Limitations on regularity lemmas for clustering graphs

Noga Alon[*]   Guy Moshkovitz[†]

Abstract

Szemerédi's regularity lemma is one instance in a family of regularity lemmas, replacing the definition of density of a graph by a more general coefficient. Recently, Fan Chung proved another instance, a regularity lemma for clustering graphs, and asked whether good upper bounds could be derived for the quantitative estimates it supplies. We answer this question in the negative, for every generalized regularity lemma.

1   Introduction

Szemerédi's regularity lemma [7] is a cornerstone of extremal combinatorics, with applications in graph theory, number theory, computer science and more. Underlying the lemma is the notion of *density* of a graph, which is the number of edges divided by the total number of vertex pairs. The lemma says, roughly, that any graph of density bounded away from zero can be approximated by a union of a constant number of bipartite graphs G_i that are regular, meaning that all induced subgraphs of a G_i with sufficiently many vertices have approximately the same density.

Recently, Chung [1] proved a variant of Szemerédi's regularity lemma that is tailor made for *clustering graphs*, which are graph with a *clustering coefficient* bounded away from zero. The clustering coefficient of a graph G is the number of triangles in G divided by the number of two-edge paths in G. It plays a key role in the definition of the "small world phenomenon", as clustering graphs come up often in real-world settings, such as social and neural networks ([5, 8]). Chung's regularity lemma says, roughly, that every graph G with a clustering coefficient bounded away from zero can be approximated by a union of a constant number of induced tripartite subgraphs G_i, such that all induced subgraphs of a G_i

with sufficiently many two-edge paths have approximately the same clustering coefficient. Importantly, this regularity lemma is meaningful even if the graph G does not have constant density, unlike Szemerédi's regularity lemma, as long as the clustering coefficient of G is constant.

More generally, for any fixed graph H and subgraph F of H on the same vertex set, one may consider the analogous (H, F)-*coefficient*, namely, the number of copies of H divided by the number of copies of F. It turns out that an "(H, F)-regularity lemma" holds in general (see [1]).

Since clustering graphs often appear "in nature", it is perhaps reasonable to suspect that they have an efficient regularity lemma, meaning one where the number of parts in the partition is a slow-growing function of the regularity parameter $1/\epsilon$. Indeed, the existence of an efficient regularity lemma was the first open question raised in Chung's paper [1], which adds that "It is of both theoretical and practical interest to see if the clustering property could be helpful for reducing the size of the partition". More generally, one might ask whether there is an efficient (H, F)-regularity lemma for *some* pairs (H, F). Here we prove that the answer to all these questions is negative.

1.1 Generalized regularity

Denote by $n_H(G)$, for a k-vertex graph H and a k-partite graph G, the number of unlabeled copies[1] of H in G where distinct vertices are mapped into distinct vertex classes. For a subgraph $F \subseteq H$, also on k vertices, the (H, F)-coefficient of a k-partite graph G is

$$c_{H,F}(G) = \frac{n_H(G)}{n_F(G)}.$$

Note that the density $d(G)$ of a bipartite graph G is a special case of a $c_{H,F}(G)$; specifically, if the vertex classes of G are (V_1, V_2) then $c_{K_2, \overline{K_2}}(G) = |G|/|V_1||V_2| = d(G)$. More generally, $c_{H, \overline{K_k}}(G)$ is the "H-density" (that is, normalized H count) of G. Moreover, $c_{K_3, P_2}(G)$ is the clustering coefficient of G, where P_2 denotes the 2-edge path graph.[2]

Let us formally define (H, F)-regularity. We use the notation $x \pm \epsilon$ for a number lying in the interval $[x - \epsilon, x + \epsilon]$.

Definition 1 ((H, F)-regular graph). *Let $F \subseteq H$ be graphs on k vertices. A k-partite graph G is ϵ-(H, F)-regular if for every induced k-partite subgraph G' with $n_F(G') \geq \epsilon n_F(G)$, we have $c_{H,F}(G') = c_{H,F}(G) \pm \epsilon$.*

---

[1]That is, subgraphs of G that are isomorphic to H. Alternatively, we may consider labeled copies (injective mappings $V(H) \to V(G)$ that map edges of H to edges G) with no real changes in our results.

[2]We note that [1] has, for technical reasons, two definitions of the clustering coefficient, depending on whether the graph is guaranteed to be tripartite or not, and in the former case the denominator is the number of two-edge paths with some fixed orientation. We opted to use a single definition.

Definition 2 ((H, F)-regular partition). *Let $F \subseteq H$ be graphs on k vertices. A vertex partition $\mathcal{P}$ of a graph G is ϵ-(H, F)-regular if all but at most $\epsilon n_H(G)$ copies of H in G lie in induced k-partite subgraphs $G[X_1, \ldots, X_k]$, with $X_1, \ldots, X_k \in \mathcal{P}$, that are ϵ-(H, F)-regular.*

Our result shows that in any proof of an (H, F)-regularity lemma, the order of the partition (that is, its number of parts) can in general be as large as a tower function, similarly to the case for Szemerédi's regularity lemma—as originally shown by Gowers [4]. In fact, the graphs we construct use Gowers' construction (or any other construction witnessing tower-type bounds) as a black box. Our proof proceeds by showing that for certain special graphs, as defined in Section 2, having an (H, F)-regular partition can be used to derive a similarly regular partition.

We are now ready to state our result. Formally, the tower function is defined recursively by $\mathrm{twr}(n) = 2^{\mathrm{twr}(n-1)}$ for $n \geq 1$ and $\mathrm{twr}(0) = 1$. To simplify the presentation, we omit all floor and ceiling signs whenever these are not crucial.

Theorem 1. *Let $F \subsetneq H$ be graphs on k vertices. There is a graph whose every ϵ-(H, F)-regular partition is of order at least $\mathrm{twr}(1/\mathrm{poly}(\epsilon k^k))$.*

2 Lower Bound Proof

2.1 Semi-blowups

The central definition in our proof is that of a *semi-blowup*. A *blowup* of a graph H is any graph obtained by replacing every vertex i of H by a (non-empty) independent set V_i of new vertices, each edge $\{i, j\}$ of H by a complete bipartite graph between V_i and V_j, and each non-edge $\{i, j\}$ of H by an empty bipartite graph between V_i and V_j. A *semi-blowup of H* is any graph obtained from a blowup of H by replacing the bipartite graph between V_i and V_j—for only one choice of $\{i, j\}$—by any bipartite graph. We write $H \odot_e G_0$ for a semi-blowup of H in which G_0 replaces the bipartite graph corresponding to $e = \{i, j\}$.

We write $\mathbb{N} = \{0, 1, 2, \ldots\}$ for the set of nonnegative integers, and $[k] = \{1, \ldots, k\}$.

Notation. *In everything that follows, we fix:*

$$\text{an integer } k \geq 2, \text{ graphs } F \subsetneq H \text{ on the vertex set } [k],$$
$$\text{and an edge } e = \{1, 2\} \in E(H) \setminus E(F).$$

Henceforth, in the semi-blowup $H \odot_e G_0$, the first vertex class of G_0 is embedded as V_1 and the second as V_2. We will repeatedly use the following easy properties of semi-blowups. The first property shows that an induced subgraph of a semi-blowup of H is again a semi-blowup of H.

Observation 1. *For all $U_1 \subseteq V_1, \ldots, U_k \subseteq V_k$, the induced subgraph $(H \odot_e G_0)[U_1, \ldots, U_k]$ is of the form $H \odot_e G_0[U_1, U_2]$.*

3

The second property shows that the "H-density" of a semi-blowup of H is exactly the (edge-) density of the replacement bipartite graph.

Observation 2. $n_H(H \odot_e G_0) = d(G_0)|V_1| \cdots |V_k|.$

Proof. Since $e = \{1, 2\} \in E(H)$, a choice of a vertex from each vertex class, $v_1 \in V_1, \ldots, v_k \in V_k$, spans a copy of H if and only if $\{v_1, v_2\} \in E(G_0)$. Moreover, such a copy is necessarily an induced copy of H, and thus the only copy of H on this set of vertices. We conclude $n_H(H \odot_e G_0) = |G_0| \cdot |V_3| \cdots |V_k| = d(G_0)|V_1| \cdots |V_k|.$ $\qquad\square$

Our choice of F and e has the following implication, which will be central to our proofs.

Lemma 2.1. *There are integers $a, b \in \mathbb{N}$, with $a \geq 1$ and $a + b \leq k!$, such that for every bipartite graph G_0 of density d,*

$$n_F(H \odot_e G_0) = (a + bd)|V_1| \cdots |V_k|. \tag{1}$$

This has the following corollaries:

$$n_F(H \odot_e G_0) \geq |V_1| \cdots |V_k|, \tag{2}$$

and, by Observation 2,

$$c_{H,F}(H \odot_e G_0) = \frac{d}{a + bd}. \tag{3}$$

Proof. Put $H^- = (V(H), E(H) \setminus \{e\})$. For every choice of vertices $v_1 \in V_1, \ldots, v_k \in V_k$, the copy of H they induce in G contains $n_F(H)$ copies of F if $\{v_1, v_2\} \in G_0$, and otherwise only $n_F(H^-)$ copies of F. It follows that

$$n_F(G) = n_F(H^-)|V_1| \cdots |V_k| + (n_F(H) - n_F(H^-))|G_0||V_3| \cdots |V_k|.$$

Put $a = n_F(H^-)$ and $b = n_F(H) - a$, so that $n_F(G) = (a + bd)|V_1| \cdots |V_k|$. We clearly have $a + b = n_F(H) \leq k!$. Moreover, and crucially, we have $a \geq 1$, since F has at least one copy in H that does not contain e, namely, F—as $e \notin E(F)$. This completes the proof. $\qquad\square$

2.2 Regularity of semi-blowups

We use Lemma 2.1 to show that a semi-blowup can be used to reduce (H, F)-regularity to Szemerédi's regularity. Recall that a bipartite graph G on (V_1, V_2) is said to be ϵ-*regular* if for all induced bipartite subgraphs $G[S_1, S_2]$ with $|S_1| \geq \epsilon|V_1|$ and $|S_2| \geq \epsilon|V_2|$ we have $d(G[S_1, S_2]) = d(G) \pm \epsilon$.

Claim 2.2. *For every bipartite graph G_0, if $H \odot_e G_0$ is ϵ-(H, F)-regular then G_0 is $\sqrt{\epsilon} \cdot k^{2k}$-regular.*

Proof. Put $G = H \odot_e G_0$ and $\epsilon' = \sqrt{\epsilon} \cdot k^{2k}$. Let $S_1 \subseteq V_1$ and $S_2 \subseteq V_2$ be subsets with $|S_1| \geq \epsilon'|V_1|$ and $|S_2| \geq \epsilon'|V_2|$, and put $d^* = d(G[S_1, S_2])$, $d = d(G[V_1, V_2])$. We will prove that $d^* = d \pm \epsilon'$, which would complete the proof as it would imply that G_0 is ϵ'-regular.

By Observation 1, the induced k-partite graph $G[S_1, S_2, V_3 \ldots, V_k]$ is a semi-blowup of H of the form $H \odot_e G[S_1, S_2]$. It follows from (3) that there are integers $a, b \in \mathbb{N}$, with $a \geq 1$ and $a + b \leq k!$, such that

$$c_{H,F}(G) = f(d) \quad \text{and} \quad c_{H,F}(G[S_1, S_2, V_3, \ldots, V_k]) = f(d^*) \tag{4}$$

where

$$f(x) = \frac{x}{a + bx}, \qquad f \colon [0, 1] \to \mathbb{R}.$$

Note that $f'(x) = a/(a+bx)^2 \geq 0$, so f is monotone increasing and $f(x) \leq f(1) = 1/(a+b)$. Let

$$g(x) = \frac{ax}{1 - bx}, \qquad g \colon [0, 1/b) \to \mathbb{R}.$$

Observe that g and $g \circ f$ are well defined and, for every $x \in [0, 1]$, we have $g(f(x)) = x$. Note that $g'(x) = a/(1 - bx)^2$ for every $x \in [0, 1/b)$, and so is monotone increasing; this implies the bound

$$g'\big(f(x) + \epsilon\big) \leq g'\left(\frac{b + a/2}{b(a+b)}\right) = a\left(\frac{a+b}{a/2}\right)^2 \leq k^{2k},$$

where the first inequality assumes $\epsilon \leq \frac{1}{k^{2k}}$ (otherwise there is nothing to prove), so $\epsilon \leq \frac{a/2}{b(a+b)}$.

Using (2) we deduce the bound

$$n_F(G[S_1, S_2, V_3, \ldots, V_k]) \geq |S_1||S_2||V_3| \cdots |V_k| \geq \epsilon'^2|V_1| \cdots |V_k| \geq (\epsilon'^2/k!)n_F(G) \geq \epsilon \cdot n_F(G).$$

Thus, since G is ϵ-H/F-regular, $c_{H,F}(G[S_1, S_2, V_3 \ldots, V_k]) = c_{H,F}(G) \pm \epsilon$, that is, $f(d^*) = f(d) \pm \epsilon$ by (4). We deduce, using the properties of f and g above, that

$$d^* = g(f(d^*)) = g(f(d)) \pm \epsilon \cdot \max_{x = f(d) \pm \epsilon} g'(x) = d \pm \epsilon \cdot k^{2k},$$

where the second step applies the mean value theorem on g in the interval between $f(d^*)$ and $f(d)$. Since $\epsilon \cdot k^{2k} \leq \epsilon'$, this completes the proof. $\square$

2.3 Regular partitions—auxiliary claims

The following claim gives an approximate restriction of a partition onto a subset.

Claim 2.3 (Approximate restriction). *Let $\mathcal{P}$ be a partition of U and let $V \subseteq U$ be a subset. For $\mathcal{P}' = \big\{X \in \mathcal{P} \,\big|\, |X \cap V| \geq \delta|X|\big\}$ with $\delta = \alpha|V|/|U|$ we have*

$$\sum_{X \in \mathcal{P}'} |X \cap V| \geq (1 - \alpha)|V|.$$

Proof. We have

$$\sum_{X \in \mathcal{P} \setminus \mathcal{P}'} |X \cap V| \le \sum_{X \in \mathcal{P} \setminus \mathcal{P}'} \delta|X| \le \sum_{X \in \mathcal{P}} \delta|X| = \delta|U| = \alpha|V|.$$

Therefore,

$$\sum_{X \in \mathcal{P}'} |X \cap V| = \sum_{X \in \mathcal{P}} |X \cap V| - \sum_{X \in \mathcal{P} \setminus \mathcal{P}'} |X \cap V| \ge (1 - \alpha)|V|.$$

$\square$

The *slicing lemma* for regular graphs shows that a sufficiently large induced subgraph of a regular bipartite graph is regular. We will need the following analogue for (H, F)-regularity.

Claim 2.4 (Slicing). *Let T be a k-partite graph on $(X_1, \ldots, X_k)$, and let $Y_1 \subseteq X_1, \ldots, Y_k \subseteq X_k$ with $|Y_i| \ge \delta_i|X_i|$ be such that $G := T[Y_1, \ldots, Y_k]$ is a semi-blowup of H. If T is ϵ-(H, F)-regular then G is ϵ'-(H, F)-regular with $\epsilon' = \epsilon \cdot k!/(\delta_1 \cdots \delta_k)$.*

Proof. Put $\Delta = \delta_1 \cdots \delta_k$. Since, by assumption, G is a semi-blowup of H, we have, using (2),

$$n_F(G) \ge |Y_1| \cdots |Y_k| \ge \Delta|X_1| \cdots |X_k| \ge \frac{\Delta}{k!} n_F(T) = \frac{\epsilon}{\epsilon'} n_F(T).$$

Let $S_1 \subseteq Y_1, \ldots, S_k \subseteq Y_k$ be subsets with $n_F(G[S_1, \ldots, S_k]) \ge \epsilon' \cdot n_F(G)$. Then, by the above inequality, $n_F(G[S_1, \ldots, S_k]) \ge \epsilon \cdot n_F(T)$. Since T is ϵ-(H, F)-regular, we thus have

$$c_{H,F}(G[S_1, \ldots, S_k]) = c_{H,F}(T) \pm \epsilon.$$

A special case of the above, obtained by taking $S_i = Y_i$, is

$$c_{H,F}(G) = c_{H,F}(T) \pm \epsilon.$$

Combining the last two estimates, we get $c_{H,F}(G[S_1, \ldots, S_k]) = c_{H,F}(G) \pm 2\epsilon$. Since $2\epsilon \le \epsilon'$ we deduce that G is ϵ'-(H, F)-regular, as needed. $\square$

We will also need the following standard fact.

Fact 2.5. *If G is a bipartite graph of density at most ϵ^3 then G is ϵ-regular.*

To see why Fact 2.5 is true, note that all subsets S and T, each of size at least an ϵ-fraction of its vertex class, satisfy $d(S, T) \le d(G)/\epsilon^2 \le \epsilon$.

2.4 Putting everything together

We are now ready to prove Theorem 1. Recall that a vertex partition $\mathcal{P}$ of a graph G on n vertices is said to be ϵ-*regular* if $\sum_{X,X'} |X||X'| \leq \epsilon n^2$ where the sum is over all cluster pairs $X, X' \in \mathcal{P}$ such that $G[X, X']$ is not ϵ-regular. We use the following standard terminology: a (semi-) blowup is said to be *balanced* if all vertex classes have the same size. The *common refinement* of partitions $\mathcal{P}$ and $\mathcal{V}$ is the partition $\{X \cap V \mid X \in \mathcal{P}, V \in \mathcal{V}, X \cap V \neq \emptyset\}$.

Our main result towards the proof of Theorem 1 is the following, which in some sense lifts Claim 2.2 to a statement about partitions. Note that in the construction here we need the blowup to be balanced, or at least not too "unbalanced", since a relatively small vertex class could be completely covered by small fragments of clusters from the regular partition, thereby making our slicing lemma in Claim 2.4 unusable.

Theorem 2. *If $\mathcal{P}$ is an ϵ-(H, F)-regular partition of a balanced semi-blowup $H \odot_e G_0$ then the common refinement of $\mathcal{P}$ and $\{V_1, V_2\}$ is an $\epsilon^{1/4} k^{2k}$-regular partition of G_0.*

Proof. Put $\epsilon' = \epsilon^{1/4} k^{2k}$. Let $G = H \odot_e G_0$ be a semi-blowup of H with $n = |V_1| = \cdots = |V_k|$, and let $\mathcal{P}$ be an ϵ-(H, F)-regular partition of G. For $i \in \{1, 2\}$, denote the common refinement of $\mathcal{P}$ and $\{V_i\}$ by $\mathcal{Q}_i = \{X \cap V_i \mid X \in \mathcal{P}, \, X \cap V_i \neq \emptyset\}$. We need to prove that

$$\sum_{\substack{(Y_1, Y_2) \in \mathcal{Q}_1 \times \mathcal{Q}_2 \\ \text{not } \epsilon'\text{-regular}}} |Y_1||Y_2| \leq \epsilon' n^2. \tag{5}$$

Let $\mathcal{P}_i = \{X \in \mathcal{P} \mid |X \cap V_i| \geq \delta_i |X|\}$ where we set, with hindsight,

$$\delta_i = \begin{cases} \frac{1}{2}\epsilon' & \text{if } i \in \{1, 2\} \\ \frac{1}{2k} & \text{if } i \in \{3, \ldots, k\}. \end{cases}$$

Moreover, let $\mathcal{P}_i^* = \{X \cap V_i \mid X \in \mathcal{P}_i\}$. Using Claim 2.3 and the fact that G is balanced,

$$\sum_{\substack{Y_3, \ldots, Y_k : \\ \forall i \, Y_i \in \mathcal{P}_i^*}} \prod_{i=3}^{k} |Y_i| = \prod_{i=3}^{k} \sum_{Y_i \in \mathcal{P}_i^*} |Y_i| \geq \left(\frac{1}{2}n\right)^{k-2}. \tag{6}$$

On the way to proving (5), our first goal will be to prove an analogous bound for $\mathcal{P}_1^* \times \mathcal{P}_2^*$, rather than for $\mathcal{Q}_1 \times \mathcal{Q}_2$, as follows;

$$\sum_{\substack{(Y_1, Y_2) \in \mathcal{P}_1^* \times \mathcal{P}_2^* \\ \text{not } \epsilon'\text{-regular}}} |G_0[Y_1, Y_2]| \leq \epsilon 2^k \cdot |G_0|. \tag{7}$$

Put $\Delta = \delta_1 \cdots \delta_k \ (\geq \epsilon'^2/(2k)^k)$. We claim that if the k-partite graph $G[X_1, \ldots, X_k]$ with $X_1 \in \mathcal{P}_1, \ldots, X_k \in \mathcal{P}_k$ is ϵ-(H, F)-regular then the bipartite graph $G[X_1 \cap V_1, X_2 \cap V_2]$ is ϵ'-regular. To see this, put $Y_i = X_i \cap V_i$ and note that $G[Y_1, \ldots, Y_k]$ is of the form

7

$H \odot_e G[Y_1, Y_2]$, by Observation 1, and since $|Y_i| \geq \delta_i |X_i|$ for every $1 \leq i \leq k$, it is $\epsilon k! \Delta^{-1}$-(H, F)-regular by Claim 2.4. Next, Claim 2.2 implies that $G[Y_1, Y_2]$ is $\sqrt{\epsilon k! \Delta^{-1}} k^{2k}$-regular. Since $\sqrt{\epsilon k! \Delta^{-1}} k^{2k} \leq (\sqrt{\epsilon}/\epsilon') k^{3k} \leq \epsilon'$, our claim follows. Now, to simplify the discussion, let us introduce two pieces of notation:

$$\mathcal{R}_{\mathcal{P}} = \left\{ (X_1, \ldots, X_k) \in \mathcal{P}_1 \times \cdots \times \mathcal{P}_k \,\middle|\, G[X_1, \ldots, X_k] \text{ is not } \epsilon\text{-}(H, F)\text{-regular} \right\},$$

$$\mathcal{R}_{\mathcal{P}}^* = \left\{ (X_1, \ldots, X_k) \in \mathcal{P}_1 \times \cdots \times \mathcal{P}_k \,\middle|\, G[X_1 \cap V_1, X_2 \cap V_2] \text{ is not } \epsilon'\text{-regular} \right\}.$$

Then the claim above, in contrapositive, says $\mathcal{R}_{\mathcal{P}}^* \subseteq \mathcal{R}_{\mathcal{P}}$. We deduce that

$$\sum_{(X_1, \ldots, X_k) \in \mathcal{R}_{\mathcal{P}}} n_H(G[X_1, \ldots, X_k]) \geq \sum_{(X_1, \ldots, X_k) \in \mathcal{R}_{\mathcal{P}}^*} n_H(G[X_1, \ldots, X_k])$$

$$\geq \sum_{(X_1, \ldots, X_k) \in \mathcal{R}_{\mathcal{P}}^*} n_H(G[X_1 \cap V_1, \ldots, X_k \cap V_k])$$

$$= \sum_{\substack{Y_3, \ldots, Y_k: \\ \forall i \, Y_i \in \mathcal{P}_i^*}} \sum_{\substack{(Y_1, Y_2) \in \mathcal{P}_1^* \times \mathcal{P}_2^* \\ \text{not } \epsilon'\text{-regular}}} n_H(G[Y_1, \ldots, Y_k])$$

$$= \sum_{\substack{Y_3, \ldots, Y_k: \\ \forall i \, Y_i \in \mathcal{P}_i^*}} \sum_{\substack{(Y_1, Y_2) \in \mathcal{P}_1^* \times \mathcal{P}_2^* \\ \text{not } \epsilon'\text{-regular}}} |G[Y_1, Y_2]| \cdot |Y_3| \cdots |Y_k|$$

$$\geq \left(\frac{1}{2} n \right)^{k-2} \sum_{\substack{(Y_1, Y_2) \in \mathcal{P}_1^* \times \mathcal{P}_2^* \\ \text{not } \epsilon'\text{-regular}}} |G[Y_1, Y_2]| \,,$$

where the last equality uses Observation 1 and Observation 2, and the last inequality uses (6). On the other hand, as $\mathcal{P}$ is an $\epsilon\text{-}(H, F)$-regular partition of G,

$$\sum_{(X_1, \ldots, X_k) \in \mathcal{R}_{\mathcal{P}}} n_H(G[X_1, \ldots, X_k]) \leq \epsilon \cdot n_H(G) = \epsilon \cdot |G_0| n^{k-2},$$

where the last equality uses Observation 2. Combining the above two inequalities implies (7).

To deduce from (7) that $\mathcal{Q}_1 \cup \mathcal{Q}_2$ is indeed an ϵ'-regular partition of G_0, we will need two more observations. First, using Fact 2.5,

$$\sum_{\substack{(Y_1, Y_2) \in \mathcal{P}_1^* \times \mathcal{P}_2^* \\ \text{not } \epsilon'\text{-regular}}} |Y_1||Y_2| \leq \sum_{\substack{(Y_1, Y_2) \in \mathcal{P}_1^* \times \mathcal{P}_2^* \\ \text{not } \epsilon'\text{-regular}}} \epsilon'^{-3} |G_0[Y_1, Y_2]| \leq \epsilon'^{-3} \epsilon 2^k |G_0| \leq \frac{1}{2} \epsilon' |G_0| \leq \frac{1}{2} \epsilon' n^2,$$

where the second inequality follows from (7), and the penultimate inequality uses the choice of ϵ'. Second, we claim that every cluster pair $(Y_1, Y_2) \in (\mathcal{Q}_1 \times \mathcal{Q}_2) \setminus (\mathcal{P}_1^* \times \mathcal{P}_2^*)$ satisfies $|Y_1||Y_2| \leq \delta_1 |X_1||X_2|$, where $X_i \in \mathcal{P}$ is uniquely defined by $Y_i = X_i \cap V_i$. Indeed, without loss of generality $Y_1 \notin \mathcal{P}_1^*$, meaning $|Y_1| \leq \delta_1 |X_1|$, so $|Y_1||Y_2| \leq \delta_1 |X_1||X_2|$. It follows that

$$\sum_{\substack{(Y_1, Y_2) \in \mathcal{Q}_1 \times \mathcal{Q}_2: \\ (Y_1, Y_2) \notin \mathcal{P}_1^* \times \mathcal{P}_2^*}} |Y_1||Y_2| \leq \sum_{X, X' \in \mathcal{P}} \delta_1 |X \cap V_1||X' \cap V_2| = \delta_1 \left(\sum_{X \in \mathcal{P}} |X \cap V_1| \right) \left(\sum_{X \in \mathcal{P}} |X \cap V_2| \right) = \delta_1 n^2.$$

8

Combining the last two inequalities, and using our choice of $\delta_1 = \frac{1}{2}\epsilon'$, we obtain (5) and thus complete the proof. $\qquad\square$

To deduce Theorem 1 from Theorem 2, we use the celebrated tower-type lower bound for Szemerédi's regularity lemma, originally proved by Gowers [4]. (See also [6] for a simpler proof.) For our application we need the bound to hold without assuming the partition is equitable.[3] As was shown by Fox, Lovász and Zhao (see Theorem 2.2 in [3], or Theorem 1.2 in [2]), this can be assumed without significant loss.

Theorem 3 ([3, 4]). *There is a bipartite graph whose every ϵ-regular partition is of order at least* $\mathrm{twr}(1/\mathrm{poly}(\epsilon))$.

Proof of Theorem 1. Let G_0 be the bipartite graph given by Theorem 3. Consider a balanced semi-blowup $G = H \odot_e G_0$ (where, implicitly, the blowup has sufficiently many vertices for the semi-blowup G to be well defined). Let $\mathcal{P}$ be an ϵ-(H, F)-regular partition of G, and let $\mathcal{Q}$ be the common refinement of $\mathcal{P}$ and $\{V_1, V_2\}$, where V_1, V_2 are the vertex classes of G_0. Apply Theorem 2 to deduce that $\mathcal{Q}$ is a $\mathrm{poly}(\epsilon k^k)$-regular partition of G_0. By Theorem 3, the order of $\mathcal{Q}$ is at least $\mathrm{twr}(1/\mathrm{poly}(\epsilon k^k))$, and we are done. $\qquad\square$

3 Concluding Remarks

We have shown that tower-type lower bounds hold for every generalized regularity lemma. One may raise the objection that, since a generalized regularity lemma can be meaningful also for sparse graphs (that is, of density tending to 0 with the number of vertices), it is not as satisfactory if our lower bound construction is dense. Indeed, the construction in Theorem 1, being a semi-blowup of H, is—for any H with at least two edges—of density at least $1/k$. However, it turns out that one can easily get arbitrarily sparse graph with the same lower bound, provided F is connected. For example, one may add any desired number of isolated vertices to the graph from Theorem 1 so as to reduce the density as needed, yet still maintain the property that any (H, F)-regular partition is large. Indeed, this is a consequence of the following observation.

Observation 3. *Suppose F is connected. Let a k-partite graph G be obtained from a k-partite graph G_0 by adding isolated vertices. If G is ϵ-(H, F)-regular then G_0 is also ϵ-(H, F)-regular.*

Proof. As F and H are connected, $n_F(G) = n_F(G_0)$ and $n_H(G) = n_H(G_0)$. Let G' be a k-partite induced subgraph of G_0 with $n_F(G') \geq \epsilon n_F(G_0)$. Then $n_F(G') \geq \epsilon n_F(G)$, and as G is ϵ-(H, F)-regular, $c_{H,F}(G') = c_{H,F}(G) \pm \epsilon$, that is, $c_{H,F}(G') = c_{H,F}(G_0) \pm \epsilon$, as needed. $\quad\square$

---

[3]In an equitable partition, all parts have the same size give or take 1.

References

[1] F. Chung, Regularity lemmas for clustering graphs, *Adv. in Appl. Math.*, to appear (http://www.math.ucsd.edu/~fan/wp/reg_cluster.pdf). 1, 2

[2] J. Fox and L.M. Lovász, A tight lower bound for Szemerédi's regularity lemma, *Combinatorica* **37** (2017), 911–951. 2.4

[3] J. Fox, L.M. Lovász and Y. Zhao, On regularity lemmas and their algorithmic applications, *Combin. Probab. Comput.* **26** (2017), 481-505. 2.4, 3

[4] T. Gowers, Lower bounds of tower type for Szemerédi's uniformity lemma, *Geom. Funct. Anal.* **7** (1997), 322–337. 1.1, 2.4, 3

[5] R. D. Luce and A. D. Perry, A method of matrix analysis of group structure, *Psychometrika* **14** (1949), 95–116. 1

[6] G. Moshkovitz and A. Shapira, A short proof of Gowers' lower bound for the regularity lemma, *Combinatorica* **36** (2016), 187–194. 2.4

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

[8] D. Watts and S. Strogatz, Collective dynamics of small-world networks, *Nature* **393** (1998), 440–442. 1