Katarzyna Mieczkowska

Uniwersytet A. Mickiewicza w Poznaniu

On Erdős' extremal problem on matchings in hypergraphs

Praca semestralna nr 2

(semestr zimowy 2011/12)

Opiekun pracy: Tomasz Łuczak

ON ERDŐS' EXTREMAL PROBLEM ON MATCHINGS IN HYPERGRAPHS

TOMASZ ŁUCZAK AND KATARZYNA MIECZKOWSKA

Abstract. In 1965 Erdős conjectured that the number of edges in k-uniform hypergraphs on n vertices in which the largest matching has s edges is maximized for hypergraphs of one of two special types. We settled this conjecture in the affirmative for $k = 3$ and n is large enough.

1. Introduction

A *k-uniform hypergraph* or, briefly, a *k-graph* $G = (V, E)$ is a set of vertices $V \subseteq \mathbb{N}$ together with a family E of k-element subsets of V, which are called edges. We denote by $v(G) = |V|$ and $e(G) = |E|$ the number of vertices and edges of $G = (V, E)$, respectively. A family of disjoint edges of G is a *matching*, and by $\mu(G)$ we mean the size of the largest matching in G. In this paper we deal with the problem of maximizing $e(G)$ given $v(G)$ and $\mu(G)$. More formally, let $\mathcal{H}_k(n, s)$ denote the set of all k-graphs $G = (V, E)$ such that $|V| = n$ and $\mu(G) = s$; moreover let

$$\mu_k(n, s) = \max\{e(G) : G \in \mathcal{H}_k(n, s)\}, \tag{1}$$

and

$$\mathcal{M}_k(n, s) = \{G \in \mathcal{H}_k(n, s) : e(G) = \mu_k(n, s)\}. \tag{2}$$

Let us describe two kinds of k-graphs from $\mathcal{H}_k(n, s)$ which are natural candidates for members of $\mathcal{M}_k(n, s)$. By $\mathrm{Cov}_k(n, s)$ we denote the family of k-graphs $G_1 = (V_1, E_1)$ such that $|V_1| = n$ and for some subset $S \subseteq V_1$, $|S| = s$, we have

$$E_1 = \{e \subseteq V_1 : e \cap S \neq \emptyset \text{ and } |e| = k\}.$$

Clearly, if $s \leq n/k$, then $\mathrm{Cov}_k(n, s) \subseteq \mathcal{H}_k(n, s)$. Furthermore, we define $\mathrm{Cl}_k(n, s)$ as the family of all k-graphs $G_2 = (V_2, E_2)$ which consists of

---

Date: February 16, 2012.
1991 *Mathematics Subject Classification.* 05C35, 05C65, 05C70.
Key words and phrases. extremal graph theory, matching, hypergraphs.
The first author partially supported by the Foundation for Polish Science.

a complete subgraph on $ks+k-1$ and some isolated vertices, i.e. if for some subset $T \subseteq V_2$, $|T| = ks+k-1$, we have

$$E_2 = \{e \subseteq T : \ |e| = k\}.$$

Again, we have $\mathrm{Cl}_k(n,s) \subseteq \mathcal{H}_k(n,s)$. In 1965 Erdős [4] conjectured that, indeed, the function $\mu_k(n,s)$ is fully determined by k-graphs of these two types, namely that for every k, n and s, where $ks \le n-k+1$, the following holds

$$\mu_k(n,s) = \max\left\{\binom{n}{k} - \binom{n-s}{k}, \binom{sk+k-1}{k}\right\}. \tag{3}$$

Although the conjecture remains widely open a few results have been proved in this direction (cf. Frankl [7]). Most of them are dealing with the case when n is large compared to s, proving that

$$\mathcal{M}_k(n,s) = \mathrm{Cov}_k(n,s) \quad \text{for} \quad n \ge g(k)s, \tag{4}$$

where $g(k)$ is some function of k. The best published bound for $g(k)$ for general k is due to Bollobás, Daykin and Erdős [3] who showed that (4) holds whenever $g(k) \ge 2k^3$; recently, Huang, Loh, and Sudakov [9] announced that (4) remains true for $g(k) \ge 3k^2$. As for the special case of $k=3$ the current record belongs to Frankl, Rödl and Ruciński [8] who verifed (4) for $k=3$ and $n \ge 4s$.

The main result of this paper states that for $k=3$ and n large enough (3) holds for every s and, moreover, the only extremal 3-graphs belong to either $\mathrm{Cov}_k(n,s)$ or $\mathrm{Cl}_k(n,s)$.

Theorem 1. *There exists n_0 such that for $n \ge n_0$ large enough and each s, $1 \le s \le (n-2)/3$, we have*

$$\mu_3(n,s) = \max\left\{\binom{n}{3} - \binom{n-s}{3}, \binom{3s+2}{3}\right\}. \tag{5}$$

Furthermore, for such parameters n and s, we have

$$\mathcal{M}_3(n,s) \subseteq Cov_3(n,s) \cup Cl_3(n,s).$$

Let us remark that although we have made no effort to get effective bounds for n_0, it seems to be of rather moderate order and it is quite conceivable that a meticulous analysis of cases (possibly, with some help of computer) can give (5) for all values of n. Note however that the second part of the statement does not hold when $n=6$ and $s=1$ (or, for general k-graphs, for $n=2k$, $k \ge 3$, and $s=1$). Indeed, in this case

$$|\mathcal{M}_k(2k,1)| = 2^{\frac{1}{2}\binom{2k}{k}},$$

while

$$|\mathrm{Cov}_k(2k,1)| = |\mathrm{Cl}_k(2k,1)| = 2k\,.$$

The structure of the paper goes as follows. First we show that if the structure of a large graph from $\mathcal{M}(n,s)$ is 'close' to a graph from $\mathrm{Cov}_k(n,s)$, then it belongs to $\mathrm{Cov}_k(n,s)$, and the same remains true for $\mathrm{Cl}_k(n,s)$. Thus, an 'asymptotic version' of Theorem 1 implies that it holds in its exact form, provided n is large enough. Then, we recall the definition and basic properties of the shift operation which is another important ingredient of our argument. Finally, in the last part of the paper, we concentrate on the case $k=3$ and show that then the required asymptotic result indeed holds for shifted 3-graphs.

2. Stability of Cov and Cl

The aim of this section is to show that if a k-graph $G \in \mathcal{M}_k(n,s)$ is, in such a way, similar to graphs from $\mathrm{Cov}_k(n,s)$ [or $\mathrm{Cl}_k(n,s)$], then in fact it belongs to this family. In order to make it precise let us introduce families of graphs $\mathrm{Cov}_k(n,s;\varepsilon)$ and $\mathrm{Cl}_k(n,s;\varepsilon)$. Let us recall that if $G=(V,E)$ belongs to $\mathrm{Cov}_k(n,s)$, then there exists a set $S \subseteq V$, $|S|=s$, which covers all edges of G. We say that $G \in \mathrm{Cov}_k(n,s;\varepsilon)$ for some $\varepsilon > 0$, if there exists a set $S \subseteq V$, $|S|=s$, which covers all but at most $\varepsilon|E|$ edges of G. Moreover, we define $\mathrm{Cl}_k(n,s;\varepsilon)$ as the set of all k-graphs G which contain a complete subgraph on at least $(1-\varepsilon)ks$ vertices. Then the main result of this section can be stated as follows.

Lemma 2. *For every $k \geq 3$ there exist $\varepsilon > 0$ and n_0 such that for every $n \geq n_0$, $1 \leq s \leq n/k$, and $G \in \mathcal{M}_k(n,s)$ the following holds:*

(i) *if $G \in Cov_k(n,s;\varepsilon)$, then $G \in Cov_k(n,s)$;*
(ii) *if $G \in Cl_k(n,s;\varepsilon)$, then $G \in Cl_k(n,s)$.*

Before we prove the lemma let us comment briefly on the formula (3). If by $s_0(n,k)$ we define the smallest s for which

$$\binom{n}{k} - \binom{n-s}{k} \leq \binom{ks+k-1}{k},$$

then it is easy to see that

$$\lim_{n\to\infty} \frac{s_0(n,k)}{n} = \alpha_k\,,$$

where $\alpha_k \in (0,1)$ is the solution of the equation

$$1-(1-\alpha_k)^k = k^k\alpha^k\,.$$

One can check that for all $k \geq 3$ we have

$$\frac{1}{k} - \frac{1}{2k^2} < \alpha_k < \frac{1}{k} - \frac{2}{5k^2}; \tag{6}$$

in fact, $(1 - k\alpha_k)k \to -\ln(1 - e^{-1}) = 0.4586...$ as $k \to \infty$.

Proof of Lemma 2. In order to show (i) let us start with the following observation.

Claim 1. *If $G \in \mathcal{M}_k(n, s)$ contains a vertex v which is contained in more than $\binom{n}{k-1} - \binom{n-ks-1}{k-1}$ edges of $G = (V, E)$, then v belongs to $\binom{n-1}{k-1}$ edges of G.*

Proof. Take a vertex v of large degree, and let us suppose that e is a k-subset of V such that $v \in e$ and $e \notin E$. Then, by the definition of $\mathcal{M}_k(n, s)$, the graph $G \cup e$ contains a matching M of size $s + 1$, where, clearly, $e \in M$. However, since the degree of v is large, there exists a $(k - 1)$-element subset $f \subseteq V \setminus \bigcup M$ such that $e' = \{v\} \cup f$ is an edge of G. But then, $M' = M \setminus \{e\} \cup \{e'\}$ is a matching of size $s + 1$ in G. This contradiction shows that each k-element subset of V which contains v is an edge of G. $\square$

Now we prove (i). Let us assume that $G = (V, E) \in \mathcal{M}_k(n, s)$ belongs to $\mathrm{Cov}_k(n, s; \varepsilon)$ and let S be the set which covers all but at most $\varepsilon|E|$ edges of G. Let $T \subseteq S$ be the set of vertices which are not contained in $\binom{n-1}{k-1}$ edges of G and let $t = |T|$. We need to show that $t = 0$.

Observe first that, because of (6), we may and shall assume that $s \leq n(1/k - 2/(5k^2))$, since otherwise there exists a k-graph $G' \in \mathrm{Cl}_k(n, s)$ with more edges than G, contradicting the fact that $G \in \mathcal{M}_k(n, s)$. Thus, by Claim 1, the number of edges in which each vertex $v \in T$ is contained in is at most

$$\binom{n}{k-1} - \binom{n-ks-1}{k-1} \leq \Big(1 - \Big(\frac{2}{5k}\Big)^{k-1}\Big)\binom{n}{k-1}.$$

Now let $\bar{G}$ denote the k-graph obtained from G by deleting all vertices from $S \setminus T$ and all edges intersecting with them. It is easy to see that $\bar{G} \in \mathcal{M}_k(n-s+t, t)$. Now, if $t \geq n/(10k^5) \geq s/(10k^4)$, for any k-graph $\hat{G} \in \mathrm{Cov}_k(n - s + t, t)$ we have

$$\begin{aligned} e(\hat{G}) - e(\bar{G}) \geq& \frac{t}{k}\Big(\frac{2}{5k}\Big)^{k-1}\binom{n}{k-1} - 2\varepsilon s\binom{n}{k-1} \\ \geq& \Big(\Big(\frac{2}{5k}\Big)^{k-1} - 20k^5\varepsilon\Big)\frac{t}{k}\binom{n}{k-1}. \end{aligned}$$

Thus, if $\varepsilon > 0$ is small enough than $\hat{G}$ has more edges than $\bar{G}$ contradicting the fact that $\bar{G} \in \mathcal{M}_k(n-s+t,t)$. Thus, $t \le n/(10k^5) \le (n-s+t)/k^3$. But in such a case, Theorem 1 holds by the result of Bollobás, Daykin, and Erdős (see (4) above), so

$$\bar{G} \in \mathcal{M}_k(n-s+t) = \mathrm{Cov}_k(n-s+t,t)$$

and, since by the definition no vertex of T has a full degree, $t = 0$. Consequently, $G \in Cov_k(n,s)$ and (i) follows.

Now let assume that $G = (V,E) \in \mathcal{M}_k(n,s)$ belongs to $Cl_k(n,s;\varepsilon)$. Let U be the set of vertices of the largest complete k-subgraph of G such that $|U| \ge ks(1-\varepsilon)$. Furthermore, let M be a matching in G of size s which maximizes $|\bigcup M \cup U|$, and $M' = \{e \in M \,:\, e \not\subseteq U\}$. Then, for n large enough, the following holds.

Claim 2.

(i) $|\bigcup M \cup U| = ks + k - 1$.
(ii) $|M'| \le 2\varepsilon ks$.
(iii) *each edge of G either is contained in U or intersects an edge of M'.*

Proof. Observe that at most $k-1$ vertices of U can remain unsaturated by M, thus $|\bigcup M \cup U| \le ks + k - 1$. On the other hand, since U induces the largest clique in G, there exists a k-element subset $e \notin E$ such that $|e \cap U| = k-1$. Then, since $G \in \mathcal{M}_k(n,s)$, the graph $G \cup \{e\}$ contains a matching $M^* \cup \{e\}$ of size $s+1$. Thus, M^* is a matching of size s, in which precisely $k-1$ vertices from U are unsaturated, so $|\bigcup M \cup U| \ge |\bigcup M^* \cup U| \ge ks + k - 1$, and (i) follows. To prove (ii) observe that $|M'| \le |V(M') \setminus U| = |U \cup \bigcup M| - |U|$ and use (i), obtaining $|M'| \le \varepsilon ks + k - 1 \le 2\varepsilon ks$ for n big enough. Finally, (iii) is a direct consequence of the choice of M. □

Let $G' = (V,E')$ denote k-graph which consists of the clique with vertex set $\bigcup M \cup U$ and isolated vertices. Clearly, the size of the largest matching in G' is s. We shall show that G' has more edges than G provided $|M'| > 0$. Thus, we must have $M' = \emptyset$ and the assertion follows.

In order to show that $e(G') > e(G)$ we need to introduce one more hypergraph. Let $H = (V \setminus U, F)$ be the hypergraph with the edge set

$$F = \{e \cap (V \setminus U) \,:\, e \in E\}\,.$$

Note that H is not a k-graph but each of its edges has size between 1 and k. We call an edge $f \in F$ with ℓ elements *thick* if it is contained in more than $3\varepsilon k^2\binom{|U|}{k-\ell}$ edges of G, contained entirely in $U \cup f$, and *thin* otherwise. Let us make an observation somewhat analogous to Claim 1.

Claim 3. *If an edge $f \in F$ of ℓ-elements is thick, then each k-element subset of $U \cup f$ containing f is an edge of G.*

Proof. Let us suppose that for thick f there exists an k-element set e such that $f \subseteq e \subseteq U \cup f$ and $e \notin E$. Then, since $G \in \mathcal{M}_k(n,s)$, the graph $G \cup e$ contains a matching M'' of size $s+1$, where $e \in M''$. Furthermore, at most $2\varepsilon k^2 s\binom{|U|}{k-l-1} \leq 3\varepsilon k^2\binom{|U|}{k-l}$ of $(k-\ell)$-elements subsets of U are covered by sets from M'' not contained in $U \cup f$. Since f is thick, there exists a $(k-\ell)$-subset h of U which is covered only by edges of M'' contained in U and such that $f \cup h \in E$. But then, one can modify $M'' \setminus \{e\} \cup \{f \cup h\}$, replacing edges of M'' which intersect h by the same number of disjoint edges contained in U, in such a way that the new set of edges is a matching of size $s+1$, contradicting the fact that $G \in \mathcal{M}_k(n,s)$. Hence, all edges e for which $f \subseteq e \subseteq U \cup f$ must already belong to G. □

Let us count edges in $|E' \setminus E|$. For every edge $e \in M'$ consider a vertex $v \in e \setminus U$. Note that G' contains all k-element sets $e \subseteq \{v\} \cup U$, such that $v \in e$. Furthermore, from Claim 3 and the fact that U is the vertex set of the largest clique, we infer that at most $3\varepsilon k^2\binom{|U|}{k-1}$ of these sets belong to G. Thus,

$$|E' \setminus E| \geq (1-3\varepsilon k^2)|M'|\binom{|U|}{k-1}. \tag{7}$$

Now we estimate the number of edges in $|E \setminus E'|$. Let us first bound the number γ of edges $e \in E \setminus E'$ such that $e \cap (V \setminus U) \neq \emptyset$ is thin. Since, as we have already mentioned, each such edge must intersect one of $2\varepsilon sk$ edges of M', we have

$$\gamma \leq |M'|k\sum_{r=0}^{k-1} 3\varepsilon k^2\binom{n}{r} \leq 3\varepsilon|M'|k^4\binom{n}{k-1}. \tag{8}$$

Finally, let us consider a hypergraph $H' = (W', F')$ such that $W' = (V \setminus U) \cup \bigcup M'$, and

$$F' = M' \cup \{e \cap W' : \ e \cap (V \setminus U) \text{ is thick}\}.$$

It is easy to see that if the largest matching in H' covers at least $k|M'|+1$ vertices than we can enlarge it to a matching in G of size $s+1$ using Claims 2(i) and 3. Furthermore, if ε is small enough,

$|M'| \le |W'|/(2k^3)$ so one can apply the result of Bollobás, Daykin, Erdős [3] (see 4) to infer that

$$\begin{aligned}|F'| &\le |M'|\left(\binom{n}{k-1} - \binom{|U|}{k-1}\right) + k|M'|\binom{n}{k-2}\\ &\le (1+3\varepsilon k^2)|M'|\left(\binom{n}{k-1} - \binom{|U|}{k-1}\right).\end{aligned} \tag{9}$$

Thus, from (7), (8), and (9), we get

$$\begin{aligned}e(G') - e(G) \ge (1-3\varepsilon k^2)|M'|\binom{|U|}{k-1} - 3\varepsilon|M'|k^4\binom{n}{k-1}\\ - (1+3\varepsilon k^2)|M'|\left(\binom{n}{k-1} - \binom{|U|}{k-1}\right)\\ \ge |M'|\left(2\binom{|U|}{k-1} - \binom{n}{k-1} - 4\varepsilon k^4\binom{n}{k-1}\right).\end{aligned}$$

Due to (6) we may assume that $|U|/n \ge 1 - 1/(2k)$ and so

$$\binom{|U|}{k-1} \ge 0.6\binom{n}{k-1}.$$

Consequently, for $|M'| > 0$ we have $e(G') > e(G)$ and the assertion follows. □

3. Shifted graphs

Let $G = (V, E)$, $V \subseteq \mathbb{N}$ be a k-graph. For vertices $i < j$, the graph $\mathbf{sh}_{ij}(G)$, called the (i, j)-shift of G, is obtained from G by replacing each edge $e \in E$, such that $j \in e$, $i \notin e$, and $f = e - \{j\} \cup \{i\} \notin E$, by f. The basic fact we shall use about $\mathbf{sh}_{ij}$ is that it acts nicely on families $\mathcal{M}_k(n, s)$, $\mathrm{Cov}_k(n, s)$ and $\mathrm{Cl}_k(n, s)$. Let us start with the following well known result (see Frankl [7]), the proof of which we give here for the completness of the argument.

Lemma 3. *For every i, j, $i < j$, if $G \in \mathcal{M}_k(n, s)$ then $\boldsymbol{sh}_{ij}(G) \in \mathcal{M}_k(n, s)$.*

Proof. Let us first observe that the shift operation can only decrease the size of the largest matching. Indeed, let us assume that $M = \{e_1, \ldots, e_\ell\}$ is a matching in $\mathbf{sh}_{ij}(G)$ but not in G, and let $i \in e_1$. Then either $j \notin \bigcup_r e_r$ and so $M' = \{e_1 - \{i\} \cup \{j\}, e_2, \ldots, e_\ell\}$ is a matching in G, or $j \in e_2$ and then $M'' = \{e_1 - \{i\} \cup \{j\}, e_2 - \{j\} \cup \{i\}, e_3, \ldots, e_\ell\}$ is a matching in G. Hence $\mu(\mathbf{sh}_{ij}(G)) \le \mu(G)$ but since $G \in \mathcal{M}_k(n, s)$ we have also $\mu(\mathbf{sh}_{ij}(G)) = \mu(G)$. □

The following simple observation will be useful in our further argument.

Fact 4. *Let $n \geq 2k-1$. If we color all $(k-1)$-element subsets of $\{1, 2, \dots, n\}$ with two colors, then either we find two disjoint sets colored with different colors or all the sets are of the same color.* □

In order to characterize the extremal graphs in $\mathcal{M}_k(n, s)$ we shall use the following observation.

Lemma 5. *Let $n \neq 2k$, $G \in \mathcal{M}_k(n, s)$, and $i < j$.*

(i) *If $\boldsymbol{sh}_{ij}(G) \in Cov_k(n, s)$, then $G \in Cov_k(n, s)$.*
(ii) *If $\boldsymbol{sh}_{ij}(G) \in Cl_k(n, s)$, then $G \in Cl_k(n, s)$.*

Proof. Let us remark first that if $n \leq sk + k - 1$, then the only graph in $\mathcal{M}_k(n, s)$ is the complete graph, and for $s = 1$ and $n \geq 2k + 1$ we have $\mathcal{M}_k(n, 1) = \mathrm{Cov}_k(n, 1)$ by the extremal version of Erdős-Ko-Rado theorem (cf. [3]), so we may assume that $s \geq 2$ and $n \geq 2k+1$. Thus, let $G = (V, E) \in \mathcal{M}_k(n, s)$, $\mathbf{sh}_{ij}(G) \in \mathrm{Cov}_k(n, s)$ and let S be the set which covers all edges of $\mathbf{sh}_{ij}(G)$. Clearly, $i \in S$. If $j \in S$ then $G = \mathbf{sh}_{ij}(G)$ so let us assume that $j \notin S$. Let us color all $(k-1)$-element subsets f of $V \setminus \{i, j\}$ into two colors: red if $\{i\} \cup f \in E$ and blue if $\{j\} \cup f \in E$. Since S covers all k-element subsets of V in $\mathbf{sh}_{ij}(G)$, each such $(k-1)$-element subsets is colored with exactly one color. Furthermore, if for a pair of disjoint subsets f' and f'', f' is red and f'' is blue, then the edges $\{i\} \cup f'$ and $\{j\} \cup f''$ can be completed to a matching of size $s+1$, contradicting the fact that $G \in \mathcal{M}_k(n, s)$. Thus, by Fact 4, all such sets are colored with one color and either S or $S - \{i\} \cup \{j\}$ is covering all edges of G, i.e. $G \in \mathrm{Cov}_k(n, s)$.

The proof of (ii) is very similar. We take a clique T in $\mathbf{sh}_{ij}(G)$, $|T| = ks + k - 1$, and observe that the only interesting case is when $i \in T$ and $j \notin T$. Then we color all $(k-1)$-subsets of T with two colors and use Fact 4 to argue that either T or $T - \{i\} \cup \{j\}$ is the clique in G. □

Now let us define $\mathbf{Sh}(G)$ as a graph which is obtained from G by the series of shifts and which is invariant under all possible shifts. Although we shall never use this fact it is worthy to remark that $\mathbf{Sh}(G)$ is uniquely determined, i.e. if we apply to G all possible shifts then the resulting graph does not depend on the order the operations (see [7]). Let us state now an immediate consequence of Lemmata 3 and 5 we use directly in our proof.

Lemma 6.

(i) *If $G \in \mathcal{M}_k(n, s)$ then $\boldsymbol{Sh}(G) \in \mathcal{M}_k(n, s)$.*

(ii) *If* $n \neq 2k$, $G \in \mathcal{M}_k(n,s)$, *and* $\boldsymbol{Sh}(G) \in Cov_k(n,s)$, *then* $G \in Cov_k(n,s)$.
(iii) *If* $n \neq 2k$, $G \in \mathcal{M}_k(n,s)$, *and* $\boldsymbol{Sh}(G) \in Cl_k(n,s)$, *then* $G \in Cl_k(n,s)$. □

4. Proof of Theorem 1

In this section we study the case when $k = 3$. The main result of this part of the paper can be stated as follows.

Lemma 7. *For every* $\varepsilon > 0$ *there exists* n_0 *such that for every* $n \geq n_0$, $1 \leq s \leq n/3$, *and* $G \in \mathcal{M}_3(n,s)$ *we have*

$$\boldsymbol{Sh}(G) \in Cov_3(n,s;\varepsilon) \cup Cl_3(n,s;\varepsilon)\,.$$

We shall show Lemma 7 by a detailed analysis of the structure of $\mathbf{Sh}(G)$ but before we do it let us observe that it implies Theorem 1.

Proof of Theorem 1. Let $G \in \mathcal{M}_3(n,s)$. Then, by Lemma 6(i), $\mathbf{Sh}(G) \in \mathcal{M}_3(n,s)$. Thus, from Lemmata 2 and 7, for n large enough we get

$$\mathbf{Sh}(G) \in \mathrm{Cov}_3(n,s) \cup \mathrm{Cl}_3(n,s)\,,$$

and so, by Lemma 6(ii),(iii)

$$G \in \mathrm{Cov}_3(n,s) \cup \mathrm{Cl}_3(n,s)\,. \quad \square$$

Let us remark that in order to show Theorem 1 it is enough to show Lemma 7 for some given $\varepsilon > 0$.

Proof of Lemma 7. Let $\varepsilon > 0$ and $G \in \mathcal{M}_3(n,s)$. By Lemma 6(i), $\mathbf{Sh}(G) \in \mathcal{M}_3(n,s)$. To simplify the notation, by writing (i,j,k) we always mean that an edge $\{i,j,k\}$ is such that $i < j < k$. Let $M = \{(i_l, j_l, k_l) : l = 1,\dots,s\}$ be the largest matching in $\mathbf{Sh}(G)$, and let partition its vertex set into three parts $V(M) = I \cup J \cup K$ such that for every edge $(i,j,k) \in M$ we have $i \in I$, $j \in J$, and $k \in K$. Moreover, let vertices of K be labeled in such a way that $k_l < k_m$ for every $l < m$, and denote $L = \{i_l, j_l, k_l \in V(M) : l \leq (1-\varepsilon)s\}$. We shall show that for n large enough either I covers all but at most $\epsilon|E|$ edges of $\mathbf{Sh}(G)$ or $\{e \in \mathbf{Sh}(G) : e \subset L\}$ is a clique.

In order to study the structure of $\mathbf{Sh}(G)$ we introduce an auxiliary hypergraph H. Denote by V' the set of vertices which are not saturated by M. Obviously, none of the edges of $\mathbf{Sh}(G)$ is contained in V'. We use $\deg_{V'}(v)$ to denote the number of pairs $u, w \in V'$ such that $\{v,u,w\}$ is an edge in $\mathbf{Sh}(G)$. Similarly, a number of vertices $w \in V'$ such that $\{v,u,w\} \in \mathbf{Sh}(G)$ is denoted $\deg_{V'}(v,u)$. Finally, we use $e(v)$ to denote an unique edge of M containing vertex v. Let $H = (W,F)$

be a hypergraph with vertices $W = V(M)$ and the edge set $F = M \cup F_1 \cup F_2 \cup F_3$, where

$$\begin{aligned} F_1 &= \{v \in W : \deg_{V'}(v) \geq 20n\}, \\ F_2 &= \{\{v,w\} \in W^{(2)} : e(v) \neq e(w) \text{ and } \deg_{V'}(v,w) \geq 20\}, \\ F_3 &= \{\{v,w,u\} \in W^{(3)} : e(v), e(w) \text{ and } e(u) \text{ are pairwise different}\}. \end{aligned}$$

Note that since $\mathbf{Sh}(G)$ is shifted, hypergraphs F_1, F_2, F_3 are shifted as well. We shall call an edge e of $\mathbf{Sh}(G)$ *traceable* if $e \cap V(M) \in F$, and *untraceable* otherwise. Observe also that the number of untraceable edges of $\mathbf{Sh}(G)$ is bounded from above by $31n^2$, so we can afford to ignore them.

We call a triple T of edges from M *bad*, if in $\bigcup T$ there are three disjoint edges of H whose union intersects I on at most 2 vertices, and *good* otherwise. We show first that there are only few bad triples in M.

Claim 4. *No three disjoint triples are bad.*

Consequently, there exist at most six edges in the matching M so that each bad triple contains one of these edges.

Proof. Let us suppose that there exist nine disjoint edges $\{(i_l, j_l, k_l) : l = 1, \ldots, 9\} \subset M$ such that in $\{i_l, j_l, k_l : l = 1, \ldots, 9\}$ one can find a set of nine disjoint edges $H' \subset H$, which do not cover vertices i_3, i_6 and i_9. One can easily see that for any ordering of the sets $\{j_3, j_6, j_9\}$ and $\{k_3, k_6, k_9\}$ there exists a permutation $\sigma(3)$, $\sigma(6)$, $\sigma(9)$ such that $j_{\sigma(9)} > j_{\sigma(6)}$ and $k_{\sigma(9)} > k_{\sigma(3)}$; to simplify the notation let us assume that $j_9 > j_6 > i_6$ and $k_9 > k_3 > i_3$. Replace in H' an edge e which contains j_9 by $e' = e \setminus \{j_9\} \cup \{i_6\}$ and the edge f containing k_9 by $f' = e \setminus \{k_9\} \cup \{i_3\}$; note that both e' and f' belong to H since $H = \mathbf{Sh}(H)$. Thus, we obtain the family of nine disjoint edges of $H'' \subseteq H$, all of which are contained in eight edges of M. Furthermore, since edges from $F_1 \cup F_2$ have large degrees, all edges from H'' which belong to $F_1 \cup F_2$ can be simultaneously extended to disjoint edges of $\mathbf{Sh}(G)$ by adding to them vertices from $V \setminus \bigcup M$. But this would lead to a matching M' of size $s+1$ in $\mathbf{Sh}(G)$ contradicting the assumption $\mathbf{Sh}(G) \in \mathcal{M}_3(n,s)$. □

Now we study properties of good triples. We start with the following simple observation.

Claim 5. *Let T be a good triple.*

(i) $(F_1 \cap \bigcup T) \subset I$.

(ii) *For any two edges of T there are at most 5 edges in F_2 contained in their vertex set.*

Moreover, the only possible configuration with exactly 5 *edges from* F_2 *is when all these edges intersect* I *(see Fig. 1).*

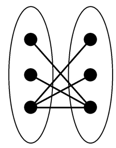

Fig. 1.

Proof. Let $T = \{(i_1, j_1, k_1), (i_2, j_2, k_2), (i_3, j_3, k_3)\}$ be a good triple.

(i) Let $j_1 < j_2 < j_3$ and assume that $(F_1 \cap \bigcup T) \not\subset I$. Then, since hypergraph F_1 is shifted, $\{j_1\} \in F_1$ and T is a bad triple because of the edges $\{j_1\}$, (i_2, j_2, k_2), (i_3, j_3, k_3), a contradiction.

(ii) Let assume by contradiction that 6 edges from F_2 are contained in $\{i_1, j_1, k_1, i_2, j_2, k_2\}$. Then $\{j_1, j_2\} \in F_2$ and at least one of the edges $\{i_1, k_2\}, \{i_2, k_1\}$ is in F_2. Let us assume that $\{i_1, k_2\} \in F_2$. Then, T is bad because of the edges $\{j_1, j_2\}, \{i_1, k_2\}, (i_3, j_3, k_3)$. □

For a triple $T \in M^{(3)}$ and for $i = 1, 2, 3$, let $f_i(T)$ be the number of edges of F_i contained in $\bigcup T$. Clearly, $f_1(T) \leq 9$, $f_2(T) \leq 27$ and $f_3(T) \leq 27$ for any triple T. However, if T is good, then, by Claim 5, we immediately infer that $f_1(T) \leq 3$ and $f_2(T) \leq 15$. Our next result shows how to bound $f_1(T)$ and $f_2(T)$ more precisely for good triples for which $f_3(T)$ is large.

Claim 6. *Let* T *be a good triple.*

(i) *If* $f_3(T) \geq 24$, *then* $f_1(T) = f_2(T) = 0$.
(ii) *If* $f_3(T) = 20$, *then* $f_1(T) \leq 1$ *and* $f_2(T) \leq 12$.
(iii) *If* $f_3(T) \leq 19$, *then* $f_1(T) \leq 3$ *and* $f_2(T) \leq 15$.
Moreover, the only triples for which $f_3(T) = 19$, $f_2(T) = 15$, *and* $f_1(T) = 3$, *are those in which each edge of* H *contained in* $\bigcup T$ *intersects* I.
(iv) *If* $f_3(T) = 21$, *then* $f_1(T) \leq 1$ *and* $f_2(T) \leq 10$.
(v) *If* $22 \leq f_3(T) \leq 23$, *then* $f_1(T) = 0$ *and* $f_2(T) \leq 7$.

Proof. Let $T = \{(i_1, j_1, k_1), (i_2, j_2, k_2), (i_3, j_3, k_3)\}$ be a good triple.

(i) Observe that since $f_3(T) \geq 24$, one of the following pairs of edges must be in H.

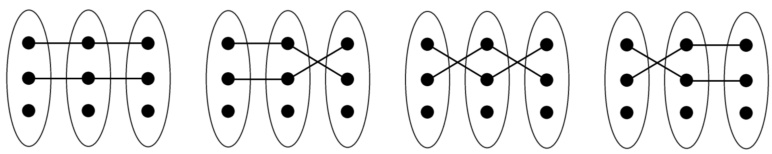

Fig. 2.

Let $e, f \in F_3$ be disjoint edges such that $e, f \subset \{j_1, j_2, j_3, k_1, k_2, k_3\}$, and let us assume that $i_1 < i_2 < i_3$. If $f_1(T) \neq 0$, then $i_1 \in F_1$ and so T is bad because of $\{i_1, e, f\}$. Similarly, if $f_2(T) \neq 0$, then $\{i_1, i_2\} \in F_2$ and again T is bad, while we assumed that T is good.

(ii) Observe that if $\{j_1, j_2, j_3\} \notin F_3$, then every edge contained in $\{j_1, j_2, j_3, k_1, k_2, k_3\}$ is not in F_3. Since there are 8 such edges, we have $f_3(T) \leq 19$. Thus, if $f_3(T) \geq 20$, then $\{j_1, j_2, j_3\} \in F_3$, and because T is good, it is easy to see that $f_1(T) \leq 1$. Now assume by contradiction that $f_2(T) \geq 13$. Then, there are two edges in T, let say $(i_1, j_1, k_1), (i_2, j_2, k_2)$, such that at least five edges of F_2 are contained in theirs set of vertices. By Claim 5, $\{i_1, k_2\}, \{k_1, i_2\} \in F_2$ and thus, T is bad because of the edges $\{i_1, k_2\}, \{k_1, i_2\}, \{j_1, j_2, j_3\}$.

(iii) It is a direct consequence of Claim 5 and the fact that $\{j_1, j_2, j_3\} \notin F_3$, since then $f_2(T) \leq 12$ (see (ii) above).

(iv) Since $f_3(T) = 21$, we know that $\{j_1, j_2, j_3\} \in F_3$ and $\{k_1, k_2, k_3\} \notin F_3$. Therefore, at least one pair of edges from Fig. 3. and Fig. 4. is in F_3.

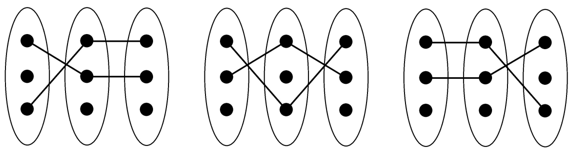

Fig. 3.

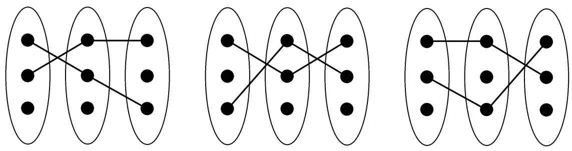

Fig. 4.

Since such a pair saturates only one vertex from I, we have $f_1(T) \leq 1$. To estimate $f_2(T)$ let assume that j_2 is not saturated by this pair of edges. Then, $\{i_1, j_2\}, \{j_2, i_3\} \notin F_2$, because T is good. Consequently, $\{i_1, k_2\}, \{j_1, j_2\}, \{j_2, j_3\}, \{k_2, i_3\}$ are also not in F_3 and thus, at most six edges of F_2 are contained in $\{i_1, j_1, k_1, i_2, j_2, k_2\}$ or in $\{i_2, j_2, k_2, i_3, j_3, k_3\}$. Now, since $\{j_1, j_2, j_3\} \in F_3$, using the same argument as in (ii), we conclude that at most four edges of F_2 are contained in $\{i_1, j_1, k_1, i_3, j_3, k_3\}$. Hence, $f_2(T) \leq 10$.

(v) From (i) we know that if in T we can find one of the pairs of edges marked on Fig. 2, then $f_1(T) = f_2(T) = 0$. Thus, let assume that for each of these pairs at least one edge is not in F_3 and $22 \leq f_3(T) \leq 23$. Hence $\{j_1, j_2, j_3\} \in F_3$ and $\{k_1, k_2, k_3\} \notin F_3$. Now consider $\{j_1, j_2, k_3\}$, $\{j_1, k_2, j_3\}$, $\{k_1, j_2, j_3\}$. It is easy to check that if at most one of them is in F_3, then $f_3(T) \leq 21$. Thus, we split our further argument into two cases.

Case 1. All three edges $\{j_1, j_2, k_3\}$, $\{j_1, k_2, j_3\}$, $\{k_1, j_2, j_3\}$ *are in* F_3.

Then, $\{j_1, k_2, k_3\}$, $\{k_1, j_2, k_3\}$, $\{k_1, k_2, j_3\}$, $\{k_1, k_2, k_3\} \notin F_3$. Therefore, as $f_3(T) \geq 22$, at least two pairs of edges shown on Fig. 3. are in F_3. Let say these are $\{i_1, k_2, k_3\}, \{k_1, j_2, j_3\}$ and $\{k_1, k_2, i_3\}, \{j_1, j_2, k_3\}$. Since T is good, edges $\{j_1, i_2\}$, $\{j_1, i_3\}$, $\{i_2, j_3\}$, $\{i_1, j_3\}$ are not in F_2, and because F_2 is shifted, the edges of F_2 contained in $\bigcup T$ are contained in the set $\{\{i_1, i_2\}, \{i_1, j_2\}, \{i_1, k_2\}, \{i_2, i_3\}, \{j_2, i_3\}, \{k_2, i_3\}, \{i_1, i_3\}\}$. Hence, $f_2(T) \leq 7$. It is also easy to observe that in that case $f_1(T) = 0$.

Case 2. Exactly two of the edges $\{j_1, j_2, k_3\}$, $\{j_1, k_2, j_3\}$, $\{k_1, j_2, j_3\}$ *are in* F_3.

Without loss of generality let $\{j_1, j_2, k_3\}$, $\{j_1, k_2, j_3\} \in F_3$. Then, $\{k_1, j_2, j_3\}$, $\{k_1, j_2, k_3\}$, $\{k_1, k_2, j_3\}$, $\{k_1, k_2, k_3\} \notin F_3$. Therefore, if $f_3(T) = 23$, then all other edges are in F_3, and so two pairs of edges shown on Fig. 3. are in F_3. Thus, as we have shown in the proof of Case 1, $f_2(T) \leq 7$. Let now consider the case when $f_3(T) = 22$. If both pairs of edges $\{j_1, k_2, j_3\}, \{k_1, i_2, k_3\}$ and $\{k_1, k_2, i_3\}$, $\{j_1, j_2, k_3\}$ are in F_3, then again $f_2(T) \leq 7$. Let now assume that only one of these pairs is in F_3, let say $\{j_1, k_2, j_3\}, \{k_1, i_2, k_3\} \in F_3$. Then also a pair $\{j_1, k_2, k_3\}, \{k_1, j_2, i_3\}$ is in F_3. Thus, $\{i_1, j_2\}$, $\{j_2, i_3\}$, $\{i_1, j_3\}$, $\{i_2, j_3\} \notin F_2$, and therefore, $f_2(T) \leq 7$. In that case we also have $f_1(T) = 0$. □

Now we bound the number of edges in $\mathbf{Sh}(G)$. First of all let us remove from M six edges so that in the remaining matching $\bar{M}$ we have only good triples (see Claim 4). In this way we omit at most $9n^2$ edges of $\mathbf{Sh}(G)$. Let us recall also that the number of untraceable edges of $\mathbf{Sh}(G)$ is at most $31n^2$. Finally, observe that for each edge $f \in F_i$ there are at most $\binom{n-3s}{3-i}$ edges $e \in \mathbf{Sh}(G)$ such that $e \cap V(M) = f$. Thus, the number of edges in $\mathbf{Sh}(G)$ is bounded from above by

$$e(\mathbf{Sh}(G)) \leq |F_1|\binom{n-3s}{2} + |F_2|(n-3s) + |F_3| + 40n^2.$$

To bound $|F_i|$, let us sum $f_i(T)$ over all $T \in \bar{M}^{(3)}$. Observe that in such a sum each edge from F_i is cour nted exactly $\binom{s-i}{3-i}$ times. Thus,

$$e(\mathbf{Sh}(G)) \le \sum_{T\in\bar{M}^{(3)}} \left(f_1(T)\frac{(n-3s)^2}{s^2} + f_2(T)\frac{n-3s}{s} + f_3(T) \right) + 41n^2.$$

Now we divide good triples into 27 groups, depending on $f_3(T)$. If $T_i = \{T \in \bar{M}^{(3)} : f_3(T) = i\}$ for $i = 1, \ldots, 27$, then

$$e(\mathbf{Sh}(G)) \le \sum_{i=1}^{27}\sum_{T\in T_i} \left(f_1(T)\frac{(n-3s)^2}{s^2} + f_2(T)\frac{n-3s}{s} + f_3(T) \right) + 41n^2.$$

Let now denote $x_1 = \sum_{i=1}^{19} |T_i|$, $x_2 = |T_{20}|$, $x_3 = |T_{21}|$, $x_4 = |T_{22}| + |T_{23}|$, $x_5 = \sum_{i=24}^{27} |T_i|$. By Claim 6, we get the following bound.

$$\begin{aligned} e(\mathbf{Sh}(G)) &\le (3x_1 + x_2 + x_3)\frac{(n-3s)^2}{s^2} \\ &+ (15x_1 + 12x_2 + 10x_3 + 7x_4)\frac{n-3s}{s} \\ &+ (19x_1 + 20x_2 + 21x_3 + 23x_4 + 27x_5) + 41n^2. \end{aligned}$$

Now it is sufficient to maximize the above function under the conditions $\sum_{i=1}^{5} x_i \le \binom{s-6}{3}$ and $x_i \ge 0$ for every $i = 1, \ldots, 5$. Then, we are to maximize a function

$$f_{s,n}(x_1, x_2, x_3, x_4, x_5) = \sum_{i=1}^{5} \alpha_i(s, n)x_i,$$

where

$$\begin{aligned} \alpha_1(s, n) &= 3(n-3s)^2/s^2 + 15(n-3s)/s + 19 \\ \alpha_2(s, n) &= (n-3s)^2/s^2 + 12(n-3s)/s + 20 \\ \alpha_3(s, n) &= (n-3s)^2/s^2 + 10(n-3s)/s + 21 \\ \alpha_4(s, n) &= 7(n-3s)/s + 23 \\ \alpha_5(s, n) &= 27\,, \end{aligned}$$

over domain $\sum_{i=1}^{5} x_i \le \binom{s-6}{3}$, $x_i \ge 0$ for $i = 1, 2, \ldots, 5$. This is a linear function of x_i's, so in order to maximize it it is enough to check which of the coefficients $\alpha_i(s, n)$ is the largest one and set the variable x_i which corresponds to this coefficient to be maximum, while the rest of the variables should be equal to zero.

It is easy to check that if $s = an$ and $a < a_0$, where $a_0 = (\sqrt{321} - 3)/52$, then $\alpha_1(s, n)$ dominates, and so for $s = an$, $a < a_0$, we have

$$e(\mathbf{Sh}(G)) \leq \frac{1}{6}(3s - 3s^2 + s^3) + O(n^2),$$

what nicely matches the lower bound for $e(\mathbf{Sh}(G))$ given by

$$e(\mathbf{Sh}(G)) \geq \binom{n}{3} - \binom{n-s}{3} = \frac{1}{6}(3s - 3s^2 + s^3) + O(n^2)\,.$$

Furthermore, in order to achieve this bound for all but $O(n^2)$ triples T we must have $f_3(T) = 19$, $f_2(T) = 15$, $f_1(T) = 3$, which is possible only if all edges of such triple intersect I (see Claim 6(iii)). Consequently, for this range of s, in $\mathbf{Sh}(G)$ there is a subset I, $|I| = s$, which covers all but at most $O(n^2)$ edges of $\mathbf{Sh}(G)$.

For $a > a_0$ the dominating coefficient is $\alpha_5(s, n) = 27$, which gives

$$e(\mathbf{Sh}(G)) \leq \frac{9}{2}s^3 + O(n^2)\,,$$

matched by the lower bound

$$e(\mathbf{Sh}(G)) \geq \binom{3s+2}{3} = \frac{9}{2}s^3 + O(n^2)\,.$$

Again to achieve this bound for all but $O(n^2)$ triples we must have $f_3(T) = 27$. But then the largest independent set contained in $\bigcup M$ has at most $O(n^{2/3})$ vertices and so, because of shifting, $\bigcup M$ contains a clique of size at least $s - O(n^{2/3}) = s - O(s^{2/3})$.

In order to complete the proof we need to consider the remaining case when $s = (a_0 + o(1))n$. Since $\alpha_1(a_0 n, n) = \alpha_5(a_0 n, n) > \alpha_i(a_0 n, n)$ for $i = 2, 3, 4$, we infer that in $\mathbf{Sh}(G)$ all triples, except of at most $O(n^2)$, must be of one of two types: either for such a triple T we have $f_3(T) = 27$, $f_2(T) = f_1(T) = 0$, or $f_3(T) = 19$, $f_2(T) = 15$, $f_3(T) = 3$ and all edges of H contained in $\bigcup T$ intersect I. It is easy to see that it is possible only when one of these two types of triples dominate. Indeed, let $M' \subseteq M$ denote the set of edges of$M$ which contain a singleton edge form F_1. Since the number of triples which are contained neither in M', nor in $M \setminus M'$ is $O(s^2)$, so $\min\{|M'|, |M \setminus M'|\}$ is bounded and, consequently, all but $O(s^2) = O(n^2)$ triples must be of one of our two types. Hence

$$\begin{aligned} \mathbf{Sh}(G) &\in \mathrm{Cov}_3(n, s; O(n^{-1})) \cup \mathrm{Cl}_3(n, s; O(n^{-1/3})) \\ &\subseteq \mathrm{Cov}_3(n, s; \varepsilon) \cup \mathrm{Cl}_3(n, s; \varepsilon)\,, \end{aligned}$$

and the assertion follows. □

References

[1] N. Alon, H. Huang, and P. Sudakov, *Nonnegative k-sums, fractional covers, and probability of small deviations*, J. Combin. Th. B, to appear.

[2] J. Akiyama and P. Frankl, *On the size of the graphs with complete factors*, J. Graph Th. (9) **27** (1985), 197–201.

[3] B. Bollobás, E. Daykin, and P. Erdős, *Sets of independent edges of a hypergraph*, Quart. J. Math. Oxford Ser. (2) **27** (1976), 25–32.

[4] P. Erdős, *A problem on independent r-tuples*, Ann. Univ. Sci. Budapest. Etvs Sect. Math. **8** (1965), 93–95.

[5] P. Erdős and T. Gallai, *On maximal paths and circuits of graphs*, Acta Math. Acad. Sci. Hungar. **10** (1959), 337–356.

[6] P. Erdős, C. Ko, and R. Rado, *Intersection theorems for systems of finite sets*, Quart. J. Math. Oxford Set. 2 **12** (1961), 313–320.

[7] P. Frankl, *The shifting technique in extremal set theory*, in: Surveys in Combinatorics, Lond. Math. Soc. Lect. Note Ser., **123** (1987), 81–110.

[8] P. Frankl, V. Rödl, and A. Ruciński, *On the maximum number of edges in a triple system not containing a disjoint family of a given size*, Combinatorics, Probability & Computing, to appear.

[9] H. Huang, P. Loh, and B. Sudakov, *The size of a hypergraph and its matching number*, Combinatorics, Probability & Computing, to appear.

Adam Mickiewicz University, Collegium Mathematicum, ul. Umultowska 87, 61-614 Poznań, Poland

E-mail address: `tomasz@amu.edu.pl`

Adam Mickiewicz University, Collegium Mathematicum, ul. Umultowska 87, 61-614 Poznań, Poland

E-mail address: `kaska@amu.edu.pl`