COMBINATORICA
Bolyai Society – Springer-Verlag

COMBINATORICA **25** (1) (2005) 77–84

AN EXTENSION OF THE RUZSA–SZEMERÉDI THEOREM

GÁBOR N. SÁRKÖZY, STANLEY SELKOW

Received February 20, 2002
Revised February 18, 2004

We let $G^{(r)}(n,m)$ denote the set of r-uniform hypergraphs with n vertices and m edges, and $f^{(r)}(n,p,s)$ is the smallest m such that every member of $G^{(r)}(n,m)$ contains a member of $G^{(r)}(p,s)$. In this paper we are interested in fixed values r,p and s for which $f^{(r)}(n,p,s)$ grows quadratically with n. A probabilistic construction of Brown, Erdős and T. Sós ([2]) implies that $f^{(r)}(n,s(r-2)+2,s)=\Omega(n^2)$. In the other direction the most interesting question they could not settle was whether $f^{(3)}(n,6,3)=o(n^2)$. This was proved by Ruzsa and Szemerédi [11]. Then Erdős, Frankl and Rödl [6] extended this result to any r: $f^{(r)}(n,3(r-2)+3,3)=o(n^2)$, and they conjectured ([4], [6]) that the Brown, Erdős and T. Sós bound is best possible in the sense that $f^{(r)}(n,s(r-2)+3,s)=o(n^2)$.

In this paper by giving an extension of the Erdős, Frankl, Rödl Theorem (and thus the Ruzsa–Szemerédi Theorem) we show that indeed the Brown, Erdős, T. Sós Theorem is not far from being best possible. Our main result is

$$f^{(r)}(n, s(r-2) + 2 + \lfloor \log_2 s \rfloor, s) = o(n^2).$$

1. Introduction

1.1. Notation and definitions

For basic graph concepts see the monograph of Bollobás [1]. We let $V(G)$ and $E(G)$ denote the vertex-set and the edge-set of the graph G, and (A,B) or (A,B,E) denote a bipartite graph $G=(V,E)$, where $V=A\cup B$, and $E\subset A\times B$. In general, given any graph G and two disjoint subsets A,B of $V(G)$, the pair (A,B) is the graph restricted to $A\times B$. $N(v)$ is the set of

Mathematics Subject Classification (2000): 05C65, 05C70

neighbors of $v \in V$. Hence the size of $N(v)$ is $|N(v)| = deg(v) = deg_G(v)$, the degree of v. For a vertex $v \in V$ and set $U \subset V - \{v\}$, we write $deg(v, U)$ for the number of edges from v to U. We denote by $e(A,B)$ the number of edges of G with one endpoint in A and the other in B. For non-empty A and B,

$$d(A,B) = \frac{e(A,B)}{|A||B|}$$

is the *density* of the graph between A and B.

Definition 1. The pair (A,B) is *ε-regular* if

$$X \subset A,\ Y \subset B,\ |X| > \varepsilon|A|,\ |Y| > \varepsilon|B|$$

imply

$$|d(X,Y) - d(A,B)| < \varepsilon,$$

otherwise it is ε-irregular.

A hypergraph $\mathcal{F}$ is called $k-uniform$ if $|F| = k$ for every edge $F \in \mathcal{F}$. A k-uniform hypergraph $\mathcal{F}$ on the set X is k*−partite* if there exists a partition $X = X_1 \cup \ldots \cup X_k$ with $|F \cap X_i| = 1$ for every edge $F \in \mathcal{F}$ and $1 \leq i \leq k$.

1.2. Turán-type hypergraph problems

We let $G^{(r)}(n,m)$ denote the set of r-uniform hypergraphs with n vertices and m edges, and let $f^{(r)}(n,p,s)$ be the smallest m such that every member of $G^{(r)}(n,m)$ contains a member of $G^{(r)}(p,s)$. The determination of $f^{(r)}(n,p,s)$ has been a longstanding open problem. Special cases of this problem appeared in [3] and [5]. For more about Turán-type hypergraph results consult the surveys by Füredi [8] and Sidorenko [12]. In this paper we are interested in fixed values r, p and s for which $f^{(r)}(n,p,s)$ grows quadratically with n.

A probabilistic construction of Brown, Erdős and T. Sós [2] implies that

$$f^{(r)}(n, s(r-2)+2, s) = \Omega(n^2).$$

In the other direction the most interesting question they could not settle was whether $f^{(3)}(n,6,3) = o(n^2)$. This was proved in the celebrated paper by Ruzsa and Szemerédi [11]. Then Erdős, Frankl and Rödl [6] extended this result to any r:

$$f^{(r)}(n, 3(r-2)+3, 3) = o(n^2),$$

and they conjectured ([4], [6]) that the Brown, Erdős and T. Sós bound is best possible in the following sense:

Conjecture 1.

$$f^{(r)}(n, s(r-2)+3, s) = o(n^2).$$

In [6] they also showed that for any $c<2$,

$$\lim_{n\to\infty} f^{(r)}(n, 3(r-2)+3, 3)/n^c = \infty.$$

In this paper by giving an extension of the upper bound of the Erdős, Frankl, Rödl Theorem (and thus the Ruzsa–Szemerédi Theorem) we show that indeed the Brown, Erdős, T. Sós Theorem is not far from being best possible.

Our main result is the following.

Theorem 1. *For all integers $r,s\geq 3$ we have*

$$f^{(r)}(n, s(r-2)+2+\lfloor \log_2 s\rfloor, s) = o(n^2).$$

In particular for $s=3$ we get the Erdős, Frankl, Rödl Theorem (and thus the Ruzsa–Szemerédi Theorem) as a special case.

Thus roughly speaking the Brown, Erdős, T. Sós Theorem is best possible apart from a $\lfloor \log_2 s\rfloor$ term. However, it still remains open whether one can eliminate this term and prove Conjecture 1; for instance, it is still left open whether it is true that $f^{(3)}(n,7,4)=o(n^2)$. What we do get from our result is that $f^{(3)}(n,8,4)=o(n^2)$, $f^{(3)}(n,9,5)=o(n^2)$, etc.

The Erdős, Frankl, Rödl approach was based on the following result: If $n\geq n_0(\varepsilon,H)$ and G is an H-free graph on n vertices, then one can remove fewer than εn^2 edges from G so that the remaining graph is K_r-free, where $r=\chi(H)$. This result in turn was proved by using Szemerédi's Regularity Lemma [13].

Another simple proof of $f^{(3)}(n,6,3)=o(n^2)$ was found later by Szemerédi (see [10]). In this paper we generalize this argument in order to prove our main theorem. Szemerédi used the following result: suppose the edges of an n-vertex graph G are properly colored by at most cn colors, and $|E(G)|\geq c'n^2$, where c and c' are positive constants. If $n\geq n_0(c,c')$, then there is a path in G on 3 edges that gets only 2 colors. In this paper we generalize this argument; we obtain certain trees that get few colors.

In the next section we provide the tools. Then in Section 3 we prove the theorem by reducing the general case to that of $r=3$.

2. Tools

We will use a simple but useful result of Erdős and Kleitman ([7], see also on page 1300 in [9]).

Lemma 1. *Every k-uniform hypergraph $\mathcal{F}$ contains a k-partite k-uniform hypergraph $\mathcal{H}$ with*

$$\frac{|\mathcal{H}|}{|\mathcal{F}|} \geq \frac{k!}{k^k}.$$

We will also use the following lemma.

Lemma 2. *For every $c_1>0$, $c_2\geq 1$ there are positive constants η, n_0 with the following properties. Let G be a graph on $n\geq n_0$ vertices with $|E(G)|\geq c_1 n^2$ that is the edge disjoint union of matchings $M_1, M_2, \ldots, M_m$ where $m\leq c_2 n$. Then there exist an $1\leq i\leq m$ and $A, B\subset V(M_i)$ such that*

- $(A\times B)\cap M_i=\emptyset$,
- $|A|=|B|\geq \eta n$,
- $|E(G|_{A\times B})|\geq \frac{c_1}{4}|A||B|$.

Proof. This lemma can be proved by a fairly standard argument using the Regularity Lemma, we leave it to the reader. A very similar argument can be found in [10]. ∎

3. Proof of Theorem 1

Let $r, s\geq 3$, $p=s(r-2)+2+\lfloor \log_2 s\rfloor$ and $l=\lfloor \log_2 s\rfloor+1$.

For any constant $c>0$ and sufficiently large n, let $\mathcal{F}\in G^{(r)}(n, \lceil cn^2\rceil)$. That is, $\mathcal{F}$ is an r-uniform hypergraph with at least $\lceil cn^2\rceil$ edges. We will show that $\mathcal{F}$ must contain a member of $G^{(r)}(p,s)$, i.e., a set of p vertices spanning at least s edges.

Using the Erdős–Kleitman theorem (Lemma 1) we find an r-partite sub-hypergraph $\mathcal{H}$ of $\mathcal{F}$ with at least

$$\frac{r!c}{r^r}n^2$$

edges. Let $X_1, \ldots, X_r$ be the vertex classes of this r-partite hypergraph $\mathcal{H}$. Consider the 3-uniform hypergraph $\mathcal{H}^*$ which is defined by the removal of $X_1, \ldots, X_{r-3}$ from the vertex set of $\mathcal{H}$ and from all edges of $\mathcal{H}$. If a 3-edge (triple) of $\mathcal{H}^*$ has multiplicity greater than 1, then we keep only one edge. Note that every triple has multiplicity less than s. Indeed, otherwise taking a triple with multiplicity at least s and s r-edges of $\mathcal{H}$ containing this triple, we get a set of at most

$$s(r-3)+3 < s(r-2)+2+\lfloor \log_2 s\rfloor = p$$

vertices that span at least s r-edges, implying that $\mathcal{F}$ contains a member of $G^{(r)}(p,s)$. Then keeping only one edge from each multiple triple in $\mathcal{H}^*$ we still have at least

$$\frac{r!c}{r^r s}n^2$$

edges.

Consider first an arbitrary $v \in X_{r-2}$ and the bipartite graph G_{bp}^v defined by v between X_{r-1} and X_r such that (u,w) is an edge in G_{bp}^v if and only if (v,u,w) is a triple in $\mathcal{H}^*$. The maximum degree in G_{bp}^v is less than s. Indeed, otherwise taking s edges from a vertex u, the vertex v and the s r-edges of $\mathcal{H}$ containing these triples, we get again a set of at most

$$s(r-2)+2 < s(r-2)+2+\lfloor \log_2 s \rfloor = p$$

vertices that span at least s r-edges, implying that $\mathcal{F}$ contains a member of $G^{(r)}(p,s)$. Then we can choose a matching M_v in G_{bp}^v such that

$$|M_v| \geq \frac{|E(G_{bp}^v)|}{s}.$$

We take the next $v' \in X_{r-2}$ and similarly as above we define $G_{bp}^{v'}$ and $M_{v'}$, but now from $M_{v'}$ we remove all the edges that are already in M_v. We continue in this fashion for all the vertices in X_{r-2}. Define the bipartite graph $G_{bp} = \bigcup_{v \in X_{r-2}} M_v$. Since every edge of G_{bp} is an edge in fewer than s of the graphs G_{bp}^v, we have

$$|E(G_{bp})| \geq \frac{r!c}{r^r s^3}n^2.$$

Next by applying Lemma 2 iteratively in G_{bp}, we will find a sequence of matchings $M_{v_1},\dots,M_{v_l}$. From these l matchings, we will construct a tree of 2^l-1 edges in the bipartite graph (X_{r-1},X_r). Leaves will be removed from this tree until it has s edges. The $s+1$ vertices of this tree, along with the corresponding $l=\lfloor \log_2 s \rfloor+1$ vertices of X_{r-2}, are then extended to a member of $G^{(r)}(p,s)$. These l vertices of X_{r-2} account for the gap between Conjecture 1 and Theorem 1. Thus in order to reduce this gap one has to construct trees (or other graphs) of similar size that are built from even fewer matchings.

To obtain M_{v_1} we apply Lemma 2 in G_{bp}. We can choose

$$c_1 = c_1^1 = \frac{r!c}{r^r s^2} \quad \text{and} \quad c_2 = c_2^1 = 1.$$

M_{v_1} is the M_i guaranteed in the lemma. Denote $M_{v_1} = (A_1, B_1)$ where $A_1 \subset X_{r-1}, B_1 \subset X_r$. Lemma 2 also guarantees that there are $A_1', B_1' \subset V(M_{v_1})$ such that

- $(A'_1 \times B'_1) \cap M_{v_1} = \emptyset$,
- $|A'_1| = |B'_1| \geq \eta_1 n$,
- $\left|E\left(G_{bp}|_{A'_1 \times B'_1}\right)\right| \geq \frac{c_1}{4}|A'_1||B'_1|$.

To obtain M_{v_2} we apply Lemma 2 again, now for $G_{bp}|_{A'_1 \times B'_1}$. Here we can choose

$$c_1 = c_1^2 = \frac{c_1^1}{16} \quad \text{and} \quad c_2 = c_2^2 = \frac{c_2^1}{2\eta_1}.$$

M_{v_2} is the M_i guaranteed in the lemma. Note that technically this M_{v_2} is not the whole M_{v_2} in G_{bp}, but it is restricted to $G_{bp}|_{A'_1 \times B'_1}$. Denote $M_{v_2} = (A_2, B_2)$ where $A_2 \subset X_{r-1}, B_2 \subset X_r$.

We continue in this fashion, satisfying

$$A_1 \supset A'_1 \supset A_2 \supset A'_2 \supset \cdots$$

and

$$B_1 \supset B'_1 \supset B_2 \supset B'_2 \supset \cdots.$$

Assume that $M_{v_j} = (A_j, B_j)$ is already defined where $A_j \subset X_{r-1}, B_j \subset X_r$. Futhermore, we have $A'_j, B'_j \subset V(M_{v_j})$ such that

- $(A'_j \times B'_j) \cap M_{v_j} = \emptyset$,
- $|A'_j| = |B'_j| \geq \eta_j(|A'_{j-1}| + |B'_{j-1}|)$,
- $\left|E\left(G_{bp}|_{A'_j \times B'_j}\right)\right| \geq \frac{c_1^j}{4}|A'_j||B'_j|$.

To obtain $M_{v_{j+1}}$ we apply Lemma 2 for $G_{bp}|_{A'_j \times B'_j}$. We can choose

$$c_1 = c_1^{j+1} = \frac{c_1^j}{16} \quad \text{and} \quad c_2 = c_2^{j+1} = \frac{c_2^j}{2\eta_j}.$$

$M_{v_{j+1}}$ is the M_i guaranteed in the lemma. Denote $M_{v_{j+1}} = (A_{j+1}, B_{j+1})$. We continue until $M_{v_1}, \ldots, M_{v_l}$ are selected.

Next using these matchings M_{v_j} we will select a set of p vertices spanning at least s r-edges of $\mathcal{H}$, implying that $\mathcal{F}$ contains a member of $G^{(r)}(p, s)$.

Lemma 3. *For any $1 \leq i \leq l = \lfloor \log_2 s \rfloor + 1$, let G_i be the graph obtained from bipartite graph $\left(X_{r-1}, X_r, \bigcup_{j=1}^{i} M_{v_j}\right)$ by removing all components which do not contain a vertex of $A_i \cup B_i$. The vertices of G_i are partitioned into $|M_{v_i}|$ trees, each with $2^i - 1$ edges.*

Proof. We use induction on i. For $i=1$, G_1 is just M_{v_1}, and each tree of G_1 has one edge. We assume the lemma to hold for $i-1$. Each endpoint of each edge $e \in M_{v_i}$ is in $A_{i-1} \cup B_{i-1}$ and thus by the inductive hypothesis belongs to exactly one tree of G_{i-1}, and each of these trees has $2^{i-1}-1$ edges. Edge e, along with the two trees it joins, comprise a new tree with 2^i-1 edges. ∎

Lemma 4. *There exist $l+s+1=\lfloor \log_2 s\rfloor+s+2$ vertices in $\mathcal{H}^*$ which span at least s 3-edges.*

Proof. In case $s=2^l-1$, then the $l=\lfloor \log_2 s\rfloor+1$ vertices $\{v_1,\dots,v_l\}$ and the $s+1$ vertices of a tree τ in G_l span at least s 3-edges of $\mathcal{H}^*$. Otherwise, we just remove leaves of τ until a total of s edges (and $s+1$ vertices) are left. Then again the $l=\lfloor \log_2 s\rfloor+1$ vertices $\{v_1,\dots,v_l\}$ and the $s+1$ vertices left in τ span at least s 3-edges of $\mathcal{H}^*$. ∎

For each of the s 3-edges in $\mathcal{H}^*$ assured by Lemma 4, we add the $r-3$ other vertices of an edge in the original hypergraph $\mathcal{H}$ which contains it. So the $s(r-2)+2+\lfloor \log_2 s\rfloor=p$ vertices span at least s edges, implying that $\mathcal{F}$ contains a member of $G^{(r)}(p,s)$.

This completes the proof of Theorem 1. ∎

References

[1] B. Bollobás: *Extremal Graph Theory*, Academic Press, London, 1978.

[2] W. G. Brown, P. Erdős and V. T. Sós: Some extremal problems on r-graphs, in *New directions in the theory of graphs, Proc. 3rd Ann Arbor Conference on Graph Theory*, pp. 55–63, Academic Press, New York, 1973.

[3] W. G. Brown, P. Erdős and V. T. Sós: On the existence of triangulated spheres in 3-graphs and related problems, *Periodica Mathematica Hungarica* **3** (1973), 221–228.

[4] P. Erdős: Problems and results on graphs and hypergraphs: similarities and differences, in *Mathematics of Ramsey Theory* (J. Nešetřil, V. Rödl, eds.), pp. 12–28, Springer Verlag, Berlin, 1990.

[5] P. Erdős: Extremal problems in graph theory, in *Theory of graphs and its applications* (M. Fiedler, ed.), pp. 29–36, Academic Press, New York, 1964.

[6] 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 and Combinatorics* **2** (1986), 113–121.

[7] P. Erdős and D. J. Kleitman: On coloring graphs to maximize the proportion of multicolored k-edges, *J. of Combinatorial Theory* **5** (1968), 164–169.

[8] Z. Füredi: Turán-type problems, *Surveys in Combinatorics*, London Math. Soc. Lecture Notes Ser., A. D. Keedwell, ed., Cambridge Univ. Press (1991), 253–300.

[9] R. L. Graham, M. Grötschel and L. Lovász: *Handbook of Combinatorics*, Elsevier Science B.V., 1995.

[10] J. Komlós and M. Simonovits: Szemerédi's Regularity Lemma and its applications in graph theory, in *Combinatorics, Paul Erdős is Eighty* (D. Miklós, V. T. Sós and T. Szőnyi, eds.), pp. 295–352, Bolyai Society Mathematical Studies, Vol. **2**, Budapest, 1996.
[11] I. Z. Ruzsa and E. Szemerédi: Triple systems with no six points carrying three triangles, in *Combinatorics*, Keszthely, 1976, *Coll. Math. Soc. J. Bolyai* **18** Volume **II.**, pp. 939–945.
[12] A. F. Sidorenko: What we do know and what we do not know about Turán numbers, *Graphs and Combinatorics* **11** (1995), 179–199.
[13] E. Szemerédi: Regular partitions of graphs, *Colloques Internationaux C.N.R.S.* № **260** – Problèmes Combinatoires et Théorie des Graphes, Orsay (1976), 399–401.

Gábor N. Sárközy

Computer Science Department
Worcester Polytechnic Institute
Worcester, MA 01609
USA
gsarkozy@cs.wpi.edu

Stanley Selkow

Computer Science Department
Worcester Polytechnic Institute
Worcester, MA 01609
USA
sms@cs.wpi.edu