The Applications of Graph Covering and Graph Packing

- ***Hung-Lin Fu Department of Applied Mathematics***
- ***National Chiao Tung University***
- ***Hsin-Chu, Taiwan***

Graph covering and packing

- Let G be a graph. A *graph covering* (*packing*) is a collection of subgraphs of G such that each edge of G is belonged to at least (at most) one subgraph of the collection.
- If each edge is belonged to the subgraphs of the collection exactly once, then we have a *graph decomposition*.

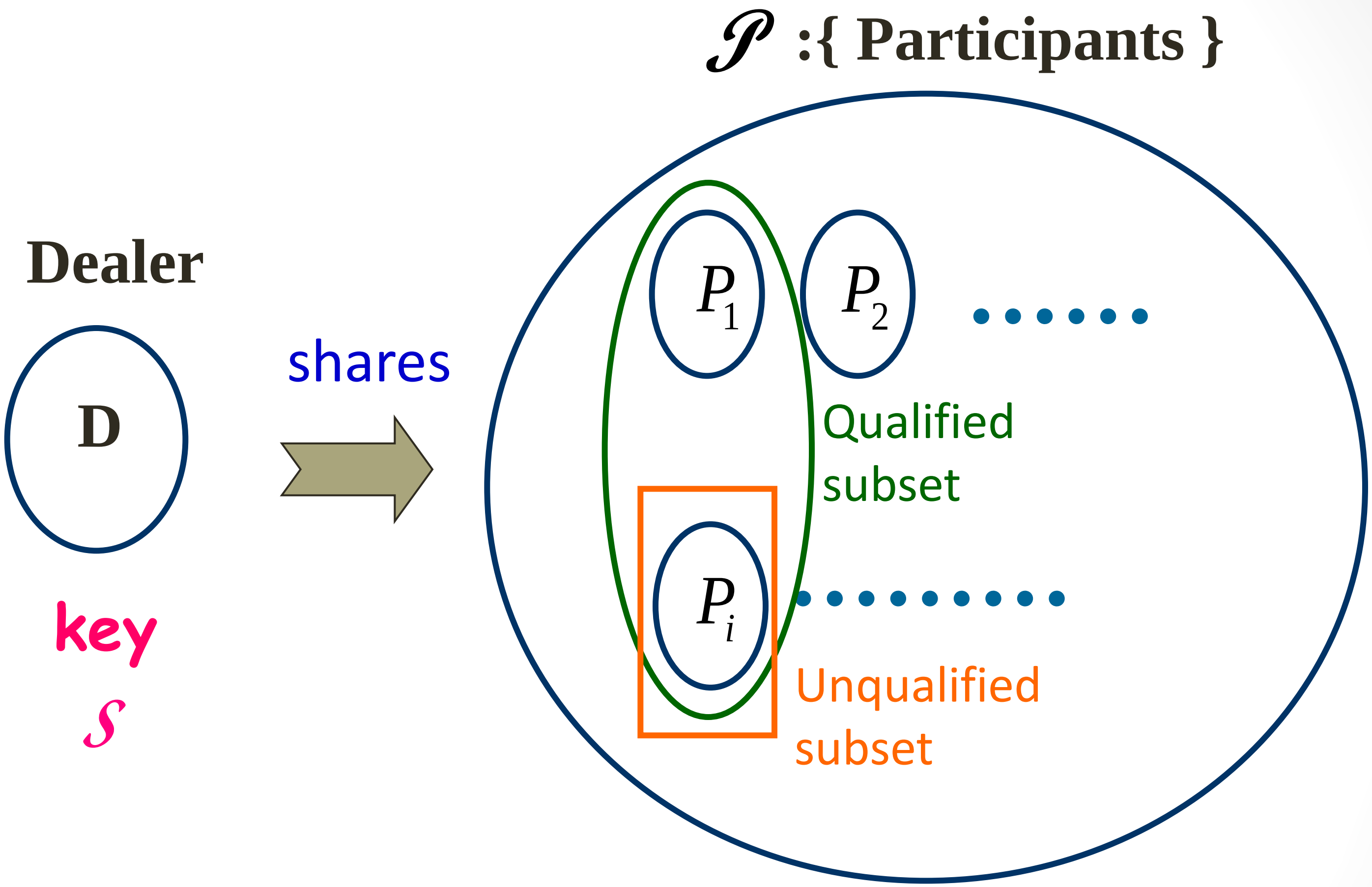
Pairs to be considered

- In case that we need to cover all the pairs considered in the problem, then graph covering is needed.
- On the other hand, if we are looking for an estimation or running an experiment, we may not need to consider the outcome of all pairs and thus packing will be enough.

Why not decomposition?

- In general, it is harder to find a suitable decomposition of a graph.
- If indeed we have a decomposition, then we are better off. In fact, this provides a better result!
- *For sure, using decomposition is the most economic way to cover a graph.*

Sharing scheme



Access structure: $\Gamma = \{A \subseteq \mathscr{P} \mid A \text{ is a qualified subset}\}$

Information ratio

$\mathscr{P} = \{P_1, P_2, \cdots, P_n\}$; $\boldsymbol{\Sigma}$: secret-sharing scheme

$\mathcal{S}$ = the set of all keys; S_i={all the shares of participant P_i }

- Γ is a *monotone* access structure if

  $\forall$ B $\subseteq \mathscr{P}$, if A $\subseteq$ B for some A$\in \Gamma$, then B $\in \Gamma$.

- Γ_0 ={A $\subseteq \mathscr{P}$ | A is a minimal qualified subset} is a *basis* of Γ

- The information ratio of $\boldsymbol{\Sigma}$ is $R_{\boldsymbol{\Sigma}} = \dfrac{\max\{\log_2|S_i| : 1 \le i \le n\}}{\log_2|\mathcal{S}|}$ ≥ 1

- The average information ratio of $\boldsymbol{\Sigma}$ is

$$R_{\boldsymbol{\Sigma}} \ge \quad AR_{\boldsymbol{\Sigma}} = \frac{\sum_{i=1}^{n} \log_2|S_i|}{n\log_2|\mathcal{S}|} \quad \ge 1$$

$$R_{\boldsymbol{\Sigma}} \ge AR_{\boldsymbol{\Sigma}} \ge 1 \qquad\qquad R_{\boldsymbol{\Sigma}} = 1 \Leftrightarrow AR_{\boldsymbol{\Sigma}} = 1$$

Graph-based access structure

G : a finite simple graph without isolated vertices.

Let $\mathscr{P} = V(G)$ and $\Gamma_0 = E(G)$

- A secret-sharing scheme for the access structure based on G is a collection of random variables $\zeta_{\mathcal{S}}$ and ζ_v for $v \in V(G)$ with a joint distribution such that

  (i) $\zeta_{\mathcal{S}}$ is the secret and ζ_v is the share of v;

  (ii) if $uv \in E(G)$, then ζ_u and ζ_v together determine the value of $\zeta_{\mathcal{S}}$;

  (iii) if $A \subseteq V(G)$ is an independent set in G, then $\zeta_{\mathcal{S}}$ and the collection $\{\zeta_v \mid v \in A\}$ are statistically independent.

Optimal (Average) Information Ratio of G

Σ : secret-sharing scheme

H : the Shannon entropy

- The information ratio of Σ is $R_\Sigma = \max_{v \in V(G)} \{H(\zeta_v)/H(\zeta_S)\}$.
- The average information ratio of Σ is

$$AR_\Sigma = \frac{\sum_{v \in V(G)} H(\zeta_v)}{nH(\zeta_S)}.$$

- The optimal information ratio of G is

$$R(G) = \inf\{R_\Sigma \mid \Sigma \text{ is a secret-sharing scheme on } G\}$$

- The optimal average information ratio of G is

$$AR(G) = \inf\{AR_\Sigma \mid \Sigma \text{ is a secret-sharing scheme on } G\}$$

$R(G) \geq AR(G) \geq 1$ $\qquad$ $R(G) = 1 \Leftrightarrow AR(G) = 1$

Known results

Theorem (Brickell and Devenport,1991)

G : connected of order n

G is a complete multipartite graph

$\Leftrightarrow R(G) = AR(G) = 1$

Theorem (Blundo et al.,1995)

G: connected of order n

G is not a complete multipartite graph

$$\Leftrightarrow R(G) \geq \frac{3}{2} \Leftrightarrow AR(G) \geq \frac{n+1}{n}$$

Covering works

Theorem (Stinson,1994)

G : a connected graph with $V(G)=\{1,2,...,n\}$. Suppose $\Pi=\{G_1, G_2, \cdots, G_h\}$ is a complete multipartite covering of G.

Let $R_i=\left|\{j : i \in V(G_j)\}\right|, 1 \le i \le n$.

Then there exists a secret-sharing scheme with information ratio R and average information ratio AR, where

$$R=\max\left\{R_i : 1 \le i \le n\right\} \text{ and } AR=\frac{\sum_{i=1}^{n} R_i}{n}=\frac{\sum_{i=1}^{h} |V(G_i)|}{n}.$$

vertex-number sum of Π

Information Ratio

Theorem (Stinson,1994)

(1) Suppose G is a connected graph of order n and d is the maximum degree of G. Then

$$R(G) \le \frac{d+1}{2}.$$

(2) $R(P_n) = \frac{3}{2}$, if $n \ge 3$; $R(C_n) = \frac{3}{2}$, if $n \ge 5$.

The exact value of $R(G)$ is known for all graphs of order ≤ 6 but some exceptions.

Average information ratio

Theorem (Stinson,1994)

$$AR(G) \le \frac{2e+n}{2n} \text{ where } e = |E(G)|.$$

Theorem (Blundo et al.,1995)

(1) $AR(C_n) = \frac{3}{2}$, if $n \ge 5$.

(2) $AR(P_n) = \begin{cases} \frac{3n}{2(n+1)}, & n: \text{even} \\ \frac{3n+1}{2(n+1)}, & n: \text{odd} \end{cases}$, if $n \ge 3$.

The exact value of $AR(G)$ is known for all graphs of order ≤ 5 but some exceptions.

Information ratio of trees

Theorem (Csirmaz and Tardas, 2007)

Let T be a tree. Then $R(T) = 2 - \frac{1}{c(T)}$

where $c(T)$ is the maximum size of a core of T.

Theorem (Csirmaz and Ligeti, 2009)

Let d be the maximum degree of G, then $R(G) = 2 - 1/d$

provided that G satisfies the following conditions:

(i) every vertex has at most one neighbor of degree one,

(ii) vertices of degree at least three are not adjacent, and

(iii) the girth of G is at least six.

Average information ratio

Theorem (with Lu, 2012)

If T be a tree of order n, then $AR(T) = \dfrac{n + in(T) - d^*(T)}{n}$

$in(T) = \#\{\, v \in V(T) \,|\, \deg_T(v) \geq 2 \,\}$

$d^*(T)$: the deduction of T.

Theorem (with Lu, 2013)

Let G be a graph in which every cycle has two degree-two vertices of distance at least four, then

$$AR(G) = \frac{n + in(G) - d^*(G)}{n}$$

General ideas

- For bipartite graphs of girth larger than 4, we use star covering. (Star is a complete multipartite graph.)
- Covering with minimum vertex-number sum will provide a good lower bound by referring to the theorem of Stinson.
- We manage to find an idea to find a good upper bound which attends the lower bound.

Core clusters

- A *core* of G is a connected subset $V_0 \subseteq IN(G)$ satisfying
  (1) each $v \in V_0$ has a neighbor $\bar{v} \notin V_0$ such that $\bar{v}$ is not a neighbor of any other vertex in V_0, and
  (2) $\{\bar{v} \mid v \in V_0\}$ is an independent set in G.

- A *core cluster* of G is a partition £ $= \{V_1, V_2, \ldots, V_k\}$ of $IN(G)$ such that each V_i is a core of G.
- The size k of £ is denoted as $c_{£}(G)$.
- $c^*(G) = \min\{c_{£} \mid £$ is a core cluster of $G\}$

IN(G) is the set of vertices of degree larger than 1.

Min-Max pair

- Theorem

  For every graph G, $c_\vartheta(G) \geq d_\Pi(G)$ holds for any core cluster ϑ and star (complete multipartite ?) covering Π of G. In particular, $c^*(G) \geq d^*(G)$.

- Fact
- If there exists a core cluster ϑ and a star covering Π such that $c_\vartheta(G) = d_\Pi(G)$, then we obtain AR(G).

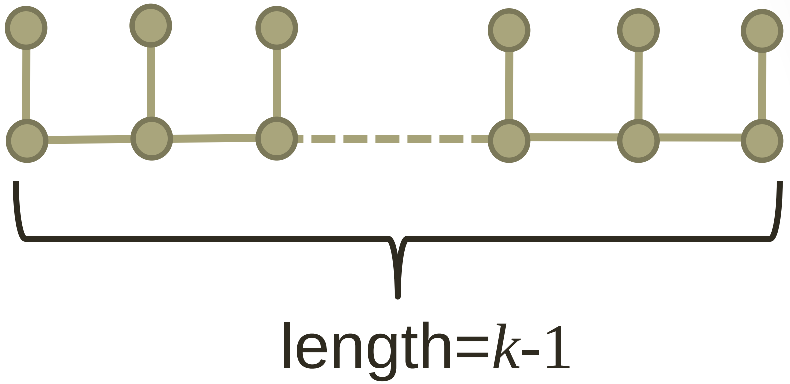
Extension to infinite graphs

Theorem (with Lu)

(1) $AR(P_\infty) = \frac{3}{2}$

(2) $AR(Comb_\infty) = \frac{3}{2}$

$Comb_k$:



$AR(G') \leq AR(Comb_k)$, for any subgraph G' of $Comb_k$.

$AR(Comb_k) = \frac{3k-1}{2k} \rightarrow \frac{3}{2}$, as $k \rightarrow \infty$

Exact ratios

Theorem

(1) $R(T_\infty) = 2$

(2) $AR(P_\infty) = \frac{3}{2}$

(3) $AR(Comb_\infty) = \frac{3}{2}$

(4) $AR(T_\infty) = \frac{3}{2}$

Application of graph packing

- Define G(r,c) as the *grid-block* with r rows and c columns where each grid point is a (distinct) vertex and two vertices are *collinear* if they are on the same row or column.
- If we define a graph from G(r,c) by letting two vertices be adjacent if and only if they are collinear, then G(r,c) is isomorphic to the *Cartesian product* of K_r and K_c denoted by $K_r \times K_c$.

An Example

- The green and pink grid-blocks pack K_9. Therefore, a *G(3,3)*-design or a 3x3 grid-block design of order 9 exists!

1	4	7
2	5	8
3	6	9

1	6	8
5	7	3
9	2	4

The Existence of G(r,c)-Designs

- If a G(r,c)-design of order n exists then the following conditions hold:

  (a) $rc \leq n$,

  (b) $r+c-2$ divides $n-1$, and

  (c) $rc(r+c-2)$ divides $n(n-1)$.

Lattice Rectangles

- A G(r,c)-design is called a *lattice square* provided that $r = c = n^{1/2}$ named by Yates, 1940.
- Construction of lattice squares for $n^{1/2}$ an odd prime power was given by Raghavarao in 1971.
- Lattice squares were extended to *lattice rectangles* ($r \neq c$ and $rc = n$) by Harshbarger in 1947.

Real World

- In most practical uses, the grid-block has size limitation and n is large.
- Thus, we have to consider the G(r,c)'s with $r < n^{1/2}$ and $c < n^{1/2}$, while preserving the *unique collinearity condition*, i.e. every pair of vertices occur at most once in the same row or column. This is one of the reasons we study *G(r,c)*-design or *G(r,c)*-packing.

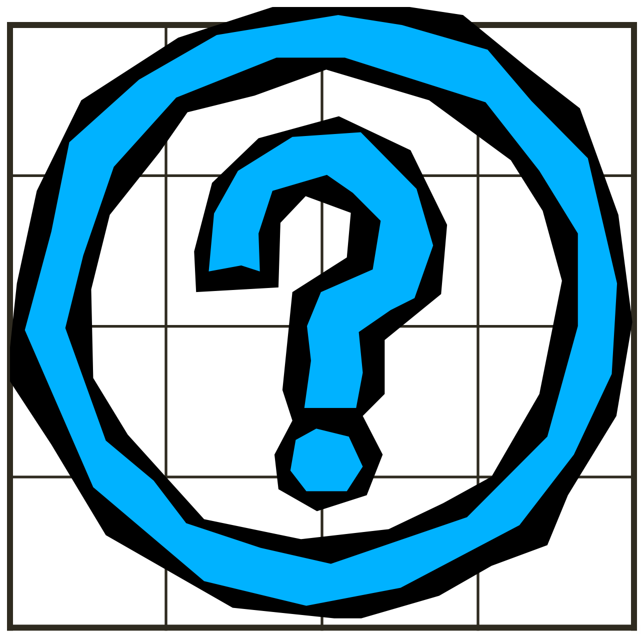
Resolvable Packing

- A *G(r,c)*-packing of order n is said to *resolvable* if the collection of grid-blocks can be partitioned into subclasses $\boldsymbol{R_1}$, $\boldsymbol{R_2}$, …, $\boldsymbol{R_t}$ such that every vertex of K_n is contained in precisely one grid-block of each class. Each $\boldsymbol{R_i}$ is called a *resolution class*.
- Clearly, such a packing exists only when rc divides n.

A G(4,4)-Packing of order 16

Resolvable

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



A Resolvable G(3,3)-Packing of order 18

0	1	2
3	4	5
6	7	8

9	10	11
12	13	14
15	16	17

0	4	8
5	6	9
7	10	14

1	3	15
12	16	11
17	2	13

0	9	13
10	17	8
12	7	3

1	5	11
6	15	2
16	14	4

Applications of packing

- For practial use, we focus on resolvable *G(r,c)*-packings. Therefore, we consider only the cases rc divides n.
- In a resolvable *G(r,c)*-packing the number of resolution classes t is at most $\lfloor (v-1)/(r+c-2) \rfloor$. If a resolvable *G(r,c)*-packing has this number of resolution classes, then it is ***optimal***.

Ready for Tests?

- In *DNA library screening*, we have a set of oligonucleotides (*clones*) and a *probe* X which is a short DNA sequence. Let X' denote the dual sequence of X obtained by first reversing the order of letters and then interchanging A with T and C with G. A clone is called *positive* if it contains X as a subsequence and *negative* if not.
- ***The goal is to identify all the positive clones.***

Group Testing

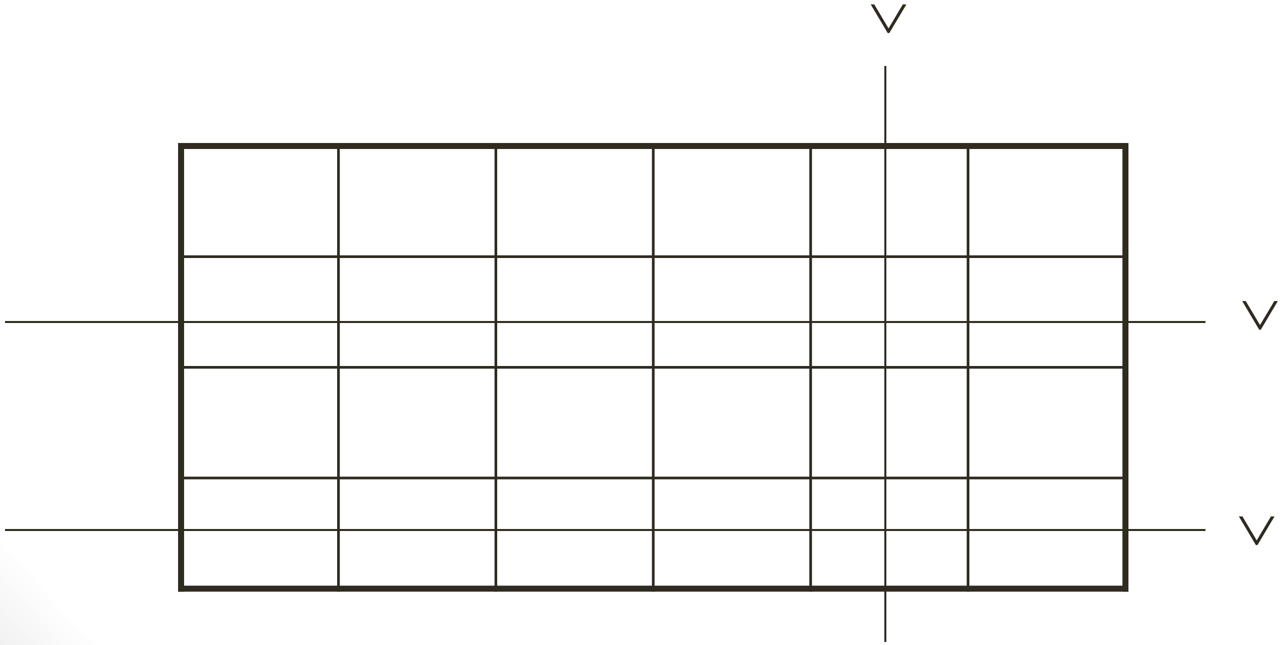
- Economy of time and costs requires that the clones be assayed in groups.
- Each group is called a *pool*.
- A pool gives a negative outcome, all clones contained in it are found to be negative. On the other hand, if a pool is positive, at the second stage we test each clone individually.
- ***Two-stage Test!***

Library Screening

- In such screening, a *microtiter* plate, which is an array with size 8 x 12 or 16 x 24, etc. is utilized and different clones are settled in each spot, called *well*, of the plate.
- *Every row and every column in a micro-titer plate is tested at the same time as a pool in the first stage. (r + c tests for a plate)*

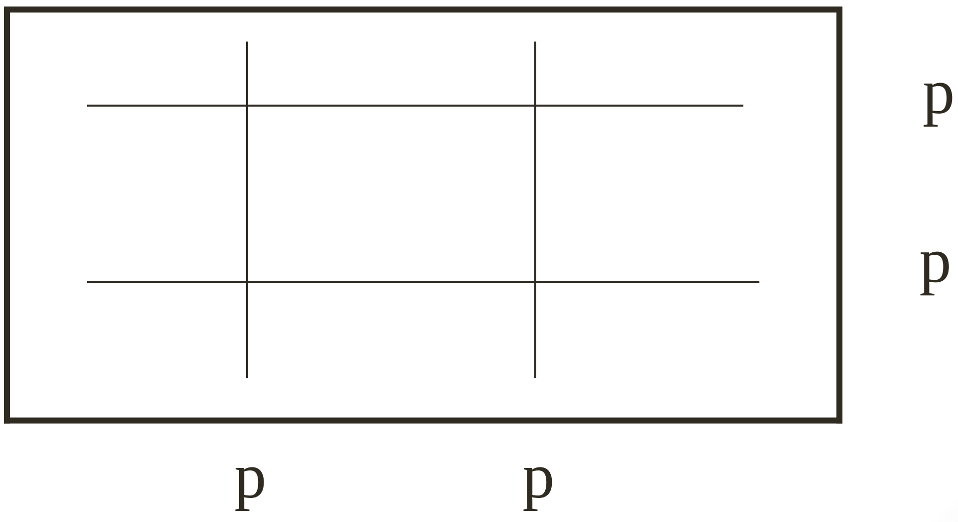
Basic Matrix Method

- If there is only one *row* (or *column*) of positive then we can determine the positive clones without the second stage test.



More Positive Clones

- **For example if two rows and two columns are positive as follows, then we *can not determine* whether the four clones settled at the crossing wells of positives are really positive or not.**



Unique Colinearity Condition

- Thus, if it is allowed to test more than twice for each clone, then it is desired that every two clones occur at most once in the same row or the same column, which is called the *unique collinearity condition (UCC)*.
- The efficiency of *UCC* was shown by Barillot et al (1991,*simulation*) and proved theoretically by Berger et al in 2000 at *Biometrics*.
- *So, corresponds to grid-block packing.*

Why Resolvable Packings?

- The *replication number* of a vertex v in a grid-block packing is the number of grid-blocks in which v is in there.
- It is a favorite property that the number of replications of each clone should be almost the same in the first stage.
- So, take a resolution class and test (group) each grid-block at the same time guarantees the above ***equal replication*** property.

Analysis of Replications

- A simulation result of a comparison between *constant replications* and *random replications* shows that we need less tests to find all the positives by using grid-block packings with constant replications.
- As an example, if we have 1,000 clones and 0.1 is the probability of positives, then it takes around 600 tests (Constant R.) and 750 tests (Random R.) to find all the positives respectively. (M-J-F, 2004)

Positives

- In general, we have an experiment (group testing) to run with the number of positives much less than the number of items, say less than 1%.
- With the idea of pooling design, we can do the experiment at the same time instead of one by one.

References

1. E. Barillot, B. Lacroix and D. Cohen, *Theoretical analysis of libraray screening using an N-dimensional pooling strategy*, Nucleic Acids Research, **19** (1991), 6241-6247.
2. T. Berger, J. W. Mandell and P. Subrahmanya, *Maximally efficient two-stage screening*, Biometrics, **56** (2000), 833-840.
3. J. E. Carter, *Designs on cubic multigraphs*, Ph.D. thesis in McMaster University, 1989.
4. H. L. Fu, F. K. Hwang, M. Jimbo, Y. Mutoh and C. L. Shiue, *Decomposing complete graphs into K_r x K_c's*, J. Statistical Planning and Inference, **119(2)** (2004), 225-236.

1. B. Harshbarger, *Rectangular lattices*, Va Agri. Exp. Stn. Memoir **1**, 1947.
2. Y. Mutoh, T. Morihara, M. Jimbo and H. L. Fu, *The existence of 2x4 grid-block designs and their applications*, SIAM. J. Discrete Math.,**16** (2003), 173-178.
3. Y. Mutoh, M. Jimbo and H. L. Fu, *A resolvable r x c grid-block packing and its application to DNA library screening*, Taiwanese J. Math., **Vol. 8, No. 4**, Dec. 2004, 713-737.
4. D. Raghavarao, *Constructions and combinatorial problems in design of experiments*, Wiley, New York, (1971).
5. F. Yates, *Lattice squares*, J. Agri. Sci., **30** (1940), 672-687.
6. R. Zhang et al., *The existence of rx4 grid-block designs with r = 3, 4*, SIAM Discrete Math., Vol. 23, Issue 2, 2009, 1045-1062.

Thank you for your attention.

謝謝!