Proofread (5-8-2005)

CMSC 27400-1/37200-1 Combinatorics and Probability **Spring 2005**

Lecture 16: May 4, 2005

Instructor: László Babai *Scribe: Hariharan Narayanan*

NOTE: Change in Monday's TA schedule; no change Tuesday and Thursday.
TA SCHEDULE: TA sessions are held in Ryerson-255, Monday 7:30-8:30,
Tuesday and Thursday 5:30–6:30pm.
INSTRUCTOR'S EMAIL: laci@cs.uchicago.edu
TA's EMAIL: hari@cs.uchicago.edu, raghav@cs.uchicago.edu

Homogenous submatrices of $(0, 1)$-matrices

Question 16.1 *How large a homogenous $t \times t$ submatrix does every $(0, 1)$-matrix contain?*

We shall show that

1. every $(0, 1)$-matrix contains a $t \times t$ submatrix where $t \sim \frac{\log_2 n}{2}$.
2. For $t \sim 2 \log_2 n$, we show via the probabilistic method, that this is not true.

We first show (2.).
Take a random $(0, 1)$-matrix. The probability that a particular $t \times t$ submatrix is homogenous is $2^{-t^2} \cdot 2$.

$$P(\exists t \times t \text{ homogenous submatrix}) < \binom{n}{t}^2 2^{1-t^2}. \text{(union bound)}$$

We use the estimate $\binom{n}{t} \leq \frac{n^t}{t!}$.

$$\begin{aligned} \binom{n}{t}^2 2^{1-t^2} &\leq \frac{n^{2t} 2^{1-t^2}}{(t!)^2} \\ &= \frac{2}{(t!)^2} (n^2 2^{-t})^t \end{aligned}$$

Thus, for the above to be < 1, it is sufficient that $n^2 2^{-t} \leq 1$, i. e. $2 \log_2 n \leq t$, which is true in (2.).
We now prove (1.).
We need to show that in every $n \times n$ $(0, 1)$-matrix, there exists a homogenous $t \times t$ submatrix

$t \sim \frac{\log_2 n}{2}$. Consider the following procedure:
Scan the first row. Let $x \in \{0,1\}$ be the majority entry. Delete all those columns of the matrix whose entry in the first row is not x. Now repeat this procedure for the $2^{nd}, 3^{rd}, \ldots, k^{th}$ rows.
$(\forall k)$ (we obtain a $k \times \rceil \frac{n}{2^k} \lceil$ matrix, with each row homogenous.)
Take all those rows that have the majority entry. This gives us a $\lceil k/2 \rceil \times \lceil n/2^k \rceil$ (or larger) submatrix. The optimal k is when $k/2 \approx n/2^k$. We approximate k by the real number x which is the exact solution of the equation $x/2 = n/2^x$. Let x be the (not necessarily integral) value such that $x/2 \sim n/2^k$.
i. e. $x 2^{x-1} \sim n$.
i. e. $\log_2 x + x - 1 \sim \log_2 n$.
i. e. $x \sim \log n$.
Therefore $\frac{n}{\log_2 n} = 2^{x-1}(1 + o(1))$, i. e. $x = \log_2 n - \log_2 \log_2 n + 1 + o(1) \sim \log_2 n$.
Let now $k = \lfloor x \rfloor$ Then $k/2 \leq n/2^k$ and $k/2 \sim \frac{\log_2 n}{2}$ □

Accountant's Principle:
The sum of the row-sums of a matrix is equal to the sum of its column-sums.

Theorem 16.2 *Every $n \times n$ $(0,1)$-matrix contains a $t \times t$ homogenous submatrix for $t \sim \log n$.*

Proof:
Without loss of generality, we may assume that we have $n/2$ rows, each of which has $\geq n/2$ $1'$s (*).
Let B be the submatrix of these rows. Consider the $n/2 \times \binom{n}{t}$ matrix C, whose columns are indexed by t-subsets T of columns of B. The entries of C are defined by:
$C_{iT} = 1$ if and only if T is in the support of row i; i. e. $j \in T \implies b_{ij} = 1$. From the accountant's principle and (*),

$$S := \sum_{i=1}^{n/2} \left(\sum_{j=1}^{\binom{n}{t}} C_{ij} \right) \geq n/2 \binom{n/2}{t}.$$

Let m be the maximum column sum.

$$S = \left(\sum_{j=1}^{\binom{n}{t}} \right) \left(\sum_{i=1}^{n/2} C_{ij} \right) \leq \binom{n}{t} m.$$

i. e. $m\binom{n}{t} \geq (n/2)\binom{n/2}{t}$.
i. e. $m \geq \frac{n/2\binom{n/2}{t}}{\binom{n}{t}}$.
Note that we have a $t \times m$ homogenous submatrix.
$\frac{(n-t)^t}{t!} < \binom{n}{t} \leq \frac{n^t}{t!}$. Dividing the upper and lower bounds, we get

$$\left(\frac{n-t}{n}\right)^t = (1 - t/n)^t > 1 - t^2/n.$$

Exercise 16.3 $(1-x)^t > 1 - tx$.

Exercise 16.4 *If* $t = t(n) = o(\sqrt{n})$*, then* $\binom{n}{t} \sim \frac{n^t}{t!}$.

If $t = (1-\epsilon)\log_2 n$,

$$(n/2)2^{-(1-\epsilon)\log_2 n} = \frac{n^\epsilon}{2} > t.$$

$\square$

Exercise 16.5 *If* $X_1, \ldots, X_m$ *are independent non-constant random variables over a sample space of size* n*, then* $2^m \leq n$.

Exercise 16.6 *If* $X_1, \ldots, X_m$ *are pairwise independent non-constant random variables over a sample space of size* n*, then* $m \leq n-1$*. Prove that this bound is tight for* $n = 2^k$.