FINDING BLOWUPS ONE VERTEX AT A TIME

JACOB FOX, YUVAL WIGDERSON, AND YUNKUN ZHOU

Abstract. An influential theorem of Nikiforov states that if an N-vertex graph G contains at least γN^h copies of some fixed h-vertex graph H, then G contains an H-blowup of order $c_H(\gamma) \log N$. We provide a new proof of this theorem, which in particular improves the best known bound on the constant $c_H(\gamma)$. In contrast to previous proofs, our proof is iterative, finding the blowup one vertex at a time.

1. Introduction

Given an integer t and a graph H, its *t-blowup* $H[t]$ is the graph obtained from H by replacing every vertex by an independent set of order t, and every edge by a complete bipartite graph between the two corresponding independent sets. Blowups occupy a central role in extremal graph theory, and many of the field's central theorems can be phrased as statements about blowups, including the Erdős–Stone theorem [9] and the Kővári–Sós–Turán theorem [21], which states that if an N-vertex graph contains at least γN^2 copies of K_2 (i.e. edges), then it contains a copy of $K_2[t]$, for $t = \Omega_\gamma(\log N)$. The following remarkable and influential theorem of Nikiforov [26, 27] extends this to yield a large blowup of any graph.

Theorem 1.1. *Let H be an h-vertex graph and let $\gamma > 0$. If N is sufficiently large, then every N-vertex graph with at least γN^h copies of H contains an H-blowup $H[t]$, where*

$$t \geqslant c_H(\gamma) \log N,$$

and $c_H(\gamma) > 0$ is a constant depending only on γ and H.

By considering a random graph with edge density $\gamma^{1/e(H)}$, it is not hard to show that this result is best possible up to the value of $c_H(\gamma)$ for nonempty H. Nikiforov's theorem has several important consequences, among them a short proof of the Bollobás–Erdős–Simonovits theorem [2, 3].

Nikiforov's original proof of Theorem 1.1 is very short and elegant, and the technique he introduced to prove it has since found many other applications (e.g. [6, 30]). Nevertheless, Theorem 1.1 occupies a bit of an unusual space in the theory, because several natural variants of it, though easy to conjecture, appear extremely difficult to prove. Most notably, the extension of Theorem 1.1 to hypergraphs remains an outstanding open problem, whose resolution would have a number of interesting applications.

Because of this, there has been a great deal of interest in providing alternative proofs of Theorem 1.1, in the hopes that new perspectives would shed greater light on the problem, and hopefully lead to techniques that are more widely applicable. In particular, there is a lot of interest in understanding the quantitative aspects of Theorem 1.1, and specifically in determining the behavior $c_H(\gamma)$ as a function of γ (this question was perhaps first explicitly raised in [29, Problem 2]). We henceforth use the notation $c_H(\gamma)$ to denote the optimal constant in Theorem 1.1, that

JF: Department of Mathematics, Stanford University, Stanford, CA 94305. Email: `jacobfox@stanford.edu`.
Research supported by NSF awards DMS-2154129 and DMS-2452737.

YW: Institute for Theoretical Studies, ETH Zürich, 8006 Zürich, Switzerland. Email: `yuval.wigderson@eth-its.ethz.ch`.
Research supported by Dr. Max Rössler, the Walter Haefner Foundation, and the ETH Zürich Foundation.

YZ: Email: `yunkunzhou@alumni.stanford.edu`.

is, the supremum over all $c > 0$ such that every N-vertex graph with at least γN^h copies of H contains an H-blowup of size $(c - o(1)) \log N$, where the $o(1) \to 0$ as $N \to \infty$. The only known upper bound on $c_H(\gamma)$, from an elementary analysis of the random graph $G(n, \gamma^{1/e(H)})$, shows that $c_H(\gamma) = O_H(1/\log \frac{1}{\gamma})$. By contrast, Nikiforov's proof yields a polynomial lower bound, namely $c_H(\gamma) = \Omega_H(\gamma^h)$ for every[1] h-vertex graph H. This is a rather wide gap, and reflects our poor understanding of the problem. In particular, if the logarithmic upper bound does not give the true value of $c_H(\gamma)$, meaning that there is a better construction than a random graph, this may explain the difficulty of extending Theorem 1.1 to other settings, as the heuristics driving the various conjectural extensions of Nikiforov's theorem often rely on the intuition coming from random graphs and hypergraphs.

Motivated by the twin goals of finding new proofs of Theorem 1.1, and of improving the bounds on $c_H(\gamma)$, many researchers have studied this problem. The first improvement was due to Rödl and Schacht [28], who gave a new proof of Theorem 1.1, and in so doing proved that[2] $c_H(\gamma) \geqslant \gamma^{1+o(1)}$ for every graph H. While still polynomial, their bound is much better than Nikiforov's bound of γ^h, and in particular the exponent is independent of H. This bound was further slightly improved by Fox, Luo, and Wigderson [11], to $c_H(\gamma) \geqslant \gamma^{1-1/e(H)+o(1)}$, yielding a small power savings depending on H. Finally, Girão, Hunter, and Wigderson [14] recently proved that $c_H(\gamma) = \Theta_H(1/\log \frac{1}{\gamma})$ whenever H is triangle-free, thus completely resolving the problem in this case.

Our main result in the present paper is a further polynomial improvement for all graphs.

Theorem 1.2. *For every graph H, we have $c_H(\gamma) \geqslant \gamma^{1/2+o(1)}$.*

More precisely, if H has h vertices and N is sufficiently large with respect to γ, then every N-vertex graph with at least γN^h copies of H contains an H-blowup $H[t]$, where

$$t \geqslant \frac{\delta_H \sqrt{\gamma}}{(\log \frac{1}{\gamma})^2} \log N$$

and $\delta_H > 0$ is a constant depending only on H.

In particular, our result is the first to obtain a power savings over the Rödl–Schacht bound that is independent of H.

However, we believe that our main contribution is not the precise bound $c_H(\gamma) \geqslant \gamma^{1/2+o(1)}$, but rather the new technique we introduce, and we are hopeful that this technique will prove useful in related settings. We provide a detailed overview of our technique in Section 2, but let us briefly remark on the differences between our approach and previous ones. Like all of the previous proofs [11, 14, 26, 27, 28], we find the parts of the blowup $H[t]$ one at a time. However, in these earlier proofs, each blowup part is found in a "one-shot" way, by applying (a variant of) the Kővári–Sós–Turán theorem to some appropriate auxiliary bipartite graph. In our approach, by contrast, we find the t vertices of a part of $H[t]$ one vertex at a time, picking them out via a greedy algorithm. In order to analyze the evolution of this algorithm, we track an associated "energy" functional. This iterative approach bears some resemblance to the recent *book algorithm* used in the spectacular improvement to the upper bound on diagonal Ramsey numbers [4] (see also [1, 18] for variants of this technique, as well as the surveys [22, 32]), as well as to the technique used to prove Szemerédi's regularity lemma [31] and its many extensions. Because such iterative approaches have found applications in a wide variety of areas, including in hypergraphs, we are cautiously optimistic that our technique could also be adapted to these settings. These connections suggest many potential avenues for further research, and we speculate on some of these in Section 7.

[1]Nikiforov actually only claimed this bound [27] for $H = K_h$, and the weaker bound $c_H(\gamma) = \Omega_H(\gamma^{h^2})$ for a general h-vertex graph [26]. However, it is not hard to see that the result for cliques immediately implies the same bound for all graphs (see Proposition 6.4 below).

[2]Here and in what follows, a $o(1)$ in the exponent represents a quantity tending to 0 as $\gamma \to 0$.

We conclude with a few further remarks on the quantitative aspects of Theorem 1.2. First, by a more careful analysis of our proof, one of the logarithmic factors in the denominator can be removed; however, since there is no reason to expect the resulting bound to be optimal, we have chosen to present a slightly weaker result with a more transparent proof. Second, while we do not expect the bound $\gamma^{1/2+o(1)}$ to be optimal, it appears to be a hard limit of our technique; concretely, if G is a random N-vertex tripartite graph with bipartite edge densities $\frac{1}{2}, \sqrt{\gamma}, \sqrt{\gamma}$, respectively, then G contains $\Omega(\gamma N^3)$ triangles as well as a copy of $K_3[\Omega(\log N/\log \frac{1}{\gamma})]$, and yet we do not know how to use our technique to find a copy of $K_3[t]$ in G for any $t = \omega(\sqrt{\gamma}\log N)$.

The rest of this paper is organized as follows. As our proof is fairly technical, we give a detailed overview of it in Section 2. Section 3 contains some preliminary lemmas we use in our proof, especially the various technical computations related to our energy function $\mathcal{E}_R$. Section 4 rigorously defines and analyzes the basic iterative algorithm discussed above, and Section 5 does the same for the more refined algorithm involving switching. Using these tools, we present the proof of Theorem 1.2 in Section 6. We conclude in Section 7 with some concluding remarks.

2. Proof overview

We now discuss our proof of Theorem 1.2 in more detail. First, while Theorem 1.2 is stated for all graphs H, all of the difficulty actually lies in proving it in the case $H = K_3$. Indeed, there is a simple and well-known reduction (Proposition 6.4) that shows that it suffices to prove Theorem 1.2 for cliques; moreover, our main argument also allows us to reduce the problem for any clique to the case of the triangle. So for the moment we focus on $H = K_3$, and at the end of the proof sketch briefly discuss how the same ideas yield the reduction from all cliques to this case.

Let now G be a graph with γN^3 triangles. Our next reduction allows us to find within G three sets A, B, C, each of size $n = \Omega_\gamma(N)$, with the following properties: each edge in $E(A, B)$ lies in at least $\tau|C|$ triangles of G, for some $\tau \geqslant \gamma^{2/3}$, and moreover the number of edges between A and B is $p|A||B|$, for some $p \geqslant \gamma/\tau$. We find this structure by applying the regularity lemma[3] to G, and passing to a regular triple of parts A, B, C spanning at least $\gamma|A||B||C|$ triangles, which must exist by averaging. By the counting lemma, the number of triangles among these three parts is very close to $d(A, B)d(A, C)d(B, C)|A||B||C|$, where d denotes the edge density between two parts. Without loss of generality $p = d(A, B)$ is the smallest of the three edge densities, implying that $\tau = d(A, C)d(B, C) \geqslant \gamma^{2/3}$. We've now found that an *average* edge in $E(A, B)$ lies on at least $\tau|C|$ triangles, and that $p\tau = d(A, B)d(A, C)d(B, C) \geqslant \gamma$, so we are essentially done; the final step is to again use properties of regular pairs to argue that nearly all the edges have the average behavior, so upon deleting a few outliers we obtain the claimed structure.

We now turn to our main argument. Our goal is to pick out vertices $c_1, \ldots, c_s \in C$, and to restrict to their common neighborhoods $A^\dagger \subseteq A, B^\dagger \subseteq B$. If we can then find a complete bipartite subgraph $K_{t,t}$ between $A^\dagger$ and $B^\dagger$, we have found a complete tripartite graph $K_{s,t,t}$ together with $C^\dagger = \{c_1, \ldots, c_s\}$; it is for this reason that we insist that all vertices in $A^\dagger \cup B^\dagger$ are adjacent to all of $c_1, \ldots, c_s$. In particular, this yields a triangle blowup $K_3[\min\{s, t\}]$.

In order to find a copy of $K_{t,t}$ in $A^\dagger \cup B^\dagger$, we will apply the Kővári–Sós–Turán theorem [21], which states that every bipartite graph (X, Y) contains a complete bipartite subgraph with parts of size $\Omega(\log \min\{|X|, |Y|\}/\log(1/d(X, Y)))$; this essentially says that for $H = K_2$, we have the optimal dependence $c_{K_2}(\gamma) \geqslant \Omega(1/\log \frac{1}{\gamma})$ in Theorem 1.1.

Note that the Kővári–Sós–Turán theorem has a logarithmic dependence on both the sizes of the sets and on the edge density, which means that we are allowed to lose a great deal when passing from the pair (A, B) to the pair $(A^\dagger, B^\dagger)$. Concretely, it is sufficient for us if $|A^\dagger|, |B^\dagger| \geqslant \sqrt{n}$, and

[3] More precisely, we apply the cylinder regularity lemma of Duke, Lefmann, and Rödl [7], which gives us a much more reasonable "sufficiently large" condition on N.

if $d(A^\dagger, B^\dagger) \geqslant p^2$, where we recall that $|A| = |B| = n$ and $d(A, B) = p$. That is, we are allowed to lose by a polynomial amount on both the sizes of the sets and on the edge density. The reason is that

$$\frac{\log \sqrt{n}}{\log \frac{1}{p^2}} = \frac{1}{4} \cdot \frac{\log n}{\log \frac{1}{p}},$$

hence any polynomial loss yields an absolute constant factor loss in the final bound.

In order to ensure we obtain these bounds on $|A^\dagger|, |B^\dagger|$, and $d(A^\dagger, B^\dagger)$, it will be helpful to track a certain potential function, whose inputs are the three numbers $|A|, |B|, d(A, B)$. Whatever potential function we pick, we would like the effect of dividing $|A|$ or $|B|$ by $\sqrt{n}$ to be roughly the same as the effect of multiplying $d(A, B)$ by p, since we are allowing ourselves to lose a factor of $\sqrt{n}$ on the former, but only a factor of p on the latter. This suggests that a natural form of the potential function multiplies $|A||B|$ by a high power of $d(A, B)$, precisely so that losses in $d(A, B)$ will be amplified. Concretely, for a real number $R > 0$ we define the *R-energy* of the pair (A, B) to be

$$\mathcal{E}_R(A, B) = |A||B|d(A, B)^R.$$

Thus, dividing $|A|$ or $|B|$ by $\sqrt{n}$ decreases $\mathcal{E}_R(A, B)$ by a factor of $\sqrt{n}$, whereas multiplying $d(A, B)$ by a factor of p decreases $\mathcal{E}_R(A, B)$ by a factor of p^R. As we want these effects to be roughly the same, in our proof, we will end up selecting $R = \Theta(\log n / \log \frac{1}{p})$.

As discussed in the introduction, we will pick out the vertices $c_1, \ldots, c_s$ one at a time. So suppose that at some point in the process, we have picked out $c_1, \ldots, c_\ell$, and restricted A and B to their common neighborhoods, say $A^{(\ell)}, B^{(\ell)}$. Inductively, we have maintained that $A^{(\ell)}, B^{(\ell)}$, and $d(A^{(\ell)}, B^{(\ell)})$ are fairly large, namely at most polynomially smaller than their starting values. Our goal is now to select $c_{\ell+1}$ in such a way that we can keep the iteration going.

To do so, let us consider some $c^* \in C$, and let $A^+ = N(c^*) \cap A^{(\ell)}, B^+ = N(c^*) \cap B^{(\ell)}$. Our hope is to show that, by selecting c^* appropriately, we can ensure that $|A^+|, |B^+|$, and $d(A^+, B^+)$ remain fairly large, so that we can set $c_{\ell+1} = c^*, A^{(\ell+1)} = A^+, B^{(\ell+1)} = B^+$, and hence continue the iteration.

First, we note that controlling the size of A^+ and B^+ is not difficult. Indeed, since every edge in $E(A, B)$ lies in at least $\tau|C|$ triangles, we see that the triple $(A^{(\ell)}, B^{(\ell)}, C)$ still contains at least $\tau e(A^{(\ell)}, B^{(\ell)})|C|$ triangles, and hence we can select[4] $c^* \in C$ which lies in at least $\tau e(A^{(\ell)}, B^{(\ell)})$ triangles. But the number of triangles containing c^* is precisely $e(A^+, B^+)$, hence

$$|A^+||B^+| \geqslant e(A^+, B^+) \geqslant \tau e(A^{(\ell)}, B^{(\ell)}) = \tau p^{(\ell)}|A^{(\ell)}||B^{(\ell)}|,$$

where we set $p^{(\ell)} = d(A^{(\ell)}, B^{(\ell)})$. Recall that at the beginning of the process, we had $\tau p \geqslant \gamma$, and by our inductive assumption we have $p^{(\ell)} \geqslant p^2$, hence $|A^+||B^+| \geqslant \gamma^2|A^{(\ell)}||B^{(\ell)}|$. Note that losing a factor of γ^2 on the sizes of the sets at every step is completely fine for us: we will end up doing s steps, so the amount we will lose on the sizes is γ^{2s}. So as long as $s \leqslant \log n/(10 \log \frac{1}{\gamma})$, say, the total amount that we will lose on the sizes of A and B is at most a small power of n, which is exactly what we are targeting. In particular, if this were the only constraint on our process, we could keep it going until $s = \Theta(\log n / \log \frac{1}{\gamma})$, thus obtaining the optimal bound $c_{K_3}(\gamma) = \Theta(1/\log \frac{1}{\gamma})$. Unsurprisingly, therefore, this is not the only, or even the main, constraint on our process.

Indeed, as discussed above, and as recorded in the form of our energy function $\mathcal{E}_R$, we are willing to tolerate much larger losses on the sizes of $|A^+|, |B^+|$ than we are on the density $d(A^+, B^+)$. Specifically, let us pick some $\alpha > 0$, and say that we are happy if we can ensure that $d(A^+, B^+) \geqslant p^{(\ell)} - \alpha$. Since we will eventually do s such steps, we will choose $\alpha \leqslant p/(2s)$, so that at the end of

[4] In reality, we also need to ensure that c^* is a new vertex, i.e. we must pick $c^* \in C \setminus \{c_1, \ldots, c_\ell\}$. But as $\ell \leqslant s = O_\gamma(\log n) \ll n$, this restriction has almost no impact, and we ignore it for this overview.

the day we will still have $d(A^\dagger, B^\dagger) \geqslant p - \alpha s \geqslant p/2$, which is even stronger than the lower bound of p^2 we were originally targeting[5].

As we are happy if $d(A^+, B^+) \geqslant p^{(\ell)} - \alpha$, let us assume contrariwise that $d(A^+, B^+) < p^{(\ell)} - \alpha$. In this case, we use a standard but crucial trick, which goes back at least to [8]: we have found a large pair of subsets $A^+ \subseteq A, B^+ \subseteq B$ which span substantially fewer edges than what you would expect based on the global average, and hence these "missing edges" must be somewhere else. More precisely, setting $A^- = A^{(\ell)} \setminus A^+, B^- = B^{(\ell)} \setminus B^+$, we learn that one of the three pairs $(A^-, B^-), (A^+, B^-), (A^-, B^+)$ must have substantially *higher* edge density than $p^{(\ell)}$. Specifically, it is not hard to see that for some pair $(X, Y) \in \{(A^-, B^-), (A^+, B^-), (A^-, B^+)\}$, we have

$$d(X, Y) \geqslant p^{(\ell)} + \frac{\alpha |A^+||B^+|}{3|X||Y|} \tag{1}$$

Because our energy functional $\mathcal{E}_R$ is highly sensitive to changes in the edge density, an elementary but tedious computation in fact implies that for one of these three pairs (X, Y), we actually have $\mathcal{E}_R(X, Y) > \mathcal{E}_R(A^{(\ell)}, B^{(\ell)})$, so long as $R \geqslant 10p/(\alpha\tau)$. That is, in the situation where we are not happy because $d(A^+, B^+) < p^{(\ell)} - \alpha$, we have the ability to restrict $A^{(\ell)}, B^{(\ell)}$ to subsets and to increase the energy. Of course, such steps are not directly useful to us because we are not able to select a new vertex $c_{\ell+1}$, but in return we obtain an energy increment.

Thus, we end up with two kinds of steps: steps where we select a new vertex $c_{\ell+1}$ and restrict to its neighborhoods, at the expense of losing at most α on the density, and steps where we do not select a new vertex but can restrict to subsets and obtain an energy increment. In the formal analysis of the algorithm, it will be more convenient to never actually perform the second kind of step, and instead, whenever we create a new pair $(A^{(\ell)}, B^{(\ell)})$, we immediately restrict these two sets to subsets which maximize the energy $\mathcal{E}_R$. By doing this, we automatically know that no energy increment is possible, hence the second scenario above can never arise, and the vertex c^* we pick necessarily satisfies $d(A^+, B^+) \geqslant p^{(\ell)} - \alpha$.

It only remains to see for how many steps s we can keep this process going. We also have a free choice to make, namely the choice of the parameter α, subject to the constraints discussed above, namely $\alpha \leqslant p/(2s)$ and $R \geqslant 10p/(\alpha\tau)$, which imply $s \leqslant p/(2\alpha) \leqslant R\tau/20$. Recall too that we require $R = \Theta(\log n / \log \frac{1}{p})$, in order to be able to conclude at the end that $|A^\dagger|, |B^\dagger|, d(A^\dagger, B^\dagger)$ are still polynomially large relative to their starting values, so that we can apply the Kővári–Sós–Turán theorem. This implies that we can keep the process going until step

$$s = \Theta(R\tau) = \Theta\left(\frac{\tau \log n}{\log \frac{1}{p}}\right) = \Omega\left(\frac{\gamma^{2/3} \log n}{\log \frac{1}{\gamma}}\right),$$

where we recall that $\tau \geqslant \gamma^{2/3}$ and $p \geqslant \gamma/\tau \geqslant \gamma$. Moreover, at the end of the process, we apply the Kővári–Sós–Turán theorem to $(A^\dagger, B^\dagger)$, finally using our guarantee that $|A^\dagger| \geqslant \sqrt{n}, |B^\dagger| \geqslant \sqrt{n}, d(A^\dagger, B^\dagger) \geqslant p^2$, to find a copy of $K_{t,t}$ for

$$t = \Omega\left(\frac{\log \sqrt{n}}{\log \frac{1}{p^2}}\right) = \Omega\left(\frac{\log n}{\log \frac{1}{\gamma}}\right).$$

In particular, t is much larger than s, and we conclude that G contains a triangle blowup $K_3[s]$, proving that $c_{K_3}(\gamma) \geqslant \gamma^{2/3+o(1)}$.

Of course, our goal is to actually prove that $c_{K_3}(\gamma) \geqslant \gamma^{1/2+o(1)}$, and we now turn to describe the remaining ingredient necessary to achieve this. At a high level, the new ingredient is to exploit

[5] As mentioned in the introduction, we can remove one of the logarithmic factors in Theorem 1.2 via a more careful analysis, and the slack in this step is the reason: to obtain this stronger result, we would pick a slightly larger value of α, so that $p - \alpha s = p^2$.

the symmetry between the three parts: the argument above treated C completely differently from A, B, and there are situations where this becomes wasteful.

More precisely, in the argument above, the weakness is that t is much larger than s, by a factor of roughly $1/\tau$. Thus, if we could somehow pass to a configuration with a larger value of τ, we would obtain a larger value of s, and hence a better lower bound on $c_{K_3}(\gamma)$.

How could we achieve this? Note that at every step of the process, the total number of triangles present is at least $e(A^{(\ell)}, B^{(\ell)}) \cdot (\tau|C|) \geqslant (p\tau/2)|A^{(\ell)}||B^{(\ell)}||C| \geqslant (\gamma/2)|A^{(\ell)}||B^{(\ell)}||C|$, recalling that we have maintained $p^{(\ell)} = d(A^{(\ell)}, B^{(\ell)}) \geqslant p/2$. Consequently, the average number of triangles that an edge of $E(A^{(\ell)}, C)$ lies in is at least

$$\frac{(\gamma/2)|A^{(\ell)}||B^{(\ell)}||C|}{e(A^{(\ell)}, C)} \geqslant \frac{\gamma/2}{d(A^{(\ell)}, C)}|B^{(\ell)}|.$$

In particular, if it ever happens that $\widehat{\tau} := \gamma/d(A^{(\ell)}, C) \gg \tau$, then we have essentially found the configuration of the type discussed above, where the value of τ has been replaced by the much larger value $\widehat{\tau}/2$. More precisely, letting $\widehat{A} = A^{(\ell)}, \widehat{B} = C, \widehat{C} = B^{(\ell)}$, we see that the triple $(\widehat{A}, \widehat{B}, \widehat{C})$ contains at least $(\gamma/2)|\widehat{A}||\widehat{B}||\widehat{C}|$ triangles, and that an average edge in $E(\widehat{A}, \widehat{B})$ lies in at least $(\widehat{\tau}/2)|\widehat{C}|$ triangles. This is not, per se, the setup we had above, since we only have a bound on the *average* number of triangles containing an edge in $E(\widehat{A}, \widehat{B})$, but by deleting the edges lying in too few triangles, we can correct this issue. By doing this and then running the argument above, we end up finding a triangle blowup $K_3[\widehat{s}]$, where

$$\widehat{s} = \Omega\left(\frac{\widehat{\tau}\log n}{\log \frac{1}{\gamma}}\right).$$

Let us pick some parameter $\mu \gg \tau$, which we will optimize later, and declare that we will do such a switch if ever $\widehat{\tau} > \mu$. In particular, if this ever happens, then we find a triangle blowup of size $\widehat{s} = \Omega(\mu \log n/\log \frac{1}{\gamma})$.

We may thus assume that we never perform such a switch, and hence that at every stage of the process we have $\widehat{\tau} \leqslant \mu$, or equivalently $d(A^{(\ell)}, C) \geqslant \gamma/\mu$. By the same argument but with the roles of $A^{(\ell)}$ and $B^{(\ell)}$ interchanged, we may also assume that $d(B^{(\ell)}, C) \geqslant \gamma/\mu$. This implies that the average degree of a vertex $c^* \in C$ in $A^{(\ell)}$ is at least $(\gamma/\mu)|A^{(\ell)}|$, and similarly for the degree into $B^{(\ell)}$. By another round of removing vertices whose degree is much lower than the average, this means that in the argument presented above, we may assume that $|A^+| \geqslant (\gamma/\mu)|A^{(\ell)}|, |B^+| \geqslant (\gamma/\mu)|B^{(\ell)}|$.

The reason this is useful to us is that in (1), the amount we gain on the density of the pair (X, Y) is proportional to $|A^+||B^+|$, and hence we are gaining more here. This, in turn, means that we can find $\mathcal{E}_R(X, Y) > \mathcal{E}_R(A^{(\ell)}, B^{(\ell)})$ under a more modest assumption on R; it now suffices that $R \gtrsim \mu/(\alpha\tau)$, where the notation $\gtrsim$ hides logarithmic factors. Recall that earlier we needed $R \geqslant 10p/(\alpha\tau)$, and hence we now obtain a milder requirement, so long as $\mu \ll p$.

It now remains to do the optimization; we again require $\alpha \leqslant p/(2s)$, so we conclude that we can continue the process up to step

$$s \leqslant \frac{p}{2\alpha} \lesssim \frac{p\tau R}{\mu} = \widetilde{\Theta}\left(\frac{\gamma}{\mu}\log n\right),$$

where we hide factors of $\log \frac{1}{\gamma}$ in the $\widetilde{\Theta}$ notation, and recall that we choose $R = \widetilde{\Theta}(\log n)$. In order to have the strongest possible bound, we should choose μ so that s and $\widehat{s}$ are of roughly the same order, meaning that we should take $\gamma/\mu = \mu$, or equivalently $\mu = \sqrt{\gamma}$. By doing so, we conclude that $c_{K_3}(\gamma) \geqslant \gamma^{1/2+o(1)}$.

Note that this argument is useful so long as $\tau \ll \mu \ll p$; we also know that $p\tau = \Omega(\gamma)$. So we can only win if $\tau \ll p$, or equivalently if $p \gg \sqrt{\gamma}$. In the regime where $p \asymp \tau \asymp \sqrt{\gamma}$, this switching argument wins us nothing. This is precisely why, in the introduction, we said that our technique does not seem able to cope with a random tripartite graph of edge densities $\frac{1}{2}, \sqrt{\gamma}, \sqrt{\gamma}$; in such a graph, no matter which of the three parts we select as C, we will have $p \geqslant \sqrt{\gamma}$.

To conclude, we briefly explain how we obtain the same result for $c_{K_{k+1}}(\gamma)$, using essentially the same argument. So let G be an N-vertex graph with γN^{k+1} copies of K_{k+1}. First, by applying the same regularity technique, we can pass to a collection of disjoint sets $A_1, \ldots, A_k, C \subseteq V(G)$, each of size $n = \Omega_{\gamma,k}(N)$, with the property that there are at least $p|A_1| \cdots |A_k|$ copies of K_k among $(A_1, \ldots, A_k)$, each of which extends to at least $\tau|C|$ copies of K_{k+1} with a vertex in C, where $\tau \geqslant \gamma^{2/(k+1)}$ and $p\tau \geqslant \gamma$. We now run exactly the same iterative argument, selecting vertices from C one by one and restricting to their common neighborhoods in $A_1, \ldots, A_k$. The analysis of this algorithm is nearly identical to the one above, and the upshot of it is that we can pass to some $C^\dagger \subseteq C, A_i^\dagger \subseteq A_i$, with the following properties: we have that $|A_i^\dagger| = \sqrt{n}$ for all i, that $|C^\dagger| = s = \Theta(\tau \log n / \log \frac{1}{p})$, and that the density of K_k-copies among $A_1^\dagger, \ldots, A_k^\dagger$ is at least $p/2$. We now note that the induced subgraph on $A_1^\dagger \cup \cdots \cup A_k^\dagger$ has $N' = k\sqrt{n} = \Omega(\sqrt{N})$ vertices and at least $\Omega(\gamma(N')^k)$ copies of K_k, so we may proceed by induction on k, with the base case $k = 2$ being precisely the triangle case handled above. The key thing to note is that for $k \geqslant 3$, we have $k + 1 \geqslant 4$, and hence

$$s = \Omega\left(\frac{\gamma^{2/(k+1)} \log n}{\log \frac{1}{\gamma}}\right) \geqslant \Omega\left(\frac{\gamma^{1/2} \log n}{\log \frac{1}{\gamma}}\right),$$

meaning that the number of vertices we select from C is indeed enough. In fact, the same argument yields the recursive bound

$$c_{K_{k+1}}(\gamma) \geqslant \min\{\gamma^{2/(k+1)+o(1)}, c_{K_k}(\Omega_k(\gamma))\}.$$

3. Preliminaries

We will use the Kővári–Sós–Turán theorem [21] in the following form. For a bipartite graph with parts (A, B), we denote by $d(A, B) = e(A, B)/(|A||B|)$ its edge density.

Theorem 3.1. *Let $p \in (0, \frac{1}{2}]$, let G be a bipartite graph with parts (X, Y), and assume that $d(X, Y) \geqslant p$. Then G contains a copy of $K_{t,t}$, where*

$$t = \left\lfloor \frac{\log(\min\{|X|, |Y|\})}{2 \log \frac{1}{p}} \right\rfloor.$$

3.1. Analytic inequalities.
In this paper, all logarithms are to base 2.

We need the following lemma, a simple analytic inequality.

Lemma 3.2. *For all real numbers $y \geqslant 0, R \geqslant 1$, we have*

$$y^{1/R} \leqslant 1 + \frac{y}{R}.$$

Proof. Recall Bernoulli's inequality, which states that $(1 + x)^R \geqslant 1 + Rx$ for all $R \geqslant 1$ and $x \geqslant 0$. Setting $x = y/R$, we find that

$$\left(1 + \frac{y}{R}\right)^R = (1 + x)^R \geqslant 1 + Rx = 1 + y \geqslant y,$$

which yields the claimed bound by taking $1/R$ powers of both sides. □

The following is a variant of Lemma 3.2, giving us a better upper bound if we know that y is either very small or very large.

Lemma 3.3. *Let $\eta \in (0, \frac{1}{4}]$ and let $R \geqslant \log \frac{1}{\eta}$. For every $y \in [1, (1-\eta)^{-2}] \cup [\frac{1}{\eta}, \infty)$, we have*

$$y^{1/R} \leqslant 1 + (\eta \log \tfrac{1}{\eta}) \frac{y}{R}.$$

Proof. First suppose that $y \in [1, (1-\eta)^{-2}]$. By applying Bernoulli's inequality to $x = y(\eta \log \frac{1}{\eta})/R$, we find

$$\left(1 + (\eta \log \tfrac{1}{\eta}) \frac{y}{R}\right)^R \geqslant 1 + (\eta \log \tfrac{1}{\eta}) y \geqslant y \left((1-\eta)^2 + \eta \log \tfrac{1}{\eta}\right),$$

where the second inequality uses our assumption $y \leqslant (1-\eta)^{-2}$. Now note that

$$(1-\eta)^2 + \eta \log \tfrac{1}{\eta} = 1 + \eta(-2 + \eta + \log \tfrac{1}{\eta}) \geqslant 1 + \eta(-2+2) = 1,$$

using our assumptions that $\eta > 0$ and $\eta \leqslant \frac{1}{4}$, the latter of which implies $\log \frac{1}{\eta} \geqslant 2$. This completes the proof in the first case, upon taking the $1/R$ power of both sides.

Now suppose that $y \geqslant 1/\eta$. Let $\theta = (\log \frac{1}{\eta})/R \leqslant 1$ and $z = \eta y \geqslant 1$. We now have that

$$\left(1 + (\eta \log \tfrac{1}{\eta}) \frac{y}{R}\right)^{1/\theta} = (1 + \theta z)^{1/\theta} \geqslant 1 + z \geqslant 2\sqrt{z} \geqslant 2 z^{1/\log \frac{1}{\eta}} = \left(\frac{z}{\eta}\right)^{1/\log \frac{1}{\eta}} = y^{1/\log \frac{1}{\eta}},$$

where the first inequality is Bernoulli's, the second is AM–GM, and the third uses that $z \geqslant 1$ and $\log \frac{1}{\eta} \geqslant 2$. Taking the θ power of both sides gives the claimed result. $\square$

3.2. **Energy and strictly balanced subgraphs.** For a k-partite k-graph $(A_1, \ldots, A_k)$, we denote by $e(A_1, \ldots, A_k)$ its edge count and by $d(A_1, \ldots, A_k) = e(A_1, \ldots, A_k)/\prod_i |A_i|$ its edge density. We will frequently use the notation $\mathcal{A}$ as a shorthand for the tuple $(A_1, \ldots, A_k)$, to minimize the number of symbols. Similarly, we will shorten tuples such as $(A_1', \ldots, A_k')$ or $(A_1^*, \ldots, A_k^*)$ as $\mathcal{A}', \mathcal{A}^*$, respectively.

For a real number $R > 1$, we also define the *R-energy* of $\mathcal{A} = (A_1, \ldots, A_k)$ to be

$$\mathcal{E}_R(\mathcal{A}) \coloneqq |A_1| \cdots |A_k| d(\mathcal{A})^R = e(\mathcal{A}) d(\mathcal{A})^{R-1},$$

with the convention that $\mathcal{E}_R(\mathcal{A}) = 0$ if $\mathcal{A}$ has no edges (and in particular if some A_i is empty). Finally, we say that $\mathcal{A}$ is *strictly R-balanced* if for all subsets $A_i' \subseteq A_i$ such that at least one inclusion is proper, we have

$$\mathcal{E}_R(\mathcal{A}) > \mathcal{E}_R(\mathcal{A}').$$

We have the following extremely simple lemma.

Lemma 3.4. *Every k-partite k-graph $\mathcal{A}$ contains a strictly R-balanced subgraph $\mathcal{A}^*$ with $\mathcal{E}_R(\mathcal{A}^*) \geqslant \mathcal{E}_R(\mathcal{A})$.*

Proof. Let $A_1^* \subseteq A_1, \ldots, A_k^* \subseteq A_k$ be subsets which maximize $\mathcal{E}_R(\mathcal{A}^*)$ among all options; if there are multiple choices achieving the same value, pick one which minimizes $\sum_i |A_i^*|$. As $\mathcal{A}$ is one valid choice, we certainly have $\mathcal{E}_R(\mathcal{A}^*) \geqslant \mathcal{E}_R(\mathcal{A})$. Moreover, any subsets $A_i' \subseteq A_i^*$ such that at least one inclusion is proper would be another valid choice with a strictly smaller value of $\sum_i |A_i'|$, hence we conclude that $\mathcal{A}^*$ is strictly R-balanced. $\square$

We remark that this notion is similar to ones used in many other contexts in extremal graph theory, e.g. in finding a sublinear expander or in almost-regularizing a graph. In such cases, one picks an appropriate functional of a graph, which is usually some power of its edge count divided by its vertex count, and then passes to a subgraph which maximizes this quantity.

The main property we will need about strictly R-balanced hypergraphs is the following lemma, which gives an upper bound on the number of edges of any subgraph of a strictly balanced hypergraph. The proof is quite simple: if a subgraph had too many edges, its energy would be higher, contradicting the assumption of strict R-balancedness.

Lemma 3.5. *Let $\mathcal{A}$ be strictly R-balanced, let $p = d(\mathcal{A})$, and let $X_i \subseteq A_i$. Let $\mathcal{X} = (X_1, \dots, X_k)$.*

(a) We have

$$e(\mathcal{X}) \leqslant p \prod_i |X_i| + \frac{p}{R} \prod_i |A_i|.$$

(b) Let $\eta \in (0, \frac{1}{4}]$ and suppose that $R \geqslant \log \frac{1}{\eta}$. If, furthermore, we have that $\frac{\prod |X_i|}{\prod |A_i|} \in [0, \eta] \cup [(1-\eta)^2, 1]$, then we have the stronger bound

$$e(\mathcal{X}) \leqslant p \prod_i |X_i| + (\eta \log \tfrac{1}{\eta}) \frac{p}{R} \prod_i |A_i|.$$

Proof. There is nothing to prove if $p = 0$, if some X_i is empty, or if $X_i = A_i$ for all i, hence we may assume that $p > 0$ and that at least one inclusion is proper. This implies that $\mathcal{E}_R(\mathcal{X}) < \mathcal{E}_R(\mathcal{A})$. Unpacking the definition, this implies that

$$\left(\frac{d(\mathcal{X})}{d(\mathcal{A})} \right)^R < \frac{\prod_i |A_i|}{\prod_i |X_i|}. \tag{2}$$

Applying Lemma 3.2, we conclude that

$$\frac{d(\mathcal{X})}{p} < \left(\frac{\prod_i |A_i|}{\prod_i |X_i|} \right)^{1/R} \leqslant 1 + \frac{\prod_i |A_i|}{R \prod_i |X_i|}$$

and therefore

$$e(\mathcal{X}) = d(\mathcal{X}) \prod_i |X_i| \leqslant \left(p + \frac{p \prod_i |A_i|}{R \prod_i |X_i|} \right) \prod_i |X_i| = p \prod_i |X_i| + \frac{p}{R} \prod_i |A_i|,$$

proving Lemma 3.5(a).

Now suppose we are in the setting of Lemma 3.5(b). There is nothing to prove if some X_i is empty, so we may assume $\prod_i |X_i| \neq 0$, and then define $y = (\prod_i |A_i|)/(\prod_i |X_i|)$. By assumption we have $1/y \in [0, \eta] \cup [(1-\eta)^2, 1]$, implying that either $y \geqslant 1/\eta$ or $1 \leqslant y \leqslant (1-\eta)^{-2}$. Hence we are in the setting of Lemma 3.3; applying this lemma to (2), we conclude that

$$\frac{d(\mathcal{X})}{p} \leqslant y^{1/R} \leqslant 1 + (\eta \log \tfrac{1}{\eta}) \frac{y}{R} = 1 + \frac{\eta \log \frac{1}{\eta}}{R} \cdot \frac{\prod_i |A_i|}{\prod_i |X_i|},$$

from which the result follows as above. ∎

4. The basic iteration

We now introduce some additional terminology. A *k-mixed graph* $\mathcal{G}$ is a (non-uniform) hypergraph with $k+1$ parts, labeled $A_1, \dots, A_k, C$. Its edges consist of some k-partite k-graph on $\mathcal{A} = (A_1, \dots, A_k)$, as well as some 2-uniform bipartite graph on $(C, A_1 \cup \cdots \cup A_k)$.

We denote by $N_1(c), \dots, N_k(c)$ the neighborhoods of a vertex $c \in C$ in the sets $A_1, \dots, A_k$, respectively. We say that a vertex $c \in C$ *extends* an edge $(a_1, \dots, a_k) \in E(\mathcal{A})$ if c is adjacent to all the vertices $a_1, \dots, a_k$. We denote by $\operatorname{ext}(\mathcal{G})$ the total number of extensions in $\mathcal{G}$, namely

$$\operatorname{ext}(\mathcal{G}) = \sum_{(a_1, \dots, a_k) \in E(\mathcal{A})} \sum_{c \in C} \mathbf{1}_{c \text{ extends } (a_1, \dots, a_k)} = \sum_{c \in C} e(N_1(c), \dots, N_k(c)). \tag{3}$$

We also denote by $\tau_C(\mathcal{G})$ the *minimum extension density* of $\mathcal{G}$, defined as

$$\tau_C(\mathcal{G}) = \min \left\{ \frac{|\{c \in C : c \text{ extends } (a_1, \dots, a_k)\}|}{|C|} : (a_1, \dots, a_k) \in E(\mathcal{A}) \right\},$$

with the convention that $\tau_C(\mathcal{G}) = 0$ if $E(\mathcal{A}) = \varnothing$. That is, for each edge in $E(\mathcal{A})$, we compute what fraction of vertices in C extend this edge, and then $\tau_C(G)$ is the minimum of these ratios. Note that $\tau_C(\mathcal{G}) \in [0,1]$ for every k-mixed graph $\mathcal{G}$. The following is the key lemma of our iterative procedure.

Lemma 4.1. *Let $k \geqslant 2$ be an integer. Fix parameters $\tau \in (0,1]$ and $R \geqslant 2^{k+1}/\tau$. Let $\mathcal{G}$ be a k-mixed graph with parts $\mathcal{A}, C$, and suppose that $\tau_C(\mathcal{G}) \geqslant \tau$ and $\mathcal{A}$ is strictly R-balanced.*

Then there exist a vertex $c^* \in C$ and non-empty subsets $A_i^* \subseteq A_i$ satisfying the following properties.

(i) c^* is adjacent to all vertices in $A_1^* \cup \cdots \cup A_k^*$.

(ii) We have

$$\mathcal{E}_R(\mathcal{A}^*) \geqslant 2^{-2^{k+2}/\tau} \cdot \mathcal{E}_R(\mathcal{A})$$

(iii) $\mathcal{A}^*$ is strictly R-balanced.

Proof. Let us set $p = d(\mathcal{A})$ and $\alpha = 2^k p/(R\tau)$, and note that $\alpha \leqslant p/2$ since $R\tau \geqslant 2^{k+1}$. Note that

$$\mathrm{ext}(\mathcal{G}) \geqslant \tau_C(\mathcal{G}) e(\mathcal{A})|C| \geqslant \tau p |A_1| \cdots |A_k||C|,$$

and hence from (3) we have that there is some vertex $c^* \in C$ for which

$$e(N_1(c^*), \ldots, N_k(c^*)) \geqslant \tau p \prod_i |A_i|.$$

We fix such a c^*, and let $A_i^+ = N_i(c^*)$ and $A_i^- = A_i \setminus A_i^+$.

Our key claim is that

$$d(\mathcal{A}^+) \geqslant p - \alpha. \tag{4}$$

Suppose for contradiction that $d(\mathcal{A}^+) < p - \alpha \leqslant p$. From the above, we find that

$$p\tau \prod_i |A_i| \leqslant e(\mathcal{A}^+) = d(\mathcal{A}^+) \prod_i |A_i^+| < p \prod_i |A_i^+|,$$

implying that

$$\frac{1}{\tau} \prod_i |A_i^+| > \prod_i |A_i|. \tag{5}$$

Let $\sigma \in \{+,-\}^k$. We now apply Lemma 3.5(a) to $(X_1, \ldots, X_k) = (A_1^{\sigma_1}, \ldots, A_k^{\sigma_k}) =: \mathcal{A}^\sigma$ to find that

$$e(\mathcal{A}^\sigma) \leqslant p \prod_i |A_i^{\sigma_i}| + \frac{p}{R} \prod_i |A_i| < p \prod_i |A_i^{\sigma_i}| + \frac{p}{R\tau} \prod_i |A_i^+|,$$

where the second inequality is from (5). Moreover, in the case $\sigma = (+, \ldots, +)$, we actually have the stronger bound

$$e(\mathcal{A}^+) < (p - \alpha) \prod_i |A_i^+| = p \prod_i |A_i^+| - \alpha \prod_i |A_i^+|,$$

using our assumption that (4) is false. Summing these 2^k inequalities, and using the fact that each edge in $\mathcal{A}$ lies in precisely one tuple $\mathcal{A}^\sigma$, we find that

$$\begin{aligned}
e(\mathcal{A}) &= \sum_{\sigma \in \{+,-\}^k} e(\mathcal{A}^\sigma) \\
&< \sum_{\sigma \in \{+,-\}^k} p \prod_i |A_i^{\sigma_i}| + \left((2^k - 1)\frac{p}{R\tau} - \alpha\right) \prod_i |A_i^+| \\
&< p \sum_{\sigma \in \{+,-\}^k} \prod_i |A_i^{\sigma_i}| = p \prod_i |A_i|,
\end{aligned}$$

using our choice of α in the second inequality. This contradicts our assumption that $p = d(\mathcal{A})$, and concludes the proof of (4).

We now show how to use (4) to complete the proof of the lemma. To do this, we first compute that

$$\begin{aligned}\mathcal{E}_R(\mathcal{A}^+) &= e(\mathcal{A}^+)d(\mathcal{A}^+)^{R-1} \\ &\geqslant \left(\tau p \prod_i |A_i|\right)(p-\alpha)^{R-1} \\ &= \left(p^R \prod_i |A_i|\right)\left(\tau p^{1-R}(p-\alpha)^{R-1}\right) \\ &= \mathcal{E}_R(\mathcal{A})\left(\tau\left(1-\frac{\alpha}{p}\right)^{R-1}\right).\end{aligned}$$

Recall that $\tau \leqslant 1$, hence the inequality $\tau \geqslant 2^{-1/\tau}$ holds. Additionally, since $\alpha/p \leqslant 1/2$, the inequality $1-\alpha/p \geqslant 2^{-2\alpha/p}$ holds. Therefore,

$$\tau\left(1-\frac{\alpha}{p}\right)^{R-1} \geqslant \tau\left(1-\frac{\alpha}{p}\right)^{R} \geqslant 2^{-1/\tau}\cdot 2^{-2R\alpha/p} \geqslant 2^{-2^{k+2}/\tau}.$$

Finally, we pass to a strictly R-balanced subgraph $(A_1^*,\dots,A_k^*)$ of $(A_1^+,\dots,A_k^+)$, as given to us by Lemma 3.4. This automatically gives us (iii), as well as (ii) by the computation above. As c^* is adjacent to every vertex of $A_i^+ \supseteq A_i^*$, we also have (i). $\square$

We remark that the proof of Lemma 4.1 would work verbatim if we only knew that the *average* number of extensions of an edge in $E(\mathcal{A})$ were at least $\tau|C|$, rather than the *minimum*. However, our next result is obtained by iterating Lemma 4.1, and controlling the minimum is much simpler than controlling the average during such an iteration.

Proposition 4.2. *Let $k \geqslant 2$ be an integer, and fix $\tau_0, p \in (0, \frac{1}{2}]$. Let $n \geqslant p^{-10^k/\tau_0}$, and let $\mathcal{G}$ be a k-mixed graph with parts $\mathcal{A}, C$, where $\min\{|A_1|,\dots,|A_k|,|C|\} \geqslant n$. If $\tau_C(\mathcal{G}) \geqslant \tau_0$ and $d(\mathcal{A}) \geqslant p$, then there are subsets $A_i^\dagger \subseteq A_i, C^\dagger \subseteq C$ satisfying*

(a) $|C^\dagger| = \lfloor(\tau_0 \log n)/(2^{k+5}\log\frac{1}{p})\rfloor$,
(b) $|A_i^\dagger| \geqslant \sqrt{n}$ for every i,
(c) $d(\mathcal{A}^\dagger) \geqslant p/2$, and
(d) Every vertex of $C^\dagger$ is adjacent to every vertex of $A_1^\dagger \cup \dots \cup A_k^\dagger$.

Proof. Let

$$s = \left\lfloor \frac{\tau_0 \log n}{2^{k+5}\log\frac{1}{p}} \right\rfloor \qquad \text{and} \qquad R = \frac{\log n}{4\log\frac{1}{p}}.$$

We will define a nested sequence of induced subgraphs $\mathcal{G} \supseteq \mathcal{G}^{(0)} \supseteq \mathcal{G}^{(1)} \supseteq \dots \supseteq \mathcal{G}^{(s)}$ on vertex sets $(A_1^{(\ell)},\dots,A_k^{(\ell)},C^{(\ell)})_{\ell=0}^s$, satisfying the following properties.

(i) $|C^{(\ell)}| = |C| - \ell$.
(ii) Every vertex in $C \setminus C^{(\ell)}$ is adjacent to all vertices in $A_1^{(\ell)} \cup \dots \cup A_k^{(\ell)}$.
(iii) $\mathcal{A}^{(\ell)}$ is strictly R-balanced and satisfies

$$\mathcal{E}_R(\mathcal{A}^{(\ell)}) \geqslant 2^{-2^{k+3}\ell/\tau_0} p^R \prod_i |A_i|.$$

To begin, we set $C^{(0)} = C$, so that properties (i) and (ii) are trivially satisfied. We also let $\mathcal{A}^{(0)}$ be a strictly R-balanced subgraph of $\mathcal{A}$, as given by Lemma 3.4, for which we have

$$\mathcal{E}_R(\mathcal{A}^{(0)}) \geqslant \mathcal{E}_R(\mathcal{A}) = d(\mathcal{A})^R \prod_i |A_i| \geqslant p^R \prod_i |A_i|,$$

so (iii) is also satisfied for $\ell = 0$.

Inductively, suppose we have defined $\mathcal{G}^{(\ell)}$ for some $\ell < s$, and aim to define $\mathcal{G}^{(\ell+1)}$. Note that $s \leqslant \tau_0 n/2$, hence $|C \setminus C^{(\ell)}| = \ell < \tau_0 n/2 \leqslant \tau_0 |C|/2$ by (i). As a consequence, every edge in $A^{(\ell)}$ has at least $\tau_0|C|/2$ extensions to vertices in $C^{(\ell)}$. That is, $\tau_{C^{(\ell)}}(\mathcal{G}^{(\ell)}) \geqslant \tau_0/2$.

We are thus in a position to apply Lemma 4.1 to $\mathcal{G}^{(\ell)}$, with parameters $\tau = \tau_0/2$ and $R = (\log n)/(4 \log \frac{1}{p})$, as defined above. In order to apply the lemma, we must check that $R \geqslant 2^{k+1}/\tau = 2^{k+2}/\tau_0$, which indeed holds by our assumption that $n \geqslant p^{-10^k/\tau_0}$. We obtain from Lemma 4.1 a vertex c^* and sets $A_i^* \subseteq A_i^{(\ell)}$. We now set $C^{(\ell+1)} = C^{(\ell)} \setminus \{c^*\}$ and $A_i^{(\ell+1)} = A_i^*$, and claim that this satisfies the desired properties.

Property (i) is immediate since we removed one vertex from $C^{(\ell)}$, as is property (ii), since the newly added vertex to $C \setminus C^{(\ell+1)}$, namely c^*, is adjacent to all vertices in $A_1^{(\ell+1)} \cup \cdots \cup A_k^{(\ell+1)}$ by the statement of Lemma 4.1. Moreover, we indeed have that $\mathcal{A}^{(\ell+1)}$ is strictly R-balanced, and satisfies

$$\mathcal{E}_R(\mathcal{A}^{(\ell+1)}) \geqslant 2^{-2^{k+3}/\tau_0} \mathcal{E}_R(\mathcal{A}^{(\ell)}) \geqslant 2^{-2^{k+3}(\ell+1)/\tau_0} p^R \prod_i |A_i|,$$

proving (iii).

Finally, we stop this process upon defining $\mathcal{G}^{(s)}$. Let $C^\dagger = C \setminus C^{(s)}$, $A_i^\dagger = A_i^{(s)}$. By construction, we have $|C^\dagger| = s$, proving (a). We also automatically get (d) from property (ii). It remains to prove (b) and (c). For both of these, we note that

$$2^{-2^{k+3}s/\tau_0} \geqslant 2^{-(\log n)/(4\log(1/p))} = 2^{-R} \geqslant n^{-1/4},$$

where the final inequality holds since $p \leqslant \frac{1}{2}$. Additionally, by our choice of R, we see that $p^R = n^{-1/4}$. Therefore,

$$d(\mathcal{A}^\dagger)^R \prod_i |A_i^\dagger| = \mathcal{E}_R(\mathcal{A}^\dagger) \geqslant 2^{-2^{k+3}s/\tau_0} p^R \prod_i |A_i| \geqslant n^{-1/4} \cdot n^{-1/4} \cdot \prod_i |A_i|,$$

by property (iii) and our choices of s and R. By combining this with the trivial upper bounds $|A_i^\dagger| \leqslant |A_i|$ and $d(\mathcal{A}^\dagger)$, we conclude that $|A_i^\dagger| \geqslant |A_i|/\sqrt{n} \geqslant \sqrt{n}$ for all i, proving (b). Similarly, we find that

$$d(\mathcal{A}^\dagger)^R \geqslant 2^{-R} \cdot p^R,$$

which yields (c). □

As a corollary, we can immediately find a large complete tripartite graph in a graph with many triangles. We will eventually apply this as an intermediate step in our switching argument, so the statement of the following result is optimized with that application in mind. We denote by $\kappa_3(G)$ the triangle density of a tripartite graph G, that is, the total number of triangles divided by the product of the sizes of the three parts.

Corollary 4.3. *Let $\kappa, \nu \in (0, \frac{1}{2}]$ be parameters, and let $m \geqslant \kappa^{-400/\nu}$. Let G be a tripartite graph with parts A, B, C, where $\min\{|A|, |B|, |C|\} \geqslant m$. If $\kappa_3(G) \geqslant \kappa$ and $\kappa_3(G)/d(A, B) \geqslant \nu$, then G contains a copy of $K_{s,t,t}$, where*

$$s = \left\lfloor \frac{\nu \log m}{512 \log \frac{1}{\kappa}} \right\rfloor \qquad \text{and} \qquad t = \left\lfloor \frac{\log m}{12 \log \frac{1}{\kappa}} \right\rfloor.$$

Proof. Let $\tau = \nu/2$, and let $F \subseteq E(A, B)$ denote the set of edges lying in at most $\tau|C|$ triangles of G. Then let G' be the subgraph of G obtained by discarding the edges of F (and not discarding any vertices); by construction, we have $\tau_C(G') \geqslant \tau$. The total number of triangles using an edge in F is at most

$$\tau|C| \cdot e(A, B) = \tau d(A, B)|A||B||C| \leqslant \frac{\nu}{2} \cdot \frac{\kappa_3(G)}{\nu}|A||B||C| = \frac{\kappa_3(G)}{2}|A||B||C|.$$

Thus, at least half the triangles in G are also present in G'. In particular,

$$e_{G'}(A, B) \geqslant \frac{\kappa_3(G)|A||B||C|}{2|C|} \geqslant \frac{\kappa}{2}|A||B|,$$

hence $d_{G'}(A, B) \geqslant \kappa/2 \geqslant \kappa^2$, using our assumption $\kappa \leqslant \frac{1}{2}$.

We may now apply Proposition 4.2 with parameters $k = 2, \tau_0 = \tau, p = \kappa^2$ to $\mathcal{G} = G'$; note that a 2-mixed graph is the same as a tripartite graph, and that m is sufficiently large to apply Proposition 4.2 since $m \geqslant \kappa^{-400/\nu} = p^{-100/\tau}$. We obtain sets $A^\dagger \subseteq A, B^\dagger \subseteq B, C^\dagger \subseteq C$, satisfying

(a) $|C^\dagger| = \lfloor (\tau \log m)/(2^7 \log \frac{1}{p}) \rfloor = s$,
(b) $|A^\dagger| \geqslant \sqrt{m}$ and $|B^\dagger| \geqslant \sqrt{m}$,
(c) $d(A^\dagger, B^\dagger) \geqslant p/2$, and
(d) every vertex of $C^\dagger$ is adjacent to every vertex of $A^\dagger \cup B^\dagger$.

In particular, given the last point, it suffices to show that $(A^\dagger, B^\dagger)$ contains a copy of $K_{t,t}$. But this is an immediate consequence of Theorem 3.1 when we note that

$$\left\lfloor \frac{\log \min\{|A^\dagger|, |B^\dagger|\}}{2 \log \frac{2}{p}} \right\rfloor \geqslant \left\lfloor \frac{\log \sqrt{m}}{2 \log \frac{1}{\kappa^3}} \right\rfloor = \left\lfloor \frac{\log m}{12 \log \frac{1}{\kappa}} \right\rfloor = t. \qquad \square$$

5. An improved bound via switching

In order to simplify the presentation of the argument, it is helpful to introduce the following definition, which precisely captures the kind of subgraph whose presence would be useful in the argument above.

Definition 5.1. Let H be a tripartite graph with parts X, Y, Z. We call H an (m, μ)-*switcher* if $\min\{|X|, |Y|, |Z|\} \geqslant m$, $\kappa_3(H) \geqslant \mu^2$, and $\kappa_3(H)/d(X, Y) \geqslant \mu$.

Note that this condition is monotone in both parameters, in the sense that every (m, μ)-switcher is also an (m', μ')-switcher for all $m' \leqslant m$ and all $\mu' \leqslant \mu$. From Proposition 4.2, we see that if $\mu \leqslant \frac{1}{2}$ and m is sufficiently large, then every (m, μ)-switcher contains a copy of $K_{s,t,t}$, where $s = \Omega((\mu \log m)/(\log \frac{1}{\mu}))$ and $t = \Omega((\log m)/(\log \frac{1}{\mu}))$.

Using these ideas, we prove the following strengthening of Lemma 4.1. Note that we have added the assumption that no subgraph of G is a switcher, and have obtained a stronger conclusion in (ii).

Lemma 5.2. *Fix parameters $\tau \in (0, \frac{1}{2}]$ and $R \geqslant 50/\tau$. Let G be a tripartite graph with parts (A, B, C), and suppose that $\tau_C(G) \geqslant \tau$. Let $m \leqslant \frac{\tau}{4} \min\{|A|, |B|, |C|\}$ be a positive integer, and let $p = d(A, B)$. Suppose further that (A, B) is strictly R-balanced, and that no subgraph of G is an $(m, \sqrt{p\tau}/4)$-switcher.*

Then there exist a vertex $c^* \in C$ and non-empty subsets $A^* \subseteq A, B^* \subseteq B$ satisfying the following properties.

(i) c^* is adjacent to all vertices in A^* and B^*.
(ii) We have

$$\mathcal{E}_R(A^*, B^*) \geqslant \tau^{20/\sqrt{p\tau}} \mathcal{E}_R(A, B).$$

(iii) (A^*,B^*) *is strictly* R*-balanced.*

Proof. Let $\kappa=\kappa_3(G)$. Note that the total number of triangles in G equals $\kappa|A||B||C|$, but is also at least $e(A,B)\cdot\tau|C|$, since every edge between A and B lies in at least $\tau_C(G)|C|\geqslant\tau|C|$ triangles; this in particular implies that $\kappa\geqslant p\tau$. Let $C_0\subseteq C$ consist of all vertices $c\in C$ which lie in at most $\frac14\kappa|A||B|$ triangles. The total number of triangles touching C_0 is at most $\frac14\kappa|A||B||C|$, hence there are at least $\frac34\kappa|A||B||C|$ triangles touching a vertex in $C\setminus C_0$. Each such vertex lies in at most $e(A,B)$ triangles, hence

$$|C\setminus C_0|\geqslant\frac{\frac34\kappa|A||B||C|}{e(A,B)}\geqslant\frac34\cdot\frac{e(A,B)\cdot\tau|C|}{e(A,B)}=\frac{3\tau}{4}|C|.$$

Let $\mu=\sqrt{p\tau}/4$, and note that $\mu\leqslant\sqrt{\kappa}/4$ since $\kappa\geqslant p\tau$. Let us call a vertex $c\in C$ A*-sparse* if $c\notin C_0$, and additionally $|N(c)\cap A|\leqslant\mu|A|$. Let C_A denote the set of A-sparse vertices in C, and suppose that $|C_A|\geqslant m$. Then consider the induced subgraph G_A on vertex set $A\cup B\cup C_A$. Since no vertex of C_A is in C_0, we see that G_A contains at least $(\frac14\kappa|A||B|)\cdot|C_A|$ triangles, and hence $\kappa_3(G_A)\geqslant\kappa/4>\mu^2$. Moreover, as every vertex in C_A has degree at most $\mu|A|$ to A, we see that $d(A,C_A)\leqslant\mu$, hence $\kappa_3(G_A)/d(A,C_A)>\mu^2/\mu=\mu$. Finally, we have $\min\{|A|,|B|,|C_A|\}\geqslant m$, hence G_A is an (m,μ)-switcher on parts $X=A,Y=C_A,Z=B$. This contradicts our assumption that no subgraph of G is an (m,μ)-switcher, so we conclude that $|C_A|<m$. By the same logic, we have $|C_B|<m$, where C_B is defined analogously.

We conclude that $|C\setminus(C_0\cup C_A\cup C_B)|\geqslant(3\tau/4)|C|-2m\geqslant\tau|C|/4$. Fix any vertex $c^*\in C\setminus(C_0\cup C_A\cup C_B)$, and let $A^\oplus=N(c^*)\cap A,B^\oplus=N(c^*)\cap B$. By the choice of c^*, we know that $e(A^\oplus,B^\oplus)\geqslant\frac14\kappa|A||B|$, and that $|A^\oplus|\geqslant\mu|A|,|B^\oplus|\geqslant\mu|B|$.

Note that if $p\leqslant16\tau$, then also $p\leqslant16(\kappa/p)$, since $\kappa\geqslant p\tau$. Therefore,

$$\left(\frac{\kappa}{p}\right)^2\geqslant\frac{\kappa}{p}\cdot\frac{p}{16}=\frac{\kappa}{16}\geqslant\mu^2$$

implying that $\kappa_3(G)/d(A,B)=\kappa/p\geqslant\mu$. Since $\min\{|A|,|B|,|C|\}\geqslant m$, we conclude that G itself is an (m,μ)-switcher, a contradiction. Hence we have $p>16\tau$.

Now let $\eta=\sqrt{\tau/p}$, and note by the above that $\eta\leqslant\sqrt{\tau/(16\tau)}=\frac14$. Let $\alpha=(15\eta p\log\frac1\eta)/(\tau R)$. Note that since $R\geqslant50/\tau$ and $\eta\leqslant\frac14$, we have that $\alpha\leqslant p/2$. We also note for future reference that $\eta|A|\geqslant4\eta/\tau>4/\sqrt{\tau}>5$, using that $1\leqslant m\leqslant\frac{\tau}{4}|A|$. This implies that $\lfloor\eta|A|\rfloor\geqslant\frac45\eta|A|$, and similarly $\lfloor\eta|B|\rfloor\geqslant\frac45\eta|B|$.

Our key claim is that

$$d(A^\oplus,B^\oplus)\geqslant p-\alpha.\tag{6}$$

Suppose for contradiction that $d(A^\oplus,B^\oplus)<p-\alpha<p$. In particular, we find that

$$p|A^\oplus||B^\oplus|>e(A^\oplus,B^\oplus)\geqslant\frac14\kappa|A||B|\geqslant\frac14(p\tau)|A||B|,$$

since $\kappa\geqslant\tau_C(G)p\geqslant\tau p$. Rearranging, we conclude that $|A^\oplus||B^\oplus|\geqslant\frac{\tau}{4}|A||B|$.

Let A^+ be a uniformly random subset of $A^\oplus$ of size $\min\{|A^\oplus|,\lfloor\eta|A|\rfloor\}$, and similarly for B^+. By averaging, we may find a specific choice of A^+,B^+ such that $d(A^+,B^+)<p-\alpha$; let us fix such a choice. The key point is that in doing this, we have lost essentially nothing on the lower bound on the product of the sizes of A^+ and B^+, since

$$|A^+||B^+|=\min\{|A^\oplus||B^\oplus|,\lfloor\eta|A|\rfloor|B^\oplus|,\lfloor\eta|B|\rfloor|A^\oplus|,\lfloor\eta|A|\rfloor\lfloor\eta|B|\rfloor\}.$$

The first term is at least $\frac{\tau}{4}|A||B|$. The next two terms are each at least $\frac45\eta\mu|A||B|=\frac{\tau}{5}|A||B|$, by our choices of $\eta=\sqrt{\tau/p}$ and $\mu=\sqrt{p\tau}/4$. Finally, the last term is at least $(\frac45)^2\eta^2|A||B|>\frac15(\tau/p)|A||B|\geqslant\frac{\tau}{5}|A||B|$. So in total, we conclude that $|A^+||B^+|\geqslant\frac{\tau}{5}|A||B|$.

Let $A^- = A \setminus A^+, B^- = B \setminus B^+$. Since $|A^+| \leqslant \eta|A|, |B^+| \leqslant \eta|B|$, we see that for every $(X,Y) \in \{(A^-,B^-),(A^+,B^-),(A^-,B^+)\}$, we have that $(|X||Y|)/(|A||B|) \in [0,\eta] \cup [(1-\eta)^2,1]$. Moreover, by our choice of R and η, we have $R \geqslant \frac{50}{\tau} \geqslant \log\frac{1}{\tau} \geqslant \log\frac{1}{\eta}$. Therefore, since (A,B) is strictly R-balanced, we may apply Lemma 3.5(b) to conclude that for each such choice of (X,Y), we have

$$e(X,Y) \leqslant p|X||Y| + (\eta \log\tfrac{1}{\eta})\frac{p}{R}|A||B| \leqslant p|X||Y| + \frac{5\eta p \log\frac{1}{\eta}}{\tau R}|A^+||B^+|.$$

Summing this inequality over all three choices of (X,Y), as well as the inequality $e(A^+,B^+) < (p-\alpha)|A^+||B^+|$, we conclude that

$$e(A,B) < p|A||B| + \left(\frac{15\eta p\log\frac{1}{\eta}}{\tau R} - \alpha\right)|A^+||B^+|.$$

But this is impossible since, by our choice of α, the second term above is 0, and the first term equals $e(A,B)$. This contradiction concludes the proof of (6).

It now only remains to estimate $\mathcal{E}_R(A^\oplus,B^\oplus)$. For this, we recall that $e(A^\oplus,B^\oplus) \geqslant \frac{1}{4}\kappa|A||B| \geqslant \frac{1}{4}p\tau|A||B|$, and compute that

$$\begin{aligned}\mathcal{E}_R(A^\oplus,B^\oplus) &= e(A^\oplus,B^\oplus)d(A^\oplus,B^\oplus)^{R-1}\\ &\geqslant \left(\frac{1}{4}p\tau|A||B|\right)(p-\alpha)^{R-1}\\ &= (|A||B|p^R)\left(\frac{\tau}{4}p^{1-R}(p-\alpha)^{R-1}\right)\\ &\geqslant \mathcal{E}_R(A,B)\left(\tau^3\left(1-\frac{\alpha}{p}\right)^{R-1}\right),\end{aligned}$$

using that $\tau \leqslant 1/2$ in the final inequality. Since $\alpha \leqslant p/2$, we may apply the inequality $1-\alpha/p \geqslant 2^{-2\alpha/p}$, which implies $(1-\alpha/p)^{R-1} \geqslant 2^{-2R\alpha/p}$. Now note that

$$\frac{2R\alpha}{p} = \frac{30\eta\log\frac{1}{\eta}}{\tau} \leqslant 30\frac{\sqrt{\tau/p}\cdot\log(1/\sqrt{\tau})}{\tau} = \frac{15\log\frac{1}{\tau}}{\sqrt{\tau p}}$$

where we use that $\eta = \sqrt{\tau/p} \geqslant \sqrt{\tau}$ in the first inequality. We may thus conclude that

$$\tau^3\left(1-\frac{\alpha}{p}\right)^{R-1} \geqslant \tau^3\cdot 2^{-15\log(1/\tau)/\sqrt{p\tau}} = \tau^3\cdot\tau^{15/\sqrt{p\tau}} \geqslant \tau^{20/\sqrt{p\tau}},$$

and hence that $\mathcal{E}_R(A^\oplus,B^\oplus) \geqslant \tau^{20/\sqrt{p\tau}}\mathcal{E}_R(A,B)$. Finally, we pass to a strictly R-balanced subgraph (A^*,B^*) satisfying $\mathcal{E}_R(A^*,B^*) \geqslant \mathcal{E}_R(A^\oplus,B^\oplus)$, as given by Lemma 3.4, which satisfies the desired properties since c^* is adjacent to all vertices in $A^\oplus \cup B^\oplus \supseteq A^* \cup B^*$. □

Our next result is obtained by iterating Lemma 5.2, and the proof is very similar to that of Proposition 4.2.

Proposition 5.3. *Fix parameters $\tau_0, p_0 \in (0,\frac{1}{2}]$. Let $n \geqslant p_0^{-400/\tau_0}$ and let G be a tripartite graph with parts A,B,C satisfying $\min\{|A|,|B|,|C|\} \geqslant n$. Suppose that $\tau_C(G) \geqslant \tau_0$ and $d(A,B) \geqslant p_0$, and that no subgraph of G is a $(\lfloor n^{1/3}\rfloor, \sqrt{p_0\tau_0}/8)$-switcher.*

Then there exist subsets $A^\dagger \subseteq A, B^\dagger \subseteq B, C^\dagger \subseteq C$ satisfying

(a) $|C^\dagger| = \lfloor(\sqrt{p_0\tau_0}\log n)/(320\log\frac{1}{p_0}\log\frac{1}{\tau_0})\rfloor$,

(b) $|A^\dagger| \geqslant \sqrt{n}$ and $|B^\dagger| \geqslant \sqrt{n}$,

(c) $d(A^\dagger,B^\dagger) \geqslant p_0/2$, and

(d) Every vertex of $C^\dagger$ is adjacent to every vertex of $A^\dagger \cup B^\dagger$.

As the proof of Proposition 5.3 is quite similar to that of Proposition 4.2, we will be brief, mostly highlighting the places where the differences lie.

Proof of Proposition 5.3. Let

$$s = \left\lfloor \frac{\sqrt{p_0\tau_0}\log n}{320\log\frac{1}{p_0}\log\frac{1}{\tau_0}} \right\rfloor, \qquad R = \frac{\log n}{4\log\frac{1}{p_0}}, \qquad m = \lfloor n^{1/3} \rfloor, \qquad \kappa_0 = \frac{p_0\tau_0}{4}, \qquad \mu = \frac{\sqrt{\kappa_0}}{4}.$$

We note that by our assumption $n \geqslant p_0^{-400/\tau_0} \geqslant 2^{400/\tau_0}$, we have that $\frac{\tau_0}{8}\sqrt{n} \geqslant m$. We will define a nested sequence of induced subgraphs $G \supseteq G^{(0)} \supseteq \cdots \supseteq G^{(s)}$ on vertex sets $(A^{(\ell)}, B^{(\ell)}, C^{(\ell)})_{\ell=0}^s$, satisfying that $|C^{(\ell)}| = |C| - \ell$, that every vertex in $C \setminus C^{(\ell)}$ is adjacent to all of $A^{(\ell)} \cup B^{(\ell)}$, and that $(A^{(\ell)}, B^{(\ell)})$ is strictly R-balanced and satisfies

$$\mathcal{E}_R(A^{(\ell)}, B^{(\ell)}) \geqslant \tau_0^{40\ell/\sqrt{\kappa_0}} p_0^R |A||B|.$$

To begin the iteration, we let $C^{(0)} = C$, and let $(A^{(0)}, B^{(0)})$ be a strictly R-balanced subgraph of (A, B), as given by Lemma 3.4, which trivially satisfies these properties.

Now suppose we have defined $G^{(\ell)}$ for some $\ell \leqslant s$. We note several properties of this graph. First, as $|C \setminus C^{(\ell)}| = \ell \leqslant s \leqslant \tau_0 n/2$, we have that $\tau_{C^{(\ell)}}(G^{(\ell)}) \geqslant \tau_0/2$. Second, using the fact that $p_0^R = n^{-1/4}$ and

$$\tau_0^{40\ell/\sqrt{\kappa_0}} \geqslant \tau_0^{40s/\sqrt{\kappa_0}} \geqslant 2^{-R} \geqslant n^{-1/4},$$

we find that

$$d(A^{(\ell)}, B^{(\ell)})^R |A^{(\ell)}||B^{(\ell)}| = \mathcal{E}_R(A^{(\ell)}, B^{(\ell)}) \geqslant 2^{-R} p_0^R |A||B| \geqslant n^{-1/2}|A||B|.$$

This implies immediately that $|A^{(\ell)}| \geqslant \sqrt{n}, |B^{(\ell)}| \geqslant \sqrt{n}$, as well as that $d(A^{(\ell)}, B^{(\ell)}) \geqslant p_0/2$.

Assuming now that $\ell < s$, we apply Lemma 5.2 to $G^{(\ell)}$, with parameters $\tau = \tau_0/2, p = d(A^{(\ell)}, B^{(\ell)}) \geqslant p_0/2$ and R, m defined as above. Our assumption $n \geqslant p_0^{-400/\tau_0}$ implies that $R \geqslant 100/\tau_0 = 50/\tau$. We've also seen that $m \leqslant \frac{\tau_0}{8}\sqrt{n} \leqslant \frac{\tau}{4}\min\{|A^{(\ell)}|, |B^{(\ell)}|, |C^{(\ell)}|\}$, using that $|C^{(\ell)}| \geqslant |C| - s \geqslant \sqrt{n}$. Finally, since $G^{(\ell)}$ is a subgraph of G, we see that no subgraph of $G^{(\ell)}$ is an $(m, \sqrt{p_0\tau_0}/8)$-switcher, hence no subgraph of $G^{(\ell)}$ is an $(m, \sqrt{p\tau}/4)$-switcher. We have thus verified all the conditions of Lemma 5.2, which produces for us a vertex c^* and sets $A^* \subseteq A^{(\ell)}, B^* \subseteq B^{(\ell)}$ satisfying

$$\mathcal{E}_R(A^*, B^*) \geqslant \tau^{20/\sqrt{p\tau}} \mathcal{E}_R(A^{(\ell)}, B^{(\ell)}) \geqslant \left(\frac{\tau_0}{2}\right)^{20/\sqrt{p_0\tau_0/4}} \mathcal{E}_R(A^{(\ell)}, B^{(\ell)}) \geqslant \tau_0^{40/\sqrt{\kappa_0}} \mathcal{E}_R(A^{(\ell)}, B^{(\ell)}).$$

We set $C^{(\ell+1)} = C^{(\ell)} \setminus \{c^*\}, A^{(\ell+1)} = A^*, B^{(\ell+1)} = B^*$, and see that all of our desired properties remain true for $G^{(\ell+1)}$. Moreover, when we stop the iteration at step s, the computations above show that we may take $A^\dagger = A^{(s)}, B^\dagger = B^{(s)}, C^\dagger = C \setminus C^{(s)}$, which satisfy all the desired properties. $\square$

6. Bounds in Nikiforov's theorem

In this section, we use the results proved in the previous sections to prove Theorem 1.2. In order to apply Propositions 4.2 and 5.3, we need a result that converts an arbitrary graph with many copies of K_{k+1} into a k-mixed graph in which we control both $\tau_C(G)$ and $d(\mathcal{A})$. This is not too hard to do via standard tools from the theory of graph regularity, and we prove the following result in Section A.

Lemma 6.1. *For every integer $k \geqslant 2$ and every $\gamma \in (0, \frac{1}{2}]$, there exists some $\zeta \geqslant \gamma^{k^{50}\gamma^{-10}} > 0$ and some $\tau \in [\gamma^{2/(k+1)}, \frac{1}{2}]$ such that the following holds. Let G be an N-vertex graph with at least γN^{k+1} copies of K_{k+1}. Then there exist disjoint sets $A_1, \ldots, A_k, C \subseteq V(G)$, each of size at least $\lfloor \zeta N \rfloor$, as well as a k-mixed graph $\mathcal{G}$ on vertex set $(\mathcal{A}, C)$ such that*

(1) Every edge of $E_{\mathcal{G}}(C, A_1 \cup \cdots \cup A_k)$ is an edge of G,
(2) Every edge of $E_{\mathcal{G}}(\mathcal{A})$ is a copy of K_k in G, and
(3) $\tau_C(\mathcal{G}) \geqslant \tau \geqslant \gamma^{2/(k+1)}$ and $d(\mathcal{A})\tau \geqslant \gamma/4$.

We begin by proving Theorem 1.2 in the case of triangles.

Theorem 6.2. $c_{K_3}(\gamma) = \Omega(\sqrt{\gamma}/(\log \frac{1}{\gamma})^2)$.

Proof. We may assume that $\gamma \leqslant \frac{1}{2}$, by choosing the implicit constant large enough. Let $N \geqslant \gamma^{-2^{60}\gamma^{-10}}$ be sufficiently large, and let G_0 be an N-vertex graph with at least γN^3 triangles. Applying Lemma 6.1 with $k = 2$, we obtain a tripartite subgraph G on parts $A, B, C \subseteq V(G)$, satisfying $n = \min\{|A|, |B|, |C|\} \geqslant \lfloor \zeta N \rfloor \geqslant \sqrt{N}$, as well as some $\tau \in [\gamma^{2/3}, \frac{1}{2}]$ such that $\tau_C(G) \geqslant \tau \geqslant \gamma^{2/3}$ and $d_G(A, B)\tau \geqslant \gamma/4$. This in particular implies $d_G(A, B) \geqslant \gamma/4$. Note that $n \geqslant \sqrt{N} \geqslant (\gamma/4)^{-400/\gamma^{2/3}}$.

We split into two cases. First, suppose that no subgraph of G is a $(\lfloor n^{1/3} \rfloor, \sqrt{\gamma}/16)$-switcher. Then we are precisely in the setting of Proposition 5.3 (with parameters $\tau_0 = \tau$ and $p_0 = \gamma/(4\tau_0)$), hence we may find sets $A^\dagger, B^\dagger, C^\dagger$ as in the statement of Proposition 5.3. We have that

$$s = |C^\dagger| \geqslant \left\lfloor \frac{\sqrt{\gamma/4}\log n}{320 \log \frac{1}{\gamma^{2/3}} \log \frac{4}{\gamma}} \right\rfloor \geqslant \left\lfloor \frac{\sqrt{\gamma}\log n}{2000(\log \frac{1}{\gamma})^2} \right\rfloor \geqslant \left\lfloor \frac{\sqrt{\gamma}\log N}{4000(\log \frac{1}{\gamma})^2} \right\rfloor,$$

as well as $|A^\dagger|, |B^\dagger| \geqslant \sqrt{n} \geqslant N^{1/4}$, and all vertices of $C^\dagger$ are adjacent to all vertices of $A^\dagger \cup B^\dagger$. We now apply Theorem 3.1 to the bipartite graph between $A^\dagger$ and $B^\dagger$, which has edge density at least $d(A, B)/2 \geqslant \gamma/8$. We conclude that $(A^\dagger, B^\dagger)$ contains a copy of $K_{t,t}$ for

$$t = \left\lfloor \frac{\log N^{1/4}}{2 \log \frac{8}{\gamma}} \right\rfloor \geqslant \left\lfloor \frac{\log N}{32 \log \frac{1}{\gamma}} \right\rfloor \geqslant s.$$

Together with $C^\dagger$, we find a copy of $K_{s,t,t} \supseteq K_{s,s,s}$ in G, completing the proof in this case.

On the other hand, if G contains a $(\lfloor n^{1/3} \rfloor, \sqrt{\gamma}/16)$-switcher, we apply Corollary 4.3 to this switcher, with parameters $m = \lfloor n^{1/3} \rfloor \geqslant N^{1/7}, \kappa = \gamma/256$, and $\nu = \sqrt{\gamma}/16$, noting that $m \geqslant \kappa^{-400/\nu}$. We conclude that G contains a complete tripartite graph $K_{s',t',t'}$, where

$$s' = \left\lfloor \frac{\nu \log m}{512 \log \frac{256}{\gamma}} \right\rfloor \geqslant \left\lfloor \frac{\sqrt{\gamma}\log N}{10^7 \log \frac{1}{\gamma}} \right\rfloor, \qquad t' = \left\lfloor \frac{\log m}{12 \log \frac{256}{\gamma}} \right\rfloor \geqslant s',$$

and we again obtain the desired result. □

Next, we use Proposition 4.2 to provide a recursive bound on $c_{K_{k+1}}(\gamma)$ in terms of $c_{K_k}(\gamma)$.

Theorem 6.3. *For every $k \geqslant 3$, we have*

$$c_{K_{k+1}}(\gamma) \geqslant \min\left\{\Omega_k\left(\frac{\gamma^{2/(k+1)}}{\log \frac{1}{\gamma}}\right), \frac{1}{8}c_{K_k}\left(\frac{\gamma}{8k^k}\right)\right\}.$$

Proof. As before, we may assume that $\gamma \leqslant 2^{-k}$, by picking the implicit constant sufficiently large. Let $\gamma' = \gamma/(8k^k)$, and let m_0 be chosen sufficiently large so that for all $m \geqslant m_0$, every m-vertex graph with $\gamma' m^k$ copies of K_k contains a copy of $K_k[t]$, where $t \geqslant \frac{1}{2}c_{K_k}(\gamma')\log m$.

Let now N be sufficiently large, and in particular at least m_0^4. Let G be an N-vertex graph with at least γN^{k+1} copies of K_{k+1}. We apply Lemma 6.1 to find disjoint sets $A_1, \ldots, A_k, C \subseteq V(G)$, as well as a k-mixed graph $\mathcal{G}$ on this vertex set. Since N is sufficiently large, we may assume $n = \min\{|A_1|, \dots, |A_k|, |C|\} \geqslant \sqrt{N}$. We now apply Proposition 4.2 to $\mathcal{G}$, with parameters $\tau_0 = \gamma^{2/(k+1)}$ and $p = \gamma/(4\tau_C(\mathcal{G})) \geqslant \gamma/4$, to obtain sets $A_i^\dagger, C^\dagger$. We have that

$$|C^\dagger| = \left\lfloor \frac{\tau_0 \log n}{2^{k+5} \log \frac{1}{p}} \right\rfloor \geqslant \left\lfloor \frac{\gamma^{2/(k+1)} \log N}{2^{k+8} \log \frac{1}{\gamma}} \right\rfloor.$$

Now, let $A_i' \subseteq A_i^\dagger$ be a randomly chosen set of size exactly $\lfloor N^{1/4} \rfloor$, which exists since $|A_i^\dagger| \geqslant \sqrt{n} \geqslant \lfloor N^{1/4} \rfloor$. In expectation, the number of k-uniform edges of $\mathcal{G}$ in $A_1' \cup \dots \cup A_k'$ is at least $(p/2)\lfloor N^{1/4} \rfloor^k$, hence we may fix an outcome with at least that many edges. This yields for us an induced subgraph of $G[A_1^\dagger \cup \dots \cup A_k^\dagger]$ of order $m = k\lfloor N^{1/4} \rfloor$ with at least $(\gamma/8)\lfloor N^{1/4} \rfloor^k = \gamma' m^k$ copies of K_k. By the definition of m_0, we can find in this subgraph a copy of $K_k[t]$, for $t = \frac{1}{2} c_{K_k}(\gamma') \log m \geqslant \frac{1}{8} c_{K_k}(\gamma') \log N$, and all of its vertices are adjacent in G to $C^\dagger$. Hence, together with $C^\dagger$, we obtain a blowup of K_{k+1} of order $\min\{t, |C^\dagger|\}$, which yields the desired result. $\square$

We note that this proof actually shows that the "sufficiently large" condition on the number of vertices for K_{k+1} is polynomial in that for K_k. In particular, since Theorem 6.2 already applies for $N \geqslant \gamma^{-O(\gamma^{-10})}$, we conclude that similarly Theorem 6.3 also holds for $N \geqslant \gamma^{-O_k(\gamma^{-10})}$.

Finally, we use the following simple and well-known argument, that shows that a bound on $c_{K_h}(\gamma)$ implies essentially the same bound on $c_H(\gamma)$ for any h-vertex graph H.

Proposition 6.4. *If H has h vertices, then $c_H(\gamma) \geqslant c_{K_h}(h^{-h}\gamma)$.*

Proof. Let G be an N-vertex graph, and suppose that G contains at least γN^h copies of H. In particular, if we identify the vertex set of H with $[h]$, then G contains at least γN^h labeled copies of H. We now pick a random partition of $V(G)$ into h parts $V_1, \dots, V_h$. Every labeled copy of H has a probability h^{-h} of being *spanning*, that is, having its ith vertex end up in V_i for all $i \in [h]$. So by linearity of expectation, there exists some h-partition of V so that at least $h^{-h}\gamma N^h$ labeled copies of H are spanning.

We now define a new h-partite graph $\widetilde{G}$ on the same vertex set as G. We obtain $\widetilde{G}$ from G by first deleting every edge contained in some V_i. Next, for every pair (i,j) such that $ij \notin E(H)$, we place a complete bipartite graph between V_i and V_j. However, if $ij \in E(H)$, then we let $\widetilde{G}$ have the same edges between V_i and V_j as are found in G.

Every spanning copy of H in G thus yields a copy of K_h in $\widetilde{G}$, and conversely, every copy of K_h in $\widetilde{G}$ arises from a spanning copy of H in G. So by the definition of $c_{K_h}(h^{-h}\gamma)$, we see that $\widetilde{G}$ must contain $K_h[t]$ as a subgraph, where $t \geqslant (c_{K_h}(h^{-h}\gamma) - o(1)) \log N$, and where the $o(1)$ term tends to 0 as $N \to \infty$. As $\widetilde{G}$ is h-partite, each of the parts of this $K_h[t]$ must lie in a distinct part V_i of $\widetilde{G}$. But if we only restrict our attention to pairs (V_i, V_j) where $ij \in E(H)$, we see that this $K_h[t]$ gives rise to a copy of $H[t]$ inside G. By the definition of $c_H(\gamma)$, this completes the proof. $\square$

With these tools, the proof of Theorem 1.2 is essentially immediate.

Proof of Theorem 1.2. The case $h = 1$ is trivial, and the case $h = 2$ is covered by Theorem 3.1, hence we may assume that $h \geqslant 3$. As before, we may also assume that $\gamma \leqslant \frac{1}{2}$. We first claim that for all $h \geqslant 3$, we have $c_{K_h}(\gamma) = \Omega_h(\sqrt{\gamma}/(\log \frac{1}{\gamma})^2)$. We prove this by induction on h, with the base case $h = 3$ being precisely the statement of Theorem 6.2. Having proved the statement for $h = k$, we obtain it for $h = k + 1 \geqslant 4$ by Theorem 6.3, since

$$c_{K_{k+1}}(\gamma) \geqslant \min\left\{\Omega\left(\frac{\gamma^{2/(k+1)}}{\log \frac{1}{\gamma}}\right), \frac{1}{8} c_{K_k}\left(\frac{\gamma}{8k^k}\right)\right\} = \min\left\{\Omega\left(\frac{\gamma^{2/4}}{\log \frac{1}{\gamma}}\right), \Omega_k\left(\frac{\sqrt{\gamma}}{(\log \frac{1}{\gamma})^2}\right)\right\},$$

and both terms are at least $\Omega_{k+1}(\sqrt{\gamma}/(\log \frac{1}{\gamma})^2)$, completing the induction. Finally, for a general graph H with h vertices, we see by Proposition 6.4 that

$$c_H(\gamma) \geqslant c_{K_h}(h^{-h}\gamma) = \Omega_H\left(\frac{\sqrt{\gamma}}{(\log \frac{1}{\gamma})^2}\right). \qquad \square$$

7. Concluding remarks

As discussed in the introduction, our iterative approach to Nikiforov's theorem is closely connected to techniques used in many other problems, and this suggests that these connections should be explored further. For example, the resemblance between our technique and the one used to bound multicolor Ramsey numbers [1] suggests that it may be useful to attack a multicolor Ramsey-theoretic variant of Nikiforov's theorem proposed in [14, Conjecture 1.6]. Moreover, in the setting of regularity lemmas, such iterative proofs can often be "reverse engineered" to show that the bound they give is optimal; this was first achieved in the breakthrough work of Gowers [16], and his technique has been adapted and extended to a wide variety of problems, e.g. [5, 10, 12, 13, 15, 17, 19, 20, 23, 24, 25]. Although we have not succeeded in doing the analogous thing here, it is conceivable that by reverse-engineering our proof, one could show that, for example, $c_{K_3}(\gamma) \leqslant \gamma^{\Omega(1)}$, demonstrating that random graphs are very far from extremal in Nikiforov's theorem.

Indeed, the most interesting question about this topic remains open, namely whether the true behavior of $c_H(\gamma)$ is polynomial or sub-polynomial. In case H is triangle-free, the result of [14] shows that the true value is $c_H(\gamma) = \Theta_H(1/\log \frac{1}{\gamma})$, but for any other H there is a wide gap between the bounds

$$\gamma^{1/2+o(1)} \leqslant c_H(\gamma) \leqslant O_H\left(\frac{1}{\log \frac{1}{\gamma}}\right),$$

where the lower bound is Theorem 1.2 and the upper bound comes from considering a random graph. Given the result of [14], it is natural to conjecture that the upper bound is closer to the truth, but we are not sure whether to believe this conjecture: as discussed above, it is natural to try to "reverse-engineer" our iterative proof to obtain a polynomial upper bound $c_H(\gamma) \leqslant \gamma^{\Omega(1)}$. While we have been unsuccessful in doing this, we believe this is a problem that deserves further attention.

Another promising direction is to take advantage of the apparent similarity between our technique and the book algorithm [4]. Recall that the main difficulty we encounter is that when we select a vertex $c^* \in C$ and define $A^+ = N(c^*) \cap A, B^+ = N(c^*) \cap B$, the density $d(A^+, B^+)$ may be smaller than $d(A, B)$. This is the main effect hurting us in our iterative algorithm, and much of the argument is about trying to limit its harm.

This is essentially the same issue one encounters when trying to prove upper bounds on diagonal Ramsey numbers with the book algorithm. Note that this problem can be described as a "negative correlation" phenomenon between the neighborhoods of different vertices. Indeed, the only way for $d(A^+, B^+)$ to be substantially smaller than $d(A, B)$ is if the neighborhoods of vertices in C are biased away from the edges in $E(A, B)$. In [1], Balister et al. introduced a novel way of controlling the effect of such negative correlations, via their so-called *geometric lemma*. Roughly speaking, they show that if there are a lot of these negative correlations, one is able to focus on a large subgraph where one actually has substantial *positive* correlation. It would be very interesting if one could adapt their technique to our setting, and ideally to thus obtain a better bound on $c_H(\gamma)$.

Finally, we note that, like most prior works on this subject [11, 26, 27, 28] (but notably not [14]), Theorem 1.2 assumes that N is sufficiently large with respect to γ. As discussed in detail in [14, Section 1.2], it would also be of great interest to prove such results without this assumption, i.e. to allow γ to tend to 0 as a function of N. Our proof shows that Theorem 1.2 holds when $N \geqslant 2^{C_H \gamma^{-11}}$ for some constant C_H depending on H, or equivalently that one can let γ tend to 0 as any function slower than $(\log N)^{-1/11}$ and still obtain meaningful results. This is unfortunately too slow for most applications, e.g. [28, Problem 3] or [6, Conjecture 1] (and in any case the bound of $\gamma^{1/2+o(1)}$ is too weak for these applications). We refer to [14, Section 1.2] for more information on this point.

References

[1] P. Balister, B. Bollobás, M. Campos, S. Griffiths, E. Hurley, R. Morris, J. Sahasrabudhe, and M. Tiba, Upper bounds for multicolour Ramsey numbers, *J. Amer. Math. Soc.* **39** (2026), 765–780.

[2] B. Bollobás and P. Erdős, On the structure of edge graphs, *Bull. London Math. Soc.* **5** (1973), 317–321.

[3] B. Bollobás, P. Erdős, and M. Simonovits, On the structure of edge graphs. II, *J. London Math. Soc. (2)* **12** (1975/76), 219–224.

[4] M. Campos, S. Griffiths, R. Morris, and J. Sahasrabudhe, An exponential improvement for diagonal Ramsey, *Ann. of Math. (2)* **203** (2026), 869–932.

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

[6] D. Conlon, J. Fox, and B. Sudakov, Erdős–Hajnal-type theorems in hypergraphs, *J. Combin. Theory Ser. B* **102** (2012), 1142–1154.

[7] R. A. 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.

[8] P. Erdős, M. Goldberg, J. Pach, and J. Spencer, Cutting a graph into two dissimilar halves, *J. Graph Theory* **12** (1988), 121–131.

[9] P. Erdös and A. H. Stone, On the structure of linear graphs, *Bull. Amer. Math. Soc.* **52** (1946), 1087–1091.

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

[11] J. Fox, S. Luo, and Y. Wigderson, Extremal and Ramsey results on graph blowups, *J. Comb.* **12** (2021), 1–15.

[12] J. Fox, H. T. Pham, and Y. Zhao, Tower-type bounds for Roth's theorem with popular differences, *J. Eur. Math. Soc. (JEMS)* **25** (2023), 3795–3831.

[13] F. Garbe and J. Hladký, A tower lower bound for the degree relaxation of the regularity lemma, *Comb. Theory* **5** (2025), Paper No. 8, 17pp.

[14] A. Girão, Z. Hunter, and Y. Wigderson, Blowups of triangle-free graphs, *Adv. Comb.* (2025), Paper No. 10, 23pp.

[15] L. Gishboliner, A. Shapira, and Y. Wigderson, Is it easy to regularize a hypergraph with easy links?, *Int. Math. Res. Not. IMRN* (2026), Paper No. rnag018, 18pp.

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

[17] B. Green, A Szemerédi-type regularity lemma in abelian groups, with applications, *Geom. Funct. Anal.* **15** (2005), 340–376.

[18] P. Gupta, N. Ndiaye, S. Norin, and L. Wei, Optimizing the CGMS upper bound on Ramsey numbers, 2024. Preprint available at arXiv:2407.19026.

[19] K. Hosseini, S. Lovett, G. Moshkovitz, and A. Shapira, An improved lower bound for arithmetic regularity, *Math. Proc. Cambridge Philos. Soc.* **161** (2016), 193–197.

[20] S. Kalyanasundaram and A. Shapira, A Wowzer-type lower bound for the strong regularity lemma, *Proc. Lond. Math. Soc. (3)* **106** (2013), 621–649.

[21] T. Kövari, V. T. Sós, and P. Turán, On a problem of K. Zarankiewicz, *Colloq. Math.* **3** (1954), 50–57.
[22] R. Morris, Some recent results in Ramsey theory, 2026. Preprint available at arXiv:2601.05221.
[23] G. Moshkovitz and A. Shapira, A short proof of Gowers' lower bound for the regularity lemma, *Combinatorica* **36** (2016), 187–194.
[24] G. Moshkovitz and A. Shapira, A sparse regular approximation lemma, *Trans. Amer. Math. Soc.* **371** (2019), 6779–6814.
[25] G. Moshkovitz and A. Shapira, A tight bound for hypergraph regularity, *Geom. Funct. Anal.* **29** (2019), 1531–1578.
[26] V. Nikiforov, Graphs with many copies of a given subgraph, *Electron. J. Combin.* **15** (2008), Note 6, 6pp.
[27] V. Nikiforov, Graphs with many r-cliques have large complete r-partite subgraphs, *Bull. Lond. Math. Soc.* **40** (2008), 23–25.
[28] V. Rödl and M. Schacht, Complete partite subgraphs in dense hypergraphs, *Random Structures Algorithms* **41** (2012), 557–573.
[29] A. Shapira and R. Yuster, On the density of a graph and its blowup, *J. Combin. Theory Ser. B* **100** (2010), 704–719.
[30] V. Souza, Blowup Ramsey numbers, *European J. Combin.* **92** (2021), Paper No. 103238, 14pp.
[31] E. Szemerédi, Regular partitions of graphs, in *Problèmes combinatoires et théorie des graphes (Colloq. Internat. CNRS, Univ. Orsay, Orsay, 1976)*, *Colloq. Internat. CNRS*, vol. 260, CNRS, Paris, 1978, 399–401.
[32] Y. Wigderson, Upper bounds on diagonal Ramsey numbers [after Campos, Griffiths, Morris, and Sahasrabudhe], *Astérisque* **462** (2025), 85–138.

Appendix A. Proof of Lemma 6.1

In this section, we present the proof of Lemma 6.1. Before doing so, we recall a number of concepts and results related to Szemerédi's regularity lemma. Given $\varepsilon \in (0,1)$, a pair of vertex sets (S,T) (not necessarily disjoint) in a graph F is said to be ε-regular if $|d(S',T') - d(S,T)| < \varepsilon$ for all $S' \subseteq S, T' \subseteq T$ with $|S'| \geqslant \varepsilon|S|, |T'| \geqslant \varepsilon|T|$. Suppose now that F is r-partite, with r-partition $V = V_1 \sqcup \cdots \sqcup V_r$. A *cylinder* K is a set of the form $W_1 \times \cdots W_r$, where $W_i \subseteq V_i$ for all $i \in [r]$. For such a cylinder K, let $V_i(K) = W_i$. We say that K is *ε-regular* if (W_i, W_j) is ε-regular for all $1 \leqslant i < j \leqslant r$. A *cylinder partition* $\mathcal{K}$ is a partition of $V_1 \times \cdots \times V_r$ into cylinders, and we say that $\mathcal{K}$ is *ε-regular* if at most an ε-fraction of the r-tuples $(v_1, \ldots, v_r) \in V_1 \times \cdots \times V_r$ are not in ε-regular cylinders of $\mathcal{K}$. The following weak regularity lemma of Duke, Lefmann, and Rödl says that every r-partite graph has an ε-regular cylinder partition with a bounded number of parts.

Lemma A.1 (Duke–Lefmann–Rödl). *For every $\varepsilon \in (0, \frac{1}{2})$ and integer $r \geqslant 2$, there exists $\beta = \beta(\varepsilon, r) > 0$ such that the following holds. Suppose that $F = (V,E)$ is an r-partite graph with an r-partition $V_1 \sqcup \cdots \sqcup V_r$. Then there exists an ε-regular cylinder partition $\mathcal{K}$ of $V_1 \times \cdots \times V_r$ such that $|V_i(K)| \geqslant \beta|V_i|$ for each $K \in \mathcal{K}$ and $i \in [r]$. Moreover, we may take $\beta \geqslant \varepsilon^{r^2\varepsilon^{-5}}$.*

We will also need the following standard counting lemma.

Lemma A.2. *Let H be a graph on vertex set $[t]$. Suppose $V_1, \ldots, V_t$ are (not necessarily distinct) vertex sets in a graph G such that (V_i, V_j) is ε-regular for all $\{i,j\} \in E(H)$. Then the number of tuples $(v_1, \ldots, v_t) \in V_1 \times \cdots \times V_t$ such that $\{v_i, v_j\} \in E(G)$ whenever $\{i,j\} \in E(H)$ is*

$$\left(\prod_{\{i,j\} \in E(H)} d(V_i, V_j) \pm \varepsilon \binom{t}{2} \right) \prod_{i=1}^{t} |V_i|.$$

Proof of Lemma 6.1. Let $\varepsilon = \gamma^2/(9\binom{k+2}{2})$. Let $\beta = \beta(\varepsilon, k+1)$ be the constant from Lemma A.1, and let $\zeta = \beta\gamma/e^{k+1}$. Note that $\zeta \geqslant \gamma^{k^{50}\gamma^{-10}}$, as claimed.

We now choose a uniformly random partition of $V(G)$ into $k+1$ parts, by assigning each vertex independently and uniformly to one of the $k+1$ parts. The probability that a given K_{k+1} in G spans the $k+1$ parts is exactly $(k+1)!/(k+1)^{k+1}$. Linearity of expectation then implies that there exists a $(k+1)$-partite subgraph of G with parts $V_1, V_2, \ldots, V_{k+1}$ which contains at least $\gamma(k+1)!(N/(k+1))^{k+1} \geqslant \gamma(k+1)! \prod_{i=1}^{k+1}|V_i|$ copies of K_{k+1}. This in particular implies that $\min_i|V_i| \geqslant \gamma(k+1)!N/(k+1)^{k+1} \geqslant \gamma N/e^{k+1}$, since the total number of K_{k+1} in this subgraph is at most $N^k \min_i|V_i|$. Let F be the $(k+1)$-partite subgraph of G with these parts. By Lemma A.1, there is an ε-regular cylinder partition $\mathcal{K}$ of $V_1 \times \cdots \times V_{k+1}$, with $|V_i(K)| \geqslant \lfloor \beta|V_i|\rfloor \geqslant \lfloor \beta\gamma N/e^{k+1}\rfloor = \lfloor \zeta N\rfloor$ for all $K \in \mathcal{K}$ and $i \in [k+1]$. By Lemma A.2, if K is an ε-regular cylinder, then the number of K_{k+1} contained in K is at most

$$\left(\prod_{1\leqslant i<j\leqslant k+1} d(V_i(K), V_j(K)) + \varepsilon\binom{k+1}{2}\right)\prod_{i=1}^{k+1}|V_i(K)|.$$

On the other hand, since at most an ε-fraction of the $(k+1)$-tuples in $V_1 \times \cdots \times V_{k+1}$ are contained in non-ε-regular cylinders, we find that the number of K_{k+1} contained in irregular cylinders is at most $\varepsilon\prod_i|V_i|$. Adding these two facts up over all the cylinders $K \in \mathcal{K}$, we find that the total number of K_{k+1} in $V_1 \times \cdots \times V_{k+1}$ is at most

$$\varepsilon\prod_{i=1}^{k+1}|V_i| + \sum_{K\ \varepsilon\text{-regular}}\left(\prod_{1\leqslant i<j\leqslant k+1} d(V_i(K), V_j(K)) + \varepsilon\binom{k+1}{2}\right)\prod_{i=1}^{k+1}|V_i(K)|.$$

On the other hand, we also know that this number is at least $\gamma(k+1)!\prod_i|V_i|$. Therefore, we find some ε-regular cylinder K which satisfies

$$\prod_{1\leqslant i<j\leqslant k+1} d(V_i(K), V_j(K)) \geqslant \gamma(k+1)! - \varepsilon - \varepsilon\binom{k+1}{2} \geqslant \gamma((k+1)! - 1), \tag{7}$$

by our choice of ε. We can then rewrite (7) as

$$\prod_{i=1}^{k+1}\left(\prod_{\substack{1\leqslant j\leqslant k+1\\ j\neq i}} d(V_i(K), V_j(K))\right)^{1/2} \geqslant \gamma((k+1)! - 1). \tag{8}$$

Suppose without loss of generality that the parenthesized term in (8) is maximized for $i = k+1$. This implies that

$$\prod_{j=1}^{k} d(V_j(K), V_{k+1}(K)) \geqslant (\gamma((k+1)! - 1))^{2/(k+1)} \geqslant \frac{5}{2}\gamma^{2/(k+1)}, \tag{9}$$

where the final inequality uses that $((k+1)! - 1)^{2/(k+1)}$ is an increasing function of k, and its value at $k = 2$ is $5^{2/3} > 5/2$. We now set $C = V_{k+1}(K)$ and $A_i = V_i(K)$ for $1 \leqslant i \leqslant k$, recalling that we proved that all these sets have size at least $\lfloor \zeta N\rfloor$, as needed. We note that by (7) and our choice of ε, we have that

$$\varepsilon\binom{k}{2} \leqslant \frac{1}{9} \prod_{1\leqslant i<j\leqslant k} d(A_i, A_j), \tag{10}$$

$$\varepsilon\binom{k+1}{2} \leqslant \frac{1}{9} \left(\prod_{1\leqslant i<j\leqslant k} d(A_i, A_j)\right) \left(\prod_{i=1}^{k} d(A_i, C)\right), \qquad \text{and} \tag{11}$$

$$\varepsilon\binom{k+2}{2} \leqslant \frac{1}{9} \left(\prod_{1\leqslant i<j\leqslant k} d(A_i, A_j)\right) \left(\prod_{i=1}^{k} d(A_i, C)^2\right). \tag{12}$$

Indeed, the final inequality follows immediately from (7), (9), and our choice of ε, and it immediately implies the other two.

Let N_k denote the number of labeled K_k with one vertex in each of $A_1, \dots, A_k$, and let N_{k+1} denote the number of labeled K_{k+1} with one vertex in each of $A_1, \dots, A_k, C$. Additionally, let K_{k+2}^- denote K_{k+2} with an edge deleted, and let N_{k+2}^- denote the number of labeled homomorphic copies of K_{k+2}^- such that the two non-adjacent vertices are in C, and the k remaining vertices are in $A_1, \dots, A_k$, respectively.

Let $\mathbf{Q}$ be a uniformly randomly chosen K_k among $A_1, \dots, A_k$, and let $\mathbf{Z}$ denote the number of extensions of $\mathbf{Q}$ in C, i.e. the number of vertices $c \in C$ such that $\mathbf{Q} \cup \{c\}$ is a K_{k+1}. We observe that

$$\mathbb{E}[\mathbf{Z}] = \frac{1}{N_k} \sum_{Q} |\{c \in C : Q \cup \{c\} \text{ is a } K_{k+1}\}| = \frac{N_{k+1}}{N_k},$$

where the sum is over all choices of $\mathbf{Q}$, that is all cliques Q among $A_1, \dots, A_k$. Similarly, we have that

$$\mathbb{E}[\mathbf{Z}^2] = \frac{1}{N_k} \sum_{Q} |\{c_1, c_2 \in C : Q \cup \{c_1\}, Q \cup \{c_2\} \text{ are both } K_{k+1}\}| = \frac{N_{k+2}^-}{N_k}.$$

By Lemma A.2 and the bounds (10), (11), and (12), we have that

$$\begin{aligned}
\frac{8}{9} \prod_{1\leqslant i<j\leqslant k} d(A_i, A_j) \prod_{i=1}^{k} |A_i| \leqslant N_k \leqslant \frac{10}{9} &\prod_{1\leqslant i<j\leqslant k} d(A_i, A_j) \prod_{i=1}^{k} |A_i|, \\
N_{k+1} &\geqslant \frac{8}{9} \left(\prod_{1\leqslant i<j\leqslant k} d(A_i, A_j)\right) \left(\prod_{i=1}^{k} d(A_i, C)\right) \left(\prod_{i=1}^{k} |A_i|\right) |C| \\
N_{k+2}^- &\leqslant \frac{10}{9} \left(\prod_{1\leqslant i<j\leqslant k} d(A_i, A_j)\right) \left(\prod_{i=1}^{k} d(A_i, C)^2\right) \left(\prod_{i=1}^{k} |A_i|\right) |C|^2.
\end{aligned}$$

Therefore, we deduce that

$$\mathbb{E}[\mathbf{Z}] = \frac{N_{k+1}}{N_k} \geqslant \frac{4}{5} \left(\prod_{i=1}^{k} d(A_i, C)\right) |C| \qquad \text{and} \qquad \frac{\mathbb{E}[\mathbf{Z}]^2}{\mathbb{E}[\mathbf{Z}^2]} = \frac{N_{k+1}^2}{N_k N_{k+2}^-} \geqslant \frac{16}{25}.$$

Let

$$\tau = \frac{2}{5} \prod_{i=1}^{k} d(A_i, C)$$

so that $\tau \geqslant \gamma^{2/(k+1)}$ by (9), and $\tau \leqslant \frac{2}{5} \leqslant \frac{1}{2}$ by definition. Let $\mathcal{G}$ be the k-mixed graph on $(\mathcal{A}, C)$ whose edges between C and $A_1 \cup \cdots \cup A_k$ are all edges of G between these sets, and whose edges in $\mathcal{A}$ consist of all K_k which have at least $\tau|C|$ extensions in C. By definition, we have $\tau_C(\mathcal{G}) \geqslant \tau \geqslant \gamma^{2/(k+1)}$.

To compute $e_{\mathcal{G}}(\mathcal{A})$, we note that $e_{\mathcal{G}}(\mathcal{A}) = N_k \Pr(\mathbf{Z} \geqslant \tau|C|)$. By the Paley–Zygmund inequality, we have

$$\Pr(\mathbf{Z} \geqslant \tau|C|) \geqslant \Pr\left(\mathbf{Z} \geqslant \frac{1}{2}\mathbb{E}[\mathbf{Z}]\right) \geqslant \frac{1}{4}\frac{\mathbb{E}[\mathbf{Z}]^2}{\mathbb{E}[\mathbf{Z}^2]} \geqslant \frac{4}{25},$$

and therefore

$$d_{\mathcal{G}}(\mathcal{A}) = \frac{N_k \Pr(\mathbf{Z} \geqslant \tau|C|)}{\prod_{i=1}^k |A_i|} \geqslant \frac{4}{25} \cdot \frac{8}{9} \prod_{1 \leqslant i < j \leqslant k} d(A_i, A_j).$$

This implies that

$$d_{\mathcal{G}}(\mathcal{A})\tau \geqslant \frac{4}{25} \cdot \frac{8}{9} \cdot \frac{2}{5} \left(\prod_{1 \leqslant i < j \leqslant k} d(A_i, A_j)\right)\left(\prod_{i=1}^k d(A_i, C)\right) \geqslant \frac{1}{20} \cdot \gamma((k+1)! - 1) \geqslant \frac{\gamma}{4},$$

where the second inequality uses (7) and the final inequality uses that $k \geqslant 2$. $\square$