Turán numbers and batch codes*

Csilla Bujtás [1] Zsolt Tuza [1,2]

[1] Department of Computer Science and Systems Technology
University of Pannonia
Veszprém, Hungary

[2] Alfréd Rényi Institute of Mathematics
Hungarian Academy of Sciences
Budapest, Hungary

Latest update on 2013–09–18

Abstract

Combinatorial batch codes provide a tool for distributed data storage, with the feature of keeping privacy during information retrieval. Recently, Balachandran and Bhattacharya observed that the problem of constructing such uniform codes in an economic way can be formulated as a Turán-type question on hypergraphs. Here we establish general lower and upper bounds for this extremal problem, and also for its generalization where the forbidden family consists of those r-uniform hypergraphs H which satisfy the condition $k \geq |E(H)| > |V(H)| + q$ (for $k > q + r$ and $q > -r$ fixed). We also prove that, in the given range of parameters, the considered Turán function is asymptotically equal to the one restricted to $|E(H)| = k$, studied by Brown, Erdős and T. Sós. Both families contain some r-partite members — often called the 'degenerate case', characterized by the equality $\lim_{n\to\infty} \mathrm{ex}(n, \mathcal{F})/n^r = 0$ — and therefore their exact order of growth is not known.

Keywords: Turán number, hypergraph, combinatorial batch code.

AMS 2000 Subject Classification: 05D05, 05C65, 68R05.

* Research supported in part by the Hungarian Scientific Research Fund, OTKA grant T-81493, moreover by the European Union and Hungary, co-financed by the European Social Fund through the project TÁMOP-4.2.2.C-11/1/KONV-2012-0004 – National Research Center for Development and Market Introduction of Advanced Information and Communication Technologies.

1 Introduction

In this paper we study a Turán-type problem on uniform hypergraphs, which is motivated by optimization of distributed data storage enabling secure data retrieval under a certain protocol.

1.1 Terminology

Hypergraphs. A *hypergraph* H is a set system with vertex set $V(H)$ and edge set $E(H)$ where every edge $e \in E(H)$ is a nonempty subset of $V(H)$. The number of its vertices and edges is the *order* and the *size* of H, respectively. A hypergraph H is called *r-uniform* if each edge of it contains precisely r vertices. For short, sometimes we shall use the term *r-graph* for r-uniform hypergraphs. Graphs without loops are just 2-uniform hypergraphs. A hypergraph H_1 is a *subhypergraph* of H_2 if $V(H_1) \subseteq V(H_2)$ and $E(H_1) \subseteq E(H_2)$ holds, moreover we say that H_1 is an *induced subhypergraph* of H_2 if also $E(H_1) = \{e : e \subseteq V(H_1) \ \wedge \ e \in E(H_2)\}$ holds. In this paper graphs and hypergraphs are meant to be simple, that is without loops and multiple edges, unless stated otherwise explicitly.

Turán numbers. Given hypergraphs H and F, H is said to be *F-free* if H has no subhypergraph isomorphic to F. Similarly, if $\mathcal{F}$ is a family of hypergraphs, H is $\mathcal{F}$-free if it contains no subhypergraph isomorphic to any member of $\mathcal{F}$. In the problems considered here, the family $\mathcal{F}$ contains r-graphs for a fixed $r \geq 2$ and the property to be $\mathcal{F}$-free is considered only for r-graphs.

In a Turán-type (hypergraph) problem there is a given collection $\mathcal{F}$ of r-uniform hypergraphs and the main goal is to determine or to estimate the *Turán number* $\mathrm{ex}(n, \mathcal{F})$ which is the maximum number of edges in an $\mathcal{F}$-free r-uniform hypergraph on n vertices. In 1941 Turán [24] determined $\mathrm{ex}(n, K_t)$, that is the maximum size of a graph G of order n such that G contains no complete subgraph on t vertices. (The spacial case of $k = 3$ was already solved in 1907 by Mantel [19].) Since then lots of famous results have been proved (see the recent surveys [15, 18]), but many problems especially among the ones concerning hypergraphs seem notoriously hard.

Combinatorial batch codes. The notion of batch code was introduced by Ishai, Kushilevitz, Ostrovsky and Sahai [17] to represent the distributed storage of m items of data on n servers such that any at most k data items are recoverable by submitting at most t queries to each server.[1] In its combinatorial version [20], 'encoding' and 'decoding' mean simply that the data items are stored on and read from the servers. Its basic case, when the parameter t equals 1, can be defined as follows.[2]

- A *combinatorial batch code (CBC-system)* with parameters (m, k, n) is a multihypergraph H of order n and size m, such that the union of any i edges contains at least i vertices for every $1 \leq i \leq k$. For given parameters r, k, n, satisfying $r \geq 2$ and $r + 1 \leq k \leq n$, let $m(n, r, k)$ denote the maximum number m of edges such that an r-uniform CBC-(m, k, n)-system exists.

Optimization problems on combinatorial batch codes (mainly for the non-uniform case and under the condition $t = 1$) were studied in [3, 6, 7, 8, 9, 10, 20]. Recently, Balachandran and Bhattacharya [2] formulated the problem of determining the maximum size of r-uniform CBC-systems as a Turán multihypergraph problem. Clearly, an r-uniform multihypergraph H is a CBC-system with parameter k if and only if it has no subhypergraph of order $i - 1$ and size exactly i for all $r + 1 \leq i \leq k$.

A problem of Brown, Erdős and T. Sós. Brown, Erdős and T. Sós started to study the problems where, for fixed integers $2 \leq r \leq v$ and $k \geq 2$, all r-graphs on v vertices and with at least k edges are forbidden to occur as a subhypergraph of an r-graph [5].[3] The maximum size of such an r-graph of order n is denoted by $f^{(r)}(n, v, k) - 1$. A general lower bound on $f^{(r)}(n, v, k)$ was proved in [5] and later further famous results were given for the cases $v \geq k$ (see, e.g., [21, 12, 22, 23, 1]). In this paper, motivated by

[1] In the main part of the literature notations n and m are used in reversed role. Here the usual notation of hypergraph Turán problems is applied for CBCs (as done also in [2]).

[2] In this definition the vertices of the hypergraph represent the n servers, the edges represent the m data items, and an edge contains exactly those vertices which correspond to the servers storing the data items represented by the edge. Parameters k and $t = 1$ express the condition that every family of at most k edges has a system of distinct representatives. Applying Hall's Theorem we obtain the definition in the form given here.

[3] On graphs, the problem was first studied by Dirac in [11].

the optimization problem on uniform CBCs, we will study a problem closely related to the case $v \le k$.

Our problem setting. We shall consider Turán-type problems for the following families of forbidden subhypergraphs. The upper index '(r)' in the notation indicates that the family consists of r-graphs.

- $\mathcal{H}^{(r)}(k,q) = \{H : |E(H)| - |V(H)| = q+1 \ \wedge \ |E(H)| \le k\}$

  To study $\mathcal{H}^{(r)}(k,q)$-free hypergraphs, we put the following restrictions on the parameters:

  - $r \ge 2$ (The problem would be trivial for the 1-uniform case.)
  - $k \ge q+r+1$ ($|E(H)| \le q+r$ would imply $|V(H)| \le r-1$ and hence $\mathcal{H}^{(r)}(k,q) = \emptyset$.)
  - $q \ge -r+1$ (Negative values can be allowed for q. But if $q \le -r$, the family $\mathcal{H}^{(r)}(k,q)$ contains an r-graph with 1 edge and with at least r vertices, and hence $\mathrm{ex}(n, \mathcal{H}^{(r)}(k,q)) = 0$ would follow.)

- $\mathcal{F}^{(r)}(k,q) = \{H : |E(H)| - |V(H)| = q+1 \ \wedge \ |E(H)| = k\}$

  In general, $r \ge 2$, $k \ge q+r+1$ and $k \ge 2$ are assumed. Here we restrict ourselves to the cases with $q \ge -r+1$. Note that $\mathcal{F}^{(r)}(k,q)$ contains exactly those r-graphs which are forbidden in the Brown-Erdős-Sós problem with $v = k-q-1$, while $\mathcal{H}^{(r)}(k,q) = \cup_{i=r+q+1}^{k} \mathcal{F}^{(r)}(i,q)$.

Moreover, for $\mathcal{H}^{(r)}(k,q)$ and $\mathcal{F}^{(r)}(k,q)$, the family of multihypergraphs with the same defining property is denoted by $\mathcal{H}_M^{(r)}(k,q)$ and $\mathcal{F}_M^{(r)}(k,q)$, respectively. When the Turán number relates to the maximum size of a multihypergraph, the lower index M is used, as well. For instance, $\mathrm{ex}_M(n, \mathcal{H}_M^{(r)}(k,q))$ denotes the maximum number of edges in a multihypergraph such that every i edges cover at least $i-q-1$ vertices subject to $q+r+1 \le i \le k$. Note that if $q = -r+1$, already the presence of edges with multiplicity 2 is forbidden and consequently $\mathrm{ex}_M(n, \mathcal{H}_M^{(r)}(k,-r+1)) = \mathrm{ex}(n, \mathcal{H}^{(r)}(k,-r+1))$.

The next facts follow immediately from the definitions:

$$m(n,r,k) \;=\; \mathrm{ex}_M(n, \mathcal{H}_M^{(r)}(k,0))$$

$$\begin{aligned} \mathrm{ex}(n, \mathcal{H}^{(r)}(k,q)) \le \mathrm{ex}(n, \mathcal{F}^{(r)}(k,q)) &= f^{(r)}(n, k-q-1, k) - 1 \le \mathrm{ex}_M(n, \mathcal{F}_M^{(r)}(k,q)) \\ \mathrm{ex}(n, \mathcal{H}^{(r)}(k,q)) &\le \mathrm{ex}_M(n, \mathcal{H}_M^{(r)}(k,q)) \end{aligned}$$

1.2 Preliminaries and our results

The following general lower bound was proved by Brown, Erdős and T. Sós [5] for $\mathcal{F}^{(r)}(k,q)$-free r-graphs under the previously given conditions ($r \ge 2$, $k \ge q+r+1$ and $k \ge 2$).

$$f^{(r)}(n, k-q-1, k) = \Omega(n^{r-1+\frac{q+r}{k-1}}). \tag{1}$$

Paterson, Stinson and Wei [20] proved that if $q = 0$ but all the r-graphs from $\mathcal{H}^{(r)}(k,0)$ are forbidden, the lower bound (1) still remains valid[4]:

$$m(n,r,k) \ge \mathrm{ex}(n, \mathcal{H}^{(r)}(k,0)) = \Omega(n^{r-1+\frac{r}{k-1}}).$$

We prove in Section 2 that the lower bound (1) can be extended also to our general case:

$$\mathrm{ex}(n, \mathcal{H}^{(r)}(k,q)) = \Omega(n^{r-1+\frac{q+r}{k-1}}). \tag{2}$$

Concerning upper bounds, our main result proved in Section 4 says that

$$\mathrm{ex}(n, \mathcal{H}^{(r)}(k,q)) = \mathcal{O}(n^{r-1+\frac{1}{\left\lfloor \frac{k}{q+r+1} \right\rfloor}}) \tag{3}$$

for every fixed $r \ge 2$ and $k \ge q+r+1$. The basis of the proof is $r = 2$ (graphs), for which the order of the upper bound follows already from a theorem of Faudree and Simonovits [13]; in fact they only forbid a subfamily of $\mathcal{F}^{(2)}(k,q)$. Under the stronger condition of excluding $\mathcal{H}^{(2)}(k,q)$ instead of $\mathcal{F}^{(2)}(k,q)$, however, a better and explicit constant can be derived on the former; and this can in turn be proved to be valid on the latter as well. For this reason, we do not simply derive the result from the one in [13] but prove the new upper bound in our Theorem 5. The more general result for hypergraphs is given in Theorem 7. In Section 4 we also prove that the same upper bound (3) is valid for multihypergraphs, in fact not only the orders of

[4]For the cases with $k - \lceil \log k \rceil \le r \le k-1$, Balachandran and Bhattacharya [2] proved the better lower bound $m(n,r,k) = \Omega(n^r)$

these upper bounds are equal but also the relatively small leading coefficients are the same.

Section 5 is devoted to exploring the connection between the Turán numbers of $\mathcal{H}^{(r)}(k,q)$ and $\mathcal{F}^{(r)}(k,q)$. The general message there is that any later improvement in the estimates concerning $\mathcal{H}^{(r)}(k,q)$ will automatically yield an improvement for $\mathcal{F}^{(r)}(k,q)$ as well, and vice versa.[5] By Theorem 11, if $r=2$ and the parameters k and q are fixed, the difference is bounded by a constant $d(k,q)$:

$$f^{(2)}(n,k-q-1,k)-\mathrm{ex}(n,\mathcal{H}^{(2)}(k,q))\leq d(k,q).$$

For $r\geq 3$, by Theorem 13 we obtain the upper bound

$$f^{(r)}(n,k-q-1,k)-\mathrm{ex}(n,\mathcal{H}^{(r)}(k,q))=\mathcal{O}(n^{r-1}),$$

which is somewhat weaker but still strong enough to prove that the Turán numbers $\mathrm{ex}(n,\mathcal{F}^{(r)}(k,q))$ and $\mathrm{ex}(n,\mathcal{H}^{(r)}(k,q))$ have the same order of growth. On the other hand, the question of sharpness of Theorem 13 remains open:

Problem 1 *For the triplets (r,k,q) of integers in the range $r\geq 2$, $q\geq -r+1$, and $k\geq q+r+1$, determine the infimum value $s(r,k,q)$ of constants $s\geq 0$ such that*

$$f^{(r)}(n,k-q-1,k)-\mathrm{ex}(n,\mathcal{H}^{(r)}(k,q))=\mathcal{O}(n^{s})$$

as $n\to\infty$.

Conjecture 2 *The infimum $s(r,k,q)$ in Problem 1 is attained as minimum.*

Our Theorem 11 shows that $s(2,k,q)=0$ holds for all pairs (k,q) in the given range, and so Conjecture 2 is confirmed for $r=2$.

At the end of this introductory section, we return to uniform combinatorial batch codes. The previous upper bound given for $m(n,r,k)$ in [20] was improved recently by Balachandran and Bhattacharya [2]:

$$m(n,r,k) \;=\; \mathcal{O}(n^{r-\frac{1}{2^{r-1}}}) \qquad \text{if} \quad 3\leq r\leq k-1-\lceil \log k\rceil. \tag{4}$$

[5] Obviously, by this principle, one should seek upper bounds for $\mathcal{H}^{(r)}(k,q)$ and lower bounds for $\mathcal{F}^{(r)}(k,q)$.

Our Corollary 9 yields a further improvement in the range $r \leq k/2 - 1$. Especially, we have

$$m(n, r, k) = \mathcal{O}\left(n^{r-1+\frac{1}{\lfloor \frac{k}{r+1} \rfloor}}\right). \tag{5}$$

Comparing (4) and (5), the difference is significant already for parameters complying with $3 \leq r = k/2-1$. For these cases, (4) gives exponent $r-1/2^{r-1}$ whilst our bound (5) yields exponent $r - 1/2$.

2 Lower bound

In this section we prove a lower bound on $\mathrm{ex}(n, \mathcal{H}^{(r)}(k, q))$ whose order is the same as proved in [5] for $f(n, k-q-1, k)$; that is, for the case when only the subhypergraphs on exactly $k-q-1$ vertices and with k edges are forbidden.

Theorem 3 *For all fixed triplets of integers r, k, q with $r \geq 2$, $q \geq -r+1$ and $k \geq r+q+1$ we have*

$$\mathrm{ex}(n, \mathcal{H}^{(r)}(k, q)) = \Omega(n^{r-1+\frac{q+r}{k-1}}) = \Omega(n^{\frac{kr-k+q+1}{k-1}}).$$

Proof. We apply the probabilistic method. Our proof technique is similar to those in [5] and [20]. We let $p = cn^{-1+\frac{q+r}{k-1}}$, where the constant $c = c(r, k, q)$ will be chosen later. Note that the lower bound $-r+1$ on q implies $pn \geq cn^{\frac{1}{k-1}}$, i.e. pn tends to infinity with n whenever r, k, q are constants.

Let $H^{(r)}_{n,p}$ be the random r-uniform hypergraph of order n with edge probability p. That is, $H^{(r)}_{n,p}$ has n vertices, and for each r-tuple S of vertices the probability that S is an edge is p, independently of (any decisions on) the other r-tuples. We denote by E the number of edges in $H^{(r)}_{n,p}$, and by F the number of forbidden subhypergraphs in $H^{(r)}_{n,p}$; by 'forbidden' we mean that for some $i \leq k$, some $i-q-1$ vertices contain at least $i > 0$ edges.

We will estimate the expected value of $E - F$, more precisely our goal is to show that the inequality $\mathbb{E}(E - F) \geq \mathbb{E}(E)/2$ on the expected values is true for a suitable choice of the constant c. Once $\mathbb{E}(E - F) \geq \mathbb{E}(E)/2$ is ensured, we obtain that there exists a (non-random) hypergraph with twice as many edges as the number of its forbidden subhypergraphs, hence removing one edge from each of the latter we obtain a hypergraph with the required structure and with at least $\mathbb{E}(E)/2 = \frac{p}{2}\binom{n}{r}$ edges.

By the additivity of expectation we have

$$\mathbb{E}(E-F)=\mathbb{E}(E)-\mathbb{E}(F),$$

moreover it is clear by definition that

$$\mathbb{E}(E)=p\cdot\binom{n}{r}=(\tfrac{1}{r!}+o(1))\cdot p\cdot n^r=(\tfrac{1}{r!}+o(1))\cdot c\cdot n^{r-1+\frac{q+r}{k-1}} \tag{6}$$

for any fixed r as $n\to\infty$. Hence we need to find an upper bound on $\mathbb{E}(F)$.

We consider the following set I of those values of i for which an $(i-q-1)$-element vertex subset is large enough to accommodate some forbidden subhypergraph:

$$I=\left\{i:\ \ i\le\binom{i-q-1}{r}\quad\wedge\quad q+r+2\le i\le k\right\}.$$

It should be noted first that if $I=\emptyset$, then also $\mathcal{H}^{(r)}(k,q)=\emptyset$ holds and hence $\mathrm{ex}(n,\mathcal{H}^{(r)}(k,q))=\binom{n}{r}$. In this case, the lower bound in the theorem is trivially valid, as the condition $k\ge r+q+1\ge 2$ implies $(q+r)/(k-1)\le 1$.

From now on, we assume that $I\neq\emptyset$. Consider any $i\in I$. On any $i-q-1$ vertices the number of ways we can select i edges is $\binom{\binom{i-q-1}{r}}{i}$, and the probability for each of those selections to be a subhypergraph of $H^{(r)}_{n,p}$ is exactly p^i. Since there are $\binom{n}{i-q-1}$ ways to select $i-q-1$ vertices, we obtain the following upper bound:

$$\begin{aligned}
\mathbb{E}(F)\ &\le\ \sum_{i\in I}\binom{\binom{i-q-1}{r}}{i}\cdot p^i\cdot\binom{n}{i-q-1}\\
&<\ \sum_{i\in I}\frac{\binom{\binom{i-q-1}{r}}{i}}{(i-q-1)!}\cdot p^i\,n^{i-q-1}\\
&<\ \left(\max_{i\in I}\frac{\binom{\binom{i-q-1}{r}}{i}}{(i-q-1)!}\right)\cdot p^k\,n^{k-q-1}\cdot\sum_{i=q+r+2}^{k}(pn)^{i-k}\\
&\le\ (C_{k,q,r}+o(1))\cdot c^k\cdot n^{k-q-1-k\,(1-\frac{q+r}{k-1})}\\
&=\ (C_{k,q,r}+o(1))\cdot c^k\cdot n^{r-1+\frac{q+r}{k-1}}
\end{aligned} \tag{7}$$

where $C_{k,q,r}$ abbreviates the maximum value of $\frac{\binom{\binom{i-q-1}{r}}{i}}{(i-q-1)!}$ taken over the range I of i.

Compare the rightmost formula of (6) with (7). The terms in parentheses containing $o(1)$ are essentially constant, while the main part of (6) is $c \cdot n^{r-1+\frac{q+r}{k-1}}$ whereas that of (7) grows with $c^k \cdot n^{r-1+\frac{q+r}{k-1}}$. Thus, choosing c sufficiently small, the required inequality $\mathbb{E}(E-F) \geq \mathbb{E}(E)/2$ will hold for n large. This completes the proof of the theorem. $\square$

Remark 4 *It can also be ensured (again by a suitable choice of c) that $\mathbb{E}(E-F)/\mathbb{E}(E)$ is arbitrarily close to 1. This is not needed for the proof above, but it may be of interest in the context of batch codes with specified rate (cf. e.g. [17]).*

3 Upper bound for graphs

First we prove an upper bound on $\mathrm{ex}(n, \mathcal{H}^{(2)}(k,q))$.

Theorem 5 *For every three integers $q \geq -1$, $k \geq 2q+6$ and $n \geq k$, we have*

$$\mathrm{ex}(n, \mathcal{H}^{(2)}(k,q)) < C \cdot n^{1+\frac{1}{\left\lfloor \frac{k}{q+3} \right\rfloor}} + (q+2)n,$$

where $C = (q+2)^{\frac{1}{\left\lfloor \frac{k}{q+3} \right\rfloor}}$.

Proof. Introduce the notation $h = \left\lfloor \frac{k}{q+3} \right\rfloor$ and assume for a contradiction that there exists a graph G of order n in which, for every $q+3 \leq i \leq k$, every i edges cover at least $i-q$ vertices and the number of edges in G is

$$|E(G)| = m \geq C \cdot n^{1+\frac{1}{h}} + (q+2)n.$$

Thus, the average degree $\bar{d}(G) = \bar{d}$ satisfies

$$\bar{d} = \frac{2m}{n} \geq 2C \cdot n^{\frac{1}{h}} + 2(q+2).$$

Moreover, every graph of average degree $\bar{d}$ has a subgraph of minimum degree greater than $\bar{d}/2$.[6] Hence, we have a subgraph F with minimum degree $\delta(F) = \delta$ such that

$$\delta > C \cdot n^{\frac{1}{h}} + q + 2. \tag{8}$$

[6]Just delete sequentially the vertices of degree smaller than or equal to $\bar{d}/2$. After each single step the average degree is greater than or equal to $\bar{d}$. Hence, finally we obtain a subgraph of minimum degree greater than $\bar{d}/2$.

Claim A. The order of F satisfies

$$|V(F)| > \frac{(\delta - q - 2)^h}{q+2}.$$

Proof. Choose a vertex x of F as a root and construct the breadth-first search tree (BFS-tree) of F rooted in x. Let L_i denote the set of vertices on the ith level of the BFS-tree, and introduce the notation $\ell_i = |L_i|$. The edges of F not belonging to the BFS-tree will be called additional edges.

First we consider the vertices of the first $h^* = \left\lfloor \frac{k-q-1}{q+3} \right\rfloor$ levels and prove that each vertex $v \in L_i$ is incident with at most $q+1$ additional edges, if $0 \le i \le h^* - 1$. Assume to the contrary that there exist $q+2$ such additional edges and consider the union of paths on the BFS-tree connecting the end-vertices of these additional edges with the root vertex x. This means $q+3$ (not necessarily edge-disjoint) paths each of length at most h^*, and at least one of them (the path between v and x) is of length at most $h^* - 1$. They form a tree, let the number of its edges be denoted by p. Together with the $q+2$ additional edges we have

$$p + q + 2 \le h^* - 1 + (q+2)h^* + q + 2 = (q+3)h^* + q + 1 \le k$$

edges, which cover only $p+1$ vertices. This contradicts the assumed property of G. Therefore, we may have at most $q+1$ additional edges incident with vertex v.

Now, we prove a bound on the number ℓ_i of vertices on the ith level if $2 \le i \le h^*$. The sum of the vertex degrees over the set L_{i-1} cannot be smaller than $\delta\ell_{i-1}$. On the other hand, each of these ℓ_{i-1} vertices is incident with at most $q+1$ additional edges, moreover there are $\ell_{i-1} + \ell_i$ edges of the BFS-tree each of them being incident with exactly one vertex from L_{i-1}. As follows,

$$\begin{aligned} \delta\, \ell_{i-1} &\le \ell_{i-1} + \ell_i + (q+1)\ell_{i-1} \\ (\delta - q - 2)\, \ell_{i-1} &\le \ell_i, \end{aligned}$$

for every $2 \le i \le h^*$. Since $\ell_1 \ge \delta - q - 2$ is also true, the recursive formula gives

$$|V(F)| \ge \ell_{h^*} \ge (\delta - q - 2)^{h^*} \ge \frac{(\delta - q - 2)^{h^*}}{q+2}. \tag{9}$$

If $h = h^*$, that is if $k \equiv q+1$ or $q+2 \pmod{q+3}$, this already proves Claim A.

In the other case we have $h = h^* + 1$ and claim that every vertex $u \in L_{h-1}$ is incident with at most $q+1$ additional edges whose other end is in $L_{h-2} \cup L_{h-1}$. Then, assume for a contradiction that there are at least $q+2$ such edges. Again, take these $q+2$ additional edges together with the paths in the BFS-tree connecting their ends with the root. In this subgraph we have only at most $(q+3)(h-1)+q+2 < k$ edges, which cover fewer vertices by $q+1$ than the number of edges. Proved by this contradiction, we have at most $q+1$ additional edges of the described type.

A similar argumentation shows that each $w \in L_h$ might be incident with at most $q+1$ additional edges whose other end is in L_{h-1}. Assuming the presence of $q+2$ such edges, we have at most $h+(q+2)(h-1)+q+2 \leq k$ edges together with the paths between their ends and the root. Moreover, this cardinality exceeds the number of covered vertices by $q+1$. Thus, we have a contradiction, which proves the property stated for w.

By these two bounds on the number of additional edges we can estimate the sum s of vertex degrees over L_{h-1} as follows:

$$\delta\,\ell_{h-1} \leq s \leq \ell_{h-1} + \ell_h + (q+1)\ell_{h-1} + (q+1)\ell_h.$$

Together with (9) this implies

$$|V(F)| \geq \ell_h \geq \frac{\delta - q - 2}{q+2}\,\ell_{h-1} \geq \frac{(\delta-q-2)^h}{q+2},$$

and proves Claim A. $\Diamond$

Turning to graph G, inequality (8) and Claim A yield the contradiction

$$n \geq |V(F)| > \frac{\left(C \cdot n^{1/h}\right)^h}{q+2} = n.$$

Therefore, in a $\mathcal{H}^{(2)}(k,q)$-free graph the number of edges must be smaller than $C \cdot n^{1+1/h} + (q+2)n$, as stated in the theorem. $\square$

Corollary 6 *For every three integers $q \geq -1$, $k \geq 2q+6$ and $n \geq k$, we have*

$$\mathrm{ex}_M(n, \mathcal{H}_M^{(2)}(k,q)) < C \cdot n^{1+\frac{1}{\lfloor \frac{k}{q+3} \rfloor}} + (q+2)n,$$

where $C = (q+2)^{\frac{1}{\lfloor \frac{k}{q+3} \rfloor}}$.

Proof. The BFS-tree of a multigraph G is meant as a simple graph. That is, if an edge uv has multiplicity $\mu > 1$ in G, and uv is an edge in the BFS-tree, then only one edge uv belongs to the tree, the remaining $\mu - 1$ copies are additional edges. With this setting every detail of the previous proof remains valid for multigraphs. $\square$

4 Upper bound for hypergraphs

In this section we study the problem for hypergraphs. The upper bound on $\mathrm{ex}(n, \mathcal{H}^{(r)}(k,q))$ will be obtained by using Theorem 5.

Theorem 7 *Let n, k, r and q be integers such that $r \geq 2$, $q \geq -r+1$ and $n \geq k \geq 2q+2r+2$, moreover let $C' = (q+r)^{\frac{1}{\lfloor \frac{k}{q+r+1} \rfloor}}$. Then,*

$$\mathrm{ex}(n, \mathcal{H}^{(r)}(k,q)) < \frac{2C'}{r!} \cdot n^{r-1+\frac{1}{\lfloor \frac{k}{q+r+1} \rfloor}} + \frac{2(q+r)}{r!} \cdot n^{r-1}.$$

Proof. Consider an $\mathcal{H}^{(r)}(k,q)$-free r-graph H. Let its order and size be denoted by n and m, respectively. For a set $S \subseteq V(H)$ denote by $d(S)$ the number of edges of H which contain S entirely. By double counting we have

$$\sum_{S \subset V(H),\ |S|=r-2} d(S) = m \binom{r}{r-2},$$

and for the average value $\bar{d}_{r-2}$ of $d(S)$ over the $(r-2)$-element subsets of $V(H)$

$$\bar{d}_{r-2} = m \, \frac{\binom{r}{r-2}}{\binom{n}{r-2}}$$

holds. Thus, there exists an $S^* \subset V(H)$ of cardinality $r-2$ satisfying

$$d(S^*) \geq m \, \frac{\binom{r}{r-2}}{\binom{n}{r-2}}.$$

Deleting the edges which do not contain S^* entirely, in addition deleting the $r-2$ vertices of S^* from the remaining edges, we obtain a graph G with $V(G) = V(H)$ and

$$E(G) = \{e \setminus S^* : S^* \subset e \ \wedge \ e \in E(H)\}, \qquad |E(G)| \geq m \, \frac{\binom{r}{r-2}}{\binom{n}{r-2}}.$$

Since every i edges ($i \le k$) cover at least $i-q$ vertices in H, every i edges cover at least $i-q-r+2$ vertices in G. Moreover, the conditions given in Theorem 5 hold for $n'=n$, $k'=k$ and $q'=q+r-2$. Then, we obtain

$$m\,\frac{\binom{r}{r-2}}{\binom{n}{r-2}} \le |E(G)| < (q+r)^{\frac{1}{\lfloor\frac{k}{q+r+1}\rfloor}}\, n^{1+\frac{1}{\lfloor\frac{k}{q+r+1}\rfloor}} + (q+r)n, \tag{10}$$

from which

$$m < \frac{2C'}{r!}\, n^{r-1+\frac{1}{\lfloor\frac{k}{q+r+1}\rfloor}} + \frac{2(q+r)}{r!}\cdot n^{r-1}$$

follows. This implies the same upper bound for $\mathrm{ex}(n,\mathcal{H}^{(r)}(k,q))$. □

The above proof remains valid if the r-graph H is allowed to have multiple edges. The only difference is that we must refer to Corollary 6 instead of Theorem 5. Hence, for multihypergraphs the same upper bound can be stated. In addition, since $m(n,r,k) = \mathrm{ex}_M(n,\mathcal{H}_M^{(r)}(k,0))$, we obtain a new upper bound for the maximum size $m(n,r,k)$ of r-uniform CBC-systems with parameters n and k.

Corollary 8 *Let n, k, r and q be integers such that $r\ge 2$, $q \ge -r+1$ and $n \ge k \ge 2q+2r+2$, moreover let $C' = (q+r)^{\frac{1}{\lfloor\frac{k}{q+r+1}\rfloor}}$. Then,*

$$\mathrm{ex}_M(n,\mathcal{H}_M^{(r)}(k,q)) < \frac{2C'}{r!}\cdot n^{r-1+\frac{1}{\lfloor\frac{k}{q+r+1}\rfloor}} + \frac{2(q+r)}{r!}\cdot n^{r-1}.$$

Corollary 9 *Let n, k, r be integers such that $r \ge 2$ and $n \ge k \ge 2r+2$, moreover let $C'' = r^{\frac{1}{\lfloor\frac{k}{r+1}\rfloor}}$. Then,*

$$m(n,r,k) < \frac{2C''}{r!}\cdot n^{r-1+\frac{1}{\lfloor\frac{k}{r+1}\rfloor}} + \frac{2}{(r-1)!}\cdot n^{r-1}.$$

5 Asymptotic equality of Turán numbers

Up to this point we were concerned with the problem of $\mathcal{H}^{(r)}(k,q)$-free hypergraphs; it is different from the one studied by Brown, Erdős and T. Sós [4, 5], where only the subhypergraphs with exactly $k-q-1$ vertices and k edges are forbidden. In this section we show that $\mathrm{ex}(n,\mathcal{H}^{(r)}(k,q))$ and $f^{(r)}(n,k-q-1,k)-1$ are asymptotically equal. For graphs ($r=2$), our result is better as there exists a constant upper bound (depending only on k and q) on their difference. As a consequence, we obtain a new upper bound on $f^{(2)}(n, v, k)$ subject to $v \geq (k+4)/2$.

First we prove the following lemma. For fixed parameters k, q and for a given graph G, a subgraph G' is said to be *forbidden (for* (k, q)*)* if $G' \in \mathcal{H}^{(2)}(k, q)$, moreover G' is *maximal forbidden (for* (k, q)*)*, if it cannot be extended into a forbidden subgraph of larger order.

Lemma 10 *Let k and q be integers such that $q \geq -1$ and $k \geq q + 3$, and let G be a graph of order at least $k - q - 1$. If a subgraph $G' \subset G$ is maximal forbidden for (k, q), then either G' has k edges or it is the union of one or more components of G.*

Proof. Assume that G' is a forbidden subgraph of G and $|E(G')| < k$. If there exists an edge $uv \in E(G)$ such that $u \in V(G')$ and $v \in V(G) \setminus V(G')$, then the subgraph G'' obtained by extending G' with the vertex v and with the edge uv satisfies $|E(G'')| - |V(G'')| = q + 1$ and $|E(G'')| = |E(G')| + 1 \leq k$. Hence G'' is forbidden for (k, q) and consequently, G' is not maximal forbidden. On the other hand, if the subgraph of G which is induced by $V(G')$ contains some edge e not in G', then with any vertex $v \in V(G) \setminus V(G')$, the subgraph $G' + e + v$ is forbidden for (k, q) and again, G' is not a maximal forbidden subgraph. Therefore, if G' is of order smaller than k and it is a maximal forbidden subgraph for (k, q), then G' is a component of G, or it is the union of some components of G. $\square$

Clearly, $f^{(2)}(n, k - q - 1, k) \geq \mathrm{ex}(n, \mathcal{H}^{(2)}(k, q))$. The following theorem states that the difference between them is bounded by a constant, once the parameters k and q are fixed.

Theorem 11 *For every pair k, q of integers satisfying $q \geq -1$ and $k \geq q+3$ there exists a constant $d = d(k, q)$ such that for every $n \geq k - q - 1$,*

$$f^{(2)}(n, k - q - 1, k) - \mathrm{ex}(n, \mathcal{H}^{(2)}(k, q)) \leq d.$$

Proof. For given parameters k and q first define $z := \min\{i : \ q + 3 \leq i \leq \binom{i-q-1}{2}\}$. If $k > z$, there is no forbidden subgraph for (k, q) and consequently, $f^{(2)}(n, k - q - 1, k) = \mathrm{ex}(n, \mathcal{H}^{(2)}(k, q)) = \binom{n}{2}$. Otherwise, z is the possible minimum size of a subgraph forbidden for (k, q). By Theorem 3

$$\mathrm{ex}(n, \mathcal{H}^{(2)}(k, q)) = \Omega(n^{1 + \frac{q+2}{k-1}})$$

holds, thus there exists an n_0 (depending only on k and q) such that for all $n \geq n_0$

$$\frac{z}{z-q-1} \cdot n \leq \mathrm{ex}(n, \mathcal{H}^{(2)}(k,q)).$$

Consequently, the following finite maximum exists:

$$d = \max\left(\left\{\frac{z}{z-q-1} \cdot n - \mathrm{ex}(n, \mathcal{H}^{(2)}(k,q)) + 1 : \ n \in \mathbb{N}\right\} \cup \{1\}\right). \quad (11)$$

We claim that d is a suitable constant for our theorem. To prove this, let us consider an $\mathcal{F}(k,q)$-free graph G on n vertices and with $f^{(2)}(n, k-q-1, k)-1$ edges. If G is $\mathcal{H}^{(2)}(k,q)$-free as well, $f^{(2)}(n,k-q-1,k)-1$ is equal to $\mathrm{ex}(n, \mathcal{H}^{(2)}(k,q))$, and since $d \geq 1$, the theorem holds for k, q and n.

In the other case, G contains a subgraph G_1 maximal forbidden for (k,q). Clearly, G_1 has fewer than k edges, hence by Lemma 10, G_1 is an induced subgraph and there is no edge between $V(G_1)$ and $V(G) \setminus V(G_1)$. Then, the remaining subgraph $G - G_1$ is either $\mathcal{H}^{(2)}(k,q)$-free or contains a subgraph G_2 of size smaller than k, which is maximal forbidden for (k,q). Iteratively applying this procedure, finally we have vertex-disjoint maximal forbidden subgraphs $G_1, \ldots G_j$ and the $\mathcal{H}^{(2)}(k,q)$-free subgraph G' induced by $V(G) \setminus \cup_{i=1}^{j} V(G_i)$, such that each edge of G is contained in exactly one of $G', G_1, \ldots G_j$. As $q+1 \geq 0$ and for every $1 \leq i \leq j$ we have $z \leq |E(G_i)| \leq k-1$, applying Lemma 10, we obtain

$$\frac{|E(G_i)|}{|V(G_i)|} = \frac{|E(G_i)|}{|E(G_i) - q - 1|} \leq \frac{z}{z-q-1}.$$

Using notations $n_1 = \sum_{i=1}^{j} |V(G_i)|$ and $n_2 = |V(G')| = n - n_1$, moreover the definition (11) of d

$$\begin{aligned}
|E(G)| &= f^{(2)}(n, k-q-1, k) - 1 \leq \frac{z}{z-q-1} \cdot n_1 + \mathrm{ex}(n_2, \mathcal{H}^{(2)}(k,q)) \\
&\leq \mathrm{ex}(n_1, \mathcal{H}^{(2)}(k,q)) + d - 1 + \mathrm{ex}(n_2, \mathcal{H}^{(2)}(k,q)) \\
&\leq \mathrm{ex}(n, \mathcal{H}^{(2)}(k,q)) + d - 1,
\end{aligned}$$

which yields

$$f^{(2)}(n, k-q-1, k) - \mathrm{ex}(n, \mathcal{H}^{(2)}(k,q)) \leq d,$$

as stated. $\square$

Corollary 12 *Let v and k be integers such that $2 \le v \le k$ and let $C = (k-v+1)^{\frac{1}{\lfloor \frac{k}{k-v+2} \rfloor}}$. Then, there exists a constant D such that for every n*

$$f^{(2)}(n,v,k) < C \cdot n^{1+\frac{1}{\lfloor \frac{k}{k-v+2} \rfloor}} + (k-v+1)n + D.$$

Proof. Let q denote $k-v-1$. Then, under the given conditions we have $-1 \le q \le k-3$ and $C = (q+2)^{\frac{1}{\lfloor \frac{k}{q+3} \rfloor}}$. Theorems 5 and 11 immediately imply the existence of a constant D such that for every n

$$f^{(2)}(n,k-q-1,k) < C \cdot n^{1+\frac{1}{\lfloor \frac{k}{q+3} \rfloor}} + (q+2)n + D.$$

This is equivalent to the statement of the corollary. □

Theorem 13 *For every four integers r,k,q and n satisfying $r \ge 2$ and $2 \le q+r+1 \le k \le n$,*

$$f^{(r)}(n,k-q-1,k) - \mathrm{ex}(n,\mathcal{H}^{(r)}(k,q)) \le (k-1)\binom{n-1}{r-1}$$

holds. Hence, for every fixed r, k, and q we have

$$f^{(r)}(n,k-q-1,k) = (1+o(1))\,\mathrm{ex}(n,\mathcal{H}^{(r)}(k,q)).$$

Proof. Consider any extremal r-graph H^* for $\mathcal{F}^{(r)}(k,q)$ on the n-element vertex set V. By definition, H^* is $\mathcal{F}^{(r)}(k,q)$-free. If H^* is also $\mathcal{H}^{(r)}(k,q)$-free, then $f^{(r)}(n,k-q-1,k) = \mathrm{ex}(n,\mathcal{H}^{(r)}(k,q))$ holds and we have nothing to prove. Otherwise we select the longest possible sequence of subhypergraphs $H_i \subset H^*$ ($i = 1,2,\ldots,\ell$) under the following conditions:

- Each H_i is isomorphic to some member of $\mathcal{H}^{(r)}(k,q) \setminus \mathcal{F}^{(r)}(k,q)$.
- Under the previous condition, H_1 is maximal in H^*.
- Under the previous conditions, H_i is maximal in $H^* \setminus \bigcup_{j=1}^{i-1} H_j$ for each $2 \le i \le \ell$.

Eventually we obtain an $\mathcal{H}^{(r)}(k,q)$-free hypergraph from H^* by removing at most $(k-1)\cdot\ell$ edges, because each H_i has at most $k-1$ edges. Thus, the proof will be done if we prove that $\ell \le \binom{n-1}{r-1}$ holds.

Let e_i be an arbitrarily chosen edge of H_i and let f_i be an $(r-1)$-element subset of e_i, which we fix (again arbitrarily) for $i = 1,2,\dots,\ell$. Should $f_i \subset e_j$ hold for some $1 \le i < j \le \ell$, the hypergraph $H_i \cup \{e_j\}$ would also be isomorphic to some member of $\mathcal{H}^{(r)}(k,q)$. This contradicts the choice (maximality) of H_i. Consequently, for all $i = 1,2,\dots,\ell$ we have:

- $|f_i| = r-1$,
- $|V \setminus e_i| = n-r$,
- $f_i \cap (V \setminus e_i) = \emptyset$,
- $f_i \cap (V \setminus e_j) \ne \emptyset$ whenever $1 \le i < j \le \ell$.

Thus, applying a theorem of Frankl [14],[7] the number of set pairs $(f_i, V\setminus e_i)$ is at most $\binom{(r-1)+(n-r)}{r-1} = \binom{n-1}{r-1}$. □

Corollary 14 *Let r, v, k be integers such that $r \ge 2$ and $(k+2r)/2 \le v \le k+r-2$ and let $C = (k+r-v-1)^{\frac{1}{\lfloor \frac{k}{k+r-v} \rfloor}}$. Then,*

$$f^{(r)}(n,v,k) \le \frac{2C}{r!} \cdot n^{r-1+\frac{1}{\lfloor \frac{k}{k+r-v} \rfloor}} + \mathcal{O}(n^{r-1}).$$

References

[1] N. Alon and A. Shapira, *On an Extremal Hypergraph Problem of Brown, Erdős and Sós*, Combinatorica 26 (2006), 627–645.

[2] N. Balachandran and S. Bhattacharya, *On an Extremal Hypergraph Problem Related to Combinatorial Batch Codes*, Manuscript (2012).

[3] S. Bhattacharya, S. Ruj and B. Roy, *Combinatorial batch codes: A Lower Bound and Optimal Constructions*, Advances in Mathematics of Communications 6 (2012), 165–174.

[7]Set pairs with prescribed intersection properties can be applied in many kinds of extremal problems (not only on graphs and hypergraphs). A detailed account on those methods and results is given in the two-part survey [25, 26].

[4] 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.

[5] W. G. Brown, P. Erdős and V. T. Sós, *Some extremal problems on r-graphs*, New directions in the theory of graphs (Proc. Third Ann Arbor Conf., Univ. Michigan, Ann Arbor, Mich., 1971), 53–63, Academic Press, New York, 1973.

[6] R. A. Brualdi, K. P. Kiernan, S. A. Meyer and M. W. Schroeder, *Combinatorial batch codes and transversal matroids*, Adv. Math. Commun., 4 (2010), 419–431. Erratum *ibid.* p. 597.

[7] Cs. Bujtás and Zs. Tuza, *Optimal batch codes: Many items or low retrieval requirement*, Adv. Math. Commun., **5** (2011), 529–541.

[8] Cs. Bujtás and Zs. Tuza, *Optimal combinatorial batch codes derived from dual systems*, Miskolc Math. Notes 12 (1) (2011), 11–23.

[9] Cs. Bujtás and Zs. Tuza, *Combinatorial batch codes: Extremal problems under Hall-type conditions*, Electr. Notes in Discrete Math. 38 (2011), 201–206.

[10] Cs. Bujtás and Zs. Tuza: *Relaxations of Hall's Condition: Optimal batch codes with multiple queries*, Applicable Analysis and Discrete Mathematics, 6 (1) (2012), 72–81.

[11] G. Dirac: *Extensions of Turán's theorem on graphs*, Acta Math. Acad. Sci. Hungar. 14 (1963), 417–422.

[12] 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.

[13] R. Faudree and M. Simonovits, *On a class of degenerate extremal graph problems*, Combinatorica 3 (1983), 83–93.

[14] P. Frankl, *An extremal problem for two families of sets*, European J. Combin. 3 (1982), 125–127.

[15] Z. Füredi and M. Simonovits, *The history of degenerate (bipartite) extremal graph problems*, in: Erdős Centennial (L. Lovász et al., Eds.), Bolyai Society Mathematical Studies 25 (2013), 169–264.

[16] J. R. Griggs, M. Simonovits and G. R. Thomas, *Extremal graphs with bounded densities of small subgraphs*, Journal of Graph Theory, 29 (1998), 185–207.

[17] Y. Ishai, E. Kushilevitz, R. Ostrovsky and A. Sahai, *Batch codes and their applications*, In: Proceedings of the 36th Annual ACM Symposium on Theory of Computing, ACM Press, New York, 2004, 262–271.

[18] P. Keevash, *Hypergraph Turan problems*, Surveys in Combinatorics, Cambridge University Press, 2011, 83–140.

[19] W. Mantel, *Problem 28*, Wiskundige Opgaven 10 (1907), 60–61.

[20] M. B. Paterson, D. R. Stinson and R. Wei, *Combinatorial batch codes*, Adv. Math. Commun., **3** (2009), 13–27.

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

[22] G. N. Sárközy and S. M. Selkow, *An extension of the Ruzsa-Szemerédi Theorem*, Combinatorica 25 (2005), 77–84.

[23] G. N. Sárközy and S. M. Selkow, *On a Turán-type hypergraph problem of Brown, Erdős and T. Sós*, Discrete Math. 297 (2005), 190–195.

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

[25] Zs. Tuza, Applications of the set-pair method in extremal hypergraph theory. “Extremal Problems for Finite Sets” (P. Frankl et al., eds.), Bolyai Society Mathematical Studies 3, 1994, 479–514.

[26] Zs. Tuza, Applications of the set-pair method in extremal problems, II., “Combinatorics, Paul Erdős is Eighty” (D. Miklós et al., eds.), Bolyai Society Mathematical Studies 2, 1996, 459–490.