

On graphs decomposable into induced matchings of linear sizes

Jacob Fox, Hao Huang and Benny Sudakov

ABSTRACT

We call a graph G an (r, t) -Ruzsa–Szemerédi graph if its edge set can be partitioned into t edge-disjoint induced matchings, each of size r . These graphs were introduced in 1978 and have been extensively studied since then. In this paper, we consider the case when $r = cn$. For $c > 1/4$, we determine the maximum possible t , which is a constant depending only on c . On the other hand, when $c = 1/4$, there could be as many as $\Omega(\log n)$ induced matchings. We prove that this bound is tight up to a constant factor. Finally, when c is fixed strictly between $1/5$ and $1/4$, we give a short proof that the number t of induced matchings is $O(n/\log n)$. We are also able to further improve the upper bound to $o(n/\log n)$ for fixed $c > 1/4 - b$ for some positive constant b .

1. Introduction

We call a graph $G = (V, E)$ an (r, t) -Ruzsa–Szemerédi graph, or (r, t) -RS graph for short, if its edge set $E(G)$ can be partitioned into t pairwise edge-disjoint induced matchings M_1, \dots, M_t , each consists of r edges. These graphs were first introduced in the famous paper by Ruzsa and Szemerédi [18], in which they show that there is no n -vertex (r, t) -RS graph for r, t , both linear in n . They also proved that this result implies the celebrated theorem of Roth [17], a subset S of $[n] = \{1, \dots, n\}$ without nontrivial 3-term arithmetic progressions has size at most $o(n)$. Ruzsa and Szemerédi also constructed $(n/e^{O(\sqrt{\log n})}, n/3)$ -RS graphs on n vertices, based on the result of Behrend [5] that there is a surprisingly large subset of $[n]$ without nontrivial 3-term arithmetic progressions. They used these graphs together with the regularity lemma of Szemerédi [20] to solve an extremal problem of Brown, Erdős and Sós [8, 9], showing that the maximum number of edges in a 3-uniform n -vertex hypergraph, in which no six vertices span at least three edges, is at least $n^{2-o(1)}$ and at most $o(n^2)$ for sufficiently large n . In this paper, we consider the following natural question.

PROBLEM 1.1. For which values of r and t does there exist an (r, t) -RS graph on n vertices?

This question has been studied in depth for decades, and has found several applications in combinatorics, complexity theory, and information theory (see, for example, [1–4, 11, 14]). One line of research was to find very dense RS graphs decomposable into large induced matchings. An early result of this type by Frankl and Füredi [13] implies that for any fixed r there are (r, t) -RS graphs on n vertices with $rt = (1 - o(1))\binom{n}{2}$ edges. However, the techniques employed in their proof cannot provide induced matchings of size larger than $\Theta(\log n)$. Later, Birk, Linial and Meshulam [6] noted that such very dense (r, t) -RS graphs with $r \rightarrow \infty$ are useful for designing a communication protocol over a shared directional multichannel. For this application, they construct (r, t) -RS graphs on n vertices with $r = (\log n)^{\Omega(\log \log n / (\log \log \log n)^2)}$ and rt is roughly equal to $n^2/24$. Note that none of the constructions mentioned so far gives an

Received 21 December 2015; revised 6 June 2016; published online 29 December 2016.

2010 *Mathematics Subject Classification* 05C35, 05C70 (primary).

Jacob Fox was supported by a Packard Fellowship, NSF Career Award DMS-135212, and an Alfred P. Sloan Fellowship. Hao Huang was supported in part by Simons Collaboration Grant. Benny Sudakov was supported in part by SNSF grant 200021-149111.

n -vertex (r, t) -RS graph with positive edge density and at the same time being an edge-disjoint union of induced matchings of size n^ε for some constant ε . This range of parameters is important for some applications when there is a tradeoff between the number of missing edges and the number of induced matchings. Indeed, Meshulam (private communication) conjectures that there exist no such graphs. This conjecture was recently disproved by Alon, Moitra and Sudakov [2] with a surprising construction of graphs with edge density $1 - o(1)$ that can be partitioned into induced matchings of size $n^{1-o(1)}$, which is nearly linear in n . Their constructions have also found a couple of interesting applications for communication problems over shared channels, linearity testing, communication complexity and the directed Steiner tree problem.

Another appealing direction is to determine the maximum number of induced matchings when their sizes are linear in n . Fischer *et al.* [11] gave a family of $(n/6 - o(n), n^{\Omega(1/\log \log n)})$ -RS graphs. They used these graphs to establish an $n^{\Omega(1/\log \log n)}$ lower bound for testing monotonicity in general posets. A slight twist of their construction gives $(cn, n^{\Omega(1/\log \log n)})$ -RS graphs for any positive constant $c < 1/4$. In this paper, we first study the range $c > 1/4$. It was noted earlier by Alon (private communication) that for this range, there can only be constantly many matchings. Our first result determines the maximum number of induced matchings for $c > 1/4$. Its proof combines techniques from coding theory together with properties of Kneser graphs.

THEOREM 1.2. *Suppose G is an (r, t) -RS graph on n vertices, then*

$$r \leq \begin{cases} \frac{n}{4} \left(1 + \frac{1}{t}\right) & \text{if } 2 \nmid t, \\ \frac{n}{4} \left(1 + \frac{1}{t+1}\right) & \text{if } 2 \mid t. \end{cases} \quad (1)$$

Moreover, these bounds are tight for every positive integer t and infinitely many n .

For $r = n/4$, it was known that there are RS graphs with logarithmically many induced matchings [16], but there were no good upper bounds in this case. Our second theorem shows that up to a constant factor, the maximum number of induced matchings when $r = n/4$ is indeed logarithmic.

THEOREM 1.3. *If an n -vertex graph G is an $(n/4, t)$ -RS graph, then $t \leq (6 + o(1)) \log_2 n$. Moreover, if such graph G is regular, then $t \leq 2 \log_2 n + 2$ and this bound is sharp for infinitely many n .*

Problem 1.1 becomes much more difficult when the size $r = cn$ of the induced matching is below the threshold $n/4$. For this range, the best-known upper bound is due to Fox [12], who improved estimates that one obtains from the Szemerédi regularity lemma [20]. His result implies that the number of matchings is at most $n/\log^{(x)} n = o(n)$, where $x = O(\log(1/c))$ and $\log^{(x)} n$ denote the x -fold iterated logarithm. This shows that the number of induced matchings has to be sublinear. On the other hand, the example of Fischer *et al.* [11] shows that for $c < 1/4$, the number of matchings could be $n^{\Omega(1/\log \log n)}$, which is much larger than $\log n$ but smaller than n^ε . It would be very interesting to close the huge gap between the upper and lower bounds for some ranges of $c < 1/4$. Here, we are able to make some progress on this problem and improve the upper bound to $O(n/\log n)$, when c is fixed strictly between $1/5$ and $1/4$.

THEOREM 1.4. *For every $\varepsilon > 0$, if G is an (r, t) -RS graph on n vertices with $r = cn$ for $1/5 + \varepsilon \leq c < 1/4$, then $t = O(n/\log n)$.*

By a more careful analysis, we are able to further improve the upper bound to $o(n/\log n)$, when c is slightly below $1/4$. To be more precise, the little- o term comes from the best-known bound in the triangle removal lemma proved in [12]. Here, \log^* is the iterated logarithm, that is, the number of times the logarithm function must be iteratively applied before the result is less than or equal to 1.

THEOREM 1.5. *There exists an absolute constant $b > 0$, such that for $r \geq (1/4 - b)n$, if G is an (r, t) -RS graph on n vertices, then $t = n/((\log n)2^{\Omega(\log^* n)}) = o(n/\log n)$.*

The rest of the paper is organized as follows. In the next section, we describe a counting proof of the upper bound in Theorem 1.2 and verify its tightness by a construction based on Kneser graphs. In Section 3, we prove Theorem 1.3 and discuss potential improvements. Section 4 consists of the proof of Theorems 1.4 and 1.5, and some remarks on the case of regular graphs. The last section contains some concluding remarks and open problems.

2. The size of matchings is cn with $c > 1/4$

In this section, we determine the maximum number of induced matchings in an (r, t) -RS graph G for $r = cn$ and $c > 1/4$. To prove an upper bound, we use ideas similar to the Plotkin bound [15] in coding theory, with a small twist. The lower bound construction is based on the properties of Kneser graphs.

Proof of Theorem 1.2. Suppose the edge set of G can be partitioned into induced matchings M_1, \dots, M_t , each containing exactly r edges. Denote by V_i the set of vertices contained in the edges of M_i . Then $|V_i| = 2r$. Moreover, each of the r edges of M_i intersects V_j in at most one vertex, since otherwise V_i and V_j must span a common edge of G . This implies that $|V_i \cap V_j| \leq r$. Let $v_i \in \{0, 1\}^n$ be the characteristic vector of V_i . Then for all $1 \leq i < j \leq t$, the Hamming distance satisfies

$$\text{dist}(v_i, v_j) = |V_i| + |V_j| - 2|V_i \cap V_j| \geq 2r + 2r - 2r = 2r.$$

This is already enough to show that t is constant, using bounds from coding theory. But one can do slightly better. Let v_0 be the all-zero vector. To get a tight upper bound, note that the above inequality can be extended to all $0 \leq i < j \leq t$ since, for $1 \leq i \leq t$, $|V_i| = 2r$. Denote by a_i (respectively, b_i) the number of vectors v_j equal to 0 (respectively, 1) in the i th coordinate, then $a_i + b_i = t + 1$. By double counting,

$$\begin{aligned} 2r \binom{t+1}{2} &\leq \sum_{0 \leq i < j \leq t} \text{dist}(v_i, v_j) \\ &= \sum_{i=1}^n a_i b_i \\ &\leq \begin{cases} n(t+1)^2/4 & \text{if } 2 \nmid t, \\ nt(t+2)/4 & \text{if } 2 \mid t. \end{cases} \end{aligned} \tag{2}$$

The last inequality follows from the fact that $a_i b_i$ is maximized when $a_i = b_i = (t+1)/2$ for odd t , and $\{a_i, b_i\} = \{(t+2)/2, t/2\}$ for even t . By simplifying the inequality, we immediately obtain (1).

To show that the bound is tight, it suffices to consider the case $t = 2k + 1$. Let H be $KG(2k + 1, k)$, the Kneser graph whose vertices correspond to all the k -subsets of a set of $2k + 1$ elements, and where two vertices are adjacent if and only if the two corresponding sets are disjoint. We define the matchings M_1, \dots, M_{2k+1} as follows: the edge (A, B) belongs to M_i , if and only

if $A \cap B = \emptyset$ and $A \cup B = [2k+1] \setminus \{i\}$. It is easy to see that B is determined after fixing A and i , which implies that M_i forms a matching. In order to show that every matching is induced in H , we take (A, B) and (C, D) both from M_i with $A \neq C, D$, then $A \cup B = C \cup D = [2k+1] \setminus \{i\}$, it is not hard to check that $A \cap C, A \cap D$ are both nonempty and therefore (A, C) and (A, D) are not contained in any M_j . By calculation, $n = \binom{2k+1}{k}$, while $r = \frac{1}{2} \binom{2k}{k} = \frac{n}{4} (1 + \frac{1}{2k+1})$. Hence, H is an $(\frac{n}{4}(1 + \frac{1}{2k+1}), 2k+1)$ -RS graph on n vertices. \square

REMARK. By taking m vertex-disjoint copies of the Kneser graph $KG(2k+1, k)$, we can construct an infinite family of n -vertex $(r, 2k+1)$ -RS graphs with $r = n(1 + 1/t)/4$. Moreover, solving $n = \binom{2k+1}{k}$ gives $k \sim \frac{1}{2} \log_2 n + \Theta(\log \log n)$, therefore we obtain an $((n/4) + \Theta(n/\log n), \log_2 n + \Theta(\log \log n))$ -RS graph H on n vertices.

3. The size of matchings is $n/4$

From Theorem 1.2, we know that in an RS graph, if every matching has size $r = cn$ where $c > 1/4$, then the number of matchings is at most a constant that depends on c . On the other hand, when $c = 1/4$, we saw in the end of the previous section that one can have logarithmic number of induced matching. The following very natural construction [16] does better and gives $2 \log_2 n$ induced matchings.

PROPOSITION 3.1. *There exists an $(n/4, 2 \log_2 n)$ -RS graph for every integer n that is the power of 2.*

Proof. Let $k = \log_2 n$ and consider the k -dimensional hypercube graph H with vertex set $\{0, 1\}^k$, where two vectors are adjacent if their Hamming distance is 1. We first partition the vertices into odd and even vectors, according to the parity of the sum of their coordinates. For $1 \leq i \leq k$, we let the i th matching M_i consist of edges between vectors \vec{v} and $\vec{v} + \vec{e}_i$, such that \vec{v} is even and its i th coordinate is 0; and the $(k+i)$ th matching M_{k+i} consist of edges between an odd vector \vec{v} whose i th coordinate equals 0 and the vector $\vec{v} + \vec{e}_i$. This construction gives $2k = 2 \log_2 n$ matchings, and obviously each matching involves exactly half of the vertices. In order to verify that the matchings are induced, we consider two distinct edges from M_i , which are $(\vec{u}, \vec{u} + \vec{e}_i)$ and $(\vec{v}, \vec{v} + \vec{e}_i)$, such that both \vec{u} and \vec{v} are even and their i th coordinates are 0. Clearly, the pairs (\vec{u}, \vec{v}) and $(\vec{u} + \vec{e}_i, \vec{v} + \vec{e}_i)$ cannot form edges of H since they have the same parity. Moreover, \vec{u} and $\vec{v} + \vec{e}_i$ (similarly, \vec{v} and $\vec{u} + \vec{e}_i$) differ in at least two coordinates. Therefore, for all $1 \leq i \leq k$, the matchings M_i , we defined, are induced. Using a similar argument, we can also show that the matchings M_{i+k} are induced. \square

When $\log_2 n$ is an even integer, we can slightly improve the above construction by adding two additional induced matchings. The first matching consists of (\vec{u}, \vec{v}) , where $\vec{u} + \vec{v} = \vec{1}$ and both are even vectors. The second matching contains all edges (\vec{u}, \vec{v}) , where $\vec{u} + \vec{v} = \vec{1}$ and both are odd vectors. This gives the following corollary.

COROLLARY 3.2. *There exist $(n/4, 2(\log_2 n + 1))$ -RS graphs on n vertices for every n that is an even power of 2.*

One should naturally ask how good are these constructions, that is, what upper bound can we prove on the number t of induced matchings in an n -vertex (r, t) -RS graph with $r = n/4$. Note that in the proof of Theorem 1.2, when $r = n/4$, the inequalities in (2) are always true for every t . Indeed, in this case there might be linearly many n -dimensional binary vectors such that their pairwise Hamming distance is equal to $n/2$ by modifying the well-known Hadamard matrix into a $(0, 1)$ -matrix and taking the row vectors. In the rest of this section, we develop a

different approach which shows that when every induced matching has size exactly $n/4$, there can be no more than $O(\log n)$ matchings. Therefore, the above constructions are best possible up to a constant factor. If we further assume that the RS graph is regular, then the construction in Corollary 3.2 is optimal. Below is a quick outline of our proof.

First, we show that in any graph G , which is decomposable into induced matchings of size $n/4$, there are many pairs of adjacent vertices whose sum of degrees is large. This gives an auxiliary graph H , which is the subgraph of G whose edges have large degree sum. Using additional properties of induced matchings, we can show that H has nice expansion properties. From there we can easily derive some estimates on the number of induced matchings. To illustrate this idea, we first prove a slightly weaker upper bound. The following lemma allows us to only consider bipartite graphs for this case.

LEMMA 3.3. *If G is an (r, t) -RS graph on n vertices, then its bipartite double cover $G \times K_2$ is an $(2r, t)$ -RS graph on $2n$ vertices.*

Proof. Denote by $G' = G \times K_2$ the bipartite double cover of the graph G . The vertices of G' are (v, i) with $v \in V(G)$ and $i \in \{0, 1\}$. Two vertices $(u, 0)$ and $(v, 1)$ are adjacent whenever u and v form an edge in G . Note that an induced matching $M_i = \{(u_j, v_j)\}_{j=1}^r$ in G corresponds to a matching $M' = \{((u_j, 0), (v_j, 1))\}_{j=1}^r \cup \{((u_j, 1), (v_j, 0))\}_{j=1}^r$ in G' , which is of size $2r$. It is also straightforward to check by definition that M' is an induced matching. Therefore, G' is an $(2r, t)$ -RS graph on $2n$ vertices. \square

THEOREM 3.4. *If G is an $(\frac{n}{4}, t)$ -RS graph on n vertices, then $t \leq 8(\log_2 n + 1)$.*

Proof. From Lemma 3.3, it suffices to show that for all n -vertex bipartite graphs G whose edges can be decomposed into induced matchings M_1, \dots, M_t , each of size $n/4$, t is at most $8 \log_2 n$.

Denote by d_v the degree of vertex v in G . We consider the subgraph H of G , with edges being the pair of vertices u, v such that $d_u + d_v \geq t$ and $(u, v) \in E(G)$. Note that $e(G) = \frac{nt}{4}$. By the Cauchy–Schwarz inequality,

$$\begin{aligned} \sum_{(u,v) \in E(G)} (d_u + d_v - t) &= \left(\sum_{v \in V(G)} d_v^2 \right) - t \cdot e(G) \geq n \left(\frac{\sum_{v \in V(G)} d_v}{n} \right)^2 - \frac{nt^2}{4} \\ &= n \left(\frac{nt/2}{n} \right)^2 - \frac{nt^2}{4} = 0. \end{aligned} \quad (3)$$

For any edge $(u, v) \in E(G)$, all the edges incident to either u or v must belong to different matchings. Therefore, $d_u + d_v \leq t + 1$. If we denote by E_i the number of edges (u, v) such that $d_u + d_v = t + i$, we have $E_1 + E_0 + E_{-1} + \dots + E_{-t} = nt/4$, while inequality (3) implies that $E_1 - \sum_{j=1}^t jE_{-j} \geq 0$. Summing these two inequalities gives $2E_1 + E_0 \geq nt/4$ and so $E_1 + E_0 \geq nt/8$. So H is a n -vertex graph with at least $nt/8$ edges and thus its average degree is at least $t/4$. Hence, H has a subgraph F of minimum degree at least $t/8$.

Set $s = t/8$. For each vertex v of G , let A_v denote the set of induced matchings containing v . Clearly, $|A_v| = d_v$. We claim that if v and u are at distance k in F , then when k is odd, $|A_u \cap A_v| \leq k$; and when k is even, $|A_u \cap A_v^c| \leq k$. This statement can be proved using induction. The base cases when $k = 0$ and 1 are obvious. Now we assume that it is true for all $k \leq i$. For $k = i + 1$, suppose u and v are at distance k . Let w be a vertex at distance 1 from v and $i = k - 1$ from u . When i is odd, from the inductive hypothesis we have $|A_v^c \cap A_w^c| = t - |A_v \cup A_w| = t - |A_v| - |A_w| + |A_v \cap A_w| \leq 1$ and $|A_w \cap A_u| \leq i$. Therefore,

$$|A_u \cap A_v^c| \leq |A_u \cap A_w| + |A_w^c \cap A_v^c| \leq i + 1.$$

Similarly, when i is even, we have $|A_v \cap A_w| \leq 1$ and $|A_w^c \cap A_u| \leq i$, and hence

$$|A_u \cap A_v| \leq |A_v \cap A_w| + |A_w^c \cap A_u| \leq i + 1.$$

Now choose an arbitrary vertex v in F , the degree of v in F is at least $s = t/8$. For every integer $i \geq 0$, let N_i be the set of vertices at distance i from v in graph F . By the assumption that G is bipartite, each N_i induces an independent set. We denote by e_i the number of edges of F that are between N_i and N_{i+1} and contained in matchings in A_v (respectively, A_v^c) when i is odd (respectively even). For odd i , we estimate the number of edges of F between N_i and N_{i+1} that are contained in matchings in A_v^c in two different ways. Note that every vertex $u \in N_i$ is contained in at least $s - |A_u \cap A_v| \geq s - i$ edges in A_v^c that are between N_i , and either N_{i-1} or N_{i+1} . On the other hand, every vertex $w \in N_{i+1}$ is contained in no more than $|A_w \cap A_v^c| \leq i + 1$ edges in A_v^c that are between N_{i+1} , and either N_i or N_{i+2} , therefore we have

$$(s - i)|N_i| - e_{i-1} \leq (i + 1)|N_{i+1}| - e_{i+1}.$$

Similarly when i is even, by bounding the number of edges between N_i and N_{i+1} that belong to matchings in A_v , we obtain the same inequality as above.

Summing up the inequalities for $i = 0, \dots, k$, we have

$$\sum_{i=0}^k (s - i)|N_i| - \sum_{i=0}^{k-1} e_i \leq \sum_{i=1}^{k+1} i|N_i| - \sum_{i=1}^{k+1} e_i.$$

Simplifying this inequality gives

$$(k + 1)|N_{k+1}| \geq \sum_{i=0}^k (s - 2i)|N_i| - e_0 + e_{k+1} + e_k \geq \sum_{i=0}^k (s - 2i)|N_i|.$$

The second inequality follows from the observation that $e_0 = 0$ since all edges between N_0 and N_1 are in A_v .

In the next step, we prove by induction that $|N_i| \geq \binom{s}{i}$. For $i = 0$ and 1 , this is obvious. Now, assuming it is true for all $i \leq k$, we have

$$\begin{aligned} |N_{k+1}| &\geq \frac{1}{k+1} \sum_{i=0}^k (s - 2i)|N_i| \geq \frac{1}{k+1} \sum_{i=0}^k (s - 2i) \binom{s}{i} \\ &= \frac{1}{k+1} \left(\sum_{i=0}^k (s - i) \binom{s}{i} - \sum_{i=0}^k i \binom{s}{i} \right) \\ &= \frac{1}{k+1} \left(\sum_{i=0}^k s \binom{s-1}{i} - \sum_{i=0}^k s \binom{s-1}{i-1} \right) \\ &= \frac{s}{k+1} \binom{s-1}{k} = \binom{s}{k+1}. \end{aligned}$$

Note that the number of vertices in $N_0 \cup N_1 \cup \dots \cup N_s$ is at most n . We therefore have

$$n \geq \sum_{k=0}^s |N_k| \geq \sum_{k=0}^s \binom{s}{k} = 2^s.$$

Solving this inequality gives $s \leq \log_2 n$, and hence $t \leq 8 \log_2 n$. \square

As we mentioned in the introduction, this bound can be further improved to $(6 + o(1)) \log_2 n$ by the following modification of the above proof.

Proof of Theorem 1.3. Recall that the proof above uses the fact that H contains $E_1 + E_0$ edges and therefore it contains a subgraph F with minimum degree at least $(E_1 + E_0)/n$, and later we obtained the inequality $n \geq 2^{(E_1 + E_0)/n}$. Let H' be the subgraph of G consisting of edges (u, v) such that $d_u + d_v = t + 1$, so $e(H') = E_1$. Graph H' contains a subgraph F' of minimum degree at least $E_1/n = s'$. By an argument similar to the one in the previous proof, we can show that if a vertex u is at distance k from v , then $|A_u \cap A_v| \leq (k + 1)/2$ if k is odd, and $|A_u \cap A_v| \leq k/2$ if k is even. Therefore, if we fix a vertex v and let N'_i be the set of vertices having distance i from v , and define e'_i in a similar way, we have

$$(s' - \lceil i/2 \rceil)|N'_i| - e'_{i-1} \leq \lceil (i + 1)/2 \rceil |N'_{i+1}| - e'_{i+1}.$$

Summing up the inequalities from $i = 0$ to k , and using $e'_{-1} = e'_0 = 0$ and $e'_k, e'_{k+1} \geq 0$, gives

$$\lceil (k + 1)/2 \rceil |N'_{k+1}| \geq \sum_{i=0}^k (s' - 2\lceil i/2 \rceil) |N'_i|.$$

We can use induction to show that $|N'_{2i}| \geq \binom{s'}{i} \binom{s'-1}{i}$ and $|N'_{2i+1}| \geq \binom{s'}{i+1} \binom{s'-1}{i}$. The $i = 0$ case follows from $|N'_0| \geq 1$ and $|N'_1| \geq s'$. Moreover, using the inductive hypothesis,

$$\begin{aligned} i \cdot |N'_{2i}| &\geq \sum_{j=0}^{i-1} (s' - 2j) |N'_{2j}| + \sum_{j=0}^{i-1} (s' - 2(j+1)) |N'_{2j+1}| \\ &\geq \sum_{j=0}^{i-1} (s' - 2j) \binom{s'}{j} \binom{s'-1}{j} + \sum_{j=0}^{i-1} (s' - 2(j+1)) \binom{s'}{j+1} \binom{s'-1}{j}. \end{aligned}$$

Substituting in the identity $\binom{s'-j}{j} \binom{s'-1}{j} = (j+1) \binom{s'}{j+1} \binom{s'-1}{j}$, and then the identity $(s' - j - 1) \binom{s'-1}{j} = (j+1) \binom{s'-1}{j+1}$, and finally computing a telescoping sum, we obtain

$$\begin{aligned} i \cdot |N'_{2i-1}| &\geq \sum_{j=0}^{i-1} (s' - j - 1) \binom{s'}{j+1} \binom{s'-1}{j} - j \binom{s'}{j} \binom{s'-1}{j} \\ &= \sum_{j=0}^{i-1} (j+1) \binom{s'}{j+1} \binom{s'-1}{j+1} - j \binom{s'}{j} \binom{s'-1}{j} \\ &= i \binom{s'}{i} \binom{s'-1}{i}. \end{aligned}$$

Similarly, it can be verified that $|N'_{2i+1}| \geq \binom{s'}{i+1} \binom{s'-1}{i}$ by the inductive hypothesis.

By counting the number of vertices in F' , we have

$$\begin{aligned} n &\geq |N'_0| + \left(\sum_{i=1}^{s'-1} |N'_{2i-1}| + |N'_{2i}| \right) + |N'_{2s'}| \\ &\geq 1 + \left(\sum_{i=1}^{s'-1} \binom{s'}{i} \binom{s'-1}{i} + \binom{s'}{i} \binom{s'-1}{i-1} \right) + 1 \\ &= 1 + \left(\sum_{i=1}^{s'-1} \binom{s'}{i}^2 \right) + 1 = \binom{2s'}{s'}. \end{aligned}$$

Therefore,

$$E_1/n = s' \leq \left(\frac{1}{2} + o(1)\right) \log_2 n.$$

Together with the previously established inequalities $(E_1 + E_0)/n = s \leq \log_2 n$ and $2E_1 + E_0 \geq nt/4$, we have

$$nt/4 \leq 2E_1 + E_0 \leq n(E_1/n + (E_1 + E_0)/n) \leq \left(\frac{3}{2} + o(1)\right) n \log_2 n,$$

and $t \leq (6 + o(1)) \log_2 n$, which completes the proof.

If the $(n/4, t)$ -RS graph G is regular, then we can improve the upper bound to $t \leq 2(\log_2 n + 1)$. In the proof of Theorem 3.4, the auxiliary graph H is the same with G , in which every vertex has degree exactly $t/2$. Therefore for bipartite graph G , we have $n \geq 2^{t/2}$ and $t \leq 2 \log_2 n$, and for general graphs we have $t \leq 2(\log_2 n + 1)$. This bound is the best we could hope for by Corollary 3.2. \square

4. The size of matchings is below $n/4$

For the purpose of simplicity, we will use \log in place of \log_2 throughout this section.

Proof of Theorem 1.4. Let $G = (V, E)$ and $E(G)$ be the disjoint union of t induced matchings from \mathcal{M} such that each matching consists of $r = cn$ edges with $1/5 + \varepsilon \leq c < 1/4$. Suppose for contradiction $t = Kn/\log n$ and $K = 100/\varepsilon$. We start by defining $V_0 = V(G)$. For each positive integer i , let v_i be a vertex of maximum degree in the induced subgraph of G on the vertex set V_{i-1} , N_i be the set of neighbors of v_i in V_{i-1} , and $V_i = V_{i-1} \setminus N_i$. Suppose $|N_i| = n_i$, then by definition $n_1 \geq n_2 \geq \dots$ and $n_i \leq d_i$, where d_i is the degree of v_i in G . Let k be the maximum integer such that $n_k \geq 2n/\log n$. Since all the functions of N_i are disjoint, we have $n \geq \sum_{i=1}^k |N_i| \geq k \cdot (2n/\log n)$, hence $k \leq (\log n)/2$. Note that $|V_i| = n - \sum_{j=1}^i n_j$, and its number of edges $e(V_i)$ is at least $e(G) - \sum_{j=1}^i n_j^2$, since every vertex in N_j has degree at most n_j in V_{j-1} . By the choice of k , we also know that $e(V_k) \leq 1/2 \cdot (2n/\log n) \cdot n = n^2/\log n$.

Here is the key observation: for every i , an induced matching $M \in \mathcal{M}$ containing the vertex v_i can only contain one vertex in N_i . Suppose this is not true, then M contains both v_i and distinct $x, y \in N_i$, then this violates the assumption that M is induced, since by construction both $v_i x$ and $v_i y$ are edges of G . It follows immediately from this observation that if S is a subset of $\{v_1, \dots, v_k\}$ and a matching $M \in \mathcal{M}$ contains precisely those vertices in S , then M has at least $r - |S|$ edges inside the set $V \setminus (\bigcup_{i \in S} N_i)$. Now we let M be a matching chosen uniformly at random from \mathcal{M} , and let S be the subset of $\{v_1, \dots, v_k\}$ that M contains. Define Y_M to be a random subset $Y_M = V \setminus (\bigcup_{i \in S} N_i)$. Observe that the probability that a random matching in \mathcal{M} contains a vertex v_i is equal to d_i/t . Therefore, using $d_i \geq n_i$,

$$\begin{aligned} \mathbb{E}[|Y_M|] &= n - \sum_{i=1}^k \mathbb{P}(v_i \in M) \cdot |N_i| \leq n - \sum_{i=1}^k n_i^2/t \\ &\leq n - 1/t \cdot (e(G) - e(V_k)) \leq n - 1/t \cdot (rt - n^2/\log n) \\ &= (1 - c + 1/K)n. \end{aligned} \tag{4}$$

Now we partition the t induced matchings from \mathcal{M} into $2^k \leq 2^{(\log n)/2} = n^{1/2}$ classes according to the set $V(M) \cap \{v_1, \dots, v_k\}$. For a fixed constant $\alpha \in (0, 1)$, note that at most αt matchings in \mathcal{M} come from those classes of sizes less than $\alpha K n^{1/2}/(2 \log n)$, since otherwise bounding the number of matchings gives $\alpha t \leq \alpha K n^{1/2}/(2 \log n) \cdot n^{1/2}$, which results in a

contradiction. Therefore, the probability that a random matching M belongs to one of the small classes is at most α . On the other hand, by Markov's inequality and (4),

$$\mathbb{P}(|Y_M| \geq (1 - c + 1/K + \alpha)n) \leq \frac{\mathbb{E}[|Y_M|]}{(1 - c + 1/K + \alpha)n} \leq 1 - \frac{\alpha}{1 - c + 1/K + \alpha}.$$

Since $c > 1/5$ and $K > 100/\varepsilon > 100$, we have

$$\frac{\alpha}{1 - c + 1/K + \alpha} > \frac{\alpha}{1 - 1/5 + 1/100 + \alpha} > \alpha,$$

if $\alpha \leq 1/10$. Therefore, by the union bound, we can find a class containing $t' \geq \alpha K n^{1/2}/(2 \log n) = n^{1/2-o(1)}$ matchings, each of size at least $r' = r - (\log n)/2 = (c - o(1))n$, on the same vertex set of size at most $n' = (1 - c + 1/K + \alpha)n$. When $c > 1/5 + \varepsilon$, $K = 100/\varepsilon$, and α is sufficiently small,

$$r'/n' = (c - o(1))/(1 - c + 1/K + \alpha) \geq 1/4.$$

However, from Theorem 1.3, we know that in this case t' is at most $(6 + o(1)) \log n' \leq (6 + o(1)) \log n$, which contradicts that $t' \geq n^{1/2-o(1)}$. \square

REMARK. When the graph G is regular or nearly regular, we can prove the $O(n/\log n)$ upper bound for a wider range $1/6 + \varepsilon \leq c < 1/4$. It is easy to see that every vertex has degree about $(2rt)/n = 2ct$ in this case. We start with all the vertices uncovered and in each step we pick a vertex covering the most number of uncovered vertices. Let X_i be the set of uncovered vertices in the i th step. The sum of degrees of vertices in X_i is equal to $2ct|X_i|$. Therefore, there exists a vertex v_i covering at least $2ct|X_i|/n$ vertices of X_i . After k steps, there are at most $n(1 - 2ct/n)^k$ uncovered vertices left. When $k = (\log n)/2$, $t > Kn/\log n$ and K is sufficiently large, this gives $o(n)$ uncovered vertices. Note that from the proof of Theorem 1.4, the probability that a random matching from \mathcal{M} covers a fixed vertex is equal to d_i/t , which is about $2ct/t = 2c$. Therefore, the expected number of vertices that are uncovered by vertices of $\{v_1, \dots, v_k\} \cap V(M)$ is equal to $n - 2c(n - o(n)) = (1 - 2c + o(1))n$. Similarly, as before, we find a large class of induced matchings that are of size $(c - o(1))n$ and on the same vertex set of size at least $(1 - 2c + o(1))n$. When $c > 1/6 + \varepsilon$, we are able to apply Theorem 1.3 and derive a contradiction.

We would also like to point out that the method of the proof of Theorem 1.4 can be applied iteratively, and yields in general a bound of $O(n/\log^{(x)} n)$ for $c > 1/(4 + x) + \varepsilon$. When $c \rightarrow 0$, this bound is weaker than the bound from triangle removal lemma, of $O(n/\log^{(x)} n)$ with $x = O(\log(1/c))$, but this alternative approach will be superior for a large range of c .

Proof of Theorem 1.5. Let $b = 10^{-9}$. Suppose for contradiction $t > 2n/(C \log n)$ and n is sufficiently large. Here $C = C(n)$ is a function that tends to infinity slowly when $n \rightarrow \infty$, which we will choose later. By Lemma 3.3, $F = G \times K_2$ is a $(2r, t)$ -RS graph with parts A and B , each of order n . Denote by \mathcal{M} the collection of t induced matchings that make up F . The number of edges of F is $2rt = (1/2 - 2b)tn$. We will pick a sequence of disjoint subsets A_1, \dots, A_s from A , such that for each i , $|A_i| = Ct$, and s is the smallest integer so that $sCt \geq n/2$ (thus $s \sim n/(2Ct) < (\log n)/4$), moreover there is a subcollection of induced matchings $\mathcal{M}_i \subset \mathcal{M}$ such that $|\mathcal{M}_i| \geq t/16$ and each matching in \mathcal{M}_i contains at most qCt vertices from A_i , where $q = 1/100$.

We first show that such a sequence of disjoint subsets would complete the proof. So suppose such a sequence of disjoint subsets exists. Define an auxiliary bipartite graph X with first part being $[s] = \{1, \dots, s\}$, and the second part being the matchings in \mathcal{M} , where (i, M) is an edge of X if and only if $M \in \mathcal{M}_i$. The minimum degree of vertices in $[s]$ is at least $t/16$, therefore

the average degree of vertices in \mathcal{M} is at least $(t/16)s/|\mathcal{M}| = s/16$. We can find a complete bipartite subgraph in X with one part $|S| = s/16$, and the other part $\mathcal{M}' \subset \mathcal{M}$ with

$$|\mathcal{M}'| \geq |\mathcal{M}| \left/ \binom{s}{s/16} \right. \geq |\mathcal{M}|/2^s \geq t/2^{(\log n)/4} > n^{1/2}.$$

Such a subgraph exists because the number of stars with a center in some $M \in \mathcal{M}$ and $s/16$ leaves in $[s]$ is equal to $\sum_{M \in \mathcal{M}} \binom{d_M}{s/16}$, where d_M is the degree of M . By convexity and $\sum_M d_M \geq |\mathcal{M}|(s/16)$, this number is at least $|\mathcal{M}|$. Therefore, there exists such a subset S of $[s]$ of size $s/16$, whose vertices have at least $|\mathcal{M}|/\binom{s}{s/16}$ common neighbors in \mathcal{M} .

Let Y be the union of A_i with $i \in S$, thus $|Y| = Ct(s/16) \geq n/32$. Each matching in \mathcal{M}' contains at most $qCts/16 \leq n/3200$ elements of Y ; thus, it contains at least $2r - n/3200 = (1/2 - 2b - 1/3200)n > 0.499n$ edges going between $A \setminus Y$ (which has size at most $31n/32$) and B (which has size n). Note that $0.499n > \frac{1}{4}|B \cup (A \setminus Y)| = \frac{63}{128}n$. Hence, by Theorem 1.2, there can only be a constant number of such induced matchings, a contradiction. So it suffices to show the existence of the desired disjoint subsets A_1, \dots, A_s of A .

We first show that the graph F is nearly regular. Let $p = 1/2000$ and $A = A' \cup A''$ be the partition of A , where $v \in A'$, if and only if $|d_F(v) - (t+1)/2| \leq p(t+1)$. We will show that A'' is quite small. Consider the paths of length 2 with both end points in A . The number of such paths is equal to

$$\begin{aligned} \sum_{v \in B} \binom{d_F(v)}{2} &= \frac{1}{2} \sum_{v \in B} d_F(v)^2 - rt \geq \frac{1}{2} \sum_{v \in B} (2rt/n)^2 - rt \\ &= 2 \left(\frac{1}{4} - b \right)^2 t^2 n - \left(\frac{1}{4} - b \right) tn > \left(\frac{1}{8} - 2b \right) (t+1)^2 n. \end{aligned} \quad (5)$$

Note that since F is a $(2r, t)$ -RS graph, every edge of F has sum of degrees of its two vertices at most $t+1$. In particular, for any $v \in A$ and neighbor $u \in B$, we have $d(u) \leq (t+1) - d(v)$. Therefore, every vertex $v \in A$ is the end point of at most $d(v)(t+1-d(v)) \leq (t+1)^2/4$ paths of length 2 in F . Moreover, each vertex in A'' is the end point of at most $(\frac{1}{4} - p^2)(t+1)^2$ paths of length 2 in F . Therefore, the number of paths of length 2 with both end points in A is at most $\frac{1}{2}(n - 4p^2|A''|)(t+1)^2/4$. This estimate together with (5) implies that $|A''| < (4b/p^2)n = 2n/125$.

Similarly, let $B = B' \cup B''$ be the partition of B , where B' consists of vertices that have degree within $p(t+1)$ of $(t+1)/2$, analogously we get $|B''| < 2n/125$. Let A^* be those vertices in A with at least $t/4$ neighbors in B'' . Since the number of edges coming out of B'' is at most $|B''|t$, then A^* has at most $|B''|t/(t/4) = 4|B''| < 8n/125$ elements. Thus, there are at least $n - 2n/125 - 8n/125 \geq 0.9n$ elements of A neither in A'' nor A^* . Let a_1 be one such element in A , so a_1 is adjacent to at most $t/4$ neighbors in B'' . Let B_1 be those vertices in B' adjacent to a_1 . So, $|B_1| \geq (1/2 - p)(t+1) - t/4 > t/8$, and $|B_1| \leq t$. Let $A^\#$ be those vertices in A adjacent to at least one vertex in B_1 . The number of edges coming out from B_1 is at least $|B_1|(1/2 - p)(t+1) \geq (t/8)(1/2 - p)(t+1) > t^2/20$. If $|A^\#| < Ct$, then we would have a bipartite graph between $A^\#$ and B_1 with parts of size at most Ct and t , respectively, but with at least $t^2/20$ edges. This graph is an RS graph on at most $(C(n)+1)t$ vertices with at least $t^2/20$ edges, which can be decomposed into t induced matchings. It is well known that this leads to a graph R on $\sim C(n)t$ vertices with $\Omega(t^2)$ edges, in which every edge is contained in exactly one triangle (see, for example, Proposition 3.1 in [2]). We take $C(n)$ to be $2^{\Omega(\log^* n)}$, where \log^* is the iterated logarithm. The best-known bound on the triangle removal lemma [12] (see also [10] for an alternative proof) guarantees that such a graph R cannot exist. We therefore have $|A^\#| \geq Ct$. Let A_1 be an arbitrary subset of $A^\#$ of size Ct .

Let \mathcal{M}' be the collection of matchings in \mathcal{M} containing a_1 , and \mathcal{M}'' be the matchings not containing a_1 , then $|\mathcal{M}'| \geq (1/2 - p)(t + 1)$ and $|\mathcal{M}''| \leq t - (1/2 - p)(t + 1)$. Each vertex $v \in B_1$ is in at least $(1/2 - p)(t + 1)$ matchings, and all but one of them is in \mathcal{M}'' . Thus, a vertex u in A_1 (which is adjacent to some $v \in B_1$) cannot be contained in these matchings except for one, so u is in at most $t - (1/2 - p)(t + 1) - (1/2 - p)(t + 1) + 1 = 2p(t + 1)$ matchings in \mathcal{M}'' . Therefore, the number of matchings in \mathcal{M}'' that contain at least qCt elements of A_1 is at most $2p(t + 1)(Ct)/(qCt) = 2p(t + 1)/q < t/8 \leq |\mathcal{M}''|/2$. Therefore, at least half the matchings in \mathcal{M}'' contain at most qCt elements of A_1 . Let \mathcal{M}_1 be the set of these matchings, we have $|\mathcal{M}_1| \geq t/16$.

We have thus pulled out the desired sets A_1 and \mathcal{M}_1 and finished the first step. In the second step, we find A_2 in $A \setminus A_1$ and \mathcal{M}_2 in a similar fashion. After the i th step, we have pulled out disjoint sets A_1, \dots, A_i and also found sets of matchings $\mathcal{M}_1, \dots, \mathcal{M}_i$. Note that $|A'' \cup A^* \cup A_1 \cup \dots \cup A_i| \leq 2n/125 + 8n/125 + sCt \leq 0.7n$. Let $A_0 = A \setminus (A'' \cup A^* \cup A_1 \cup \dots \cup A_i)$, then $|A_0| \geq 0.3n$. Note that by similar arguments, every vertex in A_0 is adjacent to at least $t/8$ neighbors in B' . Moreover, let e_v be the number of paths $vu v'$ with $v, v' \in A_0$ and $u \in B'$. Note that $e_v, v \in A_0$ counts precisely the number of edges between the neighbors of v in B' and the second neighborhood of v in A_0 . For $u \in B'$, let d'_u be the number of vertices in A_0 adjacent to u , then

$$\sum_v e_v = 2 \sum_{u \in B'} \binom{d'_u}{2} \geq 2 \binom{e(A_0, B')/|B'|}{2} |B'|.$$

Therefore, there exists a vertex $a_{i+1} \in A_0$, such that

$$e_{a_{i+1}} \geq 2 \binom{e(A_0, B')/|B'|}{2} |B'|/|A_0| = \frac{e(A_0, B')}{|A_0|} \left(\frac{e(A_0, B')}{|B'|} - 1 \right).$$

Since every vertex in A_0 is adjacent to at least $t/8$ vertices in B' , we have $e(A_0, B')/|A_0| \geq t/8$, and $e(A_0, B') \geq 0.3nt/8$. Hence, $e_{a_{i+1}} \geq (t/8)(0.3nt/8n - 1) \geq t^2/300$. Once again the triangle removal lemma implies that we can pick A_{i+1} to be any Ct vertices from A_0 that are second neighbors of a_{i+1} . Let \mathcal{M}_{i+1} be the collections of matchings in \mathcal{M} not containing a_{i+1} and containing at most qCt elements of A_{i+1} , similarly we have $|\mathcal{M}_{i+1}| \geq t/16$. In the end, we obtain the disjoint subsets A_1, \dots, A_s and collections of matchings $\mathcal{M}_1, \dots, \mathcal{M}_s$ as desired. \square

5. Concluding remarks

In this paper, we consider the problem of determining the maximum number of induced matchings t as a function in the size of matchings r and number of vertices n , and settle the problem for $r = cn$ when $c > 1/4$ and $c = 1/4$. Several intriguing problems remain open.

- When $r = n/4$, we prove that $t \leq (6 + o(1)) \log_2 n$, while the hypercube construction and its refinement only gives a $t = (2 + o(1)) \log_2 n$ lower bound. For regular graphs, we know that this lower bound is tight. It seems plausible that for r, n satisfying $r = n/4$ and n tending to infinity, we always have $t \leq 2 \log_2 n + 2$. The expansion part of our proof of Theorem 1.3 is pretty robust in that it uses nothing about the number t of matchings. However, there are two places we lost a constant factor in the proof. We first pass from G to H (respectively, H'), keeping only edges whose sum of degrees is at least t (respectively, $t + 1$). Maybe it is possible to somehow incorporate negative edges (that is, whose sum of degrees of its two vertices is less than t) in the argument. In the second step, we pass from H (or H') to its subgraph by keeping only high-degree vertices. It would be great to do a more careful analysis and eventually improve the constant before $\log_2 n$ from 6 to 2.

- When the size of matchings is close to $n/4$, there are two quite different constructions of RS graphs. One is the Kneser graph $KG(2k + 1, k)$ with $k \sim \frac{1}{2} \log_2 n$, which is an

$(n/4 + \Theta(n/\log n), (1 + o(1)) \log_2 n)$ -RS graph. The other is the hypercube $\{0, 1\}^{\log_2 n}$, which is an $(n/4, 2 \log_2 n)$ -RS graph. Can we find a family of (r, t) -RS graphs that bridge between these two examples, say with $t = c \log_2 n$ for some $c \in [1, 2]$, and $r - n/4 = \Omega(n/\log n)$?

- In Section 4, we obtained an upper bound $O(n/\log n)$ on the number of matchings when c is between $1/5$ and $1/4$, and improved the bound to $n/((\log n)^{2\Omega(\log^* n)})$ for $c > 1/4 - b$ for some constant $b > 0$. However, the best construction [11] so far only gives $n^{\Omega(1/\log \log n)}$. It remains a challenging question to close this gap for this range, and in general for all r linear in n .

- Another interesting open problem is to maximize the size of induced matchings, when the number of matchings is linear in n . It is somewhat complementary to the questions we studied in this paper. The Ruzsa–Szemerédi theorem [18] gives an upper bound $o(n)$, which in turn gives a similar bound for Roth’s theorem. On the other hand, for Roth’s theorem, much better results are known. The current best upper bound on the size of the largest subset of $[n]$ with no 3-term arithmetic progression is of the form $O(n/\log^{1-o(1)} n)$, first proved by Sanders [19] using Fourier analysis on Bohr sets, and subsequently improved slightly further by Bloom [7]. It would be very interesting to see whether a similar bound holds for the Ruzsa–Szemerédi theorem. Moreover, the best-known lower bound for both problems is still essentially $n/e^{O(\sqrt{\log n})}$ obtained by Behrend [5] about 70 years ago. Closing the gap between these bounds is an important open problem.

Acknowledgement. The authors would like to thank the anonymous referee for providing a number of insightful comments.

References

1. N. ALON, ‘Testing subgraphs in large graphs’, *Random Structures Algorithms* 21 (2002) 359–370.
2. N. ALON, A. MOITRA and B. SUDAKOV, ‘Nearly complete graphs decomposable into large induced matchings and their applications’, *J. Eur. Math. Soc.* 15 (2013) 1575–1596.
3. N. ALON and A. SHAPIRA, ‘Testing subgraphs in directed graphs’, *J. Comput. System Sci.* 69 (2004) 354–382.
4. N. ALON and A. SHAPIRA, ‘A characterization of easily testable induced subgraphs’, *Combin. Probab. Comput.* 15 (2006) 791–805.
5. F. BEHREND, ‘On sets of integers which contain no three terms in arithmetic progression’, *Proc. Natl. Acad. Sci. USA* 32 (1946) 331–332.
6. Y. BIRK, N. LINIAL and R. MESHULAM, ‘On the uniform-traffic capacity of single-hop interconnections employing shared directional multichannels’, *IEEE Trans. Inform. Theory* 39 (1993) 186–191.
7. T. F. BLOOM, ‘A quantitative improvement for Roth’s theorem on arithmetic progressions’, *J. London Math. Soc.* (2016), doi: 10.1112/jlms/jdw010.
8. W. BROWN, P. ERDŐS and V. SÓS, *Some extremal problems on r -graphs*, New Directions in the Theory of Graphs, Proceedings of the 3rd Ann Arbor Conference on Graph Theory (Academic Press, New York, 1973) 55–63.
9. W. BROWN, P. ERDŐS and V. SÓS, ‘On the existence of triangulated spheres in 3-graphs and related problems’, *Period. Math. Hungar.* 3 (1973) 221–228.
10. D. CONLON and J. FOX, ‘Graph removal lemmas’, *Surveys in combinatorics*, London Mathematical Society Lecture Note Series 409 (eds S. R. Blackburn, S. Gerke and M. Wildon; Cambridge University Press, Cambridge, 2013) 1–49.
11. E. FISCHER, E. LEHMAN, I. NEWMAN, S. RASKHODNIKOVA, R. RUBINFELD and A. SAMORODNITSKY, ‘Monotonicity testing over general poset domains’, *Proceedings of the 34th Annual ACM Symposium on Theory of Computing* (ACM, New York, 2002) 474–483.
12. J. FOX, ‘A new proof of the graph removal lemma’, *Ann. of Math.* 174 (2011) 561–579.
13. P. FRANKL and Z. FÜREDI, ‘Colored packing of sets’, *Combinatorial Design Theory, Ann. Discrete Math.* 34 (1987) 165–178.
14. J. HÅSTAD and A. WIGDERSON, ‘Simple analysis of graph tests for linearity and PCP’, *Random Structures Algorithms* 22 (2003) 139–160.
15. M. PLOTKIN, ‘Binary codes with specified minimum distance’, *IRE Trans. Inform. Theory* 6 (1960) 445–450.
16. S. RASHODNIKOVA, ‘Property testing: theory and applications’, PhD Thesis, Massachusetts Institute of Technology, Cambridge, MA, 2003.
17. K. ROTH, ‘On certain sets of integers’, *J. Lond. Math. Soc.* 28 (1953) 104–109.
18. I. RUZSA and E. SZEMERÉDI, ‘Triple systems with no six points carrying three triangles’, *Colloquia Math. Soc. János Bolyai* 18 (1978) 939–945.

19. T. SANDERS, 'On Roth's theorem on progressions', *Ann. of Math.* 174 (2011) 619–636.
20. E. SZEMERÉDI, 'Regular partitions of graphs', *Colloques Internationaux CNRS 260 — Problèmes Combinatoires et Théorie des Graphes*, Orsay, Paris (1976) 399–401.

Jacob Fox
Department of Mathematics
Stanford University
CA 94305-2125
USA

jacobfox@stanford.edu

Hao Huang
Department of Math and CS
Emory University
Atlanta, GA 30322
USA

hao.huang@emory.edu

Benny Sudakov
Department of Mathematics
ETH, Zürich
8092 Zürich
Switzerland

benjamin.sudakov@math.ethz.ch