CS 683: Advanced Design & Analysis of Algorithms	February 15, 2008

Lecture 12

Lecturer: John Hopcroft — *Scribes: Chris Provan, Anand Bhaskar*

1 Branching Processes

1.1 Review from Lecture 9

Recall that we have been investigating properties of random trees where in each generation the number of children for each leaf is an independent random variable where $p_i \in [0,1]$ is the probability of a leaf having i children for each $i \in \mathbb{Z}_0^+$. Thus the generating function for the number of children in the first generation is $f(x) = \sum_{i=1}^{\infty} p_i x^i$. Defining $f_1 = f$ and $f_j = f \circ f_{j-1}$, the generating function for the j-th generation is given by f_j.

Consider the stationary points of f. We assume that each node has a finite number of children with probability 1, so $\sum_{i=1}^{\infty} p_i = 1$ and thus $f(1) = 1$.

- If $f'(1) < 1$ then the convexity of $f(x)$ on $x \in [0,1]$ implies that $x = 1$ is the unique stationary point of $f(x)$, and the branching process dies out almost surely.
- If $f'(1) = 1$ then we have two subcases
  - $p_1 = 1$, in which case $f(x) = x$ for all x and the tree is just a chain which grows forever and the probability of extinction is 0
  - $p_1 < 1$, in which case the strict convexity of f implies that $x = 1$ is again the unique stationary point and the process dies out almost surely.
- If $f'(1) > 1$ then there is some $q \in [0,1)$ for which $f(q) = q$. This q, the smallest fixed point of f, is the probability of extinction for the branching process.

1.2 Expected Size of Extinct Family

Given that a branching process has become extinct, the size of the tree is clearly finite. This does not directly imply that the expected size of a terminated branching process is finite, but we will see in the following analysis that this is the case.

Claim: If $m := f'(1) \neq 1$ then the expected size of the family given that the process terminates is finite.

Proof. Let Z_j be the size of the jth generation and let $Z = \sum_{j=1}^{\infty} Z_j$ be the size of the family. Let A be the event that the branching process terminates. We are therefore interested in $E[Z|A] = \sum_{j=0}^{\infty} E[Z_j|A]$, where the equality holds by the non-negativity of the Z_j. Note that $Z_0 = 1$ almost surely.

Using Bayes' Theorem,

$$P[Z_1 = k, A] = P[Z_1 = k|A] \cdot P[A] = P[A|Z_1 = k] \cdot P[Z_1 = k].$$

Define q to be the minimum stationary point of $f(x)$ in $[0, 1]$. q is well-defined since $f(1) = 1$. Then as noted previously, the probability of extinction is $P[A] = q$. Furthermore, the conditional probability of extinction given the size of the first generation $P[A|Z_1 = k] = q^k$ since each of the k trees rooted at the nodes in the first generation terminates independently with probability q. Thus

$$\begin{aligned} P[Z_1 = k|A] &= \frac{P[A|Z_1 = k] \cdot P[Z_1 = k]}{P[A]} \\ &= \frac{q^k \cdot p_k}{q} \\ &= q^{k-1} \cdot p_k, \end{aligned}$$

and

$$\begin{aligned} E[Z_1|A] &= \sum_{k=0}^{\infty} k \cdot P[Z_1 = k|A] \\ &= \sum_{k=0}^{\infty} k \cdot q^{k-1} \cdot p_k \\ &= f'(q). \end{aligned}$$

Now consider the jth generation. Let Z_{ij} be the number of children of the ith node of generation $j-1$ for $i \in \{1, 2, \ldots, Z_{j-1}\}$. Each of the Z_{ij} are independent and distributed as Z_1, they are independent from Z_{j-1}, and $Z_j = \sum_{i=1}^{Z_{j-1}} Z_{ij}$. Therefore

$$\begin{aligned} E[Z_j|A] &= E\left[\sum_{i=1}^{Z_{j-1}} Z_{ij} \middle| A\right] \\ &= E[Z_{j-1}|A] \cdot E[Z_{1,j}|A] && \text{(this is Wald's identity)} \\ &= E[Z_{j-1}|A] \cdot E[Z_1|A] \\ &= E[Z_{j-1}|A] \cdot f'(q). \end{aligned}$$

Then a simple inductive argument shows $E[Z_j|A] = [f'(q)]^j$.

If $m = f'(1) < 1$ then $q = 1$ as noted earlier and $f'(q) < 1$. If $m > 1$ then $q < 1$. Since $f(1) = 1$ and f is continuously differentiable and lies below the line $y = x$ on the interval $(q, 1)$, we must have $f'(q) < 1$. Therefore $f'(q) < 1$ for all values of $m \neq 1$. This implies

$$\begin{aligned} E[Z|A] &= \sum_{j=0}^{\infty} E[Z_j|A] \\ &= \sum_{j=0}^{\infty} [f'(q)]^j \\ &= \frac{1}{1 - f'(q)} < \infty. \end{aligned}$$

Thus we have proved that if $m = f'(1) \neq 1$, then the expected size of the extinct tree is finite.

□

2 Growth Models

We now return to random graphs. $G(n, p)$ often does not model real world graphs, which can be seen by a comparison of degree distributions. Instead, we now consider the following model:

- Begin with a single vertex.
- At each iteration, add a new vertex.
- Add an edge with probability δ.
- If an edge is to be added, choose the two endpoints one after the other according to some probability distribution over the vertices.

For now, we will choose endpoints uniformly at random.

2.1 Degree Distribution

Let $d_k(t)$ be the expected number of vertices of degree k at time t. Then $d_0(1) = 1$ represents the starting state of the graph. Also, at any time t, there are t vertices in the graph since a vertex is added in each iteration.

Suppose there are $d_0(t-1)$ vertices of degree zero at the beginning of iteration t. First a new vertex of degree zero is added, bringing the total number of vertices up to t. If an edge is then added, the probability of selecting a vertex that is of degree zero is $\frac{d_0(t-1)+1}{t}$

for each endpoint since there are now a total of $d_0(t-1)+1$ degree-zero vertices. Then the approximate recurrence for the expected number of edges of degree zero at time t is

$$\begin{aligned} d_0(t) &= d_0(t-1) + 1 - 2\delta \cdot \frac{d_0(t-1)+1}{t} \\ &= d_0(t-1) + 1 - 2\delta \cdot \frac{d_0(t-1)}{t} - \frac{2\delta}{t} \end{aligned}$$

Suppose $d_0(t) \to p_0 t$ for some nonnegative constant p_0. Then we have the asymptotic dynamic

$$\begin{aligned} p_0 t &= p_0(t-1) + 1 - 2\delta \cdot \frac{p_0 t}{t} \\ &= p_0 t - p_0 + 1 - 2\delta p_0 \end{aligned}$$

which implies $p_0 = \frac{1}{1+2\delta}$.