CS683 – Lecture 4

Lecture by John Hopcroft
Notes taken by Jean-Baptiste Jeannin and Mikhaïl Lisovich

Monday January 28th, 2008

1 One more threshold on $G(n, p)$

1.1 Presentation of the problem

In this lecture we will study one more threshold having to do with $G(n, p)$ and the emergence of cycles when p grows. The cycles emerge when p is asymptotically equivanlent to $\frac{1}{n}$. For $p = \frac{C}{n}$:

- if $C < 1$, there is a number of cycles independent of n
- if $C > 1$, the number of cycles will grow with n

1.2 Proof

Let us show it. Let x be the number of cycles, and let us calculate the expectation of x, $E(x)$. Let x_k be the number of cycles of size k. The expectation $E(x_k)$, corresponding to the number of cycles of size k is then given by:

$$E(x_k) = \left(\begin{array}{c} n \\ k \end{array} \right) \frac{(k-1)!}{2} p^k$$

Let us explain how to obtain this: for each cycle, there are $\left(\begin{array}{c} n \\ k \end{array} \right)$ ways of choosing its k vertices; then pick an element (it does not matter which one), there are $k-1$ choices for the second element in the cycle,$k-2$ for the third element,...,1 choice for the last element; but we are counting twice each cycle: once when we select the vertices beginning on the right-hand side, and once on the left-hand side. Therefore there are $\frac{(k-1)!}{2}$ ways of choosing this cycle. Finally, for each of these cycles, the probability of it appearing is p^k.

Therefore the total number of cycles $E(x)$ is given by:

$$E(x) = \sum_{k=3}^{n} \left(\begin{array}{c} n \\ k \end{array} \right) \frac{(k-1)!}{2} p^k$$

1.2.1 If $C < 1$

If $C < 1$, then:

$$\begin{aligned}
E(x) &= \sum_{k=3}^{n} \binom{n}{k} \frac{(k-1)!}{2} p^k \\
&\leq \sum_{k=3}^{n} \frac{n!}{k(k-1)!(n-k)!} \frac{(k-1)!}{2} p^k \\
&\leq \sum_{k=3}^{n} \frac{n^k}{2k} p^k \\
&\leq \sum_{k=3}^{n} (np)^k \\
&\leq \sum_{k=3}^{n} C^k \\
&\leq \sum_{k=3}^{+\infty} C^k
\end{aligned}$$

which is a convergent geometric series since $C < 1$. Therefore if $C < 1$, the number of cycles $E(x)$ is finite and does not depend on n.

1.2.2 If $C > 1$

$$\begin{aligned}
E(x) &= \sum_{k=3}^{n} \binom{n}{k} \frac{(k-1)!}{2} p^k \\
&= \sum_{k=3}^{n} \frac{n(n-1)\cdots(n-k+1)}{2k} \left(\frac{C}{n}\right)^k
\end{aligned}$$

Therefore, if $C > 1$ and when n grows to $+\infty$, $E(x)$ grows to $+\infty$. Therefore the number of cycles will grow with n.

2 Another Combinatorial Structure - $N(n, p)$

2.1 Problem Definition

$N(n, p)$ is a subset of $\{1, 2, .., n\}$ obtained by choosing each integer with probability p. Every monotone property on this structure has a threshold. By monotone property we mean a property whose probability which grows with p (an example of a non-monotone property is Prob(even-sized set), whose probability does not converge as $n \to \infty$). We will show this for a specific property –

the number of arithmetic progressions within $N(n,p)$.

Definition: a particular arithmetic progression is defined by parameters a and b. A set contains a geometric progression if it contains numbers a, $a+b$, $a+2b$, and so on.

2.2 Proof

Let x be the number of algorithmic progressions of size k. We claim that $E(x) = n^2 p^k$. This can be shown by bounding the total number of arithmetic progressions of size k. The total number of progressions within a subset cannot exceed n^2, since there are only n^2 pairs (a,b) that can be chosen. To lower-bound the expected number, divide the interval $\{1,2,..,n\}$ into k consecutive subintervals of size n/k ($n \bmod k$ values will be left out). In each subinterval, there are n/k choices of a and n/k choices of b.

If $p = n^{-2/k}$, $E(x) = n^2 \left(n^{-2/k}\right)^k = 1$. If $p < n^{-2/k}$, $E(x)$ is asymptotically zero. If $p > n^{-2/k}$, $E(x) \to \infty$.

We have shown that the expectation can converge to infinity. However, we are interested in showing that every subset picked will have an increasing number of arithmetic progressions. To do this, we will need to employ a second moment argument (see previous lectures for a refresher).

2.2.1 Claim: $\lim \frac{E(x^2)}{E^2(x)} = 1$

2.2.2 Proof Sketch:

Let I_x be an indicator variable for the i^{th} arithmetic expression. $x = I_1 + I_2 + ... + I_{n^2}$, and we have shown that $E(x) = n^2 p^k$.

$\frac{E(x^2)}{E^2(x)} = \frac{\sum_{i,j}(I_i I_j)}{\sum_{i,j} E(I_i)E(I_j)}$

Note that the indicator variables are not independent because two progressions may share one or more common elements. Because of this, we can not simply write $E(I_i I_j) = E(I_i)E(I_j)$.

$$\lim \frac{E(x^2)}{E^2(x)} = \frac{p^k + n^4 p^{2k} + n^3 p^{2k-1} + n^2 p^{2k-2} + ...}{n^2 p^{2k}}$$

where the p^k term refers to identical sequences, $n^4 p^{2k}$ refers to non-overlapping sequences, the third term refers to sequences with one overlapping number, and so on. For $p > n^{-2/k}$, $p^k > n^{-2}$, and the expression can be shown to converge to 1. The full proof is left as an excercise in Homework 13.

3 Introduction to the Sudoku problems

A Sudoku game is something like that:

	1	2					4	6
		3	6	9		2		
6	4			5	2		7	
5			2	4			8	3
	7	4	8	6			9	
8								2
				8	6			4
		5		1		6		
3		9	3		7			1

The aim of the game is to fill in the table, so that to have each number from 1 to 9 in each row, in each column, and in each little square.

We can be interested in a few problems:

- For each distribution of the numbers there will be a certain number of solutions. If we put the numbers randomly in the table, there will be a threshold from which there will only be one solution left.
- How do you decide if tables are easy or difficult.
- We can see the constraints on the table as a conjonctive normal form that has to be satisfied. The probability of satisfying a CNF depends on the number of clauses as in the following figure: