| M375T/M396C: Topics in Complex Networks | Spring 2013 |
|---|---|
| Lecture 03 — 01/22/13 | |
| *Lecturer: Ravi Srinivasan* | *Scribe: Matt Lubrano* |

# 3.1 Galton-Watson branching processes (cont'd)

Recall that $\xi_i^n$ be the number of children of parent $i$ in generation $n$ of the randomly generated tree $\mathcal{T}$. Here, the $\{\xi_i^n\}_{i,n\in\mathbb{N}}$ are iid with the same distribution as some random variable $\xi$ with given distribution $p_k = \mathbb{P}(\xi = k)$, $k \in \mathbb{N}$ (i.e., the offspring distribution). Define $X_n$ to be the number of children in generation $n$, so that

$$X_{n+1} = \sum_{i=1}^{X_n} \xi_i^n.$$

Let $\mu = \mathbb{E}[\xi]$ denote the mean of the offspring distribution and $\phi_\xi = \mathbb{E}\left[s^\xi\right] = \sum_{k=0}^\infty s^k p_k$ the probability generating function of $\xi$. Furthermore, let $p_{\text{ext}} = \mathbb{P}(|\mathcal{T}| < \infty)$ be the probability that the tree $\mathcal{T}$ has finite size—that is, the probability that the population goes extinct. We will now how this extinction probability varies as a function of $\mu$.

## 3.1.1 Phase transition for extinction probability

**Theorem 3.1.** *Let $\mu$, $\phi_\xi$, and $p_{\text{ext}}$ be defined as before. Then the following statements hold:*

1. *Subcritical regime:* If $\mu < 1$, then $p_{\text{ext}} = 1$.

2. *Critical regime:* If $\mu = 1$ and $p_1 < 1$, then $p_{\text{ext}} = 1$.

3. *Supercritical regime:* If $\mu < 1$, then $p_{\text{ext}} < 1$ and $p_{\text{ext}}$ is the smallest solution of $s = \phi_\xi(s)$.

**Proof:** Let $p_{\text{ext}}^{(n)} = \mathbb{P}(X_n = 0)$ be the probability that extinction happens at or before generation $n$. Note here that $\{X_n = 0\} \subseteq \{X_{n+1} = 0\}$ for all $n \in \mathbb{N}$, since if the number of individuals in the $n^{\text{th}}$ generation is zero then the number of individuals in all subsequent generations remains zero. So $p_{\text{ext}}^{(n)}$ is increasing in $n$ and is bounded above by 1. Hence, $p_{\text{ext}} = \lim_{n\to\infty} p_{\text{ext}}^{(n)}$ exists, but it remains to be seen what its value is. As we now show, $p_{\text{ext}}$ satisfies the equation $p_{\text{ext}} = \phi_\xi(p_{\text{ext}})$ and is therefore the smallest fixed point of the map $s \mapsto \phi_\xi(s)$.

Define $\phi_n(s) = \mathbb{E}\left[s^{X_n}\right]$ to be the probability generating function of $X_n$. The number of children in generation $n$ depends on the number of children in generation $n-1$, which depends on the number of children in generation $n-2$, and so on. In particular, the number of children in generation $n$ depends on the number of children $X_1$ in the first generation. Conditioning on $X_1$ and using the tower property of conditional expectation, we have

$$\mathbb{E}\left[s^{X_n}\right] = \mathbb{E}\left[\mathbb{E}\left[s^{X_n}|X_1\right]\right] = \sum_{k=0}^{\infty} \mathbb{E}\left[s^{X_n}|X_1 = k\right]\mathbb{P}(X_1 = k).$$

Let us denote by $\tilde{X}_{n-1}^{(j)}$ the total number of individuals of generation $n$ which are descendants of the $j^{\text{th}}$ member of the first generation. Then,

$$\begin{aligned}
\mathbb{E}\left[s^{X_n}|X_1 = k\right] &= \mathbb{E}\left[s^{\tilde{X}_{n-1}^{(1)}+\cdots+\tilde{X}_{n-1}^{(k)}}\right] \\
&= \mathbb{E}\left[s^{\tilde{X}_{n-1}^{(1)}}\right]\cdots\mathbb{E}\left[s^{\tilde{X}_{n-1}^{(k)}}\right] \\
&= \left(\mathbb{E}\left[s^{X_{n-1}}\right]\right)^k.
\end{aligned}$$

Here, we have used that the subtrees $\mathcal{T}_1, \ldots, \mathcal{T}_k$ generated from individuals of the first generation are independent and have the same distributional properties, so that the $\left\{\tilde{X}_{n-1}^{(j)}\right\}_{j=1}^{k}$ are iid with the same distribution as $X_{n-1}$. Since $\mathbb{P}(X_1 = k) = p_k$ we obtain

$$\begin{aligned}
\phi_n(s) &= \sum_{k=0}^{\infty}\left(\mathbb{E}\left[s^{X_{n-1}}\right]\right)^k p_k \\
&= \sum_{k=0}^{\infty}\left(\phi_{n-1}(s)\right)^k p_k \\
&= \phi_\xi\left(\phi_{n-1}(s)\right).
\end{aligned}$$

Using that $\phi_n(0) = \mathbb{P}(X_n = 0) = p_{\text{ext}}^{(n)}$, the previous display implies $p_{\text{ext}}^{(n)} = \phi_\xi\left(p_{\text{ext}}^{(n-1)}\right)$. We take $n \to \infty$ to find $p_{\text{ext}} = \phi_\xi(p_{\text{ext}})$.

Alternatively, we can use a depth-first argument in contrast to the breadth-first argument given above to show the same result. To see this, note that

$$p_{\text{ext}} = \mathbb{P}(|\mathcal{T}| < \infty) = \sum_{k=0}^{\infty}\mathbb{P}(|\mathcal{T}| < \infty|X_1 = k)\mathbb{P}(X_1 = k).$$

Since

$$\begin{aligned}
\mathbb{P}(|\mathcal{T}| < \infty|X_1 = k) &= \mathbb{P}(|T_1| < \infty, \cdots, |T_k| < \infty) \\
&= \mathbb{P}(|T_1| < \infty)\cdots\mathbb{P}(|T_k| < \infty) \\
&= \left(\mathbb{P}(|\mathcal{T}| < \infty)\right)^k,
\end{aligned}$$

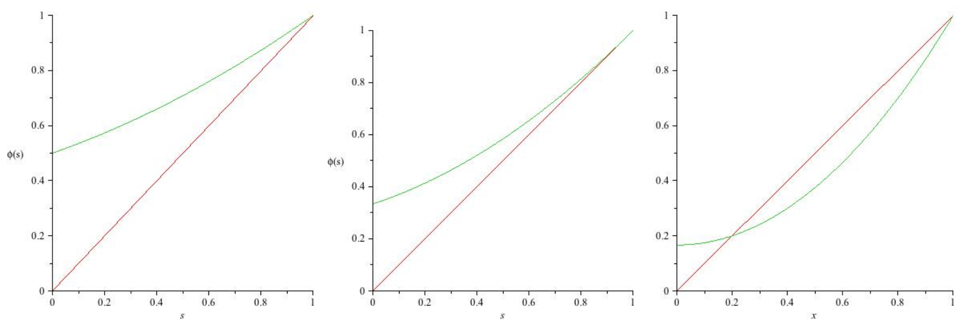

**Figure 3.1.** The graph of $\phi_\xi(s)$ versus $s$ for $0 \le s \le 1$ in the subcritical, critical, and supercritical cases, respectively.

we again find that

$$p_{\text{ext}} = \sum_{k=0}^{\infty} (p_{\text{ext}})^k \, p_k = \phi_\xi(p_{\text{ext}}).$$

To finish the proof, consider the behavior of $\phi_\xi$ for different values of $\mu$. Recall from the previous lecture that we can recover the distribution of a random variable from its probability generating function by differentiating it repeatedly. In particular,

$$\phi_\xi'(s) = \sum_{k=1}^{\infty} k s^{k-1} p_k \ge 0, \qquad \phi_\xi''(s) = \sum_{k=1}^{\infty} k(k-1) s^{k-2} p_k \ge 0$$

which implies that $\phi_\xi$ is increasing and convex for $0 \le s \le 1$. If, in addition, $p_0 + p_1 < 1$ then we must have $p_k > 0$ for some $k \ge 2$, so that $\phi''(s) > 0$ and $\phi_\xi$ is *strictly* convex. It is also easy to check that $\phi_\xi(0) = p_0$, $\phi_\xi(1) = 1$, and $\phi_\xi'(1) = \sum_{k=1}^{\infty} k p_k = \mu$. To conclude:

1. If $\mu < 1$, convexity implies that the only fixed point of the map $s \mapsto \phi_\xi(s)$ is $s = 1$. Therefore, $p_{\text{ext}} = 1$.
2. If $\mu = 1$ and $p_1 < 1$, we must have that $p_0 + p_1 < 1$ (since otherwise $p_k = 0$ for all $k \ge 2$ which implies $\mu = 0 \cdot p_0 + 1 \cdot p_1 < 1$). So $\phi_\xi$ is strictly convex and again the only fixed point of the map $s \mapsto \phi_\xi(s)$ is $s = 1$, giving $p_{\text{ext}} = 1$.
3. If $\mu > 1$, convexity implies that the map $s \mapsto \phi_\xi(s)$ has two fixed points, the smaller of which is $p_{\text{ext}} < 1$. In particular, if $p_0 = 0$ then $\phi_\xi(0) = 0$ and $p_{\text{ext}} = 0$. On the other hand, if $p_0 > 0$ then $p_{\text{ext}} > 0$ as well. □

### 3.1.2 Branching processes and random walks

Remarkably, there exists a connection between branching processes and one of the most fundamental objects in probability theory: a random walk. This correspondence will allow us to study properties of branching processes using classical results. A *random walk* $\{S_n : n \geq 0\}$ on $\mathbb{Z}$ is defined by

$$\begin{aligned} &S_n = S_{n-1} + X_n, \quad n \geq 1 \\ &S_0 = 0 \end{aligned}$$

where $\{X_n\}_{n=1}^{\infty}$ are iid random variables with some known distribution.

How can we construct a correspondence to branching processes? Recall that the two most common methods of exploring a given tree $\mathcal{T}$ are:

- *Breadth-first search:* Here, we visit all children of a node $i \in \mathcal{T}$ before visiting any grandchildren of node $i$.
- *Depth-first search:* For a given node $i \in \mathcal{T}$ we choose one of its children and recursively visit descendants of that child before visiting any other children of node $i$.

As we now show, *one-by-one exploration* of nodes in a breadth-first manner yields a random walk. This is seen as follows. Construct a queue $Q$ so that for each time $k \geq 0$, we say that a node is *active* if it resides in $Q$. At each time step, we pick any one node in $Q$ for removal (i.e., deactivate it) and add its children to $Q$. Only the root node resides in $Q$ at $k = 0$. Letting $A_k$ be the number of active nodes at time $n$, we find that $\{A_k : k \geq 0\}$ satisfies

$$\begin{aligned} &A_k = A_{k-1} - 1 + \xi_k, \quad k \geq 1 \\ &A_0 = 1 \end{aligned}$$

where $\{\xi_k\}_{k\in\mathbb{N}} \overset{\text{iid}}{\sim} \xi$ are sampled from the offspring distribution. Therefore, the number of active nodes follows a random walk and the total population equals the first time the queue is empty:

$$|\mathcal{T}| = \min\{k \geq 0 : A_k = 0\}.$$

### 3.1.3 Chernoff bounds and population size

Since when $\mu < 1$ we know that the population is finite with probability 1, it is natural to ask if we can estimate the size of the total population. In particular, can we find an upper bound on the probability that the size of the tree at extinction exceeds $k$? Indeed we can, and this bound is exponentially decreasing in $k$ with a rate given in terms of the offspring distribution.

We will prove this result using a concentration inequality known as a *Chernoff bound*. If $X_1, \ldots, X_n$ are iid random variables with the same distribution as a random variable $X$, the law of large numbers guarantees $\lim_{n\to\infty} \mathbb{P}\left(n^{-1}\sum_{i=1}^n X_i \geq \mathbb{E}[X] + \varepsilon\right) = 0$ for all $\varepsilon > 0$. Chernoff bounds allow us to quantify how quickly this probability goes to 0 in terms of a *rate function*

$$h(x) = \sup_{\theta \geq 0}\left\{\theta x - \log \mathbb{E}\left[e^{\theta X}\right]\right\}$$

(for those familiar with more advanced probability, note that this quantity is of prime importance in *large deviation theory*). Before we state and prove this, we first recall the following elementary result:

**Lemma 3.2.** (Markov's inequality) *For a non-negative random variable $X \geq 0$ and $a \geq 0$,*

$$\mathbb{P}(X \geq a) \leq \frac{\mathbb{E}[X]}{a}.$$

**Proof:** Let $I_{\{X \geq a\}}$ be the indicator function that the event $\{X \geq a\}$ occurs. Since $aI_{\{X \geq a\}} \leq X$, taking the expected value of each side implies $a\mathbb{P}(X \geq a) = a\mathbb{E}\left[I_{\{X \geq a\}}\right] \leq \mathbb{E}[X]$. $\square$

**Lemma 3.3.** (Chernoff bound) *If $\{X_i\}_{i \in \mathbb{N}} \overset{iid}{\sim} X$, then for any $a \in \mathbb{R}$ one has*

$$\mathbb{P}\left(\sum_{i=1}^{n} X_i \geq na\right) \leq e^{-nh(a)}.$$

**Proof:** Letting $\theta \geq 0$, we have

$$\mathbb{P}\left(\sum_{i=1}^{n} X_i \geq na\right) = \mathbb{P}\left(e^{\theta \sum_{i=1}^{n} X_i} \geq e^{n\theta a}\right) \leq e^{-n\theta a}\mathbb{E}\left[e^{\theta \sum_{i=1}^{n} X_i}\right]$$

by Markov's inequality (which we can apply since $e^{\theta \sum_{i=1}^{n} X_i}$ is a non-negative random variable). Since

$$\mathbb{E}\left[e^{\theta \sum_{i=1}^{n} X_i}\right] = \mathbb{E}\left[e^{\theta X_1} e^{\theta X_2} \cdots e^{\theta X_n}\right] = \left(\mathbb{E}\left[e^{\theta X}\right]\right)^n = e^{n \log \mathbb{E}\left[e^{\theta X}\right]},$$

substituting back into the last expression implies

$$\mathbb{P}\left(\sum_{i=1}^{n} X_i \geq na\right) \leq \exp\left(-n\left\{\theta a - \log \mathbb{E}\left[e^{\theta X}\right]\right\}\right).$$

The result follows by choosing the value of $\theta \geq 0$ which minimizes the r.h.s. $\square$

The Chernoff bound is only useful for $a$ such that $h(a) > 0$. Assume that there exists $b > 0$ such that $\mathbb{E}[e^{bX}] < \infty$, since otherwise $h(a) = 0$ for all $a$. This also guarantees $g(\theta) = \theta a - \log \mathbb{E}\left[e^{\theta X}\right]$ is finite and differentiable for $0 \leq \theta < b$. Then for $a > \mathbb{E}[X]$, we have $g'(0) > 0$ so $h(a)$ is strictly positive.

Applying the previous result we have the following bound on the total population of a Galton-Watson branching process:

**Theorem 3.4.** *Let* $h(x) = \sup_{\theta \geq 0} \left\{ \theta x - \log \mathbb{E}\left[e^{\theta \xi}\right] \right\}$ *be the rate function associated to the offspring distribution. Then for all* $k \in \mathbb{N}$,

$$\mathbb{P}(|\mathcal{T}| > k) \leq e^{-kh(1)}.$$

This bound is only useful when $h(1) > 0$, which we can guarantee when $\mu < 1$ (i.e., the subcritical case).

**Proof:** Let $\{A_k : k \geq 0\}$ be the one-by-one exploration process of $\mathcal{T}$. Then

$$\begin{aligned}
\mathbb{P}(|\mathcal{T}| > k) &= \mathbb{P}(A_1 > 0, \ldots, A_k > 0) \\
&\leq \mathbb{P}(A_k > 0) \\
&\leq \mathbb{P}\left( A_0 + \sum_{i=1}^{k} \xi_i > k \right) \\
&= \mathbb{P}\left( \sum_{i=1}^{k} \xi_i \geq k \right) \\
&\leq e^{-kh(1)}.
\end{aligned}$$

□

# References

1. Draief, M., & Massoulié, L. (2010). Epidemics and rumours in complex networks. Cambridge University Press.

2. Durrett, R. (2006). Random graph dynamics (Vol. 20). Cambridge University Press.

*Last edited: March 21, 2013*