ACTA ARITHMETICA
LXXV.2 (1996)

Arithmetic progressions of length three in subsets of a random set

by

Yoshiharu Kohayakawa (São Paulo),
Tomasz Łuczak (Poznań and Atlanta, Ga.) and
Vojtěch Rödl (Atlanta, Ga.)

0. Introduction. In 1936 Erdős and Turán [ET 36] asked whether for every natural number k and every positive constant α, every subset A of $[n] = \{0, 1, \ldots, n-1\}$ with at least αn elements contains a k-term arithmetic progression provided n is sufficiently large with respect to α and k. This conjecture was resolved by Roth [Ro 53] for $k = 3$, whereas for general k it was settled in the affirmative by the outstanding theorem of Szemerédi [Sz 75]. A few years after Szemerédi's paper was published, an entirely different proof of this result, based on ergodic theory, was given by Furstenberg [Fu 77]. Since then, the main open problem concerning the original question of Erdős and Turán has been to find better lower bounds for the size of A that guarantee the existence of arithmetic progressions of length k in A. Unfortunately, not much has been accomplished for $k \geq 4$. The explicit estimates that follow from Szemerédi's original proof are very poor and Furstenberg's approach does not provide such bounds at all. The case $k = 3$ is much better understood. Roth's original argument implies that it is enough to assume that $|A| \geq n/\log\log n$ and the best lower bound to date has been given independently by Heath-Brown and Szemerédi (see [H-B 87]), who showed that for some absolute constant $c > 0$ every subset of $[n]$ with at least $n/(\log n)^c$ elements contains an arithmetic progression of length three, provided n is sufficiently large.

1991 *Mathematics Subject Classification*: 11B25, 05D99, 11B05, 11K99.

Key words and phrases: Szemerédi's theorem, arithmetic progressions, combinatorial number theory, regularity lemma, random sets of integers.

Research of the first author partially supported by FAPESP (Proc. 93/0603–1) and by CNPq (Proc. 300334/93–1 and ProTeM-CC-II Project ProComb).

Research of the third author partially supported by NSF grant DMS–9401559.

In this paper we approach a related problem. Namely, here we are interested in the existence of a "small" and "sparse" set $R \subseteq [n]$ with the property that every subset $A \subseteq R$ that contains a fixed positive fraction of the elements of R contains also a 3-term arithmetic progression. The measure of sparseness here should be so that it reflects the fact that R is locally poor in 3-term arithmetic progressions. Clearly, a natural candidate for such a set R is an M-element set R_M uniformly selected from all the M-element subsets of $[n]$, where $1 \le M = M(n) \le n$ is to be chosen suitably. Our main result here confirms this appealing, intuitive idea.

For integers $1 \le M \le n$, let $\mathcal{R}(n, M)$ be the probability space of all the M-element subsets of $[n]$ equipped with the uniform measure. In the sequel, given $0 < \alpha \le 1$ and a set $R \subseteq [n]$, let us write $R \to_\alpha 3$ if any $A \subseteq R$ with $|A| \ge \alpha |R|$ contains a 3-term arithmetic progression. Our main result may then be stated as follows.

THEOREM 1. *For every constant $0 < \alpha \le 1$, there exists a constant $C = C(\alpha)$ such that if $C\sqrt{n} \le M = M(n) \le n$ then the probability that $R_M \in \mathcal{R}(n, M)$ satisfies $R_M \to_\alpha 3$ tends to 1 as $n \to \infty$.*

From Theorem 1, it is easy to deduce the following analogous result for the random sets $R_p \subseteq [n]$ $(0 \le p = p(n) \le 1)$ whose elements are chosen from $[n]$ independently with probability p. Thus, if we write $\mathcal{R}(n, p)$ for the probability space of such R_p, then for a given set $R \subseteq [n]$ the probability that $R_p = R$ is $p^{|R|}(1-p)^{n-|R|}$.

THEOREM 2. *For every constant $0 < \alpha \le 1$, there exists a constant $C = C(\alpha)$ such that if $C/\sqrt{n} \le p = p(n) \le 1$ then the probability that $R_p \in \mathcal{R}(n, M)$ satisfies $R_p \to_\alpha 3$ tends to 1 as $n \to \infty$.*

Note that Theorems 1 and 2 are, in a way, close to being best possible: if $M = M(n) = \lfloor \varepsilon \sqrt{n} \rfloor$ for some fixed $\varepsilon > 0$ then the number of 3-term arithmetic progressions in $R_M \in \mathcal{R}(n, M)$ is, with large probability, smaller than $2\varepsilon^2 |R_M|$, and hence all of them may be destroyed by deleting at most $2\varepsilon^2 |R_M|$ elements from R_M; in other words, with large probability the relation $R_M \to_\alpha 3$ does *not* hold for $\alpha = 1 - 2\varepsilon^2$. Clearly, a similar phenomenon happens for R_p with $p = p(n) = \varepsilon/\sqrt{n}$.

Our results above immediately imply the existence of "sparse" sets $S = S_\alpha$ such that $S \to_\alpha 3$ for any fixed $0 < \alpha \le 1$. The following result makes this assertion precise.

COROLLARY 3. *Suppose that $s = s(n) = o(n^{1/8})$ and $g = g(n) = o(\log n)$ as $n \to \infty$. Then, for every fixed $\alpha > 0$, there are constants C and N such that for every $n \ge N$ there exists $S \subseteq [n]$ satisfying $S \to_\alpha 3$ for which the following three conditions hold.*

(i) *For every* $k \geq 0$ *and* $l \geq 1$ *the set* $\{k, k+l, \ldots, k+sl\}$ *contains at most three elements of* S, *and therefore*, *in particular*, S *contains no* 4*-term arithmetic progression.*

(ii) *Every set* $\{k, k+l, \ldots, k+ml\}$ *with* $k \geq 0$, $l \geq 1$, *and* $m \geq \sqrt{n}\log n$ *contains at most* $Cm/\sqrt{n}$ *elements of* S.

(iii) *If* $\mathcal{F} = \mathcal{F}(S)$ *is the* 3*-uniform hypergraph on the vertex set* S *whose hyperedges are the* 3*-term arithmetic progressions contained in* S, *then* $\mathcal{F}$ *has no cycle of length smaller than* g.

In words, conditions (i) and (ii) above say that the set S intersects any arithmetic progression in a small number of elements. In particular, S contains no 4-term arithmetic progressions. Condition (iii) is more combinatorial in nature, and says that the 3-term arithmetic progressions contained in S locally form a tree-like structure, which makes the property $S \to_\alpha 3$ somewhat surprising.

Let us remark that the following extension of Szemerédi's theorem related to Corollary 3 was proved in [Rö 90], thereby settling a problem raised by Spencer [Sp 75]. Let k, $g \geq 3$ be fixed integers and $0 < \alpha \leq 1$ be a fixed real. Theorem 4.3 of [Rö 90] asserts that then, for any large enough n, there exists a k-uniform hypergraph $\mathcal{F}$ on $[n]$, all of whose hyperedges are k-term arithmetic progressions, such that $\mathcal{F}$ contains no cycle of length smaller than g but each subset $A \subseteq [n]$ with $|A| \geq \alpha n$ contains a hyperedge of $\mathcal{F}$. For other problems and results in this direction, see Graham and Nešetřil [GN 86], Nešetřil and Rödl [NR 87] and Prömel and Voigt [PV 88]. Note that Corollary 3 strengthens the above result of [Rö 90] in the case $k = 3$.

The proof of Theorem 1 is unfortunately quite long. In the next section we describe our general approach, stressing the main ideas involved and ignoring several quite technical parts. We hope that the outline of our method presented there will be of some use in following the actual proof. The organization of the paper is also discussed in the next section.

1. Outline of the method of proof. The main lemma in the proof of Theorem 1 is Lemma 19. In essence, what this lemma says is quite simple. Assume $C\sqrt{n} \leq M' = M'(n) \leq n$ for some large $C > 0$. Disregarding some technicalities, Lemma 19 states the following: if we condition on our set $R_{M'} \in \mathcal{R}(n, M')$ satisfying a certain "sparseness" condition, the probability that $R_{M'}$ *fails* to contain an arithmetic progression of length three is at most $\exp\{-cM'\}$, where we may make c arbitrarily large by picking C appropriately large.

Theorem 1 is shown to follow from Lemma 19 in two steps. Suppose C is a large constant with respect to a given fixed $\alpha > 0$ and $M = M(n) \leq \alpha^{-1}M'$. We aim at showing that $R_M \to_\alpha 3$ with probability approaching 1. Our first step consists of a quick calculation based on Lemma 19 to deduce that a

typical random set $R_M \in \mathcal{R}(n, M)$ will *not* contain a "sparse" M'-element subset R' that is free of 3-term arithmetic progressions. Our second step is then to show that our "sparseness" condition is weak enough for *every* M'-element subset R' of a typical $R_M \in \mathcal{R}(n, M)$ to be sparse. Hence Theorem 1 follows.

Thus, all our efforts go into proving Lemma 19. An important tool in the proof will be a version of Szemerédi's regularity lemma [Sz 78]. As is well known, this is an important graph theoretical component of Szemerédi's proof of his theorem on arithmetic progressions. It turns out that it is most convenient to phrase our arguments below in terms of graphs as well. Following an idea of Ruzsa and Szemerédi [RSz 78] (see also Erdős, Frankl and Rödl [EFR 86] or Graham and Rödl [GR 87]), for every subset R' of $[n]$ we construct a graph $G(n, R')$ that, roughly speaking, has the property that it contains a triangle (more precisely, a "spontaneous" triangle) if and only if R' contains a 3-term arithmetic progression (more precisely, an "arithmetic triple"). Lemma 19 is in fact stated in terms of sparse graphs and spontaneous triangles, and it asserts that sparse graphs free of such triangles are extremely rare.

Unfortunately, the proof of Lemma 19 is quite complex, and we shall not attempt to give a non-technical outline of it here. Probably any such sketch would fail to be of much help. Nonetheless, we remark that the argument below is divided into several steps, which are, to a large extent, independent of one another, and perhaps of some interest in their own right.

The organization of our paper is as follows. In Section 2 we introduce the notions of regularity, uniformity and sparseness of graphs, and state a version of Szemerédi's regularity lemma for suitably sparse graphs together with a few related results. We start Section 3 with an analogue of a theorem of Ruzsa and Szemerédi [RSz 78] for sparse graphs (cf. Theorem 8 and Lemma 9), and then give an important but rather technical lemma, Lemma 10, concerning the existence of certain structures we call "flowers" in edge-coloured sparse graphs. It is in proving Lemma 10 that we shall make use of Szemerédi's regularity lemma in the form given in Section 2.

One of the main probabilistic ingredients in the proof of Lemma 19 is given in Section 4. Roughly speaking, we show in Lemma 11 that a random induced subgraph of a bipartite uniform graph contains with very large probability a fair number of edges. In Section 5 we give a simple sufficient condition for a regular bipartite graph to be uniform. In Section 6 we relate our graph theoretical results of the previous sections to subsets of $[n]$: we define the "difference graph" $G_R = G_R(n)$ for any given subset R of $[n]$ and, using a result from Section 5, show that if R is a random set of suitably large expected size, then its difference graph G_R is uniform with large probability. Finally, the statement and proof of our main lemma, Lemma 19, and the

proof of Theorem 1 are given in Section 7, together with a sketch of the proof of Corollary 3.

2. Uniform graphs and Szemerédi's lemma. Let G be a graph with vertex set $V = V(G)$ and edge set $E(G)$. Write $|G| = |V(G)| = n$ for the order of G, and $e(G) = |E(G)|$ for its size $|E(G)|$. Furthermore, let $U, W \subseteq V$ be a pair of disjoint, non-empty subsets of G, let $E_G(U,W)$ denote the set of edges of G that have one end-vertex in U and the other in W, and set $e(U,W) = e_G(U,W) = |E_G(U,W)|$. The *density* $d_G(U,W)$ of the pair (U,W) is defined by

$$d_G(U,W) = \frac{e_G(U,W)}{|U||W|} \Big/ \frac{e(G)}{n^2} .$$

For $0 < \varepsilon \leq 1$, we say that the pair (U,W) is (ε, G)-*regular*, or simply ε-*regular*, if for all $U' \subseteq U$ and $W' \subseteq W$ with $|U'| \geq \varepsilon|U|$ and $|W'| \geq \varepsilon|W|$ we have

$$|d_G(U,W) - d_G(U',W')| \leq \varepsilon.$$

We say that an l-partite graph G is (l,ε)-*uniform*, or just ε-uniform, if all the $\binom{l}{2}$ pairs of distinct vertex classes of the l-partition of G are ε-regular.

We say that a partition $\Pi = (V^i)_0^k$ of the vertex set $V = V(G)$ of G is (ε,k)-*equitable* if $|V^0| \leq \varepsilon n$ and $|V^1| = \ldots = |V^k|$. We refer to V^0 as the *exceptional* class of Π. We say that the (ε,k)-equitable partition Π is a *subpartition* of a partition $\Pi' = (W^j)_0^s$ of V if each V^i with $1 \leq i \leq k$, that is, every *non-exceptional* class of Π, is contained in some member of the partition Π'. For Π to be a subpartition of Π' in the case where Π' is an equitable partition as well, we require every non-exceptional class of Π to be contained in some *non-exceptional* class of Π'. We say that an (ε,k)-equitable partition $\Pi = (V^i)_0^k$ is $(\varepsilon,k;G)$-*regular*, or simply (ε,G)-*regular*, if at most $\varepsilon\binom{k}{2}$ pairs (V^i,V^j) with $1 \leq i < j \leq k$ are not (ε,G)-regular.

Finally, for a given $b > 2$ and $0 < \eta \leq 1$, we say that G is (b,η)-*sparse* if, for every disjoint pair of sets $U, W \subseteq V$ such that $|U|, |W| \geq \eta n$, we have $d_G(U,W) \leq b$. Thus, roughly speaking, a graph is (b,η)-sparse if all of its large induced subgraphs are not much denser that the graph itself. We can now state our extension of Szemerédi's lemma for (b,η)-sparse graphs.

LEMMA 4. *For any given $\varepsilon > 0$, $b > 2$, $k_0 \geq 1$ and $s \geq 1$, there are constants $\eta = \eta(\varepsilon,b,k_0,s) > 0$ and $K_0 = K_0(\varepsilon,b,k_0,s) \geq k_0$ that depend only on ε, b, k_0, and s for which the following holds. For every (b,η)-sparse graph G and every partition $(W^j)_0^s$ of the vertex set of G, there exists an (ε,G)-regular (ε,k)-equitable subpartition of $(W^j)_0^s$ with $k_0 \leq k \leq K_0$.*

The proof of Lemma 4 goes along the same lines as the proof of Szemerédi's original result [Sz 78], and hence we omit it here. As a matter of

fact, in order to prove Lemma 4, one may just rewrite the argument from [Sz 78] putting the "scaled" density $d_G(U,W)$ (or, more precisely, $d_G(U,W)/b$) in place of Szemerédi's original density $d(U,W) = e_G(U,W)/(|U| \cdot |W|)$.

From Lemma 4 it immediately follows that one can uniformly partition any *fixed* number of graphs at the same time. Let $G_1, \ldots, G_m$ be a sequence of graphs defined on the same vertex set V, and let $G = \bigcup_{i=1}^m G_i$ be their union, i.e., the graph on V with the edge set $\bigcup_{i=1}^m E(G_i)$. We say that $G_1, \ldots, G_m$ is (b,η)-*sparse* if G is (b,η)-sparse. Furthermore, an (ε,k)-equitable partition $\Pi = (V^i)_0^k$ is $(\varepsilon,k;G_1,\ldots,G_m)$-*regular* if it is $(\varepsilon,k;G_i)$-regular for all $1 \le i \le m$.

One can easily deduce from Lemma 4 that every (b,η)-sparse sequence of graphs $G_1,\ldots,G_m$ admits an $(\varepsilon,k;G_1,\ldots,G_m)$-regular partition, provided η is small enough with respect to ε, b and $1/k$. Roughly speaking, we first choose a rapidly decreasing sequence $\varepsilon = \varepsilon_m \ge \varepsilon_{m-1} \ge \ldots \ge \varepsilon_1 > 0$ of constants, and then invoke Lemma 4 in turn to define a sequence $\Pi_1, \ldots, \Pi_m$ of finer and finer partitions of V. To be precise, Π_i is required to be a subpartition of Π_{i-1} for all $1 < i \le m$, and Π_i $(1 \le i \le m)$ is required to be an (ε_i, G_i)-regular partition of V with a "small" number of classes (as small as Lemma 4 can guarantee). Choosing the ε_i $(1 \le i \le m)$ carefully enough, the final partition Π_m of our sequence $\Pi_1,\ldots,\Pi_m$ will be the $(\varepsilon,k;G_1,\ldots,G_m)$-regular partition we seek.

It will be important later that, for any $1 < i \le m$, the partition Π_i above may be determined *solely* from Π_{i-1} and G_i. In other words, the graphs $G_{i+1},\ldots,G_m$ play no rôle in the definition of Π_i.

We now make the above informal discussion precise. Thus, let k_0 and $m \ge 3$ be natural numbers, let $\boldsymbol{\varepsilon}$ be a sequence of numbers $\varepsilon_1,\ldots,\varepsilon_m$ such that $0 < \varepsilon_1 \le \ldots \le \varepsilon_m < 1$ and let $G_1,\ldots,G_m$ be graphs with the same vertex set V, where $|V| \ge k_0$. In the definition below, we shall assume that the set of all equitable partitions of V have been given a fixed ordering, say $\prec$. The $(\boldsymbol{\varepsilon},k_0)$-*canonical sequence of partitions* $\widetilde{\Pi}_1,\ldots,\widetilde{\Pi}_m$ for $G_1,\ldots,G_m$ is defined recursively in the following way:

(i) Among all the $(\varepsilon_1,k;G_1)$-regular partitions which minimize $k \ge k_0$, let $\widetilde{\Pi}_1$ be the first one according to $\prec$.

(ii) Assume that $2 \le i \le m$ and that the partition $\widetilde{\Pi}_{i-1}$ has already been defined. Then we let $\widetilde{\Pi}_i$ be the $\prec$-first $(\varepsilon_i,k;G_1,\ldots,G_i)$-regular subpartition of $\widetilde{\Pi}_{i-1}$ which minimizes k.

Note that any partition of the vertices of G into singletons is $(\varepsilon,k;G)$-regular for $k = |V|$ and every $\varepsilon > 0$. Thus, for each sequence $G_1,\ldots,G_m$ as above, an (ε,k_0)-canonical sequence of partitions does exist and is of course unique by definition.

Now we can use Lemma 4 to deduce that for (b,η)-sparse sequences of graphs the sizes of the partitions in the associated canonical partition can be uniformly bounded from above.

LEMMA 5. *For every $\varepsilon > 0$, $b > 2$, $k_0 \geq 1$ and $m \geq 1$, there are constants $\eta = \eta(\varepsilon, b, k_0, m) > 0$ and $K_0 = K_0(\varepsilon, b, k_0, m) \geq k_0$ and a sequence $\boldsymbol{\varepsilon} = \boldsymbol{\varepsilon}(m, b, \varepsilon, k_0) = (\varepsilon_1, \ldots, \varepsilon_m)$ with $0 < \varepsilon_1 \leq \ldots \leq \varepsilon_m = \varepsilon$ such that η, K_0 and $\boldsymbol{\varepsilon}$ depend only on ε, b, k_0 and m and the following holds. For every (b,η)-sparse sequence of graphs $G_1, \ldots, G_m$, the $(\boldsymbol{\varepsilon}, k_0)$-canonical sequence of partitions $\widetilde{\Pi}_1, \ldots, \widetilde{\Pi}_m$ associated with $G_1, \ldots, G_m$ only contains partitions of sizes bounded by $K_0 + 1$. In fact, we have*

$$k_0 + 1 \leq |\widetilde{\Pi}_1| \leq \ldots \leq |\widetilde{\Pi}_m| \leq K_0 + 1.$$

Note that the $\boldsymbol{\varepsilon}$ in the lemma above depends only on m, b, ε, and k_0. In fact, throughout the paper we shall assume that, for any given m, b, ε, and k_0 as in Lemma 5, we have a *fixed* vector $\boldsymbol{\varepsilon} = \boldsymbol{\varepsilon}(m, b, \varepsilon, k_0)$ as in that lemma associated with this choice of m, b, ε, and k_0.

Let us conclude this section with two simple observations. The definition of ε-regularity deals with the distribution of edges between "large" sets. Nonetheless, it turns out that each ε-uniform l-partite graph G contains a large 3ε-uniform l-partite subgraph $\overline{G}$ such that *each* vertex of $\overline{G}$ has a fairly large degree. In fact, more is true as shown by the following statement.

FACT 6. *Suppose $l \geq 2$ and $0 < \varepsilon < 1/(5l)$. Let G be an ε-uniform $(l+1)$-partite graph with $(l+1)$-partition $V(G) = V_0 \cup \ldots \cup V_l$. Then there exist subsets $\overline{V}_i \subseteq V_i$ $(1 \leq i \leq l)$ such that, for every $1 \leq i \leq l$, we have $|\overline{V}_i| \geq (1 - l\varepsilon)|V_i|$ and for every vertex $v \in \overline{V}_i$ we have*

$$|N_G(v) \cap \overline{V}_j| \geq (1 - 2l\varepsilon) e_G(V_i, V_j)/|V_j|$$

for every $j \neq i$ $(1 \leq j \leq l)$. In particular, the graph $\overline{G}$ induced in G by $\bigcup_{i=1}^{l} \overline{V}_i$ is 3ε-uniform.

P r o o f. Since proofs of very similar statements can be found in [HKŁ 95], here we only mention the simple idea behind the argument. In order to find the $\overline{V}_i$ $(0 \leq i \leq l)$ we successively delete from V_i the vertices that violate the conditions we seek. Then one can easily show that, because of the ε-uniformity of G, this process finishes with the required sets $\overline{V}_i$ $(0 \leq i \leq l)$. ∎

Finally, since every graph G on n vertices contains a bipartite subgraph H whose vertex classes are of cardinality $\lfloor n/2 \rfloor$ and $\lceil n/2 \rceil$ and such that $e(H) \geq e(G)/2$, the following fact is an immediate consequence of the definition of a (b,η)-sparse graph.

FACT 7. *If G is (b,η)-sparse then every subgraph H of G with $|V(H)| \geq 2\eta|V(G)| + 2$ vertices contains at most $b|V(H)|^2 e(G)/|V(G)|^2$ edges.*

3. The Ruzsa–Szemerédi theorem for sparse graphs. In this section we state and prove an analogue of a result of Ruzsa and Szemerédi [RSz 78] (see also Erdős, Frankl and Rödl [EFR 86] or Graham and Rödl [GR 87]) which remains valid for (b,η)-sparse graphs. Let us say that a graph is 3-*decomposable* if it is the union of edge-disjoint triangles. Then the Ruzsa–Szemerédi theorem that we shall need (and generalize below) may be stated as follows.

THEOREM 8. *For every constant $c>0$ there exists a constant $\widehat{\delta}(c)>0$ such that every 3-decomposable graph G with at least cn^2 edges contains at least $\widehat{\delta}(c)n^3$ triangles.*

We would like to apply a similar result for graphs which are not so dense. Unfortunately, in this case the above theorem is no longer valid: there are 3-decomposable graphs G with n vertices and at least $n^2\exp(-3\sqrt{\log n})$ edges which contain only $e(G)/3$ triangles (see, e.g., Theorem 6.6 in [GR 87]). We are thus forced to take another approach. Very roughly speaking, our method will consist in proving a probabilistic version of Theorem 8, Lemma 19, asserting that, in some sense, "counterexamples" as above are rare. However, we need to work for a while before we may state and prove Lemma 19.

We start with a result saying that if a 3-decomposable graph G admits an ε-regular partition then, although G may contain only a few triangles, it must contain many "dense" triples of partition classes. To emphasize the difference between triangles on the one hand and triples of partition classes on the other, we shall refer to the latter as *triads*.

Thus, let G be a (b,η)-sparse graph on n vertices and let $\mathit{\Pi}=(V^i)_0^k$ be an (ε,k)-equitable (ε,G)-regular partition of the vertex set of G. We say that a pair (V^r,V^s) $(1\le r<s\le k)$ is *thick* if it is (ε,G)-regular and

$$e_G(V^r,V^s)\ge |V^r|\cdot|V^s|e(G)/(50n^2).$$

We say that a triad (V^r,V^s,V^t) $(1\le r<s<t\le k)$ is *thick* if all three pairs (V^r,V^s), (V^r,V^t) and (V^s,V^t) are thick.

LEMMA 9. *For every $b>2$ there exist constants $\overline{\delta}=\overline{\delta}(b)>0$ and $\overline{k}_0=\overline{k}_0(b)\ge 1$ that depend only on b such that the following holds. For every (b,η)-sparse 3-decomposable graph G, if $\mathit{\Pi}$ is an (ε,k)-equitable (ε,G)-regular partition of G such that $200\varepsilon b\le 1$, $k\ge\overline{k}_0$, and $0<\eta\le\min\{\varepsilon,1/(2k)\}$, then $\mathit{\Pi}$ contains at least $\overline{\delta}k^3$ thick triads.*

Proof. Let $\overline{k}_0=\overline{k}_0(b)=40000b$ and $\overline{\delta}=\overline{\delta}(b)=\widehat{\delta}(0.3/b)>0$, where $\widehat{\delta}$ is as given by Theorem 8. We shall show that these values will do for our lemma. Suppose $200\varepsilon b\le 1$ and let $\mathit{\Pi}$ be an (ε,k)-equitable (ε,G)-regular partition of a (b,η)-sparse 3-decomposable graph G, where $k\ge\overline{k}_0$ and $0<\eta\le\min\{\varepsilon,1/(2k)\}$. Then the following assertions hold:

(i) the number of edges of G that have both ends in one set of Π is, by Fact 7, less than

$$\varepsilon^2 n^2 be(G)/n^2 + k(n/k)^2 be(G)/n^2 = (\varepsilon^2 b + b/k)e(G) \le e(G)/100,$$

(ii) the number of edges of G incident to the vertices in V^0 is less than

$$\varepsilon n^2 be(G)/n^2 \le \varepsilon be(G) \le e(G)/200,$$

(iii) the number of edges between pairs (V^r, V^s) $(1 \le r < s \le k)$ that are not ε-regular is less than

$$\varepsilon \binom{k}{2} \left(\frac{n}{k}\right)^2 \frac{be(G)}{n^2} \le \varepsilon be(G) \le e(G)/200,$$

(iv) the number of edges between pairs (V^r, V^s) $(1 \le r < s \le k)$ that are not thick is less than

$$\binom{k}{2} \left(\frac{n}{k}\right)^2 \frac{e(G)}{50n^2} \le e(G)/100.$$

Hence, at least $0.97e(G)$ edges of G belong to thick pairs (V^r, V^s). Let these edges form the edge set of the spanning subgraph G^0 of G. Let $\mathcal{F}$ be a family of edge-disjoint triangles of G such that $G = \bigcup \mathcal{F}$. Clearly, just considering $\mathcal{F}$, we see that G^0 contains at least $e(G)/3 - 0.03e(G) \ge 0.3e(G)$ triangles. In particular, the partition Π has at least one thick triad, say $(V^{r(1)}, V^{s(1)}, V^{t(1)})$. Let us delete from G^0 all edges between the sets $V^{r(1)}$, $V^{s(1)}$ and $V^{t(1)}$, and let G^1 be the graph obtained in this way. Since the number of edges we delete is smaller than $3b(n/k)^2 e(G)/n^2 \le 3be(G)/k^2$, we destroy at most $3be(G)/k^2 < 0.3e(G)$ triangles from $\mathcal{F}$. Thus, the graph G^1 contains a triangle and hence the partition Π, viewed as a partition of G^1, contains at least one thick triad. We repeat the procedure above and obtain a sequence $G = G^0 \supset \ldots \supset G^l$ of spanning subgraphs of G with G^l such that Π, viewed as a partition of G^l, contains no thick triad. Since in every step we decrease the number of triangles in $\mathcal{F}$ by at most $3be(G)/k^2$, we have $l \ge 0.1k^2/b$. Hence, the graph $G(\Pi)$, whose vertices are the sets $V^1, \ldots, V^k$ and V^r, V^s $(1 \le r < s \le k)$ are joined by an edge if and only if the pair (V^r, V^s) is thick, contains at least $0.1k^2/b$ edge-disjoint triangles. Thus, by Theorem 8 and our choice of $\overline{\delta} = \overline{\delta}(b) = \widehat{\delta}(0.3/b)$, the graph $G(\Pi)$ contains at least $\overline{\delta}k^3$ triangles. Consequently, there exist at least $\overline{\delta}k^3$ thick triads in Π and Lemma 9 follows. ∎

We now turn to the main lemma of this section, Lemma 10. As already mentioned in Section 1, this is a rather technical result, and before we may state it we need to introduce a few definitions, including the definition of a "flower" in an edge-coloured graph. Let us say that a sequence $G_1, \ldots, G_m$ of graphs on the same vertex set is *balanced* if

(i) $E(G_i) \cap E(G_j) = \emptyset$ for all $1 \le i < j \le m$,
(ii) $e(G_i) = e(G_j)$ for all $1 \le i < j \le m$,
(iii) for all $1 \le i \le m$, all the vertices of G_i have the same degree.

It will be convenient to think of a balanced sequence as above as a graph G, namely $G = \bigcup_{i=1}^m G_i$, whose edges have been coloured with m colours. Let $\widetilde{G} = (G_i)_{i=1}^m$ be a balanced sequence of graphs and $G = \bigcup_{i=1}^m G_i$. Let $n = |V(G)|$. Suppose also that $\widetilde{\Pi}_1, \dots, \widetilde{\Pi}_m$ is the $(\boldsymbol{\varepsilon}, k_0)$-canonical sequence of partitions associated with $G_1, \dots, G_m$ for some given k_0 and some $\boldsymbol{\varepsilon} = (\varepsilon_1, \dots, \varepsilon_m)$. Then, for any constant $0 < \delta \le 1$, a $(\delta, k_0, \boldsymbol{\varepsilon}; \widetilde{G})$-*flower*, or, for short, a $(\delta, \widetilde{G})$-*flower*, consists of three indices $1 \le w(1) < w(2) < w(3) \le m$ together with a vertex v of G and a family $\{(X^{(i)}, Y^{(i)}) : 1 \le i \le g\}$ of pairs of non-exceptional elements $X^{(i)}$ and $Y^{(i)}$ of $\widetilde{\Pi}_{w(1)}$ such that

(i) $2g|X^{(i)}| = 2g|Y^{(i)}| \ge \delta n$,
(ii) the $2g$ sets $X^{(i)}$, $Y^{(i)}$ $(1 \le i \le g)$ are all distinct,
(iii) all pairs $(X^{(i)}, Y^{(i)})$ are $(\varepsilon_{w(i)}, G_{w(1)})$-regular and

$$\begin{aligned} e_{G_{w(1)}}(X^{(i)}, Y^{(i)}) &\ge \delta|X^{(i)}| \cdot |Y^{(i)}|e(G_{w(1)})/(10^6 n^2) \\ &= \delta|X^{(i)}| \cdot |Y^{(i)}|e(G)/(10^6 mn^2), \end{aligned}$$

(iv) the vertex v is joined to each $X^{(i)}$ $(1 \le i \le g)$ by at least $\delta|X^{(i)}| \times e(G_{w(2)})/(10^6 n^2) = \delta|X^{(i)}|e(G)/(10^6 mn^2)$ edges of $G_{w(2)}$,
(v) the vertex v is joined to each $Y^{(i)}$ $(1 \le i \le g)$ by at least $\delta|Y^{(i)}| \times e(G_{w(3)})/(10^6 n^2) = \delta|Y^{(i)}|e(G)/(10^6 mn^2)$ edges of $G_{w(3)}$.

We may now state and prove Lemma 10.

Lemma 10. *Let $b > 2$ be given. Then there exist integers $m = m(b) \ge 3$ and $k_0 = k_0(b)$, and a real number $0 < \delta = \delta(b) \le 1$ that depend only on b such that, for any $0 < \varepsilon \le 1$, there exists a constant $0 < \eta = \eta(b, \varepsilon) \le 1$ for which the following holds. Let $\boldsymbol{\varepsilon} = \boldsymbol{\varepsilon}(m, b, \varepsilon, k_0)$. If $\widetilde{G} = (G_i)_{i=1}^m$ is a balanced m-edge-colouring of a (b, η)-sparse 3-decomposable graph $G = \bigcup_{i=1}^m G_i$, then the $(\boldsymbol{\varepsilon}, k_0)$-canonical sequence of partitions $\widetilde{\Pi}_1, \dots, \widetilde{\Pi}_m$ associated with $\widetilde{G} = (G_i)_{i=1}^m$ admits a $(\delta, k_0, \boldsymbol{\varepsilon}; \widetilde{G})$-flower.*

Remark. In the sequel, when considering (b, η)-sparse sequences of graphs $\widetilde{G}$ as above, we shall often say that "a $(\delta, k_0, \boldsymbol{\varepsilon}; \widetilde{G})$-flower exists" or that "$\widetilde{G}$ contains a $(\delta, \widetilde{G})$-flower". In such cases, we are tacitly assuming that these flowers are built from the $(\boldsymbol{\varepsilon}, k_0)$-canonical sequence of partitions associated with our $\widetilde{G}$, where $\boldsymbol{\varepsilon} = \boldsymbol{\varepsilon}(m, b, \varepsilon, k_0)$ is the vector we have explicitly associated with m, b, ε and k_0.

Proof of Lemma 10. Let $\overline{\delta}(b) > 0$ be as given by Lemma 9. Set $m = m(b) = \lceil 3 \cdot 10^6/\overline{\delta}(b) \rceil$, $\delta = \delta(b) = \overline{\delta}(b)/(1600m^3) > 0$ and $k_0 =$

$k_0(b) = \max\{\overline{k}_0(b), 800m^3/\overline{\delta}(b)\} \geq 1$, where $\overline{k}_0(b)$ is given by Lemma 9. We shall show that the assertion holds with this choice of m and δ and k_0. Thus, let an arbitrary $0 < \varepsilon \leq 1$ be given. Clearly, we may assume that $\varepsilon \leq \min\{1/(200b), \overline{\delta}(b)/(800m^3)\}$. Now let us invoke Lemma 5 to obtain $\eta(\varepsilon, b, k_0, m) > 0$ and $K_0 = K_0(\varepsilon, b, k_0, m) \geq k_0$ as given by that lemma. We may and shall assume that $\eta \leq \min\{\varepsilon, 1/(2K_0)\}$. Our aim is to show that the assertion holds for $\eta = \eta(b, \varepsilon) = \eta(\varepsilon, b, k_0, m)$ given above. Therefore, let $\widetilde{G} = (G_i)_{i=1}^m$ be a balanced m-edge-colouring of a (b, η)-sparse 3-decomposable graph $G = \bigcup_{i=1}^m G_i$. We need to verify that, under these conditions, a $(\delta, k_0, \boldsymbol{\varepsilon}; \widetilde{G})$-flower does indeed exist.

Let $\widetilde{\Pi}_1, \ldots, \widetilde{\Pi}_m$ be the $(\boldsymbol{\varepsilon}, k_0)$-canonical sequence of partitions associated with $G_1, \ldots, G_m$. We first concentrate our attention on the graph $G = \bigcup_{i=1}^m G_i$ and the partition $\widetilde{\Pi}_m$. We show that each of at least half of the thick triads (V^r, V^s, V^t) in $\widetilde{\Pi}_m$, whose existence is guaranteed by Lemma 9, has the property that we may assign to its "sides", i.e., the pairs (V^r, V^s), (V^r, V^t) and (V^s, V^t), some three distinct colours so that, if a side is assigned colour $1 \leq w \leq m$, then it contains a substantial number of edges from G_w.

Let us eliminate first the triads for which an assignment as we seek is not possible. Given $1 \leq w \leq m$, we say that a thick triad (V^r, V^s, V^t) of the ε-regular partition $\widetilde{\Pi}_m$ of G is *dominated* by w if, putting $u = |V^r| = |V^s| = |V^t|$, at least two out of the three pairs of sets (V^r, V^s), (V^r, V^t) and (V^s, V^t) are joined by more than

$$u^2 e(G)/(150n^2) \geq (n/(2k_m))^2 e(G)/(150n^2) = e(G)/(600k_m^2)$$

edges of G_w, where $k_m = |\widetilde{\Pi}_m| - 1 \leq K_0$. If a thick triad is not dominated by any w $(1 \leq w \leq m)$, we say that it is *balanced*. Since every vertex of G_w has degree $2e(G_w)/n = 2e(G)/(mn)$, at most $2e(G)/(mk_m)$ edges of G_w are incident to a given set V^r of $\widetilde{\Pi}_m$. Consequently, the number of thick triads dominated by w is less than

$$k_m \left[\frac{2e(G)/(mk_m)}{e(G)/(600k_m^2)}\right]^2 \leq \frac{1.5 \cdot 10^6 k_m^3}{m^2} \leq \frac{\overline{\delta}(b) k_m^3}{2m},$$

so the number of thick triads dominated by some w $(1 \leq w \leq m)$ is less than $\overline{\delta}(b)k_m^3/2$. Thus, by Lemma 9, the graph G contains at least $\overline{\delta}(b)k_m^3/2$ balanced thick triads.

We call a pair (V^r, V^s) of partition classes of $\widetilde{\Pi}_m$ a *w-rich* pair, where $1 \leq w \leq m$, if the number of edges of G_w between V^r and V^s is larger than $|V^r| \cdot |V^s| e(G_w)/(150n^2) = |V^r| \cdot |V^s| e(G)/(150mn^2)$. We say that (V^r, V^s, V^t) is $(w(1), w(2), w(3))$-*colourable*, where the $w(i)$ $(i \in \{1, 2, 3\})$ are three distinct colours from $\{1, \ldots, m\}$, if and only if we may assign the colours

$w(1)$, $w(2)$ and $w(3)$ to the pairs (V^r, V^s), (V^r, V^t) and (V^s, V^t), in some order, in such a way that a colour w is assigned to a pair only if this pair is w-rich. Now, let (V^r, V^s, V^t) be a balanced thick triad. A moment's thought reveals that any such triad is $(w(1), w(2), w(3))$-colourable for some $1 \le w(1) < w(2) < w(3) \le m$. Consequently, there is a choice of $(w(1), w(2), w(3))$ with $1 \le w(1) < w(2) < w(3) \le m$ such that $\widetilde{\Pi}_m$ contains at least $\overline{\delta}(b)k_m^3/2\binom{m}{3}$ thick triads that are $(w(1), w(2), w(3))$-colourable.

Thus, there exist a non-exceptional partition class $\overline{V}$ of $\widetilde{\Pi}_m$ and pairs (U_m^i, W_m^i) $(1 \le i \le f = \lceil \overline{\delta}(b)k_m^2/(2m^3) \rceil)$ of non-exceptional partition classes of $\widetilde{\Pi}_m$ such that, for all $1 \le i \le f$, the triple $(\overline{V}, U_m^i, W_m^i)$ satisfies the following conditions:

(i) the pair (U_m^i, W_m^i) is $(\varepsilon, G_{w(1)})$-regular, and

$$e_{G_{w(1)}}(U_m^i, W_m^i) \ge |U_m^i| \cdot |W_m^i| e(G_{w(1)})/(150n^2),$$

(ii) the pair $(\overline{V}, U_m^i)$ is $(\varepsilon, G_{w(2)})$-regular, and

$$e_{G_{w(2)}}(\overline{V}, U_m^i) \ge |\overline{V}| \cdot |U_m^i| e(G_{w(2)})/(150n^2),$$

(iii) the pair $(\overline{V}, W_m^i)$ is $(\varepsilon, G_{w(3)})$-regular, and

$$e_{G_{w(3)}}(\overline{V}, W_m^i) \ge |\overline{V}| \cdot |W_m^i| e(G_{w(3)})/(150n^2).$$

Now, it follows from Fact 6 with $l = 3$ that, for every given i $(1 \le i \le f)$, at least half of the vertices of $\overline{V}$ are joined both by at least $|U_m^i| e(G_{w(2)})/(300n^2)$ edges of $G_{w(2)}$ to U_m^i and by at least $|W_m^i| e(G_{w(3)}) \times (300n^2)^{-1}$ edges of $G_{w(3)}$ to W_m^i. Thus, by an elementary averaging argument, there is a vertex v in $\overline{V}$ and a set $\Xi = \{(U_m^i, W_m^i) : i \in \Lambda\}$ of $w(1)$-rich pairs (U_m^i, W_m^i) of cardinality $|\Lambda| \ge f/2 \ge \overline{\delta}(b)k_m^2/(4m^3)$ such that v is joined by at least $|U_m^i| e(G_{w(2)})/(300n^2)$ edges of $G_{w(2)}$ to U_m^i and by at least $|W_m^i| e(G_{w(3)})/(300n^2)$ edges of $G_{w(3)}$ to W_m^i for all $(U_m^i, W_m^i) \in \Xi$.

We shall show that the existence of such a vertex v and such a set Ξ implies the existence of a $(\delta, \widetilde{G})$-flower. The three colours associated with the flower we shall exhibit are $w(1)$, $w(2)$, $w(3)$. Our only problem is that so far we have only considered the partition $\widetilde{\Pi}_m$, whereas the definition of the flower deals with subsets of the partition $\widetilde{\Pi}_{w(1)}$. Thus, in the rest of the proof we shall try to relate the properties of $\widetilde{\Pi}_m$ with those of $\widetilde{\Pi}_{w(1)}$. In order to do this, we shall consider the graph $\widehat{G}(\widetilde{\Pi}_m, \Xi)$ whose vertex set is the set of non-exceptional partition classes of $\widetilde{\Pi}_m$, with two such vertices U_m, W_m being connected by an edge in $\widehat{G}(\widetilde{\Pi}_m, \Xi)$ if and only if $(U_m, W_m) \in \Xi$.

Note first that $\widehat{G}(\widetilde{\Pi}_m, \Xi)$ has $N = k_m$ vertices and at least cN^2 edges, where $c = \overline{\delta}(b)/(8m^3)$. Furthermore, since $\widetilde{\Pi}_m$ is a subpartition of $\widetilde{\Pi}_{w(1)}$,

each non-exceptional partition class of $\widetilde{\Pi}_m$ is contained in some non-exceptional partition class of $\widetilde{\Pi}_{w(1)}$. Thus, the partition $\widetilde{\Pi}_{w(1)}$ of G naturally induces a partition $\widehat{\Pi}_{w(1)}$ of the vertex set of $\widehat{G}(\widetilde{\Pi}_m, \Xi)$. Unfortunately, because of the exceptional class V_m^0 of $\widetilde{\Pi}_m$, not all partition classes of $\widehat{\Pi}_{w(1)}$ need have the same size. However, since $|V_{w(1)}^0|, |V_m^0| \leq \varepsilon n$ we immediately see that all of them contain at most $2k_m/k_{w(1)}$ elements, and a little argument shows that at least $(1-3\varepsilon)k_{w(1)}$ of them have at least $k_m/(2k_{w(1)})$ elements.

Now suppose $l \geq 100/c$ and $\varepsilon \leq c/100$. Then simple calculations show that the following holds: for any graph $\widehat{G}$ with N vertices and at least cN^2 edges, and every partition $\widehat{\Pi} = (V^i)_{i=1}^l$ of the vertex set of $\widehat{G}$ into l classes such that all the partition classes V^i have cardinality at most $2N/l$ and not fewer than $(1-3\varepsilon)l$ of them have cardinality greater than or equal to $N/(2l)$, there exist at least $cl^2/100$ pairs $\{\widehat{V}, \widehat{V}'\}$ of distinct partition classes of $\widehat{\Pi}$ such that $|\widehat{V}|, |\widehat{V}'| \geq N/(2l)$ and $\widehat{V}$, $\widehat{V}'$ are joined by at least $c|\widehat{V}| \cdot |\widehat{V}'|/10$ edges. Hence, since every graph on at most k vertices and at least ck^2 edges contains a matching of size at least $ck/2$, the above partition $\widehat{\Pi}$ of $\widehat{G}$ must contain at least $g = \lceil cl/200 \rceil$ disjoint pairs $(\widehat{X}^i, \widehat{Y}^i)$ $(1 \leq i \leq g)$ such that $|\widehat{X}^i|, |\widehat{Y}^i| \geq N/(2l)$ and such that the number of edges between $\widehat{X}^i$ and $\widehat{Y}^i$ is at least $c|\widehat{X}^i| \cdot |\widehat{Y}^i|/10$ for all $1 \leq i \leq g$. Finally, we may again use the fact that dense graphs contain large matchings to deduce that there is a matching of size at least $cN/(40l)$ between $\widehat{X}^i$ and $\widehat{Y}^i$ for all $1 \leq i \leq g$.

We apply this observation to $\widehat{G} = \widehat{G}(\widetilde{\Pi}_m, \Xi)$ and the partition $\widehat{\Pi} = \widehat{\Pi}_{w(1)}$. Thus let us check the required hypotheses for our observation to apply. First note that the condition $l \geq 100/c$ becomes in our context $k_{w(1)} \geq 800m^3/\overline{\delta}(b)$, whilst $\varepsilon \leq c/100$ corresponds to $\varepsilon \leq \overline{\delta}(b)/(800m^2)$. The number of disjoint pairs $(\widehat{X}^i, \widehat{Y}^i)$ $(1 \leq i \leq g)$ guaranteed by the observation is $g = \lceil ck_{w(1)}/200 \rceil = \lceil \overline{\delta}(b)k_{w(1)}/(1600m^3) \rceil$. Moreover, $\widehat{X}^i$ is joined to $\widehat{Y}^i$ by at least

$$\frac{c|\widehat{X}^i| \cdot |\widehat{Y}^i|}{10} \geq \frac{c}{10}\left(\frac{k_m}{2k_{w(1)}}\right)^2 \geq \frac{1}{10} \cdot \frac{\overline{\delta}(b)}{8m^3} \cdot \frac{k_m^2}{4k_{w(1)}^2} = \frac{\overline{\delta}(b)k_m^2}{320k_{w(1)}^2 m^3}$$

edges of $\widehat{G}(\widetilde{\Pi}_m, \Xi)$, and hence the corresponding pair $(X_{w(1)}^i, Y_{w(1)}^i)$ of partition classes of $\widetilde{\Pi}_{w(1)}$ is joined by at least

$$\frac{\overline{\delta}(b)k_m^2}{320k_{w(1)}^2 m^3}\left(\frac{n}{2k_m}\right)^2 \frac{e(G_{w(1)})}{150n^2} \geq \frac{\overline{\delta}(b)|U_{w(1)}^i| \cdot |W_{w(1)}^i| e(G)}{2 \cdot 10^5 m^4 n^2}$$

edges of $G_{w(1)}$.

Furthermore, we know that, for every $1 \le i \le g$, the graph $\widehat{G}(\widetilde{\Pi}_m, \Xi)$ contains a matching $\widehat{M}_i$ between $\widehat{X}^i$ and $\widehat{Y}^i$ of size at least

$$\frac{c}{40} \cdot \frac{k_m}{k_{w(1)}} = \frac{\overline{\delta}(b)}{320m^3} \cdot \frac{k_m}{k_{w(1)}}.$$

Recall that every edge of $\widehat{G}(\widetilde{\Pi}_m, \Xi)$ corresponds to a pair (U_m, W_m) from Ξ. We may and shall assume that at least $(\overline{\delta}(b)/(640m^3))k_m/k_{w(1)}$ of the edges (U_m, W_m) from $\widehat{M}_i$ are such that $U_m \in \widehat{X}^i$ and $W_m \in \widehat{Y}^i$ (i.e., $U_m \subseteq X^i_{w(1)}$ and $W_m \subseteq Y^i_{w(1)}$). Recall that, for every $(U_m, W_m) \in \Xi$, the vertex v is joined to U_m by at least

$$\frac{|U_m| e(G_{w(2)})}{300n^2} \ge \frac{n}{2k_m} \cdot \frac{e(G)}{300mn^2} = \frac{e(G)}{600k_m mn}$$

edges of $G_{w(2)}$, and similarly it is joined to W_m by at least $e(G)/(600k_m mn)$ edges of $G_{w(3)}$. Therefore the vertex v is joined to $X^i_{w(1)}$ by at least

$$\frac{\overline{\delta}(b)}{640m^3} \cdot \frac{k_m}{k_{w(1)}} \cdot \frac{e(G)}{600k_m mn} \ge |X^i_{w(1)}| \frac{\overline{\delta}(b) e(G)}{4 \cdot 10^5 m^4 n^2}$$

edges of $G_{w(2)}$. Clearly, the same argument applied to $Y^i_{w(1)}$ shows that v is joined to $Y^i_{w(1)}$ by at least $|Y^i_{w(1)}| \overline{\delta}(b) e(G)/(4 \cdot 10^5 m^4 n^2)$ edges of $G_{w(3)}$. Since $\delta = \overline{\delta}(b)/(1600m^3)$, we conclude that there are at least $g = \lceil \delta k_{w(1)} \rceil$ pairwise disjoint pairs $(X^i_{w(1)}, Y^i_{w(1)})$ of non-exceptional partition classes of $\widetilde{\Pi}_{w(1)}$ such that every such pair is $(\varepsilon, G_{w(1)})$-regular, between every such pair there are at least $\overline{\delta}(b)|U^i_{w(1)}| \cdot |W^i_{w(1)}| e(G)/(10^6 m^4 n^2)$ edges of $G_{w(1)}$, and one set of every such pair is joined to v by at least

$$|X^i_{w(1)}| \frac{\overline{\delta}(b) e(G)}{4 \cdot 10^5 m^4 n^2} \ge \frac{\overline{\delta}(b) |X^i_{w(1)}| e(G)}{10^6 m^4 n^2} = \frac{\overline{\delta}(b) |Y^i_{w(1)}| e(G)}{10^6 m^4}$$

edges of $G_{w(2)}$, while the other is joined to v by at least $\overline{\delta}(b)|Y^i_{w(1)}| e(G) \times (10^6 m^4 n^2)^{-1}$ edges of $G_{w(3)}$. Finally, if u is the common cardinality of the sets $X^i_{w(1)}$, $Y^i_{w(1)}$ $(1 \le i \le g)$, then $2gu \ge \delta n$. Therefore $1 \le w(1) < w(2) < w(3) \le m$, together with $v \in V(G)$, and $\{(X^i_{w(1)}, Y^i_{w(1)}) : 1 \le i \le g\}$ form a $(\delta, \widetilde{G})$-flower, and Lemma 10 follows. ∎

4. Random subgraphs of uniform bipartite graphs. In this section we give a result that, although a little technical, may be of independent interest. Namely, we prove that, under quite weak hypotheses, with very large probability a random induced subgraph H' of a bipartite uniform graph H contains a fair number of edges. In fact, we show that, in selecting the random subgraph H' in question, we may allow an "adversary" to "mark",

during the selection process, a few vertices of H as forbidden vertices for H' so as to minimize our chances of getting many edges in our random subgraph; even with this rule the probability that we fail to get a few edges in H' is essentially super-exponentially small in the number of edges of the original bipartite graph H.

Let $H = H(u, \varrho, \varepsilon)$ be an ε-uniform bipartite graph with bipartition $V(H) = V_1 \cup V_2$, where $|V_1| = |V_2| = u \geq 1$, and with edge-density $e(H)u^{-2} = \varrho$. Let $d_1, d_2 \leq u$ be two given positive integers. Now select a random induced subgraph of H in the following manner. First, an adversary chooses a set $S_1 \subset V_1$ with $|S_1| \leq u/\log\log u$. Then we randomly pick a set $D_1 \subset V_1 \setminus S_1$ with $|D_1| = d_1$, with all the d_1-subsets of $V_1 \setminus S_1$ equiprobable. Next, under the full knowledge of the sets S_1 and D_1, our adversary picks a set $S_2 \subset V_2$ with $|S_2| \leq u/\log\log u$, and we randomly pick a set $D_2 \subset V_2 \setminus S_2$ with $|D_2| = d_2$, with all the d_2-subsets of $V_2 \setminus S_2$ equiprobable. Let us call the outcome of the above procedure a *random* $(d_1, d_2; S_1, S_2)$*-subgraph* of H, or simply a (d_1, d_2)*-subgraph* of H.

LEMMA 11. *For every constant* $0 < \beta \leq 1$, *there exist a constant* $0 < \varepsilon = \varepsilon(\beta) \leq 1$ *and a natural number* u_0 *such that, for any real* $d \geq 2(u/\varepsilon)^{1/2}$ *and any given graph* $H = H(u, \varrho, \varepsilon)$ *as above with* $u \geq u_0$ *and* $\varrho \geq d/u$, *the following assertion holds. If* $d_1, d_2 \geq d$, *regardless of the choices for* S_1 *and* S_2 *of our adversary, the probability that a random* $(d_1, d_2; S_1, S_2)$*-subgraph of* H *fails to contain at least* $d/2$ *edges is at most* β^d.

Proof. Given $0 < \beta \leq 1$, we choose $\varepsilon = \beta^2/16 > 0$ and show that this choice of $\varepsilon > 0$ will do. In the sequel, we assume that u is large enough for our inequalities to hold. Let our adversary choose the set $S_1 \subset V_1$. Recall that $|S_1| \leq u/\log\log u$. We show first that the set U of those vertices of V_2 that are adjacent to the vertices in our random set $D_1 \subset V_1 \setminus S_1$ is larger than $(1-\varepsilon)n$ with probability at least $1 - (4\varepsilon)^{d/2}$. In order to do so we generate the vertices of D_1 one by one and prove that, typically, in each step we enlarge the set U by a fair number of vertices.

Let us randomly choose a vertex v^1 among all the vertices of $V_1 \setminus S_1$ to be the first vertex of D_1. Denote by W^1 the set of the vertices of V_1 that have fewer than $d\varepsilon/2 < (1-\varepsilon)d$ neighbours in V_2. Then, by the ε-regularity of (V_1, V_2), we have $|W^1| \leq \varepsilon u$. If v^1 belongs to W^1, let us say that it is a *bad* vertex, whereas if $v^1 \notin W^1$ let us say that it is a *good* vertex. Finally, let $U^1 \subseteq V_2$ be the set of neighbours of v^1 in V_2.

Similarly, suppose that for some $2 \leq i \leq d_1$ the vertices $v^1, \ldots, v^{i-1}$ have already been put into D_1. We randomly pick a vertex v^i from $V_1 \setminus (\{v^1, \ldots, v^{i-1}\} \cup S_1)$ to be the ith vertex of D_1 and denote by $U^{\leq i-1}$ the set of neighbours of $v^1, \ldots, v^{i-1}$. Then, if $|U^{\leq i-1}| \leq (1-\varepsilon)u$, we let W^i be the set of all the vertices in V_1 that have fewer than $d\varepsilon/2 \leq (1-\varepsilon)\varepsilon d$

neighbours in $V_2 \setminus U^{\le i-1}$. Note that, by the ε-regularity of the pair (V_1, V_2), we have $|W^i| \le \varepsilon u$. As before, we call v^i *bad* if it belongs to W^i, and *good* otherwise. Moreover, in the case where $|U^{\le i-1}| > (1-\varepsilon)u$, we always say that v^i is good. This process is continued until all the d_1 elements of D_1 have been chosen.

Now, suppose that our process has terminated with a set D_1 with $U = U^{\le d_1}$ of cardinality $|U| < (1-\varepsilon)u$. Since each good vertex v^i increases the size of the neighbourhood of D_1 in V_2 by at least $d\varepsilon/2$, the number of good elements in D_1 must be less than $2u/(d\varepsilon) \le d_1/2$. Hence, at least half of all the elements of D_1 must be bad, but the probability of this happening is smaller than $2^{d_1}\varepsilon^{d_1/2} \le (4\varepsilon)^{d/2}$. Thus, $|U| \ge (1-\varepsilon)n$ with probability at least $1-(4\varepsilon)^{d/2}$.

Assume now that our process has terminated with a set D_1 with $U = U^{\le d_1}$ of cardinality $|U| \ge (1-\varepsilon)u$. We now let our adversary pick his set $S_2 \subset V_2$. Then the probability that at least half of the $d_2 \ge d$ vertices of D_2 should lie outside U is at most $2^{d_2}(\varepsilon n/(n-|S_2|))^{d_2/2}$, which, for sufficiently large u, is less than $2^{d_2}(2\varepsilon)^{d_2/2} \le (8\varepsilon)^{d/2}$.

Thus, the probability that our random (d_1, d_2)-subgraph of H contains fewer than $d/2$ edges is bounded from above by $(4\varepsilon)^{d/2} + (8\varepsilon)^{d/2} \le (16\varepsilon)^{d/2} = \beta^d$, as required. ∎

5. A sufficient condition for uniformity. Let G be an l-partite graph with l-partition $V(G) = V_1 \cup \ldots \cup V_l$ ($l \ge 2$). Recall that G is ε-uniform if all pairs (V_i, V_j) $(1 \le i < j \le l)$ are ε-regular. Moreover, observe that in order to check the ε-regularity for any such pair (V_i, V_j), in principle one must examine the density of many pairs (A, B) with $A \subseteq V_i$, $B \subseteq V_j$. However, it turns out that the ε-regularity of (V_i, V_j) is implied by a rather simple condition imposed upon the intersection of the neighbourhoods of pairs of vertices. This idea has been exploited in many places; see, e.g., Alon, Duke, Lefmann, Rödl and Yuster [ADLRY 94], Frankl, Rödl and Wilson [FRW 88] and Thomason [Th 87a] (see also [Th 87b]). The following fact is a slight refinement of earlier results in [Th 87a] and [FRW 88].

LEMMA 12. *Let G be a d-regular bipartite graph with bipartition $V(G) = V_1 \cup V_2$, where $|V_1| = |V_2| = n$ and $d = pn$ $(0 < p \le 1)$. Assume that for a subset B of V_2 with b vertices and some $\varepsilon > 0$ we have*

$$\sum_{x,x' \in B} |N(x) \cap N(x')| \le (1+\varepsilon)\binom{b}{2}np^2, \tag{1}$$

where the sum is taken over all unordered pairs $x, x' \in B$ with $x \ne x'$. Then, for every subset A of V_1 with a vertices, we have

$$(e(A,B) - abp)^2 \le \varepsilon a(n-a)b^2p^2 + abnp.$$

P r o o f. Suppose $V_1 = \{x_1, \dots, x_n\}$ and let d_i $(1 \le i \le n)$ be the number of neighbours the vertex x_i has in B. For simplicity of notation, assume that $A = \{x_1, \dots, x_a\}$. Then, since G is d-regular, we have

$$\sum_{i=1}^{n} d_i = e(V_1, B) = db = e(A, B) + e(V_1 \setminus A, B).$$

Furthermore, using (1), counting directed paths of length two with both ends in B leads to

$$(1+\varepsilon)b(b-1)np^2 \ge \sum_{i=1}^{n} d_i(d_i-1) = \sum_{i=1}^{a} d_i(d_i-1) + \sum_{i=a+1}^{n} d_i(d_i-1).$$

Since $\sum_{i=1}^{a} d_i = e(A, B)$ and $\sum_{i=a+1}^{n} d_i = e(V_1 \setminus A, B) = db - e(A, B)$, by the Cauchy–Schwarz inequality we have

$$\begin{aligned}(1+\varepsilon)b(b-1)np^2 \ge{}& \frac{1}{a}e(A,B)(e(A,B)-a)\\ &+ \frac{1}{n-a}(db-e(A,B))(db-e(A,B)-n+a),\end{aligned}$$

which, after elementary calculations, may be reduced to

$$abpn(n-a)(1-p) + \varepsilon ab(b-1)p^2n(n-a) \ge n(e(A,B)-abp)^2,$$

from which the assertion easily follows. ∎

As an immediate consequence of the above result we obtain a simple sufficient condition for the η-uniformity of a regular bipartite graph G.

Lemma 13. *Let G be a d-regular bipartite graph with bipartition $V(G) = V_1 \cup V_2$, where $|V_1| = |V_2| = n$, $d = pn$, and $p = \omega/n$ with $\omega = \omega(n) \to \infty$ as $n \to \infty$. Suppose that for some constant $0 < \varepsilon < 1/2$ independent of n we have*

$$\sum_{x,x'} (|N(x) \cap N(x')| - (1+\varepsilon)np^2) \le 4\varepsilon^3 n^3 p^2, \tag{2}$$

where the sum is taken over all unordered pairs x, $x' \in V_2$ with $x \ne x'$ and $|N(x) \cap N(x')| \ge (1+\varepsilon)np^2$. Then G is $8\varepsilon^{1/3}$-uniform provided n is large enough.

P r o o f. Take any $A \subseteq V_1$, $B \subseteq V_2$ of sizes $a = |A| \ge 2\varepsilon^{1/3}n$ and $b = |B| \ge 2\varepsilon^{1/3}n$. Then, writing $\sum'_{x,x' \in B}$ for the sum over all unordered pairs of distinct vertices $x, x' \in B$, and $\sum''_{x,x' \in V_2}$ for the sum over all unordered pairs of distinct vertices $x, x' \in V_2$ with $|N(x) \cap N(x')| \ge (1+\varepsilon)np^2$, we deduce from (2) that

$$\sum_{x,x'\in B}{}' |N(x)\cap N(x')|$$
$$\le (1+\varepsilon)\binom{b}{2}np^2 + \sum_{x,x'\in V_2}{}'' (|N(x)\cap N(x')| - (1+\varepsilon)np^2)$$
$$\le (1+\varepsilon)\binom{b}{2}np^2 + 4\varepsilon^3 n^3p^2 \le (1+\varepsilon+2\varepsilon^2)\binom{b}{2}np^2 \le (1+2\varepsilon)\binom{b}{2}np^2.$$

Furthermore, since $\omega = pn \to \infty$ as $n \to \infty$, for large enough n we have $abpn \le \varepsilon anb^2p^2$ and hence, by Lemma 12,

$$|e(A,B) - abp| \le \sqrt{4\varepsilon anb^2p^2} \le 2\varepsilon^{1/3}abp.$$

Now, to complete the proof, it is enough to observe that, according to our definition of density, $d_G(A,B) = 4e(A,B)/(abp)$ while $d_G(V_1,V_2) = 4$. Hence

$$|d_G(A,B) - d_G(V_1,V_2)| = \frac{4|e(A,B)-abp|}{abp} \le 8\varepsilon^{1/3},$$

as required. ∎

6. Difference-graphs. Let A be a subset of $[n] = \{0,1,\dots,n-1\}$. The associated *difference-graph* $G_A = G_A(n)$ is the bipartite graph with bipartition $V(G_A) = V \cup W$, where both V and W are copies of $[n]$, and for $v \in V$ and $w \in W$ the pair $\{v,w\}$ is an edge of G_A if and only if $w \equiv v + a \pmod n$ for some $a \in A$. It is immediate that each vertex of G_A has degree $|A|$, and that the number of common neighbours of two distinct vertices v and v' that belong to the same class of the bipartition depends only on the value of $v - v'$. In fact, this value is the same as the number $t_A(v-v')$ of ordered pairs $(a,a') \in A \times A$ such that $a' - a \equiv v - v' \pmod n$. Therefore, the structure of G_A is closely related to the behaviour of the numbers $t_A(j)$ $(1 \le j < n)$. Our next result deals with the distribution of the $t_R(j)$ for a random set $R \in \mathcal{R}(n,p)$. In the sequel, we write $\oplus$ and $\ominus$ for addition and subtraction modulo n, respectively.

Lemma 14. *For every fixed $0 < \varepsilon \le 1$ and $0 < \eta \le 1$ there exists a constant $C = C(\varepsilon,\eta)$ for which the following holds. For every $p = p(n) \ge C/\sqrt{n}$, the probability that $R \in \mathcal{R}(n,p)$ satisfies*

$$\sum \{t_R(j) - (1+\varepsilon)np^2\} \le \eta n^2p^2, \tag{3}$$

where the sum is taken over all $1 \le j < n$ such that $t_R(j) \ge (1+\varepsilon)np^2$, tends to 1 as $n \to \infty$.

Proof. For $i \in [n]$, let I_i be the characteristic function of the event $\{i \in R\}$. Thus I_i is a 0–1 random variable with $I_i = 1$ if and only if $i \in R$. For a given $1 \le j < n$, let us divide the set of all n pairs $(i, i\oplus j)$ $(i \in [n])$ into three classes $B^1(j)$, $B^2(j)$ and $B^3(j)$ in such a way that $\lfloor n/3 \rfloor \le |B^l(j)| \le \lceil n/3 \rceil$

for all $l \in \{1,2,3\}$, and such that, for any $i \in [n]$, no $B^l(j)$ contains both pairs $(i \ominus j, i)$ and $(i, i \oplus j)$. The fact that such a partition is always possible follows from the simple observation that every 2-regular graph admits a proper 3-edge-colouring in which the sizes of any two colour classes differ by at most one. In fact, for later convenience, we may and shall further require of the $B^l(j)$ that, for all $l \in \{1,2,3\}$, if $j_1 + j_2 = n$, then the set of unordered pairs that naturally correspond to the elements of $B^l(j_1)$ should be the same as the corresponding set for $B^l(j_2)$. Now, for all $1 \le j < n$ and $l \in \{1,2,3\}$, define the random variables $X(j,l)$ and $\widehat{X}(j,l)$ by setting

$$X(j,l) = \sum_{(i,i\oplus j)\in B^l(j)} I_i I_{i\oplus j},$$

and

$$\widehat{X}(j,l) = \begin{cases} 0 & \text{if } X(j,l) < (1+\varepsilon)|B^l(j)|p^2, \\ X(j,l) & \text{otherwise} \end{cases}$$

or, briefly, $\widehat{X}(j,l) = \mathbf{1}(j,l)X(j,l)$, where $\mathbf{1}(j,l)$ is the characteristic function of the event $\{X(j,l) \ge (1+\varepsilon)|B^l(j)|p^2\}$. Then, clearly, we have $t_R(j) = X(j,1) + X(j,2) + X(j,3)$ for all $1 \le j < n$. Notice now that the left-hand side of (3) is at most

$$\sum_{l=1}^{3}\sum_{j=1}^{n-1}(\widehat{X}(j,l) - \mathbf{1}(j,l)(1+\varepsilon)|B^l(j)|p^2).$$

Hence it is enough to show that, with probability tending to 1 as $n \to \infty$, we have

$$Z = \sum_{l=1}^{3}\sum_{j=1}^{n-1}\widehat{X}(j,l) \le \eta n^2 p^2. \tag{4}$$

Let us estimate first the expectation $\mathrm{E}\widehat{X}(j,l)$ of $\widehat{X}(j,l)$. Note that $X(j,l)$ is a sum of independent 0–1 random variables, and thus it has binomial distribution $\mathrm{Bi}(m,p^2)$, where $m = |B^l(j)|$. (It was for achieving the above independence that the classes $B^1(j)$, $B^2(j)$ and $B^3(j)$ were introduced.) Therefore, setting $r_0 = \lceil (1+\varepsilon)p^2 m \rceil$, we get

$$\begin{aligned} \mathrm{E}(\widehat{X}(j,l)) &= \sum_{r=r_0}^{m} r\binom{m}{r}p^{2r}(1-p^2)^{m-r} \\ &= mp^2\sum_{r=r_0}^{m}\binom{m-1}{r-1}p^{2r-2}(1-p^2)^{m-r} \\ &= mp^2\sum_{r=r_0-1}^{m-1}\binom{m-1}{r}p^{2r}(1-p^2)^{m-r-1} \\ &= mp^2 b_{\ge}(r_0-1; m-1, p^2), \end{aligned} \tag{5}$$

where $b_{\geq}(r_0-1; m-1, p^2)$ is the probability that a binomial random variable with parameters $m-1$ and p^2 takes values greater than or equal to r_0-1. The behaviour of the function $b_{\geq}$ is, of course, well studied, and it is known that for every choice of m', p' and k' such that $k' \leq m'p'$ we have

$$b_{\geq}(m'p'+k'; m', p') \leq \exp\Big(-\frac{(k')^2}{3m'p'}\Big).$$

(See, for instance, Section 6 in McDiarmid [McD 89].) Thus, in our case,

$$b_{\geq}(r_0-1; m-1, p^2) \leq \exp\Big(-\frac{(\varepsilon p^2 m-1)^2}{3(m-1)p^2}\Big) \leq \exp\Big\{-\frac{1}{10}\varepsilon^2 np^2\Big\};$$

here and below we assume that n is large enough for all our inequalities to hold. Hence $\mathrm{E}\widehat{X}(j,l) \leq |B^l(j)|p^2 \exp(-\varepsilon^2 np^2/10)$ and, consequently,

$$\mathrm{E}(Z) = \mathrm{E}\Big(\sum_{l=1}^{3}\sum_{j=1}^{n-1}\widehat{X}(j,l)\Big) \leq n^2p^2 \exp\Big\{-\frac{1}{10}\varepsilon^2 np^2\Big\}.$$

Thus, if $p = p(n)$ is such that $np^2 \to \infty$ as $n \to \infty$, then the right-hand side of the above inequality is $o(n^2p^2)$. Therefore, by Markov's inequality, relation (4) holds with probability $1-o(1)$ as $n \to \infty$, and hence our lemma follows in this case.

In order to complete the proof, it is enough to consider the case when $p = p(n) > C/\sqrt{n}$ for some large $C > 0$ but, say, $np^2 < \log\log n$. Thus we henceforth assume that $p = p(n)$ satisfies these conditions. In the remaining, rather technical part of the proof, we shall compute the variance of $Z = \sum_{l=1}^{3}\sum_{j=1}^{n-1}\widehat{X}(j,l)$ for such a p and then show that this random variable is concentrated around its expectation through a direct application of Chebyshev's inequality.

First, let us note that with $m = |B^l(j)|$ as before, the variance $\mathrm{Var}(\widehat{X}(j,l))$ of $\widehat{X}(j,l)$ is no greater than

$$\begin{aligned}
&\sum_{r=r_0}^{m} r^2\binom{m}{r}p^{2r}(1-p^2)^{m-r} \\
&\quad= \sum_{r=r_0}^{m} r(r-1)\binom{m}{r}p^{2r}(1-p^2)^{m-r} + \mathrm{E}\widehat{X}(j,l) \\
&\quad= m(m-1)p^4 \sum_{r=r_0-2}^{m-2}\binom{m-2}{r}p^{2r-2}(1-p^2)^{m-r-2} + \mathrm{E}\widehat{X}(j,l) \\
&\quad= m(m-1)p^4 b_{\geq}(r_0-2; m-2, p^2) + mp^2 b_{\geq}(r_0-1; m-1, p^2) \\
&\quad\leq n^2p^4 \exp\Big\{-\frac{1}{10}\varepsilon^2 np^2\Big\} = O(1).
\end{aligned}$$

Note that for $j_1, j_2 \in \{1, \dots, n-1\}$, if $j_1 + j_2 = n$, then $\widehat{X}(j_1, l) = \widehat{X}(j_2, l)$ for all $l \in \{1, 2, 3\}$. Thus the covariance $\mathrm{Cov}(\widehat{X}(j_1, l), \widehat{X}(j_2, l))$ between $\widehat{X}(j_1, l)$ and $\widehat{X}(j_2, l)$ coincides with $\mathrm{Var}(X(j_1, l)) = O(1)$. Now let $(j_1, l_1) \neq (j_2, l_2)$ and suppose also that if $j_1 + j_2 = n$ then $l_1 \neq l_2$. We shall estimate $\mathrm{Cov}(\widehat{X}(j_1, l_1), \widehat{X}(j_2, l_2))$. For simplicity, put $U_1 = \widehat{X}(j_1, l_1)$ and $U_2 = \widehat{X}(j_2, l_2)$. We have $\mathrm{Cov}(\widehat{X}(j_1, l_1), \widehat{X}(j_2, l_2)) = \mathrm{Cov}(U_1, U_2) = \mathrm{E}(U_1 U_2) - \mathrm{E}(U_1)\mathrm{E}(U_2)$. We shall first estimate from above the value of $\mathrm{E}(U_1 U_2)$. Recall that we write R for a random element of our probability space $\mathcal{R}(n, p)$. We have

$$\mathrm{E}(U_1 U_2) = \sum_{R_0 \subseteq [n]} U_1(R_0) U_2(R_0) \,\mathrm{Prob}(R = R_0). \tag{6}$$

Let us say that $R \in \mathcal{R}(n, p)$ is *exceptional* if either $U_1(R)$ or $U_2(R)$ is at least as large as $u_{\max} = \lfloor \log n \rfloor$. Standard inequalities for the tail of the binomial distribution give that the probability that $R \in \mathcal{R}(n, p)$ is exceptional is no greater than $n^{-(1/2)\log\log n}$. Since U_1 and U_2 are at most n, the exceptional R_0 contribute $n^{2-(1/2)\log\log n}$ to the sum in (6). In the sequel we concentrate our attention on non-exceptional $R \in \mathcal{R}(n, p)$.

Let us introduce some notation. Let G_s ($s \in \{1, 2\}$) be the directed graph with vertex set $[n]$ and edge set $B^{l_s}(j_s)$. Thus each G_s simply consists of some isolated edges. Moreover, the directed graph $G_1 \cup G_2$ contains no cycle of length 2. Let $H_s = H_s(R)$ ($s \in \{1, 2\}$) be the random subgraph of G_s induced by the elements of $R \in \mathcal{R}(n, p)$, and put $H = H(R) = H_1(R) \cup H_2(R) \subseteq G_1 \cup G_2$.

We now consider the structure of H. Let us say that $R \in \mathcal{R}(n, p)$ is *typical* if $H = H(R)$ is a matching. The probability that $R \in \mathcal{R}(n, p)$ is not typical is at most $np^3 \leq (\log\log n)^{3/2}/\sqrt{n}$. Thus, the contribution of the atypical, non-exceptional $R_0 \subseteq [n]$ in (6) is at most $u_{\max}^2 (\log\log n)^{3/2} n^{-1/2} = O((\log n)^2 (\log\log n)^{3/2} n^{-1/2})$. From here onwards we only take into account $R \in \mathcal{R}(n, p)$ that are typical and non-exceptional. Thus $H = H(R)$ will always consist of isolated edges.

To bound the contribution of the typical, non-exceptional $R_0 \subseteq [n]$ to the sum in (6), we shall need the following consequence of a large deviations inequality of Janson, Łuczak, and Ruciński [JŁR 90]. Let J be a graph with maximal degree at most 2, with $\overline{m}$ edges, and n_2 vertices of degree two. Let J^p be a random induced subgraph of J obtained by selecting its vertices randomly and independently, each with probability p. Then the inequality of Janson, Łuczak and Ruciński gives that the probability that J^p has no edges is at most $\exp\{-\overline{m}p^2 + 2n_2 p^3\}$.

We are now ready to bound $\mathrm{E}'(U_1 U_2) = \sum U_1(R_0) U_2(R_0) \,\mathrm{Prob}(R = R_0)$, where the sum is taken over all typical and non-exceptional $R_0 \subseteq [n]$. For

$s \in \{1,2\}$, let $m_s = |B^{l_s}(j_s)|$ and $u_s = \lceil (1+\varepsilon) m_s p^2 \rceil$. Then $\mathrm{E}'(U_1U_2)$ is at most

$$\begin{aligned}
&\sum_{u_1=r_1}^{u_{\max}} u_1 \binom{m_1}{u_1} \sum_{u_2=r_2}^{u_{\max}} u_2 \binom{m_2}{u_2} p^{2u_1+2u_2} \exp(-(m_1+m_2-3u_1-3u_2)p^2 + 2np^3) \\
&\quad \le \sum_{u_1=r_1}^{u_{\max}} u_1 \binom{m_1}{u_1} p^{2u_1}(1-p^2)^{m_1-u_1} \sum_{u_2=r_2}^{u_{\max}} u_2 \binom{m_2}{u_2} p^{2u_2}(1-p^2)^{m_2-u_2} \\
&\qquad \times \exp(2u_{\max}p^2 + np^4 + 2np^3) \\
&\quad \le (1 + O(n^{-1/2}(\log\log n)^{3/2})\mathrm{E}(U_1)\mathrm{E}(U_2).
\end{aligned}$$

Therefore

$$\begin{aligned}
\mathrm{Cov}(\widehat{X}(j_1,l_1), \widehat{X}(j_2,l_2)) &= \mathrm{Cov}(U_1,U_2) = \mathrm{E}(U_1U_2) - \mathrm{E}(U_1)\mathrm{E}(U_2) \\
&\le n^{2-(1/2)\log\log n} + O(n^{-1/2}(\log n)^2(\log\log n)^{3/2}) \\
&\quad + O(n^{-1/2}(\log\log n)^{7/2}) \\
&= O(n^{-1/2}(\log n)^2(\log\log n)^{3/2}).
\end{aligned}$$

Thus,

$$\begin{aligned}
\mathrm{Var}(Z) &= \mathrm{Var}\Big(\sum_{l=1}^{3}\sum_{j=1}^{n-1} \widehat{X}(j,l)\Big) = \sum_{l_1=1}^{3}\sum_{j_1=1}^{n-1}\sum_{l_2=1}^{3}\sum_{j_2=1}^{n-1} \mathrm{Cov}(\widehat{X}(j_1,l_1), \widehat{X}(j_2,l_2)) \\
&= 6nO(1) + 9n^2 O(n^{-1/2}(\log n)^3) = O(n^{3/2}(\log n)^3).
\end{aligned}$$

We now note that (5) and the trivial fact that

$$b_{\ge}(r_0-1; m-1, p^2) \ge \binom{m-1}{r_0-1} p^{2(r_0-1)}(1-p^2)^{m-r_0} \ge \exp\{-3np^2\}$$

imply that

$$\mathrm{E}(Z) = \sum_{l=1}^{3}\sum_{j=1}^{n-1} \mathrm{E}\widehat{X}(j,l) \ge n(n-1)p^2 \exp\{-3np^2\} \ge C^2 n/(2(\log n)^3).$$

Thus $\mathrm{Var}(Z) = o((\mathrm{E}(Z))^2)$ and, hence, from Chebyshev's inequality, with probability tending to 1 as $n \to \infty$, we have

$$Z \le 2\mathrm{E}(Z) \le 2n^2p^2 \exp(-\varepsilon^2 np^2/10).$$

Thus (4) holds with probability tending to 1 as $n \to \infty$ provided $C^2 \ge 10\varepsilon^{-2}\log(2/\eta)$. ∎

Lemmas 13 and 14 immediately imply that the difference graph $G_R = G_R(n)$ for a random subset $R \in \mathcal{R}(n,p)$ is η-uniform, provided the probability p is large enough.

FACT 15. *For every* $0 < \eta \leq 1$ *there exists a constant* $C = C(\eta)$ *such that, if* $p = p(n) \geq C/\sqrt{n}$ *and* $R \in \mathcal{R}(n,p)$, *then the bipartite graph* $G_R = G_R(n)$ *is* η*-uniform with probability tending to* 1 *as* $n \to \infty$.

7. Proof of Theorem 1. Let S be a subset of $[n] = \{0, \ldots, n-1\}$, where n is an odd natural number. Following Ruzsa and Szemerédi [RSz 78] (see also Erdős, Frankl and Rödl [EFR 86] and Graham and Rödl [GR 87]) we introduce a graph $G(n,S)$ that reflects the arithmetic structure of S. Thus, $G(n,S)$ is a tripartite graph whose vertex set consists of three copies V_1, V_2, V_3 of the set $[n]$ and $\{i,j\}$ is an edge of $G(n,S)$ if and only if one of the following three conditions holds:

(i) $i \in V_1$, $j \in V_2$ and $j = i \oplus k$ for some $k \in S$,
(ii) $i \in V_2$, $j \in V_3$ and $j = i \oplus k$ for some $k \in S$,
(iii) $i \in V_1$, $j \in V_3$ and $j = i \oplus 2k$ for some $k \in S$;

here and below $\oplus$ and $\ominus$ stand for addition and subtraction modulo n. Clearly, if $k \in S$ then the vertices $i \in V_1$, $i \oplus k \in V_2$ and $i \oplus 2k \in V_3$ induce a triangle in $G(n,S)$. A triangle of $G(n,S)$ of this type is said to be *trivial*. We are interested in the non-trivial, or *spontaneous*, triangles of $G(n,S)$, since they reflect the arithmetic structure of S in the sense made precise below.

Clearly, each $G(n,S)$ contains precisely $n|S|$ trivial triangles, and in fact $G(n,S)$ is the edge-disjoint union of those triangles. More importantly, the number of spontaneous triangles in $G(n,S)$ depends on the number of *arithmetic triples* in S, that is, triples of *distinct* elements $a,b,c \in S$ such that $a \ominus c = c \ominus b$. Indeed, for any arithmetic triple $\Delta = (a,b,c)$ and every $i \in [n]$, the graph $G(n,S)$ contains the associated spontaneous triangle with vertices $i \in V_1$, $i \oplus a \in V_2$ and $i \oplus a \oplus b \in V_3$, and, conversely, any spontaneous triangle is associated with such a pair (Δ, i). Thus, in order to verify whether S contains an arithmetic triple it is enough to look for a spontaneous triangle in $G(n,S)$.

Naturally, we shall be particularly interested in the structure of $G(n,R)$ for a random subset R of $[n]$. It will be later crucial that, for a large random set $R \subseteq [n]$, the graph $G(n,R)$ is typically uniform, and hence sparse, as shown by our next two results. In the sequel, it will be convenient to extend the definition of η-uniformity to subsets of $[n]$ in the obvious way. If $S \subseteq [n]$ is such that $G(n,S)$ is η-uniform for some $0 < \eta \leq 1$, then let us say that S itself is η*-uniform*. Moreover, given $b > 2$ and $0 < \eta \leq 1$, we define the notion of (b,η)-sparseness for S above in the analogous way.

FACT 16. *For every* $0 < \eta \leq 1$ *there exists a constant* $C = C(\eta)$ *such that, if* $p = p(n) \geq C/\sqrt{n}$, *then the probability that* $R \in \mathcal{R}(n,p)$ *is* η*-uniform tends to* 1 *as* $n \to \infty$ *along the odd integers.*

P r o o f. Since for odd n the graph $G_{2R} = G_{2R}(n)$ may be identified with $G_{\bar{R}} = G_{\bar{R}}(n)$ for some $\overline{R} \in \mathcal{R}(n,p)$, with the map $R \mapsto \overline{R}$ measure preserving, the assertion follows from Fact 15 and the definition of $G(n,R)$. ∎

From Fact 16 one may deduce the following result concerning the uniformity and sparseness of $R_M \in \mathcal{R}(n,M)$ for large M.

FACT 17. *For every $0 < \eta \le 1$ there exists a constant $C = C(\eta)$ such that, if $M = M(n) \ge C\sqrt{n}$, then the probability that $R \in \mathcal{R}(n,M)$ is η-uniform and $(4,\eta)$-sparse tends to 1 as $n \to \infty$ along the odd integers.*

P r o o f. We start by noticing that, for any $0 < \eta \le 1$, if $0 < \eta_0 \le 1$ is small enough, then for any $S \subseteq [n]$ the fact that the graph $G = G(n,S)$ is η_0-uniform implies that G is, say, $(4,\eta)$-sparse. Therefore we proceed to show that for any $0 < \eta_0 \le 1$, if $Cn^{1/2} \le M = M(n) \le n$ for some sufficiently large constant C, then $G = G(n,R_M)$ is η_0-uniform with probability $1-o(1)$ as $n \to \infty$ along the odd integers.

Pick $\eta_1 = \eta_0/6$ and $\varepsilon = \eta_1^2/3$. Let $C = (1+\varepsilon)C_1$, where $C_1 = C(\eta_1)$ is as given by Fact 16, and assume that $C\sqrt{n} \le M = M(n) \le n$. Set $p = p(n) = M/(1+\varepsilon)n$. We may generate $R_M \in \mathcal{R}(n,M)$ by picking $R_p \in \mathcal{R}(n,p)$ conditioned on R_p satisfying $|R_p| \le M$, and then by adding random elements of $[n]\setminus R_p$ to R_p to obtain a set R_M of cardinality M. Since with probability $1-o(1)$ as $n \to \infty$, we have $|R_p| \ge (1-\varepsilon)pn$, we shall assume that our R_p does satisfy this condition. Assume also that $G = G(n,R_p)$ is η_1-uniform, and recall that by Fact 16 this event also holds with probability $1-o(1)$. It now suffices to show that, under these two conditions on R_p, the set R_M is η_0-uniform whatever elements were added to R_p to generate R_M.

Write G_M for $G(n,R_M)$ and G_p for $G(n,R_p)$. Let $U, W \subset V(G_M)$ be two disjoint sets contained in two distinct vertex classes of G_M with $|U|, |W| \ge \eta_1 n$. Put $\varrho_M = M/n$ and $\varrho_p = |R_p|/n$. Note that then $(1-2\varepsilon)\varrho_M \le \varrho_p \le \varrho_M$ and that $|\varrho_p - p| \le \varepsilon p$. Moreover, we have

$$e_{G_M}(U,W) \ge e_{G_p}(U,W) \ge (1-\eta_1)\varrho_p|U|\cdot|W| \ge (1-2\eta_1)\varrho_M|U|\cdot|W|$$

and

$$\begin{aligned} e_{G_M}(U,W) &\le e_{G_p}(U,W) + 2\varepsilon pn|U| \\ &\le (1+\eta_1)\varrho_p|U|\cdot|W| + 2(\varepsilon p/\eta_1)|U|\cdot|W| \\ &\le (1+2\eta_1)\varrho_p|U|\cdot|W| \le (1+2\eta_1)\varrho_M|U|\cdot|W|. \end{aligned}$$

Now notice that

$$d_{G_M}(U,W) = 3e_{G_M}(U,W)/(\varrho_M|U|\cdot|W|)$$

and the density between two any sets of tripartition of $G(n, R_M)$ is 3. Thus,

$$|d_{G_M}(U,W) - 3| \le \frac{3|e_{G_M}(U,W) - \varrho_M|U| \cdot |W||}{\varrho_M|U| \cdot |W|} \le 6\eta_1 = \eta_0,$$

and so $G_M = G(n, R_M)$ is indeed η_0-uniform, as required. ∎

In the sequel, it will be necessary for us to view $\mathcal{R}(n, M)$ as resembling a product of a large number of spaces. Assume that the integer m divides M, and put $M_0 = M/m$. We then define the space $\widetilde{\mathcal{R}}(n, m, M_0)$ as the uniform space of m-tuples $\widetilde{R} = (R_i)_{i=1}^m$ of pairwise disjoint M_0-subsets $R_i \subseteq [n]$. Thus all m-tuples $\widetilde{R} \in \widetilde{\mathcal{R}}(n, m, M_0)$ are equiprobable, and the map $\widetilde{R} = (R_i)_{i=1}^m \in \widetilde{\mathcal{R}}(n, m, M_0) \mapsto \bigcup_{i=1}^m R_i \in \mathcal{R}(n, M)$ is measure-preserving. We shall also consider the probability space $\widetilde{\mathcal{G}} = \widetilde{\mathcal{G}}(n, m, M_0)$ of the balanced 3-decomposable m-edge-coloured graphs $\widetilde{G} = \widetilde{G}(n, \widetilde{R}) = (G_i)_{i=1}^m$ determined by the $G_i = G(n, R_i)$ $(1 \le i \le m)$, where $\widetilde{R} = (R_i)_{i=1}^m$ is a random element of $\widetilde{\mathcal{R}}(n, m, M_0)$. In this space we consider the event $\mathcal{A}(b, \eta)$ that a graph from $\widetilde{\mathcal{G}}$ should be (b, η)-sparse, and denote the conditional probability space obtained from $\widetilde{\mathcal{G}}$ by conditioning on $\mathcal{A}(b, \eta)$ by $\widetilde{\mathcal{G}}(n, m, M_0 \mid b, \eta)$.

Lemma 18. *Let $m \ge 3$, $k_0 \ge 1$, $b > 4$, $0 < \xi \le 1$ and $0 < \delta \le 1$ be given. Then there exist constants $0 < \varepsilon = \varepsilon(m, \xi, \delta) \le 1$ and $C = C(m, k_0, b, \xi, \delta)$ for which the following holds for any sufficiently large n. Suppose $Cn^{1/2} \le M = mM_0 = mM_0(n) \le n/(\log\log n)^2$, let $\widetilde{G} = (G_i)_{i=1}^m \in \widetilde{\mathcal{G}}(n, m, M_0 \mid b, \eta)$ and let $\widetilde{\Pi}_1, \dots, \widetilde{\Pi}_m$ be the $(\mathfrak{e}, k_0)$-canonical sequence of partitions associated with $\widetilde{G}$, where as usual $\mathfrak{e} = \mathfrak{e}(m, b, \varepsilon, k_0)$. Then the probability that there exists a $(\delta, k_0, \mathfrak{e}; \widetilde{G})$-flower which contains no spontaneous triangles is smaller than ξ^M.*

Proof. Put $\beta = \xi^{10^8 m\delta^{-2}}$, and let $0 < \varepsilon = \varepsilon(\beta) = \varepsilon(m, \xi, \delta) \le 1$ be as given by Lemma 11. Furthermore, let η and $K_0 \ge 1/(2\varepsilon)$ be such that Lemma 5 holds and let $C_1 = C(\eta)$ be as in Fact 17. Finally, set $C = \max\{C_1, 10^7 m\sqrt{K_0}/(\delta\sqrt{\varepsilon})\}$. We shall show that for such a choice of ε, η and C the assertion holds. We also remark that in the sequel we tacitly assume that n is sufficiently large whenever it is needed.

Let us first restate our result in terms of m-coloured graphs $\widetilde{G} = (G_i)_{i=1}^m$ from $\widetilde{\mathcal{G}}(n, m, M_0)$. Let $\mathcal{B}(\delta)$ be the event that $\widetilde{G}$ should be (b, η)-sparse and moreover it should contain a $(\delta, k_0, \mathfrak{e}; \widetilde{G})$-flower without a spontaneous triangle. We have to show that

$$\mathrm{Prob}(\mathcal{B}(\delta) \mid \mathcal{A}(b, \eta)) \le \xi^M.$$

Suppose we show that

$$\mathrm{Prob}(\mathcal{B}(\delta)) \le \xi^M/2. \tag{7}$$

Then, by our choice of C, for any large enough n we have $\mathrm{Prob}(\mathcal{A}(b,\eta)) > 1/2$. Therefore

$$\mathrm{Prob}(\mathcal{B}(\delta) \mid \mathcal{A}(b,\eta)) = \frac{\mathrm{Prob}(\mathcal{B}(\delta) \cap \mathcal{A}(b,\eta))}{\mathrm{Prob}(\mathcal{A}(b,\eta))} \le \frac{\mathrm{Prob}(\mathcal{B}(\delta))}{\mathrm{Prob}(\mathcal{A}(b,\eta))} \le \xi^M,$$

as required. Hence it only remains to prove (7).

Let us first estimate the probability that a fixed $(\delta,\widetilde{G})$-flower contains no spontaneous triangle. Thus, let the colours $1 \le w(1) < w(2) < w(3) \le m$, the vertex $v \in V(G)$, and the family $\{(X^{(i)},Y^{(i)}) : 1 \le i \le g\}$, where

$$u = |X^{(i)}| = |Y^{(i)}| \ge n/(2K_0),$$

and $2gu \ge \delta n$, form a $(\delta,\widetilde{G})$-flower. We may further assume that all the $X^{(i)}$ $(1 \le i \le g' = \lceil g/2 \rceil)$ are contained in a single vertex class of the tripartition of G, as are all the $Y^{(i)}$ $(1 \le i \le g')$. Below we shall only consider $X^{(i)}$ and $Y^{(i)}$ for $1 \le i \le g'$. Also, let us mention for completeness that G has $3n$ vertices and $e(G) = |E(G)| = 3nM = 3nmM_0$ edges.

Suppose now that the R_i $(1 \le i < w(2))$ have been chosen. Thus, in particular, the graphs $H_i = G_{w(1)}[X^{(i)},Y^{(i)}]$ $(1 \le i \le g')$, namely the bipartite subgraphs of $G_{w(1)}$ with bipartition $V(H_i) = X^{(i)} \cup Y^{(i)}$ and edge set $E_{G_{w(1)}}(X^{(i)},Y^{(i)})$, have been fixed. By the definition of a $(\delta,\widetilde{G})$-flower, each H_i $(1 \le i \le g')$ is ε-uniform, and for all $1 \le i \le g'$ we have $e(H_i) = e_{G_{w(1)}}(X^{(i)},Y^{(i)}) \ge \varrho_0 u^2$, where $\varrho_0 = 10^{-6}\delta M_0/n$.

Let us now pick $R_{w(2)}$, and study the neighbourhood $N^{(i)}_{w(2)}(v) \subset X^{(i)}$ $(1 \le i \le g')$ of the vertex v in the graph $G_{w(2)}$ inside the set $X^{(i)}$. Put $d = \varrho_0 u$. Again by the definition of a $(\delta,\widetilde{G})$-flower, we know that $d^{(i)}_1 = |N^{(i)}_{w(2)}(v)| \ge d = \varrho_0 u$. Now let us condition on the values of the $d^{(i)}_1$ $(1 \le i \le g')$. Once these g' numbers are fixed, for every $1 \le i \le g'$, all the subsets of cardinality $d^{(i)}_1$ of the set $X^{(i)} \setminus S^{(i)}_1$ are equally likely to be chosen as $N^{(i)}_{w(2)}(v)$, where $S^{(i)}_1$ is the neighbourhood of v within $X^{(i)}$ in the graph $\bigcup_{1\le k<w(2)} G_k$. Furthermore, we make the simple but important observation that, because we choose $R_{w(2)}$ randomly and uniformly from all the M_0-subsets of $[n] \setminus \bigcup_{1\le k<w(2)} R_k$, and because we have decided on the cardinalities $d^{(i)}_1$ $(1 \le i \le g')$ in advance, the sets $N^{(i)}_{w(2)}(v)$ $(1 \le i \le g')$ are all selected independently.

Suppose now that the R_i $(w(2) \le i < w(3))$ have also been chosen, and pick $R_{w(3)}$. Let $N^{(i)}_{w(3)}(v) \subset Y^{(i)}$ be the $G_{w(3)}$-neighbourhood of v within $Y^{(i)}$. Put $d^{(i)}_2 = |N^{(i)}_{w(3)}(v)|$ $(1 \le i \le g')$, and note that again $d^{(i)}_2 \ge d = \varrho_0 u$. We now condition on the values of the $d^{(i)}_2$ $(1 \le i \le g')$. As above, under this conditioning, for every $1 \le i \le g'$, all the subsets of cardinality $d^{(i)}_2$ of

the set $Y^{(i)} \setminus S_2^{(i)}$ are equally likely to be chosen as $N^{(i)}_{w(3)}(v)$, where $S_2^{(i)}$ is the neighbourhood of v within $Y^{(i)}$ in the graph $\bigcup_{1\le k<w(3)} G_k$. As before, the sets $N^{(i)}_{w(3)}(v)$ $(1 \le i \le g')$ are all selected independently.

We now apply Lemma 11 to all the H_i $(1 \le i \le g')$. Recall that $d = \varrho_0 u$ and notice that, for all $1 \le i \le g'$, the density $e(H_i)u^{-2}$ of H_i is at least $\varrho_0 = d/u$ and $d_j^{(i)} \ge d$ $(j \in \{1,2\})$. Also, by the choice of C, we have $d \ge 2(u/\varepsilon)^{1/2}$. Moreover, since $M = mM_0 \le n/(\log\log n)^2$, for large enough n we have $|S_j^{(i)}| \le u/\log\log u$ for all $1 \le i \le g'$, $j \in \{1,2\}$. We now recall that the sets $N^{(i)}_{w(2)}(v)$ $(1 \le i \le g')$ are all selected independently, as are all the $N^{(i)}_{w(3)}(v)$ $(1 \le i \le g')$. Thus, applying Lemma 11 simultaneously to all the H_i $(1 \le i \le g')$, we see that the probability that we do not have a spontaneous triangle in G is at most $\beta^{dg'} \le \xi^{2M}$.

Now, to complete the proof, it suffices to estimate the number of all possible candidates for $(\delta, \widetilde{G})$-flowers in our (b,η)-sparse m-edge-coloured graph $\widetilde{G}$. Clearly, the vertex v can be selected in at most n ways, there are at most $\binom{m}{3} \le m^3$ possible choices for the indices $w(1)$, $w(2)$ and $w(3)$ and, since $|\widetilde{\Pi}_{w(1)}| \le K_0 + 1$ (cf. Lemma 5) and $g \le (|\widetilde{\Pi}_{w(1)}| - 1)/2 < K_0$, the number of possible choices for the set of pairs $\{(X^{(i)}, Y^{(i)}) : 1 \le i \le g\}$ can be estimated very generously by $K_0 \times K_0! \le (K_0+1)!$. Thus, since M grows to infinity at least as fast as $\sqrt{n}$, we have

$$\mathrm{Prob}(\mathcal{B}(\delta)) \le nm^3(K_0+1)!\xi^{2M} \le \xi^M/2$$

whenever n is large enough. Thus (7) holds and Lemma 18 is proved. ∎

As an almost immediate consequence of the above lemma and Lemma 10, we get the following result, which will be crucial for the proof of Theorem 1.

Let $\mathcal{R}(n, M \mid b, \eta)$ denote the uniform probability space whose elements are the (b,η)-sparse M-subsets R of $[n]$. Clearly, $\mathcal{R}(n, M \mid b, \eta)$ may be obtained from $\mathcal{R}(n, M)$ by conditioning on the event that $G(n, R)$ should be (b,η)-sparse. Let the associated probability space of the $G(n, R)$ $(R \in \mathcal{R}(n, M \mid b, \eta))$ be denoted by $\mathcal{G}(n, M \mid b, \eta)$. Thus to pick an element G from $\mathcal{G}(n, M \mid b, \eta)$ we simply generate $R \in \mathcal{R}(n, M \mid b, \eta)$ and let $G = G(n, R)$. Suppose the integer m divides M. Clearly, since all graphs from $\mathcal{G}(n, M \mid b, \eta)$ are generated by M-subsets of $[n]$ and each such subset can be decomposed into m subsets of size $M_0 = M/m$ in the same number of ways, one can generate an element of $\mathcal{G}(n, M \mid b, \eta)$ by choosing a graph from $\widetilde{\mathcal{G}}(n, m, M_0 \mid b, \eta)$ and ignoring the colouring of its edges.

Now our next result can be stated as follows.

LEMMA 19. *For every $b \ge 4$ and $0 < \xi \le 1$, there exist constants $0 < \eta(b,\xi) \le 1$, $C = C(b,\xi)$, and $N = N(b,\xi)$ such that, for every $n \ge N$*

and $C\sqrt{n} \le M = mM_0 = mM_0(n) \le n/(\log\log n)^2$, *where* $m = m(b)$ *is as given in Lemma* 10, *the probability that* $G \in \mathcal{G}(n, M \mid b, \eta)$ *contains no spontaneous triangle is at most* ξ^M.

P r o o f. Let $b \ge 4$ and $0 < \xi \le 1$ be given. Choose $m = m(b) \ge 3$, $0 < \delta = \delta(b) \le 1$ and $k_0 = k_0(b) \ge 1$ as in Lemma 10, and let $0 < \varepsilon = \varepsilon(m, \xi, \delta) \le 1$ be as in Lemma 18. Finally, let $0 < \eta = \eta(b, \varepsilon) \le 1$ be as given by Lemma 10, and let $C = C(m, k_0, b, \xi, \delta)$ be as given by Lemma 18. As always, we assume that n is sufficiently large whenever it is needed.

Observe that, because of the choice of m, k_0, and δ, every element of $\widetilde{\mathcal{G}}(n, m, M_0 \mid b, \eta)$ contains an $(\boldsymbol{\varepsilon}(m, b, \varepsilon, k_0), k_0)$-flower and, by Lemma 18, with probability at least $1 - \xi^M$ every such flower must contain a spontaneous triangle. Thus the probability that an element of $\widetilde{\mathcal{G}}(n, m, M_0 \mid b, \eta)$ contains no spontaneous triangle is smaller than ξ^M. As already mentioned, the graphs from $\widetilde{\mathcal{G}}(n, m, M_0 \mid b, \eta)$ naturally correspond to elements from $\mathcal{G}(n, mM_0 \mid b, \eta)$, and hence Lemma 19 follows. ∎

We can now finally prove Theorem 1.

P r o o f o f T h e o r e m 1. Clearly, it suffices to prove that, for any given $0 < \alpha \le 1$, there is a suitable choice for $C = C(\alpha)$ such that

(†) *if* $Cn^{1/2} \le M = M(n) \le n$, *then* $\lim_{n\to\infty} \mathrm{Prob}(R \to_\alpha 3) = 1$, *where the limit is taken along o d d values of* n.

Thus henceforth we may and shall assume that n is odd. Also, below we assume that n is sufficiently large whenever it is needed. For convenience, let us say that a property holds *almost surely* if it holds with probability tending to 1 as n tends to infinity along odd integers.

Let a constant $0 < \alpha \le 1$ be given, and assume that $1 \le M = M(n) \le n$. Note that the Heath-Brown–Szemerédi result mentioned in Section 0 implies that *any* set $A \subset [n]$ with $|A| \ge n/(\log\log n)^2$ contains a 3-term arithmetic progression provided n is sufficiently large. Thus we may and shall assume that $\alpha M \le n/(\log\log n)^2$, since otherwise $R \to_\alpha 3$ for any set $R \subseteq [n]$ with $|R| = M$.

Put $\alpha' = \alpha/2$. Set $b = 6/\alpha' \ge 4$ and $0 < \xi = \alpha'/4 \le 1$, and let $m = m(b) \ge 3$ be as given in Lemma 10. Moreover, let $0 < \eta = \eta(b, \xi) \le 1$ and $C_1 = C(b, \xi)$ be as given in Lemma 19, and let $C_2 = C(\eta)$ be as given in Fact 17. We let $C = C(\alpha') = \max\{(4/(3\alpha'))C_1, C_2\}$. We now show that (†) holds with this choice of C.

Our first aim is to verify that $R \in \mathcal{R}(n, M)$ almost surely has the property that any subset $A' \subseteq R$ with at least $\alpha'|R|$ elements contains an arithmetic triple. For simplicity, let us write $R \to'_{\alpha'} 3$ if R has this property.

Let us start by picking an integer multiple $M' = M'(n)$ of m such that $(3\alpha'/4)M \le M' \le \alpha' M$ holds for any sufficiently large n. Put $M_0 =$

$M_0(n) = M'/m$. We now consider the spaces $\mathcal{R}(n, M' \mid b, \eta)$ and $\mathcal{G}(n, M' \mid b, \eta)$. Note that $C_1 n^{1/2} \le M' = mM_0 \le n/(\log\log n)^2$, and hence that by Lemma 19 the probability that $G \in \mathcal{G}(n, M' \mid b, \eta)$ contains no spontaneous triangle is at most $\xi^{M'}$. Thus, the probability that $R \in \mathcal{R}(n, M' \mid b, \eta)$ contains no arithmetic triple is at most $\xi^{M'}$ for all large enough n. Therefore, the number of (b, η)-sparse subsets A' of $[n]$ with M' elements that do not contain arithmetic triples is at most $\xi^{M'}\binom{n}{M'}$.

Let $\mathcal{D}$ be the event that $R \in \mathcal{R}(n, M)$ should contain a (b, η)-sparse subset A' with M' elements which is free of arithmetic triples. Then the probability that $\mathcal{D}$ holds is at most

$$\xi^{M'}\binom{n}{M'}\binom{n-M'}{M-M'}\binom{n}{M}^{-1} \le \left(\frac{e\xi n}{M'}\right)^{M'}\left(\frac{M}{n}\right)^{M'} \le \left(\frac{4e\xi}{3\alpha'}\right)^{M'} = o(1).$$

Let now $\mathcal{S}$ be the event that $R \in \mathcal{R}(n, M)$ is $(4, \eta)$-sparse. Then, by Fact 17, $\mathcal{S}$ holds almost surely. We now note that if R is $(4, \eta)$-sparse, then any subset $A' \subseteq R$ with $|A'| = M'$ is (b, η)-sparse. Therefore, if $\mathcal{D}$ fails and $\mathcal{S}$ holds, then $R \to'_{\alpha'} 3$. Since almost surely $\mathcal{D}$ fails and $\mathcal{S}$ holds, we conclude that a random set $R \in \mathcal{R}(n, M)$ satisfies $R \to'_{\alpha'} 3$ almost surely.

Now recall that n is odd, and write $n = 2k + 1$. Observe that if A is a subset of R with at least $\alpha|R|$ elements, then at least one of the subsets $A_1 = A \cap \{0, \dots, k\}$ and $A_2 = A \cap \{k, \dots, 2k\}$ must have at least $\alpha'|R| = \alpha|R|/2$ elements, and that A_i ($i \in \{1, 2\}$) contains an arithmetic triple if and only if it contains an arithmetic progression of length three. Thus (†) does hold and Theorem 1 is proved. ∎

Corollary 3 may be deduced from Theorem 1 in a routine manner.

Sketch of the proof of Corollary 3. Let $s = s(n)$, $g = g(n)$, and α be as in the statement of our corollary. Pick C_0 sufficiently large so that with $p = p(n) = C_0 n^{-1/2}$ and $R = R_p \in \mathcal{R}(n, p)$ we have $R \to_{\alpha/2} 3$ with probability $1 - o(1)$ as $n \to \infty$.

The probability that a fixed set as in (i) of the corollary meets R in more than 3 elements is $O(s^4 n^{-2})$. Hence the expected number of such sets is $O(s^4) = o(\sqrt{n})$. The probability that a fixed set as in (ii) meets R in more than $Cm/\sqrt{n}$ elements is at most $\exp\{-2m/\sqrt{n}\}$ for large enough C. Therefore the probability that such a set exists is $o(1)$. Now, let $G^{(3)} = G_n^{(3)}$ be the 3-uniform hypergraph on $[n]$ whose hyperedges are the 3-term arithmetic progressions contained in $[n]$. One can easily check that for every $l \ge 2$ the number of cycles of length l in $G^{(3)}$ is at most $(3n)^l$. For any such cycle of $G^{(3)}$ the probability that it appears in $\mathcal{F} = \mathcal{F}(R)$ is p^{2l}. Therefore the expected number of cycles in $\mathcal{F}$ shorter than g is at most $\sum_{l=2}^{g}(3np^2)^l = o(\sqrt{n})$.

In view of the above remarks, with probability $1-o(1)$ as $n \to \infty$, there exists a set $S \subseteq R$ with $|S| \geq |R|/2$ satisfying (i)–(iii). Finally, note that if $R \to_{\alpha/2} 3$, then $S \to_{\alpha} 3$. ∎

References

[ADLRY 94] N. Alon, R. A. Duke, H. Lefmann, V. Rödl and R. Yuster, *The algorithmic aspects of the regularity lemma*, J. Algorithms 16 (1994), 80–109.

[EFR 86] P. Erdős, P. Frankl and V. Rödl, *The asymptotic number of graphs not containing a fixed subgraph and a problem for hypergraphs having no exponent*, Graphs Combin. 2 (1986), 113–121.

[ET 36] P. Erdős and P. Turán, *On some sequences of integers*, J. London Math. Soc. 11 (1936), 261–264.

[FRW 88] P. Frankl, V. Rödl and R. Wilson, *The number of submatrices of a given type in a Hadamard matrix and related results*, J. Combin. Theory Ser. B 44 (1988), 317–328.

[Fu 77] H. Furstenberg, *Ergodic behavior of diagonal measures and a theorem of Szemerédi on arithmetic progressions*, J. Anal. Math. 31 (1977), 204–256.

[GN 86] R. L. Graham and J. Nešetřil, *Large minimal sets which force long arithmetic progressions*, J. Combin. Theory Ser. A 42 (1986), 270–276.

[GR 87] R. L. Graham and V. Rödl, *Numbers in Ramsey theory*, in: Surveys in Combinatorics 1987, C. Whitehead (ed.), London Math. Soc. Lecture Note Ser. 123, Cambridge University Press, Cambridge, 1987, 111–153.

[HKŁ 95] P. E. Haxell, Y. Kohayakawa and T. Łuczak, *The induced size-Ramsey number of cycles*, Combin. Probab. Comput. 4 (1995), 217–239.

[H-B 87] D. R. Heath-Brown, *Integer sets containing no arithmetic progressions*, J. London Math. Soc. 35 (1987), 385–394.

[JŁR 90] S. Janson, T. Łuczak and A. Ruciński, *An exponential bound for the probability of nonexistence of a specified subgraph in a random graph*, in: Random Graphs '87, M. Karoński, J. Jaworski and A. Ruciński (eds.), Wiley, 1990, 73–87.

[McD 89] C. J. H. McDiarmid, *On the method of bounded differences*, in: Surveys in Combinatorics 1989, J. Siemons (ed.), London Math. Soc. Lecture Note Ser. 141, Cambridge University Press, Cambridge, 1989, 148–188.

[NR 87] J. Nešetřil and V. Rödl, *Partite construction and Ramseyan theorems for sets, numbers and spaces*, Comment. Math. Univ. Carolin. 28 (1987), 569–580.

[PV 88] H.-J. Prömel and B. Voigt, *A sparse Graham–Rothschild theorem*, Trans. Amer. Math. Soc. 309 (1988), 113–137.

[Rö 90] V. Rödl, *On Ramsey families of sets*, Graphs Combin. 6 (1990), 187–195.

[Ro 53] K. F. Roth, *On certain sets of integers*, J. London Math. Soc. 28 (1953), 104–109.

[RSz 78] I. Z. Ruzsa and E. Szemerédi, *Triple systems with no six points carrying three triangles*, Colloq. Math. Soc. János Bolyai 18 (1978), 939–945.

[Sp 75] J. Spencer, *Restricted Ramsey configurations*, J. Combin. Theory Ser. A 19 (1975), 278–286.

[Sz 75] E. Szemerédi, *On sets of integers containing no k elements in arithmetic progression*, Acta Arith. 27 (1975), 299–345.

[Sz 78] E. Szemerédi, *Regular partitions of graphs*, in: Problèmes Combinatoires et Théorie des Graphes, Proc. Colloque Inter. CNRS, J.-C. Bermond, J.-C. Fournier, M. Las Vergnas et D. Sotteau (eds.), CNRS, Paris, 1978, 399–401.

[Th 87a] A. Thomason, *Pseudo-random graphs*, in: Random Graphs '85, M. Karoński and Z. Palka (eds.), Ann. Discrete Math. 33, North-Holland, Amsterdam, 1987, 307–331.

[Th 87b] —, *Random graphs, strongly regular graphs and pseudo-random graphs*, in: Surveys in Combinatorics 1987, C. Whitehead (ed.), London Math. Soc. Lecture Note Ser. 123, Cambridge University Press, Cambridge, 1987, 173–195.

Instituto de Matemática e Estatística
Universidade de São Paulo
Rua do Matão 1010
05508-900 São Paulo, SP, Brazil

Department of Mathematics and Computer Science
Emory University
Atlanta, Georgia 30322
U.S.A.

Department of Mathematics
and Computer Science
Adam Mickiewicz University
Matejki 48/49
60-769 Poznań, Poland

Received on 9.6.1995 (2804)