J Comb Optim (2007) 14: 121–129
DOI 10.1007/s10878-007-9067-3

An unexpected meeting of four seemingly unrelated problems: graph testing, DNA complex screening, superimposed codes and secure key distribution

H.B. Chen · D.Z. Du · F.K. Hwang

Published online: 31 March 2007

Abstract This paper discusses the relation among four problems: graph testing, DNA complex screening, superimposed codes and secure key distribution. We prove a surprising equivalence relation among these four problems, and use this equivalence to improve current results on graph testing. In the rest of this paper, we give a lower bound for the minimum number of tests on DNA complex screening model.

Keywords Group testing · Pooling designs · Superimposed codes · Graph testing

1 Introduction

In the combinatorial group testing problem (see Balding et al. 1996; Du and Hwang 2000; Kautz and Singleton 1964 for a survey), we consider a set N of n items consisting of at most d defective items and the other good items. The problem is to identify all defective items with a small number of group tests. A group test can be applied to an arbitrary subset $S \subseteq N$ with two possible outcomes; a negative outcome implies all items in S are good, while a positive outcome implies otherwise, i.e., there exists at least one defective item in S, not knowing which or how many.

The first and second author would like to dedicate this paper to professor Frank K. Hwang on the occasion of his 65th birthday.

This research is partially supported by Republic of China, National Science Council grant NSC 92-2115-M-009-014.

H.B. Chen (✉) · F.K. Hwang
Department of Applied Mathematics, National Chiao Tung University, Hsinchu 300, Taiwan
e-mail: andan.am92g@nctu.edu.tw

F.K. Hwang
e-mail: fhwang@math.nctu.edu.tw

D.Z. Du
Department of Computer Science, University of Texas at Dallas, Richardson, TX 75082, USA

The group testing problem has been extended to graph testing (see Chapter 10 of Du and Hwang 2000 for references) where a hypergraph $H(V, E)$ is given. The problem is to identify a hidden subgraph D with a small number of graph tests. A graph test can be applied to an arbitrary subset $S \subseteq V$ with two possible outcomes; a negative outcome implies that all edges in the subgraph H_S induced by S are not in D, while a positive outcome implies otherwise, i.e., H_S contains at least one edge in D, not knowing which or how many. We could have different graph testing problems according as prior knowledge of D; the usual assumption is D has at most d edges, but it can also be D is a matching (Alon et al. 2004; Beigel et al. 2001) or a Hamiltonian circuit (Grebinski and Kucherov 1998). The group testing problem is a special case of the graph testing problem where H is a 1-graph, i.e., each edge is a vertex.

In the DNA complex model (Macula et al. 1999, 2004; Macula and Popyack 2004; Torney 1999), we have a set N of n molecules and an unknown family $D = \{D_i\}$ of subsets of N where each such subset is a cause of a certain disease. The problem is to identify D through a few experiments. An experiment can be applied to an arbitrary subset $S \subseteq N$ with two possible outcomes; a negative outcome implies S does not contain any $D_i \in D$, and a positive outcome implies otherwise. A set of molecules which is a candidate of a member of D is called a complex.

It is easy to see the connection between the complex model and the graph testing model. A molecule is a vertex, a complex is an edge, D is the edge-set in the hidden subgraph, and an experiment is a graph test.

Establishing such a connection leads to two consequences. The obvious one is all results on graph testing are now available to solve the complex model problem. The less obvious one is a change of emphasis in graph-testing research due to the influence of the complex model application. An experiment in the complex model can be time-consuming. Hence it is much preferable to have a nonadaptive algorithm where all subsets for testing are specified at once (and hence can be tested at once theoretically), or at least by a k-round algorithm for some small k. The literature on nonadaptive or k-round algorithm is starting to flourish (Alon et al. 2004; Grebinski and Kucherov, 1998, 2000; Gao et al. 2005; Li et al. 2007).

In the secure key distribution problem (see Mitchell and Piper 1988; Stinson et al. 2000 for a large body of literature), n persons want to communicate securely in groups of r persons. Of course, if each member of an r-group owns a key (for coding and decoding messages) which no nonmember can have, then the communication will certainly be secure. However, too many keys (to be exact, $\binom{n}{r}$) are required. Therefore, the security requirement is relaxed to, given a group of r members and a group of d nonmembers, there exists a key owned by each of the r-group and none of the d-group.

D'yachkov et al. (2002) proposed the binary superimposed (d, r)-code which satisfies the property that for any $d + r$ codewords $C_1, C_2, \ldots, C_{d+r}$, there exists an alphabet which is in every code $C_1, C_2, \ldots, C_r$, but none of $C_{r+1}, C_{r+2}, \ldots, C_{d+r}$. This code was further studied in (Stinson and Wei 2004; Stinson et al. 2000; Engel 1996; Kim and Lebedev 2004), sometimes under the name of cover-free families.

By treating each key as an alphabet, and the set of keys owned by a person as a codeword, a secure key distribution design is a binary superimposed (d, r)-code.

However, the connection between this pair of problems and the first pair of problems is not obvious. In fact, even for an inter-pair problem, where the connection is easy, the literature on the two problems are mostly independent. In this paper we prove a surprising equivalence relation between the two pairs of problems under certain conditions. We use this equivalence to improve existing results on the complex model (hence on graph testing).

2 The equivalence and its consequences

We adopt the notation of the graph-testing model. A nonadaptive graph testing algorithm can be represented by the incidence matrix M between vertices and tests, i.e., rows are label by tests, columns by vertices, and cell (i, j) has a 1-entry if and only if vertex j is in test i. We view a column as the set of row indices where the column has 1-entries. Then we can talk about union and intersection of columns. For a set $S \subseteq V$, let $\bigcup S$ and $\bigcap S$ denote the union and intersection of all columns in S. Suppose D is the set of hidden edges. Then the outcome set (the indices of rows of positive outcome) is simply $\bigcup_{e_i \in D}(\bigcap e_i)$. A hypergraph is said to be a *rank-r graph* if each edge contains at most r vertices. And a hypergraph is an *r-graph* if each edge contains exactly r vertices.

Suppose H is an r-graph and our only knowledge of D is $|D| \leq d$. We define some properties of M relating to its ability to solve this graph testing problem:
$(d, r)_H$*-separable*. For any two distinct d-sets D, D' of edges,

$$\bigcup_{e_i \in D} \Big(\bigcap e_i\Big) \neq \bigcup_{e_i \in D'} \Big(\bigcap e_i\Big). \tag{1}$$

$(\bar{d}, r)_H$*-separable*. For any two distinct sets D, D' of edges with $|D|,\ |D'| \leq d$,

$$\bigcup_{e_i \in D} \Big(\bigcap e_i\Big) \neq \bigcup_{e_i \in D'} \Big(\bigcap e_i\Big). \tag{2}$$

$(d, r)_H$*-disjunct*. For any $d + 1$ edges $e_0, e_1, \ldots, e_d$,

$$\bigcap e_0 \not\subseteq \bigcup_{i=1}^{d} \Big(\bigcap e_i\Big). \tag{3}$$

Clearly, a $(d, r)_H$-separable matrix identifies D if $|D| = d$ is known. A $(\bar{d}, r)_H$-separable matrix and a $(d, r)_H$-disjunct matrix identify D if $|D| \leq d$ is known, while the latter has an easy decoding since every edge not in D appears in a test not covering any hidden edge, thus the outcome is negative and the edge is identified. Note that when all edges not in D are so identified, the remaining edges are the hidden edges. Thus, $(d, r)_H$-disjunct implies $(\bar{d}, r)_H$-separable implies $(d, r)_H$-separable.

Similarly we can define $(d, \bar{r})_H$*-separable*, $(\bar{d}, \bar{r})_H$*-separable* and $(d, \bar{r})_H$*-disjunct* matrix when H is a rank-r graph. When H is the complete r-graph or rank-r graph, then the subscript H will be change to K. In the $\bar{d}$-separable case, we assume that no two positive edges e and e' satisfy e belonging to e' for otherwise we

cannot tell whether only e or both are in D. In the d-disjunct case, we assume that no two edges e and e' satisfy e belonging to e' for otherwise there does not exist a row covering e' but not e.

On the other hand, the incidence matrix of a binary superimposed (d, r)-code has the property which we denote by $(d, r_\cap)$-disjunct. Namely, for any $d + r$ columns $C_1, C_2, \ldots, C_{d+r}$,

$$\bigcap_{i=1}^{r} C_i \not\subseteq \bigcup_{i=r+1}^{d+r} C_i. \tag{4}$$

Note that condition (4) looks different from any of (1), (2), (3). The only result in the literature making a connection between the two types of results is the following (given in D'yachkov et al. 2002):

Lemma 2.1 *$(d, r_\cap)$-disjunct $\Rightarrow$ $(\bar{d}, \bar{r})_K$-separable $\Rightarrow$ $(d - 1, r_\cap)$-disjunct and $(d, (r - 1)_\cap)$-disjunct.*

We will give a proof of the first implication since we believe that the original proof has a slip.

Suppose M is not $(\bar{d}, \bar{r})_K$-separable, i.e., there exist two sets D and D' with $|D| \leq d$ and $|D'| \leq d$ such that $\bigcup_{e_i \in D}(\bigcap e_i) = \bigcup_{e_i \in D'}(\bigcap e_i)$. By our assumption, neither D nor D' contains two edges one containing the other. Thus there must exist an edge e in $D \cup D'$ such that e does not contain any edge from the other set for otherwise we would have $e'' \subseteq e' \subseteq e$ where e and e'' are in the same set. Without loss of generality, assume $e \in D$. Since $e \not\supseteq e_i$ for every $e_i \in D'$, we can choose $C_i \in e_i \setminus e$. Define $S = \{C_i : 1 \leq i \leq |D'|\}$. Then S is a set of at most d columns disjoint from e.

Suppose to the contrary that M is $(d, r_\cap)$-disjunct. Then there exists a row with 1-entries in every column of e and 0-entries in every column in S. Thus this row covers e but none of $e_i \in D'$. Hence at this row

$$\bigcup_{e_i \in D} \left(\bigcap e_i\right) = 1 \neq \bigcup_{e_i \in D'} \left(\bigcap e_i\right) = 0,$$

a contradiction to our previous assumption.

The slip was made by choosing $e \in D \cup D'$ which is not contained in any edge of the other set. Then $e_i \setminus e$ can be empty and C_i cannot be chosen.

We now prove the crucial relation between the two types of results.

Theorem 2.2 *$(d, r_\cap)$-disjunct $\Leftrightarrow$ $(d, r)_K$-disjunct.*

Proof Suppose M is not $(d, r_\cap)$-disjunct. Then there exist $d+r$ columns $C_1, C_2, \ldots, C_{d+r}$ such that

$$\bigcap_{i=1}^{r} C_i \subseteq \bigcup_{i=r+1}^{d+r} C_i.$$

Let $e_0 = \{C_1, C_2, \ldots, C_r\}$ and $e_i = \{C_2, C_3, \ldots, C_r, C_{r+i}\}$, $1 \le i \le d$. Then

$$\bigcup e_0 = \bigcup\{C_2, C_3, \ldots, C_r\} \cup C_1 \subseteq \bigcup\{C_2, C_3, \ldots, C_r\} \cup \left(\bigcup_{i=r+1}^{d+r} C_i\right) = \bigcup_{i=1}^{d}(\cup e_i).$$

Hence M is not $(d, r)_K$-disjunct.

Conversely, suppose M is $(d, r_\cap)$-disjunct. Let $e, e_1, \cdots, e_d$ denote $d+1$ arbitrary edges where no e_i, $1 \le i \le d$, is contained in e. Set $C_{r+i} \in e_i \setminus e$, $1 \le i \le d$, where $e_i \in D$. Then M contains a row which covers e, but intersects none of C_{r+i}, $1 \le i \le d$, i.e., covers none of $e_i \in D$. Hence M is $(d, r)_K$-disjunct. □

We apply Theorem 2.2 to improve various results in the graph testing model.

Let $H = (V, E)$ be a rank-r graph. Gao et al. (2005) gave a construction of the $(d, r)_H$-disjunct matrices by first constructing a q-ary matrix Q and then converting it to a binary matrix M. Let $f_j(e)$ denote the set of q-ary entries in row j collected from the columns associated with the edge $e \in E$. Then Q has the property that for any $d+1$ edges $e_0, e_1, \ldots, e_d$, there exists a row j in which none of the $f_j(e_i)$, $1 \le i \le d$, is contained in $f_j(e_0)$. For row j in Q, let $c_j = |\{f_j(e) : e \in E\}|$. Then $c_j \le \min\{|E|, \Sigma_{i=1}^{r}\binom{q}{i}\}$. Their conversion is to replace row j in Q by c_j rows, and each of which labeled by the set $\{(j, f)\}$ where f is a distinct element in the set $\{f_j(e) : e \in E\}$. For row $\{(j, f)\}$ in the converted matrix M, every column in e with $f_j(e) = f($ there can be more than one such edge $e)$ has a 1-entry and all other columns have a 0-entry. They proved that such a matrix M converted from a q-ary matrix Q is $(d, r)_H$-disjunct. They also gave a construction of $Q = [q_{ij}]$ with $drm + 1$ rows and q^{m+1} columns each representing a degree-m polynomial $p_v(x)$ in $GF(q)$, where $v \in V$ and q is a prime power $\ge drm + 1$, and the value in the cell q_{ij} is defined by $p_j(i)$. Assuming $|E| \ge \Sigma_{i=1}^{r}\binom{q}{i}$, the number of tests in the converted matrix M is

$$\sum_{j=1}^{drm+1} c_j \le (drm+1)\sum_{i=1}^{r}\binom{q}{i} \le (drm+1)\binom{q+r-1}{r}.$$

We now propose a better conversion. Let $c'_x = |\{p_v(x) : v \in V\}|$ for each row x in Q, then $c'_x \le q$. For row x in Q, our conversion is to replace each element in the set $\{p_v(x) : v \in V\}$ by a distinct column of a $t' \times c'_x$ $(d, r_\cap)$-disjunct matrix. Suppose x is the row in which none of the $f_x(e_i)$ is contained in $f_x(e_0)$. Let $C_i \in e_i \setminus e_0$ such that $f_x(C_i) \notin f_x(e_0)$ for $1 \le i \le d$. Then after the conversion there exists at least a row x_j in M, converted from the row x in Q, in which the columns in e_0 all have 1-entries while C_i all have 0-entries, $1 \le i \le d$. Hence $\bigcap e_0 \not\subseteq \bigcup_{i=1}^{d}(\bigcap e_i)$. Since the choice of $e_1, e_2, \ldots, e_d$ is arbitrary, the converted matrix M is $(d, r)_H$-disjunct.

Let $t(d, r, n)_H$ denote the minimum number of rows required for a $(d, r)_H$-disjunct matrix with n columns. Similarly, we define $t(d, r_\cap, n)$. Existing results on $t(d, r_\cap, q)$ (see Stinson et al. 2000 for example) show that it is less than $\binom{q+r-1}{r}$ in general or at least asymptotically. Thus, we have

Theorem 2.3 $t(d, r, q^{m+1})_H \le (drm+1)t(d, r_\cap, q)$.

When H is the complete r-graph, M is $(d, r)_K$-disjunct. By Theorem 2.2, M is also $(d, r_\cap)$-disjunct. Then we have

Corollary 2.4 $t(d, r_\cap, q^{m+1}) \leq (drm + 1)t(d, r_\cap, q)$.

Corollary 2.4 is the same result as given by D'yachkov et al. (2002) on the construction of $(d, r_\cap)$-disjunct matrices using the MDS-code. The incidence matrix of the MDS-code with parameters (q, k, t) is a q-ary matrix of size $t \times q^k$ and the Hamming distance of any pair of columns is $d = t - k + 1$. Lemma 2.5 arises from the definition of the MDS-code.

Lemma 2.5 (Sagalovich 1994) *If $q^k \geq d + r$ and $n \geq dr(k - 1) + 1$, then any MDS-code with parameters (q, k, t) has the property that for any $d + r$ columns $C_1, C_2, \ldots, C_{d+r}$, there exists a row where the set of entries over $C_1, C_2, \ldots, C_r$ and the set of entries over $C_{r+1}, \ldots, C_{d+r}$ are disjoint.*

D'yachkov, Vilenkin, Macula and Torney used the Reed-Solomon q-ary code, which is also an MDS-code, to get a $(drm + 1) \times q^{m+1}$ q-ary matrix with the property that described in Lemma 2.5. Then they also use a $t' \times q$ $(d, r_\cap)$-disjunct matrix to transform the q-ary matrix to binary one. The requirement of this q-ary matrix is seemingly different from that given by Du et al., though the latter also corresponds to an MDS-code. We now prove that the requirements of the two of q-ary matrices are equivalent along the line of Theorem 2.2.

Let $e_0, e_1, \ldots, e_d$ be any $d + 1$ complexes. Set $\{C_1, C_2, \ldots, C_r\} = e_0$ and $C_{r+i} \in e_i \setminus e_0$, $1 \leq i \leq d$. If the Reed-Solomon q-ary code property holds, i.e., there exists a row x such that the set of entries over $C_1, C_2, \ldots, C_r$ and the set of entries over $C_{r+1}, \ldots, C_{d+r}$ are disjoint, then in the row x none of $f_x(e_i)$ is contained in $f_x(e_0)$, $1 \leq i \leq d$.

Conversely, let $C_1, C_2, \ldots, C_{d+r}$ be any $d + r$ columns. Set $e_0 = \{C_1, C_2, \ldots, C_r\}$ and $e_i = \{C_2, \ldots, C_r, C_{r+i}\}$, $1 \leq i \leq d$. If there exists a row x in which none of $f_x(e_i)$ is contained in $f_x(e_0)$, then in the row x the set of entries over $C_1, C_2, \ldots, C_r$ and the set of entries over $C_{r+1}, \ldots, C_{d+r}$ are disjoint.

Li et al. (2007) gave a construction of $(d, 2)_H$-disjunct matrix where $H = K_{n,n}$ is the complete bipartite graph with n vertices in each part. Their construction is similar to the one given in (Gao et al. 2005).

By using a $(d, 2_\cap)$-disjunct matrix for conversion, we obtain an alternative result.

Theorem 2.6 $t(d, 2, 2n)_{Kn,n} \leq (2dm + 1)t(d, 2_\cap, q)$, *where q is a prime power with $q^{m+1} \geq 2n$.*

3 An extension to error-tolerant version

Stinson and Wei (2004) first gave an error-tolerant version of the $(d, r_\cap)$-disjunct matrix. A matrix is called $(d, r_\cap; z)$-disjunct if for any $d + r$ columns $C_1, C_2, \ldots, C_{d+r}$,

$$\left| \bigcap_{i=1}^{r} C_i \setminus \bigcup_{i=r+1}^{d+r} C_i \right| \geq z,$$

i.e., there exist at least z rows in which each of the r designated columns has a 1-entry and each of the other d columns has a 0-entry. For a $(d, r_\cap; z)$-disjunct matrix, if the number of error-tests is less than $\lfloor \frac{z-1}{2} \rfloor$, then we can decode the up-to-d positive complexes. It is because that each negative complex appears in at least $z - \lfloor \frac{z-1}{2} \rfloor = \lceil \frac{z-1}{2} \rceil + 1$ negative pool while each positive complex in at most $\lfloor \frac{z-1}{2} \rfloor$ negative pools (due to errors). Therefore we can separate the negative complexes from the positive ones.

In general, it is not easy to construct a matrix with error-tolerance. A trivial, but not efficient, construction to obtain error-tolerance is by taking copies of each row of the original matrix. In the section we will extend the substitution-type construction introduced in Section 2 to the error-tolerant version. Let Q_z be constructed similar to Q except there are $drm + z$ rows for $z \geq 1$. Surprisingly, by replacing Q by Q_z and $(d, r_\cap)$-disjunct matrix by $(d, r_\cap; z')$-disjunct matrix in the substitution-type construction, we obtain a $(d, r_\cap; zz')$-disjunct matrix which can correct up to $\lfloor \frac{(zz'-1)}{2} \rfloor$ errors.

Lemma 3.1 *For any $d+1$ edges $e_0, e_1, \ldots, e_d$, there exists a set R of at least z rows in Q_z such that for each $j \in R$ none of $f_j(e_i)$ is contained in $f_j(e_0)$, where $1 \leq i \leq d$.*

Proof By construction of Q_z, each column is represented by a degree-m polynomial in $GF(q)$, and all of which are distinct. Hence any two columns have common entries in at most m rows.

Suppose to the contrary that there exist at most $z-1$ rows satisfying the condition. Then by the pigeonhole principal there exists an edge $e_x \in \{e_1, e_2, \ldots, e_d\}$ such that there exists a set N of at least $rm+1$ rows satisfying $f_j(e_x) \subseteq f_j(e_0)$ for each $j \in N$. Use the pigeonhole principal again, there exist two columns, one in e_x and the other in $e_0 \setminus e_x$, with common entries in at least $m+1$ rows, a contradiction. □

By applying the substitution-type construction to Q_z, we obtain

Theorem 3.2 *By replacing each q-ary symbol in Q_z by a distinct column of a $t' \times q$ $(d, r_\cap; z')$-disjunct matrix, we obtain a $(drm+z)t' \times q^{m+1}$ $(d, r_\cap; zz')$-disjunct matrix M.*

Proof It suffices to prove that M is $(d, r_\cap; zz')$-disjunct. Take a pair of disjoint d-set and r-set of columns, we want to show that there exist zz' rows with 1-entries in the designated r columns and 0-entries in the designated d columns.

After transformation, each row satisfying above condition can generate z' rows each of which has a 1-entry in the designated r columns and a 0-entry in the designated d columns. By Lemma 3.1, there exist z rows whose entries in the d columns are all different from the entries in the r columns. Hence there exist zz' rows with a 1-entry in each of the r columns and a 0-entry in each of the d columns. □

4 A lower bound of $t^*(d, r_\cap, n)$ for small n with a constraint

A complex X is called an isolated complex if there exists a row covering only X. As it is not an efficient test, it is customary to assume that there are no iso-

lated complexes in a $(d, r_\cap)$-disjunct matrix, i.e., each row has strictly more than r 1's.

Let M be a $(d, r_\cap)$-disjunct matrix and let M' be obtained from M by interchanging all the 1's and 0's. D'yachkov et al. (2002) proved that M' is $(r, d_\cap)$-disjunct. In other words, if M is an optimal $(d, r_\cap)$-disjunct matrix, then M' is an optimal $(r, d_\cap)$-disjunct matrix. Hence we strengthen the assumption of no isolated complex to that each row has strictly more than r 1's and d 0's in a $(d, r_\cap)$-disjunct matrix. This assumption is made throughout the rest of the paper.

Let $t^*(d, r_\cap, n)$ be the minimum number of rows over all $(d, r_\cap)$-disjunct matrices of n columns. We define a secondary parameter w_k, the minimum cardinality of $\cap X$ over all k-sets X, and use a lower bound of it to bound $t^*(d, r_\cap, n)$.

Theorem 4.1 *For a $(d, r_\cap)$-disjunct matrix with no isolated complexes, we have $w_i - w_{i+1} \geq d$ for $i = 1, 2, \ldots, r$.*

Proof Let M be a $(d, r_\cap)$-disjunct matrix. Given $k \leq r$, let $C = \{C_1, C_2, \ldots, C_k\}$ be a set of k columns in M such that $|\bigcap_{i=1}^{k} C_i| = w_k$. Define $C' = C_0 \cup C$. Suppose $|\bigcap_{i=0}^{k} C_i| = w$.

Since $k \leq r$, we can choose a set S consisting of any other $r - k$ columns. Let $S_1 = S \cup C$ and $S_2 = S \cup C' \setminus C_1$. Then $|S_1| = |S_2| = r$. Hence S_1 and S_2 are distinct complexes in M. Since $\bigcap C' \subseteq \bigcap C$, whatever in $\bigcap C'$ but not in $\bigcap(S \cup C')$ is also not in $\bigcap(S \cup C)$. So $|\bigcap S_1 \setminus \bigcap S_2| \leq |\bigcap S_1 \setminus \bigcap(S \cup C')| = w_k - w \leq w_k - w_{k+1}$. If $w_k - w \leq d - 1$, then there exist $d - 1$ other complexes $X_1, X_2, \ldots, X_{d-1}$(since M has no isolated complexes) such that $\bigcap S_1 \subseteq \bigcup_{i=1}^{d-1} (\bigcap X_i) \cup (\bigcap S_2)$, violating the $(d, r_\cap)$-disjunct property. □

Corollary 4.2 *For a $(d, r_\cap)$-disjunct matrix with no isolated complexes, we have $w_i \geq d(r - i + 1) + 1$ for $i = 1, 2, \ldots, r$.*

Proof By Theorem 4.1, $w_i - w_r = (w_i - w_{i+1}) + (w_{i+1} - w_{i+2}) + \cdots + (w_{r-1} - w_r) \geq d(r - i)$, for $i \leq r$. Since each complex is not an isolated complex, $w_r \geq d + 1$. Thus, $w_i \geq d(r - i + 1) + 1$. □

Note that w_1 is the minimum weight over all columns.

Corollary 4.3 *A $(d, r_\cap)$-disjunct matrix with no isolated complexes has column weight at least $dr + 1$.*

Without loss of generality, assume $d \geq r$.

Theorem 4.4 $t^*(d, r_\cap, n) \geq \frac{r}{2}(d + r - 1)(d - r + 2) + \frac{r}{6}(r - 1)(4r - 5) + d + r$.

Proof To prove the theorem, we delete one column and all the intersecting rows from M step by step. Let M be an optimal $(d, r_\cap)$-disjunct matrix and let M_1 be obtained from M by deleting a column and all the intersecting rows. Then a result proved by Stinson et al. (2000) implies that M_1 is a $(d - 1, r_\cap)$-disjunct matrix.

Continue this process till a $(r-1, r_\cap)$-disjunct matrix is obtained. To have a better bound, we transform the current matrix to a $(r, (r-1)_\cap)$-disjunct matrix by interchanging all the 1's and 0's in the former one, and then keep going this process till a $(1, 1_\cap)$-disjunct matrix is obtained. By Corollary 4.3 we can count easily the number of rows deleted from M. And then we get a lower bound of $t^*(d, r_\cap, n)$, that is,

$$\begin{aligned}&(dr+1)+((d-1)r+1)+\cdots+((r-1)r+1)+((r-1)^2+1)\\&\quad+((r-2)(r-1)+1)+\cdots+(1^2+1)\\&=\frac{r}{2}(d+r-1)(d-r+2)+\frac{r}{6}(r-1)(4r-5)+d+r.\end{aligned}$$

□

Theorem 4.4 gives a lower bound with parameters d and r. For $r=1$, this bound is reduced to the famous Bassalygo $\binom{d+2}{2}$ bound (see Du and Hwang 2000 for reference).

References

Alon N, Beigel R, Kasif S, Rudich S, Sudakov B (2004) Learning a hidden matching. SIAM J Comput 33:487–501

Balding DZ, Bruno WJ, Knill E, Torney DC (1996) A comparatively survey of nonadaptive pooling designs, genetic mapping and DNA sequencing, IMA volumes in mathematics and its applications. Springer, New York, pp 133–154

Beigel R, Alon N, Apaydin MS, Fortnow L, Kasif S (2001) An optimal procedure for gap closing in whole genome shotgun sequencing. In: Proc. 2001 RECOMB. ACM Press, pp 22–30

Du DZ, Hwang FK (2000) Combinatorial group testing and its applications, 2nd edn. World Scientific, Singapore

D'yachkov A, Vilenkin PA, Macula AJ, Torney DC (2002) Families of finite sets in which no intersection of ℓ sets is covered by the union of s others. J Comb Theory Ser A 99:195–218

Engel K (1996) Interval packing and covering in the boolean lattice. Comb Probab Comput 5:373–384

Gao H, Hwang FK, Thai MT, Wu W, Znati T (2006) Construction of $d(H)$-disjunct matrix for group testing in hypergraphs. J Comb Optim 12:297–301

Grebinski V, Kucherov G (1998) Reconstructing a Hamiltonian cycle by querying the graph: application to DNA physical mapping. Discret Appl Math 88:147–165

Grebinski V, Kucherov G (2000) Optimal reconstruction of graphs under the additive model. Algorithmica 28:104–124

Kautz WH, Singleton RR (1964) Nonrandom binary superimposed codes. IEEE Trans Inform Theory 10:363–377

Kim HK, Lebedev V (2004) On optimal superimposed codes. J Comb Des 12:79–91

Li Y, Thai MT, Liu Z, Wu W (2005) Protein–protein interaction and group testing in bipartite graphs. Int J Bioinf Res Appl 1:420–428

Macula AJ, Popyack LJ (2004) A group testing method for finding patterns in data. Discret Appl Math 144:149–157

Macula AJ, Torney DC, Vilenkin PA (1999) Two-stage group testing for complexes in the presence of errors. DIMACS Ser Discret Math Theor Comput Sci 55:145–157

Macula AJ, Rykov VV, Yekhanin S (2004) Trivial two-stage group testing for complexes using almost disjunct matrices. Discret Appl Math 137:97–107

Mitchell CJ, Piper FC (1988) Key storage in secure networks. Discret Appl Math 21:215–228

Sagalovich YuL (1994) On separating systems. Probl Peredachi Inform 30:14–35 (in Russian)

Stinson DR, Wei R (2004) Generalized cover-free families. Discret Math 279:463–477

Stinson DR, Wei R, Zhu L (2000) Some new bounds for cover-free families. J Comb Theory Ser A 90:224–234

Torney DC (1999) Sets pooling designs. Ann Comb 3:95–101