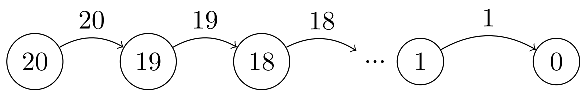
EE 126 **Probability and Random Processes**

March 20, 2018 **Random Graphs**

1 Lecture

CTMCs

Example: lightbulbs. We start with 20 lightbulbs, each of which burns out in an exponentially distributed time. We are interested in the expected time for all 20 bulbs to burn out. To solve this problem, we should draw a CTMC as follows:

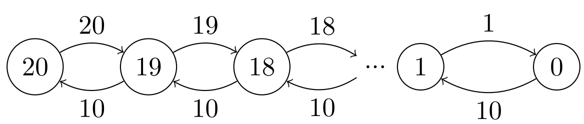


The first step equations are

$$\beta(m) = \frac{1}{m} + \beta(m-1) \quad \text{for } m = 1, ..., 20$$
$$\beta(0) = 0$$

Solving, we find that $\beta(20) = \frac{1}{20} + ... + \frac{1}{1} \approx \ln(20)$.

Example: repair. We can now replace burnt-out bulbs after an Exp(10) time (due to having to wait for a lightbulb repairman to arrive). Accordingly, the mean repair time is 0.1. What, now, is the expected time for all bulbs to burn out?



Note: the rate at which we escape state 19 is $10 + 19 = 29$. So the mean rate of escape is $1/(m + 10)$.
The first step equations are

$$\beta(20) = \frac{1}{20} + \beta(m-1)$$
$$\beta(m) = \frac{1}{m+10} + \left(\frac{m}{m+10}\right)\beta(m-1) + \left(\frac{10}{m+10}\right)\beta(m+1) \quad \text{for } m = 1, ..., 19$$
$$\beta(0) = 0$$

In general, if we have $T_E = \min\{t \geq 0 \mid X_t = E\}$ then

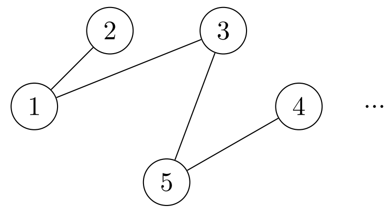
$$\beta_E(i) = \mathbb{E}[T_E \mid X_0 = i]$$

which are given by the first step equations

$$\beta_E(i) = \frac{1}{q(i)} + \sum_{j \in \mathcal{X}, j \neq i} \left[\frac{Q(i,j)}{q(i)}\right] \beta_E(j)$$

Random Graphs

Among other things, random graphs have applications in social networks, biological networks, matrix completion (e.g. Netflix challenge), and bitcoin/blockchain. The model, introduced by Erdos and Renyi, involves an undirected graph $G(n,p)$ of n vertices, where an edge between any two nodes is formed with probability p.



Basically, we flip a coin for every potential edge with heads probability p, and connect the edge if we get a heads. This process continues independently for all $\binom{n}{2}$ potential edges. Then the **Erdos-Renyi graph** is an undirected graph on n vertices such that each of the $\binom{n}{2}$ edges is present with probability p.

If $p = 0$, then $G(n,0)$ is a graph with n nodes and no edges. If $p = 1$, $G(n,1)$ is a fully connected graph.

Question: what is the expected number of edges in $G(n,p)$? Using linearity of expectation,

$$\mathbb{E}[\text{number of edges}] = \binom{n}{2}p = \frac{n(n-1)}{2}p$$

Question: what is the distribution of D, the degree of a node? What is $\mathbb{E}[D]$?

$$\begin{aligned} D &\sim \text{Bin}(n-1,p) \\ P_D(d) &= \binom{n-1}{d}p^d(1-p)^{n-d} \quad \text{for } d = 0, ..., n-1 \\ \mathbb{E}[D] &= (n-1)p \end{aligned}$$

Question: if $n \to \infty$ and $p \to 0$, while $(n-1)p = \lambda$ is a constant, how do we approximate the distribution $P_D(d)$?

$$D \sim \text{Poisson}(\lambda), \text{ and } P_D(d) = e^{-\lambda}\frac{\lambda^d}{d!}, \text{ for } d = 0, 1, ...$$

Question: what is the probability q that a node is isolated?

$$q = (1-p)^{n-1}$$

Erdos and Renyi stated a number of results that are based on the "thresholds" of p needed for certain *structural properties* to emerge. All of these properties depend on a cryptical value of p (e.g. if we made p a little lower, we would no longer have this result).

- When $p = {}^1/_{n^2}$, for example, the "first edge" appears.
- When $p = {}^1/_n$, a "giant component" emerges. We will get "puddles" (connected components) all over our graph, each with size $O(\log n)$. As the ratio $p = {}^{(1+\epsilon)}/_n$ increases, the puddles will gradually coalesce.

Today, we will focus on perhaps the most important property: *connectivity*. We will examine the threshold for connectivity. (Recall that connectivity means there's always a path from any node to any other node.)

If $p = {}^{(1-\epsilon)\log n}/_n$, the graph is (probably) not connected. If $p = {}^{(1+\epsilon)\log n}/_n$, the graph is (probably) connected. $p = {}^{\log n}/_n$ is a *sharp* threshold for connectivity (right when we pass it, we have connectivity); this is known as the *phase-transition phenomenon*.

Theorem. Let $p(n) = {}^{\lambda \log n}/_n$. Then

- If $\lambda < 1$, $P(G(n,p) \text{ is connected}) \xrightarrow{n\to\infty} 0$.
- If $\lambda > 1$, $P(G(n,p) \text{ is connected}) \xrightarrow{n\to\infty} 1$.

Proof of the first case. We want to prove that if $\lambda < 1$, then the graph is not connected. Instead, we will prove a stronger statement: at the threshold ${}^{(1-\epsilon)\log n}/_{n}$, **the probability that there are no isolated nodes goes to** 0.

Let $X =$ the number of isolated nodes. Let's find $\mathbb{E}[X]$. Call I_i the indicator variable for the event that node i is isolated. Then

$$\mathbb{E}[X] = \mathbb{E}\left[\sum_{i=1}^{n} I_i\right] = \sum_{i=1}^{n} P(\text{node } i \text{ is isolated}) = n(1-p)^{n-1}$$

Since we set p to be ${}^{\lambda \log n}/_{n}$,

$$\begin{aligned}
\mathbb{E}[X] &= n(1-p)^{n-1} \\
&= ne^{(n-1)\ln(1-p)} \\
&= ne^{(n-1)(-p+o(p))} \\
&\approx ne^{-np} \\
&= ne^{-\lambda \log n} \\
&= n^{1-\lambda}
\end{aligned}$$

Note that for $\lambda < 1$, $\mathbb{E}[X] \to \infty$. So on average we have an infinite number of isolated nodes. However, even if our mean blows up, it is not necessarily the case that we have no mass at zero. *The variance matters too.* So we are not yet done. We need a lemma.

Lemma. If X is a nonnegative integer-valued random variable, then

$$P(X = 0) \leq \frac{var(X)}{\mathbb{E}[X]^2}$$

which can be proved using Chebyshev:

$$P(X = 0) \leq P(|X - \mathbb{E}[X]| \geq \mathbb{E}[X]) \leq \frac{var(X)}{\mathbb{E}[X]^2}$$

We need to show that the variance goes to 0 faster than $\mathbb{E}[X]^2$, i.e. $var(X) < \mathbb{E}[X]^2$.

$$var(X) = var\left(\sum_{i=1}^{n} I_i\right) = \sum_{i=1}^{n} var(I_i) + \sum_{i=1}^{n} \sum_{j=1, j\neq i}^{n} cov(I_i, I_j) = n \cdot var(I_1) + n(n-1)cov(I_1, I_2)$$

Note: the I_is are not independent. If we are told that $i-1$ of i nodes are isolated, we know something about the isolation of the last node. So we cannot simply do "sum of individual variances."

$var(I_1) = nq(1-q)$, and

$$\begin{aligned}
cov(I_1, I_2) &= \mathbb{E}[I_1, I_2] - \mathbb{E}[I_1]\mathbb{E}[I_2] \\
&= P(1 \text{ and } 2 \text{ are isolated}) - q^2 \\
&= (1-p)^{n-1}(1-p)^{n-2} - q^2 \\
&= \frac{q^2}{1-p} - q^2
\end{aligned}$$

Using this, we can show that $var(X)/\mathbb{E}[X]^2 \to 0$ as $n \to \infty$, and so we conclude our argument for the first case. *The probability of "no isolated node" (/the graph being connected) goes to* 0 *when* $\lambda < 1$.