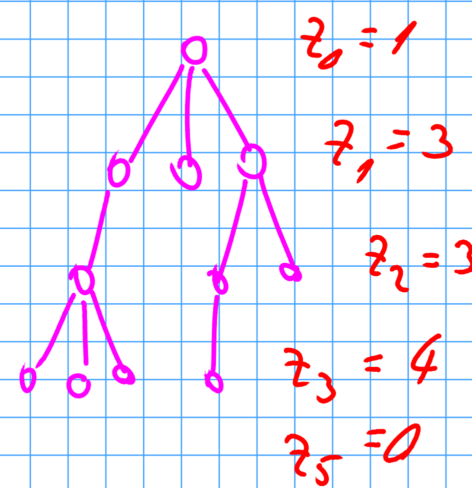
Branching processes

In the analysis of the size of the largest component of $\mathcal{G}(n, \frac{c}{n})$, we use some basic facts from the theory of branching processes. A Galton-Watson process is defined recursively by an offspring distribution $\{p_k\}$, $k=0,1,2,\dots$. We start by a root which has Z_1 children where $\mathbb{P}(Z_1 = k) = p_k$.

Now each child has a random number of children, drawn independently from the same distribution.



Let Z_i be the number of particles in the i-th generation.

Define the reproduction generating function by

$$f(s) = \mathbb{E}\left(s^{Z_1}\right) = \sum_{k=0}^{\infty} p_k \cdot s^k$$

for $s \in [0,1]$.

A key parameter is

$$m = \mathbb{E} Z_1 = \sum_{k=1}^{\infty} k p_k = f'(1)$$

We calculate

$$f_n(s) = \mathbb{E}\left(s^{Z_n}\right) = \mathbb{E}\,\mathbb{E}\left(s^{Z_n} \middle| Z_{n-1}\right)$$

$$= \mathbb{E}\,\mathbb{E}\left(s^{Y(1)+\dots+Y(Z_{n-1})} \middle| Z_{n-1}\right)$$

independent distributed as Z_1

$$= \mathbb{E} \prod_{i=1}^{Z_{n-1}} \mathbb{E}\left(s^{Y(i)}\right) = \mathbb{E}\left[f(s)^{Z_{n-1}}\right] = f_{n-1}(f(s))$$

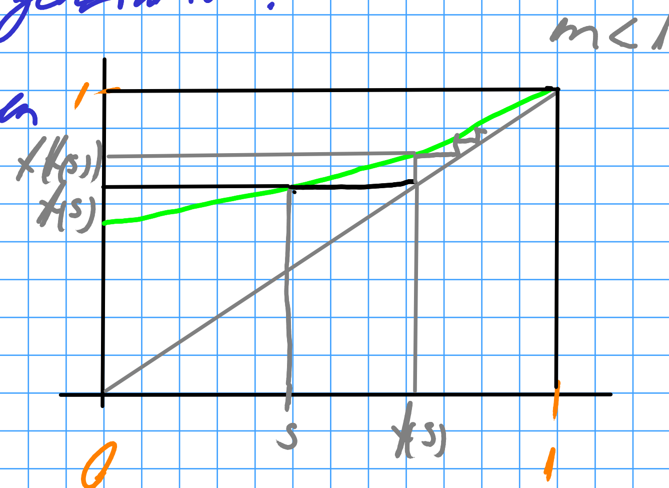
By induction, we see that

$$f_n(s) = \underbrace{f(f(\cdots f(s))\cdots)}_{n \text{ times}}$$

Note that $f(0) = p_0$, $f(1) = 1$, $f(s)$ is convex.

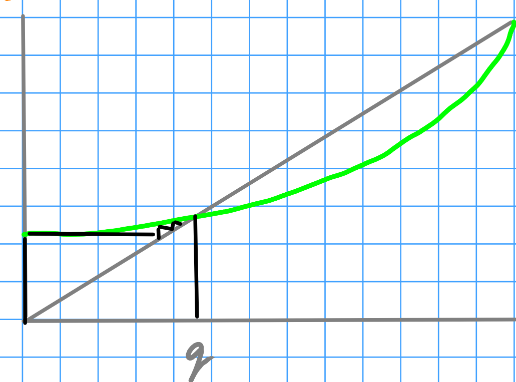
Also, $f_n(0) = \mathbb{P}(Z_n = 0)$ is the probability that the process is extinct at the n-th generation.

Suppose $m = f'(1) < 1$. Then



$f_n(0) \to 1$ as $n \to \infty$

When $m > 1$, $f_n(0) \to q$ where q is the unique solution of $f(q) = q$ such that $q < 1$.



Observe that

$$\begin{aligned} \mathbb{P}(\text{extinction}) &= \mathbb{P}(Z_n = 0 \text{ for some } n) \\ &= \lim_{n\to\infty} \mathbb{P}(Z_n = 0) \quad (\text{"continuity of measure"}) \\ &= q. \end{aligned}$$

Theorem The probability that a Galton-Watson process becomes extinct equals

1 if $m \le 1$ (except for the trivial case when $p_1 = 1$)

$q < 1$ if $m > 1$

Example Suppose $Z_1 \sim \text{Poisson}(c)$. Then

$$f(s) = \underbrace{\sum_{k=0}^{\infty} s^k \frac{c^k}{k!}}_{e^{cs}} e^{-c} = e^{c(s-1)}$$

so for $c>1$, the probability of extinction is the unique q such that $q = e^{c(q-1)}$

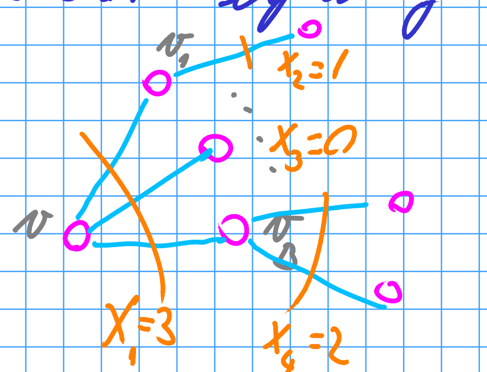
Example When $Z_1 \sim \text{Bin}(n, \frac{c}{n})$,

$$f(s) = \sum_{k=0}^{n} \binom{n}{k} s^k \left(\frac{c}{n}\right)^k \left(1-\frac{c}{n}\right)^{n-k} = \left(1 - \frac{c}{n} + \frac{sc}{n}\right)^n$$

This converges to $e^{c(s-1)}$ as $n \to \infty$

The largest component of $\mathcal{G}(n, \frac{c}{n})$

We start with the case $c<1$. We start "exploring" the graph at a given vertex v.



This vertex is connected to $X_1 \sim \text{Bin}(n-1, \frac{c}{n})$ vertices.

Mark v as saturated. Next reveal all neighbors of v_1 among the unseen vertices. Their number is X_2. Now v_1 becomes also saturated. Continue by exploring all grandchildren of v until all children are saturated.

Note that if S_k denotes the number of revealed vertices at step k, then, conditionally on the past,

$$X_k \sim \text{Bin}(n - |S_k|, \tfrac{c}{n})$$

The probability that a fixed vertex v_i belongs to a component of size $\geq k$ equals

$$\mathbb{P}\Big(\sum_{i=1}^{k} X_i \geq k-1\Big) \leq \mathbb{P}\Big(\sum_{i=1}^{k} X_i^+ \geq k-1\Big)$$

not independent — independent $\mathrm{Bin}(n, \frac{c}{n})$

But $\sum_{i=1}^{k} X_i^+ \sim \mathrm{Bin}(kn, \frac{c}{n})$ and so, by the union bound,

$$\mathbb{P}\Big(\mathcal{G}(n, \tfrac{c}{n}) \text{ has a component of size } \geq k\Big)$$

$$\leq n\, \mathbb{P}\Big(\mathrm{Bin}(nk, \tfrac{c}{n}) \geq k-1\Big)$$

n possible starting vertices

$$= n\, \mathbb{P}\Big(\mathrm{Bin}(nk, \tfrac{c}{n}) - \mathbb{E}\,\mathrm{Bin}(nk, \tfrac{c}{n}) \geq k-1-ck\Big)$$

$$\leq n \exp\Big(-\frac{(k(1-c)-1)^2}{2kc + \frac{2}{3}k(1-c)}\Big) \qquad \text{(Bernstein's inequality)}$$

$$\leq n \exp\Big(-k\frac{(1-c)^2}{2}\Big)$$

which goes to zero if $k \geq \frac{3}{(1-c)^2} \log n$. Thus,

For $c<1$, the largest component of $\mathcal{G}(n, \frac{c}{n})$ is of size at most $\frac{3}{(1-c)^2} \log n$ w.h.p.

The case when $c > 1$ is a little bit more complex.

Let $k_- = \frac{16c}{(c-1)^2} \log n$, $k_+ = n^{2/3}$. First we show that w.h.p. for every $k \in [k_-, k_+]$, for all vertices v either the exploration process becomes extinct after less than k_- steps or at the k-th step there remain at least $(c-1)k/2$ explored but not yet saturated vertices in the component of v.

To verify whether the exploration process produces a component with at least $(c-1)k/2$ unsaturated vertices, we only need to identify $k + \frac{c-1}{2}k = k\frac{c+1}{2}$ vertices.

So counting only connections outside of the component, we lower bound $X_i \geq X_i^- \sim \mathrm{Bin}\left(n - \frac{c+1}{2}k^+, \frac{c}{n}\right)$.

Then

$$\mathbb{P}\left(\begin{array}{c}\text{the process produces less than } (c-1)k/2 \text{ unsaturated} \\ \text{vertices in } k \text{ steps}\end{array}\right)$$

$$\leq \mathbb{P}\left(\sum_{i=1}^{k} X_i^- \leq k - 1 + \frac{(c-1)k}{2}\right)$$

The probability that this happens for some vertex v and for some $k \in [k^-, k^+]$ is at most

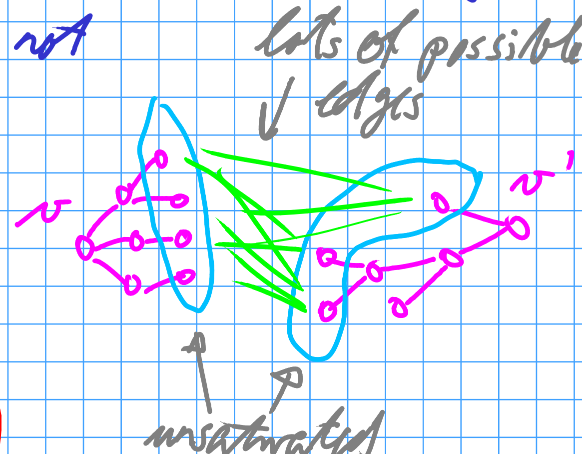
$$n \sum_{k=k^-}^{k^+} \mathbb{P}\left(\sum_{i=1}^{k} X_i^- \leq k - 1 + \frac{(c-1)k}{2}\right) = n \sum_{k} \mathbb{P}\Big(\underbrace{\mathrm{Bin}\left(nk - \tfrac{c+1}{2}k^+k, \tfrac{c}{n}\right)}_{\mathbb{E} \,\approx\, ck - \frac{c(c+1)}{2}\frac{k^+k}{n}} \leq k\frac{c+1}{2}\Big)$$

for n large enough

$$\leq n \sum_{k=k^-}^{k^+} \exp\left(-\frac{(c-1)^2 k^2}{9ck}\right) \leq n k^+ \exp\left(-\frac{(c-1)^2}{9c} k^-\right) \longrightarrow 0.$$

Thus, w.h.p., every vertex belongs to either a component of size $< k^-$ or $> k^+$.

Now let v, v' be vertices such that both of them belong to components of size $> k^+$.
Run the exploration process from both v and v' for k^+ steps. Then w.h.p. both components contain at least $(c-1)k^+/2$ unsaturated vertices.
If the two components are not connected already, the probability that they do not connect in the next step is



at most

$$\left(1-\frac{c}{n}\right)^{\left[(c-1)k_+/2\right]^2} \leq e^{-(c-1)^2 c \cdot n^{1/3}/4} = o\left(\frac{1}{n^2}\right)$$

By the union bound — for all $\binom{n}{2}$ pairs of vertices — the probability that there are two components of size at least k^+ goes to zero.
Thus, w.h.p., **all** vertices either belong to components of size at most $k^- = O(\log n)$ or to a single component of size $> k^+ = n^{2/3}$.

To estimate the size of the large component, we bound the number of vertices in small components.
The probability that a given vertex is in a small component is bounded by the extinction probability q of a Galton-Watson process with $z_i \sim \mathrm{Bin}(n-k_-, \frac{c}{n})$.

This converges to the extinction probability of a GW process with offspring distribution $\mathrm{Poi}(c)$ which satisfies $q_c = e^{c(q_c - 1)} \Rightarrow q_c < 1$.

Thus, the expected number of vertices in small components is $n(q_c + o(1))$

Exercise: Show that if Y denotes the number of vertices in components of size at most $\bar{k}$, then $\mathbb{E}(Y^2) = (1+o(1))(\mathbb{E}Y)^2$ and therefore $\mathcal{G}(n, \frac{c}{n})$ with $c>1$ contains $nq_c + o(n)$ such vertices.

Summarizing: When $c>1$, $\mathcal{G}(n,\frac{c}{n})$ contains a unique "giant" component of size $nq_c + o(n)$ and the second largest component is of size at most $\frac{16c}{(c-1)^2}\log n$.

What happens when $c = 1$?
It turns out that the correct scale to look at is
$$p = \frac{1}{n} + \frac{\lambda}{n^{4/3}}.$$
When $\lambda = \lambda_n \to -\infty$, the largest component is of size $o(n^{2/3})$ and there are many components of nearly the size of the largest.
When $\lambda \to \infty$, there is one component of size $\gg n^{2/3}$, all other components are of size $o(n^{2/3})$.
To gain some intuition, let X_k be the number of tree components of size k. Then
$$\mathbb{E}X = \binom{n}{k} k^{k-2} p^{k-1} (1-p)^{k(n-k)+\binom{k}{2}-k+1}$$
Let $p = \frac{1}{n} + \frac{\lambda}{n^{2/3}}$ and $k = c \cdot n^{2/3}$. Then — after very careful approximation —
$$\mathbb{E}X_{c\cdot n^{2/3}} \sim n^{-2/3} e^{-c^3/3 - \lambda^2 c/2 + \lambda c^2/2} \cdot c^{-5/2} / \sqrt{2\pi}$$
This goes to 0, but the expected number of tree components of size $k \in [a n^{2/3}, b \cdot n^{2/3}]$ is
$$\mathbb{E}X \longrightarrow \int_a^b \frac{1}{\sqrt{2\pi}} e^{-c^3/3 - \lambda^2 c/2 + \lambda c^2/2} \cdot c^{-5/2} \, dc$$
There are components of the order $n^{2/3}$ that are not trees though their number is very small compared to tree components.
For any fixed λ the largest components are of size $c n^{2/3}$ where the distribution of c depends on λ.