Identifying defective sets using queries of small size

Fabrício S. Benevides [*] Dániel Gerbner [†] Cory T. Palmer [‡]
Dominik K. Vu [§]

March 8, 2017

Abstract

We examine the following version of a classic combinatorial search problem introduced by Rényi: Given a finite set X of n elements we want to identify an unknown subset Y of X, which is known to have exactly d elements, by means of testing, for as few as possible subsets A of X, whether A intersects Y or not. We are primarily concerned with the non-adaptive model, where the family of test sets is specified in advance, in the case where each test set is of size at most some given natural number k. Our main results are nearly tight bounds on the minimum number of tests necessary when d and k are fixed and n is large enough.

1 Introduction

We consider a central question of combinatorial search theory in which, given a set X, we wish to identify a particular subset Y of "unknown" elements of X. This means that we know the elements of the set X and there is an oracle that knows which elements (of X) are in Y. We call the elements of Y *defective*. We say that we have identified the set Y, if we discovered precisely which elements belong to Y (and which do not). To this end, we are allowed to construct a family $\mathcal{A}$ of *queries* (i.e., questions to the oracle). Each query in $\mathcal{A}$ corresponds to a subset $A \subset X$ and the oracle responds "yes" if and only if A contains at least one of the elements of Y. Here we are concerned with the so called non-adaptive

[*]Departamento de Matemática, Universidade Federal do Ceará, Fortaleza, CE 60455-760, Brazil fabricio@mat.ufc.br. Research supported by CNPq.

[†]Hungarian Academy of Sciences, Alfréd Rényi Institute of Mathematics, P.O.B. 127, Budapest H-1364, Hungary gerbner.daniel@renyi.mta.hu. Research supported by Hungarian National Science Fund (OTKA), grant PD 109537.

[‡]Department of Mathematical Sciences, University of Montana, Missoula, Montana 59812, USA cory.palmer@umontana.edu. Research supported by Hungarian National Science Fund (OTKA), grant NK 78439.

[§]Department of Mathematical Sciences, University of Memphis, Memphis, Tennessee 38152, USA dominik.vu@memphis.edu. Research supported in part by the National Science Foundation, grants DMS-0906634 and CNS-0721983, and by the Heilbronn Fund.

case, in which the family $\mathcal{A}$ of queries is chosen in advance, so we cannot modify $\mathcal{A}$ based on the answers to some of the queries (equivalently, the queries are made simultaneously to the oracle). The typical goal is to find the minimum size of a family $\mathcal{A}$ that is required to identify any set Y. This question is related to many practical problems, amongst which are *Wasserman-type blood tests*, chemical analysis and the defective coin problem [5, 16]. A comprehensive overview of the main types of combinatorial search problems can be found in a survey by Katona [13] or the monograph of Du and Hwang [6].

In many cases we may be given some extra information about Y. For example, we may already know the number of elements in Y. Given d, in order to identify a fixed defective set of size at most (or exactly) d among n elements it is well known (see e.g. [6]) that the number of queries required, denoted $q(n,d)$, satisfies

$$\Omega\left(\frac{d^2}{\log d}\log n\right) \leq q(n,d) \leq O(d^2 \log n).$$

In other cases we have extra restrictions on the queries that we are allowed to use. In this paper, we restrict our attention to the case where the query sets may only be of size at most some given natural number k. For this model, the particular case where Y contains a single element was posed as a problem by Rényi [15] and solved by Katona [12] for $k < n/2$. Katona determined the exact form of a matrix representing an optimal search and used this to find upper and lower estimates for the minimum number of queries. While the lower bound provided is best known, the upper bound was subsequently improved by Wegener [17] and Luzgin [14]. In 2008 Ahlswede [1] proved that the lower bound is asymptotically tight.

Let X be a set of size n and $Y \subset X$ be a (unkown) set of defective elements which is known to satisfy $|Y| \leq d$. Let $q(n,\overline{d},k)$ denote the least number of queries of size at most k necessary to identify Y. In 2013, Hosszu, Tapolcai and Wiener [11] strengthened Katona's result while providing a proof entirely relying on linear algebraic methods.

Theorem 1 (Hosszu, Tapolcai and Wiener [11])**.** *For $k < n/2$, $q(n,\overline{1},k)$ is the least number q for which there exist positive integers $j \leq q-1$ and $a < \binom{q}{j+1}$ such that*

$$\sum_{i=0}^{j} i \cdot \binom{q}{i} + a(j+1) \leq kq,$$

$$\sum_{i=0}^{j} \binom{q}{i} + a = n.$$

When n is large enough this gives the following corollary.

Corollary 2 (Hosszu, Tapolcai and Wiener [11])**.** *If $n \geq \binom{k}{2} + 1$, then*

$$q(n,\overline{1},k) = \left\lceil \frac{2n-2}{k-1} \right\rceil.$$

In the same setup as above, but now for $d > 1$, D'yachkov and Rykov [7] proved a general lower bound and found conditions for when this lower bound is sharp (see also Füredi and Ruszinkó [10]).

Theorem 3 (D'yachkov and Rykov [7])**.** *If $n \geq k \geq d \geq 2$, then*

$$\left\lceil \frac{dn}{k} \right\rceil \leq q(n, \overline{d}, k).$$

Furthermore, if $d \geq 3$, $k \geq d+1$ and $n = k^d$, then

$$q(n, \overline{d}, k) = \frac{dn}{k} = dk^{d-1}.$$

In light of the above results, we focus on the case where the defective set Y is known to have size exactly d. This allows for a smaller number of queries to determine Y. Define $q(n, d, k)$ as the minimum number of queries of size at most k needed to determine any defective set Y of size exactly d among n elements. We consider this parameter when n is large compared to fixed d and k. Our main theorem gives bounds on $q(n, d, k)$, for n large enough, that are nearly sharp when d is even. In fact, when d is even, the values of $\lfloor d/2 \rfloor$ and $\lceil d/2 \rceil$ are equal, so the difference between our upper and lower bounds is at most a constant (not depending on n).

Theorem 4. *Fix an integer $d \geq 2$. If $n \geq k \geq \lfloor d/2 \rfloor + 1$, then*

$$\left\lceil \frac{(\lfloor d/2 \rfloor + 1)n}{k} \right\rceil - 1 \leq q(n, d, k).$$

Furthermore, if $k \geq 2$ and n is sufficiently large compared to k, then

$$q(n, d, k) \leq \left\lceil \frac{(\lceil d/2 \rceil + 1)(n-1)}{k} \right\rceil + (\lceil d/2 \rceil + 1).$$

Note that by querying singleton sets we can identify any defective set of size exactly d using $n-1$ queries (or a defective set of any size using n queries). So we have the trivial upper bound $q(n, d, k) \leq n-1$. When $k \leq \lceil d/2 \rceil + 1$ the bound given by Theorem 4 is worse than the trivial upper bound, so the upper bound in Theorem 4 is only of interest when $k > \lceil d/2 \rceil + 1$.

When $d = 2, 3$ these bounds can be improved (for n sufficiently large compared to k).

Theorem 5. *Let $n \geq k$ be positive integers with n sufficiently large compared to k, then*

(a)

$$q(n, 2, k) = \left\lceil \frac{2(n-1)}{k} \right\rceil,$$

(b)

$$\left\lceil \frac{3n-2}{k+3} \right\rceil \leq q(n,3,k) \leq \left\lceil \frac{3n-3}{k} \right\rceil + 3.$$

We use the notation $[n]$ to denote the set $\{1, 2, \ldots, n\}$. A family $\mathcal{A}$ of subsets of $[n]$ is called *d-separating* (*$\overline{d}$-separating*) if for any two distinct sets $D_1, D_2 \subset [n]$ of size d (at most d, respectively) there exists a member $A \in \mathcal{A}$ such that either

$$A \cap D_1 \neq \emptyset \text{ and } A \cap D_2 = \emptyset,$$

or

$$A \cap D_1 = \emptyset \text{ and } A \cap D_2 \neq \emptyset.$$

That is, there exists $A \in \mathcal{A}$ that intersects exactly one of the sets D_1, D_2. It is well known that a family $\mathcal{A}$ of queries can determine any defective set Y of size d if and only if $\mathcal{A}$ is a d-separating family.

The separating property is monotone in the following sense: a d-separating family is ℓ-separating for any $\ell \leq d$. Indeed, if $\mathcal{A}$ is d-separating, and L_1, L_2 are two distinct subsets of size ℓ from a ground set of size $n \geq d+\ell$, then fix a set S of $d-\ell$ elements from $[n] \setminus (L_1 \cup L_2)$ and let $D_1 = L_1 \cup S$ and $D_2 = L_2 \cup S$. Since $\mathcal{A}$ is d-separating, there exists $A \in \mathcal{A}$ which intersects exactly one of the sets D_i. Such a set clearly intersects exactly one of the sets L_i. In fact, the separating property is monotone even when $d \leq n < d + \ell$. In this case, when constructing D_1 and D_2 we may use some vertices from $L_1 \setminus L_2$ and $L_2 \setminus L_1$. We leave the details to the interested reader as we will only be concerned with the case in which n is large compared to d and ℓ.

For $|\mathcal{A}| = q$, we can form a $q \times n$ matrix M such that the rows of M are the characteristic vectors of the members of $\mathcal{A}$ (alternatively, thinking of $\mathcal{A}$ as a hypergraph, the transpose of M is the incidence matrix of $\mathcal{A}$). The columns of M can be thought of as characteristic vectors of a hypergraph $\mathcal{H}$ on the ground set $[q]$ (so M is the incidence matrix of $\mathcal{H}$). It is easy to see that $\mathcal{A}$ is d-separating if and only if $\mathcal{H}$ is *d-union-free*, that is, every collection of exactly d distinct hyperedges of $\mathcal{H}$ has a unique union. For a given family of queries $\mathcal{A}$ we will call such a hypergraph $\mathcal{H}$ the *dual hypergraph* of $\mathcal{A}$. Observe that the dual hypergraph $\mathcal{H}$ is defined on a ground set of size $|\mathcal{A}| = q$ and has n hyperedges. Note that in this model it is possible for $\mathcal{H}$ to have the empty set as a hyperedge[1].

Clearly if a family is $\overline{d}$-separating, then it is also d-separating. Chen and Hwang [3] provide a relationship in the other direction.

Theorem 6 (Chen and Hwang [3])**.** *If $\mathcal{A}$ is $2d$-separating family, then there exists a $\overline{d+1}$-separating family $\mathcal{A}'$ obtained by adding at most one new element to the ground set of $\mathcal{A}$.*

This theorem is the primary tool we use for proving the lower bound of Theorem 4. Chen and Hwang state that their theorem is weak in the sense that it should be possible to construct an $\overline{\ell}$-separating family $\mathcal{A}'$ for $\ell > d+1$. However, it follows from our upper bound in Theorem 4 that in general ℓ cannot be improved to $d+2$.

[1]Note that the property that $\mathcal{A}$ is d-separating, prevents the dual hypergraph from having multi-hyperedges.

2 General bounds on $q(n, d, k)$

We begin by proving the lower bound in Theorem 4. Suppose that $\mathcal{A}$ is a d-separating family on the ground set $[n]$ with query size at most k such that

$$|\mathcal{A}| < \left\lceil \frac{(\lfloor d/2 \rfloor + 1)n}{k} \right\rceil - 1.$$

Set $\ell = \lfloor d/2 \rfloor + 1$. By Theorem 6 we can add at most one new element to the ground set of $\mathcal{A}$ in order to obtain an $\overline{\ell}$-separating family $\mathcal{A}'$ such that

$$|\mathcal{A}'| < \left\lceil \frac{\ell n}{k} \right\rceil.$$

This contradicts the lower bound given by Theorem 3.

To prove the upper bound in Theorem 4 we show an explicit construction of the dual hypergraph. Recall that a hypergraph $\mathcal{H}$ is *ℓ-uniform* if all hyperedges are of size ℓ and a hypergraph $\mathcal{H}$ is *k-regular* if every vertex of $\mathcal{H}$ is contained in exactly k hyperedges. Furthermore, a hypergraph $\mathcal{H}$ is *linear* if every two hyperedges intersect in at most one vertex.

A hypergraph is said to be a *cycle* if it has at least two edges and there exists a cyclic ordering of its edges $\{e_1, \ldots, e_\ell\}$ such that there are distinct vertices $v_1, \ldots, v_\ell$ such that $v_i = e_i \cap e_{i+1}$ (where $e_{\ell+1} = e_1$). This concept of a cycle in a hypergraph is sometimes called *Berge-cycle*, after C. Berge [2]. The *length* of a cycle is the number of edges it contains and the *girth* of a hypergraph is the length of the shortest cycle it contains. We use the term *triangle* and C_4 to refer to hypergraph cycles with three and four hyperedges respectively.

We begin with a lemma relating the uniformity and the property of being union-free for hypergraphs of girth at least 5.

Lemma 7. *Let $\ell \geq 2$ and G be an ℓ-uniform linear hypergraph. If G has girth at least 5, then G is $(2\ell - 2)$-union-free.*

Proof. Suppose G is not $(2\ell-2)$-union-free, then there exist two distinct collections of edges $\mathcal{D} = \{D_1, \ldots, D_{2\ell-2}\}$ and $\mathcal{E} = \{E_1, \ldots, E_{2\ell-2}\}$ such that

$$\bigcup_{i=1}^{2\ell-2} D_i = \bigcup_{i=1}^{2\ell-2} E_i.$$

Consider the case in which there exist two edges $D_1, D_2 \in \mathcal{D}$ neither of which is a member of $\mathcal{E}$. If D_1 and D_2 are not disjoint, as G is linear, then their union has $2\ell-1$ vertices. These vertices are covered by the union of the $2\ell - 2$ hyperedges in $\mathcal{E}$ and each hyperedge E_i contains at most one vertex from each of the edges D_1 and D_2. Therefore, there must be an E_i that intersects D_1 and D_2 in distinct vertices. Hence, E_i, D_1, D_2 form a triangle; a contradiction. On the other hand, if D_1 and D_2 are disjoint, then their union has 2ℓ vertices. These vertices are covered by the union of the $2\ell-2$ hyperedges in $\mathcal{E}$ and each hyperedge E_i

contains at most one vertex from each of the sets D_1 and D_2. Therefore there must be two edges E_i and E_j whose union intersects the union of D_1 and D_2 in four distinct vertices. As each of E_i and E_j intersect each of D_1 and D_2 in at most one vertex, it is easy to see that in this case E_i, D_1, E_j, D_2 must form a C_4; a contradiction.

Consider the case in which there exists exactly one edge in $\mathcal{D}$ (say D_1) that is not a member of $\mathcal{E}$. Consequently, there is a set in $\mathcal{E}$ (say E_1) that is not in $\mathcal{D}$. Thus $\mathcal{D}\backslash D_1 = \mathcal{E}\backslash E_1$. The remaining $2\ell-3$ members of $\mathcal{D}$ and $\mathcal{E}$ are the same. The union of D_1 and E_1 has at least $2\ell - 1$ elements. All the vertices of the union except for the intersection must be covered by the edges in $\mathcal{D} \setminus D_1 = \mathcal{E} \setminus E_1$. As in the previous case we get either a triangle or a C_4 (depending on whether D_1 and E_1 intersect); a contradiction. □

Ellis and Linial [8] (using a result of Cooper, Frieze, Molloy, and Reed [4]) constructed a regular uniform hypergraph with girth at least 5.

Theorem 8 (Ellis and Linial [8])**.** *Fix integers $\ell \geq 3$ and $k \geq 2$. Then for every m large enough such that ℓ divides m, there exists a linear k-regular ℓ-uniform hypergraph on m vertices with girth at least 5.*

We now construct a d-union-free hypergraph on at most $\left\lceil \frac{(\lceil d/2 \rceil + 1)(n-1)}{k} \right\rceil + (\lceil d/2 \rceil + 1)$ vertices and with at least n hyperedges (for n large enough with respect to d and k). This will be the dual hypergraph of a d-separating family of sets which gives the upper bound in Theorem 4.

Set $\ell = \lceil d/2 \rceil + 1$ and let q be the smallest integer such that $\lceil \frac{\ell(n-1)}{k} \rceil \leq q$ and q is divisible by ℓ. Thus

$$\left\lceil \frac{\ell(n-1)}{k} \right\rceil \leq q \leq \left\lceil \frac{\ell(n-1)}{k} \right\rceil + \ell = \left\lceil \frac{(\lceil d/2 \rceil + 1)(n-1)}{k} \right\rceil + (\lceil d/2 \rceil + 1).$$

We can force q large by making n large, as $\lceil \frac{\ell(n-1)}{k} \rceil \leq q$. Therefore, by Theorem 8 we can construct a linear k-regular ℓ-uniform hypergraph $\mathcal{H}$ on q vertices with girth at least 5. The number of hyperedges in $\mathcal{H}$ is

$$\frac{kq}{\ell} \geq \frac{k}{\ell} \left\lceil \frac{\ell(n-1)}{k} \right\rceil \geq n - 1.$$

By Lemma 7, $\mathcal{H}$ is d-union-free (in fact, when d is odd $\mathcal{H}$ is $(d+1)$-union-free). Now let us add the empty set (as a hyperedge) to $\mathcal{H}$ to get a hypergraph with at least n hyperedges. This new hypergraph is still d-union-free.

3 Bounds for small defective sets

In this section we prove Theorem 5.

3.1 Two defective elements – Proof of Theorem 5(a)

Instead of applying the theorem of Ellis and Linial as above, we can use a version of the classic result of Erdős and Sachs [9] on the existence of regular graphs of arbitrary girth. This allows for a concrete bound on the threshold for n.

Theorem 9 (Erdős and Sachs [9])**.** *Fix integers $k \geq 2$ and $g \geq 4$, and let $m \geq 4k^g$ be an even integer. Then there exists a k-regular graph on m vertices with girth at least g.*

The following proposition gives the upper bound in Theorem 5(a).

Proposition 10. *Fix an integer $k \geq 2$ and let $n > 2k^7$. Then*

$$q(n, 2, k) \leq \left\lceil \frac{2(n-1)}{k} \right\rceil.$$

Proof. Let $q = \left\lceil \frac{2(n-1)}{k} \right\rceil \geq 4k^6$. We will construct a graph G with girth at least 5 on q vertices with $n-1$ edges. By Lemma 7 we have that G is a 2-union-free (hyper)graph. Then we add the empty set (as a hyperedge) to G to get a hypergraph on q vertices with n hyperedges. It is easy to see that if G is 2-union-free, then adding the empty set cannot destroy the 2-union-free property.

We distinguish two cases based on the parity of q. First let us assume q is even. By Theorem 9 there exists a k-regular graph G on q vertices with girth at least 5. The number of edges in G is $qk/2 \geq n-1$.

Now suppose q is odd. By Theorem 9 there exists a k-regular graph G' on $q+1$ vertices with girth at least 6. Let us remove an arbitrary vertex x from G'. Let X be the neighborhood of x. The graph G' is triangle-free, so X is an independent set. Furthermore, G' is k-regular, so $|X| = k$, so we can add a matching of size $\lfloor k/2 \rfloor$ to the vertices of X. Let the resulting graph be G. It is easy to see that as G' had girth at least 6, the graph G will have girth at least 5. The number of edges in G is at least

$$\frac{k(q+1)}{2} - k + \left\lfloor \frac{k}{2} \right\rfloor \geq n - 1 - \frac{k}{2} + \left\lfloor \frac{k}{2} \right\rfloor \geq n - 1 - \frac{1}{2}.$$

Therefore, the number of edges in G is at least $n-1$. □

We now prove the lower bound on $q(n, 2, k)$. Fix k and n and let q be the minimum size of a 2-separating family on the ground set $[n]$ with query size at most k. Let $\mathcal{H}$ be the dual hypergraph, so $\mathcal{H}$ has n hyperedges and has maximum degree k. In $\mathcal{H}$, let $e \leq 1$ be the number of hyperedges of size 0, let s be the number of hyperedges of size 1 that are contained in a hyperedge of size at least 3, let s' be the number of remaining hyperedges of size 1 and let t be the number of hyperedges of size at least 3. Therefore, the number of hyperedges of size 2 is $n - e - s - s' - t$. We need two simple claims relating these values.

Claim 11. *Every hyperedge of size at least 3 of $\mathcal{H}$ contains at most one hyperedge of size 1. Thus $s \leq t$.*

Proof. Assume (to the contrary) that the hyperedge h contains two hyperedges, say $\{a\}$ and $\{b\}$. Then $h \cup \{a\} = h = h \cup \{b\}$ contradicting the 2-union-free property of $\mathcal{H}$. □

Claim 12. *If $\{a\}$ and $\{a, b\}$ are hyperedges of $\mathcal{H}$, then the degree of the vertex b is 1.*

Proof. Assume there is an edge incident to b, called h, different from $\{a, b\}$. Then $h \cup \{a\} = h \cup \{a, b\}$ contradicting the 2-union-free property of $\mathcal{H}$. □

Now let us count the number of pairs (v, h) where v is a vertex and h is a hyperedge of $\mathcal{H}$ such that $v \in h$.

By definition, the maximum degree in $\mathcal{H}$ is k. Furthermore, for each hyperedge $\{a\}$ of size 1 not contained in a hyperedge of size at least 3, there is a unique vertex of degree 1 in $\mathcal{H}$. Indeed, either the hyperedge $\{a\}$ is not contained in another hyperedge and thus a is of degree 1 or $\{a\}$ is in some hyperedge $\{a, b\}$ and by Claim 12 we have that b is of degree 1. Thus we have at least s' vertices of degree 1. This implies that the number of pairs (v, h) is at most

$$k(q - s') + s' = kq - (k - 1)s'.$$

On the other hand, by counting the size of all hyperedges we get that the number of pairs (v, h) is at least

$$s + s' + 3t + 2(n - e - s - s' - t).$$

Applying Claim 11 gives

$$s+s'+3t+2(n-e-s-s'-t) \geq 2s+s'+2t+2(n-e-s-s'-t) = 2n-2e-s' \geq 2n-2-s'.$$

Combining the upper and lower estimates for the number of pairs (v, h) yields

$$2n - 2 - s' \leq kq - (k - 1)s'.$$

Rearranging these terms (and using the assumption that $k \geq 2$) gives

$$q \geq \frac{2n - 2 + (k - 2)s'}{k} \geq \frac{2(n - 1)}{k},$$

which completes the proof.

3.2 Three defective elements – Proof of Theorem 5(b)

The upper-bound follows from Theorem 4. For the lower bound fix k and n and let q be the minimum size of a 3-separating family $\mathcal{A}$ on the ground set $[n]$ with query size at most k. Let $\mathcal{H}$ be the dual hypergraph for $\mathcal{A}$. By definition $\mathcal{H}$ is defined on a ground set of size q, has n hyperedges and has maximum degree k. Note that this hypergraph is not necessarily uniform.

As in the case where $d = 2$, we sum the sizes of all hyperedges in $\mathcal{H}$. There is at most 1 hyperedge of size 0 and at most q hyperedges of size 1. First we show that there are not too many hyperedges of size 2 in $\mathcal{H}$.

Claim 13. *The graph G formed by the hyperedges of size 2 in $\mathcal{H}$ is a forest.*

Proof. We show that G contains no triangle, no C_4 and no path of length at least 4. Such a graph is clearly a forest.

Recall that a 3-separating family is also 2-separating. Let e, f, g be the edges of a triangle in G, then it is immediate that $e \cup f = e \cup g$ which violates the 2-union-free property of $\mathcal{H}$. Similarly, if e, f, g, h are the edges of a C_4 in G (such that e and g are disjoint), then $e \cup g = f \cup h$ which violates the 2-union-free property of $\mathcal{H}$. Finally, if e, f, g, h are the edges of a path of length 4 (in this order), then $e \cup f \cup h = e \cup g \cup h$ which violates the 3-union-free property of $\mathcal{H}$. □

Thus we have at most $q-1$ hyperedges of size 2 in $\mathcal{H}$. There are n total edges, so the sum of the edge sizes is minimized when there is one edge of size 0, q edges of size 1, $q-1$ edges of size 2, and the rest of size 3. Therefore, the sum of the sizes of the hyperedges of $\mathcal{H}$ is at least

$$q + 2(q-1) + 3(n-1-2q+1) = 3n - 3q - 2$$

The maximum degree in $\mathcal{H}$ is k so the above sum is at most qk. Combining these two estimates and solving for q yields

$$\frac{3n-2}{k+3} \leq q.$$

4 Further results

4.1 Fixed query size

Throughout the paper we have allowed queries to have size at most k. Katona [12] showed that when searching for a fixed defective set of size at most 1 there is no difference in the minimum number of necessary queries whether we restrict the queries size to be at most k or to be exactly k. Therefore it is somewhat unexpected that in the case of searching for a fixed defective set of size exactly d, for $d \geq 2$, we can have different answers depending on whether the query size is at most k or exactly k.

To illustrate, let us examine the simplest case, say $k = 2$ and the defective set is of size exactly d, for a given $d \geq 3$. We distinguish between two kinds of restrictions: (i) each query set is of size at most 2 or (ii) each query set is of size exactly 2. By asking queries of size 1 we can identify the defective set with $n-1$ queries, so $q(n, d, 2) \leq n-1$. On the other hand, if a family of queries is d-separating, then it is 2-separating and we can use Theorem 5 to get

$$q(n, d, 2) \geq q(n, 2, 2) = \left\lceil \frac{2(n-1)}{2} \right\rceil = n - 1.$$

Therefore, we have the following simple corollary.

Corollary 14. *If $n \geq d \geq 3$, then*

$$q(n, d, 2) = n - 1.$$

However, if we only allow queries of size exactly 2 we cannot determine such a defective set with only $n-1$ queries.

Proposition 15. *Let q be the minimum number such that there exists a family of queries of size exactly 2 that can determine any defective set of size d, for a given $d \geq 3$. Then $q \geq n$.*

Proof. Let $\mathcal{H}$ be the dual hypergraph for the family of queries in the statement of the proposition. Then $\mathcal{H}$ is a 2-regular hypergraph on q vertices with n hyperedges. The hyperedges of size 2 in $\mathcal{H}$ form a forest (see Claim 13 in Subsection 3.2). Therefore, the number of hyperedges of size 2 is at most $q-1$. Thus, if there is at most 1 hyperedge of size other than 2, then we are done.

Therefore, we can restrict ourselves to the case where $\mathcal{H}$ has at least two hyperedges of size other than 2. In $\mathcal{H}$, let $e \leq 1$ be the number of hyperedges of size 0, let s be the number of hyperedges of size 1, and let t be the number of hyperedges of size at least 3.

Clearly as $\mathcal{H}$ is 2-regular, every hyperedge of size 1 is contained in exactly one other hyperedge. Furthermore, $\mathcal{H}$ cannot contain two hyperedges of the form $\{a\}$ and $\{a, b\}$ as b would have degree 1 in this case (see Claim 12 in Subsection 3.1). Therefore, every hyperedge of size 1 is contained in a hyperedge of size at least 3. On the other hand, a hyperedge of size at least 3 cannot contain two hyperedges of size 1 as this would violate the 2-union-free property (see Claim 11 in Subsection 3.1). Therefore, the number of hyperedges of size 1 is at most the number of hyperedges of size at least 3, i.e. $s \leq t$.

The sum of degrees in $\mathcal{H}$ is $2q$. By counting the sizes of all edges in $\mathcal{H}$ we obtain that

$$2q \geq s + 3t + 2(n - e - s - t).$$

If $e = 0$, then using the fact that $s \leq t$ it follows that

$$2q \geq s + 3t + 2(n - s - t) \geq 2s + 2t + 2(n - s - t) = 2n$$

and we are done.

Now let us suppose that $e = 1$, i.e. $\mathcal{H}$ contains the empty set as a hyperedge. In this case it is easy to see that $\mathcal{H}$ cannot contain any hyperedges of size 1, i.e $s = 0$. Indeed, if $\{a\}$ is a hyperedge in $\mathcal{H}$, then there must be another hyperedge $H \in \mathcal{H}$ that contains a, but then

$$\{a\} \cup H = \emptyset \cup H$$

violates the 2-union-free property of $\mathcal{H}$. As there are at least two hyperedges of size other than 2 (and none of size 1), there must be at least one hyperedge of size at least 3, i.e. $t \geq 1$. Thus,

$$2q \geq 3t + 2(n - 1 - t) \geq 2n - 2 + t \geq 2n - 1.$$

Therefore $q \geq n$ and we are done. □

4.2 Adaptive search

We call the search model *adaptive* if we ask the query sets in a sequence and allow that each query set A may depend on the answer given for previous queries. As in the previous sections we are particularly interested in the case where the query sets are of size at most k. Let Y be a defective set of at most d elements. The minimum number of queries required to determine Y among a set of size n in the adaptive model is denoted by $t(n, \overline{d}, k)$. In the case of $d = 1$ the question was solved completely by Katona [13].

Theorem 16 (Katona [13])**.** *Let n, k be integers, such that $k < n/2$, then*

$$t(n, \overline{1}, k) = \left\lceil \frac{n}{k} \right\rceil - 2 + \left\lceil \log \left(n - k \left\lceil \frac{n}{k} \right\rceil + 2k \right) \right\rceil .$$

The proof of Theorem 16 can be easily generalized for larger defective sets.

Theorem 17. *For any integers $k, n > k, d > 1$ it holds*

$$\left\lceil \frac{n}{k} \right\rceil - 2 + \log \binom{k+1}{d} \leq t(n, \overline{d}, k) \leq \left\lceil \frac{n}{k} \right\rceil - 2 + d \left\lceil 1 + \log k \right\rceil .$$

Proof sketch. Assume we want to find the (at most) d defective elements among $[n]$. Let $s = \lceil \frac{n}{k} \rceil - 2$. Notice that we can partition $[n]$ into sets, say $A_1, \ldots, A_s, A_{s+1}$, such that $|A_1| = \cdots = |A_s| = k$ and $|A_{s+1}| \leq 2k$. Fix some i, with $1 \leq i \leq s$. We query the set A_i, that is, we ask if any of the d defective elements is in A_i. If we get a positive answer, we perform a standard binary search on A_i to determine one of the defective elements (this takes at most $\lceil \log k \rceil$ queries). An extra query is done to determine if among the remaining $k - 1$ elements of A_i there is another defective element. If so we can perform another binary search among those $k - 1$ elements to find another defective element. We repeat this (removing one element at a time) until we find all defective elements inside A_i. After performing this search for every $1 \leq i \leq s$, if we have found less than d defective elements, then we search for defective elements in A_{s+1} using the same procedure applied to the previous A_is. Altogether, we made one query for each A_i with $1 \leq i \leq s$, and for each defective element we asked at most $\lceil 1 + \log k \rceil$ more queries.

For the lower bound, notice that in any search procedure it is possible that we get negative answers for the first $\lceil \frac{n}{k} \rceil - 2$ queries. Whenever we get a negative answer we may ignore the at most k elements that were queried, but we gain no information about the remaining elements. After those queries there remain at least $k + 1$ elements that have not yet been queried. It is well-known that to identify d defective elements among (at least) $k+1$ elements we need at least $\log \binom{k+1}{d}$ queries. □

Acknowledgments

The authors would like to thank G.O.H. Katona for guidance and helpful discussions. Additionally, the authors would like to acknowledge the hospitality of the organizers of the fifth Emléktábla workshop, during which the majority of this research was conducted.

References

[1] R. Ahlswede. Ratewise-optimal non-sequential search strategies under constraints on the tests. *Discrete Appl. Math.*, 156(9):1431–1443, May 2008.

[2] C. Berge. Hypergraphs. In *Combinatorics of Finite Sets*. North-Holland, Amsterdam, 1989.

[3] H.-B. Chen and F. K. Hwang. Exploring the missing link among d-separable, $\overline{d}$-separable and d-disjunct matrices. *Discrete Appl. Math.*, 155(5):662–664, 2007.

[4] C. Cooper, A. Frieze, M. Molloy, and B. Reed. Perfect matchings in random r-regular, s-uniform hypergraphs. *Combin. Probab. Comput.*, 5(1):1–14, 1996.

[5] R. Dorfman. The detection of defective members of large populations. *Ann. Math. Statist.*, 14, 1943.

[6] D.-Z. Du and F. K. Hwang. *Combinatorial group testing and its applications*, volume 12 of *Series on Applied Mathematics*. World Scientific Publishing Co., Inc., River Edge, NJ, second edition, 2000.

[7] A. G. D'yachkov and V. V. Rykov. Optimal superimposed codes and designs for Renyi's search model. *J. Statist. Plann. Inference*, 100(2):281–302, 2002. International Conference (of the Forum for Interdisciplinary Mathematics) on Combinatorics, Information Theory and Statistics (Portland, ME, 1997).

[8] D. Ellis and N. Linial. On regular hypergraphs of high girth. *Electron. J. Combin.*, 21(1)(P1.54), 2014.

[9] P. Erdős and H. Sachs. Reguläre Graphen gegebener Taillenweite mit minimaler Knotenzahl (in German). *Wiss. Z. Martin-Luther–Univ. Halle–Wittenberg Math.-Natur. Reihe*, 12:251–257, 1963.

[10] Z. Füredi and M. Ruszinkó. Uniform hypergraphs containing no grids. *Adv. Math.*, 240:302–324, 2013.

[11] É. Hosszu, J. Tapolcai, and G. Wiener. On a problem of Rényi and Katona. In *Proc. 8th Jpn.-Hungar. Symp. Disc. Math. Appl.*, pages 229–232, 2013.

[12] G. Katona. On separating systems of a finite set. *J. Combin. Theory*, 1:174–194, 1966.

[13] G. O. H. Katona. Combinatorial search problems. In J. N. Srivastava et al., editor, *A Survey of Combinatorial Theory*. North Holland Publishing Company, Amsterdam, 1973.

[14] V.N. Luzgin. Separating systems of partitions of a finite set. *Kombinatornyĭ Analiz.*, 5:39–45, 1980.

[15] A. Rényi. On a problem of information theory. *Publ. Math. Inst. Hungar. Acad. Sci.*, 6:505–516, 1961.

[16] A. Sterett. On the detection of defective members of large populations. *Ann. Math. Statist.*, 28, 1957.

[17] I. Wegener. On separating systems whose elements are sets of at most k elements. *Discrete Mathematics*, 28(2):219–222, 1979.