2-10-2003

An Extension of the Ruzsa-Szemerédi Theorem

Gábor N. Sárközy
Worcester Polytechnic Institute, gsarkozy@cs.wpi.edu

Stanley Selkow
Worcester Polytechnic Institute, sms@cs.wpi.edu

An extension of the Ruzsa-Szemerédi Theorem

Gábor N. Sárközy, Stanley Selkow
Computer Science Department
Worcester Polytechnic Institute
Worcester, MA 01609
gsarkozy@cs.wpi.edu, sms@cs.wpi.edu

February 10, 2003

Abstract

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 $f^{(r)}(n, s(r-2)+2, s) = \Omega(n^2)$. Erdős, Frankl and Rödl ([4], [6]) conjectured that this is best possible in the sense that $f^{(r)}(n, s(r-2)+3, s) = o(n^2)$. This was first proved for $r = s = 3$ by Ruzsa and Szemerédi [11]. Then Erdős, Frankl and Rödl [6] extended this result to any r: $f^{(r)}(n, 3(r-2)+3, 3) = o(n^2)$. In this paper by giving an extension of this 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

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

1 Introduction

1.1 Notation and definitions

For basic graph concepts see the monograph of Bollobás [1]. $V(G)$ and $E(G)$ denote the vertex-set and the edge-set of the graph G. (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 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 $1 \leq i \leq k$. In this paper $\log n$ denotes the base 2 logarithm.

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 $f^{(r)}(n,p,s)$ is the smallest m such that every member of $G^{(r)}(n,m)$ contains a member of $G^{(r)}(p,s)$. The determination of $f^{(r)}(n,p,s)$ has been a longstanding open problem. Special cases of this problem appeared in [3], [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

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

Erdős, Frankl and Rödl ([4], [6]) conjectured that this is best possible in the following sense:

Conjecture 1.

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

This was first proved for $r = s = 3$ by Ruzsa and Szemerédi [11]. Erdős, Frankl and Rödl [6] extended this result to any r:

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

In this paper by giving an extension of this 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*

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

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 $\lfloor \log s \rfloor$ term. However, it still remains open whether one can eliminate this term and prove Conjecture 1.

In the next section we provide the tools including the Regularity Lemma. Then in Section 3 we apply the Regularity Lemma to obtain our main lemma. Finally in Section 4 we prove the theorem.

2 Tools

In the proof the Regularity Lemma of Szemerédi ([13]) plays a central role. Here we will use the following variation of the lemma.

Lemma 1 (Regularity Lemma – Degree form). *For every $\varepsilon > 0$ there is an $M = M(\varepsilon)$ such that if $G = (V, E)$ is any graph and $d \in [0, 1]$ is any real number, then there is a partition of the vertex-set V into $l + 1$ sets (so-called clusters) $C_0, C_1, ..., C_l$, and there is a subgraph $G' = (V, E')$ with the following properties:*

- $l \leq M$,
- $|C_0| \leq \varepsilon |V|$,
- *all clusters C_i, $i \geq 1$, are of the same size,*
- $deg_{G'}(v) > deg_G(v) - (d + \varepsilon)|V|$ *for all* $v \in V$,
- $G'|_{C_i} = \emptyset$ *(C_i are independent in G'),*
- *all pairs $G'|_{C_i \times C_j}$, $1 \leq i < j \leq l$, are ε-regular, each with a density 0 or exceeding d.*

This form (see [10]) can easily be obtained by applying the original Regularity Lemma (with a smaller value of ε), adding to the exceptional set C_0 all clusters incident to many irregular pairs, and then deleting all edges between any other clusters where the edges either do not form a regular pair or they do but with a density at most d.

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

Lemma 2. *Every k-uniform hypergraph $\mathcal{F}$ contains a k-partite k-uniform hypergraph $\mathcal{H}$ with*

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

3 Applying the Regularity Lemma

We will prove the following lemma by applying the Regularity Lemma.

Lemma 3. *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: Let us apply the degree form of the Regularity Lemma (Lemma 1) with

$$d = \frac{c_1}{2} \quad \text{and} \quad \varepsilon = \frac{c_1}{6c_2}. \tag{1}$$

Let $G'' = G' \setminus C_0$. Then we have

$$deg_{G''}(v) > deg_G(v) - (d+\varepsilon)n - |C_0| \geq deg_G(v) - (d+2\varepsilon)n \quad \text{for all} \quad v \in V(G'').$$

Thus using (1)

$$|E(G'')| = \frac{1}{2}\sum_{v\in V(G'')} deg_{G''}(v) > \frac{1}{2}\sum_{v\in V(G'')} deg_G(v) - \frac{d+2\varepsilon}{2}n^2 =$$

$$= \frac{1}{2}\sum_{v\in V(G)} deg_G(v) - \frac{1}{2}\sum_{v\in C_0} deg_G(v) - \frac{d+2\varepsilon}{2}n^2 \geq |E(G)| - \frac{d+3\varepsilon}{2}n^2 \geq \frac{c_1}{2}n^2.$$

Hence there is an $1 \leq i \leq m$ such that

$$|M_i|_{G''}| > \frac{c_1}{2c_2}n = 3\varepsilon n. \tag{2}$$

Write $U = V(M_i|_{G''})$ for the vertex set of $M_i|_{G''}$. (2) implies that $|U| > 6\varepsilon n$. Write also $U_i = U \cap C_i$. Define $I = \{i \mid |U_i| > 3\varepsilon|C_i|\}$, and set $U' = \cup_{i\in I} U_i$ and $U'' = U \setminus U'$. Clearly $|U''| \leq 3\varepsilon n$. Since $|U| > 6\varepsilon n$, we have two vertices $u, v \in U'$ adjacent in $M_i|_{G''}$. Let $u \in C_i$ and $v \in C_j$. In G'' we have at least one edge between C_i and C_j, and hence we must have a density more than $d = \frac{c_1}{2}$ between them. Consider U_i and U_j. A is an arbitrary subset of U_i with $|A| = \lfloor \varepsilon|C_i| \rfloor + 1 > \varepsilon|C_i|$. B is an arbitrary subset of U_j with $|B| = \lfloor \varepsilon|C_j| \rfloor + 1 > \varepsilon|C_j|$ and $(A \times B) \cap M_i = \emptyset$. This is possible since

$$|U_j| > 3\varepsilon|C_j| > 2\lfloor \varepsilon|C_j| \rfloor + 2,$$

if $n \geq n_0$. Then the first property of A, B in the lemma is clearly satisfied. For the second property we can choose $\eta = \frac{\varepsilon(1-\varepsilon)}{M(\varepsilon)}$. Finally for the third property, ε-regularity of the pair (C_i, C_j) implies that the density between A and B is more than $d - \varepsilon \geq \frac{c_1}{4}$. This means

$$|E(G|_{A\times B})| \geq \frac{c_1}{4}|A||B|,$$

and thus completing the proof of the lemma. □

4 Proof of Theorem 1

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

Assume indirectly that there is a constant $c > 0$ such that

$$f^{(r)}(n,p,s) > \lceil cn^2 \rceil. \tag{3}$$

From this assumption we will get a contradiction. (3) means that there exists an r-uniform hypergraph $\mathcal{F}$ with

$$f^{(r)}(n,p,s) - 1 \geq \lceil cn^2 \rceil \geq cn^2$$

edges that does not contain a member of $G^{(r)}(p,s)$, i.e. a set of p vertices spanning at least s edges. Let us assume that n is sufficiently large.

Using the Erdős-Kleitman theorem (Lemma 2) 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

$$s(r-3) + 3 \leq s(r-2) + 2 + \lfloor \log s \rfloor = p$$

vertices that span at least s r-edges, a contradiction. Then if in $\mathcal{H}^*$ we keep only one edge from each multiple triple 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_b^v defined by v between X_{r-1} and X_r such that (u,w) is an edge in G_b^v if and only if (u,v,w) is a triple in $\mathcal{H}^*$. The maximum degree in G_b^v is at most 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

$$s(r-2) + 2 \leq s(r-2) + 2 + \lfloor \log s \rfloor = p$$

vertices that span at least s r-edges, a contradiction. Then we can choose a matching M_v in G_b^v such that

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

We take the next $v' \in X_{r-2}$, and similarly as above we define $G_b^{v'}$ 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 $G_b = \cup_{v \in X_{r-2}} M_v$. Since every edge of G_b is an edge in at most s of the graphs G_b^v, we have

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

Next by applying Lemma 3 iteratively in G_b, we will find a sequence of matchings $M_{v_1}, \ldots, M_{v_l}$. To obtain M_{v_1} we apply Lemma 3 in G_b. We can choose

$$c_1 = c_1^1 = \frac{r!c}{r^r s^3} \quad \text{and} \quad c_2 = c_2^1 = 1.$$

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 3 also guarantees that there are $A_1', B_1' \subset V(M_{v_1})$ such that

- $(A_1' \times B_1') \cap M_{v_1} = \emptyset$,
- $|A_1'| = |B_1'| \geq \eta_1 n$,
- $\left|E\left(G_b|_{A_1' \times B_1'}\right)\right| \geq \frac{c_1}{4}|A_1'||B_1'|$.

To obtain M_{v_2} we apply Lemma 3 again, now for $G_b|_{A_1' \times B_1'}$. 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_b, but it is restricted to $G_b|_{A_1' \times B_1'}$. 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. 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

- $(A_j' \times B_j') \cap M_{v_j} = \emptyset$,
- $|A_j'| = |B_j'| \geq \eta_j(|A_{j-1}'| + |B_{j-1}'|)$,
- $\left|E\left(G_b|_{A_j' \times B_j'}\right)\right| \geq \frac{c_1^j}{4}|A_j'||B_j'|$.

To obtain $M_{v_{j+1}}$ we apply Lemma 3 for $G_b|_{A_j' \times B_j'}$. 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 $\mathcal{H}$, a contradiction.

Lemma 4. *For any $1 \le i \le l = \lceil \log s \rceil$, let G_i be the graph obtained from bipartite graph $\left(X_{r-1}, X_r, \cup_{j=1}^{i} M_{v_j}\right)$ 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 5. *There exist $\lfloor \log s \rfloor + s + 2$ vertices in $\mathcal{H}^*$ which span at least s 3-edges.*

Proof: In case there exists an integer k such that $s = 2^k - 1$, then the $\lfloor \log s \rfloor + 1$ vertices $\left\{v_1, \ldots, v_{\lceil \log s \rceil}\right\}$ and the $s+1$ vertices of any tree in $G_{\lceil \log s \rceil}$ span at least s 3-edges of $\mathcal{H}^*$. Otherwise, we select any two trees τ_1 and τ_2 of $G_{\lfloor \log s \rfloor}$ assured by Lemma 4. We remove leaves of τ_1 or τ_2 until a total of s edges (and $s + 2$ vertices) are left. Then the $\lfloor \log s \rfloor$ vertices $\left\{v_1, \ldots, v_{\lfloor \log s \rfloor}\right\}$ and the $s + 2$ vertices of τ_1 and τ_2 span at least s 3-edges of $\mathcal{H}^*$. □

For each of the s 3-edges in $\mathcal{H}^*$ assured by Lemma 5, we add the $r - 3$ other vertices of an edge in the original hypergraph $\mathcal{H}$ which contains it. So the $s(r-2) + 2 + \lfloor \log s \rfloor = p$ vertices span at least s edges, a contradiction.

This completes the proof of Theorem 1. □

References

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

[2] W.G. Brown, P. Erdős, 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*, Academic Press, New York, 1973, 55-63.

[3] W.G. Brown, P. Erdős, 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* (J. Nesetril, V. Rödl, eds.), Springer Verlag, Berlin, 1990, 12-28.

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

[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, 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, 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, E. Szemerédi, Triple systems with no six points carrying three triangles, in *Combinatorics (Keszthely, 1976), Coll. Math. Soc. J. Bolyai 18, Volume II.* 939-945.

[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. N$^{\underline{o}}$ 260 - Problèmes Combinatoires et Théorie des Graphes, Orsay (1976), 399-401.