Families with no perfect matchings

Mihir Singhal*[1]

[1]*Massachusetts Institute of Technology, Cambridge, MA 02139, U.S.A.*
mihirs@mit.edu

Submitted: Nov 25, 2020; Accepted: Mar 25, 2021; Published: Dec 15, 2021

Abstract. We consider families of k-subsets of $\{1,\ldots,n\}$, where n is a multiple of k, which have no perfect matching. An equivalent condition for a family $\mathcal{F}$ to have no perfect matching is for there to be a *blocking set*, which is a set of b elements of $\{1,\ldots,n\}$ that cannot be covered by b disjoint sets in $\mathcal{F}$. We are specifically interested in the largest possible size of a family $\mathcal{F}$ with no perfect matching and no blocking set of size less than b. Frankl resolved the case of families with no singleton blocking set (in other words, the $b = 2$ case) for sufficiently large n and conjectured an optimal construction for general b. Though Frankl's construction fails to be optimal for $k = 2, 3$, we show that the construction is optimal whenever $k \geqslant 100$ and n is sufficiently large.

Mathematics Subject Classifications. 05D05

1. Introduction

Let $n = ks$ where $k, s \geqslant 2$. Consider a family $\mathcal{F}$ of k-element subsets of $[n] = \{1,\ldots,n\}$. A *matching* in $\mathcal{F}$ is a collection of sets in $\mathcal{F}$ which are all pairwise disjoint. A *perfect matching* is one which covers all elements of $[n]$ (equivalently, a matching of k-sets is a perfect matching if it has size s). We are interested in families $\mathcal{F}$ with no perfect matching. Specifically, we explore the question about how large such a family can be. This question is related to the Erdős matching conjecture [2], which regards the largest possible size of a family with no matching of size r, where n is not necessarily equal to rk.

Kleitman [10] showed that largest possible size of $\mathcal{F}$ given that it has no perfect matching is $\binom{n-1}{k}$, which is achieved by any family $\mathcal{F}$ which consists of all k-sets that do not contain a single vertex. Frankl [4] later showed that these are the only families which achieve this maximum. Such families can be considered *trivial* in the sense that there is a vertex in none of the sets of $\mathcal{F}$, which immediately implies that $\mathcal{F}$ does not have a perfect matching. Frankl considered the question of finding the largest possible size of a nontrivial $\mathcal{F}$ which has no perfect matching, and proved the following:

*Supported by NSF-DMS grant 1949884 and NSA grant H98230-20-1-0009.

Theorem 1.1 ([6])**.** *Suppose s is sufficiently large (relative to k). Then, if $\mathcal{F}$ is a nontrivial family of k-subsets of $[n]$, then the following family of k-sets attains the maximum possible size of $\mathcal{F}$:*

$$\mathcal{E}(k,n) = \big\{\{1\}\cup[n-k+2,n]\big\}\cup\big\{S : S\cap[n-k+2,n] \neq \varnothing, 1 \notin S, 2 \in S\big\}\cup\big\{T : T \subset [3,n]\big\}.$$

The family $\mathcal{E}(k,n)$ has no matching for the simple reason that there are no two disjoint sets in $\mathcal{E}(k,n)$ which cover both 1 and 2 (in particular, there is also no single set which covers both). This is an example of the notion of a *blocking set*.

In general, a *blocking set* of $\mathcal{F}$ is defined to be any set $B \subset [n]$ of size at most s which cannot be covered by a matching of size $|B|$. Note that a blocking set also cannot be covered by a larger matching since one can delete sets from any such matching which do not intersect B. (One can also define a blocking set to be one that is not coverable by any matching, regardless of its size; we will later see in Theorem 2.4 that it does not matter for our purposes which definition we choose.) It is clear that if $\mathcal{F}$ has a blocking set, then it cannot have a perfect matching; a small blocking set might also be considered to be a "trivial reason" that $\mathcal{F}$ does not have a matching. Indeed, as shown in Theorem 1.1, the smallest nontrivial $\mathcal{F}$ with no perfect matching has a blocking set of size 2. It is thus natural to ask what the largest possible size of $\mathcal{F}$ is if it has no small blocking set. Indeed, Frankl posed the following question:

Question 1.2 ([6])**.** Suppose $\mathcal{F}$ has no perfect matching and no blocking set of size less than b. What is the largest possible size of $\mathcal{F}$?

The case $b = 1$ is equivalent to having no restriction on blocking sets of $\mathcal{F}$, so Kleitman's bound of $\binom{n-1}{k}$ is the maximum. The $b = 2$ case restricts to nontrivial $\mathcal{F}$, which is just Theorem 1.1 (for large enough n). In fact, Frankl described a construction of a family for general b which he conjectured to be optimal. The construction is as follows.

Example 1.3 ([6])**.** Suppose $n \geqslant bk + b$. Let $E_1, \dots, E_{b-1}$ be pairwise disjoint $(k-1)$-sets, none of which intersect $[b]$. Define the families $\mathcal{E}_i$, for $1 \leqslant i \leqslant b$, as follows:

$$\mathcal{E}_i = \big\{\{i\} \cup E_i\big\} \cup \big\{\{i\} \cup S : S \subset [b+1,n],\ |S| = k-1,\ S \cap (E_1 \cup \cdots \cup E_{i-1}) \neq \varnothing\big\}.$$

(If $i = b$, then ignore the $\{\{i\} \cup E_i\}$.) Then we define the family

$$\mathcal{E}(k,n,b) = \big\{T : T \subseteq [b+1,n],\ |T| = k\big\} \cup \left(\bigcup_{i=1}^{b} \mathcal{E}_i\right).$$

If there is any matching in $\mathcal{E}(k,n,b)$ covering $[b]$, then by induction on i it will need to contain the set $\{i\} \cup E_i$ for each i. But there is no such set for $i = b$, a contradiction, so there can be no matching covering $[b]$. Therefore $[b]$ is a blocking set of $\mathcal{E}(k,n,b)$, and it is also easy to check that there is no smaller blocking set. It is also simple to check that when $k > 2$, this family $\mathcal{E}(k,n,b)$ is maximal in the sense that if any sets are added to it, then there will be a perfect matching. (For this last fact, we must use the assumption that $n \geqslant bk + b$.)

When $b = 2$, the set $\mathcal{E}(k,n,2)$ is the same (up to permutation of $[n]$) as $\mathcal{E}(k,n)$. Indeed, Frankl conjectured that the above construction is optimal when n is large:

Conjecture 1.4 ([6])**.** Suppose that $n = ks$ and $k, s \geqslant 2$ and b is a positive integer. There exists $s_0(k, b)$ such that whenever $s \geqslant s_0(k, b)$, the following holds: If $\mathcal{F}$ is a family of k-subsets of $[n]$ with no perfect matching and no blocking set of size less than b, then

$$|\mathcal{F}| \leqslant |\mathcal{E}(k, n, b)|.$$

It turns out that Theorem 1.4 actually fails when $b > 2$ and k is 2 or 3. When $k = 2$, $\mathcal{F}$ is just an ordinary graph, and it turns out that one can just add edges to $\mathcal{E}(2, n, b)$. In particular, noting that E_i are all singleton sets, we can add all the edges between each $i \in [b]$ and each E_j (only some of these edges exist in $\mathcal{E}(2, n, b)$). Then, elements of $[b]$ are only adjacent to elements of a $(b-1)$-element set, so there is still no matching covering $[b]$. This is actually optimal, though the proof of this is not completely trivial, and will require some of our later lemmas, so we will prove this in Section 5.

When $k = 3$, the situation is a little more subtle, but there is still a construction that has slightly more sets than $\mathcal{E}(3, n, b)$.

Example 1.5. Suppose $k = 3$ and $n \geqslant bk + b$. Then define the family of k-sets

$$\mathcal{E}'(3, n, b) = \big\{S : |S \cap [b]| = 1,\ |S \cap [b+1, 2b-1]| \geqslant 1\big\} \cup \big\{T : T \subset [b+1, n]\big\}.$$

The set $[b]$ is a blocking set of $\mathcal{E}'(3, n, b)$ because any matching which covers $[b]$ would need to contain one set which covers each element of $[b]$, and each such set would need to contain at least one element of $[b+1, 2b-1]$. Thus these b sets cannot be pairwise disjoint, so no such matching can exist. It is also easy to check that there is no smaller blocking set. Then, the size of $\mathcal{E}'(3, n, b)$ exceeds that of $\mathcal{E}(3, n, b)$ by $(b^3 - 7b + 6)/6$, which is positive whenever $b > 2$. We defer the calculation to verify this to Section 2.1.

Our main theorem is that Theorem 1.4 holds whenever $k \geqslant 100$, and also when $k \geqslant \max\{b+1, 6\}$. Specifically, there exists a function $s_0(k, b)$ such that the following theorem holds.

Theorem 1.6. *Let k, b be positive integers such that either $k \geqslant 100$ or $k \geqslant \max\{b+1, 6\}$. Also let $n = ks$, where $s \geqslant s_0(k, b)$. Suppose $\mathcal{F}$ is a family of subsets of $[n]$ of size k. Then, if $\mathcal{F}$ has no perfect matching and no blocking set of size less than b, then*

$$|\mathcal{F}| \leqslant |\mathcal{E}(k, n, b)|.$$

This essentially resolves Theorem 1.4, except in the case where k is small. Note that Theorem 1.1 for $k \geqslant 6$ can be deduced as a special case of Theorem 1.6. The remainder of this paper will be dedicated to proving Theorem 1.6. We first briefly describe our method of proof. Central to our method is the technique of *shifting*, first introduced by Erdős, Ko, and Rado [3]. We will define shifting in Section 2.2. Essentially, shifting is an operation that can be performed on a family of sets and introduces structure into the family while preserving the absence of a perfect matching. We first show in Section 3, using a method similar to that used by Frankl in proving Theorem 1.1, that the family $\mathcal{F}$ can be shifted on all but a constant number of elements of $[n]$ while maintaining the property that it has no small blocking set. Then, in Section 4, we exploit the shifted structure of $\mathcal{F}$ to show that in order for it to be big enough, it must have a blocking set of size exactly b. This result will be enough to resolve the $k = 2$ case (the case of an ordinary graph), where Theorem 1.6 does not apply but we can still find a different optimal family $\mathcal{F}$. In Section 5, we will resolve the $k = 2$ case. Then, in Section 6, we continue with the proof of Theorem 1.6, using the blocking set of size b in addition to the mostly shifted structure of $\mathcal{F}$ to reduce proving the theorem to proving Theorem 6.2, a much simpler statement which is about ordinary graphs. To do this we use techniques related to juntas, which were introduced in [1]. Finally, we prove Theorem 6.2 in Section 7.

2. Definitions and preliminary observations

We start with some definitions. We will refer to the elements of $[n]$ as *vertices*. As mentioned in the introduction, a *matching* is any collection of disjoint sets, and is said to *cover* a set if all elements of the set are contained in one of the elements in the matching. We will frequently abuse notation by saying that a collection of sets contains a vertex x to mean that x belongs to one of the sets of the collection. Also, if $\mathcal{F}$ is a family of sets, we define $\mathcal{F}(x)$ to be the family of all sets F such that $x \notin F$ and $F \cup \{x\} \in \mathcal{F}$.

We use standard asymptotic notation throughout, including big O and little o, big Ω and little ω, and the relation $\sim$ (we say that $f \sim g$ if $f = (1 + o(1))g$). All such asymptotic notation will be with respect to n (or equivalently s); in other words, k and b are considered to be constant for all asymptotic notation. We will also assume throughout that n is at least a sufficiently large function of k and b, using this assumption implicitly on many occasions.

2.1. Size of $\mathcal{E}(k, n, b)$

We now approximate the size of $\mathcal{E}(k, n, b)$, which will be useful for later comparisons. First, we show an approximation for the binomial coefficient $\binom{n-t}{k}$. We have, for any constant t, that

$$\binom{n-t}{k} = \binom{n}{k} - \sum_{i=1}^{t} \binom{n-i}{k-1} = \binom{n}{k} - t\binom{n}{k-1} + O(n^{k-2}).$$

Now, note that the size of each $\mathcal{E}_i$ (defined in Theorem 1.3) is

$$|\mathcal{E}_i| = \binom{n-b}{k-1} - \binom{n-b-(k-1)(i-1)}{k-1} + 1.$$

(In the case that $i = b$, the final 1 term should be omitted.) Then we have

$$\begin{aligned} |\mathcal{E}(k, n, b)| &= \binom{n-b}{k} + b\binom{n-b}{k-1} - \sum_{i=1}^{b} \binom{n-b-(k-1)(i-1)}{k-1} + b - 1 \\ &= \binom{n}{k} - \sum_{i=1}^{b} \binom{n-i}{k-1} + b\binom{n-b}{k-1} - \sum_{i=1}^{b} \binom{n-b-(k-1)(i-1)}{k-1} + b - 1 \\ &= \binom{n}{k} - b\binom{n}{k-1} + \sum_{i=1}^{b} i\binom{n}{k-2} - b^2\binom{n}{k-2} \end{aligned}$$

$$
\begin{aligned}
&\quad + \sum_{i=1}^{b} (b + (k-1)(i-1)) \binom{n}{k-2} + O(n^{k-3}) \\
&= \binom{n}{k} - b\binom{n}{k-1} + \left(\frac{b(b+1)}{2} - b^2 + b^2 + \frac{b(b-1)}{2}(k-1) \right) \binom{n}{k-2} \\
&\quad + O(n^{k-3}) \\
&= \binom{n}{k} - b\binom{n}{k-1} + \left(b + \frac{b(b-1)k}{2} \right) \binom{n}{k-2} + O(n^{k-3}). \qquad (2.1)
\end{aligned}
$$

Thus, $\mathcal{E}(k,n,b)$ is missing

$$
b\binom{n}{k-1} - \left(b + \frac{b(b-1)k}{2} \right) \binom{n}{k-2} + O(n^{k-3}) \qquad (2.2)
$$

sets.

We are now able to check the claim in Theorem 1.5 that when $k = 3$, then $\mathcal{E}'(3,n,b)$ is actually larger than $\mathcal{E}(3,n,b)$. When $k = 3$ we have

$$
\begin{aligned}
|\mathcal{E}(3,n,b)| &= \binom{n-b}{3} + b\binom{n-b}{2} - \sum_{i=1}^{b} \binom{n-b-2(i-1)}{2} + b - 1 \\
&= \binom{n-b}{3} + b\binom{n-b}{2} + b - 1 \\
&\quad - \sum_{i=1}^{b} \frac{(n-b)(n-b-1) - 2(2n-2b-1)(i-1) + 4(i-1)^2}{2} \\
&= \binom{n-b}{3} + b\binom{n-b}{2} - \frac{b(n-b)(n-b-1)}{2} + \frac{(2n-2b-1)(b-1)b}{2} \\
&\quad - \frac{(b-1)b(2b-1)}{3} + b - 1 \\
&= \binom{n-b}{3} + b\binom{n-b}{2} - \frac{1}{2}bn^2 + \frac{1}{2}(4b^2-b)n + \frac{1}{6}(-13b^3 + 6b^2 + 7b - 6).
\end{aligned}
$$

We can also compute

$$
\begin{aligned}
|\mathcal{E}'(3,n,b)| &= \binom{n-b}{3} + b\binom{n-b}{2} - b\binom{n-2b+1}{2} \\
&= \binom{n-b}{3} + b\binom{n-b}{2} - \frac{1}{2}bn^2 + \frac{1}{2}(4b^2-b)n + (-2b^3 + b^2).
\end{aligned}
$$

Thus, the difference in sizes is

$$
|\mathcal{E}'(3,n,b)| - |\mathcal{E}(3,n,b)| = \frac{1}{6}(b^3 - 7b + 6),
$$

which is strictly positive whenever $b > 2$.

2.2. Shifting

The operation of shifting was originally defined by Erdős, Ko, and Rado [3], and has been useful for many different problems related to the intersections of families of sets. We refer the reader to Frankl and Tokushige's survey [9], as well as Frankl's older survey [4], for an exposition of the use of shifting on various problems.

For a family $\mathcal{F}$ and vertices x, y such that $x \neq y$ and $|\mathcal{F}(x)| \geqslant |\mathcal{F}(y)|$, we will define the *shift operator* $S_{x,y}$ on $\mathcal{F}$. For each $F \in \mathcal{F}$ define the shift of F as follows:

$$S_{x,y}(F) = \begin{cases} (F \setminus \{y\}) \cup \{x\}, & \text{if } y \in F, x \notin F, \text{ and } (F \setminus \{y\}) \cup \{x\} \notin \mathcal{F}, \\ F, & \text{else} \end{cases}$$

Then, the shift of $\mathcal{F}$ is defined as

$$S_{x,y}(\mathcal{F}) = \{S_{x,y}(F) : F \in \mathcal{F}\}.$$

Intuitively, the shift can be thought of as changing y to x in all sets of $\mathcal{F}$ such that doing so would not create a collision with another set already in $\mathcal{F}$. We say that the shift is *meaningful* if $S_{x,y}(\mathcal{F}) \neq \mathcal{F}$.

Note that shifting is more commonly applied for all x, y such that $x < y$, rather than the condition $|\mathcal{F}(x)| \geqslant |\mathcal{F}(y)|$ that we use. Our redefinition here will be convenient for us because we will frequently permute the vertices in $[n]$, and would like to maintain an "ordering" for shifting which is independent of these permutations. This redefinition does not have an impact on any of the main properties of shifting, as the proofs still follow through.

Shifting is a useful operation primarily due to two properties: that it never creates a perfect matching and that it can only be performed finitely many times. Specifically, we have the following two facts, whose proofs can be found in [4] and [6], respectively.

Fact 2.1. *If $\mathcal{F}$ has no matching of size r, then $S_{x,y}(\mathcal{F})$ also has no matching of size r. In particular, letting $r = s$, shifting preserves the property of having no perfect matching.*

Fact 2.2. *If $\mathcal{F}' = S_{x,y}(\mathcal{F})$ is a meaningful shift of $\mathcal{F}$, then*

$$\sum_{i=1}^{n} |\mathcal{F}'(i)|^2 > \sum_{i=1}^{n} |\mathcal{F}(i)|^2.$$

From the above monovariant, we deduce that only a finite number of meaningful shifts can be performed on any family (even if the vertices of $[n]$ are permuted in between shifts).

Now, we say a family $\mathcal{F}$ is *shifted* on a set of vertices Y if $|\mathcal{F}(x)|$ is decreasing in x for all $x \in Y$, and there are no meaningful shifts that can be performed on $\mathcal{F}$ which use two vertices in Y. Shifted families are useful because they have a lot of structure, due to the following fact which follows directly from the definition of a shifted family.

Fact 2.3. *Suppose $\mathcal{F}$ is shifted on Y. Then, if $F \in \mathcal{F}$ contains y but not x, where $x < y$ and $x, y \in Y$, then $(F \setminus \{y\}) \cup \{x\} \in \mathcal{F}$.*

Eventually, we will be able to show that any large enough $\mathcal{F}$ satisfying the conditions of Theorem 1.6 can be made to be shifted on a large number of vertices, and we will then be able to use this structure to deduce a lot about $\mathcal{F}$.

2.3. Blocking sets have no covering matching

Before beginning the proof of Theorem 1.6, we prove one last lemma. The following lemma states that if $\mathcal{F}$ is large enough, then a small blocking set of $\mathcal{F}$ cannot be covered by any matching (even a matching with size smaller than the size of the blocking set). This fact will be useful several times.

Lemma 2.4. *Suppose that $|\mathcal{F}| > |\mathcal{E}(k,n,b)|$ and let B be a blocking set of $\mathcal{F}$, where $|B| \leqslant s/2$. Then there is no matching that covers B.*

Proof. Suppose otherwise, and let the matching M cover B. Furthermore let M be of maximal size; note that its size must be less than $|B|$ since B is a blocking set. Then $\mathcal{F}$ cannot contain a set disjoint from M, since otherwise we could append that set to M to get a larger matching covering B. M contains at most $k|B| \leqslant n/2$ vertices, so the number of sets missing from $\mathcal{F}$ is at least

$$\binom{n/2}{k} = \Omega(n^k),$$

so $\mathcal{F}$ is missing more sets than $\mathcal{E}(k,n,b)$ (since the number of missing sets of $\mathcal{E}(k,n,b)$ is given by (2.2)), a contradiction. □

3. Making $\mathcal{F}$ shifted on most of $[n]$

We will now begin with the proof of Theorem 1.6. We proceed by induction on b. The base case $b = 1$ is Kleitman's result in [10]. Assume true for all smaller values of b, and also suppose for the sake of contradiction that there exists a family $\mathcal{F}$ of k-subsets of $[n]$ with no perfect matching and no blocking set of size less than b such that $|\mathcal{F}| > |\mathcal{E}(k,n,b)|$. Further assume that $\mathcal{F}$ is of maximum possible size.

In this section we will prove the following proposition, which implies that we can assume that $\mathcal{F}$ is shifted on all but a constant number of vertices. Our proof of this proposition is based on the proof of Proposition 4.4 in [6].

Proposition 3.1. *There exists a constant $c = c(k,b)$ such that it is possible to apply shifts to $\mathcal{F}$ and permute the vertices in $[n]$ so that $\mathcal{F}$ has no blocking set of size less than b and is shifted on $[n-c]$.*

Proof. Suppose that $\mathcal{F}$ has been shifted as much as possible while maintaining the property that $\mathcal{F}$ has no blocking set of size less than b. Thus, every meaningful shift of $\mathcal{F}$ creates a blocking set of size less than b.

Permute the vertices, sorting by the size of $\mathcal{F}(i)$, so that $\mathcal{F}(1)$ has the largest size. Then, let $A = [n-2b, n]$ be the last $2b+1$ vertices in $[n]$. Note that no b'-subset of A, for $b' < b$, can form a blocking set, so for each such b'-subset there is a matching of b' sets which covers it. Let T consist of all the vertices in $[n]$ which are covered by any such matching. We then have $|T| \leqslant 2^{2b+1}(b-1)k$.

We now claim that $\mathcal{F}$ is shifted on $[n] \setminus T$. Suppose otherwise, so that there is a meaningful shift $S_{x,y}$, where $x < y$ and $x, y \notin T$. By assumption, this means that the shifted family

$\mathcal{F}' = S_{x,y}(\mathcal{F})$ must contain a blocking set B of size b', for some $b' < b$. We may furthermore pick B so that b' is minimal. Note that any b'-subset of A is covered by a matching which is contained entirely in T, and thus does not contain y, and is therefore not affected by the shift $S_{x,y}$. Thus, every b'-subset of A is still not a blocking set of $\mathcal{F}'$, so B must contain a vertex not in A. We show that every vertex in B is contained in a small number of sets of $\mathcal{F}'$.

Lemma 3.2. *For every $z \in B$, $|\mathcal{F}'(z)| = O(n^{k-2})$.*

Proof. Note that by Theorem 2.4, there is no matching that covers B. Thus we may add all k-sets disjoint from B to $\mathcal{F}'$, and B will still be a blocking set of minimal size. By the inductive hypothesis, after adding these sets, the size of $\mathcal{F}$ will be at most

$$\binom{n}{k} - b'\binom{n}{k-1} + O(n^{k-2}),$$

by (2.1). Then, since none of the sets which are disjoint from B contain z, we have

$$|\mathcal{F}'(z)| \leqslant \binom{n}{k} - b'\binom{n}{k-1} + O(n^{k-2}) - \binom{n-b'}{k} = O(n^{k-2}),$$

as desired. □

Thus, since B contains a vertex not in A, we have some vertex $z \in [n-2b-1]$ so that $|\mathcal{F}'(z)| = O(n^{k-2})$. If $z \neq y$, then

$$|\mathcal{F}(z)| \leqslant |\mathcal{F}'(z)| = O(n^{k-2}).$$

If on the other hand $z = y$, then

$$|\mathcal{F}(x)| + |\mathcal{F}(y)| = |\mathcal{F}'(x)| + |\mathcal{F}'(y)| \leqslant \binom{n}{k-1} + O(n^{k-2}),$$

so either $|\mathcal{F}(x)|$ or $|\mathcal{F}(y)|$ is at most $\frac{1}{2}\binom{n}{k-1} + O(n^{k-2})$. In any case, we have some $z' \in [n-2b-1]$ such that $|\mathcal{F}(z')| \leqslant \frac{1}{2}\binom{n}{k-1} + O(n^{k-2})$. By monotonicity of $|\mathcal{F}(i)|$, we also have that $|\mathcal{F}(w)| < |\mathcal{F}(z')|$ for all $w \in A$. Thus, the total number of sets in $\mathcal{F}$ which intersect A is at most

$$\sum_{w \in A} |\mathcal{F}(w)| \leqslant \frac{2b+1}{2}\binom{n}{k-1} + O(n^{k-2}).$$

Then, since there are at most $\binom{n-2b-1}{k}$ sets not intersecting A, the total size of $\mathcal{F}$ is at most

$$\begin{aligned}\binom{n-2b-1}{k} + \frac{2b+1}{2}\binom{n}{k-1} + O(n^{k-2}) &= \binom{n}{k} - \frac{2b+1}{2}\binom{n}{k-1} + O(n^{k-2}) \\ &< |\mathcal{E}(k,n,b)|.\end{aligned}$$

This is a contradiction, so we must indeed have that $\mathcal{F}$ is shifted on $[n] \setminus T$. Permuting the vertices so that T is at the end of $[n]$, we have that $\mathcal{F}$ is shifted on $[n-c]$, as desired, concluding the proof of Theorem 3.1. □

We henceforth assume that $\mathcal{F}$ is shifted on $[n-c]$.

4. $\mathcal{F}$ has a blocking set of size exactly b

In this section we will show that $\mathcal{F}$ has a blocking set of size exactly b. Suppose for a contradiction otherwise, and let B be the smallest blocking set of $\mathcal{F}$, where $|B| = b' > b$. We will split into cases based on whether B is small or large.

4.1. B small

Here assume that $b' \leqslant s/2 = n/2k$.

Claim 4.1. *$\mathcal{F}$ contains a set of the form $\{z\} \cup C$, where $z \in B$ and $C \subset [b'k+1, n-c]$.*

Proof. For any fixed $z \in B$, the number of possible such sets $\{z\} \cup C$ is at least the number of such sets which contain no vertex in B other than z. Thus the total number of possible $\{z\} \cup C$ is at least

$$b'\binom{n-c-b'k-b'}{k-1} \sim \frac{b'(n-(k+1)b')^{k-1}}{(k-1)!}, \tag{4.1}$$

where we have used that $n-(k+1)b' \geqslant (1-(k+1)/2k)n = \omega(1)$.

Now, if $b' \leqslant \sqrt{n}$, then the quantity in (4.1) is at least

$$(1+o(1))(b+1)n^{k-1}/(k-1)!$$

(since $b' > b$). If on the other hand $b' > \sqrt{n}$, then the same quantity is at least

$$(1+o(1))\frac{\sqrt{n}((1-(k+1)/2k)n)^{k-1}}{(k-1)!} = \Omega(n^{k-1/2}).$$

In either case, the number of such sets $\{z\} \cup C$ is greater than the number of missing sets of $\mathcal{F}$, which is at most the quantity (2.2). Thus, $\mathcal{F}$ must contain some such set $\{z\} \cup C$. $\square$

Consider a matching M with $b'-1$ sets covering $B \setminus \{z\}$ (this exists since $B \setminus \{z\}$ is not a blocking set). By Theorem 2.4, M cannot contain z. Then this matching covers at most $(b'-1)k$ vertices from $[b'k]$, so it is possible by repeated application of Theorem 2.3 to shift $\{z\} \cup C$ to a set $\{z\} \cup C' \in \mathcal{F}$ so that C' is disjoint from M. But then we can add $\{z\} \cup C'$ to M to get a matching covering B, a contradiction to Theorem 2.4.

Thus we conclude that the smallest blocking set B cannot have size b', where $b < b' \leqslant s/2$.

4.2. B large

Assume now that $b' > s/2$. We will then find a perfect matching in $\mathcal{F}$ in order to derive a contradiction.

Lemma 4.2. *$\mathcal{F}$ contains every k-vertex subset of $[n-b'+1]$.*

Proof. By repeated application of Theorem 2.3, it is enough to show that $\mathcal{F}$ contains some subset of $[n-b'+2, n-c]$. But if this is not the case, then the number of missing sets of $\mathcal{F}$ is at least

$$\binom{b'-c-1}{k} \geqslant \binom{s/2-c-1}{k} = \Omega(n^k),$$

a contradiction. □

Now, since there are no blocking sets of size smaller than b', there is a matching M in $\mathcal{F}$ which covers $[n-b'+2, n]$. By Theorem 4.2, we can then add an arbitrary matching on the vertices not covered by M to obtain a perfect matching in $\mathcal{F}$, a contradiction.

We may thus conclude that B also cannot have size $b' > s/2$. Therefore, combining the results from the two cases, the smallest blocking set of $\mathcal{F}$ has size exactly b.

5. The $k = 2$ case

We will now take a brief digression from the proof of Theorem 1.6 to address the case where $k = 2$, i.e., where $\mathcal{F}$ is an ordinary graph. First, we define the optimal family as follows. Define

$$\mathcal{E}''(2,n,b) = \{\{u,v\} : u \in [b], v \in [b+1, 2b-1]\} \cup \{\{u,v\} : u,v \in [b+1, n]\}.$$

In other words, $\mathcal{E}''(2,n,b)$ is the graph which consists of the complete graph on $[b+1,n]$, in addition to all edges between $[b]$ and $[b+1, 2b-1]$. Note that $[b]$ is a blocking set of this graph. Also, note that the total number of edges is $\binom{n-b}{2} + (b-1)b$, which is more than the number of edges in $\mathcal{E}(2,n,b)$. We will now show that this family is optimal.

Proposition 5.1. *Suppose $n = 2s$ is even, and let b be a positive integer. Then there exists $s_0(b)$ such that whenever $s \geqslant s_0(b)$, then the following holds: if $\mathcal{F}$ is a graph on $[n]$ with no perfect matching and no blocking set of size less than b, then*

$$|\mathcal{F}| \leqslant |\mathcal{E}''(2,n,b)|.$$

Proof. Let $\mathcal{F}$ be such a graph with the maximum possible number of edges. The results from the previous sections still apply in the case $k = 2$, so we can assume that $\mathcal{F}$ has a blocking set B of size exactly b, but still no smaller blocking set.

Consider a vertex $v \in B$. We claim that $\deg v \leqslant b-1$. To see this, first note that since $B \setminus \{v\}$ is not a blocking set, there is a matching M which covers $B \setminus \{v\}$. Now, consider an arbitrary neighbor u of v. If u is not a neighbor in M of any element of $B \setminus \{v\}$, then note that we can add the edge $\{u,v\}$ to M, thus covering all of B. It might be the case that u was already in an edge in M, but the other vertex is not in $B \setminus \{v\}$, but we can remove that edge and still have a matching covering B. This is a contradiction to the fact that B is a blocking set, so we must thus have that any neighbor of v must be adjacent in M to a vertex in $B \setminus \{v\}$. But there are only $b-1$ such vertices, so $\deg v \leqslant b-1$.

Now, to bound the total size of $\mathcal{F}$, note that the number of edges which contain a vertex in B is at most the total degree of B, which is $(b-1)b$. On the other hand, the number of edges which are disjoint from B is at most $\binom{n-b}{2}$. Thus, the total number of edges in $\mathcal{F}$ is at most $\binom{n-b}{2} + (b-1)b = |\mathcal{E}''(2,n,b)|$, as desired. □

6. Reduction to graph problem

We now continue with the proof of Theorem 1.6. In this section we will reduce the problem of proving Theorem 1.6 to proving a simpler statement about graphs. The techniques we use to do this reduction are closely related to the technique of juntas, which were introduced by Dinur and Friedgut in [1], and more recently used in [8] to prove results related to the Erdős matching conjecture. Specifically, it will turn out that a maximal family $\mathcal{F}$ can be well-approximated by a junta, and that the highest-order contributions come from 2-element sets in the junta.

Permute the vertices in $[n]$ again, so that $[b]$ is a blocking set of $\mathcal{F}$ and $\mathcal{F}$ is shifted on $[b+c+1,n]$. We show the following fact which allows us to ignore most vertices in $[n]$:

Fact 6.1. *Suppose M is a collection of at most b k-sets in $\mathcal{F}$ which covers $[b]$ and such that no two sets in M both contain the same vertex in $[b+c+bk]$. Then, there exists a matching in $\mathcal{F}$ which covers $[b]$.*

Proof. For $F \in M$, write $F = F_0 \sqcup F_1$, where $F_0 = F \cap [b+c+bk]$ and $F_1 = F \cap [b+c+bk+1,n]$. Then, by repeated application of Theorem 2.3, we also have $F_0 \sqcup F_1' \in \mathcal{F}$ for any $F_1' \subset [b+c+1, b+c+bk]$ which has the same size as F_1 and is disjoint from F_0. We can furthermore pick F_1' to be disjoint from all of M, since M contains at most bk vertices in total. Thus, we can replace F with $F_0 \sqcup F_1'$ in M. Doing this successively to all $F \in M$, we get that M becomes a matching. Note that all vertices in $[b]$ were in F_0 and not in F_1 at each of these steps, so M still covers $[b]$. □

Now, define $\mathcal{G}$ to be the set of all sets S which contain at least one vertex in the blocking set $[b]$ and such that there exists a set $T \subset [b+c+bk+1,n]$ satisfying $S \cup T \in \mathcal{F}$. (Note that all sets in $\mathcal{F}$ which intersect $[b]$ are automatically in $\mathcal{G}$.) By Theorem 6.1, if there is a matching in $\mathcal{G}$ which covers $[b]$, then there is also a matching in $\mathcal{F}$ which covers $[b]$, and there can be no such matching by Theorem 2.4. (Though $\mathcal{G}$ contains sets of different sizes, a matching is similarly defined to be a subcollection of disjoint sets.) Thus, there is no matching in $\mathcal{G}$ which covers $[b]$.

Also, note that $\mathcal{G}$ contains no single-element sets: if it contained the single-element set $\{z\}$, then there would be no matching in $\mathcal{G}$ (and thus no matching in $\mathcal{F}$) covering $[b] \setminus \{z\}$, and thus $[b] \setminus \{z\}$ would be a blocking set of size $b-1$, a contradiction.

Now, any missing set of size r in $\mathcal{G}$ corresponds to roughly $\binom{n}{k-r} = \Theta(n^{k-r})$ missing sets in $\mathcal{F}$. Since $\mathcal{G}$ has no 1-sets, the highest-order contribution comes from the 2-sets in $\mathcal{G}$. Thus we will be interested in bounding the number of 2-sets in $\mathcal{G}$. To this end, define the graph G on the vertices $[n]$ with an edge between any two vertices whenever the set containing those two vertices is in $\mathcal{G}$. (Note that every edge in G intersects $[b]$.) Without loss of generality (by permuting $[b]$) assume that the vertex 1 has the highest degree in G out of any vertex in $[b]$.

Now, $\mathcal{F}$ must contain a matching which covers $[2,b]$ (and which contains only sets touching $[2,b]$) since it is not a blocking set. But every set in $\mathcal{F}$ which intersects $[b]$ is also a set in $\mathcal{G}$. Thus, there exists a matching $S_1, \dots, S_a$ of sets in $\mathcal{G}$ which covers $[2,b]$ (note that it cannot contain the vertex 1). We will now use the following proposition about the graph G.

Proposition 6.2. *Suppose that either $k \geqslant \max\{b+1, 6\}$ or $k \geqslant 100$. Let G be a (simple, undirected) graph on $[n]$ such that all edges of G touch $[b]$, and 1 has the greatest degree out of* *any vertex in $[b]$. Let there also be a collection of pairwise disjoint k-sets $\alpha_1, \ldots, \alpha_a \subset [n]$ which each intersect $[b]$ and whose union contains all vertices in $[2, b]$. Suppose that it is impossible to cover $[b]$ by a matching which consists only of edges of G and some of the α_i. Then, the total number of edges in G is at most $\frac{b(b-1)}{2}(k-1)$.*

Moreover, equality holds only in the following case: $a = b-1$, and the k-sets $\alpha_1, \ldots, \alpha_a$ each contain only one element of $[2, b]$. Up to permutation of the vertices in $[2, b]$, there is an edge from i to every element of the k-set that contains j, except j itself, for every $1 \leqslant i < j \leqslant b$. These are the only edges in G. This equality case is illustrated in Figure 6.1.

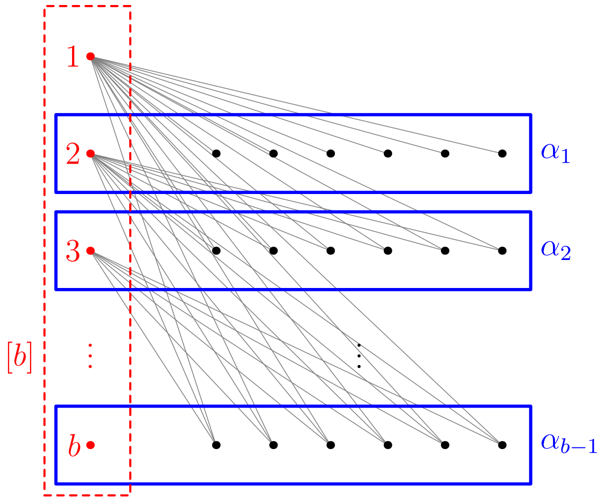

Figure 6.1: The only equality case of Theorem 6.2, up to permutation of the vertices in $[2, b]$.

We will prove Theorem 6.2 in the next section, but first we will show how it can be used to deduce Theorem 1.6. By setting $\alpha_i = S_i$ and applying Theorem 6.2 to the graph G as defined above, we may conclude that G has at most $\frac{b(b-1)}{2}(k-1)$ edges, and we also know what the equality case is. We claim that equality must in fact hold. For a contradiction, suppose that G has strictly fewer than $\frac{b(b-1)}{2}(k-1)$ edges.

First note that $\mathcal{F}$ is missing every set of the form $\{z\} \cup T$, where $z \in [b]$ and $T \subset [b+c+bk+1, n]$. The number of such missing sets is

$$b\binom{n-b-c-bk}{k-1}.$$

Furthermore, $\mathcal{F}$ is also missing every set of the form $\{z, w\} \cup T$, where $z \in [b]$, $w \in [b+c+bk]$, $\{z, w\}$ is not an edge in G, and $T \subset [b+c+bk+1, n]$. The number of such edges $\{z, w\}$ is

$$\frac{b(b-1)}{2} + b(c+bk) - |E(G)| \geqslant b(c+bk) - \frac{b(b-1)}{2}(k-2) + 1.$$

This causes the following additional number of missing sets of $\mathcal{F}$:

$$\left(b(c+bk) - \frac{b(b-1)}{2}(k-2) + 1\right)\binom{n-b-c-bk}{k-2}.$$

Therefore, the total number of missing sets of $\mathcal{F}$ is at least

$$\begin{aligned}
&b\binom{n-b-c-bk}{k-1} + \left(b(c+bk) - \frac{b(b-1)}{2}(k-2)+1\right)\binom{n-b-c-bk}{k-2} \\
&= b\left(\binom{n}{k-1} - (b+c+bk)\binom{n}{k-2} + O(n^{k-3})\right) \\
&\qquad + \left(b(c+bk) - \frac{b(b-1)}{2}(k-2)+1\right)\binom{n}{k-2} + O(n^{k-3}) \\
&= b\binom{n}{k-1} - \left(b + \frac{b(b-1)k}{2} - 1\right)\binom{n}{k-2} + O(n^{k-3}),
\end{aligned}$$

which is more than the number of missing sets of $\mathcal{E}(k,n,b)$ (as seen in (2.2)), a contradiction.

Therefore, equality must hold in Theorem 6.2, so $\mathcal{G}$ must contain 2- and k-sets of the prescribed form. Now, note that for any 2-set $E \in \mathcal{G}$, we can add $E \cup S$ to $\mathcal{F}$ for any set S disjoint from $[b]$ and E and such that $|S| = k-2$. Recall that $\mathcal{G}$ has no matching which covers $[b]$ and $\mathcal{G}$ also contains all sets in $\mathcal{F}$; thus, $[b]$ is still a blocking set of $\mathcal{F}$ after these sets are added. Similarly we can add all k-subsets of $[b+1,n]$ to $\mathcal{F}$. Finally note that $\mathcal{F}$ contains the sets $\alpha_1,\dots,\alpha_{b-1}$. But now, note that $\mathcal{F}$ contains all the sets in $\mathcal{E}(k,n,b)$, where in the construction of $\mathcal{E}(k,n,b)$ in Theorem 1.3, the sets E_i are chosen to be $E_i = \alpha_i \setminus \{i\}$. But then by the maximality of $\mathcal{E}(k,n,b)$, we have that $\mathcal{F}$ cannot contain any more sets, so $\mathcal{F} = \mathcal{E}(k,n,b)$, concluding the proof of Theorem 1.6.

7. Proof of graph problem

We will now prove Theorem 6.2. We proceed by induction on b. The cases $b = 1, 2$ are easy to check, so let $b \geqslant 3$ and assume Theorem 6.2 holds for all smaller values of b. Suppose that G satisfies the conditions of Theorem 6.2 and has at least $\frac{b(b-1)}{2}(k-1)$ edges.

First we introduce some notation. We call the vertices in $[b]$ *special*. We refer to the k-sets $\alpha_1,\dots,\alpha_a$ as *cells* (so that there are a cells), and if a cell contains two or more special vertices we say that the special vertices it contains are *siblings*. On the other hand, if a cell contains only one special vertex, we say that special vertex is *lonely*. Also, define the *exterior*, R, to be the set of vertices which are not 1 and are also not in any set α_i.

We also define the edge count $e(A,B)$, where A and B are sets of vertices, to be the number of edges between A and B. We will sometimes abuse notation by writing a single vertex as an argument of e instead of the set containing only that vertex.

Also, we note that G has no edge connecting the vertex 1 to any vertex which is not in some α_i – otherwise that edge would form a covering matching of $[b]$ along with all the sets $\alpha_1,\dots,\alpha_a$. Also, there is no edge connecting 1 to any lonely vertex, since otherwise that edge would form a covering matching along with the sets α_i which do not contain that vertex.

We now make use of the following lemma of graph theory, which is a generalization of Hall's marriage theorem.

Lemma 7.1. *Let G be a graph, and let A be a set of vertices of G. Suppose that G has no* *matching which covers A. Then, there exists a set $S \subset A$ such that $|N(S)| \leqslant 2|S| - 1$. (Here $N(S)$ is the set of neighbors of S, including all points of S itself.)*

Proof. Let M be a matching in G which contains the maximum possible number of vertices in A; we may assume that M contains no edges disjoint from A. By assumption, G contains a vertex v which is not in M.

We construct S in steps, along with a set T which will eventually contain all neighbors of S, except those in S itself. Initially, let S contain only the vertex v. While there exists any edge from an element of S to another vertex u which is not in S or T, if u has a neighbor in M which is also in A, add u to T and the neighbor to S. (Note that whenever we add one vertex each to S and T, they are adjacent in M. Thus, the neighbor of u in M cannot already be in S or T.)

Now, suppose that after this process has ended, there is still an edge from a vertex $w \in S$ to a vertex $u \notin S, T$. Note that u cannot be adjacent in M to any vertex in A, otherwise it would have previously been added to T. Now by construction, there is an even-length path connecting v to w, where every other vertex (including v and w) is in S, and the rest are in T, and every other edge is an edge of M (so that the last edge touching w is in M, but the edge touching v is not). Then, we can flip the parity of the edges on the path which are in M, so that the edge touching v is now in M, but the edge touching w is not. Now, M contains v but does not contain w, but otherwise contains the same vertices. Finally, add the edge $\{w, u\}$ to M, and remove any edge touching u in M; since an edge touching u in M cannot contain any vertex in A, this does not remove any vertex of A from M. Thus, M now covers all the vertices in A which it originally covered, in addition to v, contradicting its original maximality.

Thus, there is no edge from any vertex in S to a vertex not in S or T. Therefore, $N(S) = S \cup T$. Since one vertex is added to S for each vertex added to T, we have $|T| = |S| - 1$, so $|N(S)| = 2|S| - 1$ as desired. □

Using this lemma on G with $A = [b]$, we can find a set $S \subset [b]$ such that $|S| = d$ and $|N(S)| \leqslant 2d - 1$. If $1 \in S$, then since the vertex 1 has the maximal degree out of $[b]$, this means that there are at most

$$b(2d-1) \leqslant b(2b-1) \leqslant \frac{5b(b-1)}{2} \leqslant \frac{b(b-1)}{2}(k-1)$$

edges in G. Equality cannot hold since that would require $b = d$, and then we would be double-counting the edge between the vertices 1 and 2. Thus we are immediately done if $1 \in S$. We will assume henceforth that $1 \notin S$.

Now let Γ denote the set of cells α_i which intersect S, and let $d' = |\Gamma|$. Clearly we have $d' \leqslant d$. Also, let T denote the set of special vertices which are siblings of some vertex in S. Then, $S \cup T$ is the set of all special vertices in the cells of Γ. Let $\ell = |T|$. Each cell α_i contains a vertex which is not in T (and which is also not 1), so we have

$$a \leqslant b - \ell - 1. \tag{7.1}$$

Now, we wish to induct by removing the cells in Γ. The idea is that since S has few neighbors and therefore low total degree, we will not be removing too many edges. After removing the cells in Γ and all edges connected to vertices in these cells, there will still be no matching that covers the remaining special vertices, since otherwise that matching, along with the cells of Γ, would cover $[b]$. Thus, as long as the vertex 1 still has the highest degree, we will be able to apply the inductive hypothesis.

We delete the edges touching Γ in a specific order that will allow us to more easily bound the number of deleted edges. Specifically, we delete the edges in four steps as follows.

1. Delete all edges touching S. Let X be the set of deleted edges. Since $|N(S)| \leqslant 2d-1$, we have
$$|X| \leqslant d(d-1) + \binom{d}{2} = \frac{3d(d-1)}{2}. \tag{7.2}$$

2. Delete all remaining edges from $[b] \setminus (S \cup T)$ to Γ. Note that this deletes at most $d'k - d$ edges from each vertex in $[b] \setminus (S \cup T)$.

3. If vertex 1 is no longer the maximum-degree vertex in $[b] \setminus (S \cup T)$, then arbitrarily delete edges from any such vertices with higher degree (except those touching 1) until 1 is again the maximum-degree vertex in $[b] \setminus (S \cup T)$. Let Y be the set of edges deleted in steps 2 and 3 combined.

   Let p be the total number of elements of S which are adjacent to the vertex 1. Then, the number of edges touching 1 which were deleted in steps 1 and 2 combined is at most $d'k - d + p$. This means that Y contains at most $d'k - d + p$ edges touching each vertex in $[b] \setminus (S \cup T)$. Thus,
$$|Y| \leqslant (b-d-\ell)(d'k - d + p). \tag{7.3}$$

   We will now bound p. Note that a cell $\gamma \in \Gamma$ which contains t vertices of S contributes at most t to p, but only if γ contains a sibling, since 1 is not adjacent to any lonely vertices. If $t \geqslant 2$, then γ also contributes at least $t-1 \geqslant t/2$ to the quantity $d - d'$. Otherwise, $t = 1$ and γ must contain a special vertex not in S, in which case γ contributes 1 to ℓ. Therefore, we have the inequality
$$p \leqslant 2(d-d') + \ell.$$

   Thus, the above bound on Y becomes
$$\begin{aligned} |Y| &\leqslant (b-d-\ell)(d'k - d + 2(d-d') + \ell) \\ &= (b-d-\ell)(d'(k-2) + d) + \ell(b-d-\ell) \\ &\leqslant (b-d-\ell)d(k-1) + \ell(b-d-\ell). \end{aligned} \tag{7.4}$$

4. Delete all remaining edges touching T. Note that since 1 was the vertex of maximum degree, this deletes at most $\ell \deg(1)$ edges (where the degree is calculated before any edges were deleted). Let r be the total number of cells α_i which contain siblings. Since 1 does not have an edge to any lonely vertex, it has at most $k-1$ edges to each cell except for these r cells, so
$$\deg(1) \leqslant a(k-1) + r.$$

Now, each of the r cells which contain siblings contain at least 2 special vertices (while all the others still contain 1), so since $b-1$ is the number of special vertices which are contained in some cell, this means that $b-1 \geqslant 2r+(a-r)$. Rearranging, $r \leqslant b-a-1$. Therefore, using this and (7.1), the above bound then becomes

$$\begin{aligned}
\deg(1) &\leqslant a(k-1)+r && (7.5)\\
&\leqslant a(k-1)+(b-a-1)\\
&= a(k-2)+b-1\\
&\leqslant (b-\ell-1)(k-2)+b-1 && (7.6)\\
&= (b-\ell)(k-1)+\ell-(k-1).
\end{aligned}$$

Thus, if Z is the set of edges deleted in this step, then

$$|Z| \leqslant \ell \deg(1) \leqslant \ell((b-\ell)(k-1)+\ell-(k-1)). \tag{7.7}$$

Now, we will apply the inductive hypothesis as described earlier. Note that all vertices touching Γ are deleted in steps 1, 2, and 4, and step 3 ensures by deleting some more edges that vertex 1 still has maximal degree. Thus the inductive hypothesis can be applied; since there are $b-d-\ell$ special vertices remaining, the number of edges remaining is at most $\frac{(b-d-\ell)(b-d-\ell-1)}{2}(k-1)$. Then, using (7.2), (7.4), and (7.7), the total number of edges originally in G is at most

$$\begin{aligned}
&\frac{(b-d-\ell)(b-d-\ell-1)}{2}(k-1)+|X|+|Y|+|Z|\\
&\leqslant \frac{(b-d-\ell)(b-d-\ell-1)}{2}(k-1)+\frac{3d(d-1)}{2}+(b-d-\ell)d(k-1)+\ell(b-d-\ell)\\
&\qquad +\ell((b-\ell)(k-1)+\ell-(k-1))\\
&\leqslant \frac{(b-d-\ell)(b-d-\ell-1)}{2}(k-1)+\frac{d(d-1)}{2}(k-1)+(b-d-\ell)d(k-1)\\
&\qquad +\ell((b-\ell)(k-1)+\ell-(k-1)+(b-d-\ell)) && (7.8)\\
&= \frac{(b-\ell)(b-\ell-1)}{2}(k-1)+\ell(b-\ell)(k-1)+\ell(\ell-(k-1)+(b-d-\ell))\\
&= \frac{b(b-1)}{2}(k-1)-\frac{\ell(\ell-1)}{2}(k-1)+\ell(\ell-(k-1)+(b-d-\ell))\\
&= \frac{b(b-1)}{2}(k-1)-\frac{\ell(\ell-1)}{2}(k-3)+\ell(-(k-2)+(b-d-\ell))\\
&\leqslant \frac{b(b-1)}{2}(k-1)-\frac{\ell(\ell-1)}{2}(k-3)+\ell(-(k-2)+(b-1-1))\\
&= \frac{b(b-1)}{2}(k-1)-\frac{\ell(\ell-1)}{2}(k-3)+\ell(-k+b). && (7.9)
\end{aligned}$$

If $k \geqslant b+1$, then this is at most $\frac{b(b-1)}{2}(k-1)$, the desired bound. To complete the proof of the $k \geqslant \max\{b+1,6\}$ case, it remains to check that the equality case is as described in Theorem 6.2.

To this end, suppose equality holds. In order to achieve equality in the very last step, we must have $\ell = 0$. Also, in order to have equality at (7.8), we also have $d = 1$. Without loss of generality suppose that $S = \{b\}$, and that α_{b-1} contains the vertex b (and no other special vertices, since $\ell = 0$). Also, by the inductive hypothesis, the edges not touching α_{b-1} take exactly the form in Theorem 6.2 (but for $b-1$ instead of b). It remains to check that the edges touching α_{b-1} are exactly those that go from $[b-1]$ to vertices in α_{b-1}, except the vertex b itself. However, note that 1 has no edge to b (since b is lonely), and any vertex i in $[2, b-1]$ cannot have an edge to b since then one could make a matching consisting of the edge from i to b, all the cells except α_{b-1} and the one containing i, and any edge from 1 to the cell containing i (this exists by the inductive hypothesis), and this would form a covering matching of $[b]$. Thus, there are no edges containing b.

On the other hand, equality must hold in (7.3). Since $d' = d = 1$ and $p = 0$ (since all vertices are lonely), this means that in steps 2 and 3, $k-1$ edges were deleted from each vertex in $[b-1]$. Recall that (by the inductive hypothesis) after deletion of edges, every vertex in $[2, b-1]$ has strictly lower degree than 1. Therefore, no edges could have been deleted in step 3, so $k-1$ edges were deleted from each vertex in $[b-1]$ in step 2. Thus, every vertex in $[b-1]$ must have been connected to every vertex in α_{b-1} except b. Finally, note that (7.2) implies that the vertex b has no edges at all. This completes the analysis of the equality case; we have shown that the only deleted edges are those from $[b-1]$ to α_{b-1} (but not to b), and thus the graph G takes exactly the form described in Theorem 6.2. This completes the proof of Theorem 6.2 in the $k \geqslant \max\{b+1, 6\}$ case.

Now, to prove the $k \geqslant 100$ case of Theorem 6.2, we assume henceforth that $100 \leqslant k \leqslant b$ (since $b < k$ is already covered by the previous case). We bound $|Y|$ in a slightly different way; let Φ be the set of cells in Γ whose vertices are all adjacent to 1, and let $q = |\Phi|$. (Note that every cell in Φ must contain siblings, since 1 is not adjacent to any lonely vertex; we will use this fact later). Then, the number of edges touching 1 which were deleted in steps 1 and 2 is at most $d'(k-1) + q$ (since there are at most $k-1$ edges from 1 to any cell whose vertices aren't all adjacent to it). It is still the case that $d'(k-1) + q$ is at least $d'k - d$, which is the maximum number of edges deleted from any vertex in step 2. Thus, as before, there are at most $d'(k-1) + q$ edges deleted from each vertex in $[b] \setminus (S \cup T)$ in steps 2 and 3 combined, so we have the bound

$$|Y| \leqslant (b-d-\ell)(d'(k-1)+q). \tag{7.10}$$

Using the bound (7.10) instead of (7.4), the bound on the number of edges in G becomes

$$\begin{aligned}
&\frac{(b-d-\ell)(b-d-\ell-1)}{2}(k-1) + |X| + |Y| + |Z| \\
&\leqslant \frac{(b-d-\ell)(b-d-\ell-1)}{2}(k-1) + \frac{3d(d-1)}{2} + (b-d-\ell)(d'(k-1)+q) \\
&\qquad + \ell((b-\ell)(k-1) + \ell - (k-1)) \qquad (7.11) \\
&\leqslant \frac{(b-d-\ell)(b-d-\ell-1)}{2}(k-1) + \frac{d(d-1)}{2}(k-1) + (b-d-\ell)(d(k-1)+q) \\
&\qquad + \ell((b-\ell)(k-1) + \ell - (k-1)) \qquad (7.12)
\end{aligned}$$

$$\begin{aligned}
&= \frac{(b-\ell)(b-\ell-1)}{2}(k-1) + \ell(b-\ell)(k-1) + (b-d-\ell)q + \ell(\ell-(k-1)) \\
&= \frac{b(b-1)}{2}(k-1) - \frac{\ell(\ell-1)}{2}(k-3) - \ell(k-2) + (b-d-\ell)q \\
&\leqslant \frac{b(b-1)}{2}(k-1) - \frac{\ell(\ell-1)}{2}(k-3) - \ell(k-2) + bq \\
&\leqslant \frac{b(b-1)}{2}(k-1) + bq. \qquad (7.13)
\end{aligned}$$

Now, if $q = 0$, then again the desired bound is achieved, and it remains to check the equality case. The analysis of the equality case is exactly the same as in the $k \geqslant \max\{b+1, 6\}$ case, and so we are done if $q = 0$.

Thus, suppose henceforth that $q > 0$. We will show that either the bound (7.13) can be reduced by more than bq, or that we can find a matching which covers $[b]$, in both cases obtaining a contradiction. To do the first of these, it is enough to show that any step of the above chain of inequalities has a difference of more than bq.

Now, note that (7.13) is within bq of its assumed lower bound; this means that there is a lot of rigidity in the above chain of inequalities. This means that we will be able to find a lot of structure in G, or else one of the steps in the chain of inequalities will be able to be reduced by more than bq. To this end, we show some claims about G. (All claims that follow are assumed to be stated before any edges in G are deleted.)

Claim 7.2. $q \leqslant d < 0.031b$.

Proof. Clearly $q \leqslant d$, since every cell in Φ (which is a subset of Γ) must have a vertex in S. Suppose for a contradiction that $d \geqslant 0.031b$. Then we also have $d - 1 \geqslant 0.021b$. Then the quantity in (7.12) exceeds the previous quantity by

$$\frac{d(d-1)}{2}(k-4) \geqslant \frac{(k-4)(d-1)}{2}q \geqslant \frac{96(0.021b)}{2}q > bq,$$

since $k, b \geqslant 100$. This is a contradiction. □

Claim 7.3. $\ell < 0.025b$.

Proof. Suppose $\ell \geqslant 0.025b$. Then $\ell - 1 \geqslant 0.015b$. Then, (7.13) exceeds the previous line by at least

$$\frac{\ell(\ell-1)}{2}(k-3) > 0.000375(k-3)b^2 > 0.031b^2 \geqslant bq,$$

a contradiction. □

Claim 7.4. *Fewer than $q/2$ cells contain two siblings both in S.*

Proof. Suppose otherwise. Then we have $d - d' \geqslant q/2$, so we can bound $d'(k-1) + q < d(k-1)$, and thus we can replace the $d(k-1)+q$ term in (7.12) with just $d(k-1)$ (and make the inequality strict). If we follow through the chain of inequalities, this means we can remove the bq term in (7.13), thus obtaining a contradiction (since the inequality was strict). □

Let Φ' then be the set of cells in Φ which contain only one vertex of S, so that $|\Phi'| > q/2$ by the preceding claim.

Corollary 7.5. $\ell > q/2$. *(In particular, ℓ is positive.)*

Proof. Recall that all cells in Φ contain siblings, so every cell in Φ' must contain a vertex in T. Thus $\ell = |T| > q/2$. ☐

Claim 7.6. *Fewer than $0.1b$ special vertices have siblings.*

Proof. Suppose that at least $0.1b$ special vertices have siblings. Note that $b-1-a$ is equal to the sum of $t-1$ over all cells, where t is the number of special vertices in the cell. Since $t-1 \geqslant t/2$ whenever the cell has siblings, this means that $b-1-a \geqslant 0.05b$. Rearranging, $a \leqslant b-0.05b-1$, so the bound (7.6) can be reduced by $(0.05b-\ell)(k-2) > 0.025(k-2)b$. Consequently, the bound (7.7) (and thus also (7.11)) can be reduced by $0.025(k-2)\ell b > 0.0125(k-2)bq > bq$, a contradiction. ☐

Corollary 7.7. $a > 0.89b$.

Proof. Each lonely vertex is in its own cell, so a, the number of cells, is at least the number of lonely siblings, which by Theorem 7.6 is greater than $0.9b - 1 > 0.89b$. ☐

Claim 7.8. *One of the vertices in T has degree greater than $ak - 3b$.*

Proof. Suppose otherwise, and that all vertices in T have degree at most $ak-3b$. Since $a \leqslant b-r$, we have $ak-3b \leqslant a(k-1)-r-2b$, so the bound of (7.5) exceeds the number of edges deleted from each vertex in T by at least $2b$. Thus the bound (7.7) can be reduced by at least $2\ell b > bq$, a contradiction. ☐

Corollary 7.9. $\deg(1) > ak - 3b > 0.89kb - 3b$.

Proof. Since 1 has maximum degree in $[b]$, this follows directly from Theorem 7.8 and Theorem 7.7. ☐

Define a vertex in $[b] \setminus (S \cup T)$ to be *good* if its degree is less than $0.89kb - 4b$. In particular, if a vertex is good, then its degree is less than $\deg(1) - b$.

Claim 7.10. *There are at least $0.334b$ good vertices.*

Proof. Suppose otherwise. Then, since $|S \cup T| = d + \ell \leqslant 0.056b$, this means that at least $0.61b$ special vertices have degree at least $0.89kb - 4b$. This means that the total number of edges in G is (subtracting $b(b-1)/2$ to account for possible double counting of edges)

$$0.61b(0.89kb - 4b) - \frac{b(b-1)}{2} > 0.54kb^2 - 3b^2 > 0.5kb^2 + b^2.$$

But we have already shown in (7.13) that the number of edges in G is at most $\frac{b(b-1)}{2}(k-1)+bq \leqslant 0.5kb^2 + 0.031b^2$, which is less than the above quantity, which is a contradiction. ☐

Now, consider any good vertex v. Recall that when we derived the bound (7.10), we used that at most $d'(k-1)+q$ edges are deleted from vertex 1 in steps 1 and 2. Therefore, since v has degree at least b less than that of vertex 1, step 3 cannot increase the number of deleted edges from v farther than $d'(k-1)+q-b$.

Claim 7.11. *There are at most $0.031b$ good vertices v such that at most $d'(k-1)+q-b$ edges are deleted from v in step 2.*

Proof. Suppose otherwise. For any such vertex v, the number of deleted edges touching v in steps 2 and 3 combined cannot exceed $d'(k-1)+q-b$. Thus, for each such v, we can lower the bound (7.10) by at least b. Since there are more than $0.031b$ such v, in total we can lower the bound (7.10) by at least $0.031b^2 > bq$, a contradiction. ☐

Now say a vertex is *very good* if it is lonely and good, and more than $d'(k-1)+q-b$ edges are deleted from v in step 2. By the argument preceding the previous claim, no edges are deleted from any very good vertex in step 3.

Claim 7.12. *There are at least $0.203b$ very good vertices.*

Proof. This follows directly from Theorem 7.10, Theorem 7.6, and Theorem 7.11. ☐

Now, for any very good vertex v, let j_v be such that $d'k-d-j_v$ vertices are deleted from v in step 2. Since $d'k-d-j_v \leqslant d'(k-1)+q-j_v$, this means that we can reduce the bound in (7.10) by j_v for each such v. Thus, the sum of j_v over all very good vertices v is at most bq. But note that j_v counts the number of missing edges from v to all vertices in Γ except those in S. Thus, there are at most bq missing edges from all good vertices v to all vertices in Γ except those in S. Therefore, we can deduce the following.

Claim 7.13. *There are at most $0.011bq$ pairs (γ, v), where $\gamma \in \Phi'$ and v is a very good vertex, such that v has no edge to any vertex in γ except the one that is in S.*

Proof. Any such pair (γ, v) contains $k-1$ missing edges from v to the vertices in γ except the ones in S, so by the preceding argument, there are at most $bq/(k-1) < 0.011bq$ such pairs. ☐

Define a pair (γ, β) of cells, where $\gamma \in \Phi'$ and β contains a lonely vertex $v \in [b] \setminus (S \cup T)$, to be *happy* if there is an edge from v to some vertex in γ (except the vertex in γ that is in S). There are at least $(q/2)(0.203b)$ pairs (γ, β) where $\gamma \in \Phi'$ and β contains a very good vertex, so by Theorem 7.13, there are at least $0.09bq$ happy pairs.

We will now show that happy pairs can be used to reduce the bound in (7.5) in order to obtain a contradiction. In order to do so, we must first introduce some notation to measure how much the bound can be reduced.

We had used (7.5) as a bound for the number of edges deleted from any $w \in T$ in step 4. Note the edges deleted from w in step 4 are exactly those which do not touch in $[b] \setminus T$ (since the edges to $[b] \setminus T$ were deleted in steps 1 and 2). To this end, define

$$F(w) = a(k-1) - e(w, (\alpha_1 \cup \cdots \cup \alpha_a) \setminus ([b] \setminus T)),$$

where the latter term is just the number of edges from w to all cells, excluding those edges to $[b] \setminus T$. We also define

$$f(w) = e(w, R),$$

recalling that the exterior R is defined as the complement of $\alpha_1 \cup \cdots \cup \alpha_a \cup \{1\}$ (and thus is also disjoint from $[b] \setminus T$). Then, the total number of edges deleted from w in step 4 is

$$a(k-1) - F(w) + f(w).$$

Consequently, we can lower the bound on the number of edges deleted from w given by (7.5) (and thus also lower the bound (7.7)) by $r + F(w) - f(w) \geqslant F(w) - f(w)$, for each $w \in T$. Let

$$E(w) = \max\{F(w) - f(w), 0\}.$$

Since the bound (7.7) cannot be lowered by more than bq, we have

$$\sum_{w \in T} E(w) \leqslant bq. \tag{7.14}$$

Now, define the function $F(w, \beta)$ to be the contribution of the cell β toward $F(w)$. Formally, we define

$$F(w, \beta) = (k-1) - e(w, \beta \setminus ([b] \setminus T)).$$

Since each cell has at least one vertex in $[b] \setminus T$, we have $|\beta \setminus ([b] \setminus T)| \leqslant k-1$, so $F(w, \beta)$ is always nonnegative. Then, we have by definition that

$$F(w) = \sum_{i \in [a]} F(w, \alpha_i).$$

Now, similarly define

$$E(w, \beta) = \max\{F(w, \beta) - f(w), 0\}.$$

Then, we have

$$E(w) \geqslant \sum_{i \in [a]} E(w, \alpha_i).$$

This is because if any term in the sum is positive, then the right hand side is immediately less than $F(w)$ by at least $f(w)$. Therefore, (7.14) implies that

$$\sum_{w \in T,\, i \in [b]} E(w, \alpha_i) \leqslant bq.$$

Furthermore, for a block $\gamma \in \Gamma$, define $E(\gamma, \beta)$ to be the sum over all $w \in \gamma$ which are also in T of the quantity $E(w, \beta)$. Then, we have

$$\sum_{\gamma \in \Gamma,\, i \in [b]} E(\gamma, \alpha_i) \leqslant bq. \tag{7.15}$$

Now we will bound the value of $E(\gamma, \beta)$ when (γ, β) is a happy pair. First, we will need the following lemma.

Lemma 7.14. *If (γ, β) forms a happy pair, then there exists no matching in G consisting only of edges from γ to $\beta \cup R$ which covers all vertices in $\gamma \cap T$.*

Proof. Suppose otherwise. We will construct a matching M which covers $[b]$. Let x denote the vertex in γ which is also in S (there is only one since $\gamma \in \Phi'$), and let v denote the special vertex in β.

Since $\gamma \in \Phi$, there exists an edge connecting 1 with x; add this edge to M. Also, add all edges of the matching described in the lemma statement to M. Then M now covers all special vertices in γ.

If v is not already covered by M, then note that since (γ, β) is happy, there must be an edge from v to some vertex in γ other than x. Then, we can just add this vertex to M, and if there is already an edge containing the other vertex, we can just delete it.

Now M still covers all special vertices in γ, in addition to v (which is the only special vertex in β) and 1. Also, all edges in M only contain vertices in $\{1\} \cup \beta \cup \gamma \cup R$. Thus we can add all cells α_i other than β, γ to M, so that M is a matching which covers all special vertices, a contradiction. □

This allows us to bound $E(\gamma, \beta)$.

Lemma 7.15. *If (γ, β) is a happy pair, then $E(\gamma, \beta) \geqslant k - 1$.*

Proof. Suppose (γ, β) is happy. Then, applying Hall's marriage theorem along with Theorem 7.14, there must exist a set $W \subset \gamma \cap T$ with at most $|W| - 1$ neighbors in $\beta \cup R$. Note that $1 \leqslant |W| \leqslant k - 1$ (the second inequality is because γ has one vertex in S which is thus not in T). Then, for any $w \in W$, we have

$$\begin{aligned} E(w, \beta) \geqslant F(w, \beta) - f(w) &= (k-1) - e(w, \beta \setminus ([b] \setminus T)) - e(w, R) \\ &\geqslant (k-1) - e(w, \beta) - e(w, R) \\ &= (k-1) - e(w, \beta \cup R) \\ &\geqslant (k-1) - (|W| - 1) \\ &= k - |W|. \end{aligned}$$

Summing over w in W, this means that

$$E(\gamma, \beta) \geqslant \sum_{w \in W} E(w, \beta) \geqslant |W|(k - |W|).$$

Since $1 \leqslant |W| \leqslant k - 1$, the above quantity is at least $k - 1$. □

Finally, note that Theorem 7.15, along with the previous observation that there are at least $0.09bq$ happy pairs, means that

$$\sum_{\gamma \in \Gamma,\, i \in [b]} E(\gamma, \alpha_i) \geqslant (0.09bq)(k-1) > bq.$$

This is at odds with (7.15), so we finally have a contradiction. This completes the proof of Theorem 6.2.

8. Concluding remarks

This problem is closely related to the Erdős matching conjecture, which states the following.

Conjecture 8.1 (Erdős matching conjecture, [2])**.** Let n, k, r be positive integers such that $n \geqslant kr$. Suppose that $\mathcal{F}$ is a family of k-subsets of $[n]$. If $\mathcal{F}$ has no matching of size r, then

$$|\mathcal{F}| \leqslant \max\left\{\binom{kr-1}{k}, \binom{n}{k} - \binom{n-r+1}{k}\right\}.$$

Here $\binom{kr-1}{k}$ is the number of k-subsets of $[kr-1]$, and $\binom{n}{k} - \binom{n-r+1}{k}$ is the number of k-subsets of $[n]$ which contain an element of $[r-1]$.

The cases where $n \geqslant 5rk/3 - 2r/3$ and where $n \leqslant (r+1)(k+\varepsilon_k)$ (where $\varepsilon_k > 0$ is some function of k) have been proven by Frankl and Kupavskii [7, 5]. In these proofs, the authors crucially used the fact that $\mathcal{F}$ can be assumed to be a shifted family. We wonder whether our methods may be generalized to make progress on this conjecture. The notion of a blocking set is no longer helpful as is, since we care about matchings of size smaller than n/k. However, maybe it is possible to define a more general notion of a blocking set in the case that $n > kr$, and use methods similar to ours to show that such a blocking configuration must be of a certain form.

We also wonder whether our methods can be used to prove Theorem 1.6 in the case where k is small, in order to completely resolve Theorem 1.4. In particular, the only obstacle to extending our proof to all k is Theorem 6.2, the statement about graphs. Though we were only able to prove it for large enough k, based on examination of small cases, we conjecture that this statement is true for small k as well.

Conjecture 8.2. The statement of Theorem 6.2 holds whenever $k \geqslant 4$.

This conjecture would immediately imply Theorem 1.4 for all $k \geqslant 4$, since the rest of our proof still holds up. We believe there may still be a simple proof of this conjecture, since the statement is primarily about ordinary graphs, which are substantially simpler to deal with than set systems.

In the case $k = 3$, the construction $\mathcal{E}'(3, n, b)$ corresponds to a second equality case in Theorem 6.2, shown in Figure 8.1. If this equality case is in fact the only other equality case of Theorem 6.2 (as small cases seem to suggest), then one can use an identical argument to ours to prove that the optimal $\mathcal{F}$ is either $\mathcal{E}(3, n, b)$ or $\mathcal{E}'(3, n, b)$. By the computation in Section 2.1, we would then know that it must be $\mathcal{E}'(3, n, b)$, leading to the following conjecture.

Conjecture 8.3. For each positive integer b there exists $s_0(b)$ such that the following holds: Let $n = ks$, where $s \geqslant s_0(b)$, and let $\mathcal{F}$ be a family of 3-subsets of $[n]$ with no perfect matching and no blocking set of size less than b. Then,

$$|\mathcal{F}| \leqslant |\mathcal{E}'(3, n, b)|.$$

In [6], Frankl also conjectured that the result of Theorem 1.1 is actually true for all $s \geqslant 6$ (as opposed to s needing to be larger than some function of k). We may similarly expect that the dependence on k may be removed in Theorem 1.6. Thus we conjecture the following.

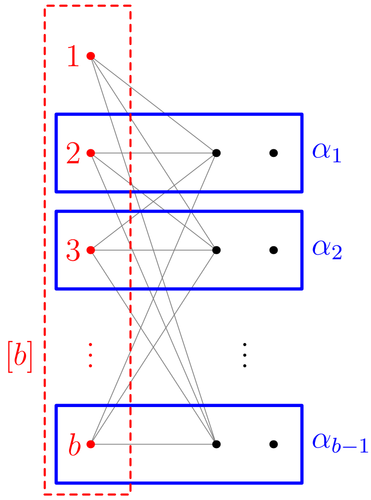

Figure 8.1: The other equality case of Theorem 6.2 in the case where $k = 3$.

Conjecture 8.4. There exists a function $s_0(b)$ such that whenever k, b, s are positive integers such that $k \geqslant 4$ and $s \geqslant s_0(b)$, the following holds: Let $n = ks$, and suppose that $\mathcal{F}$ is a family of k-subsets of $[n]$. Then, if $\mathcal{F}$ has no perfect matching and no blocking set of size less than b, then

$$|\mathcal{F}| \leqslant |\mathcal{E}(k, n, b)|.$$

Acknowledgements

This research was conducted at the University of Minnesota, Duluth REU. Thanks to Joe Gallian for running the REU program. Thanks also to Zachary Chroman for helpful discussions, and to Evan Chen for reviewing a draft of this paper and providing suggestions and assistance in making diagrams.

References

[1] I. Dinur and E. Friedgut, *Intersecting families are essentially contained in juntas*, Comb. Probab. Comput. **18** (2009), 107–122.

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

[3] P. Erdős, C. Ko, and R. Rado, *Intersection theorems for systems of finite sets*, Q. J. Math. **12** (1961), no. 1, 313–320.

[4] P. Frankl, *The shifting technique in extremal set theory*, Surveys in combinatorics **123** (1987), 81–110.

[5] ———, *Proof of the Erdős matching conjecture in a new range*, Israel J. Math. **222** (2017), no. 1, 421–430.

[6] P. Frankl, *On non-trivial families without a perfect matching*, European J. Combin. **84** (2020), 103044.

[7] P. Frankl and A. Kupavskii, *The Erdős matching conjecture and concentration inequalities*, 2018.

[8] ———, *Simple juntas for shifted families*, (2019).

[9] P. Frankl and N. Tokushige, *Invitation to intersection problems for finite sets*, J. Combin. Theory Ser. A **144** (2016), 157 – 211.

[10] D. J. Kleitman, *Maximal number of subsets of a finite set no k of which are pairwise disjoint*, J. Combin. Theory **5** (1968), no. 2, 157 – 163.