# 6.7720/18.619/15.070 Lecture 2
# The First & Second Moment Methods: Branching Processes

Kuikui Liu

February 5, 2025

**Acknowledgements & Disclaimers** *Several of the topic choices and lecture notes were partly inspired by materials created by Guy Bresler and David Gamarnik, who taught previous iterations of this course. We are grateful for the materials they shared and for the discussions we had with them. Please be advised that these notes have not been subjected to the usual scrutiny reserved for formal publications. If you do spot an error, please contact the instructor.*

## 1 A Refined Second Moment Method

Chebyshev's Inequality is a useful way to bound the probability that a random variable deviates from its expectation. We used it several times in the previous lecture when analyzing Erdös–Rényi random graphs. However, if our goal is to establish lower bounds on the probability that a random variable is positive (e.g. a certain event occurs or a certain structure exists), the following version of the second moment method tends to be more useful.

**Theorem 1.1** (Paley–Zygmund Inequality; Second Moment Method)**.** *Let $X \geq 0$ be a nonnegative random variable. For any $0 \leq \theta \leq 1$, we have*

$$\Pr\left[X > \theta \cdot \mathbb{E}[X]\right] \geq (1-\theta)^2 \cdot \frac{\mathbb{E}[X]^2}{\mathbb{E}[X^2]}$$

*Proof.* A standard trick for bounding expectations of random variables is to insert indicator functions for when the random variable lies in a certain range. In our setting, we have

$$\begin{aligned}\mathbb{E}[X] &= \mathbb{E}\left[X \cdot \mathbf{1}_{X \leq \theta \cdot \mathbb{E}[X]}\right] + \mathbb{E}\left[X \cdot \mathbf{1}_{X > \theta \cdot \mathbb{E}[X]}\right] \\ &\leq \theta \cdot \mathbb{E}[X] + \sqrt{\mathbb{E}[X^2] \cdot \Pr[X > \theta \cdot \mathbb{E}[X]]}. \qquad \text{(Cauchy–Schwarz)}\end{aligned}$$

Rearranging then completes the proof. $\square$

It is instructive to compare this with Chebyshev's Inequality, which yields

$$\Pr\left[X > \theta \cdot \mathbb{E}[X]\right] \geq 1 - \frac{\mathrm{Var}(X)}{(1-\theta) \cdot \mathbb{E}[X]^2}.$$

This is a vacuous bound if $\mathbb{E}[X^2] > (2-\theta) \cdot \mathbb{E}[X]^2$. The advantage of Theorem 1.1 is that it remains meaningful whenever $\mathbb{E}[X^2] \leq C \cdot \mathbb{E}[X]^2$, where $C \geq 1$ is allowed to be *any* constant.

## 2 Galton–Watson Branching Processes

Let us now introduce a fundamental class of stochastic processes called *branching processes*, which we will analyze using the first and second moment methods. These processes are extremely useful as a way to make predictions about the behavior of more complex stochastic processes, including the local structure of sparse Erdös–Rényi random graphs and various natural statistical inference problems; we'll discuss these in more depth later.

**Definition 1** (Galton–Watson Branching Process)**.** *Let $\xi$ be a probability distribution over $\mathbb{N} = \{0, 1, 2, \dots\}$. The* Galton–Watson Branching Process with offspring distribution $\xi$ *is an infinite sequence $\mathcal{Z}$ of discrete $\mathbb{N}$-valued random variables $Z_0, Z_1, \dots$ generated as follows:*

- *We initialize $Z_0 = 1$ with probability 1.*
- *For each $\ell \in \mathbb{N}$, given $Z_\ell$, we let $Z_{\ell+1}$ be the sum of $Z_\ell$-many independent samples from $\xi$. More precisely, we let $Z_{\ell+1} = \sum_{i=1}^{Z_\ell} X_{\ell+1,i}$ where $X_{\ell+1,1}, \ldots, X_{\ell+1,Z_t} \sim \xi$ are independent.*

One should imagine a Galton–Watson process as generating a (possibly infinite) random tree rooted at a vertex $r$. In each step of the process of generating this tree, we take each of the vertices in the most recently generated level $\ell$ and spawn a random number of children, each independently drawn from $\xi$. This defines the set of vertices in the next level $\ell + 1$ of the tree. Note that for each $\ell$, the random variable $Z_\ell$ counts the number of vertices in level $\ell$.

You can take the offspring distribution $\xi$ to be your favorite probability distribution over $\mathbb{N}$. Some common ones include the following:

- **Uniform:** $\xi = \mathsf{Unif}\{0, \ldots, d\}$, i.e. $\xi(k) = \frac{1}{d+1}$ for $k = 0, \ldots, d$.
- **Poisson:** $\xi = \mathsf{Poi}(\lambda)$, i.e. $\xi(k) = \frac{\lambda^k e^{-\lambda}}{k!}$ for $k \in \mathbb{N}$, where $\lambda \in \mathbb{R}_{\geq 0}$ is the mean of $\xi$.
- **Binomial:** $\xi = \mathsf{Bin}(d, p)$, i.e. $\xi(k) = \binom{d}{k} p^k (1-p)^{d-k}$ for $k = 0, \ldots, d$ (more on this in Section 3).
- **Geometric:** $\xi = \mathsf{Geo}(p)$, i.e. $\xi(k) = (1-p)^k \cdot p$ for $k \in \mathbb{N}$.

Let us begin by computing the means and variances of a Galton–Watson branching process.

**Lemma 2.1.** *Let $\{Z_\ell\}_{\ell \in \mathbb{N}}$ be a Galton–Watson branching process with offspring distribution $\xi$. If $\xi$ has finite expectation $\mu = \mathbb{E}_{X \sim \xi}[X] < \infty$, then,*

$$\mathbb{E}[Z_\ell] = \mu^\ell, \qquad \forall \ell \in \mathbb{N}.$$

*If, in addition, $\xi$ has finite variance $\sigma^2 = \mathbb{E}_{X \sim \xi}\left[(X - \mu)^2\right] < \infty$, then*

$$\mathrm{Var}\left(Z_\ell\right) = \sigma^2 \mu^{\ell-1} \sum_{k=0}^{\ell-1} \mu^k, \qquad \forall \ell \in \mathbb{N}.$$

*Proof.* We prove the claims by induction on $\ell \in \mathbb{N}$; the base cases $\ell = 0, 1$ are immediate. Suppose the claims hold for some $\ell \in \mathbb{N}$. Then

$$\begin{aligned}
\mathbb{E}\left[Z_{\ell+1}\right] &= \sum_{k=0}^{\infty} \Pr[Z_\ell = k] \cdot \mathbb{E}\left[Z_{\ell+1} \mid Z_\ell = k\right] \\
&= \sum_{k=0}^{\infty} \Pr[Z_\ell = k] \cdot k \cdot \mu \\
&= \mu \cdot \mathbb{E}[Z_\ell] \\
&= \mu^{\ell+1}. && \text{(Induction Hypothesis)}
\end{aligned}$$

Similarly,

$$\begin{aligned}
\mathbb{E}\left[Z_{\ell+1}^2\right] &= \sum_{k=0}^{\infty} \Pr[Z_\ell = k] \cdot \mathbb{E}\left[Z_{\ell+1}^2 \mid Z_\ell = k\right] \\
&= \sum_{k=0}^{\infty} \Pr[Z_\ell = k] \cdot \left(k \cdot \mathbb{E}_{X \sim \xi}[X^2] + k(k-1) \cdot \mathbb{E}_{X \sim \xi}[X]^2\right) \\
&= \sigma^2 \cdot \mu^\ell + \mu^2 \mathbb{E}[Z_\ell^2].
\end{aligned}$$

It follows that

$$\begin{aligned}
\mathrm{Var}\left(Z_{\ell+1}\right) &= \mathbb{E}\left[Z_{\ell+1}^2\right] - \mu^{2\ell} \\
&= \sigma^2 \cdot \mu^\ell + \mu^2 \,\mathrm{Var}\left(Z_\ell\right) \\
&= \sigma^2 \cdot \mu^\ell + \sigma^2 \mu^2 \mu^{\ell-1} \sum_{k=0}^{\ell-1} \mu^k && \text{(Induction Hypothesis)} \\
&= \sigma^2 \mu^\ell \sum_{k=0}^{\ell} \mu^k && \text{(Simplifying)}
\end{aligned}$$

as desired. $\square$

## 2.1 The Extinction Phase Transition

If we imagine $Z_\ell$ as the number of individuals in the $\ell$th generation of some population of organisms, then a key event of interest is *extinction.* Note that if $Z_\ell = 0$ for some $\ell \in \mathbb{N}$, then $Z_t = 0$ for all $t \geq \ell$.

**Definition 2** (Extinction Event)**.** *We define the* extinction event *of a branching process as* $\bigcup_{\ell \in \mathbb{N}} E_\ell$*, where* $E_\ell$ *is the event that* $Z_\ell = 0$*.*

**Fact 2.2.** *We always have* $\Pr_{\mathcal{Z}}[\mathsf{Extinction}] \geq \xi(0)$*.*

**Definition 3** (Criticality)**.** *We say a Galton–Watson branching process with offspring distribution* $\xi$ *having finite mean* $\mu = \mathbb{E}_{X\sim\xi}[X] < \infty$ *is* subcritical *if* $\mu < 1$*,* critical *if* $\mu = 1$*, and* supercritical *if* $\mu > 1$*.*

One of the most fundamental results on the behavior of branching processes is the phase transition for the extinction probability. In a sense, it is a shadow of the "primordial" phase transition concerning the behavior of the exponential function $x \mapsto b^x$ as the base $b$ is varied around 1. We begin with the following lemma.

**Lemma 2.3.** *For a probability distribution* $\xi$ *over* $\mathbb{N}$*, define its* generating function $\psi = \psi_\xi$ *by the formal power series*

$$\psi_\xi(s) \stackrel{\mathsf{def}}{=} \mathbb{E}_{X\sim\xi}\left[s^X\right] = \sum_{k=0}^{\infty} \xi(k) \cdot s^k,$$

*which is well-defined as a function over* $[0,1]$ *(i.e. the power series converges), and satisfies* $\psi(1) = 1$*. Then for a Galton–Watson branching process* $\mathcal{Z}$ *with offspring distribution* $\xi$*, the extinction probability* $\Pr_{\mathcal{Z}}[\mathsf{Extinction}]$ *must be a solution to the fixed point equation* $s = \psi_\xi(s)$*.*[1]

*Proof.* One way to think about generating $\mathcal{Z} = \{Z_\ell\}_{\ell \in \mathbb{N}}$ is via recursion:

- First, we set $Z_0 = 1$ and sample $Z_1 \sim \xi$.
- We then independently generate $Z_1$-many Galton–Watson branching processes, i.e. for each $k = 1, \ldots, Z_1$, we independently sample $\mathcal{Z}^{(k)} = \left\{Z_\ell^{(k)}\right\}_{\ell \in \mathbb{N}}$.
- Finally, we set $Z_{\ell+1} = \sum_{k=1}^{Z_1} Z_\ell^{(k)}$ for each $\ell \in \mathbb{N}$.

This self-similarity implies that

$$\begin{aligned}
\Pr\left[\mathcal{Z} \text{ goes extinct}\right] &= \sum_{k=0}^{\infty} \Pr[Z_1 = k] \cdot \Pr\left[\mathcal{Z} \text{ goes extinct} \mid Z_1 = k\right] \\
&= \sum_{k=0}^{\infty} \xi(k) \cdot \Pr\left[\mathcal{Z}^{(1)}, \ldots, \mathcal{Z}^{(k)} \text{ all go extinct}\right] \\
&= \sum_{k=0}^{\infty} \xi(k) \cdot \prod_{i=1}^{k} \Pr\left[\mathcal{Z}^{(i)} \text{ goes extinct}\right], \qquad \text{(Independence)}
\end{aligned}$$

since the only way for $\mathcal{Z}$ to go extinct is if each $\mathcal{Z}^{(k)}$ goes extinct. But since $\mathcal{Z}$ and the $\mathcal{Z}^{(k)}$ all have the same law, the above simplifies to

$$\Pr_{\mathcal{Z}}[\mathsf{Extinction}] = \sum_{k=0}^{\infty} \xi(k) \cdot \Pr_{\mathcal{Z}}[\mathsf{Extinction}]^k.$$

This shows that $\Pr_{\mathcal{Z}}[\mathsf{Extinction}]$ must be a fixed point of $\psi$. □

**Theorem 2.4.** *Let* $\mathcal{Z} = \{Z_\ell\}_{\ell \in \mathbb{N}}$ *be a Galton–Watson branching process with offspring distribution* $\xi$ *having finite mean* $\mu$ *and variance. Let* $\psi$ *denote the generating function of* $\xi$*.*

[1] It is an easy exercise to prove that for every $\ell \in \mathbb{N}$, the generating function of the random variable $Z_\ell$ is precisely the iterated composition $\psi^{\circ \ell}$ of $\psi$.

- *If the process is subcritical (i.e. $\mu < 1$), then $s = 1$ is the unique fixed point of $\psi$, and* $\Pr_{\mathcal{Z}}[\mathsf{Extinction}] = 1$.
- *If the process is supercritical (i.e. $\mu > 1$), then there exists $p^* = p^*(\xi) \in [0, 1)$ such that the set of fixed points of $\psi$ is precisely $\{p^*, 1\}$. Moreover,* $\Pr_{\mathcal{Z}}[\mathsf{Extinction}] = p^*$.
- *If the process is critical (i.e. $\mu = 1$), and $\sigma^2 > 0$, then $s = 1$ is again the unique fixed point of $\psi$, and* $\Pr_{\mathcal{Z}}[\mathsf{Extinction}] = 1$.
- *If the process is critical (i.e. $\mu = 1$), and $\sigma^2 = 0$, then the set of fixed points of $\psi$ is the entire interval $[0, 1]$, and* $\Pr_{\mathcal{Z}}[\mathsf{Extinction}] = 0$.

To classify the fixed points of $\psi$ and prove that the extinction probability is equal to the smallest such fixed point, we crucially take advantage of the following analytic properties of $\psi$, which can be immediately deduced by differentiating $\psi$.

**Lemma 2.5.** *Let $\xi$ be an offspring distribution over $\mathbb{N}$. Then its generating function $\psi$ satisfies the following properties.*

- *$\psi$ is strictly increasing (unless $\xi(0) = 1$, in which case $\psi \equiv 1$).*
- *$\psi$ is convex, i.e. $\psi\left(\frac{s+t}{2}\right) \le \frac{\psi(s)+\psi(t)}{2}$ for all $s, t \in [0, 1]$. Moreover, if $\xi(k) > 0$ for some $k \ge 2$, then this convexity is strict, i.e. the inequality is strict whenever $s \neq t$.*

*Proof of Theorem 2.4.* We establish each case in turn.

- If $\mu < 1$, then since $\psi$ lies above its tangent at 1, we have that $\psi(s) > s$ for all $s \in [0, 1)$. Hence, $s = 1$ is the unique fixed point of $\psi$ and $\Pr_{\mathcal{Z}}[\mathsf{Extinction}] = 1$. Another way to see that the extinction probability must be 1 is by the fact that
$$\sum_{\ell=0}^{\infty} \mathbb{E}[Z_\ell] = \sum_{\ell=0}^{\infty} \mu^\ell = \frac{1}{1-\mu} < \infty,$$
which is impossible if $\Pr[Z_\ell > 0, \forall \ell \in \mathbb{N}] > 0$.
- If $\mu > 1$, then $\psi(s) < s$ for $s$ in a neighborhood of 1. Since $\psi(0) \ge 0$ as well, there must exist $p^* \in [0, 1)$ such that $p^* = \psi(p^*)$ by the Intermediate Value Theorem. This other $p^* \neq 1$ is unique by strict convexity of $\psi$. To show that $\Pr_{\mathcal{Z}}[\mathsf{Extinction}] = p^*$, it suffices to establish $\Pr_{\mathcal{Z}}[\mathsf{Extinction}] < 1$, or equivalently, $\Pr[Z_\ell > 0, \forall \ell \in \mathbb{N}] > 0$. Note that
$$\Pr[Z_\ell > 0, \forall \ell \in \mathbb{N}] = \lim_{L\to\infty} \Pr[Z_\ell > 0, \forall 0 \le \ell \le L] = \lim_{\ell\to\infty} \Pr[Z_\ell > 0].$$
Hence, it suffices to show that there is some constant $C > 0$, possibly depending on $\mu, \sigma^2$, such that $\Pr[Z_\ell > 0] \ge C$ for all $\ell \in \mathbb{N}$. For this, let us use the second moment method. Observe that
$$\begin{aligned}
\Pr[Z_\ell > 0] &\ge \frac{\mathbb{E}[Z_\ell]^2}{\mathbb{E}[Z_\ell^2]} && \text{(Paley–Zygmund Inequality; Theorem 1.1)}\\
&= \frac{\mu^{2\ell}}{\mu^{2\ell} + \sigma^2\mu^{\ell-1}\sum_{k=0}^{\ell-1}\mu^k} && \text{(Lemma 2.1)}\\
&= \frac{1}{1 + \frac{\sigma^2}{\mu-1}\left(\frac{1}{\mu} - \frac{1}{\mu^{\ell+1}}\right)} && \text{(Using } \mu > 1 \text{ and } \textstyle\sum_{k=0}^{\ell-1}\mu^k = \frac{\mu^\ell-1}{\mu-1}\text{)}\\
&\ge \frac{1}{1 + \sigma^2/\mu^2}. && \text{(Monotonicity in } \ell\text{)}
\end{aligned}$$
This establishes that $\Pr[Z_\ell > 0, \forall \ell \in \mathbb{N}] > 0$ and so $\Pr_{\mathcal{Z}}[\mathsf{Extinction}] = p^*$.
- If $\mu = 1$ and $\sigma^2 > 0$, then $\xi(k) > 0$ for some $k \ge 2$. This means $\xi$ is strictly convex, and therefore lies strictly above its tangent line at 1 over $[0, 1)$. This tangent line is precisely the identity function, and so $s = 1$ is again the unique fixed point of $\psi$. This means $\Pr_{\mathcal{Z}}[\mathsf{Extinction}] = 1$.
- If $\mu = 1$ and $\sigma^2 = 0$, then $\psi(s) = s$ and the set of fixed points is all of $[0, 1]$. Furthermore, $\xi(1) = 1$ and so $Z_\ell = 1$ with probability 1 for all $\ell \in \mathbb{N}$. This means $\Pr_{\mathcal{Z}}[\mathsf{Extinction}] = 0$.

$\square$

# 3 Percolation in the $d$-ary Tree

Recall that *(bond) percolation with parameter $p$* on a graph $G = (V, E)$ is given by a random subgraph $H = (V, F)$ where each edge $e \in E$ is included in $F$ independently with probability $p$. The study of percolation originated in mathematical physics, where it was used as a basic mathematical model of porous materials. The idea was to represent a block of material as an infinite graph like the integer lattice $\mathbb{Z}^3$, where each edge allows liquid to "flow through it" independently with some probability $p$. The main question physicists were interested in was whether or not there exists an infinite path permitting water to flow through the entire block of material.

There is an extremely rich history and theory of percolation on all sorts of lattices (and graphs more broadly). In this section, we'll be looking at percolation on the infinite $d$-ary tree, denoted as $\widehat{\mathbb{T}}_d$. This is the infinite tree rooted at a distinguished vertex $r$ where every vertex in the tree has exactly $d$ children. We will be interested in the *percolation event*, that is, the event $E_\infty$ that there exists an infinite component within the randomly generated subgraph; note that we do not require that this component contains the root $r$. Writing $\Pr_p[\cdot]$ for the probability of an event under bond percolation on $\widehat{\mathbb{T}}_d$ with parameter $p$, we will be interested in the critical probability

$$p_{\mathsf{c}}\left(\widehat{\mathbb{T}}_d\right) \stackrel{\mathsf{def}}{=} \sup\left\{p : \Pr_p\left[E_\infty\right] = 0\right\}.$$

From the theory of Galton–Watson branching processes we discussed above, it is natural to guess that $p_{\mathsf{c}}\left(\widehat{\mathbb{T}}_d\right) = 1/d$, since this is the threshold at which the expected number of children a vertex is connected to is equal to 1. Let us now make this precise.

**Theorem 3.1.** *For every $d \in \mathbb{N}$, we have $p_{\mathsf{c}}\left(\widehat{\mathbb{T}}_d\right) = 1/d$.*

*Proof.* For each vertex $v \in \widehat{\mathbb{T}}_d$, let $\mathcal{C}_v$ denote the unique connected component of the sampled subgraph containing $v$, restricted to the subtree of $\widehat{\mathbb{T}}_d$ rooted at $v$. Note that

$$E_\infty = \bigcup_{v \in \widehat{\mathbb{T}}_d} \{\mathcal{C}_v \text{ is infinite}\}.$$

Since the subtree rooted at $v$ is isomorphic to $\widehat{\mathbb{T}}_d$ itself, all the components $\mathcal{C}_v$ have the same law as $\mathcal{C}_r$ itself. Given this, it is enough to study $\mathcal{C}_r$.

Observe that if we let $Z_\ell$ denote the number of vertices which are at distance $\ell$ from the root vertex $r$, then $\{Z_\ell\}_{\ell \in \mathbb{N}}$ is distributed as a Galton–Watson branching process with offspring distribution $\xi$ given by $\xi(k) = \mathsf{Bin}(d, p)$. Moreover, $\mathcal{C}_r$ is infinite if and only if this branching process does not go extinct. Since the mean of this offspring distribution is $\mu = d \cdot p$, if $p > 1/d$, then the branching process is supercritical and there already is strictly positive probability that $\mathcal{C}_r$ is infinite by Theorem 2.4. This tells us that $p_{\mathsf{c}}\left(\widehat{\mathbb{T}}_d\right) \le p$ for every $p > 1/d$, i.e. $p_{\mathsf{c}}\left(\widehat{\mathbb{T}}_d\right) \le 1/d$, since

$$\Pr_p\left[E_\infty\right] \ge \Pr_p[\mathcal{C}_r \text{ is infinite}] > 0, \qquad \forall p > 1/d.$$

On the other hand, if $p < 1/d$, then this branching process is subcritical and $\Pr_p\left[\mathcal{C}_r \text{ is infinite}\right] = 0$ by Theorem 2.4. Hence, by the Union Bound, we have

$$\Pr_p\left[E_\infty\right] \le \sum_{v \in \widehat{\mathbb{T}}_d} \Pr_p\left[\mathcal{C}_v \text{ is infinite}\right] = 0, \qquad \forall p < 1/d.$$

This shows that $p_{\mathsf{c}}\left(\widehat{\mathbb{T}}_d\right) \ge p$ for every $p < 1/d$. In particular, $p_{\mathsf{c}}\left(\widehat{\mathbb{T}}_d\right) \ge 1/d$. □

Note that as a corollary, we get that $\Pr_p[E_\infty]$ exhibits a sharp phase transition at $p_{\mathsf{c}}\left(\widehat{\mathbb{T}}_d\right)$.

**Corollary 3.2.** *For every $d \in \mathbb{N}$, we have that*

$$\Pr_p\left[E_\infty\right] = \begin{cases} 1, & \text{if } p > p_{\mathsf{c}}\left(\widehat{\mathbb{T}}_d\right) \\ 0, & \text{if } p < p_{\mathsf{c}}\left(\widehat{\mathbb{T}}_d\right) \end{cases}.$$

Note that the $p < p_{\mathsf{c}}\left(\widehat{\mathbb{T}}_d\right)$ case just follows from the definition of $p_{\mathsf{c}}\left(\widehat{\mathbb{T}}_d\right)$. In the case $p > p_{\mathsf{c}}\left(\widehat{\mathbb{T}}_d\right)$, the claim just follows from Kolmogorov's zero-one law. This is actually quite intuitive, since the fact that our branching process is supercritical means that there is some positive constant probability that the unique connected component containing the root is itself infinite. On the off chance that it is finite, then it has some maximum depth $L$. But the vertices at depth $L+1$ then each independently spawn their own branching process within their own subtree, each of which again has some positive probability of extending to infinity. This process continues indefinitely and so eventually, at least one of the branching processes must grow an infinite connected component.

# 4 Foreshadowing: Local Structure of Sparse Erdös–Rényi

The branching process perspective is also useful for studying Erdös–Rényi random graphs in the *sparse* regime, where $p_n = \frac{d}{n}$ for constant $d$. This is well below the connectivity threshold, and so a randomly sampled graph is disconnected with high probability. However, in a future lecture, we will prove that another fascinating phase transition occurs at $d = 1$. If $d < 1$, not only is the graph disconnected with high probability, but all its connected components are tiny (of size $O(\log n)$). On the other hand, when $d > 1$, the graph contains a unique connected component of size $\Omega(n)$; this is often referred to as the *giant component* in Erdös–Rényi.

We can actually already give an intuitive explanation for this phenomenon using the theory of branching processes we've developed. Let us examine the size of the connected component containing some arbitrarily fixed vertex $v \in V$. We can do this by performing a breadth-first search in the graph and revealing the randomly chosen edges on an as-needed basis. For $\ell \in \mathbb{N}$, let $N_\ell$ denote the number of vertices at distance exactly $\ell$ away from $v$. Then the size of the connected component containing $v$ is $\sum_{\ell=0}^{n} N_\ell$. Moreover, $N_\ell = 0$ implies $N_j = 0$ for all $j \geq \ell$.

- The size $N_1$ of the depth-1 neighborhood is a random variable drawn from $\mathsf{Bin}(n-1, p_n) \approx \mathsf{Bin}(n, d/n)$.
- Given $N_1$, the number of vertices $N_2$ at distance 2 away from $v$ is drawn from $\mathsf{Bin}(n - N_1 - 1, p_n)$, which is again approximately $\mathsf{Bin}(n, d/n)$ since $N_1 \leq o(1)$ with very high probability.
- These approximations hold while the number of visited vertices in the breadth-first search is $o(n)$. We expect these balls to grow at most exponentially fast in the radius, so we can expect to continue this analysis for $o(\log n)$ steps.

Now, it is well-known that the binomial $\mathsf{Bin}(n, d/n)$ is extremely well-approximated by the Poisson distribution $\mathsf{Poi}(d)$ with the same expectation in the sparse regime $d = O(1)$;[2] we will prove this in a future lecture. Hence, at least up to $\ell \leq o(\log n)$, we expect $N_0, \ldots, N_\ell$ to evolve as a Galton–Watson branching process with offspring distribution $\mathsf{Poi}(d)$. If $d < 1$, the process is subcritical and the expected size of the component is $\frac{1}{1-d}$. We also expect the branching process to die off quickly. If $d > 1$, the process is supercritical and the process will survive until $\Omega(n)$ vertices have been included in the component. We will prove these claims rigorously in a future lecture.

[2]The Central Limit Theorem applies when $d/n$ is a constant independent of $n$.