

A new approach for the Brown-Erdős-Sós problem

Asaf Shapira *

Mykhaylo Tyomkyn †

Abstract

The celebrated Brown-Erdős-Sós conjecture states that for every fixed e , every 3-uniform hypergraph with $\Omega(n^2)$ edges contains e edges spanned by $e + 3$ vertices. Up to this date all the approaches towards resolving this problem relied on highly involved applications of the hypergraph regularity method, and yet they supplied only approximate versions of the conjecture, producing e edges spanned by $e + O(\log e / \log \log e)$ vertices.

In this short paper we describe a completely different approach, which reduces the problem to a variant of another well-known conjecture in extremal graph theory. A resolution of the latter would resolve the Brown-Erdős-Sós conjecture up to an absolute additive constant.

1 Introduction

1.1 Background and previous results

Some of the most well studied problems in extremal combinatorics are those asking which objects are guaranteed to appear in “dense” objects. Among notable examples are Roth’s Theorem [18] on 3-term arithmetic progressions in dense sets of integers, and the Kővári-Sós-Turán Theorem [16] on bipartite subgraphs of dense graphs. In this paper we consider a question raised by Brown, Erdős and Sós in 1973 [3, 2], which is one of the most famous open problems of this type.

Given an integer $e \geq 3$, one would expect a dense 3-uniform hypergraph (3-graph for short) to contain e edges spanned by a small number of vertices. To quantify this, let (v, e) -configuration denote a set of e edges spanned by at most v vertices. The Brown–Erdős–Sós Conjecture (BESC) states that for every fixed $e \geq 3$ and all large enough n , every 3-graph with $\Omega(n^2)$ edges contains an $(e + 3, e)$ -configuration. Despite a lot of effort over the past 50 years, the BESC is only known to hold for $e = 3$, due to a result of Ruzsa and Szemerédi [21].

Since even the $e = 4$ case of the BESC seems hopeless, it is natural to try to prove approximate versions of the conjecture, namely that 3-graphs with $\Omega(n^2)$ edges contain $(e + f(e), e)$ -configurations, for some slowly growing function f . The first result of the above type was obtained by Sárközy and Selkó [22] who showed that every 3-graph with $\Omega(n^2)$ edges contains for every fixed e an $(e + 2 + \lfloor \log_2 e \rfloor, e)$ -configuration. This was improved by Solymosi and Solymosi [23] for the special case $e = 10$ from 15 to 14 vertices. A general asymptotic improvement of the result of [23] was obtained recently by Conlon, Gishboliner, Levanzov and Shapira [8], who proved the existence of $(e + O(\log e / \log \log e), e)$ -configurations.

*School of Mathematics, Tel Aviv University, Tel Aviv 69978, Israel. Email: asafico@tau.ac.il. Supported in part by ERC Consolidator Grant 863438 and NSF-BSF Grant 20196.

†Department of Applied Mathematics, Charles University. Email: tyomkyn@kam.mff.cuni.cz. Supported in part by ERC Synergy Grant DYNASNET 810115 and GAČR Grant 22-19073S.

Besides its intrinsic interest, the BESC turned out to be one of the most influential problems in extremal combinatorics. For example, the proof of the case $e = 3$ [21] was one of the first applications of Szemerédi’s regularity lemma [24], and further introduced the famous graph removal lemma. One of the main motivations for the development of the celebrated hypergraph regularity method [11, 17, 19, 20, 26] was the hope that it will lead to a resolution of BESC. While this did not materialize, the hypergraph regularity method was instrumental in the latest works [8, 23]. However, although the above proofs rely on highly involved applications of the hypergraph regularity method, it appears that the following natural approximate version of the BESC is beyond their reach.

Conjecture 1.1 (Constant deficiency BESC). *There is an absolute constant d so that for every e and every large enough n , every 3-graph with $\Omega(n^2)$ edges contains an $(e + d, e)$ -configuration.*

1.2 A new approach for Conjecture 1.1

Our aim in this paper is to reduce Conjecture 1.1 to a problem involving graphs. Let us denote by $\text{ex}(n, H)$ the maximum number of edges in an n vertex graph not containing a copy of H as a subgraph. The Kővári-Sós-Turán Theorem [16] which we mentioned above, states that for every fixed $t \leq s$, we have $\text{ex}(n, K_{s,t}) = O(n^{2-1/t})$ where $K_{s,t}$ is the complete bipartite graph with parts of size t and s . This bound is known to be tight for large s , see [4] for recent progress and references. One of the main research directions in extremal graph theory is to obtain better bounds for sparser bipartite graphs. One such problem was raised by Erdős [9], who conjectured that if H is a t -degenerate bipartite graph then $\text{ex}(n, H) = O(n^{2-1/t})$. While there are some approximate results towards this conjecture [1, 10, 13, 15], the question is open even for $t = 2$. Note that in general, the conjectured bound $O(n^{2-1/t})$ for t -degenerate bipartite graphs cannot be improved since the aforementioned $K_{s,t}$ is t -degenerate. In particular, the bound is tight for every t -degenerate H which contains a copy of $K_{s,t}$. In light of this, Conlon [5] conjectured that if we assume that a t -degenerate bipartite graph H has no $K_{t,t}$ then we have $\text{ex}(n, H) = O(n^{2-1/t-\delta})$ for some $\delta = \delta(H) > 0$. Lending plausibility to this conjecture, Sudakov and Tomon [25] showed that if all vertices in one of the parts of H have degree at most t but H has no $K_{t,t}$ then $\text{ex}(n, H) = o(n^{2-1/t})$. For $t = 2$ Conlon’s conjecture can be stated as:

Conjecture 1.2 (Conlon [5]). *For every 2-degenerate C_4 -free bipartite graph H there exists a constant $\delta = \delta(H) > 0$ such that*

$$\text{ex}(n, H) = O(n^{3/2-\delta}).$$

There are several results supporting Conjecture 1.2. For example, Conlon and Lee [7] proved that if H is a bipartite graph so that each vertex in one of H ’s sides has maximum degree 2 (such a graph is clearly 2-degenerate) and H is C_4 -free then $\text{ex}(n, H) = O(n^{3/2-\delta})$ for some $\delta = \delta(H) > 0$. Further results in this direction were obtained in [6, 14].

Let $\mathcal{H}_{k,t}$ be the family of 2-degenerate graphs on k vertices and $2k - t$ edges. We raise the following weaker version of Conjecture 1.2.

Conjecture 1.3. *There are absolute constants t, k_0 such that for every $k \geq k_0$ and large enough n , every graph with $\Omega(n^{3/2})$ edges contains a copy of **some** $H \in \mathcal{H}_{k,t}$.*

Let us briefly explain why Conjecture 1.3 is indeed weaker than Conjecture 1.2. It is not hard to see that for every t and large enough k , the family $\mathcal{H}_{k,t}$ contains C_4 -free graphs (see Claim 3.1). Conjecture 1.2 then states that if G has $\Omega(n^{3/2})$ edges then G should contain a copy of **every** $H \in \mathcal{H}_{k,t}$ which is C_4 -free, while Conjecture 1.3 only asks G to contain a copy of **some** $H \in \mathcal{H}_{k,t}$. Note also

that Conjecture 1.3 is weaker than the statement that for every $k \geq k_0$ we have $\text{ex}(n, H) = o(n^{3/2})$ for some $H \in \mathcal{H}_{k,t}$, which is itself weaker than Conjecture 1.2.

Our main result in this paper is the following alternative approach for resolving Conjecture 1.1.

Theorem 1.4. *Conjecture 1.3 implies Conjecture 1.1.*

Before turning to the proof of Theorem 1.4, we mention that it might very well be the case that in Conjecture 1.3 we can replace the lower bound $\Omega(n^{3/2})$ by $\Omega(n^{3/2-\delta})$ for some $\delta = \delta(k) > 0$. Indeed, this bound is implied by Conjecture 1.2. It is not hard to see that in this case the proof of Theorem 1.4 would give that for some absolute constant d and for every e there is $\varepsilon = \varepsilon(e) > 0$ that one can find $(e + d, e)$ -configurations in every 3-graph with $n^{2-\varepsilon}$ edges. Such a result would be an approximate version of a conjecture suggested by Gowers and Long [12], stating that 3-graphs with $n^{2-\varepsilon}$ edges contain $(e + 4, e)$ -configurations.

2 Proof of Theorem 1.4

To avoid confusion, we will refer to edges of a 3-graph as hyperedges. Fix $e \geq 3$ and let \mathcal{G} be a 3-graph with n vertices and $\Omega(n^2)$ hyperedges. We will rely on the well known observation that in the context of the BESC one can assume that \mathcal{G} is linear and 3-partite on vertex sets $(\mathcal{A}, \mathcal{B}, \mathcal{C})$. We now apply a variant of the construction of Solymosi and Solymosi [23]. Given \mathcal{G} , define an auxiliary bipartite multigraph G' as follows. Set $V(G') = (A, B)$ where $A = \binom{\mathcal{A}}{2}$ and $B = \binom{\mathcal{B}}{2}$. For two vertices $\{a_1, a_2\} \in A$ and $\{b_1, b_2\} \in B$ put an edge between them if there is a $c \in \mathcal{C}$ so that $a_1 b_1 c$ and $a_2 b_2 c$ are hyperedges of \mathcal{G} , and (independently) put an edge between them if there is a $c' \in \mathcal{C}$ such that $a_1 b_2 c'$ and $a_2 b_1 c'$ are hyperedges of \mathcal{G} . Since \mathcal{G} is linear, each pair of vertices in G' are connected by at most 2 edges. If we let $d(c)$ denote the degree of a vertex $c \in \mathcal{C}$ in \mathcal{G} then

$$|E(G')| = \sum_{c \in \mathcal{C}} \binom{d(c)}{2} \geq |\mathcal{C}| \left(\frac{1}{|\mathcal{C}|} \sum_{c \in \mathcal{C}} d(c) \right) = |\mathcal{C}| \binom{|E(\mathcal{G})|/|\mathcal{C}|}{2} \geq \frac{|E(\mathcal{G})|^2}{4|\mathcal{C}|}.$$

Since $e(\mathcal{G}) = \Omega(n^2)$, $|\mathcal{C}| \leq n$, and $|V(G')| \leq n^2$, we obtain $|E(G')| = \Omega(|V(G')|^{3/2})$. Since, as noted above, each pair of vertices in G' are connected by at most 2 edges, G' has a simple subgraph G which also contains $\Omega(|V(G')|^{3/2})$ edges. Therefore, if k_0 and t are the constants from Conjecture 1.3 and n is large enough, then we may assume the following.

Claim 2.1. *For every $k_0 \leq k \leq e$, the graph G contains a 2-degenerate bipartite graph F on k vertices with at least $2k - t$ edges.*

We would now like to understand what kind of (v, e) -configuration in \mathcal{G} we get by “unpacking” each of the graphs F in Claim 2.1. Optimistically, if v_1, \dots, v_k is the ordering of $V(F)$ certifying its 2-degeneracy, then every time we add a vertex v_i to v_1, \dots, v_{i-1} of degree 2 to the previous vertices, we expect to get 4 new vertices in \mathcal{G} ; these are c_1, c_2 and either a_1, a_2 (if $v_i \in A$) or b_1, b_2 (if $v_i \in B$). We also expect to get 4 new hyperedges in \mathcal{G} ; these are the 4 hyperedges that correspond to the 2 new edges in G that connect v_i to 2 of the vertices v_1, \dots, v_{i-1} . If this holds for all but a bounded number of F 's vertices, then we will get a $(4k, 4k - O_k(1))$ configuration, hence taking $k \approx e/4$ would finish the proof. Unfortunately, we do not know how to prove such a statement, since in certain cases (see below) some of the 4 vertices/hyperedges might have already appeared when adding one of the previous vertices v_j . Instead, the main idea in Lemma 2.2 below is to show that F gives rise to a $(e' + d, e')$ -configuration, so that if e' is not very close to $4k$ (as in the optimistic analysis above)

then we have $d \leq 0$. It is then easy to show how repeated applications of Lemma 2.2 give Theorem 1.4. In what follows G and \mathcal{G} are those we discussed above.

Lemma 2.2. *Let $k \geq t \geq 4$ be integers, and suppose F is a 2-degenerate subgraph of G with k vertices and $2k - t$ edges. Then \mathcal{G} contains a subgraph \mathcal{F} such that*

- (1) $|V(\mathcal{F})| - 4t \leq |E(\mathcal{F})| \leq 4k$, and
- (2) Either $|E(\mathcal{F})| \geq 4k - 10^4 t^3$ or $|E(\mathcal{F})| \geq |V(\mathcal{F})| > 0$.

We first derive Theorem 1.4 from Lemma 2.2. Assuming Conjecture 1.3 holds with constants t, k_0 we show that Conjecture 1.1 holds with $d = \max\{24k_0, 3(4t + 10^4 t^3)\}$. Indeed, we claim that for every $0 \leq e' \leq e$ we can find e' hyperedges in \mathcal{G} spanned by at most $e' + d$ vertices. If $e' \leq \max\{8k_0, 4t + 10^4 t^3\}$, we just take e' arbitrary hyperedges from \mathcal{G} . For larger e' we apply Lemma 2.2 with the above t and with $k = \lfloor e'/4 \rfloor \geq k_0$ (by Claim 2.1 we know that G contains an F with these parameters). If the lemma returns a configuration \mathcal{F}' whose number of edges satisfies $e' - 10^4 t^3 - 4 \leq |E(\mathcal{F}')| \leq e'$ (and is on at most $e' + 4t$ vertices), we just add to \mathcal{F}' arbitrarily chosen $e' - |E(\mathcal{F}')| \leq 10^4 t^3 + 4$ hyperedges to get a set of e' edges on at most $e' + d$ vertices. Otherwise, we have $|E(\mathcal{F}')| \geq |V(\mathcal{F}')|$ so we can remove \mathcal{F}' from \mathcal{G} and then restart the process with $e'' = e' - |E(\mathcal{F}')|$ (the 3-graph $\mathcal{G} \setminus \mathcal{F}'$ still has $\Omega(n^2)$ hyperedges assuming n is large). We will obtain a set \mathcal{F}'' of e'' hyperedges on at most $e'' + d$ vertices, and can then return $\mathcal{F}'' \cup \mathcal{F}'$ as the set of e' hyperedges on at most $e' + d$ vertices.

Proof of Lemma 2.2. Suppose G contains a subgraph F as above. Let v_1, \dots, v_k be the vertices of F in the order that certifies its 2-degeneracy. For each $i \in [k]$ let $F_i = F[v_1, \dots, v_i]$ be the induced subgraph on the first i vertices. Let $\mathcal{F}_i \subseteq \mathcal{G}$ be a subgraph of \mathcal{G} that corresponds to F_i . That is,

$$V(\mathcal{F}_i) \cap (\mathcal{A} \cup \mathcal{B}) = \{p \in \mathcal{A} \cup \mathcal{B} : \{p, q\} \in V(F_i) \text{ for some } q \in \mathcal{A} \cup \mathcal{B}\},$$

and for every edge uv of F_i , where $u = \{a_1, a_2\}$ and $b = \{b_1, b_2\}$ let $c \in \mathcal{C}$ be the (unique) vertex certifying that $uv \in E(F)$ (in particular $\{a_1 b_1 c, a_2 b_2 c\} \subseteq E(\mathcal{G})$ or $\{a_1 b_2 c, a_2 b_1 c\} \subseteq E(\mathcal{G})$). We include c in $V(\mathcal{F}_i)$ and the corresponding pair of hyperedges in $E(\mathcal{F}_i)$, and applying the same procedure for each edge of F_i we take the union of the resulting hyperedges.

Proof of assertion (1): Initially we have a graph $F_0 := (\emptyset, \emptyset)$ with 0 edges and vertices. Given some $i \in [k]$, let $F^- := F_{i-1}$ and $\mathcal{F}^- := \mathcal{F}_{i-1}$. Suppose without loss of generality that $v_i \in \mathcal{A}$, that is, $v := v_i$ corresponds to a pair $\{a_1, a_2\} \in \binom{\mathcal{A}}{2}$. Let $d(v)$ denote the degree of v in F_i , by our assumptions we have $d(v) \leq 2$. Let $\Delta_E(i) := |E(\mathcal{F}_i) \setminus E(\mathcal{F}^-)|$ and $\Delta_V(i) := |V(\mathcal{F}_i) \setminus V(\mathcal{F}^-)|$.

Note that

$$0 \leq \Delta_E(i) \leq 2d(v) \leq 4,$$

which, summing over all i gives the inequality $|E(\mathcal{F})| \leq 4k$ stated in assertion (1). To prove the second inequality, we need to consider the degree of v : if $d(v) = 2$, let us call v a *regular* vertex, otherwise (if $d(v)$ is 0 or 1) we say that v is *singular*. Accordingly, we are speaking of a regular or singular step i . A crucial observation is that since F is 2-degenerate and has $2k - t$ edges, then the total number of singular steps is at most $2t$.

Suppose first that v is regular, and let u and w be the two neighbours of v in F^- . Let u and w correspond to $\{b_1, b_2\} \in \binom{\mathcal{B}}{2}$ and $\{b_3, b_4\} \in \binom{\mathcal{B}}{2}$ respectively, with $\{b_1, b_2\} \neq \{b_3, b_4\}$ (note that some individual b_1, b_2, b_3, b_4 may coincide). Furthermore, we have vertices $c_1, c_2 \in \mathcal{C}$ such that (after relabelling) $a_1 b_1 c_1, a_2 b_2 c_1, a_1 b_3 c_2$ and $a_2 b_4 c_2$ are hyperedges of \mathcal{F}_i . Note that we must have $c_1 \neq c_2$

for otherwise, by linearity of \mathcal{G} , we would have $b_1 = b_3$ and $b_2 = b_4$, and so $\{b_1, b_2\} = \{b_3, b_4\}$. Since all other hyperedges of \mathcal{F}_i were already contained in \mathcal{F}^- , we have

$$E(\mathcal{F}_i) \setminus E(\mathcal{F}^-) \subseteq \{a_1b_1c_1, a_2b_2c_1, a_1b_3c_2, a_2b_4c_2\}. \quad (2.1)$$

Similarly, $V(\mathcal{F}_i) \setminus V(\mathcal{F}^-) \subseteq \{a_1, a_2, c_1, c_2\}$, and so we also have $0 \leq \Delta_V(i) \leq 4$.

We now claim that $\Delta_V(i) \leq \Delta_E(i)$, and that in fact $\Delta_V(i) < \Delta_E(i)$ when $\Delta_E(i) \in \{1, 2, 3\}$ (this will be used in the proof of assertion (2)). Indeed, if $\Delta_E(i) = 4$, then there is nothing to prove since $\Delta_V(i) \leq 4$. If $\Delta_E(i) = 3$, then without loss of generality the hyperedge $a_1b_1c_1$ was already contained in \mathcal{F}^- . Hence, $\{a_1, c_1\} \subseteq V(\mathcal{F}^-)$, implying $\Delta_V(i) \leq 2$. Similarly, if $\Delta_E(i) = 2$, we have $\Delta_V(i) \leq 1$ (if $a_1b_1c_1$ and $a_2b_2c_1$ were in \mathcal{F}^- then only c_2 can be a new vertex, and if $a_1b_1c_1$ and one of the hyperedges containing c_2 were already in \mathcal{F}^- then only a_2 can be a new vertex), and if $\Delta_E(i) = 1$, then $\Delta_V(i) = 0$ (if only $a_1b_1c_1$ is a new hyperedge then c_1 was added with $a_2b_2c_1$ and a_1 was added with $a_1b_3c_2$). Finally, if $\Delta_E(i) = 0$ then $\{a_1, a_2, c_1, c_2\} \subseteq V(\mathcal{F}^-)$ so $\Delta_V(i) = 0$. So, we obtain $|E(\mathcal{F}_i)| - |V(\mathcal{F}_i)| \geq |E(\mathcal{F}_{i-1})| - |V(\mathcal{F}_{i-1})|$.

If v is singular, a similar case analysis shows that $|E(\mathcal{F}_i)| - |E(\mathcal{F}_{i-1})| \geq |V(\mathcal{F}_i)| - |V(\mathcal{F}_{i-1})| - 2$. Since there are at most $2t$ singular steps in total, summing over all i yields $|E(\mathcal{F})| \geq |V(\mathcal{F})| - 4t$ as desired.

Proof of assertion (2): In order to prove the second assertion we need to study the above process in more detail.

Suppose a step i is regular. If $\Delta_E(i) = 0$ we call it a *0-step*, if $\Delta_E(i) = \Delta_V(i) = 4$ we say this is a *4-step*. If we have $\Delta_V(i) < \Delta_E(i)$, then we call this step a *good* regular step. Note that by the argument in the paragraph following (2.1), every regular step which is not a 0-step or a 4-step is a good step. Note also that at each good regular step the difference $|E(\mathcal{F}_i)| - |V(\mathcal{F}_i)|$ strictly increases and, as we have seen in the proof of (1), this difference decreases only at singular steps, in which it decreases by at most 2. Hence, if the total number of good regular steps is at least $4t$ we would have $|E(\mathcal{F})| \geq |V(\mathcal{F})| > 0$ as needed. So let us assume for the rest of the proof that we have fewer than $4t$ good regular steps. Let us say that a (regular or singular) step is *good* if it is either good regular in the above sense or singular. So, the total number of good steps is less than $6t$.

If the number of 0-steps is at most $s := 6t(12t + 2)^2$, then all but $s + 6t$ of the steps are 4-steps and so we have $|E(\mathcal{F})| \geq 4k - 4(s + 6t) \geq 4k - 10^4t^3$ as needed. So suppose towards contradiction that this is not the case, i.e., that the number of 0-steps is greater than s . We will now show that this means that the total number of good and steps is at least $6t$, contradicting the statement made in the previous paragraph.

We say that a vertex $c \in \mathcal{C}$ is *involved* in step i (or equivalently, step i *involves* c) if c plays the role of either c_1 or c_2 in the extension of \mathcal{F}_{i-1} to \mathcal{F}_i described above. Note that each regular step involves precisely two vertices of \mathcal{C} . Similarly, we say that a hyperedge $e \in E(\mathcal{G})$ is involved in step i if it plays the role of one of the hyperedges arising in the extension of \mathcal{F}_{i-1} to \mathcal{F}_i (we stress that this is regardless of whether e had already been contained in \mathcal{F}_{i-1}).

Observation 2.3. *A pair of hyperedges $e_1 = a_1b_1c$ and $e_2 = a_2b_2c$, where $a_1, a_2 \in A, b_1, b_2 \in B, c \in C$, can simultaneously be involved in at most one step.*

Indeed, for every step i involving both hyperedges there must be vertices $u, w \in V(F)$ with $u = \{a_1, a_2\}$ and $w = \{b_1, b_2\}$ such that one of u and w is the vertex v_i and the other is v_j for some $j < i$.

We now claim that every 0-step involving some vertex $c \in \mathcal{C}$ must be preceded by a good step involving c . Indeed, suppose that c is involved in a 0-step at time i . Suppose that v_i repre-

sents some $\{a_1, a_2\} \in \binom{A}{2}$ with $E(F_i) \setminus E(F_{i-1}) = \{v_i u, v_i w\}$ for some $u, w \in V(F)$ representing $\{b_1, b_2\}, \{b_3, b_4\} \in \binom{B}{2}$ respectively (the case when $v_i \in B$ is identical), and that the hyperedges of \mathcal{G} certifying that $\{v_i u, v_i w\} \subseteq E(F_i)$ (after relabelling) are $\{a_1 b_1 c, a_2 b_2 c, a_1 b_3 c', a_2 b_4 c'\}$ for some $c' \in \mathcal{C}$ (and note that since this is a 0-step, all these hyperedges are already contained in $E(\mathcal{F}_{i-1})$). Let j_1 be the first step involving the hyperedge $e_1 := a_1 b_1 c$, i.e. $j_1 < i$ is the unique j such that $e_1 \in E(\mathcal{F}_j) \setminus E(\mathcal{F}_{j-1})$. Let j_2 be defined analogously with respect to $e_2 := a_2 b_2 c$. If step j_1 or j_2 are singular, we have proved the claim (since singular steps are good by definition). So, let us assume they are both regular. If $j_1 \neq j_2$ then at time $\max(j_1, j_2)$ (say, this is j_2) we have a good step involving c , since $\Delta_V(j_2) < 4$ yet $\Delta_E(j_2) \geq 1$, so this cannot be a 0-step or a 4-step and thus must be a good step. On the other hand we cannot have $j_1 = j_2$ since that would mean both e_1 and e_2 would be involved in two different steps, contradicting Observation 2.3. This proves the above claim.

Now let $\mathcal{Z} \subseteq \mathcal{C}$ be the set of all vertices in \mathcal{C} involved in 0-steps. Suppose first that $|\mathcal{Z}| > 12t$. Then, as for every $z \in \mathcal{Z}$ each 0-step involving z is preceded by a good step also involving z , the number of vertices of \mathcal{C} involved in good steps is greater than $12t$. Since every step involves at most 2 vertices of \mathcal{C} , we obtain that the total number of good steps is greater than $6t$, as needed.

So, let us assume that $|\mathcal{Z}| \leq 12t$. Then, by pigeonhole, some $z \in \mathcal{Z}$ was involved in at least $(2s)/(12t) = (12t + 2)^2$ of the 0-steps. This implies that z must be contained in at least $12t + 2$ hyperedges of \mathcal{F} , as each 0-step involving z involves two hyperedges containing z , and no such pair may be involved twice by Observation 2.3.

Let now $J \subset [k]$ be the set of all $j \in [k]$ such that at step j for some hyperedge $e \in \mathcal{F}$ with $z \in e$ we have $e \in E(\mathcal{F}_j) \setminus E(\mathcal{F}_{j-1})$. Since at any given step j we can have at most 2 such hyperedges e , we have $|J| \geq (12t + 2)/2 = 6t + 1$. On the other hand for every step in $j \in J$ except $j_0 = \min J$ we have $\Delta_V(j) < 4$, since $z \in \mathcal{F}_{j_0}$, and $\Delta_E(j) > 0$, by definition of J . This means that each of these $|J| - 1 \geq 6t$ steps is not a 0-step or a 4-step, and therefore must be a good step.

We have thus shown that if the number of 0-steps is at most s then the number of good steps is at least $6t$, which completes the proof of the lemma. \square

3 C_4 -free graphs in $\mathcal{H}_{k,t}$

We say that a graph is *exactly-(2, t)-degenerate* if it can be obtained from a set of t isolated vertices by repeatedly adding new vertices of degree exactly 2. Note that every exactly-(2, t)-degenerate graph belongs to $\mathcal{H}_{k,t}$. The following claim shows that $\mathcal{H}_{k,t}$ contains not only C_4 -free graphs, but in fact graphs of arbitrary large girth.

Claim 3.1. *For every g there is $t = t(g)$ so that for every $k \geq t$, there is a k -vertex exactly-(2, t)-degenerate bipartite graph of girth at least g .*

Proof. We claim that starting with an independent set of size $t = t(g)$ (to be chosen later), we can repeatedly add vertices so that each k -vertex graph in the sequence is exactly-(2, t)-degenerate, bipartite, of girth at least g , and in addition satisfies the following two conditions: (i) it has maximum degree at most 8 and (ii) it has a bipartition into two set of sizes $\lceil k/2 \rceil$ and $\lfloor k/2 \rfloor$. The initial independent set under a balanced bipartition clearly satisfies these two conditions, so let us show how to add a vertex and maintain them. Suppose the graph has $k - 1$ vertices and bipartition into sets A, B satisfying $|A| \leq |B|$. Since it has maximum degree at most 8, it contains $O(k)$ pairs of vertices connected by a path of length at most $g - 2$. Since the average degree of the vertices in B is less than 4, at least half the vertices have degree at most 7. Hence, at least $\binom{k-1}{2}^{1/4} \geq \frac{k^2}{50}$ of the

pairs of vertices in B both have degree at most 7. Assuming t is large enough so that $k \geq t$ satisfies $\frac{k^2}{50} - O(k) > 1$, we thus have a pair of vertices $u, v \in B$ so that both of them have degree at most 7 and there is no path of length at most $g - 2$ connecting them. Hence, we can add a new vertex to A and connect it to u and v . \square

Acknowledgement: We would like to thank David Conlon for useful discussions.

References

- [1] N. Alon, M. Krivelevich, and B. Sudakov. Turán numbers of bipartite graphs and related Ramsey-type questions, *Comb. Probab. Comput* 12 (2003), 477–494. [1.2](#)
- [2] W. G. Brown, P. Erdős and V.T. Sós, Some extremal problems on r -graphs, *New Directions in the Theory of Graphs, Proc. 3rd Ann Arbor Conference on Graph Theory*, Academic Press, New York, 1973, 55-63. [1.1](#)
- [3] W. Brown, P. Erdős and V. Sós. On the existence of triangulated spheres in 3-graphs, and related problems, *Periodica Mathematica Hungarica*, 3(3–4) (1973), 221-228. [1.1](#)
- [4] B. Bukh, Extremal graphs without exponentially-small bicliques, manuscript 2022. [1.2](#)
- [5] D. Conlon, private communication, 2022. [1.2](#), [1.2](#)
- [6] D. Conlon, O. Janzer and J. Lee, More on the extremal number of subdivisions, *Combinatorica* 41 (2021), 465-494. [1.2](#)
- [7] D. Conlon and J. Lee, On the extremal number of subdivisions, *Int. Math. Res. Not.* 2021, 9122-9145. [1.2](#)
- [8] D. Conlon, L. Gishboliner, Y. Levanzov and A. Shapira, A new bound for the Brown–Erdős–Sós problem, *J. Combin. Theory Ser. B.* 158 (2023), 1-35. [1.1](#)
- [9] P. Erdős, Some recent results on extremal problems in graph theory. *Results, Theory of Graphs (Internat. Sympos., Rome, 1966)*, pages 117–123, 1967. [1.2](#)
- [10] Z. Füredi, On a Turán type problem of Erdős, *Combinatorica* 11 (1991), 75–79. [1.2](#)
- [11] W. T. Gowers, Hypergraph regularity and the multidimensional Szemerédi theorem, *Ann. of Math.* 166 (2007), 897–946. [1.1](#)
- [12] W. T. Gowers and J. Long, The length of an s -increasing sequence of r -tuples, *Combin. Probab. Comput.* 30 (2021), 686–721. [1.2](#)
- [13] A. Grzesik, O. Janzer and Z. L. Nagy, The Turán number of blow-ups of trees, *J. Combin. Theory Ser. B.* 156 (2022), 299-309. [1.2](#)
- [14] O. Janzer, The extremal number of the subdivisions of the complete bipartite graph, *SIAM J. Discrete Math.* 34 (2020), 241-250. [1.2](#)
- [15] O. Janzer, Disproof of a conjecture of Erdős and Simonovits on the Turán number of graphs with minimum degree 3, *Int. Math. Res. Not.*, to appear. [1.2](#)

- [16] T. Kővári, V. T. Sós and P. Turán, On a problem of K. Zarankiewicz, *Colloquium Math.* 3 (1954), 50-57. [1.1](#), [1.2](#)
- [17] B. Nagle, V. Rödl and M. Schacht, The counting lemma for regular k -uniform hypergraphs, *Random Structures Algorithms* 28 (2006), 113–179. [1.1](#)
- [18] K.F. Roth, On certain sets of integers, *J. London Math. Soc.* 28 (1953), 104-109. [1.1](#)
- [19] V. Rödl and J. Skokan, Regularity lemma for k -uniform hypergraphs, *Random Structures Algorithms* 25 (2004), 1–42. [1.1](#)
- [20] V. Rödl and J. Skokan, Applications of the regularity lemma for uniform hypergraphs, *Random Structures Algorithms* 28 (2006), 180–194. [1.1](#)
- [21] I. Ruzsa and E. Szemerédi, Triple systems with no six points carrying three triangles, in *Combinatorics (Keszthely, 1976)*, *Coll. Math. Soc. J. Bolyai* 18, Volume II, 939-945. [1.1](#)
- [22] G. N. Sárközy and S. Selkow, An extension of the Ruzsa-Szemerédi theorem, *Combinatorica* 25 (2004), 77-84. [1.1](#)
- [23] D. Solymosi and J. Solymosi, Small cores in 3-uniform hypergraphs, *J. Combin. Theory Ser. B.* 122 (2017), 897-910. [1.1](#), [2](#)
- [24] E. Szemerédi, Regular partitions of graphs, In: *Proc. Colloque Inter. CNRS*, 1978, 399-401. [1.1](#)
- [25] B. Sudakov and I. Tomon, Turán number of bipartite graphs with no $K_{t,t}$, *Proc. Amer. Math. Soc.* 148 (2020), 2811-2818. [1.2](#)
- [26] T. Tao, A variant of the hypergraph removal lemma, *J. Combin. Theory Ser. A* 113 (2006), 1257–1280. [1.1](#)