On an Extremal Hypergraph Problem Related to Combinatorial Batch Codes

Niranjan Balachandran

Department of Mathematics,
Indian Institute of Technology Bombay,
Mumbai.
E-mail: niranj@math.iitb.ac.in

Srimanta Bhattacharya*

Centre of Excellence in Cryptology,
Indian Statistical Institute,
Kolkata.
E-mail: srimanta_r@isical.ac.in

Abstract

Let n, r, k be positive integers such that $3 \leq k < n$ and $2 \leq r \leq k-1$. Let $m(n,r,k)$ denote the maximum number of edges an r-uniform hypergraph on n vertices can have under the condition that any collection of i edges spans at least i vertices for all $1 \leq i \leq k$. We are interested in the asymptotic nature of $m(n,r,k)$ for fixed r and k as $n \to \infty$. This problem is related to the forbidden hypergraph problem introduced by Brown, Erdős, and Sós and very recently discussed in the context of *combinatorial batch codes* (CBCs). In this short paper we obtain the following results.

(i) Using a result due to Erdős we are able to show $m(n,r,k) = o(n^r)$ for $7 \leq k$, and $3 \leq r \leq k-1-\lceil \log k \rceil$. This result is best possible with respect to the upper bound on r as we subsequently show through explicit construction that for $6 \leq k$, and $k - \lceil \log k \rceil \leq r \leq k-1, m(n,r,k) = \Theta(n^r)$. This explicit construction improves on the non-constructive general lower bound obtained by Brown, Erdős, and Sós for the considered parameter values.

(ii) For 2-uniform CBCs we obtain the following results.

(a) We provide exact value of $m(n,2,5)$ for $n \geq 5$.

(b) Using a result of Lazebnik *et al.* regarding maximum size of graphs with large girth, we improve the existing lower bound on $m(n,2,k)$ $(\Omega(n^{\frac{k+1}{k-1}}))$ for all $k \geq 8$ and infinitely many values of n.

(c) We show $m(n,2,k) = O(n^{1+\frac{1}{\lfloor \frac{k}{4} \rfloor}})$ by using a result due to Bondy and Simonovits, and also show $m(n,2,k) = \Theta(n^{\frac{3}{2}})$ for $k = 6, 7, 8$ by using a result of Kővári, Sós, and Turán.

Keywords: extremal hypergraphs, Turán numbers. combinatorial batch codes.

*Corresponding author. Tel.: +9133 2575 4316; fax: +9133 2577 3104

1 Preliminaries

Hypergraphs and Turán numbers

We briefly recall some definitions, and set up our notation.

A hypergraph F is a tuple $F := (\mathcal{V}, \mathcal{F})$, where $\mathcal{V}$ is a set of *vertices* and $\mathcal{F}$ is a family of subsets of $\mathcal{V}$. Sets of $\mathcal{F}$ are called *edges* of the hypergraph and cardinality of $\mathcal{F}$ is called *size* of the hypergraph; we will denote the size of F by $|F|$. A hypergraph is called *simple* if it does not contain repeated edges, i.e., there are no multiple copies of any edge; it is called *r-uniform* if each of its edges has cardinality r. For a vertex $x \in \mathcal{V}$, degree of x is the number of edges in $\mathcal{F}$ containing x. Further, by $K_n^{(r)}$ we will denote the complete r-uniform hypergraph on n vertices, and by $K^{(r)}(l, \dots, l)$ we will denote the complete r-uniform r-partite hypergraph with l vertices in each part. We will denote by $K(s,t)$ the complete bipartite graph with partite sets of size s and t respectively and by C_i, a cycle of length i.

Let $\mathcal{H}$ be a family of r-uniform hypergraphs. By the *Turán number* of the family $\mathcal{H}$ denoted by $\mathrm{ex}(n, \mathcal{H})$, we mean the maximum size of an r-uniform hypergraph on n vertices that does not contain a copy of any of the hypergraphs of $\mathcal{H}$ as a sub-hypergraph; such a hypergraph (which may not be unique) will be termed *extremal* for the family $\mathcal{H}$. A word on notation here: Following their common use in literature, we will use the notation $\mathrm{ex}(n, \mathcal{H})$ to denote Turán number of the family $\mathcal{H}$ taken over simple hypergraphs only, i.e., to denote maximum size of a simple hypergraph without containing any member of $\mathcal{H}$, and use $\mathrm{ex}^*(n, \mathcal{H})$ when we allow considered hypergraphs to have repeated edges.

For functions f and g, we will write $f = O(g)$ if there is an absolute constant c such that, $f(n)/g(n) \leq c$ for sufficiently large n; $f = o(g)$ if $\lim_{n\to\infty} f(n)/g(n) = 0$. Also we will write $f = \Omega(g)$ if $g = O(f)$; $f = \Theta(g)$ if $f = O(g)$ and $g = O(f)$.

Combinatorial Batch Codes and an extremal problem

The notion of *Batch Codes* was introduced in [15] as an abstraction of a load balancing problem in a distributed database setup. Loosely speaking, an (m, N, k, n, t)*-batch code models the problem of storing m data items into n servers in such a way that any k (we will refer to this parameter as *retrievability parameter*) of the n data items may be retrieved by reading at most t items from each server and the overall storage to be limited to N. In this article, we will exclusively consider the case of $t = 1$ as that seems to capture the essence of the problem and this is the case for most of the work done in this area. So, henceforth, we will not mention this parameter in any expression with the understanding that $t = 1$ case is considered.

Combinatorial Batch Codes (CBCs), also introduced in [15], and subsequently studied in [18, 6, 8, 9, 10, 1], is a subclass which models the scenario when each of the N stored data items is a copy of each of the m input data items, i.e., the m input data items are replicated among n servers. This restriction makes the problem a purely combinatorial one which can be studied in the setting of a hypergraph. Without providing further details (cf. [18, 8, 1]) we state the following theorem of [18] which characterizes CBCs in the setting of a hypergraph.

Theorem 1.1 ([18]). *A hypergraph $(\mathcal{V}, \mathcal{F})$ represents an (m, N, k, n)-CBC if $|\mathcal{V}| = n$, $|\mathcal{F}| = m$, $\sum_{F\in\mathcal{F}} |F| = N$ and every collection of i edges of $\mathcal{F}$ contains at least i vertices for $1 \leq i \leq k$.*

*In batch code literature, number of data items is denoted by n and number of servers is denoted by m. In this article, we deviate from this and reverse the roles of m and n to make it consistent with the common notations used in the hypergraph setting

Henceforth, we will refer to the hypergraph representing a CBC as a CBC with corresponding parameters. Naturally a CBC will be termed r-uniform if the corresponding hypergraph is r-uniform.

The problem that we will address in this paper concerns maximizing the number of data items (m) of a uniform CBC for given values of the number of servers (n) and the retrievability parameter (k). Equivalently, and more formally we have the following extremal problem.

> *Let n, r, k be positive integers such that $3 \leq k < n$ and $2 \leq r \leq k-1$. Determine $m(n, r, k)$, the maximum number of edges an r-uniform hypergraph on n vertices can have under the condition that any collection of i edges, $1 \leq i \leq k$, spans at least i vertices.*

This is a forbidden hypergraph problem, where we have the following family of forbidden hypergraphs:

$$\mathcal{G}_r(k) = \{H : H \text{ is an } r\text{-uniform hypergraph with } i \text{ edges and } < i \text{ vertices for } 1 \leq i \leq k\}, \tag{1}$$

and we are interested in the asymptotic nature of $m(n, r, k) = \mathrm{ex}^*(n, \mathcal{G}_r(k))$ for fixed r and k as $n \to \infty$. Here it is important to note that members of $\mathcal{G}_r(k)$, as well as extremal hypergraphs for $\mathcal{G}_r(k)$ contain repeated edges. However, in the following, we show that for the purpose of understanding the asymptotic nature of $m(n, r, k)$ it is sufficient to restrict our attention to simple hypergraphs only.

Let H be an r-uniform hypergraph that is extremal for the collection $\mathcal{G}_r(k)$, and let H' be a simple r-uniform hypergraph that has maximum number of edges among all simple hypergraphs without having any member of $\mathcal{G}_r(k)$. Then we have the following.

Proposition 1.2. *Let $\mathcal{G}_r(k), H, H'$ be as defined above, then $|H'| \geq \frac{1}{r}|H|$.*

Proof. Since H does not have any member of $\mathcal{G}_r(k)$, an edge of H can have at most r copies. Hence, for the maximal simple sub-hypergraph of H'' of H it follows that $|H''| \geq \frac{1}{r}|H|$. Now, both H' and H'' are simple hypergraphs without any member of $\mathcal{G}_r(k)$; so, following the defintion of H' we have $|H'| \geq |H''| \geq \frac{1}{r}|H|$. □

Let

$$\mathcal{H}_r(k) = \{H : H \text{ is an } r\text{-uniform simple hypergraph with } i \text{ edges and } < i \text{ vertices for } 1 \leq i \leq k\} \subseteq \mathcal{G}_r(k). \tag{2}$$

Then, following above proposition, we have

$$\mathrm{ex}^*(n, \mathcal{G}_r(k)) \leq r\,\mathrm{ex}(n, \mathcal{H}_r(k)). \tag{3}$$

Also, since any simple hypergraph not containing any member of $\mathcal{H}_r(k)$ does not contain any member of $\mathcal{G}_r(k)$, we have

$$\mathrm{ex}^*(n, \mathcal{G}_r(k)) \geq \mathrm{ex}(n, \mathcal{H}_r(k)). \tag{4}$$

So, from (3) and (4) we have

$$m(n, r, k) = \mathrm{ex}^*(n, \mathcal{G}_r(k)) = \Theta(\mathrm{ex}(n, \mathcal{H}_r(k))). \tag{5}$$

Hence, to understand asymptotic nature of $m(n, r, k)$ it is sufficient to consider Turán number of the family $\mathcal{H}_r(k)$ over simple hypergraphs. In fact, as we show in the next proposition, we can also take the subfamily

$$\mathcal{I}_r(k) = \{H : H \text{ is a simple } r\text{-uniform hypergraph with } i \text{ edges and } i-1 \text{ vertices for } r+3 \leq i \leq k\} \subseteq \mathcal{H}_r(k). \tag{6}$$

as the forbidden family for our problem at hand; though the subfamily is not essential for our results and discussed here only for the sake of completeness.

Proposition 1.3. *Let $\mathcal{H}_r(k)$ and $\mathcal{I}_r(k)$ be defined as above. Then* $\mathrm{ex}(n, \mathcal{I}_r(k)) = \mathrm{ex}(n, \mathcal{H}_r(k))$.

Proof. Since $\mathcal{H}_r(k) \supseteq \mathcal{I}_r(k)$, it trivially follows that $\mathrm{ex}(n, \mathcal{H}_r(k)) \leq \mathrm{ex}(n, \mathcal{I}_r(k))$. To prove the other direction, we claim the following.

Claim. *Let $H = (\mathcal{V}, \mathcal{F})$ be a simple r-uniform hypergraph such that $|\mathcal{F}| = k$, and $|\mathcal{V}| < k$. Then there is sub-hypergraph (not necessarily an induced one) $H' = (\mathcal{V}', \mathcal{F}')$ of H, such that $|\mathcal{F}'| = l$, $|\mathcal{V}'| = l - 1$, for some $r + 2 < l \leq k$.*

Proof. Indeed, and even in a stronger sense, one can arbitrarily delete edges from H until the condition is satisfied and guaranteed to get the desired sub-hypergraph H'. To see why this should always hold, first note that any collection of $r + 2$ edges of H spans at least $r + 2$ vertices, so we have $k > r + 2$. Next, consider the sequence of sub-hypergraphs $H = H_k \supset H_{k-1} \supset \ldots \supset H_{r+2}$, obtained by arbitrarily deleting edges one by one, where $|H_i| = i$, and let v_i be the number of vertices of H_i, $r + 2 \leq i \leq k$. Then we have $k - v_k \geq 1$, and $r+2-v_{r+2} \leq 0$. Now, since $v_i \geq v_{i-1}$ for $r+3 \leq i \leq k$, we have $(i - v_i) - (i - 1 - v_{i-1}) \leq 1$. Hence, there must be some $H' = H_l$, $r + 2 < l \leq k$, such that $l - v_l = 1$. This proves the claim.

So, following above claim, any simple hypergraph not containing any member of $\mathcal{I}_r(k)$ does not contain any member of $\mathcal{H}_r(k)$. Hence, $\mathrm{ex}(n, \mathcal{H}_r(k)) \geq \mathrm{ex}(n, \mathcal{I}_r(k))$, and hence, the proposition. □

So, henceforth, unless otherwise stated, all the considered hypergraphs will be simple only.

This type of extremal problem was introduced by Brown, Erdős, and Sós in [5], where the authors considered as forbidden family the following family of hypergraphs:

$$\mathcal{H}_r(p, q) = \{H : H \text{ is an } r\text{-uniform hypergraph with } p \text{ vertices and } q \text{ edges } \}. \tag{7}$$

They showed, through non-constructive arguments, the following lower bound:

$$\mathrm{ex}(n, \mathcal{H}_r(p, q)) = \Omega(n^{\frac{rq-p}{q-1}}). \tag{8}$$

A lower bound for $m(n, r, k)$ was given in [15], where the authors obtained the following result using a simple probabilistic argument:

$$m(n, r, k) = \Omega(n^{r-1}). \tag{9}$$

In [18], the authors using a method (method of alteration, same as in the proof of (8)) of [5], improved this lower bound:

$$m(n, r, k) = \Omega(n^{\frac{kr}{k-1}-1}). \tag{10}$$

On the other hand, in [18], the authors obtained the following upper bound:

$$m(n, r, k) \leq \frac{(k-1)}{\binom{k-1}{r}} \binom{n}{r}. \tag{11}$$

Now, it is trivial to see that this bound is met exactly for $r = 1$. In [18], it was shown by explicit construction, that this bound is tight (in exact terms) for the cases $r = k - 1$ and $r = k - 2$. Indeed, $k - 1$ copies of $K_n^{(k-1)}$ and $k - 2$ copies of $K_n^{(k-2)}$, respectively do the job.

So, (11) essentially shows an upper bound $O(n^r)$ for $m(n, r, k)$. We are interested in the values of r in the range $2 \leq r \leq k - 3$, for $k \geq 5$.

In this article we obtain the following results.

(i) In Section 2, we improve the upper bound (11) in an asymptotic sense. In particular, using a result due to Erdős ([13]) we are able to show $m(n,r,k) = o(n^r)$ for $7 \le k$, and $3 \le r \le k-1-\lceil \log k \rceil$. This result is best possible with respect to the upper bound on r as we subsequently show through explicit construction that for $6 \le k$, and $k - \lceil \log k \rceil \le r \le k-1, m(n,r,k) = \Theta(n^r)$.

   This explicit construction improves on the general lower bound (8) obtained by Brown, Erdős, and Sós for the parameters $p = k-1, q = k, k - \lceil \log k \rceil \le r \le k-1$, where $k \ge 6$. In this case their lower bound is the same as (10), i.e., $\Omega(n^{\frac{kr}{k-1}-1})$.

(ii) In Section 5, we deal with the graph case, i.e., 2-uniform CBCs and obtain the following results.

   (a) We provide exact value of $m(n,2,5)$ for $n \ge 5$.

   (b) Using a result of Lazebnik *et al.* [17] regarding maximum size of graphs with large girth we improve the existing lower bound (10) ($\Omega(n^{\frac{k+1}{k-1}})$) of [18] for all $k \ge 8$ and infinitely many values of n.

   (c) We show $m(n,2,k) = O(n^{1+\frac{1}{\lfloor \frac{k}{4} \rfloor}})$ by using a result due to Bondy and Simonovits [4], and also show $m(n,2,k) = \Theta(n^{\frac{3}{2}})$ for $k = 6,7,8$ by using a result of Kővári, Sós, and Turán.

2 r-UNIFORM CASE FOR $r \ge 3$

We begin this section by stating the following result due to Erdős that will be crucial in our proof of Theorem 2.2.

Theorem 2.1 ([13]). *Let $n > n_0(r,l), l > 1$. Then for sufficiently large C (C is independent of n, r, l),*

$$n^{r-\frac{C}{l^{r-1}}} < \mathrm{ex}(n, K^{(r)}(l,\ldots,l)) \le n^{r-\frac{1}{l^{r-1}}}.$$

Remark 2.1.1. The number n_0 in the above theorem is a constant depending only on r and l.

We first show that $m(n,r,k) = o(n^r)$ for $7 \le k$, and $3 \le r \le k-1-\lceil \log k \rceil$. All the logarithms mentioned in this paper are to the base 2.

Theorem 2.2. *Let $k \ge 7$, and $3 \le r \le k-1-\lceil \log k \rceil$. Then for sufficiently large n ($n > n_1(r)$, where n_1 is a constant depending only on r), $m(n,r,k) \le rn^{r-\frac{1}{2^{r-1}}}$.*

Proof. Let $7 \le u \le k$, $1 \le v \le u - 2\lceil \log u \rceil$ be such that $r = u - v - \lceil \log u \rceil$. It is clearly possible to find such u, v for the range of values of r as in the hypothesis. Consider the r-uniform, complete r-partite hypergraph $\mathcal{H} = (\mathcal{V}, \mathcal{F})$, where

$$\mathcal{V} := \{x_1, \ldots, x_{u-v-2\lceil \log u \rceil}, \ldots, x_r, y_1, \ldots, y_{u-v-2\lceil \log u \rceil}, \ldots, y_r\}$$

and

$$\mathcal{F} := \Big\{ \{z_1, \ldots, z_r\} : z_i \in \{x_i, y_i\}, 1 \le i \le r \Big\}.$$

So, applying Theorem 2.1 for $\mathcal{H} = K^{(r)}(2,\ldots,2)$, we get, for sufficiently large n, i.e., for $n > n_0(r,2) = n_1(r)$

$$\mathrm{ex}(n, \mathcal{H}) \le n^{r-\frac{1}{2^{r-1}}}. \tag{12}$$

Next, consider the r-uniform sub-hypergraph $\mathcal{H}' = (\mathcal{V}', \mathcal{F}')$ of $\mathcal{H}$, where

$$\mathcal{V} \supseteq \mathcal{V}' := \{x_1, \dots, x_{u-v-2\lceil \log u \rceil}, \dots, x_r, y_{u-v-2\lceil \log u \rceil+1}, \dots, y_r\}$$

and

$$\mathcal{F}' := \left\{ \{x_1, \dots, x_{u-v-2\lceil \log u \rceil}, z_{u-v-2\lceil \log u \rceil+1}, \dots, z_r\} : z_j \in \{x_j, y_j\}, u - v - 2\lceil \log u \rceil + 1 \leq j \leq r \right\}.$$

Since $v \geq 1$,

$$|\mathcal{V}'| = u - v \leq u - 1, \tag{13}$$

and

$$|\mathcal{F}'| = 2^{\lceil \log u \rceil} \geq u. \tag{14}$$

Now, since $u \leq k$, $\mathcal{H}' \in \mathcal{H}_r(k)$, where $\mathcal{H}_r(k)$ is the family defined in (2). So, we have

$$\mathrm{ex}(n, \mathcal{H}_r(k)) \leq \mathrm{ex}(n, \mathcal{H}') \leq \mathrm{ex}(n, \mathcal{H})$$

. Hence, from (3) and (12), we get for sufficiently large n, i.e., for $n > n_1(r)$

$$m(n, r, k) \leq r n^{r - \frac{1}{2^{r-1}}}.$$

□

A few remarks regarding the theorem are in order.

Remark 2.2.1.
1. Each edge of $\mathcal{F}'$ has the fixed set of vertices $\{x_1, \dots, x_{u-v-2\lceil \log u \rceil}\}$. This choice is arbitrary and any fixed set of $u - v - 2\lceil \log u \rceil$ vertices $\{z_1, \dots, z_{u-v-2\lceil \log u \rceil}\}$ could have been chosen maintaining the condition $z_j \in \{x_j, y_j\}, 1 \leq j \leq u - v - 2\lceil \log u \rceil$.
2. One can also see that the same construction with partite sets of size l, in conjunction with Theorem 2.1 gives us a corresponding result for $r \leq k - 1 - (l-1)\lceil \log_l k \rceil$.
3. Inequalities (13) and (14) are tight when u is a power of 2 and $v = 1$. In particular, when k is a power of 2 and $r = k - 1 - \log k$, we have $|\mathcal{V}'| = k - 1$ and $|\mathcal{F}'| = k$. So, k edges of $\mathcal{H}'$ span exactly $k - 1$ vertices.

In the following example, we demonstrate the construction for small parameter values.

Example 2.1. Let $r = 4, k = 8$. We choose $u = 8$. Which yields $v = u - \lceil \log u \rceil - r = 1$. So, $\mathcal{H} = (\mathcal{V}, \mathcal{F})$, where $\mathcal{V} = \{x_1, x_2, x_3, x_4, y_1, y_2, y_3, y_4\}$, and $\mathcal{F} = \{x_1, y_1\} \times \{x_2, y_2\} \times \{x_3, y_3\} \times \{x_4, y_4\}$. Similarly, as forbidden hypergraph we can consider $\mathcal{H}' = (\mathcal{V}', \mathcal{F}') \subset \mathcal{H}$, where $\mathcal{V}' = \{x_1, x_2, x_3, x_4, y_2, y_3, y_4\}$, and $\mathcal{F}' = \{x_1\} \times \{x_2, y_2\} \times \{x_3, y_3\} \times \{x_4, y_4\}$. Consequently, $7 = |\mathcal{V}'| \leq |\mathcal{F}'| = 8$.

Theorem 2.3. *$m(n, r, k) = \Theta(n^r)$ for $6 \leq k$, $k - \lceil \log k \rceil \leq r \leq k - 1$.*

Proof. Here we show $m(n, r, k) = \Omega(n^r)$ for the stated ranges of values of r and k. This, together with $m(n, r, k) = O(n^r)$ from (11), would imply $m(n, r, k) = \Theta(n^r)$. Also, first we prove the above for $r = k - \lceil \log k \rceil$, as this turns out to be the tight case and the same argument can easily be applied for the rest of the range of values of r. Note that the cases $r = k - 1$ and $r = k - 2$ have already been settled in

[18].
Construction: Consider the complete r-uniform, r-partite hypergraph $\mathcal{H} = (\mathcal{V}, \mathcal{F})$, where $\mathcal{V} = \mathcal{V}_1 \cup \mathcal{V}_2 \cup \ldots \cup \mathcal{V}_r$, such that $\mathcal{V}_i \cap \mathcal{V}_j = \emptyset$ for $i \neq j$, and $|\mathcal{V}_i| = \lfloor \frac{n+i-1}{r} \rfloor$, for $1 \leq i \leq r$. Clearly $|\mathcal{F}| = \Omega(n^r)$.

In the following claim, we show that $\mathcal{H}$ does not contain any member of $\mathcal{I}_r(k)$ †, where $\mathcal{I}_r(k)$ is as defined in (6). This, together with Proposition 1.3 and (4) would imply $m(n, r, k) = \Omega(n^r)$.

Claim. *$\mathcal{H}$ does not contain a sub-hypergraph $\mathcal{H}' = (\mathcal{V}', \mathcal{F}')$ such that $|\mathcal{V}'| = i - 1$ and $|\mathcal{F}'| \geq i$ for $r + 3 \leq i \leq k$.*

Proof. First, we observe that if there is a sub-hypergraph $\mathcal{H}' = (\mathcal{V}', \mathcal{F}')$ such that $|\mathcal{V}'| = i - 1$ and $|\mathcal{F}'| \geq i$ for some $r + 3 \leq i < k$ then there is another sub-hypergraph $\mathcal{H}'' = (\mathcal{V}'', \mathcal{F}'')$, such that $\mathcal{H}' \subseteq \mathcal{H}'' \subseteq \mathcal{H}$, $|\mathcal{V}''| = k - 1$, and $|\mathcal{F}''| \geq k$. To get $\mathcal{H}''$ from $\mathcal{H}'$ we simply add $k - i$ edges to $\mathcal{F}'$, where each edge contains exactly one unique vertex not contained in $\mathcal{V}'$, so that these newly added $k - i$ vertices are pairwise distinct for the newly added $k - i$ edges. This is always possible due to the structure of $\mathcal{H}$, provided there are $k - i$ distinct vertices in $\mathcal{V} \setminus \mathcal{V}'$. But this can be safely assumed because n is large enough; in fact, $n \geq k$ would suffice. Hence it is sufficient to establish that $\mathcal{H}$ does not contain a sub-hypergraph $\mathcal{H}' = (\mathcal{V}', \mathcal{F}')$ such that $|\mathcal{V}'| = k - 1$ and $|\mathcal{F}'| \geq k$.

Let $\mathcal{V}' \subseteq \mathcal{V}$ with $|\mathcal{V}'| = k-1$. Then $\mathcal{V}' = \mathcal{V}'_1 \cup \mathcal{V}'_2 \cup \ldots \cup \mathcal{V}'_r$, where $\mathcal{V}'_i \subseteq \mathcal{V}_i$, for $1 \leq i \leq r$. Furthermore, note that $|\mathcal{V}'_i| \geq 1$ for each i, otherwise $\mathcal{F}'$ would be empty, contradicting the assumption $|\mathcal{F}'| \geq k$. Then,

$$\sum_{i=1}^{r} |\mathcal{V}'_i| = |\mathcal{V}'| = k - 1, \text{ and } |\mathcal{F}'| = \prod_{i=1}^{r} |\mathcal{V}'_i|.$$

Here, it is important to note that $|\{\mathcal{V}'_i : |\mathcal{V}'_i| = 1, 1 \leq i \leq r\}| \geq 2r - k + 1 \geq 1$ for $6 \leq k$, and $r = k - \lceil \log k \rceil$.

Now, observe that maximum of $|\mathcal{F}'|$ is attained when there are exactly $(k - r - 1)$ $\mathcal{V}'_i$s, $1 \leq i \leq k$ with $|\mathcal{V}'_i| = 2$, and for the remaining $2r - k + 1$ $\mathcal{V}'_i$s, $|\mathcal{V}'_i| = 1$. This can be seen by the following shifting argument. Assuming without loss of generality that $|\mathcal{V}'_1| = 1$ and $|\mathcal{V}'_2| = l$, where $l \geq 3$, it is easy to see that by shifting a vertex from $\mathcal{V}'_2$ to $\mathcal{V}'_1$, $|\mathcal{V}'_1||\mathcal{V}'_2|$ becomes $2(l - 1) > l$, for $l \geq 3$, thereby increasing $\prod_{i=1}^{r} |\mathcal{V}'_i|$.

So, following the above argument, $|\mathcal{F}'|$ is maximized when the values $|\mathcal{V}'_i|$ are as equal as possible for $1 \leq i \leq r$, and since as observed earlier we always have $|\{\mathcal{V}'_i : |\mathcal{V}'_i| = 1, 1 \leq i \leq r\}| \geq 1$, it follows that $|\mathcal{V}'_i| \in \{1, 2\}$ for $1 \leq i \leq r$. This implies

$$|\mathcal{F}'| \leq 2^{k-r-1} = 2^{\lceil \log k \rceil - 1} \leq k - 1.$$

This proves the claim.

Now, applying the same argument for the cases $k - \lceil \log k \rceil < r \leq k - 1$, one can easily see that $|\mathcal{F}'| \leq 2^{k-r-1} < 2^{\lceil \log k \rceil - 1} \leq k - 1$. Hence, the theorem. □

3 2-uniform case

For 2−uniform (graph) CBCs, we provide the following improvements over existing results.

†Here $\mathcal{H}_r(k)$, as defined in (2), can as well be used as forbidden family.

4 An Exact Result

Theorem 4.1. *$m(n,2,5)=\lfloor\frac{n^2}{4}\rfloor$ for $n\geq 5$.*

Proof. We first remark that in [18], Paterson *et al.* observed that the complete bipartite graph on n vertices satisfies the batch condition for $k=5$, and hence concluded that $m(n,2,5)\geq\lfloor\frac{n^2}{4}\rfloor$. Hence it suffices to prove that $m(n,2,5)\leq\lfloor\frac{n^2}{4}\rfloor$. To be more precise, since this is an exact result, here we need to show that $m(n,2,5)=\mathrm{ex}^*(n,\mathcal{G}_2(5))\leq\lfloor\frac{n^2}{4}\rfloor$, where $\mathcal{G}_2(5)$ is the forbidden family of multigraphs, i.e., graphs with repeated edges, defined according to (1). So, we need to show that any multigraph with n vertices and at least $\lfloor\frac{n^2}{4}\rfloor+1$ edges contains a sub-multigraph with 4 vertices and 5 edges.

We prove this by induction on n. This is clearly true for $n=4$. Now suppose we have a multigraph with n vertices satisfying the given condition. We may assume that the multigraph has exactly $\lfloor\frac{n^2}{4}\rfloor+1$ edges by throwing any extra edges away. Observe that the multigraph contains a vertex of degree at most $\lfloor\frac{n}{2}\rfloor$. Removing this vertex along with all its incident edges leaves a multigraph with $n-1$ vertices and at least $\lfloor\frac{(n-1)^2}{4}\rfloor+1$ edges, which by the induction hypothesis contains a sub-multigraph with 4 vertices and 5 edges. □

5 Improvement of the lower bound

For improvement of the lower bound we need the following simple lemma which also appears as an exercise in [3]. We include the proof here for the sake of completeness.

Lemma 5.1. *Let $k\geq 6$ be a positive integer. If a graph has k edges and at most $k-1$ vertices then it has girth at most $\lfloor\frac{2k}{3}\rfloor$. This bound is tight.*

Proof. For $k=6$ the statement is clear. Suppose the statement does not hold for some $k>6$. Pick k minimum so that the statement does not hold, i.e., suppose that for some k we have a graph G with k edges, and at most $k-1$ vertices, and the girth of G is at least $\geq\lfloor\frac{2k}{3}\rfloor+1$. We may assume that the graph is connected. Since G has k edges and at most $k-1$ vertices, it contains at least 2 distinct cycles, C_1,C_2, say. Now, if C_1 and C_2 are edge disjoint then one of them will have length at most $\lfloor\frac{k}{2}\rfloor<\lfloor\frac{2k}{3}\rfloor$ which contradicts the assumption on the girth of G.

If C_1 and C_2 are not edge disjoint let $l_0=|E(C_1)\cap E(C_2)|$ be the number of common edges and l_1,l_2 be the number of edges that exclusively belonging to the C_1, and C_2 respectively. Clearly, $l_0+l_1=|E(C_1)|$, and $l_0+l_2=|E(C_2)|$. Now, consider the subgraph of G consisting of the edges of $C_1\Delta C_2$; here Δ refers to the symmetric difference of the corresponding edge sets. Every vertex in this subgraph has even degree, so in particular it contains a cycle. By the assumption on the girth of G, we have

$$l_i+l_j\geq\lfloor\frac{2k}{3}\rfloor+1,$$

for $0\leq i<j\leq 2$, which gives

$$l_0+l_1+l_2>k,$$

a contradiction.

This bound is best possible since we may consider two vertices joined by three pairwise vertex-disjoint paths of length $\frac{k}{3}$ (for k a multiple of 3) to give a tight example. □

Next we mention the following result of Lazebnik *et al.* and apply it to improve the lower bound 10 on $m(n,2,k)$ of [18], which in this case is $\Omega(n^{\frac{k+1}{k-1}})$, for all k and infinitely many values of n.

Theorem 5.2 ([17])**.** *For $s \geq 2$, $\mathrm{ex}(n, \{C_3, C_4, \ldots, C_{2s+1}\}) = \Omega(n^{1+\frac{2}{3s-3+\epsilon}})$ for infinitely many values of n, where $\epsilon = 0$ if s is odd, and $= 1$ if s is even.*

Corollary 5.3. *Let $k \geq 8$ then*

$$m(n,2,k) = \begin{cases} \Omega(n^{\frac{k-3}{k-5}}) & \text{if } k = 5 \mod 6 \\ \Omega(n^{\frac{k-2}{k-4}}) & \text{if } k = 2 \mod 6 \text{ or } k = 4 \mod 6 \\ \Omega(n^{\frac{k-1}{k-3}}) & \text{if } k = 1 \mod 6 \text{ or } k = 3 \mod 6 \\ \Omega(n^{\frac{k}{k-2}}) & \text{if } k = 0 \mod 6 \end{cases}$$

for infinitely many values of n.

Proof. The proof follows directly from Lemma 5.1 and Theorem 5.2. We remark here that for the cases where $k = 0 \mod 3$ or $k = 1 \mod 3$ the bounds of the corollary may be improved as Lemma 5.1 requires the girth to be $\lfloor \frac{2k}{3} \rfloor + 1$, which is odd in these cases. Whereas the bounds obtained were by applying Theorem 5.2 for graphs of girth $\lfloor \frac{2k}{3} \rfloor + 2$.

6 Improvement of the upper bound

Next, we improve the upper bound on $m(n, 2, k)$. We begin with the following theorem due to Bondy and Simonovits.

Theorem 6.1. *([4]) If in a graph of order n, the number of edges $> 100kn^{1+\frac{1}{k}}$, then the graph contains a C_{2l} for every $l \in [k, kn^{\frac{1}{k}}]$.*

Remark 6.1.1. There have been improvements (cf. [21], [20]) in the constant term ($100k$) of this important theorem. However, since we do not require best of the constants for our result, we will use the above theorem only.

Theorem 6.2. *For $k \geq 4, m(n, 2, k) = O(n^{1+\beta})$, where $\beta = \frac{1}{\lfloor \frac{k}{4} \rfloor}$.*

Proof. Here we show that $m(n, 2, k) \leq 100kn^{1+\beta}$. Consider a graph G with $100kn^{1+\beta}$ edges. Now, it is well known that any graph has a subgraph whose minimum degree is at least half of the average degree of the original graph. Hence, G has a subgraph H with minimum vertex degree at least $100kn^{\beta}$.

Now, Theorem 6.1 implies that H has a cycle C of length at most $2\lfloor \frac{k}{4} \rfloor$. Let $v \in C$ be an arbitrary vertex, and consider all the walks of length $\lfloor \frac{k}{4} \rfloor$ in H starting at v in which no edge repeats consecutively in the walk. It is easy to see that the number of such walks is

$$100kn^{\beta}(100kn^{\beta} - 1)^{\lfloor \frac{k}{4} \rfloor - 1} > n.$$

Consequently, there is a vertex v', such that at least two distinct walks of length $\lfloor \frac{k}{4} \rfloor$ starting at v, terminate at v'. These two walks along with C give rise to a forbidden configuration consisting of at most k edges spanning at most $k - 1$ vertices. □

Theorem 6.2 leads to trivial upper bound $O(n^2)$ on $m(n, 2, k)$ for $k = 6, 7$. However, it turns out that we can improve this trivial upper bound on $m(n, 2, k)$for $k = 6, 7$ by the following well-known theorem due to Kővári *et al.*.

Theorem 6.3 ([16], see also [2])**.** *Suppose* $2 \le s,\ 2 \le t$. *Then*

$$\mathrm{ex}(n, K(s,t)) \le \frac{1}{2}(s-1)^{\frac{1}{t}}(n-t+1)n^{1-\frac{1}{t}} + \frac{1}{2}(t-1)n.$$

Corollary 6.4. $m(n,2,k) = \Theta(n^{\frac{3}{2}})$ *for* $k = 6,7,8$.

Proof.

(a) Theorem 6.3 clearly implies an upper bound of $O(n^{\frac{3}{2}})$ for $\mathrm{ex}(n, K(s,2))$. More precisely, it implies there is a constant $c_{s,2}$ such that for all $n > n_0$, any graph of order n with more than $c_{s,2}n^{\frac{3}{2}}$ edges will contain a $K(s,2)$. Now, $s = \lceil \frac{k}{2} \rceil, (k \ge 6)$ serves our purpose. Because $K(\lceil \frac{k}{2} \rceil, 2)$ has $\ge k$ edges and $\le k-1$ vertices for $k \ge 6$, and hence contains the forbidden structure of k edges spanning $\le k-1$ vertices as subgraph. This implies $m(n,2,k) = O(n^{\frac{3}{2}})$ for $k \ge 6$.

To prove the asymptotic tightness of the upper bound for the cases $k = 6,7,8$, we note that Lemma 5.1 implies a graph which is $\{C_3, C_4, C_5\}$-free, gives rise to 2-uniform CBC with retrievability parameter k for $k \le 8$. Now, it is well-known that for q a prime power, the incidence graph of $PG(2,q)$ is a $q+1$-regular bipartite graph with $2(q^2+q+1)$ vertices and girth 6. In fact, and it was shown in [14], for sufficiently large n this construction leads to a graph on n vertices having $\Omega(n^{\frac{3}{2}})$ edges whose girth is 6. Incidentally, this construction also improves on the non-constructive lower bound of (10) for $m(n,2,7)$ and $m(n,2,8)$. So, finally we have $m(n,2,k) = \Omega(n^{\frac{3}{2}})$ for $k = 6,7,8$.

□

7 Concluding Remarks

While the draft was under preparation, the authors were informed by Cs. Bujtás and Zs. Tuza [12] about their simultaneous and independent work in similar direction. The family of forbidden configurations they consider is the following.

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

where $r \ge 2, q \ge -r+1, k \ge q+r+1$ are fixed integers, and $|V(H)|$ and $|E(H)|$ denote respectively the number of vertices and edges of the hypergraph H.

Their results improve on the upper bound on $m(n,r,k)$ given in Theorem 2.2 for some ranges of values of r. Further, their results show that $m(n,2,k) = O(n^{1+\frac{1}{\lfloor \frac{k}{3} \rfloor}})$ which improves on our upper bound of $O(n^{1+\frac{1}{\lfloor \frac{k}{4} \rfloor}})$, and also implies the following asymptotically exact results.

(i) $m(n,2,k) = \Theta(n^{\frac{4}{3}})$, for $k = 9,10,11$. This follows from Corollary 5.3.

(ii) $m(n,2,k) = \Theta(n^{\frac{6}{5}})$, for $k = 15,16,17$. This follows by considering the incidence graph of finite generalized hexagon (see [19]) of order q, where q is a prime power; the graph is a $(q+1)$-regular bipartite graph of girth 12 and has partite sets of size $q^5+q^4+q^3+q^2+q+1$. This together with Lemma 5.1 implies the result.

Acknowledgement: The authors would like to thank Prof. Cs. Bujtás and Prof. Zs. Tuza for some comments on an earlier draft of this manuscript and later sharing information about their work. The second author would also like to thank Prof. N. M. Singhi and Dr. A. Bhattacharya of TIFR, Mumbai for helpful discussions on this problem. Lastly, and by no means least, the authors thank the anonymous reviewers for their useful comments and suggestions to improve the overall quality of the paper.

References

[1] S. Bhattacharya, S. Ruj and B. Roy, "Combinatorial Batch Codes: A Lower Bound and Optimal Constructions", *Advances in Mathematics of Communications,* vol. 6, pp. 165–174, 2012.

[2] B. Bollobás "*Extremal Graph Theory*", Dover, New York, 1978.

[3] B. Bollobás "*Modern Graph theory*", Graduate Texts in Mathematics 184, Springer. New York, 1998, pp. 136.

[4] J. A. Bondy and M. Simonovits "Cycles of Even Length in Graphs", *Journal of Combinatorial Theory(B),* vol. 16, pp. 97–105, 1974.

[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 conference on graph theory at Ann Arbor, pp. 53–63, 1973.

[6] R. A. Brualdi, K. P. Kiernan, S. A. Meyer and M. W. Schroeder, "Combinatorial Batch Codes and Transversal Matroids", *Advances in Mathematics of Communications,* vol. 4, pp. 419–431, 2010.

[7] R. A. Brualdi, K. P. Kiernan, S. A. Meyer and M. W. Schroeder, "Erratum to Combinatorial Batch Codes and Transversal Matriods", *Advances in Mathematics of Communications,* vol. 4, pp. 597–597, 2010. *Advances in Mathematics of Communications,* To appear.

[8] Cs. Bujtás and Zs. Tuza, "Optimal combinatorial batch codes derived from dual systems", *Miskolc Mathematical Notes,* vol. 12, pp. 11–23, 2011.

[9] Cs. Bujtás and Zs. Tuza, "Optimal batch codes: Many items or low retrieval requirement", *Advances in Mathematics of Communications,* vol. 5, pp. 529–541, 2011.

[10] Cs. Bujtás and Zs. Tuza, "Combinatorial batch codes: Extremal problems under Hall-type conditions", *Electronic Notes in Discrete Mathematics,* vol. 38, pp. 201–206, 2011.

[11] Cs. Bujtás and Zs. Tuza, "Relaxations of Hall's condition: Optimal batch codes with multiple queries", *Appl. Anal. Discrete Math.* vol. 6, pp. 72–81, 2012.

[12] Cs. Bujtás and Zs. Tuza, "Turán numbers and batch codes", *manuscript*, 2012.

[13] P. Erdős, "On extremal problems of graphs and generalized graphs", *Israel Journal of Mathematics,* vol. 2, pp. 183–190, 1964.

[14] P. Erdős, A. Rényi, and V. T. Sós, "On a Problem of Graph Theory", *Stud. Sci. Math. Hung.,* vol. 1, pp. 215–235, 1966.

[15] Y. Ishai, E. Kushilevitz, R. Ostrovsky and A. Sahai, "Batch codes and their applications", *Proceedings of the thirty-sixth annual ACM symposium on Theory of computing,* vol. 36, pp. 262–271, 2004.

[16] P. Kővári,V. T. Sós, and P. Turán, " On a problem of K. Zarankiewicz", *Colloquium Mathematicum,* vol. 3, pp. 50–57, 1954.

[17] F. Lazebnik, V.A. Ustimenko and A.J. Woldar "A New Series of Dense Graphs of High Girth" "*Bulletin of the AMS*, Volume 32, Number 1, pp. 73–79, 1995

[18] M. B. Paterson, D. R. Stinson and R. Wei, "Combinatorial Batch Codes", *Advances in Mathematics of Communications,* vol. 3, pp. 13–27, 2009.

[19] S. E. Payne and J. A. Thas, "*Finite Generalized Quadrangles*", EMS Series of Lectures in Mathematics, European Mathematical Society, 2009, pp. 36.

[20] O. Pikhurko, "A Note on the Turán Function of Even Cycles", *Proceedings of the American Mathematical Society*, 140 (2012) 3687-3992.

[21] J. Verstraëte, "On arithmetic progressions of cycle lengths in graphs", *Combinatorics, Probability and Computing*, 9 (2000), no. 4, 369-373.