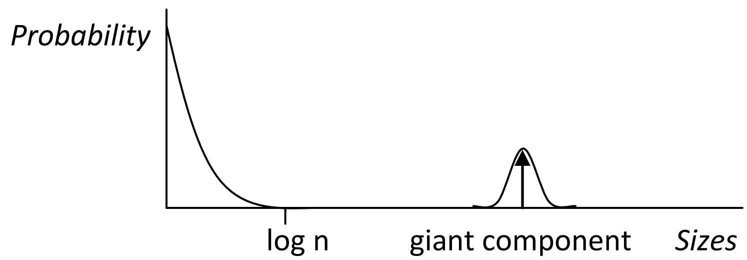
CS 4850 Lecture 10

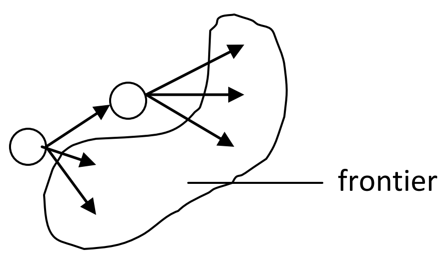
February 9, 2009 – Scribe: Eric First (ELF24)

Distribution of Component Sizes in G(n,p)

Consider $G(n,p)$ with $p = \frac{d}{n}$ and $d > 1$.

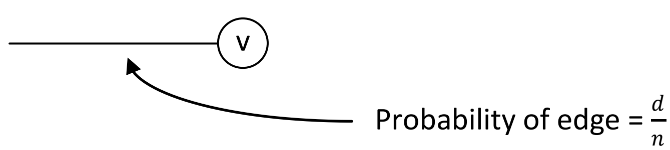


- Pick a vertex
- Explore component that the vertex is in
- Mark vertices as discovered or processed
- Run algorithm until frontier is empty



Initially, the frontier grows very large. After half the vertices are discovered, it grows more slowly. Eventually, the frontier doesn't grow at all and starts shrinking as vertices are processed.

We want the size of the frontier as a function of time.
Let z_i be the number of discovered vertices at time i.



$1 - (1 - \frac{d}{n})^i$ = probability of finding a given vertex within i steps

$$E(z_i) = n(1 - (1 - \frac{d}{n})^i) = n(1 - e^{-\frac{di}{n}})$$

Normalize by dividing by n

- Size of components
- Time

$S = E(z_i)$

Frontier = $S - i = n(1 - e^{-\frac{di}{n}}) - i$

$$\frac{S - i}{n} = 1 - e^{-\frac{di}{n}} - \frac{i}{n}$$

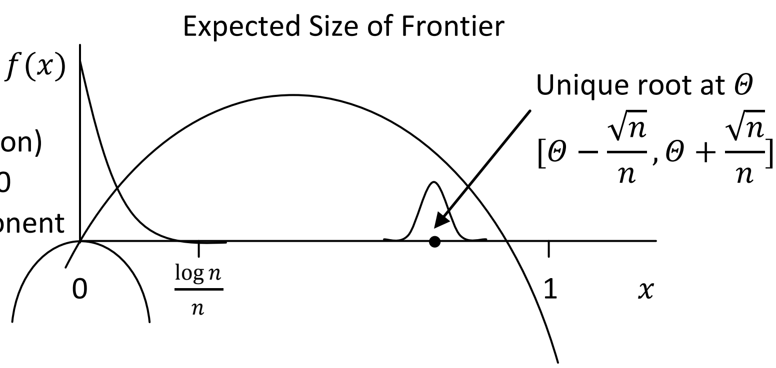
Normalized frontier:
$f(x) = 1 - e^{-dx} - x$
(Normalized time $x = \frac{i}{n}$)

What does this curve look like?

$f(x) = 1 - e^{-dx} - x$
$f'(x) = de^{-dx} - 1$
$f''(x) = -d^2 e^{-dx} < 0$

If $d = 1$ (phase transition)

- Double root at 0
- No giant component



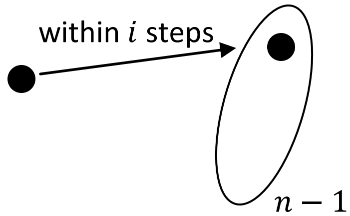
$f(0) = 0$
$f'(0) = d - 1 > 0$
$f(1) = -e^{-d}$

z_i is a random variable so its value will be close to but not exactly its expected value.

For small i, $i = O(\log n)$, and the expected size of the frontier grows as $(d - 1)i$.

$$n\left(1 - e^{-\frac{di}{n}}\right) - i = n\left(1 - \left(1 - \frac{di}{n} + \cdots\right)\right) - i = n\left(\frac{di}{n} - \cdots\right) - i = (d - 1)i + \cdots$$

What is the probability distribution of z_i?



$\text{prob} = 1 - e^{-\frac{di}{n}}$

$\text{binomial}(n-1, 1 - e^{-\frac{di}{n}}) \simeq \text{binomial}(n, \frac{di}{n})$

In this range, the distribution can be approximated by the Poisson, $\text{Prob}(k) = e^{-di}\,\frac{(di)^k}{k!}$

$$
\begin{aligned}
\text{Prob}(k) &= \binom{n}{k}\left(\frac{di}{n}\right)^k\left(1-\frac{di}{n}\right)^{n-k} \\
&= \frac{n(n-1)\ldots(n-k)}{k!}\left(\frac{di}{n}\right)^k\left(1-\frac{di}{n}\right)^{\frac{n}{di}-di} \\
&= \frac{n^k}{k!}\frac{(di)^k}{n^k}\,e^{-di} \\
&= \frac{(di)^k}{k!}\,e^{-di}
\end{aligned}
$$

$=$ Poisson Distribution

- Expected value of the distribution increases linearly as you move out from the origin.
- The Poisson Distribution drops off exponentially fast.
- The probability that a random variable will be zero drops off exponentially fast.

As you approach Θ, the Binomial Distribution can no longer be approximated by the Poisson because k is of order n, e.g. $k \simeq \frac{n}{2}$). Instead, approximate with the Gaussian.

$\binom{n}{k}p^k(1-p)^{n-k}$

$e^{-\frac{(m-k)^2}{\sigma^2}}$

$E = np$
$\sigma^2 = np(1-p) = O(n)$

Starts dropping off when $m-k$ gets to $\sqrt{n}$.

Simpler case:

$\binom{n}{k}(\frac{1}{2})^k(\frac{1}{2})^{n-k}$ $\qquad$ $E = \frac{n}{2}$

Binomial Distribution