Cover-free families on hypergraphs

Lucia Moura, University of Ottawa
lmoura@uottawa.ca
(joint work with Thaís Bardini Idalino)

RICCOTA, Rijeka, July 6, 2023

1. Definition and applications of cover-free families

2. CFFs on hypergraphs with non-overlapping edges

3. CFFs on hypergraphs with overlapping edges

1. Definition and applications of cover-free families

2. CFFs on hypergraphs with non-overlapping edges

3. CFFs on hypergraphs with overlapping edges

A d-cover-free family is a set system where no set is covered by the union of d others.

Definition (d-CFF(t, n))

Given $d < t \leq n$ positive integers, a *d-cover-free family*, denoted d-CFF(t, n), is a set system $(X, \mathcal{B})$ with $|X| = t$ and $|\mathcal{B}| = n$ such that for any block B_{i_0} and any other d blocks $B_{i_1}, \ldots, B_{i_d} \in \mathcal{B}$, we have

$$\left| B_{i_0} \setminus \bigcup_{j=1}^{d} B_{i_j} \right| \geq 1.$$

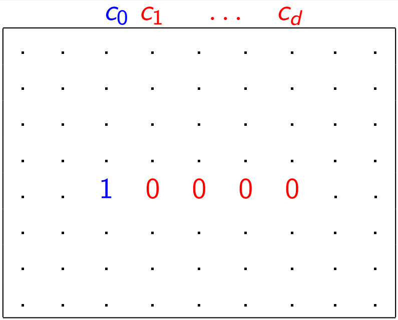
Studied under different names:

- Kautz and Singleton (1964) - superimposed codes
- Erdös, Frankl and Füredi (1985) - cover-free families
- Hwang and Sós (1987) - d-complete designs for use in group testing
- Du and Hwang book on CGT: d-disjunct matrix

A d-cover-free family is a set system where no set is covered by the union of d others.

Definition (d-CFF(t, n), alternative definition)

A d-CFF(t, n) is a $t \times n$ binary array such that any $d + 1$ columns must contain a subarray that is a permutation matrix.

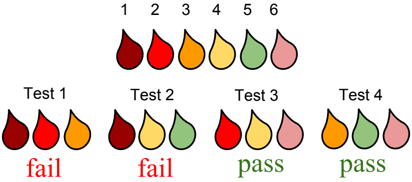


The goal is to construct arrays with minimum t for given n and d:

$$t(d, n) = \min\{t : \exists d\text{-CFF}(t, n)\}$$

Combinatorial group testing

- Identify d defective items from a set of n items pooled into t groups, where $t < n$.
- The groups are tested, instead of all elements individually.
- Use a d-CFF(t, n):
  - the n columns represent elements
  - the t rows represent the groups to be tested.



1-CFF(4,6) Matrix

	1	2	3	4	5	6
$test_1$	1	1	1	0	0	0
$test_2$	1	0	0	1	1	0
$test_3$	0	1	0	1	0	1
$test_4$	0	0	1	0	1	1

Combinatorial group testing - example $d = 2$

2 defective items, 2-CFF, the "strength" is $d + 1 = 3$

item	1	2	3	4	5	6	7	8	9	10	11	12	
input	?	?	?	?	?	?	?	?	?	?	?	?	result:
t1	1	0	0	1	0	0	1	0	0	1	0	0	?
t2	1	0	0	0	1	0	0	1	0	0	1	0	?
t3	1	0	0	0	0	1	0	0	1	0	0	1	?
t4	0	1	0	1	0	0	0	0	1	0	1	0	?
t5	0	1	0	0	1	0	1	0	0	0	0	1	?
t6	0	1	0	0	0	1	0	1	0	1	0	0	?
t7	0	0	1	1	0	0	0	1	0	0	0	1	?
t8	0	0	1	0	1	0	0	0	1	1	0	0	?
t9	0	0	1	0	0	1	1	0	0	0	1	0	?

Defective: items ?

Combinatorial group testing - example $d = 2$

2 defective items, 2-CFF, the "strength" is $d + 1 = 3$

item	1	2	3	4	5	6	7	8	9	10	11	12	
input	?	?	?	?	?	?	?	?	?	?	?	?	result:
t1	1	0	0	1	0	0	1	0	0	1	0	0	0
t2	1	0	0	0	1	0	0	1	0	0	1	0	0
t3	1	0	0	0	0	1	0	0	1	0	0	1	1
t4	0	1	0	1	0	0	0	0	1	0	1	0	0
t5	0	1	0	0	1	0	1	0	0	0	0	1	1
t6	0	1	0	0	0	1	0	1	0	1	0	0	0
t7	0	0	1	1	0	0	0	1	0	0	0	1	1
t8	0	0	1	0	1	0	0	0	1	1	0	0	1
t9	0	0	1	0	0	1	1	0	0	0	1	0	1

Defective: items ?

Combinatorial group testing - example $d = 2$

2 defective items, 2-CFF, the "strength" is $d + 1 = 3$

item	1	2	3	4	5	6	7	8	9	10	11	12	
input	0	0	?	0	0	0	0	0	0	0	0	?	result:
t1	1	0	0	1	0	0	1	0	0	1	0	0	0
t2	1	0	0	0	1	0	0	1	0	0	1	0	0
t3	1	0	0	0	0	1	0	0	1	0	0	1	1
t4	0	1	0	1	0	0	0	0	1	0	1	0	0
t5	0	1	0	0	1	0	1	0	0	0	0	1	1
t6	0	1	0	0	0	1	0	1	0	1	0	0	0
t7	0	0	1	1	0	0	0	1	0	0	0	1	1
t8	0	0	1	0	1	0	0	0	1	1	0	0	1
t9	0	0	1	0	0	1	1	0	0	0	1	0	1

Defective: items ?

Combinatorial group testing - example $d = 2$

2 defective items, 2-CFF, the "strength" is $d + 1 = 3$

item	1	2	3	4	5	6	7	8	9	10	11	12	
input	0	0	1	0	0	0	0	0	0	0	0	1	result:
t1	1	0	0	1	0	0	1	0	0	1	0	0	0
t2	1	0	0	0	1	0	0	1	0	0	1	0	0
t3	1	0	0	0	0	1	0	0	1	0	0	1	1
t4	0	1	0	1	0	0	0	0	1	0	1	0	0
t5	0	1	0	0	1	0	1	0	0	0	0	1	1
t6	0	1	0	0	0	1	0	1	0	1	0	0	0
t7	0	0	1	1	0	0	0	1	0	0	0	1	1
t8	0	0	1	0	1	0	0	0	1	1	0	0	1
t9	0	0	1	0	0	1	1	0	0	0	1	0	1

Defective: items 3 and 12

CFF SAMPLE HIGHLIGHT OF KNOWN RESULTS

DEFINITION (SAME AS BEFORE)

A d-cover-free family is a set system where no set is covered by the union of d others.

1-CFF is equivalent to a Sperner system, thus by Sperner theorem:

$$t(1, n) = \min\{t : \binom{t}{\lfloor t/2 \rfloor} \geq n\} \sim c \log n$$

Examples of constructions for $d \geq 2$ from designs and codes:

- any s-$(v, k, 1)$-packing is a d-CFF(v, b) with $d = \lfloor \frac{k-1}{s-1} \rfloor$,
- any BIBD$(v, k, 1)$ gives a $(k-1)$-CFF(v, b), where $b = \binom{v}{k} / \binom{k}{2}$.
- any code $(L, n, D)_q$ with gives a d-CFF(qL, n) with $d = \lfloor \frac{L-1}{L-D} \rfloor$,

Asymptotics: There exists c_1 and c_2 such that

$$c_1 \frac{d^2}{\log d} \log n \leq t(d, n) \leq c_2 d^2 \log n$$

News in focus

NICOLAS ASFOURI/AFP/GETTY

Nasal or throat swabs from several people can be combined in a single test.

THE MATHEMATICAL STRATEGY THAT COULD TRANSFORM CORONAVIRUS TESTING

To save time and money, several countries are using a technique called group testing, which pools samples from many people.

By Smriti Mallapaty

Scientists say that widespread testing is needed to get outbreaks of the new coronavirus under control. But, in many regions, there's a shortage of the chemicals needed to run diagnostics. Researchers are scrambling to devise faster, simpler tests (see page 506). Now, in several countries, officials have started using a strat- are already using group testing.

There are many ways to conduct group testing, and scientists in several countries are experimenting with the best method for doing this during a pandemic. Their ideas largely come from a field of mathematics that has been applied to a wide range of problems, from detecting faulty Christmas-tree lights to estimating the prevalence of HIV in a population. "There has been a flurry of innovation in this mixed together and tested once (see 'Group testing', Method 1). Groups of samples that test negative are ruled out. But if a group tests positive, every sample in that group is then retested individually. Researchers estimate the most efficient group size – the one that uses the least number of tests – on the basis of the prevalence of the virus in the community.

In May, officials in Wuhan, China, used Method 1 as part of their efforts to test the

Group testing application

Determine a set of up to d defective items among n items, using t tests.

	1	2	3	4	5	6	
Test 1	1	1	1	0	0	0	FAIL
Test 2	1	0	0	1	1	0	FAIL
Test 3	0	1	0	1	0	1	PASS
Test 4	0	0	1	0	1	1	PASS

Group testing using a 1-CFF(4, 6)

For **non-adaptive** group testing, we can use a d-CFF(t, n).

- Good when the probabilities of different items being defective are independent.
- CFFs on hypergraphs can be better when these probabilities are highly dependent (e.g. infectious diseases)

"no column is covered by the union of columns **that form an edge**"

Definition (CFF on hypergraph)

A CFF on a hypergraph H, denoted H-CFF(t, n), is a $t \times n$ binary matrix so that for every edge $e = \{c_1, \ldots, c_d\}$ and any vertex c_0 not on the edge, we must see the occurrence of row with $[10\ldots0]$ in the submatrix defined by $c_0, c_1, \ldots, c_d$.

		c_0	c_1	$\ldots$		c_d		
.	.	.	.	.	.	.	.	.
.	.	.	.	.	.	.	.	.
.	.	.	.	.	.	.	.	.
.	.	1	0	0	0	0	.	.
.	.	.	.	.	.	.	.	.
.	.	.	.	.	.	.	.	.
.	.	.	.	.	.	.	.	.

When this property holds for each subset of an edge we call it an

$$\overline{H}\text{-CFF}(t, n)$$

Definitions of CFFs on hypergraphs

Let $H = (V(H), E(H))$ be a hypergraph with $V(H) = n$ and $\text{rank}(H) = \max\{|E| : e \in E\} < n$.
A CFF on H is a $t \times n$ binary array satisfying one of the following properties (depending on the CFF type being defined):

- H-CFF: for each set of columns corresponding to an edge E and each vertex $v \notin E$, there must be a row with 0...0 on the columns of E and a 1 on the column of v.
- $\overline{H}$-CFF: for each set of columns corresponding to $S \subseteq E$, for edge E, and each vertex $v \notin S$, there must be a row with 0...0 on columns of s and a 1 on the column of v.
- (H, r)-CFF: similar definition as H-CFF with union of up to r edges $E_1, \ldots, E_r$ in place of E.
- $(\overline{H}, r)$-CFF: similar definition as $\overline{H}$-CFF with union of up to r edges $E_1, \ldots, E_r$ in place of E.

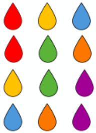
1. Definition and applications of cover-free families

2. CFFs on hypergraphs with non-overlapping edges

3. CFFs on hypergraphs with overlapping edges

ordinary *d*-CFF

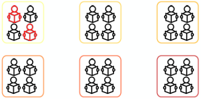
	1	2	3	4	5	6
Test 1	1	1	1	0	0	0
Test 2	1	0	0	1	1	0
Test 3	0	1	0	1	0	1
Test 4	0	0	1	0	1	1



FAIL
FAIL
PASS
PASS

$\overline{H}$-CFF (disjoint edges)

1	1	1	1	1	1	1	1	1	1	1	1	0	0	0	0	0	0	0	0	0	0	0	0	FAIL
1	1	1	1	0	0	0	0	0	0	0	0	1	1	1	1	1	1	1	1	0	0	0	0	FAIL
0	0	0	0	1	1	1	1	0	0	0	0	1	1	1	1	0	0	0	0	1	1	1	1	PASS
0	0	0	0	0	0	0	0	1	1	1	1	0	0	0	0	1	1	1	1	1	1	1	1	PASS
1	0	0	0	1	0	0	0	1	0	0	0	1	0	0	0	1	0	0	0	1	0	0	0	FAIL
0	1	0	0	0	1	0	0	0	1	0	0	0	1	0	0	0	1	0	0	0	1	0	0	PASS
0	0	1	0	0	0	1	0	0	0	1	0	0	0	1	0	0	0	1	0	0	0	1	0	FAIL
0	0	0	1	0	0	0	1	0	0	0	1	0	0	0	1	0	0	0	1	0	0	0	1	PASS



Non-overlapping edges, $r = 1$

Let $H = (V(H), E(H))$ be a hypergraph such that any two edges are disjoint.
Let $n = |V(G)|$, $m = |E(G)|$, $k = rank(H)$.

Theorem (Idalino and M. 2022)

Let $t_1 = \min\{s : \binom{s}{\lfloor s/2 \rfloor} \geq m\}$. *Then,*

- $\exists$ *H-CFF*(t_1, n); *in addition,* $t_1 \sim \Theta(\log m)$ *is best possible;*
- $\exists$ $\overline{H}$*-CFF*(t, n) *with* $t = t_1 + k \sim \Theta(\log m + k)$.

1	1	1	1	1	1	1	1	1	1	1	1	0	0	0	0	0	0	0	0	0	0	0	0	FAIL
1	1	1	1	0	0	0	0	0	0	0	0	1	1	1	1	1	1	1	1	0	0	0	0	FAIL
0	0	0	0	1	1	1	1	0	0	0	0	1	1	1	1	0	0	0	0	1	1	1	1	PASS
0	0	0	0	0	0	0	0	1	1	1	1	0	0	0	0	1	1	1	1	1	1	1	1	PASS
1	0	0	0	1	0	0	0	1	0	0	0	1	0	0	0	1	0	0	0	1	0	0	0	FAIL
0	1	0	0	0	1	0	0	0	1	0	0	0	1	0	0	0	1	0	0	0	1	0	0	PASS
0	0	1	0	0	0	1	0	0	0	1	0	0	0	1	0	0	0	1	0	0	0	1	0	FAIL
0	0	0	1	0	0	0	1	0	0	0	1	0	0	0	1	0	0	0	1	0	0	0	1	PASS

Examples:

- 900 students = 30 classes of 30 students: 37 tests.
- 1000 people = 250 households of 4 people: 14 tests.

NON-OVERLAPPING EDGES, $r \geq 2$

Let $H = (V(H), E(H))$ be a hypergraph such that any two edges are disjoint. Let $n = |V(G)$, $m = |E(G)|$, $k = rank(H)$.

THEOREM (IDALINO AND M. 2022)

Let $r \geq 2$.
Suppose $\exists$ r-CFF(t_A, m), A_1, and $\exists$ $(r-1)$-CFF(t_B, m), A_2. Then,

- $\exists$ (H, r)-CFF(t_A, n);
- $\exists$ $(\overline{H}, r)$-CFF$(t_A + kt_B, n)$.

PROOF:
Let $B_1 = A_1 \otimes R_k$, where $R_k = [1...1]$ with k columns.
Let $B_2 = A_2 \otimes I_k$. In both cases, remove some columns for e with $|e| < k$.
B_1 is an (H, r)-CFF(t_A, n), and
$\begin{bmatrix} B_1 \\ B_2 \end{bmatrix}$ is an $(\overline{H}, r)$-CFF$(t_A + kt_B, n)$. $\square$

$r = 2,\ H = 12$ DISJOINT TRIPLES OF VERTICES

2-CFF(9, 12):

$$A = \begin{pmatrix} 1&0&0&1&0&0&1&0&0&1&0&0\\ 1&0&0&0&1&0&0&1&0&0&1&0\\ 1&0&0&0&0&1&0&0&1&0&0&1\\ 0&1&0&1&0&0&0&0&1&0&1&0\\ 0&1&0&0&1&0&1&0&0&0&0&1\\ 0&1&0&0&0&1&0&1&0&1&0&0\\ 0&0&1&1&0&0&0&1&0&0&0&1\\ 0&0&1&0&1&0&0&0&1&1&0&0\\ 0&0&1&0&0&1&1&0&0&0&1&0 \end{pmatrix} \qquad R_3 = \begin{pmatrix} 1 & 1 & 1 \end{pmatrix},$$

1-CFF(6, 12):

$$B = \begin{pmatrix} 1&1&1&1&1&1&0&0&0&0&0&0\\ 1&1&1&1&0&0&1&1&0&0&0&0\\ 1&0&0&0&1&1&0&0&1&1&1&0\\ 0&1&0&0&1&0&1&0&1&1&0&1\\ 0&0&1&0&0&1&0&1&1&0&1&1\\ 0&0&0&1&0&0&1&1&0&1&1&1 \end{pmatrix} \qquad I_3 = \begin{pmatrix} 1&0&0\\ 0&1&0\\ 0&0&1 \end{pmatrix},$$

$(\overline{H}, 2)$-CFF(27, 36):

$$\left(\frac{A \otimes R_3}{B \otimes I_3}\right) = \left(\begin{array}{cccccccccccc} R_3&0&0&R_3&0&0&R_3&0&0&R_3&0&0\\ R_3&0&0&0&R_3&0&0&R_3&0&0&R_3&0\\ R_3&0&0&0&0&R_3&0&0&R_3&0&0&R_3\\ 0&R_3&0&R_3&0&0&0&0&R_3&0&R_3&0\\ 0&R_3&0&0&R_3&0&R_3&0&0&0&0&R_3\\ 0&R_3&0&0&0&R_3&0&R_3&0&R_3&0&0\\ 0&0&R_3&R_3&0&0&0&R_3&0&0&0&R_3\\ 0&0&R_3&0&R_3&0&0&0&R_3&R_3&0&0\\ 0&0&R_3&0&0&R_3&R_3&0&0&0&R_3&0\\ \hline I_3&I_3&I_3&I_3&I_3&I_3&0&0&0&0&0&0\\ I_3&I_3&I_3&I_3&0&0&I_3&I_3&0&0&0&0\\ I_3&0&0&0&I_3&I_3&0&0&I_3&I_3&I_3&0\\ 0&I_3&0&0&I_3&0&I_3&0&I_3&I_3&0&I_3\\ 0&0&I_3&0&0&I_3&0&I_3&I_3&0&I_3&I_3\\ 0&0&0&I_3&0&0&I_3&I_3&0&I_3&I_3&I_3 \end{array}\right).$$

Non-overlapping edges - asymptotics

Define $t(H, r) = \min\{t : \exists\text{(H,r)}-CFF(t, |V(H)|)\}$.

Let $H_{m,k} =$ hypergraph with m disjoint edges of cardinality k.

Then, for $r \geq 2$,

$$
\begin{aligned}
t(H_{m,k}, 1) &= \Theta(\log m) \\
t(\overline{H}_{m,k}, 1) &= O(k + \log m) \\
t(H_{m,k}, r) &= O(((r+1)^2 \log m) \\
t(\overline{H}_{m,k}, r) &= O(((r+1)^2 + kr^2) \log m)
\end{aligned}
$$

1. Definition and applications of cover-free families

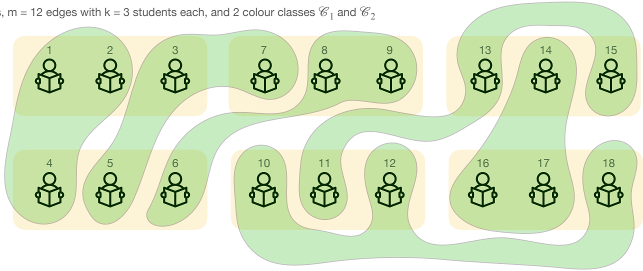
2. CFFs on hypergraphs with non-overlapping edges

3. **CFFs on hypergraphs with overlapping edges**

The high school problem

Construction

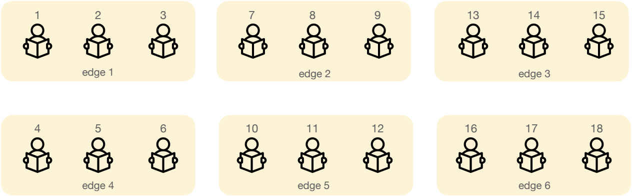
Total:
n = 18 vertices, m = 12 edges with k = 3 students each, and 2 colour classes $\mathscr{C}_1$ and $\mathscr{C}_2$



The high school problem

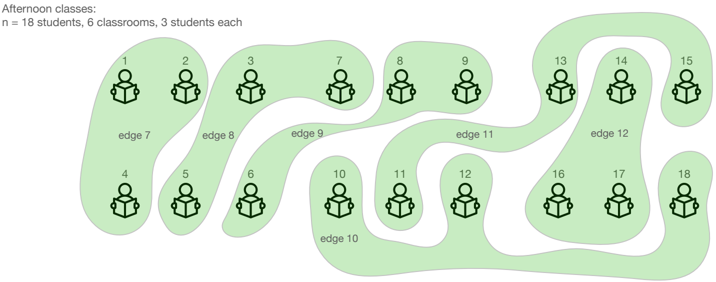
Construction

Morning classes:
n = 18 students, 6 classrooms, 3 students each



The high school problem

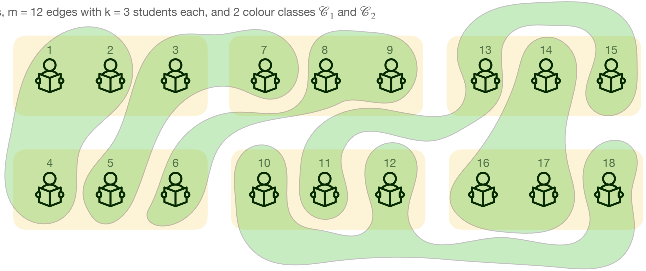
Construction



The high school problem

Construction

Total:
n = 18 vertices, m = 12 edges with k = 3 students each, and 2 colour classes $\mathscr{C}_1$ and $\mathscr{C}_2$

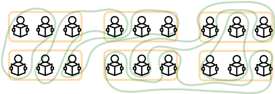


Overlapping edge construction

1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18
1	1	1	1	1	1	1	1	1	0	0	0	0	0	0	0	0	0
1	1	1	0	0	0	0	0	0	1	1	1	1	1	1	0	0	0
0	0	0	1	1	1	0	0	0	1	1	1	0	0	0	1	1	1
0	0	0	0	0	0	1	1	1	0	0	0	1	1	1	1	1	1
1	0	0	1	0	0	1	0	0	1	0	0	1	0	0	1	0	0
0	1	0	0	1	0	0	1	0	0	1	0	0	1	0	0	1	0
0	0	1	0	0	1	0	0	1	0	0	1	0	0	1	0	0	1

1	1	1	1	1	1	1	1	1	0	0	0	0	0	0	0	0	0
1	1	0	1	0	0	0	0	0	1	1	1	1	0	1	0	0	1
0	0	1	0	1	0	1	0	0	1	0	1	0	1	0	1	1	1
0	0	0	0	0	1	0	1	1	0	1	0	1	1	1	1	1	0
1	0	1	0	0	1	0	0	0	1	1	0	0	1	0	0	0	0
0	1	0	0	1	0	0	1	0	0	0	1	1	0	0	1	0	0
0	0	0	1	0	0	1	0	1	0	0	0	0	0	1	0	1	1

$(\mathcal{S},1) - CFF(t,n)$



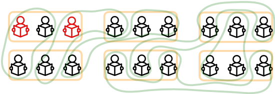
Overlapping edge construction

Edges 1, 7 and 8 are infected

1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18
1	1	1	1	1	1	1	1	1	0	0	0	0	0	0	0	0	0
1	1	1	0	0	0	0	0	0	1	1	1	1	1	1	0	0	0
0	0	0	1	1	1	0	0	0	1	1	1	0	0	0	1	1	1
0	0	0	0	0	0	1	1	1	0	0	0	1	1	1	1	1	1
1	0	0	1	0	0	1	0	0	1	0	0	1	0	0	1	0	0
0	1	0	0	1	0	0	1	0	0	1	0	0	1	0	0	1	0
0	0	1	0	0	1	0	0	1	0	0	1	0	0	1	0	0	1

1	1	1	1	1	1	1	1	1	0	0	0	0	0	0	0	0	0
1	1	0	1	0	0	0	0	0	1	1	1	1	0	1	0	0	1
0	0	1	0	1	0	1	0	0	1	0	1	0	1	0	1	1	1
0	0	0	0	0	1	0	1	1	0	1	0	1	1	1	1	1	0
1	0	1	0	0	1	0	0	0	1	1	0	0	1	0	0	0	0
0	1	0	0	1	0	0	1	0	0	0	1	1	0	0	1	0	0
0	0	0	1	0	0	1	0	1	0	0	0	0	0	1	0	1	1

$(\mathcal{S},1) - CFF(t,n)$



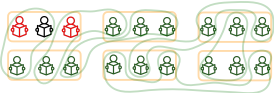
Overlapping edge construction

Edges 1, 7 and 8 are infected,
students 4-18 are cleared out

1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18
1	1	1	1	1	1	1	1	1	0	0	0	0	0	0	0	0	0
1	1	1	0	0	0	0	0	0	1	1	1	1	1	1	0	0	0
0	0	0	1	1	1	0	0	0	1	1	1	0	0	0	1	1	1
0	0	0	0	0	0	1	1	1	0	0	0	1	1	1	1	1	1
1	0	0	1	0	0	1	0	0	1	0	0	1	0	0	1	0	0
0	1	0	0	1	0	0	1	0	0	1	0	0	1	0	0	1	0
0	0	1	0	0	1	0	0	1	0	0	1	0	0	1	0	0	1

1	1	1	1	1	1	1	1	1	0	0	0	0	0	0	0	0	0
1	1	0	1	0	0	0	0	0	1	1	1	1	0	1	0	0	1
0	0	1	0	1	0	1	0	0	1	0	1	0	1	0	1	1	1
0	0	0	0	0	1	0	1	1	0	1	0	1	1	1	1	1	0
1	0	1	0	0	1	0	0	0	1	1	0	0	1	0	0	0	0
0	1	0	0	1	0	0	1	0	0	0	1	1	0	0	1	0	0
0	0	0	1	0	0	1	0	1	0	0	0	0	0	1	0	1	1

$$(\mathcal{S},1) - CFF(t,n)$$



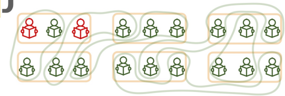
Overlapping edge construction

Edges 1, 7 and 8 are infected, students 4-18 are cleared out

1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18
1	1	1	1	1	1	1	1	1	0	0	0	0	0	0	0	0	0
1	1	1	0	0	0	0	0	0	1	1	1	1	1	1	0	0	0
0	0	0	1	1	1	0	0	0	1	1	1	0	0	0	1	1	1
0	0	0	0	0	0	1	1	1	0	0	0	1	1	1	1	1	1
1	0	0	1	0	0	1	0	0	1	0	0	1	0	0	1	0	0
0	1	0	0	1	0	0	1	0	0	1	0	0	1	0	0	1	0
0	0	1	0	0	1	0	0	1	0	0	1	0	0	1	0	0	1
1	1	1	1	1	1	1	1	1	0	0	0	0	0	0	0	0	0
1	1	0	1	0	0	0	0	0	1	1	1	1	0	1	0	0	1
0	0	1	0	1	0	1	0	0	1	0	1	0	1	0	1	1	1
0	0	0	0	0	1	0	1	1	0	1	0	1	1	1	1	1	0
1	0	1	0	0	1	0	0	0	1	1	0	0	1	0	0	0	0
0	1	0	0	1	0	0	1	0	0	0	1	1	0	0	1	0	0
0	0	0	1	0	0	1	0	1	0	0	0	0	0	1	0	1	1

Inside each infected edge, we can identify precisely the infected students

$$(\bar{\mathcal{S}},1) - CFF(t = 2 \times (4+3), n = 18)$$
$$(\mathcal{S},1) - CFF(t' = 2 \times 4, n = 18)$$



Defect cover and (H, r)-CFFs

An edge is *defective* if it contains a defective vertex, it is *non-defective*, otherwise. A set of edges is a *defect cover* if the set of defective vertices is contained in the union of these edges; such a defect cover is *minimal* if no proper subset is a defect cover.

The number of defective edges may be much larger than the size of a defect cover!!!

The next proposition shows that a $(\overline{H}, r)$-CFF's ability to detect defectives only requires that r is an upper bound on the cardinality of a minimum defect cover.

Proposition (Idalino and M. 2022)

Let H be a hypergraph and let r be an upper bound on the size of a minimal defect cover.

A simple algorithm using an $(\overline{H}, r)$-CFF determines the set of defective items.

A simple algorithm using an (H, r)-CFF returns a defect cover of size at most r.

CONSTRUCTION 1 ($\exists$ DEFECT COVER OF SIZE 1)

An ℓ-edge-colouring of a hypergraph $H = (V(H), E(H))$ is a mapping from $E(H)$ to $\{1, \ldots, \ell\}$ such that no vertex is incident to more than one edge mapping to the same colour.

Let $H_1, H_2, \ldots, H_\ell$ be the hypergraphs on $V(H)$ with edges consisting of each colour class.

THEOREM (IDALINO AND M. 2022; DEFECT COVER SIZE $r = 1$)

For each i, $1 \leq i \leq \ell$,
let $k_i = rank(H_i) \max\{|e| : e \in E(H_i)\}$,
let $f_i = |E(H_i)| + \delta_i$, where $\delta_i = 0$ if the $E(H_i)$ spans $[1, n]$ and $\delta_i = 1$, ow.
For $i = 1, 2, \ldots, \ell$, use previous theorem to build $(\overline{H_i}, 1)$-CFF$(t_i + k_i, n)$ and $(H_i, 1)$-CFF$(t_i + k_i, n)$, where $t_i = \min\{s : \binom{s}{\lfloor s/2 \rfloor} \geq f_i\}$.
Then,
the horiz. pasting of all $(H_i, 1)$-CFFs gives $(H, 1)$-CFF$(\sum_{i=1}^{\ell} t_i, n)$; and
the horiz. pasting of all $(\overline{H_i}, 1)$-CFFs gives $(\overline{H}, 1)$-CFF$(\sum_{i=1}^{\ell}(t_i + k_i), n)$.

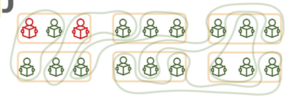
Overlapping edge construction

1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18
1	1	1	1	1	1	1	1	1	0	0	0	0	0	0	0	0	0
1	1	1	0	0	0	0	0	0	1	1	1	1	1	1	0	0	0
0	0	0	1	1	1	0	0	0	1	1	1	0	0	0	1	1	1
0	0	0	0	0	0	1	1	1	0	0	0	1	1	1	1	1	1
1	0	0	1	0	0	1	0	0	1	0	0	1	0	0	1	0	0
0	1	0	0	1	0	0	1	0	0	1	0	0	1	0	0	1	0
0	0	1	0	0	1	0	0	1	0	0	1	0	0	1	0	0	1
1	1	1	1	1	1	1	1	1	0	0	0	0	0	0	0	0	0
1	1	0	1	0	0	0	0	0	1	1	1	1	0	1	0	0	1
0	0	1	0	1	0	1	0	0	1	0	1	0	1	0	1	1	1
0	0	0	0	0	1	0	1	1	0	1	0	1	1	1	1	1	0
1	0	1	0	0	1	0	0	0	1	1	0	0	1	0	0	0	0
0	1	0	0	1	0	0	1	0	0	0	1	1	0	0	1	0	0
0	0	0	1	0	0	1	0	1	0	0	0	0	0	1	0	1	1

Edges 1, 7 and 8 are infected, students 4-18 are cleared out

Inside each infected edge, we can identify precisely the infected students

$$(\overline{\mathcal{S}},1) - CFF(t = 2 \times (4+3), n = 18)$$
$$(\mathcal{S},1) - CFF(t' = 2 \times 4, n = 18)$$



CONSTRUCTION 2 ($\exists$DEFECT COVER OF SIZE r, $r \geq 2$)

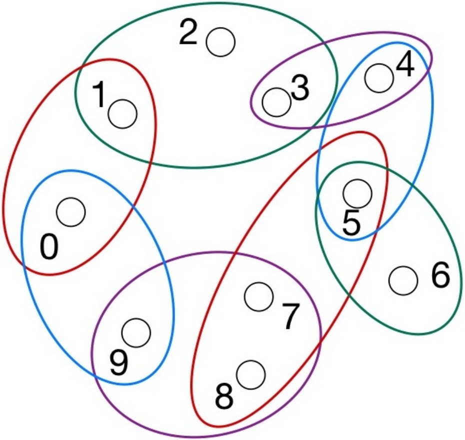
A *strong edge-coloring* of a hypergraph H is an edge-coloring such that any two vertices belonging to distinct edges with the same colour are not adjacent.

THEOREM (IDALINO AND M. 2022; DEFECT COVER SIZE $r \geq 2$)

Let H be a hypergraph and let $r \geq 2$ be an upper bound on the number of edges of a minimal defective cover.
Let $H_1, H_2, \ldots, H_s$ be the hypergraphs on $V(H)$ and sets of edges equal each colour class of an s-strong-edge-colouring of H.
Let $k_i = \max\{|S| : S \in E(H_i)\}$ and let $m_i = |E(H_i)|$.
Then, there exists an $(\overline{H}, r)$-CFF(t, n) with

$$t \leq \sum_{i=1}^{s'} (t(r, m_i) + k_i \cdot t(r-1, m_i)).$$

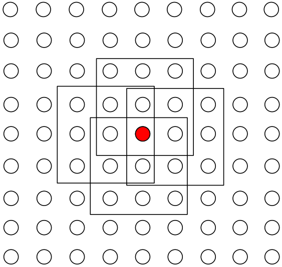
$$t(H_{k,m,s}, r) = O(s \cdot k \cdot r^2 \cdot \log m_{max})$$



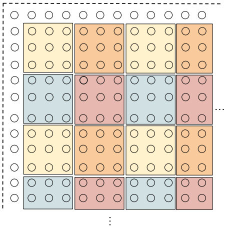

```
0123456789
1100000000
0000010110   M1
-------------
1000000000
0100000000
0000010000   N1
0000000100
0000000010
0111000000   M2
0000011000
-------------
0100000000
0010000000
0001000000   N2
0000010000
0000001000
0001100000   M3
0000000111
-------------
0001000000
0000100000
0000000100   N3
0000000010
0000000001
0000110000   M4
1000000001
-------------
0000100000
0000010000   N4
1000000000
0000000001
```

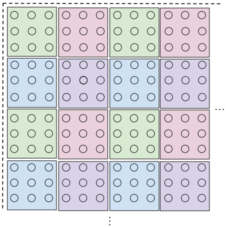
Consecutive Hypergraph (dimension 2)

Example 5. Consider a venue with 4356 people sitting in a square auditorium of 66 rows with 66 seats per row. Edges are sets of individuals sitting nearby. We consider edges of size 9 consisting of all possible contiguous 3×3 squares. There is a strong colouring with $\ell = 36$ colour classes of $11 \times 11 = 121$ edges each: we need 4 colours to "tile" the room with edges and 9 such tilings to cover all edges. Thm. 5 constrution vertically concatenates matrices $M_1, \ldots, M_{36}, N_1, \ldots, N_{36}$. For each colour class i, we use a 2-CFF(25, 125) using the polynomial construction from Proposition 8 for $q = 5$ and $k = 2$ to build each M_i, totalling $36 \times 25 = 900$ tests. For $N_1, \ldots, N_{36}$, each of which is supposed to be a 1-CFF$(t, 121)$ multiplied by I_9, we use instead a single matrix N built as follows. Take A as a 1-CFF(12, 484) obtained from the Sperner construction and do $N = A \otimes I_9$ with 108 rows. Carefully assign vertices in the grid to the columns of matrix N so that each 3×3 square corresponds to a block of identity matrix I_9 in a tiling fashion. This is enough to identify each non-defective vertex that lies inside one of the two defective edges, which is the purpose of N. Therefore with a total of 1008 tests we can screen 4356 people for any 18 infected people that appear within any 2 regions of size 3×3. See pictures for this example in [18].



Repeat the following pattern 9 times (total 36 colour classes):

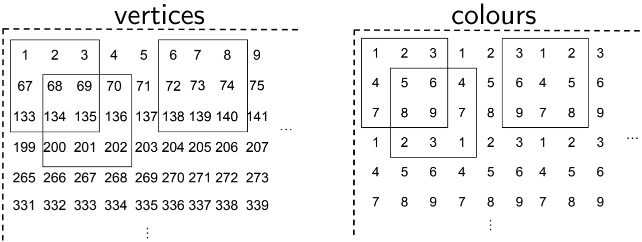




$(H, 2)$-CFF$(900, 4356)$ can be built from 36 copies of 2-CFF$(25, 125)$

Consecutive Hypergraph (dimension 2), cont'd

But we can improve the construction of $\overline{H}$-CFF, for the consecutive hypergraph.
Use a vertex colour with $k = 9$ colours:



An $(\overline{H}, 2)$-CFF$(1008, 4356)$ can be built from:
36*($(H, 2)$-CFF$(25, 125)$) (900 rows)
$+$ 1-CFF$(12, 484) \otimes I_9$ (108 rows)

In general, we reduce $t(\overline{H}, r)$ from $s \times (t(r, n/k) + k \times t(r-1, n/k))$ to $s \times t(r, n/k) + k \times t(r-1, n/k)$.

Ongoing research and future work

More research is required for the overlapping case.

- Strong colouring can be prohibitively large, so we also need constructions that don't use strong colouring.
- New constructions can decrease the number of rows for special classes of hypergraphs, like we did for consecutive hypergraphs.
- We plan to extend the model with a limit of d defective items inside an edge.
- We are studying constructions for products of hypergraphs, and characterizations via directed-hypergraph homomorphisms (under study with Prangya Parida).

Thank you !

References

1. Idalino, T.B., Moura, L., Structure-Aware Combinatorial Group Testing: A New Method for Pandemic Screening, Combinatorial Algorithms. IWOCA 2022. LNCS **13270** (2022) https://arxiv.org/abs/2202.09264
2. Gonen, M., Langberg, M., Sprintson, A., Group Testing on General Set-Systems, arXiv manuscript, (2022). https://doi.org/10.48550/arXiv.2202.04988
3. Nikolopoulos, P., Rajan Srinivasavaradhan, S., Guo, T., Fragouli, C., Diggavi, S. Group testing for connected communities. In: 24th International Conference on Artificial Intelligence and Statistics on Proc. of Machine Learning Research, pp. 2341–2349. PMLR, (2021).
4. Nikolopoulos, P., Rajan Srinivasavaradhan, S., Guo, T., Fragouli, C., Diggavi, S. Group testing for overlapping communities. In: ICC 2021 - IEEE Intern. Confer. Communic., pp. 1–7. IEEE (2021).
5. Mallapaty, S.: The mathematical strategy that could transform coronavirus testing. Nature, **583**(7817), 504–505 (2020).
6. T. B. Idalino, L. Moura, A survey of cover-free families: constructions, applications and generalizations. Stinson66 - New Advances in Designs,