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

Set A: Bait proteins
Set B: Prey proteins

Test1 = positive
Test2 = negative

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

Motivation

Model group testing as binary incidence matrices

<table>
<thead>
<tr>
<th>test/item</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
<th>Output</th>
</tr>
</thead>
<tbody>
<tr>
<td>test1</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>test2</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>test3</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>test4</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>1</td>
</tr>
</tbody>
</table>

$d$-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

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

A 3(H)-disjunct matrix

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 $e_1, \ldots, e_d$.

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

Output
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 $t_A \times |A|$ matrix with columns labeled by the vertices in $A$.
- $M_B$ is a $d$-disjunct $t_B \times |B|$ matrix with columns labeled by the vertices in $B$.

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

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

Row $<i, i'>$ in $M$ contains $e_i$, but not $e_{i+1}, ..., e_d$.
If \( t \geq d(k-1)+1 \), 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 \( C_1 \ldots C_d \).

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 \( GF(q) \).

\[
\begin{align*}
M'_{G}(q, k, t) &= \begin{cases} 1 & \text{if } f_1(x) \neq f_2(x) \text{ or } g_3(x) \neq g_4(x) \text{ for } x \in [1, q) \setminus \{0\} \\
0 & \text{otherwise}
\end{cases} \\
&= \begin{pmatrix}
1 & 0 & 0 & 1 \\
0 & 1 & 0 & 0 \\
0 & 0 & 1 & 0 \\
1 & 0 & 0 & 1
\end{pmatrix}
\end{align*}
\]

Row \( x \) in \( M'_{G}(q, k, t) \) contains \( e_x \) if and only if \( f_1(x) \neq f_2(x) \) or \( g_3(x) \neq g_4(x) \).
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.

Step 1:
Construct a $t \times |E|$ matrix $M'(q, k, t)$ 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 $M'(q, k, t)$, there exists a row at which the entry of $C_0$ does not contain the entry of $C_j$ for $j=1, \ldots, d$.

Step 2:
Construct a matrix $M(q, k, t)$ from $M'(q, k, t)$. $M(q, k, t)$ has $|V|$ columns labeled with all the vertices in $G$. For each row $x$ of $M'(q, k, t)$ and each entry $Q$ at row $x$, construct a row with label $<x, Q>$ for $M(q, k, t)$ such that the cell $(<x, Q>, u)$ contains a 1-entry if and only if $u$ is in color $i$ and $p_u(x) = y$ for $(y, i) = Q$.

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