Available online at www.sciencedirect.com
ScienceDirect
Discrete Mathematics 306 (2006) 3097 – 3111
DISCRETE MATHEMATICS
www.elsevier.com/locate/disc

On set systems with a threshold property

Zoltán Füredi[a,b,1], Robert H. Sloan[c,2], Ken Takata[d,2,3], György Turán[e,f,2,4]

[a] *Department of Mathematics, University of Illinois at Urbana-Champaign, Urbana, IL 61801, USA*
[b] *Rényi Institute of Mathematics of the Hungarian Academy of Sciences, Budapest, P.O. Box 127, Hungary 1364*
[c] *Department of Computer Science, University of Illinois at Chicago, Chicago, IL 60607-7053, USA*
[d] *Department of Mathematics, Hamline University, Saint Paul, MN 55104-1284, USA*
[e] *Department of Mathematics, Statistics and Computer Science, University of Illinois at Chicago, Chicago, IL 60607-7045, USA*
[f] *Research Group on Artificial Intelligence, Hungarian Academy of Sciences, Szeged, Hungary 6720*

Received 17 November 2003; received in revised form 21 December 2005; accepted 21 December 2005
Available online 16 October 2006

Abstract

For n, k and t such that $1<t<k<n$, a set $\mathscr{F}$ of subsets of $[n]$ has the (k,t)-*threshold property* if every k-subset of $[n]$ contains at least t sets from $\mathscr{F}$ and every $(k-1)$-subset of $[n]$ contains less than t sets from $\mathscr{F}$. The minimal number of sets in a set system with this property is denoted by $m(n,k,t)$. In this paper we determine $m(n,4,3)$ *exactly* for n sufficiently large, and we show that $m(n,k,2)$ is asymptotically equal to the generalized Turán number $T_{k-1}(n,k,2)$.

Keywords: Set system; Extremal problem; Packing

1. Introduction

For n, k and t such that $1<t<k<n$, a set $\mathscr{F}$ of subsets of $[n]=\{1,\ldots,n\}$ has the (k,t)-*threshold property* if

- every k-subset of $[n]$ contains at least t sets from $\mathscr{F}$,
- every $(k-1)$-subset of $[n]$ contains less than t sets from $\mathscr{F}$.

Let $m(n,k,t)$ denote the minimal number of sets in a set system with the (k,t)-threshold property. The set of all $(k-1)$-subsets over $[n]$ satisfies the requirements for every t, and thus it follows that

$$m(n,k,t)\leqslant\binom{n}{k-1}. \tag{1}$$

E-mail addresses: z-furedi@math.uiuc.edu, furedi@renyi.hu (Z. Füredi), sloan@uic.edu (R.H. Sloan), takata@adelphi.edu (K. Takata), gyt@uic.edu (G. Turán).

[1] Research supported in part by the Hungarian National Science Foundation under Grants OTKA T 032452, T 037846 and by the National Science foundation under Grant DMS 0140692.
[2] Research supported in part by the National Science Foundation under Grants CCR-9800070, CCR-0100336 and CCF-0431059.
[3] Part of this work was done while Takata was a Ph.D. student at University of Illinois at Chicago.
[4] Research supported in part by the Hungarian National Science Foundation under Grant OTKA T-25721.

0012-365X/$ - see front matter © 2006 Published by Elsevier B.V.
doi:10.1016/j.disc.2006.06.001

It was shown in [22] that for every fixed k it holds that

$$m(n,k,t)=\Omega(n^{k-1}) \tag{2}$$

and

$$m(n,3,2)=\binom{n-1}{2}+1. \tag{3}$$

Also, Sloan et al. [22] gave bounds for $m(n,4,2)$ and $m(n,4,3)$. The function $m(n,k,t)$ was studied in [22] in the context of *frequent sets* of Boolean matrices, a concept used in knowledge discovery and data mining [1]. Jukna [12] studied the same function in circuit complexity theory, in a different range of the parameters, for establishing trade-offs between the size of threshold circuits and the size of the thresholds used in the gates of the circuit.

In this paper we determine $m(n,4,3)$ *exactly* for n sufficiently large. The optimal solution turns out to be rather complicated (see Section 4.1). There are only a few exact results known for this kind of extremal problem (such as [3,9]). We also show that $m(n,k,2)$ is asymptotically equal to the generalized Turán number $T_{k-1}(n,k,2)$.

The paper is organized as follows. In Section 2 we give some preliminaries. Section 3 gives the bound for $m(n,k,2)$. The result for $m(n,4,3)$ is contained in Sections 4 and 5, the former giving the proof without the proofs of the lemmas, and the latter proving the lemmas. Finally, in Section 6 we mention some open problems.

2. Preliminaries

For a set system $\mathscr{F}$, let $\mathscr{F}_i$ denote the set of i-element subsets in $\mathscr{F}$.[5]

Proposition 1. *For every n, k and every $t\geqslant 3$ there are optimal set systems over $[n]$ with the (k,t)-threshold property for which $\mathscr{F}_k=\emptyset$.*

Proof. Let $\mathscr{F}$ be an optimal set system with the (k,t)-threshold property, and let H be a k-element set in $\mathscr{F}$. As H cannot be deleted from $\mathscr{F}$, it must be the case that besides H itself, H contains exactly $t-1$ sets from $\mathscr{F}$. At least one of these sets is nonempty. Let one of its elements be x. Then $H\backslash\{x\}$ contains at most $t-2$ sets from $\mathscr{F}$. If $H\backslash\{x\}\notin\mathscr{F}$ then deleting H from $\mathscr{F}$ and adding $H\backslash\{x\}$ instead, one maintains the (k,t)-threshold property and decreases the number of k-sets in $\mathscr{F}$. Otherwise, as $t-1\leqslant k-2$, there is a $y\neq x$ such that $H\backslash\{y\}\notin\mathscr{F}$. Deleting H from $\mathscr{F}$ and adding $H\backslash\{y\}$ instead, one again maintains the (k,t)-threshold property and decreases the number of k-sets in the system $\mathscr{F}$. Repeating this operation, all k-element sets can be eliminated from $\mathscr{F}$. □

For $t=2$ the situation is somewhat different, as the family consisting of $\emptyset$ and all k-subsets of $[n]$ has the $(k,2)$-threshold property, and it contains fewer sets than the family of all $(k-1)$-sets, for small values of n. Proposition 1 remains valid, though, for large values of n, and one can also note that optimal systems cannot contain the empty set.

Proposition 2. *For every k and $n>2k$ there are optimal set systems over $[n]$ with the $(k,2)$-threshold property for which*

$$\mathscr{F}_0=\mathscr{F}_k=\emptyset.$$

Proof. Let $\mathscr{F}$ be an optimal set system with the $(k,2)$-threshold property. The assumption $n>2k$ implies that $\mathscr{F}$ cannot contain $\emptyset$. Otherwise, the nonempty sets in $\mathscr{F}$ all have to be of size k, thus $|\mathscr{F}|\geqslant 1+\binom{n}{k}$, and so $\mathscr{F}$ cannot be optimal. In order to show k-subsets can be eliminated from $\mathscr{F}$, one can now argue as in the first case of the proof of Proposition 1. □

The generalized Turán number $T_r(n,k,t)$ is the minimal number of r-subsets of $[n]$ such that every k-subset of $[n]$ contains at least t of these subsets. Then clearly

$$m(n,k,t)\leqslant T_{k-1}(n,k,t). \tag{4}$$

[5] We use capitals $A, B, \ldots$ for sets of elements and pairs, and script letters $\mathscr{H}, \mathscr{M}, \mathscr{T}\ldots$ for sets of triples, quadruples, etc., with the exception of $\mathscr{F}_0, \mathscr{F}_1$ and $\mathscr{F}_2$.

The best known bounds for $T_3(n,4,2)$ are due to Mubayi [15] (the lower bound, improving a previous result of de Caen [5]) and Frankl and Füredi [9] and independently by Sidorenko [19] (the upper bound).

Lemma 3 (*Mubayi [15], Frankl and Füredi [9], Sidorenko [19]*).

$$\left(\frac{2}{3}+10^{-6}\right)\binom{n}{3}\leqslant T_3(n,4,2)\leqslant\frac{5}{7}(1+o(1))\binom{n}{3}.$$

The best known bounds for $T_{k-1}(n,k,2)$ for $k\geqslant 5$ are the following:

$$\frac{2}{k-1}(1+o(1))\binom{n}{k-1}\leqslant T_{k-1}(n,k,2)\leqslant C\frac{\log k}{k}\binom{n}{k-1}.$$

Here the lower bound is due to de Caen [6] and the upper bound can be obtained modifying the constructions of [10,21].

A (3, 2)-*packing* is a set of triples such that every pair is contained in at most one triple. In this paper by a *packing* we always mean a (3, 2)-packing. The packing number $P(r,3,2)$ is the maximal number of triples of $[r]$ which form a packing. A *maximum* packing is a packing of $P(r,3,2)$ triples. The size of maximum packings of triples were determined by Spencer [23]. The following results are also presented, e.g., in Rogers and Stanton [18].

The size of a maximum packing is given by

$$P(r,3,2)=\begin{cases}\left\lfloor\frac{r}{3}\left\lfloor\frac{r-1}{2}\right\rfloor\right\rfloor & \text{if } r\not\equiv 5 \pmod 6,\\ \left\lfloor\frac{r}{3}\left\lfloor\frac{r-1}{2}\right\rfloor\right\rfloor-1 & \text{if } r\equiv 5 \pmod 6.\end{cases}$$

Given a packing $\mathscr{P}$, a pair is called *missing* if it is not contained in any triple of $\mathscr{P}$. If $r\equiv 1$ or $3 \pmod 6$ then maximum packings are Steiner systems, containing every pair *exactly* once, thus there are no missing pairs. If $r\equiv 0$ or $2 \pmod 6$ then missing pairs in maximum packings form a perfect matching. If $r\equiv 4 \pmod 6$ then the graph of missing pairs is a star on four vertices plus a perfect matching on the remaining vertices. If $r\equiv 5 \pmod 6$ then the missing pairs form a cycle of length four.

3. The bound for $m(n,k,2)$

In this section we show that $m(n,k,2)$ is asymptotically equal to a generalized Turán number. This shows that in these cases using subsets of size *less than* $k-1$ does not help, at least asymptotically. For $k=3$, the optimal system consists of a singleton plus a complete graph on the remaining vertices, while the best solution using only pairs is a complete graph minus a perfect matching. Thus, for the exact solution it *does* help to use smaller sets, at least in this special case.

Theorem 4. *It holds that* $m(n,k,2)=(1+o(1))T_{k-1}(n,k,2)$.

Proof. In view of (4), we only need to prove

$$m(n,k,2)\geqslant(1+o(1))T_{k-1}(n,k,2).$$

First we state a result of Erdős and Simonovits [8]. If $\mathscr{H}$ is any r-uniform hypergraph then $\mathrm{ex}_r(n,\mathscr{H})$ denotes the maximal number of r-tuples over $[n]$ that do not contain a copy of $\mathscr{H}$. It was shown by Katona et al. [13] that

$$\lim_{n\to\infty}\frac{\mathrm{ex}_r(n,\mathscr{H})}{\binom{n}{r}}$$

always exists (see, e.g., [4,11]). The limit is denoted by $\pi_r(\mathscr{H})$.

Erdős and Simonovits [8] showed that if there are significantly more r-tuples than $\mathrm{ex}_r(n,\mathscr{H})$, then there are *many* copies of $\mathscr{H}$. Let v denote the number of vertices of $\mathscr{H}$. □

Lemma 5 (*Erdös and Simonovits [8]*). *For every* $\varepsilon>0$ *there is a* $\delta=\delta(\varepsilon,\mathscr{H})>0$, *such that every set of at least* $(\pi_r(\mathscr{H})+\varepsilon)\binom{n}{r}$ *r-subsets over* $[n]$ *contains at least* $\delta\binom{n}{v}$ *copies of* $\mathscr{H}$.

In order to apply this result, let us consider the hypergraph $\mathscr{H}_k$ consisting of $k-1$ of the $(k-1)$-subsets on k vertices. Then, by definition, it holds that

$$T_{k-1}(n,k,2)=\binom{n}{k-1}-\mathrm{ex}_{k-1}(n,\mathscr{H}_k).$$

Thus, the statement of the theorem can be reformulated as

$$m(n,k,2)\geqslant(1+\mathrm{o}(1))(1-\pi_{k-1}(\mathscr{H}_k))\binom{n}{k-1}.$$

Now let us assume that for some $\varepsilon>0$ there are arbitrarily large values of n such that

$$\begin{aligned} m(n,k,2) &< (1-\varepsilon)\,(1-\pi_{k-1}(\mathscr{H}_k))\binom{n}{k-1} \\ &< \left(1-\pi_{k-1}(\mathscr{H}_k)-\varepsilon'\right)\binom{n}{k-1}, \end{aligned} \tag{5}$$

for some $\varepsilon'>0$ depending on ε and k. Here we used that

$$\pi_r(\mathscr{H})\leqslant\frac{e(\mathscr{H})-1}{e(\mathscr{H})}$$

holds for every hypergraph $\mathscr{H}$ which follows from

$$\mathrm{ex}(n,\mathscr{H})\leqslant\frac{e(\mathscr{H})-1}{e(\mathscr{H})}\binom{n}{r}.$$

This can be proved by a direct averaging argument [20].

Let $\mathscr{F}$ be an optimal system over $[n]$ with the $(k,2)$-threshold property. By Proposition 2 we may assume that $\mathscr{F}$ contains no k-subsets or the empty set. Inequality (5) implies

$$|\mathscr{F}_{k-1}|<\left(1-\pi_{k-1}(\mathscr{H}_k)-\varepsilon'\right)\binom{n}{k-1}.$$

Thus, by Lemma 5 (switching over to the set of $(k-1)$-subsets *not* contained in $\mathscr{F}_{k-1}$),

$$\begin{aligned} &\text{there are at least } \delta\binom{n}{k} \text{ many } k\text{-subsets of } [n] \\ &\quad\text{containing at most one set from } \mathscr{F}_{k-1}. \end{aligned} \tag{6}$$

It follows from the definition of the $(k,2)$-threshold property that $\mathscr{F}$ contains at most one set of size at most $\lfloor (k-1)/2\rfloor$, as the union of two such sets would be of size at most $k-1$, and it would contain two sets from $\mathscr{F}$. Thus, the number of k-subsets of $[n]$ containing any set from $\mathscr{F}$ of size at most $\lfloor (k-1)/2\rfloor$ is $\mathrm{O}(n^{k-1})$.

If k is even, than any two sets in $\mathscr{F}$ of size $k/2$ must be disjoint, for the same reason, thus their number is $\mathrm{O}(n)$. Hence, the number of k-subsets of $[n]$ containing any set from $\mathscr{F}_{k/2}$ is $\mathrm{O}(n\,n^{k/2})=\mathrm{O}(n^{k-1})$.

Furthermore, if $k/2<j\leqslant k-2$ then the intersection of any two sets of size j in $\mathscr{F}$ is at most $2j-k$, again for the same reason. (Note that for $j=k-1$ this restriction is meaningless.) Thus, every subset of $[n]$ of size at most $2j-k+1$ can be contained in at most one set belonging to $\mathscr{F}_j$, and so $|\mathscr{F}_j|=\mathrm{O}(n^{2j-k+1})$.

Hence, the number of k-subsets of $[n]$ containing any set from $\mathscr{F}_j$ is at most $\mathrm{O}(n^{2j-k+1}n^{k-j})=\mathrm{O}(n^{j+1})$. Therefore, the number of k-subsets of $[n]$ containing any set from $\mathscr{F}_j$ is $\mathrm{O}(n^{k-2})$ (resp., $\mathrm{O}(n^{k-1})$) for every j such that $k/2<j<k-2$ (resp., $j=k-2$).

In summary, the number of k-subsets of $[n]$ containing any set from $\mathscr{F}$ of size at most $k-2$ is $\mathrm{O}(n^{k-1})$. Thus, there are $(1-\mathrm{o}(1))\binom{n}{k}$ many k-subsets of $[n]$ which contain at least two sets from $\mathscr{F}_{k-1}$. This contradicts (6), thus proving the theorem. □

4. Determining $m(n, 4, 3)$

In order to formulate our main result, the exact determination of $m(n, 4, 3)$ below as Theorem 8, we define an integer valued function *f*. Then a number of properties of the function *f* are noted. The proofs of these properties, as well as the proofs of the other lemmas used in this section, are postponed to the next section.

Define

$$f(n) = \left\lfloor \frac{n}{2} \right\rfloor \left\lceil \frac{n}{2} \right\rceil + \binom{\lfloor \frac{n}{2} \rfloor}{3} + \binom{\lceil \frac{n}{2} \rceil}{3} - P\left(\left\lfloor \frac{n}{2} \right\rfloor, 3, 2\right) - P\left(\left\lceil \frac{n}{2} \right\rceil, 3, 2\right)$$
$$= \frac{n^3}{24} - \frac{n^2}{12} + \mathrm{O}(n).$$

Lemma 6. *For every fixed constant* ℓ *it holds that*

$$f(n) = f(n - \ell) + \frac{n^2}{8}\ell + \mathrm{O}(n).$$

Consider the function

$$g(n_1, n_2) = n_1 n_2 + \binom{n_1}{3} + \binom{n_2}{3} - P(n_1, 3, 2) - P(n_2, 3, 2). \tag{7}$$

Note that $f(n) = g(\lfloor n/2 \rfloor, \lceil n/2 \rceil)$.

Lemma 7. *If n is sufficiently large and* $n_1 + n_2 = n$ *then* $g(n_1, n_2)$ *is minimal iff* $|n_1 - n_2| \leqslant 1$.

As $g(2, 1) = 2$, $g(3, 0) = 0$, $g(2, 2) = 4$ and $g(3, 1) = 3$, equal division does not give the minimal value of g if n is 3 or 4.

Now we formulate the result determining $m(n, 4, 3)$.

Theorem 8. *If n is sufficiently large then* $m(n, 4, 3) = f(n)$.

The rest of this section presents the proof of the theorem. First, we describe a construction proving the upper bound, and then we give the lower bound. At the end of the section we discuss optimal systems.

4.1. A construction

A set system over $[n]$ of size $f(n)$ with the (4, 3)-threshold property can be constructed as follows. Divide the ground set into two halves of sizes $\lfloor n/2 \rfloor$, and $\lceil n/2 \rceil$. Include in $\mathscr{F}$ all pairs that contain one element from each half, and all triples contained in either half *except* a maximum packing in both halves. It is easy to check that this system has size $f(n)$ and it indeed has the (4, 3)-threshold property. Lemma 7 means that the same construction for unequal halves gives no improvement. As there are many non-isomorphic maximum packings, this construction gives many non-isomorphic systems of size $f(n)$. We will refer to all these as *packing constructions*.

4.2. The lower bound

Let $\mathscr{F}$ be a set system over $[n]$ with the (4, 3)-threshold property. By Proposition 1, it may be assumed that $\mathscr{F}$ contains no quadruples. It must be the case that $\emptyset \notin \mathscr{F}$, as otherwise $\mathscr{F}$ without $\emptyset$ has the (4, 2)-threshold property, and from the lower bound of Lemma 3 it follows that $|\mathscr{F}| > f(n)$ if n is sufficiently large. An important property, which follows directly from the definition of the (4, 3)-threshold property, is that $\mathscr{F}_2$ is a *triangle-free graph*.

Now $\mathscr{F}$ may contain at most two singletons. If the singleton $\{x\}$ belongs to $\mathscr{F}$ then the (4, 3)-threshold property implies that there can be at most one pair in $\mathscr{F}$ containing x. Also, if there is such a pair $\{x, y\}$, then y cannot be contained in any further pair in $\mathscr{F}$. If there are two singletons in $\mathscr{F}$, then, again from the definition of the (4, 3)-threshold property, neither can be contained in a pair belonging to $\mathscr{F}$.

Thus, there is a set C of at most two vertices such that C contains all singletons of $\mathscr{F}$, and there is no pair in $\mathscr{F}$ that contains exactly one element from C. Let us assume w.l.o.g. that the vertices *not* in C are $\{1, \ldots, m\}$, where $n-2 \leqslant m \leqslant n$. Consider the graph G induced by $\mathscr{F}_2$ on $[m]$. Thus, $G = ([m], \mathscr{F}_2|[m])$. Let $d(x)$ denote the degree of vertex x in G.

We distinguish two cases, based on the degrees of G. These two cases are discussed in separate subsections below.

4.2.1. *Case I:* $\sum_{x=1}^{m}((m-1)/2 - d(x))^2 > 2(m-1)^2$

Let $t()$ denote the number of triangles in a graph. We use the identity

$$\sum_{x=1}^{m} \binom{d(x)}{2} + \binom{m-1-d(x)}{2} = \binom{m}{3} + 2(t(G) + t(\bar{G})). \tag{8}$$

This holds as the left-hand side counts triangles in G, resp., in $\bar{G}$, three times, and all other triples once (see, e.g., Lovász [14]). As noted above, the (4, 3)-threshold property implies that $t(G) = 0$. Using the identity[6]

$$\binom{a}{2} + \binom{b-a}{2} = 2\binom{\frac{b}{2}}{2} + \left(\frac{b}{2} - a\right)^2,$$

Eq. (8) can be rewritten as

$$\begin{aligned} t(\bar{G}) &= \frac{1}{2}\left(2m\binom{\frac{m-1}{2}}{2} - \binom{m}{3} + \sum_{x=1}^{m}\left(\frac{m-1}{2} - d(x)\right)^2\right) \\ &= \frac{m^3}{24} - \frac{m^2}{4} + \mathrm{O}(m) + \frac{1}{2}\sum_{x=1}^{m}\left(\frac{m-1}{2} - d(x)\right)^2 \\ &> \frac{m^3}{24} + \frac{3}{4}m^2 + \mathrm{O}(m). \end{aligned}$$

In the last step we used the assumption of Case I to estimate the last term from below.

Now let $\mathscr{S}$ be the set of triples over $[m]$ containing no set from $\mathscr{F}$. It follows from the definitions that

$$t(\bar{G}) \leqslant |\mathscr{S}| + |\mathscr{F}_3|.$$

Thus one gets

$$|\mathscr{F}_3| \geqslant \frac{m^3}{24} + \frac{3}{4}m^2 + \mathrm{O}(m) - |\mathscr{S}|.$$

Hence, one can get a lower bound to $|\mathscr{F}_3|$ by proving an upper bound for $|\mathscr{S}|$. This is obtained in the following lemma.

Lemma 9. $|\mathscr{S}| \leqslant m(m-1)/3$.

Thus we get

$$|\mathscr{F}_3| \geqslant \frac{m^3}{24} + \frac{3}{4}m^2 - \frac{m^2}{3} + \mathrm{O}(m) > \frac{m^3}{24} + \frac{m^2}{3} + \mathrm{O}(m),$$

which is *greater than* $f(n)$ if n is sufficiently large. This completes the proof of the theorem for Case I.

---

[6] Here we write $\binom{x}{2}$ for $x(x-1)/2$, where x is not necessarily an integer.

4.2.2. Case II: $\sum_{x=1}^{m}((m-1)/2)-d(x))^2 \leqslant 2(m-1)^2$

In order to consider this case, let us introduce some further notation. Let us consider the set

$$D=\{x : d(x)<0.45(m-1)\}.$$

By the assumption of Case II, the size of D is bounded by a constant. In particular, it holds that $|D|\leqslant 800$. Assume, w.l.o.g. that $[m]\backslash D=[r]$.

We use the following result of Andrásfai et al. [2].

Lemma 10 (*Andrásfai et al. [2]*). *Every triangle-free non-bipartite graph on r vertices has a vertex of degree at most* $\frac{2}{5}r$.

Let G' be the graph induced by G on $[r]$, and let the set of edges of G' be E. Now every vertex of G' has degree at least $0.45(m-1)-800$. This is greater than $\frac{2}{5}r$ if n is sufficiently large. Hence, if n is sufficiently large then G' is *bipartite*. Let the two halves of this bipartite graph be A and B, with $|A|=r_1$ and $|B|=r_2$.

In summary, so far we have partitioned the ground set into four sets: the *exceptional sets* C and D, which are of size bounded by a constant and the sets A and B. The graph $G'=([r],E)$ formed by the pairs in $\mathscr{F}_2$ over $[r]$ is bipartite, with bipartition (A,B).

The assumption of Case II implies that G' is 'almost' a complete bipartite graph with equal halves. In particular, we have the following.

Lemma 11. *If n is sufficiently large then it holds that*

(1) $|r_1-n/2|\leqslant n^{3/4}$ *and* $|r_2-n/2|\leqslant n^{3/4}$,
(2) *there are* $O(n^{7/4})$ *edges missing between A and B.*

In fact, $O(\sqrt{n})$ and $O(n^{3/2})$ can be proved (instead of $n^{3/4}$ and $O(n^{7/4})$), but it would not simplify the main line of the arguments later.

Now we introduce some more notation for different types of triples in $\mathscr{F}_3$. For $x\in C\cup D$ let $\mathscr{W}_x$ denote the set of triples in $\mathscr{F}$ containing x which have their other two elements in $A\cup B$. Let $\mathscr{T}_1$ (resp., $\mathscr{T}_2$) be the set of triples in $\mathscr{F}$ with two points in A and one in B (resp., two in B and one in A), and put $\mathscr{T}=\mathscr{T}_1\cup\mathscr{T}_2$. Let $\mathscr{H}_1$ (resp., $\mathscr{H}_2$) be the set of triples in $\mathscr{F}$ contained in A (resp., B), and let $\mathscr{M}_1$ (resp., $\mathscr{M}_2$) be the set of triples contained in A (resp., B) which do *not* belong to $\mathscr{F}$. Finally, let M_0 be the set of pairs $\{x,y\}\notin\mathscr{F}$, for which $x\in A$ and $y\in B$.

As quadruples in A, resp., B, do not contain any singletons or pairs from $\mathscr{F}$, it must be the case that they contain at least three triples from $\mathscr{F}$. Hence $\mathscr{M}_1$ and $\mathscr{M}_2$ are *packings*, implying $|\mathscr{M}_i|\leqslant P(r_i,3,2)$. Thus,

$$|\mathscr{H}_i|=\binom{r_i}{3}-|\mathscr{M}_i|\geqslant\binom{r_i}{3}-P(r_i,3,2). \tag{9}$$

Hence, by the definition of the function g, and by Lemma 7, it holds that

$$r_1r_2+\sum_{i=1}^{2}|\mathscr{H}_i|\geqslant g(r_1,r_2)\geqslant f(r). \tag{10}$$

Lemma 7 also implies that (10) holds with *strict* inequality *unless* $|r_1-r_2|\leqslant 1$, for large enough n.

The lower bound estimates for remainder of the proof begin with the following inequality, which is a direct consequence of the definitions:

$$|\mathscr{F}|\geqslant\sum_{x\in C\cup D}|\mathscr{W}_x|+\sum_{i=1}^{2}(|\mathscr{T}_i|+|\mathscr{H}_i|)+|E|. \tag{11}$$

We start by formulating two lemmas that are used to treat the exceptional sets C and D.

First, let us consider an element $x \in C$, and $y_1, y_2, y_3 \in A$. The quadruple $\{x, y_1, y_2, y_3\}$ contains no pairs from $\mathscr{F}_2$. It may be the case that $\{x\} \in \mathscr{F}$ and $\{y_1, y_2, y_3\} \in \mathscr{F}$, but we still need one more triple in $\{x, y_1, y_2, y_3\}$. Hence, at least one of the triples $\{x, y_1, y_2\}, \{x, y_1, y_3\}, \{x, y_2, y_3\}$ belongs to $\mathscr{F}_3$. If, in addition, it holds that $\{y_1, y_2, y_3\} \notin \mathscr{F}_3$, then at least *two* of the triples $\{x, y_1, y_2\}, \{x, y_1, y_3\}, \{x, y_2, y_3\}$ belong to $\mathscr{F}_3$.

Let us consider the graph $G_A^x = (A, E_A^x)$, where $\{y_1, y_2\} \in E_A$ iff $\{x, y_1, y_2\} \notin \mathscr{F}$. It follows from the remarks above that G_A^x is *triangle-free*, and furthermore, it contains *at most one* edge from every triple $\{y_1, y_2, y_3\}$ belonging to the packing $\mathscr{M}_1$. The following lemma shows that every triangle-free graph, which contains at most one edge from every triple in a large packing, has relatively few edges. This lemma implies that either there are only a few triples missing in $\mathscr{F}$ from A or B, or there are many triples incident to x. In both cases one obtains a lower bound for the number of triples in $\mathscr{F}$.

Lemma 12. *Let H be a triangle-free graph on s vertices and $\mathscr{P}$ be a packing of size at least $\frac{1}{8}s^2$ over the vertex set of H, such that every triple in $\mathscr{P}$ contains at most one edge from H. Then H has at most $(1 - 1/50)(s^2/4)$ edges, if s is sufficiently large.*

The next lower bound shows that there are many triples incident to the vertices of D.

Lemma 13. *For every $x \in D$ it holds that $|\mathscr{W}_x| \geqslant (1 + \frac{1}{120})(n^2/8)$, if n is sufficiently large.*

Recall that M_0 denotes the set of pairs between A and B that are missing from $\mathscr{F}_2$. We consider the bipartite graph $G_0 = (A \cup B, M_0)$. Note that G_0 is the 'bipartite complement' of G' above. Depending on the degrees of this graph, we distinguish two subcases of Case II. The maximal degree of G_0 is denoted by $\Delta(G_0)$.

Case II.1: $\Delta(G_0) \geqslant 10$.

In this case we use the following lemma.

Lemma 14. *If $\Delta(G_0) \geqslant 10$ then $|M_0| < |\mathscr{T}|$.*

As $|M_0| = r_1 r_2 - |E|$, it follows that

$$|\mathscr{T}| + |E| > r_1 r_2. \tag{12}$$

Rearranging (11), using Lemma 13 and (12) we get

$$\begin{aligned} |\mathscr{F}| &\geqslant \sum_{x \in D} |\mathscr{W}_x| + |\mathscr{T}| + |E| + \sum_{x \in C} |\mathscr{W}_x| + \sum_{i=1}^{2} |\mathscr{H}_i| \\ &> (1 + \frac{1}{120}) \frac{n^2}{8} |D| + r_1 r_2 + \sum_{x \in C} |\mathscr{W}_x| + \sum_{i=1}^{2} |\mathscr{H}_i|. \end{aligned} \tag{13}$$

We distinguish two subcases, depending on the sizes of the sets $\mathscr{M}_1, \mathscr{M}_2$.

Case II.1.1: $|\mathscr{M}_i| \geqslant (r_i^2/8)$ holds for $i = 1, 2$.

In this case Lemmas 11 and 12, and (10) imply that we can continue inequality (13) by writing, for large enough n,

$$\begin{aligned} |\mathscr{F}| &> \left(1 + \frac{1}{120}\right) \frac{n^2}{8} |D| + \left(\sum_{i=1}^{2} \left(\binom{r_i}{2} - \left(1 - \frac{1}{50}\right) \frac{r_i^2}{4} \right) \right) |C| + f(r) \\ &\geqslant \left(1 + \frac{1}{120}\right) \frac{n^2}{8} |D| + \left(1 + \frac{1}{60}\right) \frac{n^2}{8} |C| + f(r). \end{aligned} \tag{14}$$

Using Lemma 6 it follows that (14) is *greater than* $f(n)$, for large enough n.

Case II.1.2: $|\mathscr{M}_i| < (r_i^2/8)$ holds for some $i = 1, 2$.

Assume w.l.o.g. that $|\mathscr{M}_1| < (r_1^2/8)$. Then, Eq. (9) can be rewritten as

$$|\mathscr{H}_1| = \binom{r_1}{3} - |\mathscr{M}_1| > \binom{r_1}{3} - \frac{r_1^2}{8}$$
$$\geqslant \binom{r_1}{3} - P(r_1, 3, 2) + \frac{1}{24} r_1^2 + \mathrm{O}(r_1). \tag{15}$$

In this case we cannot apply Lemma 12. However, as G_A^x is triangle-free, it still holds that $|E_A^x| \leqslant (r_1^2/4)$ by Turán's theorem. Hence, in this case, using (15), we can continue inequality (13) with

$$|\mathscr{F}| \geqslant \left(1 + \frac{1}{120}\right) \frac{n^2}{8} |D| + r_1 r_2$$
$$+ \left(\sum_{i=1}^{2} \left(\binom{r_i}{2} - \frac{r_i^2}{4}\right)\right) |C| + \sum_{i=1}^{2} |\mathscr{H}_i|$$
$$\geqslant \left(1 + \frac{1}{120}\right) \frac{n^2}{8} |D| + \left(\sum_{i=1}^{2} \left(\binom{r_i}{2} - \frac{r_i^2}{4}\right)\right) |C|$$
$$+ f(r) + \frac{1}{24} r_1^2 + \mathrm{O}(r_1)$$
$$\geqslant \left(1 + \frac{1}{120}\right) \frac{n^2}{8} |D| + \frac{n^2}{8} |C| + f(r) + \frac{1}{96} n^2 + \mathrm{o}(n^2). \tag{16}$$

Using Lemma 6 it again follows that (16) is *greater than* $f(n)$ for large enough n.

Case II.2: $\Delta(G_0) < 10$.

In this case we replace Lemma 14 with the following.

Lemma 15. *If* $\Delta(G_0) < 10$ *then*

$$|\mathscr{T}_i| + P(r_i, 3, 2) \geqslant |\mathscr{M}_i| + \tfrac{1}{2} |M_0|.$$

If equality holds then $\mathscr{T}_i = M_0 = \emptyset$ *and* $\mathscr{M}_i$ *is a maximum packing*.

Using $|\mathscr{M}_i| = \binom{r_i}{3} - |\mathscr{H}_i|$ and $|M_0| = r_1 r_2 - |E|$, Lemma 15 implies

$$|\mathscr{T}_i| + |\mathscr{H}_i| + \frac{1}{2} |E| \geqslant \binom{r_i}{3} - P(r_i, 3, 2) + \frac{1}{2} r_1 r_2. \tag{17}$$

Similar to Case II.1 above, we distinguish two subcases depending on the sizes of the sets $\mathscr{M}_1$, $\mathscr{M}_2$.

Case II.2.1: $|\mathscr{M}_i| \geqslant (r_i^2/8)$ holds for $i = 1, 2$.

Summing (17) for i, we get

$$\left(\sum_{i=1}^{2} (|\mathscr{T}_i| + |\mathscr{H}_i|)\right) + |E| \geqslant f(r). \tag{18}$$

Thus, using (11), (18), Lemmas 12 and 13, it follows that

$$|\mathscr{F}| \geqslant \sum_{x \in C \cup D} |\mathscr{W}_x| + \left(\sum_{i=1}^{2} (|\mathscr{T}_i| + |\mathscr{H}_i|)\right) + |E|$$
$$\geqslant \left(1 + \frac{1}{120}\right) \frac{n^2}{8} |D| + \left(1 + \frac{1}{60}\right) \frac{n^2}{8} |C| + f(r),$$

which is *at least* $f(n)$ for large enough n.

It also follows that $|\mathscr{F}| > f(n)$ for large enough n, *unless* $C = D = M_0 = \mathscr{T}_1 = \mathscr{T}_2 = \emptyset$, $\mathscr{M}_1$ and $\mathscr{M}_2$ are maximum packings and $|r_1 - r_2| \leqslant 1$. In this case $\mathscr{F}$ is one of the *packing constructions* described in Section 4.1.

Case II.2.2: $|\mathscr{M}_i| < (r_i^2/8)$ holds for some $i = 1, 2$.

Assume w.l.o.g. that $|\mathscr{M}_1| < (r_1^2/8)$. In this case we further modify Lemma 15.

Lemma 16. *If* $\Delta(G_0) < 10$ *and* $|\mathscr{M}_1| < (r_1^2/8)$ *then*

$$|\mathscr{T}_1| + \tfrac{4}{5} P(r_1, 3, 2) \geqslant |\mathscr{M}_1| + \tfrac{1}{2}|M_0|.$$

Rewriting as in the previous subcase, we get

$$|\mathscr{T}_1| + |\mathscr{H}_1| + \frac{1}{2}|E| \geqslant \binom{r_1}{3} - P(r_1, 3, 2) + \frac{1}{2} r_1 r_2 + \frac{1}{5} P(r_1, 3, 2). \tag{19}$$

For $i = 2$ inequality (17) holds. Adding it to inequality (19), one gets

$$\sum_{i=1}^{2} (|\mathscr{T}_i| + |\mathscr{H}_i|) + |E| \geqslant f(r) + \frac{1}{5} P(r_1, 3, 2). \tag{20}$$

Hence, arguing as in Case II.1.2,

$$\begin{aligned} \mathscr{F} &\geqslant \sum_{x \in C \cup D} |\mathscr{W}_x| + \left(\sum_{i=1}^{2} (|\mathscr{T}_i| + |\mathscr{H}_i|) \right) + |E| \\ &\geqslant \left(1 + \frac{1}{120}\right) \frac{n^2}{8} |D| + \frac{n^2}{8} |C| + f(r) + \frac{1}{5} P(r_1, 3, 2) + \mathrm{o}(n^2), \end{aligned}$$

which is again *greater than* $f(n)$ for large enough n. This completes the proof of the theorem.

4.3. Optimal systems

As noted in Section 4.1, there are many non-isomorphic packing constructions, thus there are many different optimal systems with the (4, 3)-threshold property. One can ask whether *all* optimal systems are packing constructions? In this section we show that the answer to this question depends on the (mod 12) remainder of n.

Theorem 17. *Every optimal system with the* (4, 3)*-threshold property is a packing construction iff* $n \equiv 0, 1, \ldots, 6 \pmod{12}$, *for n sufficiently large*.

Proof. It was noted at the beginning of the proof of Theorem 8 that optimal systems cannot contain $\emptyset$.

The proof of the theorem also shows that, assuming that an optimal system contains no quadruples, it also cannot contain any singletons. The proof of Proposition 1 shows that quadruples can be eliminated by repeatedly replacing a quadruple by a triple contained in it. Combining these two observations, it follows that an optimal system cannot contain any singletons.

We concluded $|\mathscr{F}| > f(n)$ in all cases of Theorem 8 *except* Case II.2.1. In that case, we noted that only packing constructions can provide equality. The proof of Theorem 8 assumed that $\mathscr{F}$ contains no quadruples.

Hence, the remaining question is to decide if there are optimal systems which contain quadruples? We recall from Section 2 that a pair is *missing* from a maximum packing $\mathscr{P}$ if it is not contained in any triple of $\mathscr{P}$. The results on maximum packings described there imply that the graph formed by missing pairs is the same for every maximum packing on a given ground set. Inspecting the graph of missing pairs in all cases, it can be observed that maximum packings over $[r]$ contain incident missing pairs iff $r \equiv 4$ or $5 \pmod 6$. If $r \equiv 4 \pmod 6$ then the missing pairs form a star on four vertices, and if $r \equiv 5 \pmod 6$ then the missing pairs form a cycle of length four.

Now let us assume that there is an optimal system containing quadruples as well. Then, using the procedure described in the proof of Proposition 1, let us eliminate all quadruples. Let the system obtained before eliminating the last quadruple be $\mathscr{F}'$. Then it holds that

$$\mathscr{F} = \mathscr{F}' \backslash \{Q\} + \{T\},$$

where $\mathscr{F}$ is a packing construction, $Q = \{x, y, z, u\}$ and $T = \{x, y, z\}$.

As $\mathscr{F}$ is a packing construction, it must be the case that x, y, z belong to the same half. The element u must also belong to the same half, as otherwise $\{u, x\}$, $\{u, y\}$, $\{u, z\}$ would be in $\mathscr{F}'$, and so Q could have been deleted from $\mathscr{F}'$. It also follows that $\mathscr{F}'$ contains exactly two of the triples $\{u, x, y\}$, $\{u, x, z\}$ and $\{u, y, z\}$. Assume w.l.o.g. that $\mathscr{F}'$ contains $\{u, x, z\}$ and $\{u, y, z\}$.

Consider the maximum packing $\mathscr{P}$ consisting of the triples missing from $\mathscr{F}'$ in the half containing Q. We claim that $\{x, z\}$ and $\{y, z\}$ are missing pairs for $\mathscr{P}$.

Let us assume that this is not the case. Thus, there is a v in the same half such that, w.l.o.g., $\{v, x, z\} \in \mathscr{P}$, or, equivalently, $\{v, x, z\} \notin \mathscr{F}$. Then v must be different from u or y. But then the quadruple $\{x, y, z, v\}$ contains at most two sets from $\mathscr{F}'$, contradicting the (4, 3)-threshold property. Thus, we can conclude that if there is an optimal system which is not a packing construction, then either $\lfloor n/2 \rfloor$ or $\lceil n/2 \rceil$ is congruent to 4 or 5 $\pmod 6$. This happens when n is congruent to 7, 8, 9, 10 or 11 $\pmod{12}$.

To complete the proof of the theorem, we also have to show that in these cases there is an optimal system which is not a packing construction. Reversing the above argument, let us assume that $\mathscr{F}$ is a packing construction, and $\{x, z\}$ and $\{y, z\}$ are missing pairs for some maximum packing $\mathscr{P}$ in one of the halves. Then $T = \{x, y, z\} \in \mathscr{F}$. Also, as $\{x, y\}$ cannot be a missing pair, there is a triple $\{u, x, y\} \in \mathscr{P}$, or, equivalently, $\{u, x, y\} \notin \mathscr{F}$. Then, let us form the system

$$\mathscr{F}' = \mathscr{F} \backslash \{T\} + \{Q\},$$

where $Q = \{x, y, z, u\}$. We claim that $\mathscr{F}'$ has the (4, 3)-threshold property. It only has to be shown that every quadruple $\{x, y, z, v\}$ contains at least three triples from $\mathscr{F}'$, where $u \neq v$ and v is in the same half as u. But $\{v, x, z\}$ and $\{v, y, z\}$ are in $\mathscr{F}$, and hence, in $\mathscr{F}'$, as $\{x, z\}$ and $\{y, z\}$ are missing pairs. Also, $\{v, x, y\}$ is in $\mathscr{F}$, and hence, in $\mathscr{F}'$, as $\mathscr{P}$ is a packing and $\{u, x, y\} \notin \mathscr{F}$. This completes the proof of the theorem. □

5. Proofs of the lemmas for Theorem 8

In this section we give the proofs of the lemmas used in the proof of Theorem 8.

Proof of Lemma 6. Follows by direct calculation. □

Proof of Lemma 7. Let us assume that $n_1 \geqslant n_2 > 0$. Considering cases corresponding to the possible parities of n_1 and n_2, one gets

$$P(n_1, n_2) - P(n_1 + 1, n_2 - 1) = \beta n_2 - \alpha n_1 + \gamma,$$

where α and β are $\frac{1}{6}$ or $\frac{1}{2}$, and $\gamma \geqslant -\frac{13}{3}$. Using this, it follows again by direct calculation that

$$g(n_1 + 1, n_2 - 1) - g(n_1, n_2) \geqslant \frac{n_1^2 - n_2^2}{2} - 2n_1 + \frac{8}{3}n_2 - \frac{13}{3},$$

which is positive if, e.g., $n \geqslant 14$. □

Proof of Lemma 9. It is sufficient to show that every pair x, y is contained in at most two triples from $\mathscr{S}$, as then the size of $\mathscr{S}$ is at most $2\binom{m}{2}/3$. Assume that $\{x, y, z_1\}$, $\{x, y, z_2\}$ and $\{x, y, z_3\}$ are different triples in $\mathscr{S}$.

The quadruple $\{x, y, z_1, z_2\}$ contains at least three sets from $\mathscr{F}$. But neither of the triples $\{x, y, z_1\}$, $\{x, y, z_2\}$ or the pairs $\{x, y\}$, $\{x, z_1\}$, $\{x, z_2\}$, $\{y, z_1\}$, $\{y, z_2\}$ are in $\mathscr{F}$. Thus, it must be the case that $\{z_1, z_2\}$ is in $\mathscr{F}$. Similarly, $\{z_1, z_3\}$ and $\{z_2, z_3\}$ are in $\mathscr{F}$. But this is a contradiction, as then $\{z_1, z_2, z_3\}$ contains three sets from $\mathscr{F}$. □

Proof of Lemma 11. Let us call a vertex *bad* if

$$\left|\frac{m-1}{2} - d(x)\right| \geqslant \frac{1}{2} m^{3/4},$$

and *good* otherwise. The assumption of Case II implies that, for large enough n, there are at most $m^{3/4}$ bad vertices. For the first part of the claim, assume that $r_1 > m/2 + m^{3/4}$. Then, as A is an independent set, all the vertices of A have degree at most $m/2 - m^{3/4}$, contradicting the above inequality. If $r_1 < m/2 - m^{3/4}$, then $r_2 > m/2 + m^{3/4} - 800$, and we get a similar contradiction.

The second part follows by noting that missing edges can be of two types. Either their endpoint in A is bad, or it is good. For the first type, the bound follows as there are at most $m^{3/4}$ bad vertices, each incident to at most m missing edges. For the second type, for each of the at most good m vertices in A, there are $\mathrm{O}(m^{3/4})$ missing edges by the degree condition and the upper bound on the size of B. □

Proof of Lemma 12. Let us assume that H has more than $(1-\frac{1}{50})(s^2/4)$ edges. If H is not bipartite then let us iteratively delete vertices of minimum degree until the remaining graph becomes bipartite. Let the number of remaining vertices be denoted by j. By Lemma 10 and Turán's theorem, it holds that

$$\sum_{\ell=j+1}^{s} \frac{2}{5}\ell + \frac{j^2}{4} \geqslant \left(1 - \frac{1}{50}\right)\frac{s^2}{4}. \tag{21}$$

Direct computation gives $j > 0.94s$ if s is sufficiently large. Thus, the number of vertices deleted is at most $0.06s$, and the number of edges deleted is at most $\frac{2}{5}0.06s^2 < 0.1s^2/4$. Hence, the final bipartite graph has at least $(0.98 - 0.1)(s^2/4) = 0.88(s^2/4)$ edges. Therefore, there are at most $(1-0.88)(s^2/4) = 0.12(s^2/4)$ edges missing from between the two halves of the bipartition.

The number x of vertices in one half of the bipartition satisfies $x(s-x) \geqslant 0.88(s^2/4)$. Thus, it follows that both halves have at least $0.32s$ vertices. The number of triples of the packing $\mathscr{P}$ contained within the two halves is at most $\frac{1}{6}(0.32^2 + 0.68^2)s^2 + \mathrm{O}(s \leqslant 0.0942s^2$. Hence, at least $(\frac{1}{8} - 0.0942)s^2 > 0.123(s^2/4)$ triples of $\mathscr{P}$ contain points from both halves. Now H has at most one edge in each such triple. Thus, for each triple there is at least one edge missing from between the two halves. As $\mathscr{P}$ is a packing, the edges corresponding to different triples are different. So there are at least $0.123(s^2/4)$ triples missing from between the two halves, contradicting the upper bound $0.12(s^2/4)$ given above. □

One can obtain a slightly stronger statement using a more involved argument and applying some results of Erdős, Faudree, Pach, and Spencer [7] concerning how to make a graph bipartite. But, again, these improvements do not seem to simplify the proof of our main theorem.

Proof of Lemma 13. Consider a vertex $x \in D$, and let the number of pairs $\{x, y\} \notin \mathscr{F}_2$ with $y \in A$ (resp., $y \in B$) be a (resp., b).

If $\{x, z_1\} \in \mathscr{F}_2$ and $\{x, z_2\} \in \mathscr{F}_2$ for $z_1 \in A$ and $z_2 \in B$ then, as $\mathscr{F}_2$ is triangle free, it must be the case that $\{z_1, z_2\} \notin \mathscr{F}_2$. Hence, the second part of Lemma 11 implies

$$(r_1 - a)(r_2 - b) = \mathrm{O}(n^{7/4}). \tag{22}$$

Thus at least one of the terms, say, $r_1 - a$, is $\mathrm{O}(n^{7/8})$, and so $a \geqslant n/2 - \mathrm{O}(n^{7/8})$. By assumption, $a + b > 0.55(m-1)$, so the first part of Lemma 11 implies that $b > \frac{1}{21}n$, if n is sufficiently large.

Now consider a quadruple $\{x, y_1, y_2, y_3\}$ with $y_1, y_2, y_3 \in A$ such that $\{x, y_i\} \notin \mathscr{F}_2$ for $i = 1, 2, 3$. This quadruple contains no singletons or pairs from $\mathscr{F}$. Hence, it must contain at least two of the triples $\{x, y_1, y_2\}$, $\{x, y_1, y_3\}$, $\{x, y_2, y_3\}$. In other words, only one of these three triples can be missing from $\mathscr{F}_3$. So the number of triples $\{x, y, z\} \in \mathscr{F}_3$ with $y, z \in A$ is at least

$$\binom{a}{2} - \left\lfloor \frac{a}{2} \right\rfloor. \tag{23}$$

Of course, the same argument applies to B as well. Hence the number of triples containing x is at least

$$\binom{a}{2} - \left\lfloor \frac{a}{2} \right\rfloor + \binom{b}{2} - \left\lfloor \frac{b}{2} \right\rfloor \geqslant \binom{\frac{1}{21}n}{2} + \binom{\frac{n}{2} - \mathrm{O}(n^{7/8})}{2} - \mathrm{O}(n)$$
$$\geqslant \left(1 + \frac{1}{120}\right)\frac{n^2}{8},$$

if n is sufficiently large. □

Proof of Lemma 14. Consider vertices $x \in A$ and $y_1, y_2, y_3 \in B$ such that $\{x, y_i\} \in M_0$ for $i = 1, 2, 3$. Then the quadruple $\{x, y_1, y_2, y_3\}$ contains no singleton or pair from $\mathscr{F}$. Hence, even if $\{y_1, y_2, y_3\} \in \mathscr{F}$, at least two of the triples $\{x, y_1, y_2\}, \{x, y_1, y_3\}, \{x, y_2, y_3\}$ belong to $\mathscr{F}$. This shows that in general, if $x \in A$ has ℓ neighbors $y_1, \ldots, y_\ell$ in $G_0 = (A \cup B, M_0)$, then at least $\binom{\ell}{2} - \lfloor \ell/2 \rfloor$ of the $\binom{\ell}{2}$ triples $\{x, y_i, y_j\}$ are in $\mathscr{F}$. Otherwise, two triples not belonging to $\mathscr{F}$ would share a y_i, contradicting the above argument.

Hence, as

$$\binom{\ell}{2} - \left\lfloor \frac{\ell}{2} \right\rfloor \geqslant \ell \tag{24}$$

for $\ell \geqslant 4$, it follows that if $x \in A$ (resp., $x \in B$) and $d_{M_0}(x) \geqslant 4$, then $d_{\mathscr{T}_2}(x) \geqslant d_{M_0}(x)$ (resp., $d_{\mathscr{T}_1}(x) \geqslant d_{M_0}(x)$).

Now, starting with G_0 and the set of triples $\mathscr{T}$, let us iteratively delete a vertex of maximal degree, as long as its degree is at least 4. Then (24) guarantees that in each step we delete at least as many M_0-pairs as $\mathscr{T}$-triples. Furthermore, as

$$\binom{\ell}{2} - \left\lfloor \frac{\ell}{2} \right\rfloor \geqslant \ell + 19 \tag{25}$$

if $\ell \geqslant 10$, and we assumed $\Delta(G_0) \geqslant 10$, it follows that in the first phase we delete at least 19 more M_0-pairs than $\mathscr{T}$-triples.

Let the graph obtained from G_0 at the end of this process be $G_1 = (A_1 \cup B_1, M_1)$, and the set of remaining triples be $\mathscr{U}$. By construction, it holds that $\Delta(G_1) \leqslant 3$.

Consider two independent pairs $\{x_1, y_1\}$ and $\{x_2, y_2\}$ in G_1. The quadruple $\{x_1, x_2, y_1, y_2\}$ contains no singletons, and at most two pairs from $\mathscr{F}$. Thus, it contains at least one triple from $\mathscr{F}$. If $\{x_i, y_i\}$, $i = 1, \ldots, v$, is a maximum set of independent pairs in M_1, then this construction gives at least $\binom{v}{2}$ different triples, hence $|\mathscr{U}| \geqslant \binom{v}{2}$. On the other hand, $|M_1| \leqslant 3v$ by König's theorem.

If $v \geqslant 7$ then $3v \leqslant \binom{v}{2}$, hence in this case $|M_1| \leqslant |\mathscr{U}|$. If $v < 7$ then it follows that $|M_1| \leqslant 18$. The 'advantage' 19 gained by M_0-pairs over $\mathscr{T}$-triples in the first phase guarantees that the lemma holds in both cases. □

Proof of Lemma 15. We may assume that $M_0 \neq \emptyset$. Otherwise the statement follows directly, as $\mathscr{M}_1$ is a packing. Also, we assume w.l.o.g that $i = 1$. We show that under these assumptions, the statement of the lemma holds with *strict* inequality.

Let $\{x, y\}$ be a pair in M_0 for $x \in A$, $y \in B$, and let $\{x, z_1, z_2\}$ be a triple in $\mathscr{M}_1$. Then $\{x, y, z_1, z_2\}$ contains at most two pairs from $\mathscr{F}$. Thus, $\mathscr{F}$ contains at least one of the triples $\{x, y, z_1\}, \{x, y, z_2\}, \{y, z_1, z_2\}$. In this way, for every triple $\{x, z_1, z_2\} \in \mathscr{M}_1$ we find a corresponding triple in $\mathscr{T}_1$, incident to y. As $\mathscr{M}_1$ is a packing, for different triples in $\mathscr{M}_1$ we get different triples in $\mathscr{T}_1$. Thus,

$$d_{\mathscr{T}_1}(y) \geqslant d_{\mathscr{M}_1}(x), \tag{26}$$

where d denotes the degrees in the corresponding hypergraphs. In order to apply (26), let B_1 be the set of non-isolated vertices of M_0 in B. For every $y \in B_1$, select an $x \in A$ such that $\{x, y\} \in M_0$. Let the set of x's selected be A_1. As $\Delta(G_0) < 10$, it holds that

$$|M_0| < 10|B_1| \leqslant 100|A_1|. \tag{27}$$

As the assumption $M_0 \neq \emptyset$ implies $100|A_1| \leqslant \frac{2}{3}\lfloor (r_1-1)/2 \rfloor |A_1| - 4$ if n is large enough, (27) implies

$$\frac{2}{3}\left\lfloor \frac{r_1-1}{2} \right\rfloor |A_1| - 4 > |M_0|. \tag{28}$$

Adding up the inequalities (26) for every selected pair it follows that

$$\sum_{y\in B_1} d_{\mathscr{T}_1}(y) \geqslant \sum\nolimits_{x\in A_1} d_{\mathscr{M}_1}(x). \tag{29}$$

(After the summation, every term on the right-hand side appears with a coefficient $\geqslant 1$, which is then changed to 1.) Using again the fact that $\mathscr{M}_1$ is a packing, $d_{\mathscr{M}_1}(x) \leqslant \lfloor r_1 - 1/2 \rfloor$, and so

$$|A \backslash A_1| \left\lfloor \frac{r_1-1}{2} \right\rfloor \geqslant \sum_{x\in A\backslash A_1} d_{\mathscr{M}_1}(x). \tag{30}$$

Now adding up $\frac{1}{2}$ times (28) and $\frac{1}{3}$ times (29) and (30), we get

$$\frac{1}{3}\sum_{y\in B_1} d_{\mathscr{T}_1}(y) + \frac{1}{3} r_1 \left\lfloor \frac{r-1}{2} \right\rfloor - 2 > \frac{1}{3}\sum_{x\in A} d_{\mathscr{M}_1}(x) + \frac{1}{2}|M_0|. \tag{31}$$

The lemma now follows by noting that the first term on the left-hand side is at most $|\mathscr{T}_1|$, and the second term on the left-hand side is at most $P(r_1, 3, 2)$. □

Proof of Lemma 16. The assumption of the lemma implies

$$|\mathscr{M}_1| \leqslant \tfrac{3}{4} P(r_1, 3, 2) + \mathrm{O}(r_1).$$

Hence (30) can be replaced by

$$\sum_{x\in A\backslash A_1} d_{\mathscr{M}_1}(x) \leqslant \sum_{x\in A} d_{\mathscr{M}_1}(x) = 3|\mathscr{M}_1| \leqslant \frac{9}{4} P(r_1, 3, 2) + \mathrm{O}(r_1). \tag{32}$$

Let us add up $\frac{1}{3}$ times (29), (32) and $\frac{1}{2}$ times

$$|M_0| \leqslant 10 r_1.$$

We get

$$\frac{1}{3}\sum_{y\in B_1} d_{\mathscr{T}_1}(y) + \frac{3}{4} P(r_1, 3, 2) + \mathrm{O}(r_1) \geqslant \frac{1}{3}\sum_{x\in A} d_{\mathscr{M}_1}(x) + \frac{1}{2}|M_0|, \tag{33}$$

and the statement of the lemma follows. □

6. Remarks and open problems

The results of Section 4.3 have been completed by Takata [24], who gives a description of all optimal systems for n sufficiently large.

The general lower bound (2) uses a repeated application of the Ramsey theorem, and thus the constants involved decrease rapidly with k. Hence there are no lower bounds for $m(n, k, t)$ when k increases with n. In view of the detailed results for $m(n, 4, t)$, it would also be interesting to get bounds for the case $k = 5$.

In general, one expects that for fixed n and k, the function $m(n, k, t)$ decreases with t in the range $1 < t < k$. It would be interesting to prove this. Another general question is to consider the case when $\mathscr{F}$ may be a multiset. The results of [22] apply to this more general case, but in the present paper we did not consider this version.

The case $t \geqslant k$ was recently investigated by Mubayi and Zhao [16,17]. Among other results they solved (asymptotically) all cases $k = 4$, $t \leqslant 16$. They also proposed a series of new problems.

In the opposite range of parameters, it would be interesting to look at $m(n, n/2, t)$. The results of Jukna [12] imply that $m(n, n/2, t)$ is superpolynomial if $t = o(n/\log n)$. This can be also showed in a simpler way by noting that if $\mathscr{F}$ is a system over $[n]$ with the $(n/2, t)$-threshold property then every $n/2$-subset of $[n]$ contains a different family of t sets from $\mathscr{F}$. Thus it must be the case that

$$\binom{n}{\frac{n}{2}} \leqslant \binom{m}{t}.$$

A specific question here is, motivated by the fact that for fixed k one has $m(n, k, k-1) = \Omega(n^{k-1})$, whether $m(n, n/2, n/2 - 1)$ is exponential?

Finally, we mention an open problem related to Lemma 12, giving an upper bound for the number of edges of triangle-free graphs which contain at most one edge in every triple of a large packing. This appears to be an extension of the Turán theorem in a direction which has not been studied yet. Thus, one could ask for the maximal number of edges in a triangle-free graph which contains at most one edge from every triple in some fixed maximum packing.

References

[1] R. Agrawal, H. Mannila, R. Srikant, H. Toivonen, A.I. Verkamo, Fast discovery of association rules, in: U.M. Fayyad, G. Piatetsky-Shapiro, P. Smyth, R. Uthurusamy (Eds.), Advances in Knowledge Discovery and Data Mining, AAAI Press, Menlo Park, CA, 1996, pp. 307–328.
[2] B. Andrásfai, P. Erdős, V.T. Sós, On the connection between chromatic number, maximal clique and minimal degree of a graph, Discrete Math. 8 (1974) 205–218.
[3] B. Bollobás, Three-graphs without two triples whose symmetric difference is contained in a third, Discrete Math. 8 (1974) 21–24.
[4] B. Bollobás, Combinatorics: Set Systems, Hypergraphs, Families of Vectors, and Combinatorial Probability, Cambridge University Press, Cambridge, 1986.
[5] D. de Caen, Extension of a theorem of Moon and Moser on complete subgraphs, Ars Combin. 16 (1983) 5–10.
[6] D. de Caen, A note on the probabilistic approach to Turán's problem, J. Combin. Theory Ser. B 34 (1983) 340–349.
[7] P. Erdős, R. Faudree, J. Pach, J. Spencer, How to make a graph bipartite, J. Combin. Theory Ser. B 45 (1988) 86–98.
[8] P. Erdős, M. Simonovits, Supersaturated graphs and hypergraphs, Combinatorica 3 (1983) 181–192.
[9] P. Frankl, Z. Füredi, An exact result for 3-graphs, Discrete Math. 50 (1984) 323–328.
[10] P. Frankl, V. Rödl, Lower bounds for Turán's problem, Graphs Combin. 1 (1985) 213–216.
[11] Z. Füredi, Turán type problems, Surveys in Combinatorics, London Mathematical Society Lecture Note Series, vol. 166, Cambridge University Press, Cambridge, 1991, pp. 255–300.
[12] S. Jukna, Computing threshold functions by depth-3 threshold circuits with smaller thresholds of their gates, Inform Process Lett. 56 (1995) 147–150.
[13] G. Katona, T. Nemetz, M. Simonovits, A new proof of a theorem of P. Turán and some remarks on a generalization of it, Mat. Lapok 15 (1964) 228–238 (In Hungarian).
[14] L. Lovász, Combinatorial Problems and Exercises, North-Holland, Amsterdam, 1979.
[15] D. Mubayi, On hypergraphs with every four points spanning at most two triples, Electron. J. Combin. 10 (2003) Research Paper 10, 4.
[16] D. Mubayi, Y. Zhao, On a two-sided Turán problem, Electron. J. Combin. 10 (2003) Research Paper 42, 17.
[17] D. Mubayi, Y. Zhao, Nonuniform Turán-type problems, J. Combin Theory A 111 (2005) 106–110.
[18] M.J. Rogers, R.G. Stanton, Packings and coverings by triples, Ars Combin. 13 (1982) 61–69.
[19] A.F. Sidorenko, Extremal constants and inequalities for distributions of sums of random vectors, Ph.D. Thesis, Moscow University, 1982, (in Russian).
[20] A.F. Sidorenko, An analytic approach to extremal problems for graphs and hypergraphs, Extremal problems for finite sets (Visegrád, 1991), Bolyai Math. Studies 3 (1994) 423–455.
[21] A.F. Sidorenko, Upper bounds for Turán numbers, J. Combin. Theory Ser. A 77 (1997) 134–147.
[22] R.H. Sloan, K. Takata, Gy. Turán, On frequent sets of Boolean matrices, Ann. Math. Artificial Intelligence 24 (1998) 193–209.
[23] J.H. Spencer, Maximal consistent families of triples, J. Combin. Theory 5 (1968) 1–8.
[24] K. Takata, Listing algorithms for combinatorial objects and related combinatorial problems, Ph.D. Thesis, University of Illinois at Chicago, 2004.