Combinatorics, Probability and Computing (2006) 15, 185–191. © 2006 Cambridge University Press
doi:10.1017/S0963548305007108 Printed in the United Kingdom

Triangle-Free Hypergraphs

ERVIN GYŐRI[†]

Alfréd Rényi Institute of Mathematics, Budapest, Hungary
(e-mail: gyori@renyi.hu)

Received 6 September 2004; revised 15 May 2005

For Béla Bollobás on his 60th birthday

In this paper, we give sharp upper bounds on the maximum number of edges in very unbalanced bipartite graphs not containing any cycle of length 6. To prove this, we estimate roughly the sum of the sizes of the hyperedges in triangle-free multi-hypergraphs.

Introduction

In [2], Erdős, Sárközy and Sós studied some number-theoretic problems on product representations of squares. To prove a number-theoretic conjecture, they needed the following conjecture.

Let $G(X, Y)$ be a bipartite graph with colour classes X, Y where $|X| = m \leqslant |Y| = n$. If the graph G does not contain any cycle of length 6 then $e(G) \leqslant 2n + c(mn)^{2/3}$ for some constant c.

In [4], the present author proved that $e(G) \leqslant 2n + m^2/2$. (We note that this result was proved for any m and n but it is weak if $n = o(m^2)$. De Caen and Székely [1] and Faudree and Simonovits [3] proved $e(G) \leqslant c(mn)^{2/3}$ for some constant c if $m^2 \geqslant n$.) The result above was sufficient to prove the number-theoretic conjecture but not sharp. It is clear that the coefficient 2 of n cannot be improved, but the term involving m can be improved. It was conjectured in [4] that $e(G) \leqslant 2n + m - 2$ holds and this is the best possible. However, we show that this conjecture is false. In this paper, we prove the estimate

$$e(G) \leqslant 2n + m^2/8$$

which is sharp if $n \geqslant m^2/16$.

Note that bipartite graphs can be considered as incidence bipartite graphs of multi-hypergraphs. ('Multi' means that we may have several copies of the same hyperedge.) The incidence bipartite graph is C_6-free if and only if the multi-hypergraph is triangle-free.

[†] Research partially supported by OTKA Research Fund 34702.

https://doi.org/10.1017/S0963548305007108 Published online by Cambridge University Press

We use the weakest possible definition of triangle: three hyperedges h_1, h_2, h_3 constitute a *triangle* if there are vertices v_1, v_2, v_3 such that $v_1, v_2 \in h_1$, $v_2, v_3 \in h_2$ and $v_3, v_1 \in h_3$.

Let the weight $w(e)$ of a hyperedge e be $|e| - 2$ and the total weight $w(H)$ of the hypergraph H be the sum of the weights of the hyperedges. Then we prove the following.

Theorem 1. *Let $H = (V, E)$ be a triangle-free multi-hypergraph. Then*

$$w(H) = \sum_{e \in E} (|e| - 2) \leqslant n^2/8$$

if n is large enough.

Remarks. **(1)** We prove the statement for $n \geqslant 100$, but for the sake of brevity we do not make special efforts to get the best estimate for n. On the other hand, we skip the proof of the upper estimate $(n^2 + c)/8$ with the best constant $c = 16$ for $n \leqslant 84$.

(2) There are some counterexamples if n is small.

Example 1. For $n = 9$, let the elements be arranged in a 3×3 grid and take two copies of the triplets in each row and each column. The resulting 3-uniform multi-hypergraph has 12 edges, and its total weight is 12, which is more than the required upper bound 10.

Example 2. Let $n = 12$, let the elements be arranged in a 3×4 grid and take two copies of each row and each column. The resulting multi-hypergraph has 14 edges, and its total weight is 20, which is more than the required upper bound 18.

Remark. Theorem 1 is sharp, as the following example shows.

Example 3. For n divisible by 4, take the hypergraph H with

$$V(H) = \{x_1, x_1', x_2, x_2', \ldots, x_{n/4}, x_{n/4}', y_1, y_1', y_2, y_2', \ldots, y_{n/4}, y_{n/4}'\},$$
$$E(H) = \{\{x_i, x_i', y_j, y_j'\} : 1 \leqslant i \leqslant n/4, 1 \leqslant j \leqslant n/4\}.$$

If 4 does not divide n then take the hypergraph above with $4\lceil n/4 \rceil$ vertices and truncate H by deleting the first $4\lceil n/4 \rceil - n$ vertices of x_1, x_1', y_1, y_1' from all hyperedges of H. Further examples can be obtained by taking some vertex pairs x_i, x_i' and replacing all 4-edges $\{x_i, x_i', y_j, y_j'\}$ by the 3-edges $\{x_i, y_j, y_j'\}, \{x_i', y_j, y_j'\}$.

Taking the incidence bipartite graph of the appropriate hypergraph, we obtain the following result.

Corollary. Let $G(X, Y)$ be a C_6-free bipartite graph with colour classes X, Y, such that $|X| = m > 100$, $|Y| = n > 100$. Then

$$e(G) \leqslant 2m + n^2/8.$$

This estimate is sharp too if $m \geqslant n^2/16$. Take the incidence bipartite graph of the hypergraph in Example 3 if $m = n^2/16$. If $m > n^2/16$ then first add $m - n^2/16$ copies

https://doi.org/10.1017/S0963548305007108 Published online by Cambridge University Press

of the 2-edge $\{x_1, x_1'\}$ and again take the incidence bipartite graph of the resulting hypergraph.

Before the proof of Theorem 1, we prove a special case, in which the idea of the proof is very clear and neat.

Theorem 2. *Let $H = (V, E)$ be a triangle-free 3-uniform hypergraph without multiple edges. Then*

$$|E| = \sum_{e \in E} (|e| - 2) \leqslant n^2/8.$$

Remark added in proof. We learned that P. Frankl, A. Füredi and G. Simonyi also proved Theorem 2 later, but independently.

Proof of Theorem 2. Note that for an arbitrary hyperedge $e \in E$, if $f \in E$ and $g \in E$ are hyperedges such that $|e \cap f| = |e \cap g| = 2$ then $e \cap f = e \cap g$, otherwise e, f, and g constitute a triangle on the vertices of e, a contradiction.

Now, if $e = xyz \in E$ is an arbitrary hyperedge, then we may assume that, say, $\{x, y\}, \{x, z\} \nsubseteq f$ for any hyperedge $f \neq e$ in H. Then replace the hyperedge e by the 2-edges xy, xz, and similarly every hyperedge is replaced by two 2-edges. Note that we do not take the same edge to replace two different hyperedges, since then this edge is the intersection of these two hyperedges and they must have been replaced by other 2-edge pairs.

Claim 1. *The resulting graph G is triangle-free.*

Proof of Claim 1. Suppose that G contains a triangle xyz. If, say, xy and yz are obtained from the same hyperedge e and xz is obtained from a hyperedge $\{x, z, u\}$, then $e \cap f = \{x, z\}$, and then by definition f should have been replaced by the edges ux, uz, a contradiction. If xy, yz, and xz are obtained from three different hyperedges, then these hyperedges constitute a triangle on the vertices x, y, z, a contradiction.

Thus, Turán's [5] theorem implies that G has at most $n^2/4$ edges and so H has at most $n^2/8$ hyperedges. □

Proof of Theorem 1

Obviously, we may assume that each hyperedge contains at least three vertices, and we cannot have 3 or more copies any of the hyperedges because they would constitute a triangle on three vertices of the hyperedge.

We will proceed by induction on n. Note that we cannot start the induction with some small values of n since, as we have seen, there are counterexamples for small n. Instead, when $n \geqslant 84$, we use an upper bound $((n-k)^2 + c)/8$ for $n - k$, with some small constant c and use this to prove the upper bound $(n^2 + c - k)/8$ for n (we call this 'strong' induction). After a small number of inductive steps, we will have $w(H) \leqslant n^2/8$. The existence of such a constant c is trivial by finiteness, and we skip the proof that $w(H) \leqslant (n^2 + c)/8$ with $c = 16$ for $n \leqslant 84$ because it is lengthy (although $c \leqslant 43$ falls out of the proof).

First, we prove a simple, but very useful lemma.

https://doi.org/10.1017/S0963548305007108 Published online by Cambridge University Press

Lemma 1. *Let x be a vertex of a triangle-free multi-hypergraph H and let $h_1, h_2, \dots, h_m$ be edges such that $|h_i| \geqslant 3$ and $x \in h_i$ for $i = 1, \dots, m$. Then $|\cup_{i=1}^{m} h_i| > m$.*

Proof. Take two hyperedges $h_1, h_2 \in H$ and suppose that $h_1 \cap h_2 \neq \{x\}$, say $y \in h_1 \cap h_2 - \{x\}$. Then for any hyperedge h_3 we have $h_1 \cap h_3, h_2 \cap h_3 \subseteq \{x, y\}$, since otherwise h_1, h_2, h_3 constitute a triangle, a contradiction. It implies that if $|h_1 \cap h_2| \geqslant 3$ then for any hyperedge $h_3 \in H$, we have $h_1 \cap h_3 = h_2 \cap h_3 = \{x\}$. So, the components of the hypergraph H' with edge set $E(H') = \{h - \{x\} : h \in E(H)\}$ either have at most two hyperedges or the hyperedges of the component constitute a 'star', *i.e.*, every two hyperedges meet in the same vertex. Thus, the number of hyperedges in H' is at most $|\cup_{h' \in E(H')} h'|$ and equality holds if and only if each component of H' is a double copy of a 2-edge. □

Now, we prove Theorem 1 by distinguishing a few cases concerning the hyperedge sizes.

Case 1. *For $k > 4$, there is a k-edge $e = x_1 \cdots x_k$ in H.*

Let H_i denote the subhypergraph of the hyperedges containing x_i and let $n_i + 1 = |X_i|$, where $X_i = \cup_{h \in H_i} h$ for $i = 1, \dots, k$. Note that if $f \in H_i$ and $g \in H_j$ are distinct hyperedges (two identical copies are OK!) for some $i \neq j$ then $f \cap g \subseteq \{x_i, x_j\}$, since otherwise e, f and g constitute a triangle.

If x_i is not contained in any hyperedge $h \in H_j, j \neq i$, then X_i does not meet any other X_j and Lemma 1 implies that if we delete x_i from all hyperedges $h \neq e$ of H then the weight decreases by at most $|X_i| - 1$.

Let us consider the at least two element intersections of e with the hyperedges of H, *i.e.*, the sets $\{S_f = e \cap f : f \in H, |e \cap f| > 1\}$. (If $S_f = S_g$ then we take two, or in general, several copies of the same set!)

If $|S_f| > 2$ then S_f is not met by another set S_g, because then e, f and g constitute a triangle in H, a contradiction. Then by applying Lemma 1 for the pairwise vertex-disjoint hypergraphs $H_i - f$, $x_i \in S_f$ we obtain that $n_i \leqslant |X_i| - |f|$, and if we delete all the vertices $x_i \in S_f$ from all hyperedges $h \neq e$ of H then the weight decreases by at most $|\cup_{x_i \in S_f} X_i| - 2$.

If $|S_f| = 2$ and S_f is met by another set S_g then $S_f = S_g$ (actually $f \cap g = S_f$ since otherwise e, f and g constitute a triangle in H). Suppose that exactly the sets $f_1 = f, f_2, \dots, f_s$ have $e \cap f_i = S_f$. Then, by applying Lemma 1 for the vertex-disjoint hypergraphs $H_i - \{f_1, \dots, f_s\}$, $x_i \in S_f$, we obtain that if we delete all the vertices $x_i \in S_f$ from all hyperedges $h \neq e$ of H then the weight decreases again by at most $|\cup_{x_i \in S_f} X_i| - 2$.

So if we delete all vertices of e from all hyperedges of $H - e$ then the weight decreases by at most $n - 2$ and equality can hold only if there is a hyperedge f containing e. Applying the induction hypothesis estimate $((n-k)^2 + c)/8$ for the resulting hypergraph, we obtain that $w(H) \leqslant ((n-k)^2 + c)/8 + n - 2 + k - 2$, which is at most $(n^2 + c - k)/8$ if $n \geqslant 19$. (The difference increases as k increases and it is easy to get the bound for $k = 5$.)

So we may assume that every hyperedge in H has three or four vertices.

https://doi.org/10.1017/S0963548305007108 Published online by Cambridge University Press

Case 2. *There are* 4*-edges* $e = x_1x_2x_3x_4$ *and* $f = x_1x_2x_3x_5$ *of* H *sharing exactly three vertices, or there is a* 4*-edge* $e = x_1x_2x_3x_4$ *containing a* 3*-edge* $f = x_1x_2x_3$.

Then it is easy to see that replacing f by one more copy of e we would get a triangle-free hypergraph of not smaller weight.

So we may assume that H contains only 3- and 4-edges, and apart from the double copies, all intersections contain at most two vertices. In this case we prove a more general lemma than the statement of Theorem 1.

Lemma 2. *Let* H *be a triangle-free hypergraph of* 2*-,* 3*- and* 4*-edges on* n *vertices such that there are no parallel* 2*-edges; moreover, no* 2*-edge is contained in any other hyperedge and the intersection of any two hyperedges contains at most two vertices unless they are two copies of the same* 3*- or* 4*-edge. Let us define the weight* f *of the a hyperedge* e *by* $2(|e| - 2)$ *if* e *is a* 3*- or* 4*-edge and by 1 if it is a* 2*-edge. Then the total weight* $f(H)$ *of the edges is at most* $n^2/4$ *if* n *is large enough.*

Note that Lemma 2 implies the required statement. Theorem 1 is just the special case with no 2-edge. Again we have some counterexamples for small n (see the examples above). We try prove the lemma again by 'strong' induction, *i.e.*, if we have the induction hypothesis upper bound $((n-m)^2 + c)/4$ for $n - m$ with some integer c then we prove the upper bound $(n^2 + c - m)/4$ for n. We find a direct proof when it does not work.

Proof of Lemma 2. The proof relies on the idea of the proof of Theorem 2. Take all the single (= no parallel copy) edges e met by another edge f in two vertices and fix such an f. Take the sets $X = e \cap f$ and $Y = e - X$. Replace the hyperedge e by the 2-edges joining X and Y. Note that e is replaced by $|X||Y| \geqslant 2(|e| - 2)$ 2-edges, and these 2-edge sets are pairwise disjoint we do not create parallel 2-edges. (If not, then there is a hyperedge g meeting both X and Y, and then e, f, g constitute a triangle on some vertices $x \in X \cap g$, $y \in Y \cap g$ and $z \in X - \{x\}$, a contradiction.) If e is a single 3- or 4-edge not met by any other hyperedge in two vertices, then take an arbitrary set $X \subset e$ of two vertices and replace e again by the 2-edges joining X and $Y = e - X$. These 2-edge sets are still pairwise disjoint and each hyperedge is replaced by 2-edges except the double 3- or 4-edges. What we have obtained is special hypergraph H_0: a set of single 2-edges and some double 3- and 4-edges. (If there is no such double 3- or 4-edge then we are done, by Turán's theorem and the forthcoming Claim 2.)

Claim 2. *The hypergraph* H_0 *is triangle-free.*

Proof of Claim 2. If H_0 contains a triangle of the hyperedges e, f, g on the vertices x, y, z then the appropriate hyperedges clearly constitute a triangle on the same vertex set in H unless two 2-edges, say $e = xy$ and $f = xz$, are obtained from the very same hyperedge h. By the construction, hyperedge g is obtained from another hyperedge. If g is a 3- or 4-edge containing vertices y, z, w, then h,g and the other copy g' of g constitute a triangle on $\{y, z, w\}$, a contradiction. If yz is a 2-edge obtained from a hyperedge e_0, then h meets

https://doi.org/10.1017/S0963548305007108 Published online by Cambridge University Press

e_0 in at least two elements, and we have replaced the hyperedge e_0 by 2-edges because it was met by another hyperedge in two vertices. So there was a hyperedge f_0 containing, say, y but not containing z, which defined the edges replacing e_0. Then there is a vertex $w \in e_0 \cap f_0 - \{y\}$ and e_0, f_0, h constitute a triangle on y, z, w, a contradiction. □

The obtained hypergraph H_0 has a nice structure: it has double 3- and 4-edges and single 2-edges such that the intersection of any two hyperedges has at most one vertex unless these hyperedges are parallel, *i.e.*, have the same vertex set.

Now we show that we may assume that H_0 does not contain any 4-edge because the 'strong' induction step can be done in this case. (As usual, the proof works only if n is large.)

Suppose that e and f are 4-edges of H_0 on the same vertex set $\{x_1, x_2, x_3, x_4\}$. Let H_i denote the subhypergraph of $H_0 - \{e, f\}$ consisting of the hyperedges containing x_i and let $X_i = \cup_{h \in H_i} h$ for $i = 1, 2, 3, 4$. Note that if $g \in H_i$ and $h \in H_j$ are distinct hyperedges (two identical copies are OK!) for some $i \neq j$ then $g \cap h = \emptyset$, since otherwise e, g and h constitute a triangle. So the sets X_i are pairwise disjoint and, say, X_1 has at most $n/4$ vertices by the pigeonhole principle. Suppose that H_1 contains m_1 2-edges, $2m_2$ 3-edges (m_2 parallel pairs) and $2m_3$ 4-edges (again, m_3 parallel pairs). Then we have $m_1 + 2m_2 + 3m_3 \leqslant n/4 - 1$ by Lemma 1, the fact that no 2-edge is contained in any other edge and $|X_1| \leqslant n/4$. Let us delete x_1 from all hyperedges of H_0 and keep just one copy of the arising 2-edges if we get two parallel 2-edges. Then the resulting hypergraph H_0' is triangle-free and has only double 3- and 4-edges and single 2-edges such that the intersection of any two hyperedges has at most one vertex unless these hyperedges are parallel, *i.e.*, have the same vertex set. So we can apply the induction hypothesis that $f(H_0') \leqslant ((n-1)^2 + c)/4$. Deleting x_1, the total weight of the hyperedges decreases by $4 + m_1 + 3m_2 + 4m_3 \leqslant 4 + 3(m_1 + 2m_2 + 3m_3)/2 \leqslant 3n/8 + 5/2$. So $f(H_0) \leqslant ((n-1)^2 + c)/4 + 3n/8 + 5/2$, which is at most $(n^2 + c - 1)/4$ if $n \geqslant 24$.

What remains to be proved is as follows.

Claim 3. *If H_0 is a triangle-free multi-hypergraph of n vertices, n_3 double 3-edges (altogether $2n_3$ 3-edges) and n_2 single 2-edges then $f(H_0) = 4n_3 + n_2 \leqslant n^2/4$ (if n is large).*

Remark. There are some counterexamples for small n. Example 1 is just one of them, for which $4n_3 + n_2 = 24 > n^2/4 = 81/4$.

Proof of Claim 3. If the hypergraph is a graph (*i.e.*, there are no 3-edges) then we are done, by Turán's theorem. Suppose that H_0 has a double hyperedge $e = x_1x_2x_3$. Then we try to proceed by 'strong' induction.

Let $e = f = x_1x_2x_3$ be a double 3-edge of H_0. The other hyperedges in H_0 meet e in at most one vertex. Let $2k$ be the number of 3-edges meeting e in exactly one vertex. If g and h are any two hyperedges meeting e in one vertex, but not two copies of the same triple, then clearly $(g - e) \cap (h - e) = \emptyset$ by the triangle-freedom of H_0. If we delete the vertices x_1, x_2, x_3 from the hyperedges of H_0 then we get 2-edges from the triples meeting e in one vertex, so the sum $4n_3 + n_2$ decreases by at most $4 + 3k + (n - 2k - 3)$ and by the the induction hypothesis, the total weight is at most $((n-3)^2 + c)/4$ in the resulting

https://doi.org/10.1017/S0963548305007108 Published online by Cambridge University Press

multi-hypergraph. So $f(H_0) \leqslant (n^2 + c - 3)/4$ if $4 + 3k + (n - 2k - 3) \leqslant (6n - 12)/4$, *i.e.*, if $2k \leqslant n - 8$.

So, we may assume that $2k \geqslant n - 7$ for any choice of $x_1x_2x_3$ and at least $n - 4$ vertices are contained in 3-edges.

Again, let H_1, H_2 and H_3 denote the hypergraphs of the hyperedges $h \neq e, f$ of H_0 containing x_1, x_2 and x_3, respectively, and let X_1, X_2 and X_3 be the (pairwise disjoint) vertex sets of H_1, H_2 and H_3, resp., with $|X_1| = n_1, |X_2| = n_2, |X_3| = n_3$. Suppose that $2k_1, 2k_2$ and $2k_3$ are the numbers of 3-edges in H_1, H_2 and H_3, respectively. Deleting the vertices x_1, x_2 from the hyperedges of H_0 and applying the induction hypothesis estimate $((n-2)^2 + c)/4$, we get $f(H_0) \leqslant 4 + 3k_1 + 3k_2 + (n_1 - 2k_1 - 1) + (n_2 - 2k_2 - 1) + ((n-2)^2 + c)/4 \leqslant 2 + 3n_1/2 + 3n_2/2 + ((n-2)^2 + c)/4$, which is at most $(n^2 + c - 2)/4$ if $3n_1 + 3n_2 \leqslant 2n - 7$. So, we may assume that $n_1 + n_2 \geqslant 2n/3 - 2$ and so $n_3 \leqslant n/3 + 2$. Similarly, we may assume that $n_1, n_2 \leqslant n/3 + 2$, and in general, if x is a vertex contained in a 3-edge, then the union of the hyperedges containing x has at most $n/3 + 4$ vertices.

Finally, deleting the vertex x_1 from the hyperedges of H_0 and applying the induction hypothesis estimate $((n-1)^2 + c)/4$, we get $f(H_0) \leqslant 3 + 3k_1 + (n_1 - 2k_1 - 1) + ((n-1)^2 + c)/4 \leqslant 3/2 + 3n_1/2 + ((n-1)^2 + c)/4$, which is at most $(n^2 + c - 1)/4$ if $n_1 \leqslant n/3 - 4/3$. So, we may assume that $n_1 \geqslant n/3 - 1$, and similarly we may assume that $n_2, n_3 \geqslant n/3 - 1$, and in general, if x is a vertex contained in a 3-edge, then the union of the hyperedges containing x has at least $n/3 - 1$ vertices.

Let y be a vertex not contained in any 3-edge. (There are at most four such vertices, as we have seen.) The vertex y is contained in at least four 2-edges, otherwise the induction goes trivially. So there is a 2-edge yz such that z is contained in a double 3-edge. So the hypergraph of the hyperedges containing z has at least $n/3 + 1$ vertices. Actually, it can be improved to $n/3 + 4/3$, since $n_1 - 2k_1 \geqslant 1$ in this case. From among these at least $n/3 + 4/3$ vertices just z is joined to y by a 2-edge by the triangle-freedom, so the number of 2-edges containing y is at most $2n/3 - 4/3$.

Now, we add up the weights of the hyperedges vertexwise: $4/3$ for the double 3-edges, $1/2$ for 2-edges containing the vertex. Since this sum is smaller for vertices contained in 3-edges, we may assume that there are four vertices not contained in a 3-edge. We obtain that $f(H_0) \leqslant 4(n-4)(n/6 + 3/2)/3 + 4(n/3 - 2/3)$, which is less than $n^2/4$ if $n \geqslant 84$. (It also yields that $f(H_0) \leqslant n^2/4 + 43$ for every n but the constant 43 as well the constant 84 can be improved easily, but it is very tiresome to get the best constants.) □

References

[1] de Caen, D. and Székely, L. A. (1992) The maximum size of 4- and 6-cycle free bipartite graphs on m, n vertices. In *Sets, Graphs and Numbers* (Proc. Coll. dedicated to the 60th birthday of A. Hajnal and V. T. Sós, Budapest, 1991), North-Holland, Amsterdam, pp. 135–142.

[2] Erdős, P., Sárközy, A. and Sós, V. T. (1995) On product representation of powers I. *European J. Combin.* **16** 323–331.

[3] Faudree, R. and Simonovits, M. On a class of degenerate extremal graph problems II. In preparation

[4] Győri, E. (1997) C_6-free bipartite graphs and product representation of squares. *Discr. Math.* **165/166** 371–375.

[5] Turán, P. (1941) On an extremal problem in graph theory (in Hungarian). *Mat. Fiz. Lapok* **48** 436–452.

https://doi.org/10.1017/S0963548305007108 Published online by Cambridge University Press