College of Computer & Information Science Spring 2012
Northeastern University 24 February 2012
CS 7880: Network Algorithms and Analysis Scribe: Ravishankar Rajagopal

The following notes are heavily based on the treatment of branching processes in the book by Easley and Kleinberg [EK10] (this treatment, in turn, is fairly standard and is similar to that in many books on branching processes and random graph dynamics).

1 Branching Processes

Branching processes can be used to model the spread of diseases, innovations, computer viruses, gossip etc from one agent to another. A branching process can be modeled by an infinite tree characterized by two parameters, the branching factor k and a probability value p. The branching factor determines the number of child nodes that each node is connected to, starting from the root and p is the probability that the disease is transmitted across a given edge, independent of the other edges in the graph.

The most important question that we ask here is the following: Given values of k and p, what is the probability that the process would eventually terminate?

For example if $k = 1$ and $p = 1$, then it is easy to see that the process continues indefinitely with each person infecting one new person.

Theorem:

In general if we let R_0 - called the basic reproductive number of the disease - be the expected number of new cases of the disease caused by a single individual and q^* is the probability that the process goes on indefinitely then

1. If $R_0 < 1$, then $q^* = 0$
2. If $R_0 > 1$, then $q^* > 0$ and
3. If $R_0 = 1$, then if $k = p = 1$, $q^* = 1$ else $q^* = 0$

Proof:

Consider a node v at level n. Then since each of the n nodes leading from the root to v must successfully transmit the disease in order for this node to be reached.

$$\Pr[v \text{ is reached}] = p^n$$

If we let X_n be the number of infected individuals at level n, then since there are k^n nodes at level n, by linearity of expectations we have

$$\mathrm{E}[X_n] = (pk)^n$$

Note that the total number of nodes reached by this process is therefore $\sum_{n \geq 0} X_n$. Hence again by the linearity of expectations we have

$$\mathrm{E}[\text{Number of nodes reached}] = \textstyle\sum_n (pk)^n$$

If $R_0 = pk < 1$, then as n grows, R_0^n converges to 0 and hence the probability that the disease infects people at the n^{th} level must also be converging to 0 as n increases. Thus $q^* = 0$.

If $R_0 = pk > 1$, then E[Number of nodes reached] tends to infinity. However this does not mean that Pr[number of nodes reached is infinite] > 0. For example, let X be a random variable which takes the value 2^n with probability $1/2^n$. Here $\mathrm{E}[X] = \infty$ though $\Pr[X = \infty] = 0$.

Hence instead of using the expected number of infected individuals, we will develop a formula for q_n in terms of q_{n-1} where q_n is the probability that the epidemic survives at least n levels. Note that

$$q^* = \lim_{n \to \infty} q_n$$

Thus the probability that the disease does not survive the first n levels is given by

$$1 - q_n = [(1-p) + p(1 - q_{n-1})]^k$$

where $1-p$ is the probability that the disease does not suvive until level 1 while $p(1-q_{n-1})$ is the probability that it survives the first level but does not survive the next $n-1$ levels. The exponent k above arises since this happens independently for each of the k branches at this level. We thus have

$$q_n = 1 - (1 - q_{n-1})^k \tag{1}$$

where $q_0 = 1$ since by assumption the node at the root has the disease.

If we define the function $f(x) = 1 - (1 - px)^k$, then we can write equation (1) as $q_n = f(q_{n-1})$. Hence we have to study the sequence of values $1, f(1), f(f(1)), \ldots$ obtained by applying f repeatedly and look at the limiting value of this sequence.

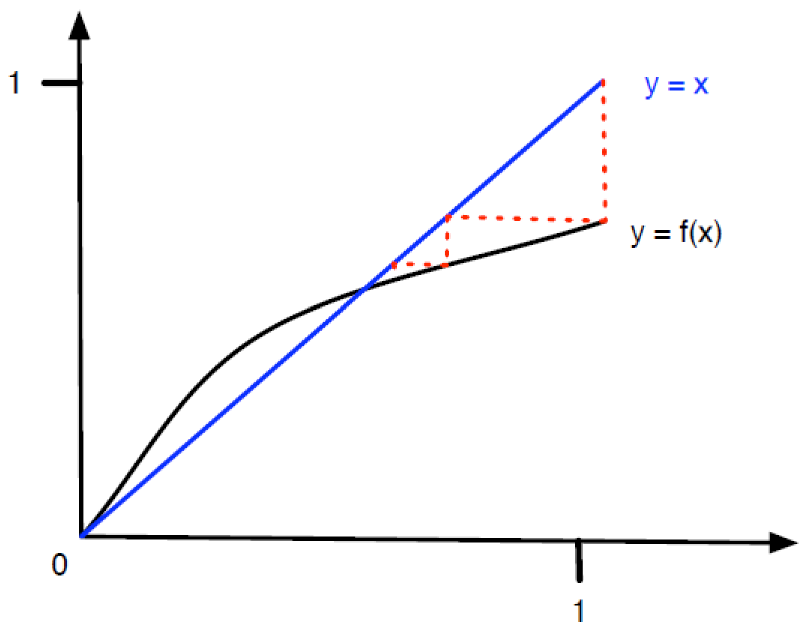
If we plot the function on a pair of $x - y$ axes, we can use the following facts

1. $f(0) = 0$ and $f(1) = 1 - (1 - p)^k < 1$. This means that the plot of f passes through the origin, but lies below the line $y = x$ at $x = 1$.
2. The first derivative $f^{'}(x) = pk(1 - px)^k >= 0$ while the second derivative $f^{''}(x) = -p^2k(k-1)(1-px)^{k-2} <= 0$. This means that $f(x)$ is an increasing function but takes a concave shape.
3. The slope of f at $x = 0$ is equal to $f^{'}(0) = pk = R_0$. So in the case where $R_0 > 1$, f starts out above the line $y = x$ but ends below it by the time we get to $x = 1$. Hence $y = f(x)$ must cross $y = x$ somewhere in the interval between 0 and 1 at a point $x^* > 0$.

Now using this plot, we can take a geometric view of the sequence of values

$$1, f(1), f(f(1)), f(f(f(1))) \ldots \tag{2}$$

In particular we can track this sequence on the line $y = x$. If we are currently at a particular point (x, x) on the line $y = x$, and we want to get to the point $(f(x), f(x))$ we can do that as follows. We first move vertically to the curve $y = f(x)$; this puts us at the point $(x, f(x))$. We then move horizontally back to the line y = x; this puts us at the point $(f(x), f(x))$ as desired. Continuing this process we pass through all points in the sequence (2) along the line $y = x$.



If we start this from $x = 1$, the process converges to the point (x^*, x^*) where the line $y = x$ meets the cuve $y = f(x)$. Now we can go back to the interpretation of this in terms of the branching process. The sequence (2) is precisely the sequence $q_0, q_1, q_2 \ldots$ as we argued above and so we have concluded that it converges to $x^* > 0$; the unique point at which $f(x) = x$ in the interval strictly between 0 and 1.

Also, when $R_0 = 1$ (provided it is not the case that $r = 1$ and $p = 1$) one can note that the only point at which the two curves intersect would be at $x = 0$. Thus the probability that the process goes on indefinitely is zero.

2 Percolation Theory

The branching process discussed above used a k-ary tree. In general we can extend this by considering an infinite graph where we pick each edge with probability p and again ask the question - What is the probability that there is an infinite connected component?

The initial motivation for percolation theory was the following question posed by Broabent and Hammersly in 1957. If a large porous stone is immersed in a bucket of water, does the centre of the stone get wet?

We can model the stone as 2-d grid. Let q be the probability that there exists an infinite connected component.

Claim 1: For any p; $q = 0$ or $q = 1$

This follows from the Kolmorogorov 0-1 law which is defined as folllows

Let E_n be a sequence of independent events. Let S be an event in the tail σ-field generated by E_n. Then Pr[S] = 0 or 1

Given a collection of events, the σ-field associated with this collection includes

1. The events in the collection
2. Their complements
3. Events such that the σ-field is closed with respect to union and intersection.

The tail σ-field generated by E_n is given by $\cap_{n \geq 1} \sigma(E_n, E_{n+1}, \ldots)$. Intuitively we can look at the tail σ-field as the set of events that are independent of any finite subset of E_n.

Now in order to prove claim 1 above, we consider a sequence of independent edges where E_i is the event that the ith edge is open. Let S be the event that there is an infinite size connected component consisting of open edges. Then $S \in \sigma\text{-field}(E_n, E_{n+1}, \ldots)\ \forall n$. Therefore $S \in$ tail σ-field generated by E_n. Intuitively a finite number of edges being closed can disconnect only a finite number of vertices from the rest of the lattice. Hence the existence or non-existence of an infinite component cannot be determined by an finite subsequence of E_n. Hence its probability must be 0 or 1.

Since P the probability of the existence of an infinite size component containing the origin is either 0 or 1 depending on the value of p, what we are most interested in is that value of p at which P goes from 0 to 1 (and stays there). The value of p at which this is achieved is called critical probability p_c.

3 Critical Probability for a 1-D Grid

The critical probability $p_c(1)$ for a 1-D grid is 1. Since the origin is part of the infinite connected component, there must exist an infinite connected component on at least one side of the origin. But in general Pr[Path of length k consisting of open edges] = p^k, so we have

$$P \leq lim_{n \to \infty} p^n$$

which equals zero if $p < 1$.

4 Critical Probability for a 2-D Grid

Theorem:

$$1/3 \leq p_c(2) \leq 2/3$$

We first prove that $p_c(2) > 1/3$. Let $N(n)$ be the number of open paths of length n. Now if the origin is a part of an infinite cluster there there exists paths of all possible lengths starting at the origin.

$$\forall n, P \leq Pr[N(n) \geq 1] \leq E(N(n))$$

The second inequality follows from the fact that the number of paths is always a positive integer.

The probability of given path of length n being open in p^n while the number of paths of length $n \leq 4 \cdot 3^{n-1}$. This is because every path that starts at the origin has four choices for the first edges and at most 3 choices for every vertex after that, eliminating the choice of going back to where it came from. Hence

$$P \leq 4/3(3p)^n$$

Hence if $p < 1/3$, $P \to 0$ as $n \to \infty$. Hence $p_c(2)$ must be at least 1/3.

References

[EK10] D. Easley and J. Kleinberg. *Networks, Crowds, and Markets: Reasoning About a Highly Connected World.* Cambridge University Press, 2010.