Branching processes

In the analysis of the size of the largest component of $G(n, c/n)$, we use some basic facts from the theory of branching processes. A Galton-Watson process is defined recursively by an offspring distribution $\{p_k\}$, $k = 0, 1, 2, \ldots$. We start by a root which has $2_0$ children where $P(2_i = k) = p_k$.

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

Let $\xi_i$ be the number of particles in the $i$-th generation.

Define the reproduction generating function by

$$X(s) = \mathbb{E}(s^{\xi_0}) = \sum_{k=0}^{\infty} p_k s^k$$

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

A key parameter is

$$m = \mathbb{E} \xi_0 = \sum_{k \geq 1} k p_k = X'(1)$$

We calculate

$$X_{n}(s) = \mathbb{E}(s^{\xi_n}) = \mathbb{E}\mathbb{E}(s^{\xi_n}/\xi_{n-1})$$

$$= \mathbb{E}\mathbb{E}\left(s^{\xi_1 + \cdots + \xi_{n-1}}/\xi_{n-1}\right)$$

indipendently distributed as $\xi_i$.

$$= \mathbb{E}\prod_{i=1}^{n-1} \mathbb{E}(s^{\xi_i}) = \mathbb{E}\left(s^{\mathbb{E}[\xi_1 + \cdots + \xi_{n-1}]}ight) = X_{n-1}(s)$$
By induction, we see that
\[ X_n(s) = X(s)^n (X(s))^{n-1} \]

Note that \( X(0) = p_0, \ X(1) = 1, \ A(s) \) is convex.

Also, \( X_n(0) = P(Z_n = 0) \) is the probability that the process is extinct at the \( n \)-th generation.

Suppose \( m = X'(1) < 1 \). Then
\[ X_n(0) \to 1 \text{ as } n \to \infty \]

When \( m > 1 \), \( X_n(0) \to g \) where \( g \) is the unique solution of \( X(g) = g \) such that \( g < 1 \).

Observe that
\[ P(\text{extinction}) = P(\exists n \text{ s.t. } Z_n = 0) \]
\[ = \lim_{n \to \infty} P(Z_n = g) \quad \text{(continuity of measure)} \]

**Theorem:** The probability that a Galton-Watson process becomes extinct equals
\[ 1 \text{ if } m \leq 1 \quad (\text{except for the trivial case when } \rho = 1) \]
\[ g < 1 \text{ if } m > 1 \]
Example
Suppose \( Z_i \sim \text{Poisson}(c) \). Then
\[
X(s) = \sum_{k=0}^{\infty} s^k \frac{c^k e^{-c}}{k!} = e^{-sc} e^{-c} = e^{-c(s-1)}
\]
so for \( c > 1 \), the probability of extinction is the unique \( \theta \) such that \( \theta = e^{-c(s-1)} \).

Example
When \( Z_i \sim \text{Bin}(n, \frac{\xi}{n}) \),
\[
X(s) = \sum_{k=0}^{n} \binom{n}{k} s^k (1 - \frac{\xi}{n})^{n-k} = \left( 1 - \frac{s}{n} + \frac{\xi}{n} \right)^n
\]
This converges to \( e^{-c(s-1)} \) as \( n \to \infty \).

The largest component of \( \tilde{G}(n, \frac{\xi}{n}) \)

We start with the case \( c < 1 \). We start "exploring" the graph at a given vertex \( v \).
This vertex is connected to \( X \sim \text{Bin}\left(n - 1, \frac{\xi}{n}\right) \) vertices.
Mark \( v \) as saturated. Next reveal all neighbors of \( v \), among the unseen vertices. Their number is \( X_2 \). Now \( v \) 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,
\[
X_k \sim \text{Bin}\left(n - 1, \frac{\xi}{n}\right)
\]
The probability that a fixed vertex $i$ belongs to a component of size $\geq k$ equals

$$P\left(\sum_{i=1}^{n} X_i \geq k-1\right) \leq P\left(\sum_{i=1}^{n} X_i \geq k-1\right) \text{ (not independent)}$$

$$P(\text{Bin}(n, 1/2) \leq k-1) \leq \text{Bin}(n, 1/2)$$

But $\frac{1}{2} \leq X_i < 1$ and so, by the union bound,

$$P(\text{Bin}(n, 1/2) \leq k-1)$$

$$\leq n \cdot P(\text{Bin}(nk, 1/2) \leq k-1)$$

in possible starting vertices

$$= n \cdot P(\text{Bin}(nk, 1/2) - n \text{Bin}(nk, 1/2) \leq k-1 - nk)$$

$$\leq n \exp\left(-\frac{(k(1-c)-1)^2}{2nk + 3nk(1-c)}\right)$$

(Bernstein's inequality)

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

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

For $c < 1$, the largest component of $G(n, 1/2)$ is of size at most $\frac{2}{(1-c)^2} \log n$ w.h.p.
The case when \( c > 1 \) is a little bit more complex.

Let \( \kappa_0 = \frac{16c}{(c-1)^3} \log n \), \( \kappa_2 = n^{\frac{2}{3}} \). First we show that w.h.p. for every \( \kappa \in [\kappa_0, \kappa_2] \), for all vertices \( v \) either the exploration process becomes extinct after less than \( \kappa \) steps or all the \( \kappa \)-th step there remain at least \((c-1)\kappa/2\) explored but not yet saturated vertices in the component of \( v \).

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

So counting only connections outside of the component, we lower bound \( X_i \geq X_i^- + \text{Bin}(n - \frac{c+1}{2} \kappa, 1/2) \).

Then

\[
P( \text{the process produces less than } (c-1)\kappa/2 \text{ unsaturated vertices in } \kappa \text{ steps} ) \leq \prod_{i=1}^{\kappa} P(X_i^- \leq \kappa - 1 + \frac{c-1}{2} \kappa) \]

The probability that this happens for some w.h.p. \( n \) and for some \( \kappa \in [\kappa_0, \kappa_2] \) is at most

\[
\sum_{\kappa_0}^{\kappa_2} \prod_{i=1}^{\kappa} P(\text{Bin}(nk - \frac{c+1}{2} \kappa, 1/2) \leq \kappa \frac{c+1}{2}) \leq \sum_{\kappa_0}^{\kappa_2} \exp\left(-\frac{(c-1)^2 \kappa^2}{3c} \right) \leq \exp\left(-\frac{(c-1)^2 \kappa^2}{3c} \right) \rightarrow 0.
\]
Thus, w.h.p., every vertex belongs to either a component of size \( < k^* \) or \( > k^* \).

Now let \( u, u' \) be vertices such that both of them belong to components of size \( > k^* \).

Run the exploration process from both \( u \) and \( u' \) for \( k^* \) steps. Then w.h.p., both components contain at least \( (c-\frac{1}{2})k^*/2 \) unsaturated vertices.

If the two components are not connected already, the probability that they do not connect in the next step is at most

\[
\left( \frac{c-\frac{1}{2}}{c} \right)^2 \left( 1 - \frac{1}{c} \right)^{k^*} \leq e^{-\frac{k^*}{2c}} = o\left( \frac{1}{n} \right)
\]

By the union bound, in 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 \( > k^* \approx n^{2/3} \).

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 \( ? \sim \text{Bin}(n-k, \frac{1}{n}) \).
This converges to the extinction probability of a GW process with offspring distribution \( \text{Poi}(c) \) which satisfies \( \lambda = c(\lambda - 1) \implies \lambda < 1 \).

Thus, the expected number of vertices in small components is \( \sqrt{n} \lambda + o(1) \).

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

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

It turns out that the correct scale to look at is

$$p = \frac{1}{n} + \frac{1}{\sqrt{n}}$$

when $1 = \frac{1}{n} \Rightarrow -\infty$, the largest component is of size $o(n^{1/3})$ and there are many components of nearly the size of the largest.

When $1 \Rightarrow \infty$, there is one component of size $\gg n^{1/3}$, all other components are of size $o(n^{1/3})$.

To gain some intuition, let $X_k$ be the number of size components of size $k$. Then

$$EX_k = (\binom{n}{k})^2 p^{k-1} \left(1 - \binom{k}{2} - p^k\right)$$

Let $p = \frac{1}{n} + \frac{1}{\sqrt{n}}$ and $k = c \cdot n^{1/3}$. Then, after very careful approximation,

$$EX_k \approx e^{-2/3 - c^{2/3} - 1/2 + 6c^{2/3}}$$

This goes to 0, but the expected number of size components of size $k \in \left[\frac{n^{1/3}, \sqrt{n}}{2}\right]$ is

$$EX \Rightarrow \int_{\frac{n^{1/3}}{2}}^{\sqrt{n}} e^{-2/3 - c^{2/3} - 1/2 + 6c^{2/3}} \cdot \frac{1}{\sqrt{n}} dc$$

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 $\lambda$, the largest component are of size $\lambda n^{1/3}$ where the distribution of $\lambda$ depends on $\lambda$. 