Branching Processes

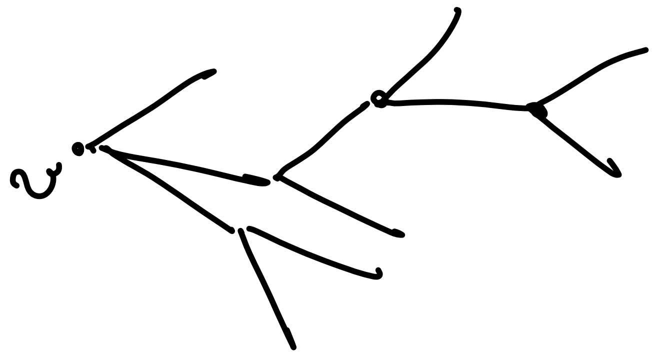
If $p = c/n$ and $d(v)$ is the degree of vertex v then

$$\Pr(d(v) = k) = \binom{n-1}{k} p^k (1-p)^{n-1-k}$$

$$= (1+o(1)) \frac{c^k e^{-c}}{k!}$$

i.e. the degree distribution is asymptotically Poisson with mean c.

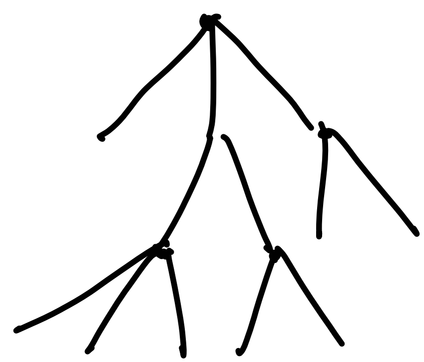
Since there are few "small" cycles, locally, $G_{n,p}$ should look like



and this has led to a comparison with Branching Processes.

It is not really so useful a method for here, but it can be the right approach for other models of a random graph.

In a simple branching process there is an initial individual who "gives birth" to X_1 children and then dies. Each of the X_1 individuals give birth and die and so on.



The number of children X produced by an individual is a random variable independent of the number produced by any other.

Let

$$p_k = P_r(X = k), \qquad k = 0, 1, 2, \ldots$$

and

$$G(z) = \sum_{k=0}^{\infty} p_k z^k$$

is the probability generating function (p.g.f.) of X.

Let

$$\mu = E(X)$$
$$= G'(1).$$

Let X_t be the number of individuals in generation t. Thus

$$X_0 = 1$$

$$E(X_{t+1}) = \sum_{k=0}^{\infty} E(X_{t+1} \mid X_t = k) \Pr(X_t = k)$$

$$= \sum_{k=0}^{\infty} k\mu \Pr(X_t = k)$$

$$= \mu E(X_t)$$

and so

$$E(X_t) = \mu^t.$$

Let T denote the total size of the set of individuals produced.

$T = \infty$ is allowed and $P_r(T=\infty)$ is one of the important parameters of the process.

Theorem

$P_r(T < \infty) = y$ where y is the smallest non-negative root of

$$y = G(y).$$

In particular, $y=1$ if $\mu \leq 1$.

Before proving this, let us consider the case where X has Poisson distribution with mean c.

$$G(z) = \sum_{k=0}^{\infty} \frac{c^k e^{-c}}{k!} z^k$$

$$= e^{c(z-1)}$$

From the theorem, the "extinction probability" y satisfies

$$y = e^{c(y-1)}$$

But then
$$cy e^{-cy} = c e^{-c}$$
Assume $c > 1$ and then $x = cy < 1$.

If we choose a vertex v and look at the BFS tree grown from v then (as we will check) this looks like our branching process.

If $T = \infty$ corresponds to being in the giant and v is chosen randomly, then
$$\Pr(v \in \text{Giant}) \simeq 1 - y = 1 - \frac{x}{c}.$$

Proof of Theorem

Let G_t be the p.g.f. for X_t. Thus

$$G_{t+1}(z) = \sum_{k=0}^{\infty} \sum_{\ell=0}^{\infty} P_r(X_{t+1}=k \mid X_t=\ell) P_r(X_t=\ell)\, z^k$$

$$= \sum_{\ell=0}^{\infty} G_t(z)^{\ell} P_r(X_t=\ell) \quad {**}$$

$$= G(G_t(z))$$

** If X, Y have p.g.f.'s f, g then X + Y has p.g.f. f × g.

Let $y_t = Pr(X_t = 0)$ so that

$$y_t = G_t(0) = G(G_{t-1}(0)) = G(y_{t-1}).$$

Now y_t is monotone increasing to $Pr(T<\infty)$ and so the continuity of G implies

$$y = G(y).$$

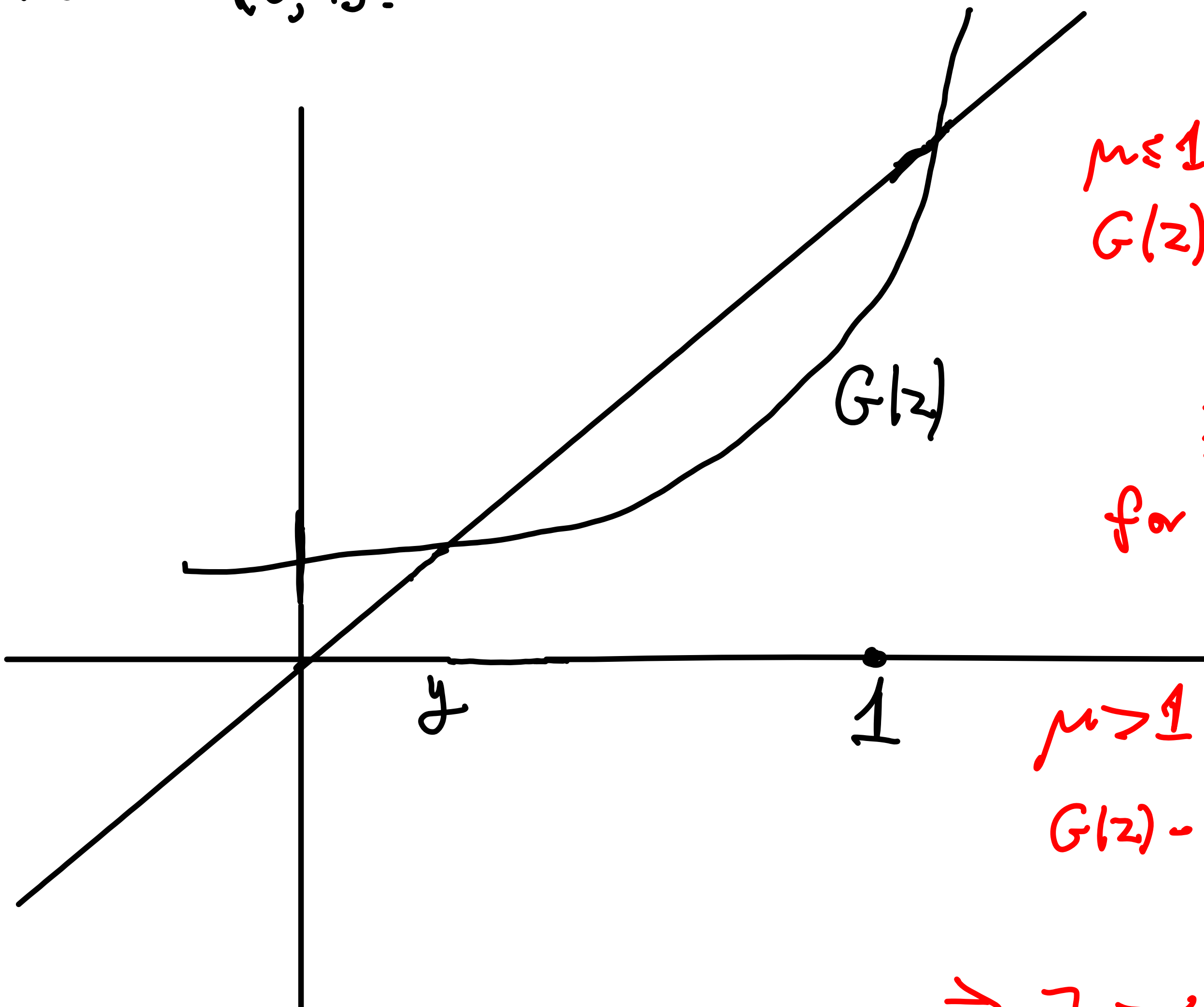
If ξ is any non-negative root of $z = G(z)$ then

$$y_1 = G(0) \le G(\xi) = \xi$$

and

$$y_t \le \xi \Rightarrow y_{t+1} = G(y_t) \le G(\xi) = \xi.$$

G is strictly convex on $[0,1]$ — $G''(z) = \sum_{k=2}^{\infty} k(k-1) p_k z^k > 0$ for $z \in (0,1]$.



$\mu \leq 1$:

$$G(z) > G(1) - G'(1)(1-z)$$
$$= 1 - \mu(1-z)$$
$$\geq z$$

for $0 \leq z < 1$

$\mu > 1$:

$G(z) - z$ < 0 $z = 1-\epsilon$

≥ 0 $z = 0$

$\Rightarrow \exists\, z < 1-\epsilon$ s.t. $G(z) - z = 0$. $\square$

Thus

$$y = Pr(T < \infty) = \lim_{t \to \infty} Pr(T \le t)$$

and we can write

$$Pr(T \le t) = y - \sigma(t)$$

where $\sigma(t) \geq 0$ and $\lim_{t \to \infty} \sigma(t) = 0$.

Back to $G_{n,p}$, $p = c/n$ $c > 1$.

Suppose we choose a vertex a and do a BFS from a until either

(i) we have explored the component C_a containing a

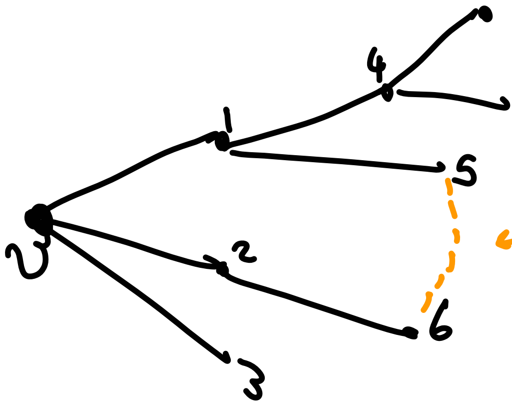
or

(ii) explored $\omega \to \infty$ vertices.*

Let T_a be the (partial) BFS tree produced.

$\boxed{\text{We are going for ease of proof rather than best possible}}$

Now fix a tree H with $\leq \omega = n^{1/2}(\log n)^3$ vertices and maximum degree $(\log n)^2$.



We do not include these edges in def. of T_a

Let d_i = degree of i, $i = 0, 1, 2, \ldots, \ell$

$$\Pr(H = T_a) = \prod_{i=0}^{\ell} \binom{n_i}{d_i} p^{d_i} (1-p)^{n_i - d_i - (s_i)}$$

where

$$n_i = n - 1 - d_1 - \cdots - d_{i-1}$$

$$= \left(\prod_{i=0}^{\ell} \frac{c^{d_i} e^{-c}}{d_i!} \right) \left(1 + O\left(\frac{\omega}{n} \right) \right)$$

$$= \left(\prod_{i=0}^{\ell} \frac{c^{d_i} e^{-c}}{d_i!} \right) \left(1 + O\left(\frac{\omega}{n} \right) \right)$$

$$= \Pr(H \text{ is branching process tree}) \times (1 + o(1))$$

Thus,

$$\Pr(|C_a| < \omega) =$$

$$\Pr(|C_a| < \omega \wedge \Delta \geq \overbrace{(\log n)^2}^{L}) +$$

$$\Pr(|C_a| < \omega \wedge \Delta \geq (\log n)^2)$$

$$\leq n \binom{n-1}{L} \left(\frac{c}{n}\right)^{L}$$

$$\leq n \left(\frac{ce}{L}\right)^{L} = o(1)$$

$$= o(1) + \sum_{H : |H| < \omega} \Pr\left((T_a = H) \wedge \left(\Delta(G \setminus C_a) \leq (\log n)^2 \right) \right)$$

$$= o(1) + \sum_{H : |H| < \omega} \Pr(T_a = H) \Pr\left(\Delta(G \setminus C_a) \leq (\log n)^2 \right)$$

$$= o(1) + (1 + o(1)) \sum_{H : H < \omega} \Pr(T_a = H)$$

$$= o(1) + (1 + o(1)) \sum_{H : |H| < \omega} \Pr(H \text{ is branching process tree})$$

$$= o(1) + (1 + o(1)) \Pr(T_a < \omega) \sim y .$$

Thus if

$$X_0 = \# v : |C_a| < \omega, \quad \omega \to \infty$$

then

$$E(X) = ny\left(1 - O(\omega/n) - o(\omega)\right).$$

We next show, via Chebychef, that X_0 is concentrated around its mean.

C_v

In constructing C_v we do not look at edges here i.e. they are unconditioned

We claim that for $b \neq a$

$$\Pr(|C_b| < \omega \mid |C_a| < \omega) \qquad (*)$$

$$\leq \frac{|C_a|}{n} + (1+o(1)) \Pr(|C_b| < \log n)$$

$\Pr(w \in C_a)$

Fixing C_a, we replace n by $n - (\leq \omega)$ in computing $\Pr(|C_a| < \omega)$.

It follows from (*) on previous page that

$$E(X_0^2) \leq$$
$$E(X_0) + E(X_0) \times \frac{\omega}{n} + (1+o(1))\,E(X_0)^2$$

i.e.

$$\mathrm{Var}(X_0) \leq 2E(X_0) + \eta E(X_0)^2$$

where $\eta \longrightarrow 0$.

Then

$$\Pr\left(\left|X_0 - E(X_0)\right| \geq \theta n\right)$$

$$\leq \frac{2E(X_0)}{\theta^2 n^2} + \frac{\eta\, E(X_0)^2}{\theta^2 n^2}$$

$$\longrightarrow 0 \quad \text{if} \quad \theta = \eta^{1/3}.$$

Our aim now is to show that REST is connected, without using previous analysis.

Suppose $|C_v| \geq n^{1/2}(\log n)^3$ and we stop our DFS from v when we reach $n^{1/2}(\log n)^3$.

We argue next that whp

$$|N(S_a)| \geq n^{1/2}(\log n)$$

Indeed

$$\Pr\Big(\exists S, T : |S| = \underbrace{n^{1/2}(\log n)^3}_{k},\ |T| = \underbrace{n^{1/2}\log n}_{\ell} :$$

S induces a connected subgraph and

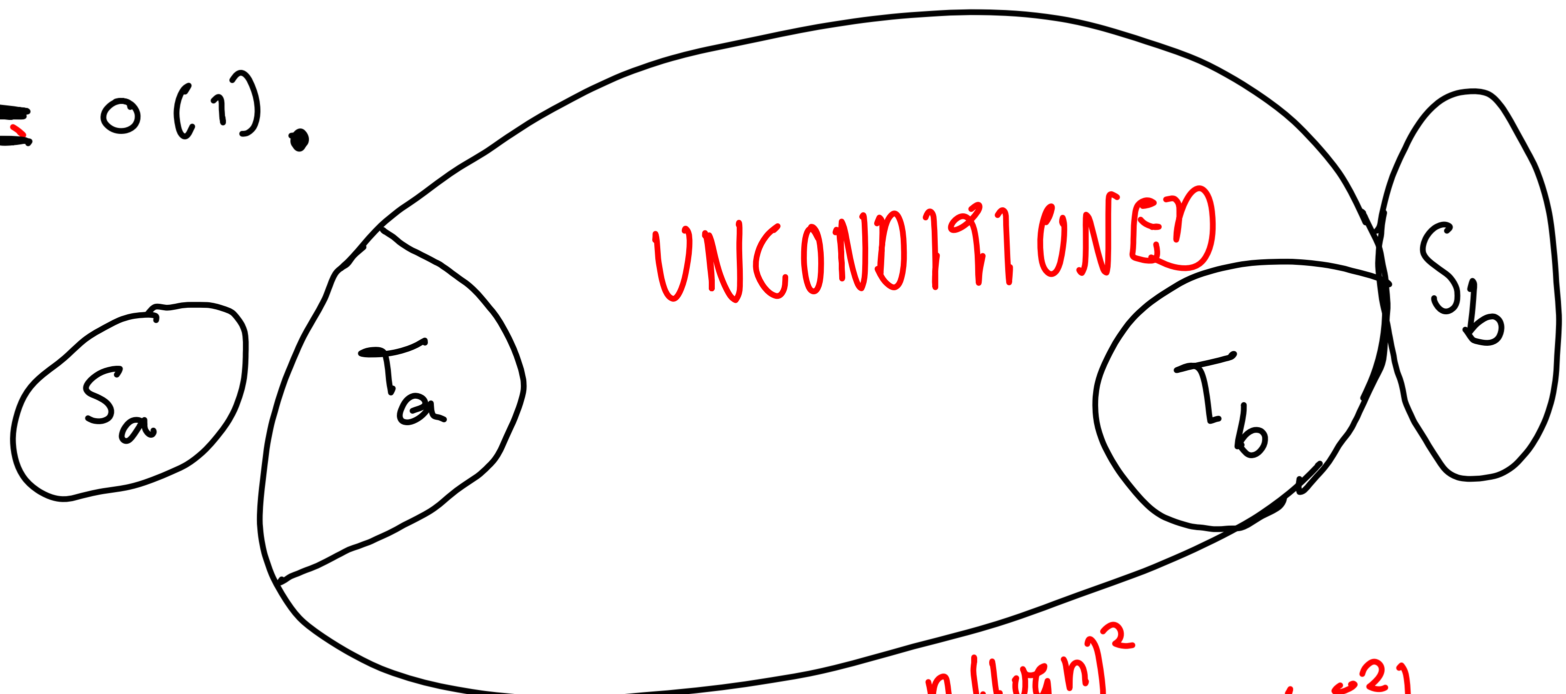
there are no $S : [n] \setminus S \cup T\Big)$

$$\leq \binom{n}{k}\binom{n}{\ell} k^{k-2} p^{k-1} (1-p)^{k(n-k-\ell)}$$

$$\leq \left(\frac{ne}{k}\right)^k \cdot \left(\frac{ne}{\ell}\right)^\ell \cdot k^{k-2} \left(\frac{c}{n}\right)^{k-1} e^{-ck(1-o(1))}$$

$$\leq n\left(ce^{1-c} \cdot n^{1/(\log n)^2}\right)^k \qquad \ell \leq \frac{k}{(\log n)^2}$$

$$= o(1).$$



$$\Pr(\text{no } T_a, T_b \text{ edges}) \leq (1-p)^{n(\log n)^2} = o(n^{-2}).$$

This shows that vertices a, $|C_a| \geq n^{\frac{1}{2}} (\log n)^3$ form a connected component.