Protein-Protein Interaction and Group Testing in Bipartite Graphs

Presented by Yingshu Li
5-17-2006

Outline

- Motivation
- Problem definition
- Construction methods
- Generalization

Motivation

Protein-Protein Interactions

- Interactions between bait proteins and prey proteins.
- Critical in many biological processes
  - Formation of macromolecular complexes
  - Transduction of signals in biological pathways

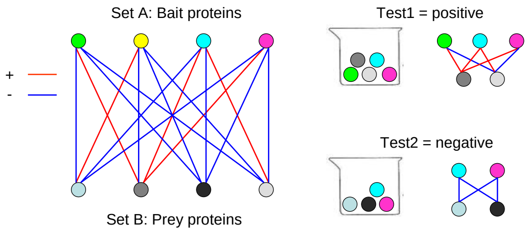
Motivation

Group Testing

- Dates back to World War II.
- Now being used in many applications.
- An efficient method to identify 'protein-protein interactions' between a finite number of bait proteins and prey proteins, through conducting tests on subsets of bait and prey proteins.

Motivation

Group testing for protein-protein interactions



Group testing in a complete bipartite graph $K_{A,B}$

Motivation

Model group testing as binary incidence matrices

test/item	1	2	3	4	5	6	7	8	Output
test1	1	0	1	1	0	1	0	1	**1**
test2	0	0	1	1	1	0	1	0	**0**
test3	1	0	0	0	1	0	0	0	**0**
test4	0	1	0	0	0	1	0	0	**1**

Motivation

d-disjunct matrix

An $t \times n$ binary matrix is *d-disjunct* ($d < t$) if for any $d+1$ columns $C_0, C_1, \ldots, C_d$, there exists a row such that C_0 has an *1*-entry and all $C_1, \ldots, C_d$ have *0*-entries.

A 2-disjunct matrix

$$\begin{pmatrix} 0 & 0 & 1 & 0 & 0 \\ 1 & 0 & 0 & 1 & 0 \\ 0 & 1 & 0 & 0 & 0 \\ 1 & 0 & 1 & 0 & 1 \\ 1 & 1 & 0 & 0 & 0 \end{pmatrix}$$

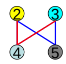
Motivation

d-disjunct matrix for bipartite graph

The binary incidence matrix *M* for a bipartite graph *H* is *d(H)-disjunct* if for any $d+1$ edges $e_0, e_1, \ldots, e_d$ of *H*, there exists a row in *M* indicating that a test contains e_0, but not $e_1, \ldots, e_d$.

A 3(H)-disjunct matrix

2	3	4	5	Output
0	1	1	0	**1**
1	0	0	1	**0**
1	1	1	0	**0**
1	0	1	0	**1**
0	1	0	1	**0**



Problem Definition

How to construct a *d(G)-disjunct* matrix for a bipartite graph *G*?

- Input: a bipartite graph *G=(A, B, E).*
- Output: a *d(G)-disjunct* matrix (A test regimen).

The First Construction

- *G=(A, B, E)* is a bipartite graph.
- M_A is a *d-disjunct* t_A X |*A*| matrix with columns labeled by the vertices in *A.*
- M_B is a *d-disjunct* t_B X |*B*| matrix with columns labeled by the vertices in *B.*

M_A =

	①	②	③
1_A	1	0	0
2_A	0	1	0
3_A	0	0	1

M_B =

	④	⑤	⑥
1_B	0	1	0
2_B	1	0	0
3_B	0	0	1

The First Construction

M_A =

$\langle i_A \rangle$	①	②	③
1_A	1	0	0
2_A	0	1	0
3_A	0	0	1

M_B =

$\langle i_B \rangle$	④	⑤	⑥
1_B	0	1	0
2_B	1	0	0
3_B	0	0	1

M

$\langle i_A, i_B \rangle$	①	②	③	④	⑤	⑥
$\langle 1_A, 1_B \rangle$	1	0	0	0	1	0
$\langle 1_A, 2_B \rangle$	1	0	0	1	0	0
$\langle 1_A, 3_B \rangle$	1	0	0	0	0	1
$\langle 2_A, 1_B \rangle$	0	1	0	0	1	0
$\langle 2_A, 2_B \rangle$	0	1	0	1	0	0
$\langle 2_A, 3_B \rangle$	0	1	0	0	0	1
$\langle 3_A, 1_B \rangle$	0	0	1	0	1	0
$\langle 3_A, 2_B \rangle$	0	0	1	1	0	0
$\langle 3_A, 3_B \rangle$	0	0	1	0	0	1

Cell ($\langle i_A, i_B \rangle$, *u*) = 1 if and only if

- $u \in A$, cell (i_A, *u*) in M_A= 1 or
- $u \in B$, cell (i_B, *u*) in M_B= 1

The First Construction

***M* is a *d(G)-disjunct* matrix.**

i=1_A i'=1_B

M =

	①	②	③	④	⑤	⑥
$\langle 1_A, 1_B \rangle$	1	0	0	0	1	0
$\langle 1_A, 2_B \rangle$	1	0	0	1	0	0
$\langle 1_A, 3_B \rangle$	1	0	0	0	0	1
$\langle 2_A, 1_B \rangle$	0	1	0	0	1	0
$\langle 2_A, 2_B \rangle$	0	1	0	1	0	0
$\langle 2_A, 3_B \rangle$	0	1	0	0	0	1
$\langle 3_A, 1_B \rangle$	0	0	1	0	1	0
$\langle 3_A, 2_B \rangle$	0	0	1	1	0	0
$\langle 3_A, 3_B \rangle$	0	0	1	0	0	1

M_A and M_B are *2-disjunct*.

$e_0 = (x_0, y_0) = $(①, ⑤)
$e_1 = (x_1, y_1) = $(②, ④)
$e_2 = (x_2, y_2) = $(③, ⑥)

In M_A, $(i, x_0)=1$, $(i, x_j)=0$, $x_j=\{x_1, \ldots, x_d\} \setminus \{x_0\}$

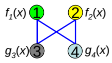
In M_B, $(i', y_0)=1$, $(i', y_0)=0$, $y_j=\{y_1, \ldots, y_d\} \setminus \{y_1\}$

For edges $e_0, \ldots, e_d$, is there a row indicating a test contains e_0, but not $e_1, \ldots, e_d$?

Row $\langle i, i' \rangle$ in *M* contains e_0, not $e_1 \ldots e_d$!

The Second Construction

- $G=(A, B, E)$ is a bipartite graph.
- $GF(q)$ be a finite field of order q.
- Associate each edge $e = (u, v)$ of G a pair of polynomials (f_u, g_v). f_u and g_v are of degree k-1 over $GF(q)$.



Cell $(x, e) = (f_u(x), g_v(x))$

$M'_G(q, k, t) =$

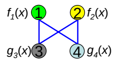
	(①,③)	(①,④)	(②,③)	(②,④)
1	$(f_1(1), g_3(1))$	$(f_1(1), g_4(1))$	$(f_2(1), g_3(1))$	$(f_2(1), g_4(1))$
2	$(f_1(2), g_3(2))$	$(f_1(2), g_4(2))$	$(f_2(2), g_3(2))$	$(f_2(2), g_4(2))$
3	$(f_1(3), g_3(3))$	$(f_1(3), g_4(3))$	$(f_2(3), g_3(3))$	$(f_2(3), g_4(3))$

If $t \geqslant k$, then for any 2 columns C_0, C_1 , there exists a row at which the entry of C_0 does not equal the entries of C_1 .

13

The Second Construction

$G(A, B, E)$



$M'_G(q,k,t)$

$<x>$	(①,③)	(①,④)	(②,③)	(②,④)
1	$(f_1(1), g_3(1))$	$(f_1(1), g_4(1))$	$(f_2(1), g_3(1))$	$(f_2(1), g_4(1))$
2	$(f_1(2), g_3(2))$	$(f_1(2), g_4(2))$	$(f_2(2), g_3(2))$	$(f_2(2), g_4(2))$
3	$(f_1(3), g_3(3))$	$(f_1(3), g_4(3))$	$(f_2(3), g_3(3))$	$(f_2(3), g_4(3))$

In M, cell $(<x, (y, z)>, u) = 1$ if and only if

- $u \in A$, $f_u(x) = y$ or
- $u \in B$, $g_v(x) = z$

$M_G(q,k,t)$

$<x, (y, z)>$	①	②	③	④
$<1, (f_1(1), g_3(1))>$	1	0	1	0
$<1, (f_1(1), g_4(1))>$	1	0	0	1
$<1, (f_2(1), g_3(1))>$	0	1	1	0
$<1, (f_2(1), g_4(1))>$	0	1	0	1
$<2, (f_1(2), g_3(2))>$	1	0	1	0
$<2, (f_1(2), g_4(2))>$	1	0	0	1
$<2, (f_2(2), g_3(2))>$	0	1	1	0
$<2, (f_2(2), g_4(2))>$	0	1	0	1
$<3, (f_1(3), g_3(3))>$	1	0	1	0
$<3, (f_1(3), g_4(3))>$	1	0	0	1
$<3, (f_2(3), g_3(3))>$	0	1	1	0
$<3, (f_2(3), g_4(3))>$	0	1	0	1

14

The Second Construction

Cell $(x, e) = (f_u(x), g_v(x))$

$M'_G(q, k, t) =$

	(①,③)	(①,④)	(②,③)	(②,④)
1	$(f_1(1), g_3(1))$	$(f_1(1), g_4(1))$	$(f_2(1), g_3(1))$	$(f_2(1), g_4(1))$
2	$(f_1(2), g_3(2))$	$(f_1(2), g_4(2))$	$(f_2(2), g_3(2))$	$(f_2(2), g_4(2))$
3	$(f_1(3), g_3(3))$	$(f_1(3), g_4(3))$	$(f_2(3), g_3(3))$	$(f_2(3), g_4(3))$

If $t \geqslant d(k-1)+1$, then for any $d + 1$ columns $C_0 \ldots C_d$, there exists a row at which the entry of C_0 does not equal the entries of $C_1 \ldots C_d$.

➔ $f_1(1) \neq f_2(1)$ & $g_3(1) \neq g_4(1)$

15

The Second Construction

$x=1, (y, z)= (f_1(1), g_3(1))$

$f_1(1) \neq f_2(1)$
$g_3(1) \neq g_4(1)$

$f_u(x) = y?$
$g_v(x) = z?$

$M_G(q,k,t)=$

	①	②	③	④
$<1, (f_1(1), g_3(1))>$	1	0	1	0
$<1, (f_1(1), g_4(1))>$	1	0	0	1
$<1, (f_2(1), g_3(1))>$	0	1	1	0
$<1, (f_2(1), g_4(1))>$	0	1	0	1
$<2, (f_1(2), g_3(2))>$	1	0	1	0
$<2, (f_1(2), g_4(2))>$	1	0	0	1
$<2, (f_2(2), g_3(2))>$	0	1	1	0
$<2, (f_2(2), g_4(2))>$	0	1	0	1
$<3, (f_1(3), g_3(3))>$	1	0	1	0
$<3, (f_1(3), g_4(3))>$	1	0	0	1
$<3, (f_2(3), g_3(3))>$	0	1	1	0
$<3, (f_2(3), g_4(3))>$	0	1	0	1

If $t \geqslant d(k - 1) + 1$, then $M_G(q, k, t)$ is *d(G)-disjunct*.

$M'_G(q, k, t)$ has a row x such that the entry (y, z) in cell (x, e_0) does not equal the entry in cell (x, e_j) for all $j = 1 \ldots d$.

Row $<x, (y, z)>$ in M contains e_0, not $e_1 \ldots e_d$!

For edges $e_0, \ldots, e_d$, is there a row indicating a test contains e_0, but not $e_1, \ldots, e_d$?

16

Generalization

- $G=(V, E)$ is a hyper-graph and c-colorable.
- $GF(q)$ is a finite field of order q.
- Associate $u \in V$ a polynomial p_u of degree $k-1$ over $GF(q)$ such that for u and v with the same color, p_u and p_v are distinct.

Generalization

Step 1:

Construct a t X $|E|$ matrix $M'_G(q, k, t)$ with the rows labeled by t elements in $GF(q)$ and the columns labeled by all the edges of G such that each cell (x, e) contains a set $\{(p_u(x),i) \mid u \in V \text{ with color } i\}$.

- Property of M': For any $d+1$ columns $C_0, \ldots, C_d$ in $A_G(q, k, t)$, there exists a row at which the entry of C_0 does not contain the entry of C_j for $j=1, \ldots, d$.

Generalization

Step 2:

Construct a matrix $M_G(q, k, t)$ from $M'_G(q, k, t)$. $M_G(q, k, t)$ has $|V|$ columns labeled with all the vertices in G. For each row x of $M'_G(q, k, t)$ and each entry Q at row x, construct a row with label $\langle x, Q \rangle$ for $M_G(q, k, t)$ such that the cell $(\langle x, Q \rangle, u)$ contains a 1-entry if and only if u is in color i and $p_u(x)=y$ for $(y, i) \in Q$.

- If $t \geqslant d(k - 1) + 1$, then $B_G(q, k, t)$ is $d(G)$-disjunct.

Thank You !