goal is in the setting of the (non induced) removal lemma. Indeed, as we now show, Fox's famous improved bound for the removal lemma [7], follows as a simple corollary of Theorem 1.

[1] Here, and in similar statements, the O notation hides constants that depend on h, but not on ϵ or p (or n).

Corollary 4. *For every H,* $\mathrm{Rem}_H(\epsilon) \leq \mathrm{twr}(O(\log 1/\epsilon))$.

Proof. Suppose G is ϵ-far from being H-free. It is easy to see that G must contain $(\epsilon/|E(H)|)n^2$ copies of H that are edge disjoint. Let G' be the graph obtained by taking just those copies of H. Then G' has density ϵ and is ϵ/h^2-far from being H-free (since the copies are edge disjoint). Let $\mathcal{H}$ be the family of supergraphs of H on h vertices. Then G' is ϵ/h^2-far from being induced $\mathcal{H}$-free. Hence, by Theorem 1 it contains at least $n^h/\,\mathrm{twr}(O(\log 1/\epsilon))$ induced copies of some $H' \in \mathcal{H}$. Since such an H' contains H as a subgraph we are done. □

1.3 Paper Organization

As in [2] and [3], our proof of Theorem 1 will rely on a variant of the regularity lemma, stated in Lemma 3.2, from which the proof will follow rather easily. The proof of this lemma will rely on some tools previously used in [3] and [9], which we describe in Section 2. The proofs of Lemma 3.2 and Theorem 1 are given in Section 3.

2 Preliminary lemmas

2.1 Cylinder regularity lemma

Given a graph $G = (V, E)$ and two subsets of vertices $X, Y \subseteq V$ we write $d(X, Y)$ for the density of edges between X and Y, that is $d(X, Y) = |\{(x, y) \in E(G)|x \in X, y \in Y\}|/|X||Y|$. We say that the pair[2] of vertex sets (X, Y) is ϵ-regular if for all $A \subseteq X$ and $B \subseteq Y$ with $|A| \geq \epsilon|X|$ and $|B| \geq \epsilon|Y|$, we have $|d(A, B) - d(X, Y)| < \epsilon$. A partition $V_1, \ldots, V_k$ of $V(G)$ is equitable if the sizes of the parts differ by at most one, that is, if $||V_i| - |V_j|| \leq 1$ for all $i \neq j$. We use $|\mathcal{P}|$ to denote the number of parts in $\mathcal{P}$.

Suppose $G = (V, E)$ is a k-partite[3] graph with k-partition $V_1, \ldots, V_k$. A *cylinder* in G is any collection of k-tuples obtained by picking (non-empty) subsets $W_1 \subseteq V_1, \ldots, W_k \subseteq V_k$ and taking the product set $W_1 \times \ldots \times W_k$. Note that the collection of all k-tuples of vertices $(x_1, \ldots, x_k)$ with $x_i \in V_i$ is the cylinder $V_1 \times \ldots \times V_k$. A *cylinder partition* $\mathcal{K} = \{K_1, K_2, \ldots\}$ of the cylinder $V_1 \times \ldots \times V_k$ is a collection of cylinders so that every k-tuple $(x_1, \ldots, x_k) \in V_1 \times \ldots \times V_k$ belongs to precisely one $K \in \mathcal{K}$. We use $|\mathcal{K}|$ to denote the number of cylinders in the cylinder partition $\mathcal{K}$. We say that a cylinder K is ϵ-regular if all the $\binom{k}{2}$ pairs of subsets (W_i, W_j), $1 \leq i < j \leq k$, are ϵ-regular. A cylinder K is strongly ϵ-regular if in addition, all the pairs (W_i, W_i) are ϵ-regular, that is, if every W_i is ϵ-regular with itself. The cylinder partition $\mathcal{K}$ is ϵ-regular if all but a $\frac{1}{4}$-fraction of the k-tuples $(v_1, \ldots, v_k) \in V_1 \times \ldots \times V_k$ are in ϵ-regular cylinders of $\mathcal{K}$, and it is strongly ϵ-regular if the same applies for strong ϵ-regular cylinders.

The key idea in the proof of [3] was an application of the following lemma, the so called *Cylinder regularity lemma* of Duke, Lefmann and Rödl [5], which will also be instrumental in our proof.

[2] In all cases, the sets X, Y will either be disjoint or identical, that is, we will either consider bipartite ϵ-regular graphs, or graphs that are ϵ-regular. The latter are sometimes referred to as ϵ-quasi-random.

[3] We will always assume that in such a k-partition all the sets V_i are non-empty.

Lemma 2.1. *For each $0 < \alpha \le \frac{1}{2}$ and integer k, set $c(\alpha, k) = \alpha^{-k^2\alpha^{-5}}$. If $G = (V, E)$ is a k-partite graph with k-partition $V = V_1, \ldots, V_k$, then there is an α-regular cylinder partition $\mathcal{K}$ of $V_1 \times \ldots \times V_k$ satisfying $|\mathcal{K}| \le c(\alpha, k)$. Moreover, every $K = W_1 \times \ldots \times W_k \in \mathcal{K}$ and $1 \le i \le k$ satisfy $|W_i| \ge |V_i|/c(\alpha, k)$.*

We will also need[4] the following lemma from [3].

Lemma 2.2. *For each $0 < \gamma \le \frac{1}{2}$ set $r(\gamma) = 2^{\gamma^{-(\frac{20}{\gamma})^4}}$. Every graph $G = (V, E)$ has a vertex partition $\{V_1, V_2, \ldots\}$ into at most $r(\gamma)$ parts, such that each of the parts V_i is γ-regular with itself, i.e., each pair (V_i, V_i) is γ-regular.*

2.2 The potential function

Let the function $H : \mathbb{R}^+ \to \mathbb{R}^+$ be given by $H(x) = x \ln x$, where henceforth $0 \ln 0 = 0$. We will use H to define a potential function for vertex partitions. For the rest of this subsection let G be an n-vertex graph. We define the potential of a partition $\mathcal{P}$ of $V(G)$ by

$$H(\mathcal{P}) := \sum_{V, V' \in \mathcal{P}} \frac{|V||V'|}{n^2} H(d(V, V')) .$$

Observe that the summation is over ordered pairs (V, V'). It will be convenient to generalize the above definition. Henceforth, let $\mathcal{P}$ be a partition of $A \subseteq V(G)$ and $\mathcal{P}'$ be a partition of $A' \subseteq V(G)$. We more generally define

$$H(\mathcal{P}, \mathcal{P}') := \sum_{V \in \mathcal{P}, V' \in \mathcal{P}'} \frac{|V||V'|}{|A||A'|} H(d(V, V')) ,$$

and in particular $H(\mathcal{P}) = H(\mathcal{P}, \mathcal{P})$ if $\mathcal{P}$ is a partition of $V(G)$. We will also need the following two claims which follow easily from the convexity of the function $x \ln x$ (see either [7] or [9]).

Claim 2.3. *If $\mathcal{Q}$ refines $\mathcal{P}$ and $\mathcal{Q}'$ refines $\mathcal{P}'$ then $H(\mathcal{Q}, \mathcal{Q}') \ge H(\mathcal{P}, \mathcal{P}')$.*

Claim 2.4. *If G has density p then $p \ln p \le H(\mathcal{P}) \le 0$.*

Another observation we will use is the following technical lemma.

Lemma 2.5. *Let $G = (V, E)$ be a graph, and let $\mathcal{P} = \{V_1, \ldots, V_k\}$ be a partition of V. Then for every $t \ge k$, there is an equipartition $\mathcal{P}_0$ of V into t parts so that $H(\mathcal{P}_0) \ge H(\mathcal{P}) - \frac{k}{t}$.*

Proof. Let $x = |V| \pmod t$ and note that an equipartition of V into t parts has x parts of size $\left\lceil \frac{|V|}{t} \right\rceil$ and $t - x$ parts of size $\left\lfloor \frac{|V|}{t} \right\rfloor$. Let us then iteratively remove from V subsets of size $\left\lceil \frac{|V|}{t} \right\rceil$ so that every such set should be a subset of one of the sets V_i. We stop once we have picked x such sets or once each of the sets has fewer than $\left\lceil \frac{|V|}{t} \right\rceil$ vertices left in it. We now iteratively remove from the remaining vertices subsets of size $\left\lfloor \frac{|V|}{t} \right\rfloor$ so that every such set should be a subset of one

[4] We remark that if one is only interested in obtaining Corollary 4, then one can do away with Lemma 2.2 and its application in the proof of Lemma 3.1.

of the sets V_i. We stop once we have picked $t-x$ such sets or once each of the sets has fewer than $\left\lfloor \frac{|V|}{t} \right\rfloor$ vertices left in it.

Let us call the new partition we have thus created $\mathcal{Q}$, and let V_i' be the vertices that were not removed from V_i in the process of creating $\mathcal{Q}$. Let x' denote the number of subsets of size $\left\lceil \frac{|V|}{t} \right\rceil$ we have pulled out of the sets V_i and set $U = \bigcup_{i=1}^k V_i'$. We now pull $x - x'$ sets of size $\left\lceil \frac{|V|}{t} \right\rceil$ out of U and then partition the remaining vertices into sets of size $\left\lfloor \frac{|V|}{t} \right\rfloor$ (thus obtaining $t-x$ sets of this size). We thus get an equipartition of V which we take to be $\mathcal{P}_0$.

Noting that the (intermediate) partition $\mathcal{Q}$ refines $\mathcal{P}$, Claim 2.3 implies that $H(\mathcal{Q}) \geq H(\mathcal{P})$. It is also easy to see that the process of producing $\mathcal{Q}$ guarantees that each $|V_i'| \leq n/t$ and so $|U| \leq kn/t$. We thus have

$$
\begin{aligned}
H(\mathcal{P}_0) &= \sum_{\substack{X,Y\in\mathcal{Q}:\\ X,Y\subseteq V\setminus U}} H(d(X,Y))\frac{|X||Y|}{|V|^2} \\
&\quad + 2\sum_{\substack{X\in\mathcal{Q},Y\in\mathcal{P}_0:\\ X\subseteq V\setminus U,\, Y\subseteq U}} H(d(X,Y))\frac{|X||Y|}{|V|^2} + \sum_{\substack{X,Y\in\mathcal{P}_0:\\ X,Y\subseteq U}} H(d(X,Y))\frac{|X||Y|}{|V|^2} \\
&\geq \sum_{\substack{X,Y\in\mathcal{Q}:\\ X,Y\subseteq V\setminus U}} H(d(X,Y))\frac{|X||Y|}{|V|^2} - e^{-1}\frac{2|U|(|V|-|U|)+|U|^2}{|V|^2} \\
&\geq H(\mathcal{Q}) - \frac{k}{t} \geq H(\mathcal{P}) - \frac{k}{t}\,,
\end{aligned}
$$

where we used the fact that for every $0 \leq x \leq 1$ we have $-e^{-1} \leq H(x) \leq 0$. □

Suppose $\mathcal{P} = \{V_1,\ldots,V_k\}$ and $\mathcal{Q}$ are two partitions of $V(G)$ and $\mathcal{Q}$ refines $\mathcal{P}$. Then we will use $\mathcal{Q}|_{V_i}$ to denote the partition $\mathcal{Q}$ restricted to V_i, that is, the collection of parts of $\mathcal{Q}$ contained in V_i. We now write

$$
\ell_1(\mathcal{Q},\mathcal{P}) = \frac{1}{2}\sum_{i,j=1}^{k}\sum_{U\in\mathcal{Q}|_{V_i},U'\in\mathcal{Q}|_{V_j}} |U||U'||d(U,U') - d(V_i,V_j)|\,.
$$

We now turn to an important lemma from [9] (Lemma 3.4 in [9]), which follows from an appropriate application of Pinsker's Inequality.

Lemma 2.6. *Suppose $G=(V,E)$ is an n vertex graph of density p, $\mathcal{P}=\{V_1,\ldots,V_k\}$ is an equipartition of V, and $\mathcal{Q}$ is a partition refining $\mathcal{P}$. If $H(\mathcal{Q})-H(\mathcal{P}) < 2x^2p$, then $\ell_1(\mathcal{Q},\mathcal{P}) < xpn^2$.*

We now return to cylinder partitions, discussed above. Let $G=(V,E)$ and $\mathcal{P}=\{V_1,\ldots,V_k\}$ be an equipartition, and $\mathcal{K}$ be a cylinder partition of $V_1\times\ldots\times V_k$. For $K = W_1\times\ldots\times W_k \in \mathcal{K}$, we define

$$
d(K) = \frac{|W_1|\times\ldots\times|W_k|}{|V_1|\times\ldots\times|V_k|} \quad \text{and} \quad \ell_1(K,\mathcal{P}) = \frac{1}{k^2}\sum_{i\neq j=1}^{k}|d(W_i,W_j) - d(V_i,V_j)|\,.
$$

The reader should note that unlike the definition of ℓ_1 we used above, here we define the distance between a partition and a cylinder without summing on $i = j$.

We say that a cylinder partition $\mathcal{K}$ is ϵ-close to vertex partition $\mathcal{P}$ if

$$\sum_{K\in\mathcal{K}} d(K)\cdot \ell_1(K,\mathcal{P}) \le \epsilon\ .$$

If $K = W_1\times\ldots\times W_k$ is a cylinder in $V_1\times\ldots\times V_k$ then we write K_i for W_i. Given a cylinder partition $\mathcal{K}$ of $V_1\times\ldots\times V_k$, let $Q(\mathcal{K})$ be the partition of $V_1\cup\ldots\cup V_k$ obtained by taking the common refinement of all the sets K_i with $i\in[k]$ and $K\in\mathcal{K}$. Note that this partition is a refinement of the partition $V_1,\ldots,V_k$.

Lemma 2.7. *Let $G=(V,E)$ be of density p and $\mathcal{P}=\{V_1,\ldots,V_k\}$ be an equipartition of V. If a cylinder partition $\mathcal{K}$ of $V_1\times\ldots\times V_k$ is such that $\mathcal{Q}=\mathcal{Q}(\mathcal{K})$ satisfies $H(\mathcal{Q})<H(\mathcal{P})+\epsilon$, then $\mathcal{K}$ is $\sqrt{2p\epsilon}$-close to $\mathcal{P}$.*

Proof. We will assume that all sets V_i are of equal size since it will have no (real) affect on the following calculations. For the rest of this paragraph, let us fix $1\le i\ne j\le k$. For $K\in\mathcal{K}$, we have $d(K_i,K_j)=\sum_{U,U'} d(U,U')\frac{|U||U'|}{|K_i||K_j|}$, where here and in the rest of this paragraph, $\sum_{U,U'}$ is a sum over all $U\in\mathcal{Q}|_{V_i}$ with $U\subseteq K_i$ and $U'\in\mathcal{Q}|_{V_j}$ with $U'\subseteq K_j$. Hence, by the triangle inequality and the fact that $\sum_{U,U'}\frac{|U||U'|}{|K_i||K_j|}=1$ we get that

$$|d(K_i,K_j)-d(V_i,V_j)| \le \sum_{U,U'} |d(U,U')-d(V_i,V_j)|\frac{|U||U'|}{|K_i||K_j|}\ .$$

Averaging this inequality over all $K\in\mathcal{K}$ with weights $d(K)$, we have,

$$\begin{aligned}\sum_{K\in\mathcal{K}} |d(K_i,K_j)-d(V_i,V_j)|d(K) &\le \sum_{K\in\mathcal{K}}\sum_{U,U'} |d(U,U')-d(V_i,V_j)|\frac{|U||U'|}{|K_i||K_j|}d(K)\\ &= \sum_{U,U'}\left(|d(U,U')-d(V_i,V_j)|\sum_{K\in\mathcal{K}}\frac{|U||U'|}{|K_i||K_j|}d(K)\right)\\ &= \sum_{U,U'} |d(U,U')-d(V_i,V_j)|\frac{|U||U'|}{|V_i||V_j|}\ ,\end{aligned}$$

where in the first equality we just switched the order of summation meaning that here (and in the third line) the $\sum_{U,U'}$ is over all $U\in\mathcal{Q}|_{V_i}$ and $U'\in\mathcal{Q}|_{V_j}$ and the $\sum_{K\in\mathcal{K}}$ is only over $K\in\mathcal{K}$ satisfying $U\subseteq K_i$ and $U'\subseteq K_j$. The second equality follows from the fact that for every $U\subseteq V_i$ and $U'\subseteq V_j$, we have $\sum_K\prod_{t\ne i,j}|K_t|=\prod_{t\ne i,j}|V_t|$ with the sum being over all $K\in\mathcal{K}$ with $U\subseteq K_i$ and $U'\subseteq K_j$.

Summing the above inequality over all $i\ne j$, dividing by k^2 and recalling that all sets V_i are of size n/k we get

$$\begin{aligned}\sum_{K\in\mathcal{K}} \ell_1(K,\mathcal{P})d(K) &\le \frac{1}{n^2}\sum_{i\ne j}\sum_{U\in\mathcal{Q}|_{V_i},U'\in\mathcal{Q}|_{V_j}} |d(U,U')-d(V_i,V_j)||U||U'|\\ &\le \frac{2}{n^2}\cdot\ell_1(\mathcal{Q},\mathcal{P}) \le \sqrt{2p\epsilon}\ ,\end{aligned}$$

where the third inequality follows from the lemma's assumption that $H(\mathcal{Q}) < H(\mathcal{P}) + \epsilon$ together with Lemma 2.6. This completes the proof. $\square$

2.3 A counting and a slicing lemma

We will need a standard version of the counting lemma. See, e.g. [2] for a proof.

Lemma 2.8. *Suppose H is a graph with vertices $1, \dots, h$ and G is a graph with not necessarily disjoint vertex subsets $U_1, \dots, U_h$ such that*

- *For every $1 \le i < j \le h$, the pair (U_i, U_j) is $\frac{\epsilon^h}{4h}$-regular,*
- *For $1 \le i \le h$ we have $|U_i| \ge \frac{4h}{\epsilon^h}$,*
- *For $1 \le i < j \le k$, $d(U_i, U_j) > \epsilon$ if (i, j) is an edge of H and $d(U_i, U_j) < 1 - \epsilon$ otherwise.*

Then G contains at least $\frac{\epsilon}{4}^{\binom{h}{2}} |U_1| \times \ldots \times |U_h|$ induced copies of H with the copy of vertex i taken from U_i.

We will also need the following standard fact regarding ϵ-regular graphs, sometimes referred to as the slicing lemma.

Lemma 2.9. *If (X, Y) is an ϵ-regular pair, and $X' \subseteq X$ and $Y' \subseteq Y$ satisfy $|X'| \ge c|X|$ and $|Y'| \ge c|Y|$ for some $\epsilon \le c \le \frac{1}{2}$, then (X', Y') is an ϵ/c-regular pair.*

3 Proof of the main result

Our goal in this section is to prove Lemma 3.2, from which the proof of Theorem 1 will easily follow. We remind the reader that we always assume that $p \le 1/2$.

Lemma 3.1. *For every $0 < \epsilon \le p \le 1/2$ and h there is $z(\epsilon, h, p) = \mathrm{twr}(\frac{600hp^2}{\epsilon^2} \ln \frac{1}{p})$ such that the following holds: every graph G of density p on at least $1/\epsilon$ vertices has an equitable partition $\mathcal{P} = \{V_1, \dots, V_k\}$ with $1/\epsilon \le k \le z(\epsilon, h, p)$, and a strongly $\frac{\epsilon^h}{4h}$-regular cylinder partition $\mathcal{K}$ of $V_1 \times \ldots \times V_k$ so that $|\mathcal{Q}(\mathcal{K})| \le z(\epsilon, h, p)$ and $H(\mathcal{Q}(\mathcal{K})) \le H(\mathcal{P}) + \frac{\epsilon^2}{32p}$.*

Proof. First, if $1/\epsilon \le |V| < z(\epsilon, h, p)$ we can take $\mathcal{P}$ to be the trivial partition into singletons, and $\mathcal{K}$ to be the trivial cylinder partition consisting of a single $|V|$-tuple (in this case $\mathcal{Q}(\mathcal{K}) = \mathcal{P}$). It is easy to see that the lemma holds in this case. We can thus assume henceforth that $|V| \ge z(\epsilon, h, p)$. This fact will guarantee that in various equipartitions $\mathcal{P}$ of V we will obtain during the course of the proof, none of the parts of $\mathcal{P}$ will be empty[5]. In what follows, we will use $\mathrm{twr}(y, x)$ to denote a tower of x 2's with y on top (so, e.g., $\mathrm{twr}(y, 2) = 2^{2^y}$). We will prove that it is enough to take $z(\epsilon, h, p) = \mathrm{twr}(\frac{1}{\epsilon}, \frac{576hp^2}{\epsilon^2} \ln \frac{1}{p})$. It is easy to see that this is at most the $\mathrm{twr}(\frac{600hp^2}{\epsilon^2} \ln \frac{1}{p})$ bound stated in the lemma.

[5] Recall that when dealing with cylinders we assume that none of the parts is empty.

Set $\alpha = \frac{\epsilon^h}{4h}$. We first show that if $\mathcal{P} = \{V_1, \dots, V_k\}$ is an equipartition of V then we can find a strongly α-regular cylinder partition $\mathcal{K}$ of $V_1 \times \ldots \times V_k$ satisfying

$$|\mathcal{Q}(\mathcal{K})| \leq k2^{(r(\gamma))^k \cdot c(\alpha,k)} \,, \tag{1}$$

where $\gamma = \alpha/c(\alpha,k)$ and the functions $c(\alpha,k)$ and $r(\gamma)$ are those defined in the statements of Lemmas 2.1 and 2.2. We start by applying Lemma 2.2 with error parameter γ on each part V_i of $\mathcal{P}$. We get a partition $\{V_{i,1}, \dots, V_{i,g_i}\}$ of each part V_i into at most $r(\gamma)$ parts so that each part $V_{i,i'}$ is γ-regular. We now apply Lemma 2.1 on every possible cylinder, that consist of only one part from each V_i. More precisely, for each k-tuple $\ell = (\ell_1, \dots, \ell_k) \in [g_1] \times \ldots \times [g_k]$, we obtain an α-regular cylinder partition $\mathcal{K}_\ell$ of the cylinder $V_{1,\ell_1} \times \ldots \times V_{k,\ell_k}$ into at most $c(\alpha,k)$ cylinders, where for each $K = W_{1,\ell_1} \times \ldots \times W_{k,\ell_k} \in \mathcal{K}_{\ell_k}$ and each $1 \leq j \leq k$ we have $|W_{j,\ell_j}| \geq |V_{j,\ell_j}|/c(\alpha,k)$. Since $\gamma \leq c(\alpha,k)^{-1}$ Lemma 2.9 implies that each W_{j,ℓ_j} is $\gamma \cdot c(\alpha,k) = \alpha$-regular (with itself). Taking the union of the $\mathcal{K}_\ell$ we get a cylinder partition $\mathcal{K}$ of $V_1 \times \ldots \times V_k$ which is strongly α-regular. It is now easy to see that $\mathcal{Q}(\mathcal{K})$ satisfies (1).

We now turn to construct the partitions $\mathcal{P}$, $\mathcal{K}$ satisfying the statement of the lemma. To this end, we iteratively construct a sequence of partitions $\mathcal{P}_1, \mathcal{P}_2, \dots$ using the following process: We first set $\mathcal{P}_1$ to be an arbitrary equipartition of V into k parts, for $k = 1/\epsilon$. Assuming we have already constructed equipartition $\mathcal{P}_j$, we let $\mathcal{K}_j$ be the strongly α-regular cylinder partition one gets from the process described in the previous paragraph (so $\mathcal{Q}(\mathcal{K}_j)$ satisfies (1)). We then take $\mathcal{P}_{j+1}$ to be an equitable partition of $\mathcal{Q}(\mathcal{K}_j)$ into $\frac{64p}{\epsilon^2}|\mathcal{Q}(\mathcal{K}_j)|$ parts satisfying $H(\mathcal{P}_{j+1}) \geq H(\mathcal{Q}(\mathcal{K}_j)) - \frac{\epsilon^2}{64p}$, which exists by Lemma 2.5. We know from Claim 2.4 that every $j \geq 1$ satisfies $p\ln(p) \leq H(\mathcal{P}_j) \leq 0$, so there will be a $j \leq \frac{64p^2}{\epsilon^2} \ln \frac{1}{p}$ satisfying $H(\mathcal{P}_{j+1}) - H(\mathcal{P}_j) < \frac{\epsilon^2}{64p}$. Such a j will thus satisfy $H(\mathcal{P}_j) > H(\mathcal{P}_{j+1}) - \frac{\epsilon^2}{64p} \geq H(\mathcal{Q}(\mathcal{K}_j)) - \frac{\epsilon^2}{32p}$ so we can take $\mathcal{P} = \mathcal{P}_j$ and $\mathcal{K} = \mathcal{K}_j$ to be the two partitions in the statement of the lemma.

We now turn to prove the required upper bound on $|\mathcal{P}|$ and $|\mathcal{Q}(\mathcal{K})|$. We start by analysing the process of constructing $\mathcal{Q}_j$ and $\mathcal{P}_{j+1}$ in the iterative process described above. Suppose we start with a partition $\mathcal{P}_j$ satisfying $|\mathcal{P}_j| = k \geq 1/\epsilon$. Since $c(\alpha,k) \leq \alpha^{-(\frac{k}{\alpha})^5}$ we see that $\gamma = \alpha/c(\alpha,k) \geq 2^{-(\frac{k}{\alpha})^7}$ implying that $r(\gamma) \leq \text{twr}((k/\alpha)^{36}, 3)$. These facts and (1) imply that $|\mathcal{Q}(\mathcal{K}_j)| \leq \text{twr}((k/\alpha)^{37}, 4)$. Since $k \geq 1/\epsilon$ we can further simplify this bound to $|\mathcal{Q}(\mathcal{K}_j)| \leq \text{twr}(k, 8h)$. Recalling that $|\mathcal{P}_{j+1}| = \frac{64p}{\epsilon^2}|\mathcal{Q}_j|$ we conclude that $|\mathcal{Q}(\mathcal{K}_j)|, |\mathcal{P}_{j+1}| \leq \text{twr}(k, 9h)$. Finally, since $j \leq \frac{64p^2}{\epsilon^2} \ln \frac{1}{p}$ we are guaranteed that $|\mathcal{P}|, |\mathcal{Q}(\mathcal{K})| \leq \text{twr}(\frac{1}{\epsilon}, \frac{576hp^2}{\epsilon^2} \ln \frac{1}{p})$, as needed. □

We are now ready to state (and prove) our key lemma.

Lemma 3.2. *For every $0 < \epsilon \leq p \leq 1/2$ and h there is $S = S(\epsilon, h, p) = \text{twr}(O(\frac{p^2}{\epsilon^2} \ln \frac{1}{p}))$ so that the following holds: every graph $G = (V, E)$ of density p on at least $1/\epsilon$ vertices has an equitable partition $V_1, \dots, V_k$ where $k \geq 1/\epsilon$ and vertex subsets $W_i \subseteq V_i$ such that $|W_i| \geq |V|/S$, each pair (W_i, W_j) with $1 \leq i \leq j \leq k$ is $\frac{\epsilon^h}{4h}$-regular and*

$$\frac{1}{k^2} \sum_{i \neq j = 1}^{k} |d(W_i, W_j) - d(V_i, V_j)| < \epsilon \,. \tag{2}$$

Proof. We will prove the lemma with $S = 4z^2$ where $z = z(\epsilon, h, p)$ is given by Lemma 3.1.

Given G as in the statement, we apply Lemma 3.1 (with the same ϵ, h and p) and get an equipartition $\mathcal{P} = \{V_1, \ldots, V_k\}$ with $1/\epsilon \le k \le z$ and a strongly $\frac{\epsilon^h}{4h}$-regular cylinder partition $\mathcal{K}$ of $V_1 \times \ldots \times V_k$ such that the refinement $\mathcal{Q} = \mathcal{Q}(\mathcal{K})$ of $\mathcal{P}$ has at most z parts and satisfies $H(\mathcal{Q}) \le H(\mathcal{P}) + \frac{\epsilon^2}{32p}$. By Lemma 2.7, the cylinder partition $\mathcal{K}$ is $\frac{\epsilon}{4}$-close to $\mathcal{P}$. Hence

$$\sum_{K \in \mathcal{K}} d(K) \cdot \ell_1(K, \mathcal{P}) \le \frac{\epsilon}{4} \, .$$

Therefore, we get that $\sum d(K) \le 1/4$ where the sum is over all $K \in \mathcal{K}$ such that $\ell_1(K, \mathcal{P}) \ge \epsilon$. In other words, at most a $\frac{1}{4}$-fraction of the k-tuples $(v_1, \ldots, v_k) \in V_1 \times \ldots \times V_k$ belong to a $K \in \mathcal{K}$ that does not satisfy (2). Since $\mathcal{Q}(\mathcal{K})$ has at most z parts, the fraction of k-tuples $(v_1, \ldots, v_k) \in V_1 \times \ldots \times V_k$ that belong to parts $K = W_1 \times \ldots \times W_k$ of $\mathcal{K}$ with $|W_i| < \frac{1}{4z}|V_i|$ for at least one $i \in [k]$ is at most $\frac{1}{4z} z = \frac{1}{4}$. Finally, since $\mathcal{K}$ is a strongly $\frac{\epsilon^h}{4h}$-regular cylinder partition, we get that at most a $\frac{1}{4}$-fraction of the k-tuples belong to $K \in \mathcal{K}$ that is not strongly $\frac{\epsilon^h}{4h}$-regular. Hence, at least a $\frac{1}{4}$-fraction of the k-tuples $(v_1, \ldots, v_k) \in V_1 \times \ldots \times V_k$ belong to $K \in \mathcal{K}$ that is strongly $\frac{\epsilon^h}{4h}$-regular, satisfies (2), and every $1 \le i \le k$ satisfies $|W_i| \ge |V_i|/4z \ge |V|/4kz \ge |V|/4z^2 = |V|/S$. Hence, we can pick (at least) one K satisfying the assertion of the lemma. □

Proof of Theorem 1. Suppose G is an n-vertex graph of density p and G is ϵ-far from being induced $\mathcal{H}$-free, for some family of graphs $\mathcal{H}$ on h vertices. Note that this means that $\epsilon \le p$. First, notice that because $S(\epsilon/8, h, p) = \operatorname{twr}(O(\frac{p^2}{\epsilon^2} \ln \frac{1}{p}))$, if $n \le 2S(\epsilon/8, h, p)\frac{4h}{(\epsilon/8)^h}$, then $n^h / \operatorname{twr}(O(\frac{p^2}{\epsilon^2} \log \frac{1}{p})) < 1$. Since G must clearly contain at least one induced copy of H (as it is not induced-H free) the theorem holds in this case. So assume from now on that $n > 2S(\epsilon/8, h, p)\frac{4h}{(\epsilon/8)^h}$. Applying Lemma 3.2 on G with the same h and p but with $\epsilon/8$ instead of ϵ, we get the partition $V_1, \ldots, V_k$ and the sets $W_1, \ldots, W_k$ satisfying the conditions of the lemma with $k \ge 8/\epsilon$, and $|W_i| \ge \frac{n}{S(\epsilon/8,h,p)}$. For every i, j we now delete all the edges between V_i and V_j if $d(W_i, W_j) \le \epsilon/8$ and add all the edges between V_i and V_j if $d(W_i, W_j) \ge 1 - \epsilon/8$. The total number of edges removed between disjoint parts is at most

$$\begin{aligned}
\sum d(V_i, V_j)|V_i||V_j| &= \frac{n^2}{k^2} \sum d(V_i, V_j) \\
&\le \frac{n^2}{k^2} \left(\sum |d(V_i, V_j) - d(W_i, W_j)| + \sum d(W_i, W_j) \right) \\
&< \frac{n^2}{k^2} \left(\frac{\epsilon}{8} k^2 + \frac{\epsilon}{8} k^2 \right) = \frac{\epsilon}{4} n^2 \, ,
\end{aligned}$$

where the sums are over all $1 \le i \ne j \le k$ such that $d(W_i, W_j) \le \epsilon/8$, and in the last inequality we relied on (2). The same calculation gives that the total number of edges added between disjoint parts is also at most $\frac{\epsilon}{4} n^2$. Now, for each $1 \le i \le k$ we have $|V_i| < \frac{2n}{k}$, so the number of edges changed in each V_i is at most $\frac{4n^2}{k^2}$, and summing for all V_i we get $\frac{4n^2}{k}$. Since $k \ge \frac{8}{\epsilon}$, the total number of edges changed within parts is at most $\frac{\epsilon}{2} n^2$. All together we changed less than ϵn^2 edges. Since the graph was ϵ-far from being induced $\mathcal{H}$-free, we get that a copy of H for some $H \in \mathcal{H}$ remained. If vertex $i \in V(H)$ belongs to part $V_{i'}$ we set $U_i = W_{i'}$ (note that it might be the case that $U_i = U_j$ for $i \neq j$). So for every $i \neq j \in V(H)$, $d(U_i, U_j) \geq \epsilon/8$ if $(i,j) \in E(H)$ and $d(U_i, U_j) \leq 1 - \epsilon/8$ otherwise, because if not, this copy has already been removed. We also have that for each $i \neq j \in V(H)$ the pair (U_i, U_j) is $\frac{(\epsilon/8)^h}{4h}$-regular, and $|U_i| \geq \frac{n}{2S(\epsilon/8,h,p)} \geq \frac{4h}{(\epsilon/8)^h}$.

Now, all the conditions of the counting Lemma 2.8 are satisfied (with $\epsilon/8$ instead of ϵ), and we get that G contains at least $\frac{\epsilon}{32}^{\binom{h}{2}} \cdot |U_1| \cdot \ldots \cdot |U_h| \geq \frac{\epsilon}{32}^{\binom{h}{2}} (\frac{n}{2S(\epsilon/8,h,p)})^h \geq n^h / \operatorname{twr}(O(\frac{p^2}{\epsilon^2} \log \frac{1}{p}))$ induced copies of H, where the last inequality is true since $S(\epsilon/8, h, p) = \operatorname{twr}(O(\frac{p^2}{\epsilon^2} \ln \frac{1}{p}))$. □

Acknowledgement: We would like to thank Guy Moshkovitz for helpful discussions.

References

[1] N. Alon, R. A. Duke, H. Lefmann, V. Rodl, and R. Yuster, The algorithmic aspects of the regularity lemma, J. Algorithms **16** (1994), 80–109. 1

[2] N. Alon, E. Fischer, M. Krivelevich and M. Szegedy, Efficient testing of large graphs, Combinatorica **20** (2000), 451–476. 1, 1.3, 2.3

[3] D. Conlon and J. Fox, Bounds for graph regularity and removal lemmas, Geom. Funct. Anal. **22** (2012), 1191–1256. 1, 1.1, 1.1, 1.3, 2.1, 2.1

[4] D. Conlon and J. Fox, Graph removal lemmas, Surveys in Combinatorics, Cambridge university press, 2013, 1–50. 1, 1.1, 1.1, 1.2

[5] R. Duke, H. Lefmann and V. Rödl, A fast approximation algorithm for computing the frequencies of subgraphs in a given graph, SIAM J. Comput. **24** (1995), 598–620. 2.1

[6] P. Erdos, P. Frankl, and V. Rodl, The asymptotic number of graphs not containing a fixed subgraph and a problem for hypergraphs having no exponent, Graphs Combin. **2** (1986), 113–121. 1

[7] J. Fox, A new proof of the graph removal lemma, Ann. of Math. **174** (2011), 561–579. 1, 1.1, 1.2, 2.2

[8] W. T. Gowers, Lower bounds of tower type for Szemerédi's uniformity lemma, Geom. Funct. Anal. **7** (1997), 322–337. 1

[9] G. Moshkovitz and A. Shapira, A sparse regular aproximation lemma, Trans. Amer. Math. Soc., to appear. 1, 1.3, 2.2, 2.2

[10] I.Z. Ruzsa and E. Szemerédi, Triple systems with no six points carrying three triangles, in Combinatorics (Keszthely, 1976), Coll. Math. Soc. J. Bolyai 18, Volume II, 939–945. 1

[11] E. Szemerédi, Regular partitions of graphs, In: *Proc. Colloque Inter. CNRS*, 1978, 399–401. 1