10/24: Giant Component

Tuesday, October 24, 2006
1:07 PM

Erdos-Renyi

If $p = (1-\varepsilon)n$, then $\exists\, c_1$ s.t.

(1) $\Pr[G(n,p)$ has component larger than $c\lg n] \xrightarrow{n\to\infty} 0$

If $p = (1+2\varepsilon)n$ then $\exists\, c_2$ s.t.

(2) $\Pr[G(n,p)$ has component larger than $c_2 n] \xrightarrow{n\to\infty} 1$

Will first show $\exists\, c_3$ s.t.

(3) $\Pr[G(n,p)$ has component larger than $c_2 n] \geq c_3$ for n suff large

Will use percolation on trees.

Recall for inf binary tree, proved that
for $p < \frac{1}{2}$ $\Pr[\exists$ inf component$] = 0$
$p > \frac{1}{2}$ $\Pr[\exists$ inf compont with $\bar{0}] > 0$

I'll tell you result for k-ary tree
For $p < \frac{1}{k}$ $\Pr[\exists$ inf component$] = 0$
For $p = \frac{(1+\varepsilon)}{k}$ $\Pr[\exists$ inf component from $\bar{0}] > C(\varepsilon)$

where $C(\varepsilon)$ is a constant depending on ε,
not k.

This will be in Problem Set 3.

Let's look at this as a branching process.
For each vertex, its # of children has distribution
$B(n,p)$: binomial with probability p

if $X \leftarrow B(n,p)$ then $\Pr[X = i] = \binom{n}{i} p^i (1-p)^{n-i}$

= prob flipping n p-biased coins gives "heads" i times

So, label root node 1, and let it have X_1 children, numbered 2,.., X_1+1. Then, pick node 2, let it have X_2 children, etc.

So, node X_i has i children, $X_i \leftarrow B(n,p)$, if there is a node i.

Percolation $\Longleftrightarrow$ this process goes on forever, and

finite $\Longleftrightarrow$ this process terminating

We now know if $p \geq \frac{1+\varepsilon}{n}$, $Pr[\text{goes forever}] \geq C(\varepsilon)$

As a statement just about random variables, this says if

$X_1, X_2, \ldots$ is a seq, each from $B(n,p)$

Then $Pr[\forall i, X_1 + \cdots + X_i \geq i] \geq C(\varepsilon)$

Now, to prove (3).

Consider process on a graph in which each vertex is either "asleep" "active" or "retired"
All vertices begin asleep, and end retired or asleep

At time t=0, set vertex 1 to be active, all others asleep.

At time $t \geq 1$, pick the lowest numbered (or arbitrary)

At time $t \geq 1$, pick the lowest numbered (or arbitrary) active vertex. Set all of its nbrs active, asleep and proclaim it to be retired.

When no active nodes remain, stop
Let T be # of steps = size of component of node 1.

let X_i = # of asleep nbrs of node considered at time i.
Y_i = # retired or active nodes at start of time i
Claim: $X_i \leftarrow B(n-A_i, p)$

because are $n-A_i$ asleep nodes at this time, have a probability p of each being a nbr.

Moreover, X_i is independent of $X_1, \ldots, X_{i-1}$ because edges from node considered at time i are independent of edges from previously considered nodes.

Now, for $Y_i \leq \left(\frac{\varepsilon}{1+2\varepsilon}\right) n$,

$$p \geq (1+\varepsilon)(n-Y_i)$$

So, for $i : Y_i \leq \left(\frac{\varepsilon}{1+2\varepsilon}\right) n$, can apply analysis of Branching process to say is prob $\geq C(\varepsilon)$ that branching process continues forever.

Here, it cannot go forever, but it keeps going until $p \geq (1+\varepsilon)(n-Y_i)$ is violated.

So, $\Pr\left[\exists i : Y_i \geq \left(\frac{\varepsilon}{1+2\varepsilon}\right) n\right\} \geq C(\varepsilon)$

$\Rightarrow \Pr\left[\text{compont of 1 has at least } \left(\frac{\varepsilon}{1+2\varepsilon}\right) n \text{ verts}\right\} \geq C(\varepsilon)$

$$c_2 = \frac{\varepsilon}{1+2\varepsilon} \qquad c_3 = C(\varepsilon)$$

In terms of random variables, we have

$$X_i \leftarrow B(n - \sum_{j<i} X_i - 1, p)$$

And, for $\sum_{j<i} X_i - 1 \leq c_2 n$, we use

$$\Pr[\exists i : X_1' + \dots + X_i' < i] \leq 1 - C(\varepsilon)$$

where $X_i' \leftarrow B((1-c_2)n, p)$

As, in this case, $\Pr[X_1 + \dots + X_i < i] < \Pr[X_1' + \dots + X_i' < i]$, we set.

Proving #1: if $p = \frac{(1-\varepsilon)}{n}$ all components have size $\leq c_1 \lg n$ for some c_1

First, re-examine percolation using Chernoff bound.

Recall: if $X_1, \dots, X_k$ are 0/1-valued random variables $X = \sum X_i$, $\mu = E[X]$, then $\forall \delta < 1$

$$\Pr[X > (1+\delta)\mu] \leq e^{-\frac{\delta^2 \mu}{3}}.$$

Now, our variables are sums of n 0/1-valued random variables.

So, if $Y_1, \dots, Y_t \leftarrow B(n, p)$

Then $Y = \sum Y_t$, $\mu = E[Y]$

$$\Pr[Y > (1+\delta)\mu] < e^{-\frac{\delta^2 \mu}{3}}$$

Branching process dies if $Y_1 + \cdots + Y_t < t$ for some t.

But, $\Pr[Y_1 + \cdots + Y_t \geq t]$ is small for $p = \frac{(1-\varepsilon)}{n}$,

as $\mu[Y] = (1-\varepsilon)t$, so

$$\Pr[Y_1 + \cdots + Y_t \geq t] = \Pr\left[Y_1 + \cdots + Y_t \geq \frac{1}{(1-\varepsilon)}\mu\right]$$

For $\delta = 1 - \frac{1}{1-\varepsilon} = \frac{\varepsilon}{1-\varepsilon}$

$$\leq e^{\frac{-\left(\frac{\varepsilon}{1-\varepsilon}\right)^2 \mu}{3}}$$

$$= \left(e^{\frac{-\varepsilon^2}{3(1-\varepsilon)}}\right)^t \xrightarrow[t\to\infty]{} 0$$

In particular, for $t = c_1 \ln n$, where

$$c_1 = 2 \cdot \frac{3(1-\varepsilon)}{\varepsilon^2}$$

$$\leq \frac{1}{n^2}$$

Returning to our graph process,

we find $\Pr[\text{comp of node } 1 \text{ has } \geq c_1 \ln n \text{ nodes}]$

$$\leq \frac{1}{n^2}$$

Could do for component of any node, so

$$\Pr[\text{any component} \geq c_1 \ln n \text{ nodes}] \leq \frac{1}{n}$$

Proving #2.

Let $B(G)$ = size of largest connected component

wts $\forall \varepsilon > 0$ if $p = \frac{(1+\varepsilon)}{n}$, $\exists c_2$ s.t.

$$\Pr[B(G) \le c_2 n] \xrightarrow{n \to \infty} 0$$

Again, consider our sequence $Y_1, Y_2, \ldots$

will show that, given that grow to some size, is very unlikely die out.

First, note $\Pr[Y_1 + \cdots + Y_t < t]$

$$E[Y_1 + \cdots + Y_t] = (1+\varepsilon)t = \mu$$

$$= \Pr\left[Y_1 + \cdots + Y_t < \left(\frac{1}{1+\varepsilon}\right)\mu\right] \le e^{-\frac{\delta^2 \mu}{2}}$$

$$1 - \delta = \frac{1}{1+\varepsilon}, \quad \delta = 1 - \frac{1}{1+\varepsilon} = \frac{\varepsilon}{1+\varepsilon}$$

$$= e^{-\frac{\varepsilon^2 (1+\varepsilon) t}{(1+\varepsilon)^2 \cdot 2}}$$

$$= \left(e^{\frac{-\varepsilon^2}{2(1+\varepsilon)}}\right)^t$$

So, for t suff large $\longrightarrow 0$

Similarly, $\Pr[\exists t \ge t_0 : Y_1 + \cdots + Y_t < t]$

$$\le \sum_{t \ge t_0} \left(e^{\frac{-\varepsilon^2}{2(1+\varepsilon)}}\right)^t \le \frac{\alpha^{t_0}}{1-\alpha}$$

where $\alpha = e^{-\frac{\varepsilon^2}{2(1+\varepsilon)}}$

$$\xrightarrow[t_0 \to \infty]{} 0$$

In fact, for $t_0 = C_4 \ln n$, $< \frac{1}{n^2}$

Now, to return to Graphs modify our earlier process so that when are still sleeping nodes, but none active, we pick an arbitry one to wake up.

What can happen?

Could find components of size $< t_0 = C_4 \ln n$

But, prob of this $< 1 - C(\varepsilon)$,

So chance find $2\ln(n)/c(\varepsilon)$ components this size

$$\leq (1 - C(\varepsilon))^{2\ln(n)/c(\varepsilon)} \leq e^{-2\ln n} \leq \frac{1}{n^2},$$

so unlikely.

Chance find component of size between t_0 and

cn where $c = \frac{\varepsilon}{1+\varepsilon}$ is also

small:

Set $Z_t = Y_1 + \cdots + Y_t$

$$\Pr[\exists t : t_0 \leq t \leq cn \; Z_t < t \mid Z_t \geq t \text{ for } t \leq t_0]$$

$$= \frac{\Pr[(\exists t : t_0 \leq t \leq cn \; Z_t < t) \wedge (Z_t \geq t \text{ for } t \leq t_0)]}{\Pr[Z_t \geq t \text{ for } t \leq t_0]}$$

$$- \frac{\qquad\qquad}{\Pr[Z_t \geq t \text{ for } t \leq t_0\}}$$

$$\leq \frac{\Pr[(\exists t : t_0 \leq t \leq cn \;\; Z_t < t)\}}{\Pr[Z_t \geq t \text{ for } t \leq t_0\}}$$

$$\leq \frac{1/n^2}{c(\varepsilon)} \leq \frac{1}{c(\varepsilon)\, n^2}$$