Randomness and regularity

Tomasz Łuczak

Abstract. For the last ten years the theory of random structures has been one of the most rapidly evolving fields of discrete mathematics. The existence of sparse structures with good 'global' properties has been vital for many applications in mathematics and computer science, and studying properties of such objects led to many challenging mathematical problems. In the paper we report on recent progress on this subject related to some variants of Szemerédi's Regularity Lemma.

Mathematics Subject Classification (2000). Primary 05C80, 05D05; Secondary 05C35, 05C65, 05D40.

Keywords. Random graphs, extremal graph theory, regularity lemma, removal lemma, density theorems.

1. Introduction

In the paper we consider 'extremal' properties of families of sets, i.e., we study the size of maximal subfamilies of a given family with certain property. Let $[A]^r$ be the family of all r-sets (i.e. sets of r elements) contained in A; if $A = [n] = \{1, 2, \dots, n\}$ we put $[n]^r = [[n]]^r$. Two classical examples of extremal results for $[n]^r$ are Szemerédi's and Turán's theorems. Let us recall that Szemerédi's density theorem [17] states that if $r_k(n)$ denote the maximum size of a subset of $[n] = [n]^1$ which contains no non-trivial arithmetic progression of length k, then, $r_k(n) = o(n)$. In order to formulate Turán's theorem, we need some notation. For given r, s, n, $n \ge s > r \ge 2$, let $\mathrm{ex}([s]^r, [n]^r)$ denote the size of the largest family $\mathcal{A} \subseteq [n]^r$ such that for no set $B \subseteq [n]$, $|B| = s$, we have $[B]^r \subseteq \mathcal{A}$. Furthermore, let

$$\alpha(m, r) = \limsup_{n\to\infty} \frac{\mathrm{ex}([m]^r, [n]^r)}{\binom{n}{r}} \,. \tag{1}$$

Turán's theorem [21] states that $\alpha(m, 2) = \frac{m-2}{m-1}$ for $m \ge 2$. Let us remark that we do not know the value of $\alpha(m, r)$ for any pair (m, r) with $m > r > 2$; e.g., the question whether $\alpha(4, 3) = 5/9$ is a well known open problem of extremal set theory.

The main problem we are concerned in this paper is the existence of families of r-sets which are 'sparse', or, at least, 'locally sparse', yet preserve some of the properties of $[n]^r$ stated in the theorems above. In the following section, we state a few specific problems on the existence of locally sparse structures with good 'global'

Proceedings of the International Congress
of Mathematicians, Madrid, Spain, 2006
© 2006 European Mathematical Society

properties. Then, we explain why the standard probabilistic method cannot be directly used to study extremal properties of graphs and hypergraphs. Next we recall another important result of modern combinatorics: Szemerédi's Regularity Lemma and show how it could help in dealing with such problems. We conclude with a few remarks on possible generalizations of known results and some speculation on developments which are still to come.

2. Locally sparse structures with good extremal properties

Let us introduce first some notation. An r-uniform hypergraph is a pair $H = (V, E)$, where V is the set of vertices of H and $E \subseteq [V]^r$ denotes the set of its edges. We say that a hypergraph $H' = (V', E')$ is a subhypergraph of a hypergraph $H'' = (V'', E'')$, if $V' \subseteq V''$ and $E' \subseteq E''$. The complete r-uniform hypergraph $([m], [m]^r)$ we denote by $K_m^{(r)}$, and set $K_m^{(2)} = K_m$. A 2-uniform hypergraph is called a graph.

Let $H = (V, E)$ be an r-uniform hypergraph, and let $C = \{W_1, \ldots, W_t\}$ be a family of s-subsets of V such that $[W_i]^r \subseteq E$, for $i = 1, \ldots, t$. We say that C is a loose (s, t)-circuit if $t \geq 3$, $(s, t) \neq (3, 3)$, $W_i \cap W_{i+1} \neq \emptyset$, for $i = 1, 2, \ldots, t-1$, and $W_1 \cap W_t \neq \emptyset$. We call C a tight (s, t)-circuit if either $t = 2$ and $|W_1 \cap W_2| \geq r+1$, or $t \geq 3$ and $|W_i \cap W_{i+1}| \geq r$, for $i = 1, 2, \ldots, t-1$, as well as $|W_1 \cap W_t| \geq r$.

Finally, for r-uniform hypergraphs $H = (V, E)$ and $H' = (V', E')$, let $\mathrm{ex}(H', H)$ be the number of edges in the largest subhypergraph of H which contains no copies of H', and $\overline{\mathrm{ex}}(H', H) = \mathrm{ex}(H', H)/|E|$. It is easy to see (e.g., [8], Prop.8.4.), that for a given H' the function $\overline{\mathrm{ex}}(H', H)$ is maximized for complete H, i.e., $\overline{\mathrm{ex}}(H', H) \leq \overline{\mathrm{ex}}(H', [V]^r)$.

One of the first results on the existence of locally sparse structures with good extremal properties was proved by Erdős [1] nearly fifty years ago. It states that there are graphs with large girth and no large independent sets (and so with large chromatic number).

Theorem 2.1. *For each ℓ and $\varepsilon > 0$ there exists a graph $G(\ell, \varepsilon) = (V, E)$ such that $G(\ell, \varepsilon)$ contains no $(2, t)$-circuits with $t \leq \ell$, but each subset $W \subseteq V$ such that $|W| \geq \varepsilon |V|$ contains an edge of $G(\ell, \varepsilon)$.*

In the following section we present Erdős' elegant non-constructive proof of this fact. Then we shall try to use a similar idea to get the following sparse version of Turan's theorem.

Conjecture 2.2. For any $r, s, \varepsilon > 0$, and ℓ, there exist an $n = n(r, m, \ell, \varepsilon)$ and an r-uniform hypergraph $G^{(r)}(s, \ell, \varepsilon) = (V, E)$ such that

(i) $G^{(r)}(s, \ell, \varepsilon)$ contains no tight (s, t)-cycles with $2 \leq t \leq \ell$;

(ii) each subhypergraph $H^{(r)} \subseteq G^{(r)}(s, \ell, \varepsilon)$ with at least $(\alpha(s, r) + \varepsilon)|E|$ edges contains a subset B, $|B| = s$, such that $[B]^r \subseteq H^{(r)}$, i.e.,

$$\overline{\mathrm{ex}}([n]^s, G^{(r)}(s, \ell, \varepsilon)) \leq \alpha(s, r) + \varepsilon\,.$$

In Sections 3-5 below we describe how to approach Conjecture 2.2 using a special version of the Regularity Lemma. Here we remark only that the existence of $G^{(r)}(s, \ell, \varepsilon)$ has been shown only for $r = 2$, $s = 3$ (Frankl and Rödl [2] and Haxell *et al.* [7]), $r = 2$, $s = 4$ (Kohayakawa *et al.* [10]), and recently for $r = 2$, $s = 5$ (Gerke *et al.* [4]).

We conclude this section with a conjecture on a sparse version of Szemerédi's density theorem. Here a (k, t)-arithmetic circuit is a family of t non-trivial arithmetic progressions $A_1, \ldots, A_t$ of length k such that $A_i \cap A_{i+1} \neq \emptyset$ for $i = 1, 2 \ldots, t-1$, and $A_1 \cap A_t \neq \emptyset$.

Conjecture 2.3. For any k, ℓ, and $\alpha > 0$, there exist an $\varepsilon = \varepsilon(\alpha, k, \ell) > 0$, $n = n(k, \alpha, \ell)$, and a set $A = A(k, \ell, \alpha, n) \subseteq [n]$ such that

(i) A contains no (k, t)-arithmetic circuits for $t \leq \ell$;

(ii) any non-trivial arithmetic progression of length n^ε in $[n]$ contains at most k elements of A;

(iii) each subset B of A with at least $\alpha|A|$ elements contains a non-trivial arithmetic progression of length k.

Kohayakawa *et al.* [9] showed the existence of $A = A(3, \ell, \alpha, n)$ for any $\alpha > 0$ and ℓ. Their proof was based on the idea used by Ruzsa and Szemerédi [16] to show that $r_3(n) = o(n)$. Let us also mention that since Szemerédi's density theorem can be deduced from some extremal results for hypergraphs (see Frankl and Rödl [3], Nagle *et al.* [13], and Rödl and Skokan [15]) it is in principle possible, although somewhat unlikely, that one can imitate the argument from [9] and verify Conjecture 2.3 for all $k \geq 4$ (cf., Conjecture 6.2 below).

3. Random structures

For $0 \leq p \leq 1$ and natural numbers n, r, let $\mathbb{G}^{(r)}(n, p)$ denote the random r-uniform hypergraph with vertex set $[n]$, where edges $\mathbb{G}^{(r)}(n, p)$ are chosen from $[n]^r$ independently with probability p. Thus, the number of edges of $\mathbb{G}^{(r)}(n, p)$ is a binomially distributed random variable with parameters $\binom{n}{r}$ and p. Typically, we are interested only in the asymptotic behavior of $\mathbb{G}^{(r)}(n, p)$ when $n \to \infty$ and the probability p may depend on n. In particular, we say that for a given function $p = p(n)$ the hypergraph $\mathbb{G}^{(r)}(n, p)$ has a property $\mathcal{A}$ a.a.s. if the probability that $\mathbb{G}^{(r)}(n, p)$ has $\mathcal{A}$ tends to 1 as $n \to \infty$. Since in this note we deal mainly with graphs, instead of $\mathbb{G}^{(2)}(n, p)$ we write briefly $\mathbb{G}(n, p)$.

Let us recall Erdős' proof of Theorem 2.1. Fix ℓ and $\varepsilon > 0$. Let n be very large and p be the probability which is neither too small (so $\mathbb{G}(n, p)$ contains no large independent sets) nor too large (so $\mathbb{G}(n, p)$ is locally sparse). A good choice for p is, say, $p = p(n) = n^{-1+1/2\ell}$, but, for n large enough, any $p = p(n)$ such that $10/\varepsilon \leq np \leq n^{1/\ell}/10$ would do.

Let $X = X(n,\ell)$ be the random variable which counts $(2,t)$-circuits with $t \le \ell$ in $\mathbb{G}(n,p)$. Then, for n large enough, we have

$$\mathbb{E}X \le \sum_{i=3}^{\ell} \binom{i}{2}\binom{i}{2}^{2i} p^i \le \ell^{2\ell+3}(np)^\ell \le n^{1/2}\,. \tag{2}$$

Thus, from Markov's inequality, $\Pr(X \ge n/2) \le 2n^{-1/2}$, and so, for large enough n, with probability at least $2/3 > 1 - 2n^{-1/2}$, we have $X \le n/2$. On the other hand, for the number $Y = Y(n,k)$ of independent sets of size k in $\mathbb{G}(n,p)$ we have

$$\Pr(Y > 0) \le \mathbb{E}X = \binom{n}{k}(1-p)^{\binom{k}{2}} \le 2^n \exp\left(-p\binom{k}{2}\right). \tag{3}$$

If $k = \varepsilon n/2$ then $\Pr(Y > 0)$ tends to 0 as $n \to \infty$, i.e., for n large enough we have $\Pr(Y > 0) < 2/3$. Now, let $G(\ell,\varepsilon)$ be a graph obtained from $\mathbb{G}(n,p)$ by removing one vertex from each $(2,t)$-circuit with $t \le \ell$. Then, with probability at least $1/3$, $G(\ell,\varepsilon)$ fulfills the assertion of Theorem 2.1.

The main goal of this paper is to discuss how one can verify Conjecture 2.2 using a modified version of Erdős' approach. We shall concentrate on the simplest non-trivial case of Conjecture 2.2, when $s = 3$ and $r = 2$. In order to deduce the existence of $G^{(2)}(3,\ell,\varepsilon)$ from appropriate properties of the random graph $\mathbb{G}(n,p)$ first we need to guess what value of p we are to use. More specifically, we should find the smallest possible value of $p_0 = p_0(n)$ such that a.a.s. in each subgraph of $\mathbb{G}(n,p_0)$ which contains, say, 51% of its edges one can find a triangle. Note that if a graph G contains m edges and t triangles, then there is a triangle-free subgraph of G with at least $m - t$ edges. Thus, it seems that in $\mathbb{G}(n,p_0)$ the expected number of triangles (equal to $\binom{n}{3}p_0^3$) must be at least of the order of the expected number of edges (equal $\binom{n}{2}p_0$), i.e., $p_0 = p_0(n)$ should be at least as large as $\Omega(n^{-1/2})$. It turns out that this necessary condition is also sufficient and the following holds (see Frankl and Rödl [2], Haxell *et al.* [7]).

Theorem 3.1. *For every $\delta > 0$ there exists $c = c(\delta)$ such that if $p = p(n) \ge cn^{-1/2}$, then a.a.s. each subgraph of $\mathbb{G}(n,p)$ with at least $(1/2 + \delta)\binom{n}{2}p$ edges contains a triangle.*

Let us try to prove Theorem 3.1 using Erdős' argument. To this end one has to bound the expected number of triangle-free subgraphs H of $\mathbb{G}(n,p)$, containing 51% of edges of $\mathbb{G}(n,p)$, using a formula similar to (3). In order to do that one needs to estimate the probability that such a large subgraph H of $\mathbb{G}(n,p)$ contains no triangles. The first problem which immediately emerges is the fact that our argument must depend strongly on the fact that H has more than 51% of edges of $\mathbb{G}(n,p)$, since in every graph G one can find a large bipartite subgraph which contains more than half of its edges. Thus, we have to use some property of H shared by all subgraphs of $\mathbb{G}(n,p)$ with more than half of its edges, and does not hold for, say, bipartite subgraphs of $\mathbb{G}(n, p)$; i.e., we should consider only graphs H which are 'essentially non-bipartite'. Then, we need to show that a.a.s. each subgraph of $\mathbb{G}(n, p)$ containing at least 51% of its edges is 'essentially non-bipartite', and estimate the probability that a 'random essentially non-bipartite' graph is triangle-free.

However, now we face another, more serious obstacle. The number of subsets of the set of edges of $\mathbb{G}(n, p)$ is much larger then the number of subsets of the set of vertices of $\mathbb{G}(n, p)$. Consequently, the factor 2^n in (3) should be replaced by $\exp(\Omega(n^2 p))$. Hence, we should estimate the probability that a 'random essentially non-bipartite' subgraph H of $\mathbb{G}(n, p)$ is triangle-free by a quantity which is much smaller than the probability that $\mathbb{G}(n, p)$ contains no edges at all! This is the crucial and most difficult part of the whole argument. It is also precisely the reason why we can show Conjecture 2.2 only for $r = 2$ and $s = 3, 4, 5$; for all other cases the proof breaks at this point.

Finally, it is easy to check that if $p = p(n) = n^{-1/2+o(1)}$, then for any given ℓ a.a.s. $\mathbb{G}(n, p)$ contains fewer than $o(n^2 p)$ tight $(3, t)$-circuits for $t \le \ell$ which can be removed from $\mathbb{G}(n, p)$ without affecting much its extremal properties. Unfortunately, one cannot deal in the same way with loose $(3, t)$-circuits. The reason is quite simple: for $t \ge 4$ the number of loose $(3, t)$-circuits grows much faster than the number of triangles, because, roughly speaking, two triangles of $\mathbb{G}(n, p)$ are much more likely to share a vertex than an edge. Clearly, the same is true if instead of $\mathbb{G}(n, p)$ we consider a shadow of $\mathbb{G}^{(3)}(n, p)$, i.e., we randomly generate triples of vertices and then replace each of them by a triangle. Still, it is not inconceivable that Conjecture 2.2 can be settled in the affirmative by a non-constructive method using more sophisticated models of random hypergraphs; there have been a fair amount of attempts in this direction but so far all of them have failed miserably.

4. Regularity Lemma

One of the main ingredients of Szemerédi's ingenious proof of the density theorem was the Regularity Lemma which for the last thirty years has become one of the most efficient tools of modern graph theory. In order to formulate it rigorously we need a few technical definitions.

For a graph $G = (V, E)$ and $W, W' \subseteq V$ let $e(W, W')$ denote the number of edges joining W and W'. A pair (A, B) of disjoint subsets of vertices of G is called an ε-regular pair, if for every subsets $A' \subseteq A$, $|A'| \ge \varepsilon|A|$, $B' \subseteq B$, $|B'| \ge \varepsilon|B|$,

$$\left| \frac{e(A', B')}{|A'||B'|} - \frac{e(A, B)}{|A||B|} \right| \le \varepsilon. \tag{4}$$

An ε-regular pair behaves in many respects as the bipartite random graph $\mathbb{G}(A, B, \rho)$, in which edges between A and B appear independently with probability $\rho = e(A, B)/|A||B|$. In particular, it is easy to check, that if a pair (A, B)

is ε-regular then the number of subgraphs of a given size in the bipartite subgraph $G[A, B]$ induced by $A \cup B$ in G is close to the expected number of such subgraphs in $\mathbb{G}(A, B, \rho)$. For instance, if (A, B) is ε-regular, then the number of cycles of length four in $G[A, B]$ is equal $(\rho^4/4 \pm h(\varepsilon))|A|^2|B|^2$, where $h(\varepsilon)$ is a function which tends to 0 as $\varepsilon \to 0$. The implication in the other direction holds as well: if the number of cycles of length four in $G[A, B]$ is smaller than $(\rho^4/4 + \varepsilon)|A|^2|B|^2$, then the pair (A, B) is $h'(\varepsilon)$-regular for some function $h'(\varepsilon)$ which tend to 0 as $\varepsilon \to 0$. Let us also mention that ε-regularity implies the correct number of small subgraphs even if we consider more than one ε-regular pair. For instance, if three disjoint sets $A_1, A_2, A_3 \subseteq V$ are such that each of the pairs (A_1, A_2), (A_2, A_3), (A_1, A_3) is ε-regular with density ρ, the number of triangles in the tripartite graph induced in G by these sets is $(\rho^3 \pm h''(\varepsilon))|A_1||A_2||A_3|$, where $h''(\varepsilon) \to 0$ as $\varepsilon \to 0$.

A partition $V = V_1 \cup \cdots \cup V_k$ of the vertex set of a graph $G = (V, E)$ is called a (k, ε)-partition if for all $i, j = 1, 2, \ldots, k$ we have $\big||V_i| - |V_j|\big| \leq 1$ and all except at most εk^2 pairs (V_i, V_j), $1 \leq i < j \leq k$, are ε-regular. Now Szemerédi's Regularity Lemma (see [17] and [18]) can be stated as follows.

Lemma 4.1 (Szemerédi's Regularity Lemma). *For every $\varepsilon > 0$ there exists K such that each graph G with more than $1/\varepsilon$ vertices admits a (k, ε)-regular partition for some k, $1/\varepsilon < k < K$.*

Note that if $k > 1/\varepsilon$, then for a (k, ε)-regular partition there are at most $n^2/k + \varepsilon n^2 \leq 2\varepsilon n^2$ edges of G which are either contained inside sets V_i or join pairs which are not ε-regular. Thus, Szemerédi's Regularity Lemma says that all but $2\varepsilon n^2$ edges of any graph G can be partitioned into at most k ε-regular pairs, for some $1/\varepsilon \leq k \leq K$, where K does not depend on the number of vertices in G. Unfortunately, $K = K(\varepsilon)$ grows very fast to infinity as $\varepsilon \to 0$ (see Gowers [5]), so most of the applications of the Regularity Lemma give very poor bounds of estimated quantities.

The Regularity Lemma can be reformulated and generalized in several ways. For instance, one can view it as a statement on the compactness of certain metric space (Lovász and Szegedy [12]); an information-theoretic approach to it can be found in Tao [19]. Another versions of the Regularity Lemma ensure the existence of 'weak' (k, ε)-partitions, or 'partial ε-covers' consisting of reasonably large ε-pairs. However the two most important developments in this area are, in my opinion, generalizations of the Regularity Lemma to sparse graphs and to hypergraphs. In the following sections we discuss how the Regularity Lemma can be modified to work efficiently for sparse graphs; here we say a few words on a much harder (both to state and to prove) version of the Regularity Lemma for hypergraphs. Several years ago Frankl and Rödl [3] generalized the Regularity Lemma to r-uniform hypergraphs and proved it, together with a supplementary 'counting lemma', for $r = 3$. The case $r \geq 4$ has been dealt with by Rödl and Skokan [15] and Nagle *et al.* [13], and, independently, by Gowers [6]. As was noticed by Frankl and Rödl [3], their version of the Regularity Lemma implies the following Removal Lemma which, in turn, can be used to show Szemerédi's density theorem (for details see Rödl *et al.* [14]).

Lemma 4.2 (Removal Lemma). *For $m \geq r \geq 2$ and every $\delta > 0$ there exist $\eta > 0$ and n_0 so that for every r-uniform hypergraph F on m vertices and r-uniform hypergraph H on n, $n \geq n_0$, vertices the following holds. If H contains at most ηn^m copies of F, then one can delete δn^r edges of H to destroy all copies of F.*

Thus, the Removal Lemma states that if the number of copies of F is large enough, then they must be, in some sense, uniformly distributed in H, i.e., the large number of copies of F makes a hypergraph H, or at least parts of it, close to a random graph. In fact all known proofs of Lemma 4.2 are based on this idea. We should apply the Regularity Lemma to H, and then show that, if the number of copies of F in H is large, then there exists a big random-like subgraph H' of H which contains an anticipated number of copies of F.

5. Regularity Lemma: sparse graphs

Note that Lemma 4.1 is basically meaningless for sparse graphs since the definition (4) of ε-regular pair (A, B) does not say much on the distribution of edges between A and B if the density $\rho = e(A, B)/|A||B|$ is smaller than ε. Thus, let us modify the definition of an ε-regular pair by 'scaling' the density of the pair by d which, typically, is the density of the graph $G = (V, E)$. Hence, we say that a pair (A, B) of disjoint subsets of vertices of a graph $G = (V, E)$ is (d, ε)-regular, if for each pair of subsets $A' \subseteq A$, $|A'| \geq \varepsilon|A|$, $B' \subseteq B$, $|B'| \geq \varepsilon|B|$, we have

$$\left| \frac{e(A', B')}{|A'||B'|} - \frac{e(A, B)}{|A||B|} \right| \leq d\varepsilon. \tag{5}$$

If $d_G = |E|/\binom{|V|}{2}$ we call a (d_G, ε)-regular pair strongly ε-regular. A strongly (k, ε)-regular partition of vertices of G is defined in a similar way as (k, ε)-regular partition. Moreover, we say that a graph $G = (V, E)$ with density $d_G = |E|/\binom{|V|}{2}$ is (η, b)-bounded if each subgraph H of G with $r \geq \eta|V|$ vertices contains not more than br^2 edges. As was observed independently by Kohayakawa and by Rödl (see Kohayakawa and Rödl [11] and references therein), one can mimic the proof of Szemerédi's Regularity Lemma to get the following result.

Lemma 5.1. *For every $\varepsilon > 0$ and b there exist η and K such that each (η, b)-bounded graph G with more than $1/\varepsilon$ vertices admits a strongly (k, ε)-regular partition for some k, $1/\varepsilon < k < K$.*

The assumption that G is (η, b)-bounded is typically not very restrictive. For instance, if $\eta > 0$ and $np \to \infty$ as $n \to \infty$, then the random graph $\mathbb{G}(n, p)$ is a.a.s. $(\eta, 2)$-bounded. Consequently, a.a.s. each subgraph of such $\mathbb{G}(n, p)$ which contains at least half of its edges is $(\eta, 4)$-bounded.

A more serious problem is that, unlike in the dense case, from the fact that a sparse pair is strongly ε-regular it does not follow that the number of cycles of length four in that pair is close to the number of cycles of length four in the random bipartite graph of the same density. In a similar way, for every $\varepsilon > 0$ there exists $\delta > 0$ and a tripartite graph G with vertex set $V_1 \cup V_2 \cup V_3$, $|V_1| = |V_2| = |V_3| = n$ such that all three pairs (V_1, V_2), (V_2, V_3), and (V_1, V_3) are strongly ε-regular pairs with densities larger than δ, yet G is triangle-free. Nevertheless, Kohayakawa, Łuczak, and Rödl conjectured in [10] that such triangle-free tripartite graphs consisting of dense ε-regular triples are so rare that a.a.s. the random graph $\mathbb{G}(n, p)$ contains none of them as a subgraph.

In order to state the conjecture rigorously we need one more definition. Let $\mathbb{G}(n, p; \varepsilon, s)$ be a graph chosen at random from the family of all s-partite graphs with vertex set $V_1 \cup V_2 \cup \cdots \cup V_s$, $|V_1| = |V_2| = \cdots = |V_s| = n$, such that for each i, j, $1 \le i < j \le s$, the pair (V_i, V_j) spans a bipartite strongly ε-regular graph with $\lceil pn^2 \rceil$ edges. Then the conjecture of Kohayakawa, Łuczak and Rödl for complete graphs goes as follows (for a more general statement see [10]).

Conjecture 5.2. For every s and $\delta > 0$ there exist $\varepsilon > 0$ and C such that if $n^s p^{\binom{s}{2}} > Cn^2 p$, then the probability that $\mathbb{G}(n, p; \varepsilon, s)$ contains no copies of K_s is smaller than $\delta^{n^2 p}$.

A stronger 'counting' version of Conjecture 5.2 goes as follows.

Conjecture 5.3. For every s and $\delta > 0$ there exist $\varepsilon > 0$ and C such that if $n^s p^{\binom{s}{2}} > Cn^2 p$, then the probability that $\mathbb{G}(n, p; \varepsilon, s)$ contains fewer than $n^s p^{\binom{s}{2}}/2$ copies of K_s is smaller than $\delta^{n^2 p}$.

So far Conjectures 5.2 and 5.3 have been shown only for $s = 3, 4, 5$ (see Gerke *et al.* [4] and the references therein).

Let us observe that Theorem 3.1 follows immediately from the fact that Conjecture 5.2 holds for $s = 3$. Indeed, let us fix $\delta > 0$ and let $p = C/\sqrt{n}$, where C is a large constant. Take a subgraph H of $\mathbb{G}(n, p)$ with at least $(1/2 + \delta)\binom{n}{2} p$ edges. Choose $\varepsilon > 0$ much smaller than δ and apply Lemma 5.1 to H to find in it a strong (k, ε)-partition with $1/\varepsilon < k < K$ (as we have already pointed out for every $\eta > 0$, a.a.s. $\mathbb{G}(n, p)$ is $(2, \eta)$-bounded and so H is $(4, \eta)$-bounded and fulfills assumptions of the lemma). Since H contains more than half of the edges of $\mathbb{G}(n, p)$, and edges in $\mathbb{G}(n, p)$ are uniformly distributed around the graph, there exist three sets V', V'', V''' of the partition such that each of the pairs (V', V''), (V'', V'''), (V', V'''), is strongly ε-regular and has density at least $\delta p/10$. (Let us remark that now a vague notion of an 'essentially non-bipartite' subgraph H we have used in Section 3 can be made precise: a graph H is essentially non-bipartite if it contains a balanced tripartite graph on $\Omega(n)$ vertices which consists of three dense strongly ε-regular pairs.) Now, one can use Conjecture 5.2 and argue as in (3) that a.a.s. each tripartite subgraph of $\mathbb{G}(n, p)$ of such a type contains a triangle. Thus, H contains a triangle and Theorem 3.1 follows.

Finally, let us also note that if, say, $p = \log n/\sqrt{n}$, then elementary calculations similar to that used by Erdős (cf. (3)) reveal that for every fixed ℓ a.a.s. the number of tight $(3, t)$-circuits in $\mathbb{G}(n, p)$ with $t \leq \ell$ is $o(n^2 p)$. Thus, one can obtain a graph $G^{(2)}(3, \ell, \delta)$ with all the properties specified in Conjecture 2.2 by deleting from $\mathbb{G}(n, p)$ all edges which belong to tight $(3, t)$-circuits, $t \leq \ell$.

6. Final remarks

It is easy to see that, arguing as in the proof of Theorem 3.1 above, one can show the existence of a graph $G^{(2)}(s, \ell, \delta)$ (see Conjecture 2.2) for every s for which Conjecture 5.2 holds. A precise formulation of analogs of Conjectures 5.2 and 5.3 for hypergraphs would become very technical, thus we only mention that if appropriately stated hypergraph version of Conjecture 5.3 is true then the following straightforward 'probabilistic' generalization of the Removal Lemma holds.

Conjecture 6.1. For $s > r \geq 2$ and every $\delta > 0$ there exist $\eta > 0$ such that a.a.s. in each subhypergraph H of the random r-uniform hypergraph $\mathbb{G}^{(r)}(n, p)$ which contains fewer than $\eta n^s p^{\binom{s}{r}}$ copies of $K_s^{(r)}$ one can destroy all these copies by removing fewer than $\delta n^r p$ hyperedges of H.

An analogous question on the validity of a probabilistic version of Szemerédi's density theorem can be stated as follows.

Conjecture 6.2. For every $\delta > 0$ and k there exists $\eta > 0$ such that a.a.s. in each subset A of $\mathbb{G}^{(1)}(n, p)$ with fewer than $\eta n^2 p^k$ non-trivial arithmetic progressions of length k all these progressions can be destroyed by removing fewer than δnp elements from A.

Conjecture 6.1 states that a.a.s. a random hypergraphs $\mathbb{G}^{(r)}(n, p)$ has a property $\mathcal{A}$ such that if a hypergraph G has $\mathcal{A}$ each subgraph H of G with fewer than $\eta n^s p^{\binom{s}{r}}$ copies of $K_s^{(r)}$ can be made $K_s^{(r)}$-free by deleting fewer than $\delta n^r p$ hyperedges. One can ask if $\mathcal{A}$ follows from some simple property $\mathcal{A}'$, i.e., whether there is a compact characterization of 'pseudorandom' sparse hypergraphs. A natural candidate for $\mathcal{A}'$ is the property that the number of some special subhypergraphs in G is close to the expected value of the number of such subhypergraphs in the random hypergraph with the same density. In the case of graphs a good choice for 'probing' graphs seem to be cycles of length four. It is known (see Thomasson [20]) that if the number of cycles of length four in a graph G is close to the anticipated one, then edges in G are 'uniformly distributed' around G. Nonetheless we do not know if the 'correct' number of cycles of length four, possibly matched with some additional requirements ensuring that G is locally sparse, can guarantee that G has good 'extremal' properties like those described in Conjecture 6.1. Another challenging problem is to strengthen the definition of a strongly ε-regular pair to, say, a 'super ε-regular pair' such that the analog of Lemma 5.1 remains valid in this setting (i.e., each dense subgraph of a random-like graph G admits a 'super (k, ε)-partition') and furthermore, each tripartite graph which consists of three dense super ε-regular pairs contains a triangle. Similar questions can be asked for hypergraphs, as well as for the subsets of $[n]$ (or, in somewhat more natural setting, for subsets of $\mathbb{Z}_n$, where n is a prime).

References

[1] Erdős, P. Graph theory and probability. *Canad. J. Math.* **11** (1959), 34–38.

[2] Frankl, P., and Rödl, V., Large triangle-free subgraphs in graphs without K_4. *Graphs Combin.* **2** (1986), 135–144.

[3] Frankl, P., and Rödl, V., Extremal problems on set systems. *Random Structures Algorithms* **20** (2002), 131–164.

[4] Gerke, S., Schickinger, T., Steger, A., K_5-free subgraphs of random graphs. *Random Structures Algorithms* **24** (2004), 194–232.

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

[6] Gowers, W. T., Quasirandomness, counting and regularity for 3-uniform hypergraphs. *Combin. Probab. Comput.* **15** (2006), 143–184.

[7] Haxell, P. E., Kohayakawa, Y., Łuczak, T., Turán's extremal problem graphs: forbidding odd cycles. *Combinatorica* **16** (1996), 107–122.

[8] Janson, S., Łuczak, T., Ruciński, A., *Random Graphs*. Wiley-Intersci. Ser. Discrete Math. Optim., Wiley, New York 2000.

[9] Kohayakawa, Y., Łuczak, T., and V. Rödl, Arithmetic progressions of length three in subsets of a random set. *Acta Arith.* **75** (1996), 133–163.

[10] Kohayakawa, Y., Łuczak, T., and Rödl, V., On K_4-free subgraphs of random graphs. *Combinatorica* **17** (1997), 173–213.

[11] Kohayakawa, Y., and Rödl, V., Regular pairs in sparse random graphs. *Random Structures Algorithms* **22** (2003), 359–434.

[12] Lovasz, L., Szegedy, B., Szemerédi's Lemma for the analyst. Preprint.

[13] Nagle, B., Rödl, V., Schacht, M., The counting lemma for regular k-uniform hypergraphs. *Random Structures Algorithms* **28** (2006), 113–179.

[14] Rödl, V., Nagle, B., Skokan, J., Schacht, M., Kohayakawa, Y., The hypergraph regularity method and its applications. *Proc. Nat. Acad. Sci. USA* **102** (2005) 8109–8113.

[15] Rödl, V., Skokan, J., Regularity lemma for k-uniform hypergraphs. *Random Structures Algorithms* **25** (2004), 1–42.

[16] Ruzsa, I., Szemerédi, E., Triple systems with no six points carrying three triangles. In *Combinatorics* (Proc. Fifth Hungarian Colloq. Keszthely, 1976), Vol. II, Colloq. Math. Soc. János Bolyai 18, North-Holland, Amsterdam, New York 1978, 939–945.

[17] Szemerédi, E., On sets of integers containing no k elements in arithmetic progression. *Acta Arith.* **28** (1975), 299–345.

[18] Szemerédi, E., Regular partitions of graphs. In *Problèmes Combinatoires et Théorie des Graphes* (ed. by J. Bermond et al.), Colloq. Internat. CNRS 260, CNRS, Paris 1978, 399–401.

[19] Tao, T., An information-theoretic proof of Szemerédi's regularity lemma. Preprint

[20] Thomason, A., Pseudorandom graphs. In *Random graphs '85* (Poznań, 1985), North-Holland Math. Stud. 144, North-Holland, Amsterdam 1987, 307–331.

[21] Turán, P., Egy gráfelméleti szélsöérték feladotról. *Mat. Fiz. Lapok* **48** (1941), 436–452; see also On the theory of graphs. *Colloq. Math.* **3** (1954), 19–30.

Adam Mickiewicz University, Faculty of Mathematics and CS, 61-614 Poznań, Poland
E-mail: tomasz@amu.edu.pl