Sid Banerjee (sbanerjee@cornell.edu)

Problem 1: (Distributed Resource Allocation (ALOHA!))

(Adapted from M& U, Problem 5.11) In this problem, we study a simple distributed protocol for allocating agents to shared resources, wherein agents contend for resources but 'back off' in the face of contention. One example application is in the design of *multiple-access protocols* in wireless networks, where this scheme forms the basis of the ALOHA protocol and its variants.

Consider the following simplified model of a cell-phone network: the wireless spectrum is divided into a set of n channels, where every cell-phone can access every channel, but each channel can support at most one cell-phone conversation. Now suppose n users are trying to simultaneously download a file using their cell-phones – we want a way to match them to channels in a distributed manner.

Part (a)

We start with n users and n channels, and assume the system evolves over rounds. In every round, users pick from the n channels independently and uniformly at random. A user who picks a channel not picked by any other user, is served in that round (i.e., finishes downloading the file) and removed from consideration. The remaining users again pick one of the n channels in the next round. We finish when every user is served.

If there are m users at the start of a round. what is the expected number of users at the start of the next round?

Solution: If there are m users at the start of the round, the probability that a particular user is served by itself is $(1-1/n)^{m-1}$. By linearity of expectation, the expected number of users that get served is $m(1-1/n)^{m-1}$, and the expected number of those that remain is $m - m(1-1/n)^{m-1}$.

Part (b)

Suppose that in every round, the number of users removed is exactly the expected number of users rounded up to the nearest integer. Show that all the users are served in $O(\log\log n)$ rounds.
Hint: If x_j is the expected number of users left after j rounds, show that $x_{j+1} \leq \frac{(x_j)^2}{n}$. Now define $y_j = \log x_j$, and solve the iteration.

Solution: We use part a) together with the assumption that in each round the expected number of users are served. Also, notice that $(1-x)^k \geq 1 - kx$ for $k \geq 0$. Therefore,

$$x_{j+1} = x_j(1-(1-1/n)^{x_j-1}) \leq x_j(1-(1-(x_j-1)/n)) = x_j(x_j-1)/n \leq x_j^2/n.$$

This is equivalent to write:

$$\frac{x_{j+1}}{n} \leq \left(\frac{x_j}{n}\right)^2.$$

Now, let $y_j = \log(x_j/n)$. Then, by recursion we get:

$$y_{j+1} \leq 2^j y_1.$$

And therefore,

$$\frac{x_{j+1}}{n} \leq \left(\frac{x_1}{n}\right)^{2^j} \Rightarrow x_{j+1} \leq n\left(\frac{x_1}{n}\right)^{2^j}.$$

Now, notice that $x_1/n = 1/\alpha = const$, where $\alpha \geq 1$. Hence,

$$x_{j+1} \leq n/\alpha^{2^j}.$$

This value is 1, when $j = O(\log\log n)$.

Problem 2: (The SIS Epidemic on Random Graphs)

Consider an SIS epidemic on an undirected graph G, where neighboring nodes meet at rate λ. In class, we showed the following thresholds for a long infection lifetime:

$$\lambda < \frac{1}{d_{\max}} \Rightarrow \mathbf{E}[T_{ext}] = O(1) \quad , \quad \lambda > \frac{1}{\eta(k)} \Rightarrow \mathbf{E}[T_{ext}] = \Omega\left(e^k\right)$$

where $d_{\max}$ is the maximum degree of nodes in G, and $\eta(k)$ is the generalized isoperimetric constant, which for any $k \leq n$ is defined as:

$$\eta(k) = \min_{S \subset V, |S| \leq k} \frac{|E(S, \overline{S})|}{|S|}$$

(where $|E(S,\overline{S})|$ is the number of edges between S and $\overline{S} = V \setminus S$).

We now study these thresholds in the case where G is generated as a $G(n,p)$ random graph in the connected regime (i.e., with average degree $(n-1)p > \ln n$) – in particular, we want to show that the phase transition threshold is $\lambda = \Theta(1/(n-1)p)$ (w.h.p, as $n \to \infty$).

Part (a)

Consider a graph G generated from the $G(n,p)$ model with $p = 6\ln n/(n-1)$, and let $d = (n-1)p$ and $D_{\max}$ be the average and maximum degrees respectively. Prove that:

$$\mathbf{P}\left[D_{\max} > 2d\right] \leq \frac{1}{n}$$

Thus, if we choose $\lambda < 1/2d$, then, with high probability, an SIS epidemic on G has fast extinction.
Hint: For $X = \sum X_i$ with $X_i \in [0,1]$ i.i.d, recall the Chernoff bounds on the upper and lower tails:

$$\mathbf{P}[X > (1+\epsilon)\mathbf{E}[X]] < e^{-\epsilon^2 \mathbf{E}[X]/3} \quad , \quad \mathbf{P}[X < (1-\epsilon)\mathbf{E}[X]] < e^{-\epsilon^2 \mathbf{E}[X]/2}$$

Solution: For any node $v \in V$, let D_v denote the node's degree. Then we have that $\mathbf{E}[D_v] = (n-1)p = d$. Moreover, by the Chernoff bound for the upper tail, we have (using $\epsilon = 1$):

$$\mathbf{P}[D_v > 2d] < e^{-d/3} = e^{-2\ln n} = \frac{1}{n^2}$$

Now using the union bound, we get:

$$\begin{aligned}\mathbf{P}[D_{\max} > 2d] &= \mathbf{P}[\cup_{v\in V}\{D_v > 2d\}] \\ &\leq \sum_{v\in V}\mathbf{P}[D_v > 2d] \\ &\leq n\cdot\frac{1}{n^2} = \frac{1}{n}\end{aligned}$$

Part (b)

We now want to show a similar threshold for λ to ensure a long-lasting epidemic. As before, consider a graph G generated from the $G(n,p)$ model, and average degree $d = (n-1)p$. For any $k \leq n$, and any ϵ, argue that:

$$\mathbf{P}[\eta(k) < d/3] = \sum_{i=1}^{k}\binom{n}{i}\mathbf{P}\left[\delta_i < d\cdot i/3\right]$$

where $\delta_i \sim Binomial(i(n-i), p)$.
Hint: For any set S with $|S| = i$, what is the distribution of the cut-size $|E(S,\overline{S})|$?

Solution: From the definition of $\eta(k)$, and using the union bound, we have:

$$\begin{aligned}\mathbf{P}[\eta(k) > d/3] &= \mathbf{P}\left[\bigcup_{S\subset V,|S|\leq k}\left\{\frac{|E(S,\overline{S})|}{|S|} > \frac{d}{3}\right\}\right] \\ &\leq \sum_{S\subset V,|S|\leq k}\mathbf{P}\left[\frac{|E(S,\overline{S})|}{|S|} > \frac{d}{3}\right]\end{aligned}$$

Moreover, from the symmetry of the $G(n,p)$ model, we have that for all sets S of the same size, the probability that $\frac{|E(S,\overline{S})|}{|S|} > \frac{d}{3}$ is the same – moreover, we also have that for any set S of size $k \leq n$, $E(S,\overline{S}) = Binomial(k(n-k), p)$ (since each potential edge between a node in S and a node in $\overline{S}$ exists independently with probability p). Hence, we can rewrite the above RHS as:

$$\mathbf{P}[\eta(k) < d/3] = \sum_{i=1}^{k}\binom{n}{i}\mathbf{P}\left[\delta_i < d\cdot i/3\right]$$

Part (c)

Let $k = n/3$, and define $\epsilon_i = 1 - \frac{(n-1)}{3(n-i)}$ – argue that for any $n > 1$ and $i \leq n/3$, we have that $\epsilon_i \in (1/2, 1)$. Now, using the above equation and a Chernoff bound, show that:

$$\mathbf{P}[\eta(n/3) < d/3] \leq \sum_{i=1}^{n/3} \binom{n}{i} e^{-i(n-i)p\epsilon_i^2/2} \leq \sum_{i=1}^{n/3} \binom{n}{i} e^{-i(n-1)p/12}$$

Note: The problem has an 8 instead of 12 in the denominator of the exponent - although tighter bounds are possible, I had intended this solution.

Solution: First, for $\delta_i \sim Binomial(i(n-i), p)$, note that $\mathbf{E}[\delta_i] = i(n-i)p$, and the Chernoff bound for the lower tail gives:

$$\mathbf{P}\left[\delta_i < d \cdot i/3\right] = \mathbf{P}\left[\delta_i < i(n-i)p\left(\frac{n-1}{3(n-i)}\right)\right] = \mathbf{P}\left[\delta_i < \mathbf{E}[\delta_i]\left(1 - \epsilon_i\right)\right] \leq e^{-\frac{i(n-i)p\epsilon_i^2}{2}}$$

To further upper bound the expression, we note that since $i \leq n/3$, then $(n-i)/(n-1) \geq 2/3$ and $\epsilon_i = 1 - \frac{(n-1)}{3(n-i)} \geq \frac{1}{2}$. Thus, we get:

$$\mathbf{P}[\eta(n/3) < d/3] \leq \sum_{i=1}^{n/3} \binom{n}{i} e^{-\frac{i(n-i)p\epsilon_i^2}{2}} \leq \sum_{i=1}^{n/3} \binom{n}{i} e^{-\frac{i(n-1)p}{12}}$$

Part (d) (Optional)

Now suppose we choose $p = 24 \ln n/(n-1)$. Using the inequality $\binom{n}{i} \leq \frac{n^i}{i!}$, show that:

$$\mathbf{P}[\eta(n/3) < d/3] \leq \sum_{i=1}^{n/3} \frac{n^{-i}}{i!} \leq e^{1/n} - 1$$

Thus, argue that choosing $\lambda > 3/d$ ensures that as $n \to \infty$, with probability going to 1, the SIS epidemic has $\mathbf{E}[T_{ext}] = \Omega(e^{n/3})$.

Solution: Substituting p in the inequality in the previous part, along with the inequality given, we get:

$$\begin{aligned}\mathbf{P}[\eta(n/3) < d/3] &\leq \sum_{i=1}^{n/3} \binom{n}{i} e^{-\frac{24i \ln n}{12}} \leq \sum_{i=1}^{n/3} \frac{n^i}{i!} n^{-2i} \\ &\leq \sum_{i=1}^{n/3} \frac{n^{-i}}{i!} \leq e^{1/n} - 1\end{aligned}$$

Thus, as $n \to \infty$, we have that $\mathbf{P}[\eta(n/3) < d/3] \to 0$ – in other words, with probability tending to 1, we have that $\eta(n/3) > d/3$. Thus, choosing $\lambda \geq 3/d$ ensures that $\eta(n/3)\lambda > 1$ w.h.p., and hence $\mathbf{E}[T_{ext}] = \Omega(e^{n/3})$.

Problem 3: (Differential Equation Approximations of CTMCs)

In class, for any CTMC $X(t)$, we defined the rate matrix A (where for $i \neq j$, $A_{ij} = \lambda_{ij}$, the rate of transitions from state i to state j; moreover $A_{ii} = -\sum_j A_{ij}$), and transition matrix $P(t)$ (where $P_{i,j}(t) = \mathbf{P}[X(t) = j | X(0) = i]]$). We also discussed that these matrices satisfy Kolmogorov's forward equation: $\frac{dP(t)}{dt} = P(t)A$. We now formally show one example of how we can use this to get exact differential equations for $\mathbf{E}[X(t)]$.
(Warning: This is a somewhat roundabout way of deriving the differential equation for $\mathbf{E}[X(t)]$, but it will give you a feel for the more general theory behind CTMC and differential equations.)

Part (a)

Consider the simple *birth process* of cellular growth: we start with $X(0) = 1$ cell, and each cell splits into two at rate λ (i.e., after Exponential(λ) time). Let $X(t)$ be the number of cells at time t – clearly this is a CTMC, taking values in $\mathbb{N} = \{1, 2, \ldots\}$. Write down the rate matrix A and the state transition diagram for $X(t)$.
Note that the state-transition diagram and rate matrix are infinite – we did not explicitly deal with this in class, but most things we discussed extend to this case. For this question, you need to specify the state-transitions and rate matrix for any single state i.

Solution: $X(t)$ is a simple birth-death chain, with death rate 0 for all states, and birth rate of λi for state i. The rate matrix A has the form:

$$\begin{aligned} A_{i,i+1} &= \lambda i \quad \forall i \in \{1, 2, \ldots\} \\ A_{ii} &= -\lambda i \quad \forall i \in \{1, 2, \ldots\} \\ A_{i,j} &= 0 \quad \forall i \in \{1, 2, \ldots\}, j \notin \{i, i+1\} \end{aligned}$$

Part (b)

Let $\pi(t)$ be the distribution of the number of cells $X(t)$ at time t. Since $X(0) = 1$, this means $\pi_1(0) = 1, \pi_k(0) = 0$ for $k > 0$. Moreover, we have $\pi(t) = \pi(0)P(t)$. Now argue that for any $k \in \mathbb{N}$:

$$\frac{d\pi_k(t)}{dt} = \left[\pi(0)\frac{dP(t)}{dt}\right]_k = -k\lambda\pi_k(t) + (k-1)\lambda\pi_{k-1}(t),$$

Solve this equation explicitly for $\pi_1(t)$ – does the answer make sense?
Hint: Note that for any $n \times 1$ row vector π and $n \times n$ matrix P, we have $[\pi P]_k = \sum_j \pi_j P_{jk}$ – for an infinite vector $\{\pi_k\}_{k \in \mathbb{N}}$ and matrix $\{P_{ij}\}_{i,j \in \mathbb{N}}$, we can define multiplication in a similar manner.

Solution: First note because $\pi_1(0) = 1, \pi_k(0) = 0$ for $k > 0$, then we have

$$\pi(t) = \pi(0)P(t) \Rightarrow \pi_k(t) = [\pi(0)P(t)]_k = \sum_j \pi_j(0)P_{jk}(t) = P_{1k}(t) \Rightarrow \pi_k(t) = P_{1k}(t).$$

Now, for any $k \in \mathbb{N}$,

$$\frac{d\pi_k(t)}{dt} = \left[\pi(0)\frac{dP(t)}{dt}\right]_k = \sum_j \pi_j(0)[P(t)A]_{jk} = [P(t)A]_{1k} = \sum_i P_{1i}(t)A_{ik} = \sum_i \pi_i(t)A_{ik} =$$
$$= A_{kk}\pi_k(t) + A_{(k-1)k}\pi_{k-1}(t) = -k\lambda\pi_k(t) + (k-1)\lambda\pi_{k-1}(t).$$

Part (c)

For any random variable X, recall we defined the *generating function* $\phi_X(s) = \mathbf{E}[s^X]$. Let $F(s,t) = \phi_{X(t)}(s)$. Argue that $F(s,t) = \sum_{k\in\mathbb{N}} \pi_k(t)s^k$, and also that $\mathbf{E}[X(t)] = \frac{\partial F(s,t)}{\partial s}\Big|_{s=1}$ (i.e., evaluate the partial derivative $\frac{\partial F(s,t)}{\partial s}$ at $s = 1$).

Solution: First, note that given $X(0) = 1$, we can use the fact from part b) that $\pi_k(t) = P_{1,k}(t) = \mathbf{P}[X(t) = k|X(0) = 1|] = \mathbf{P}(X(t) = k)$. Therefore,

$$F(s,t) = \phi_{X(t)}(s) = \mathbf{E}[s^{X(t)}] = \sum_{k\in\mathbb{N}} \mathbf{P}(X(t) = k)s^k = \sum_{k\in\mathbb{N}} \pi_k(t)s^k.$$

Now, it's easy to see

$$\frac{\partial F(s,t)}{\partial s}\bigg|_{s=1} = \sum_{k\in\mathbb{N}} \pi_k(t)ks^{k-1}\bigg|_{s=1} = \sum_{k\in\mathbb{N}} k\pi_k(t) = \sum_{k\in\mathbb{N}} k\mathbf{P}(X(t) = k) = \mathbf{E}[X(t)].$$

Part (d)

Next, prove that:

$$\frac{\partial F(s,t)}{\partial t} = \lambda(s^2 - s)\frac{\partial F(s,t)}{\partial s}$$

Hint: Write out both $\frac{\partial F(s,t)}{\partial t}$ and $\frac{\partial F(s,t)}{\partial s}$ by differentiating each term in the summation, and compare coefficients of s^k. You do not need to justify the correctness of operations on infinite sums.

Solution: From part c), we have

$$\frac{\partial F(s,t)}{\partial s} = \sum_{k\in\mathbb{N}} \pi_k(t)ks^{k-1}.$$

Same way, using the result from part b), we can get:

$$\frac{\partial F(s,t)}{\partial t} = \sum_{k\in\mathbb{N}} \frac{d\pi_k(t)}{dt}s^k = \sum_{k\in\mathbb{N}} [-k\lambda\pi_k(t) + (k-1)\lambda\pi_{k-1}(t)]s^k =$$
$$= \lambda\left[-s\sum_{k\in\mathbb{N}} \pi_k(t)ks^{k-1} + s^2\sum_{k\in\mathbb{N}} \pi_{k-1}(t)(k-1)s^{k-2}\right].$$

Now we just need to compare the coefficients of s^k, and we get the desired result.

Part (e)

Finally, use the fact that $\frac{\partial^2 F(s,t)}{\partial s \partial t} = \frac{\partial^2 F(s,t)}{\partial t \partial s}$ (i.e., for the function $F(s,t)$, you can take partial derivatives in any order – note that is not true in general!), to show that:

$$\frac{d\mathbf{E}[X(t)]}{dt} = \lambda \mathbf{E}[X(t)]$$

Hint: Differentiate the above equation w.r.t. s, and set $s = 1$.

Solution: Using the result of part d), we can write

$$\mathbf{E}[X(t)] = \left.\frac{\partial F(s,t)}{\partial s}\right|_{s=1} \Rightarrow \frac{d\mathbf{E}[X(t)]}{dt} = \left.\frac{\partial^2 F(s,t)}{\partial t \partial s}\right|_{s=1} = \frac{\partial}{\partial s}\left.\left(\lambda(s^2 - s)\frac{\partial F(s,t)}{\partial s}\right)\right|_{s=1} =$$
$$= \left.\left(\lambda(2s-1)\frac{\partial F(s,t)}{\partial s}\right)\right|_{s=1} + \left.\left(\lambda(s^2 - s)\frac{\partial^2 F(s,t)}{\partial s^2}\right)\right|_{s=1} = \lambda \left.\frac{\partial F(s,t)}{\partial s}\right|_{s=1} = \lambda \mathbf{E}[X(t)].$$

So,

$$\frac{d\mathbf{E}[X(t)]}{dt} = \lambda \mathbf{E}[X(t)].$$

Problem 4: (Designing a Viral Advertising Campaign)

The SIS epidemic can be used as a model for brand recall amongst consumers. Suppose we have a population of n people, wherein each peron knows exactly d other people (i.e., the social network is d-regular), and pairs who know each other meet at rate λ. At time 0, n_0 people are aware a certain brand – for example, they are using a certain new app. Subsequently, people who are using the app forget about it at rate 1, while people who have forgotten about it are reminded of it when they meet someone who currently is using it.

Part (a)

Let $X(t)$ be the number of people using the app at time t – in class we discussed that $X(t)$ can be stochastically dominated from above by a birth-death Markov chain $Y(t)$ (i.e., $Y(t) \geq X(t)$ at all t) with the following transitions – $Y(t)$ goes from $n \to n-1$ at rate n, and for all other states $Y(t) \in \{1, \dots, n-1\}$, we have:

$$Y(t) \to Y(t) - 1 \text{ at rate } Y(t) \quad , \quad Y(t) \to Y(t) + 1 \text{ at rate } Y(t) \cdot \lambda \cdot d \tag{1}$$

Note in particular that under the above transitions, $Y(t) = 0$ is an absorbing state. Now we use the differential equation method to find the phase transition for the SIS process. Suppose we ignore the boundary condition at n, and instead let the transitions in Equation 1 hold for all $Y(t) > 0$ (note that this maintains the stochastic dominance). Argue that:

$$\frac{d\mathbf{E}[Y(t)]}{dt} = (\lambda d - 1)\mathbf{E}[Y(t)],$$

and hence $\mathbf{E}[Y(t)] = Y(0)e^{(\lambda d-1)t}$. From this we see that if $\lambda d < 1$, then all people stop using the app after $O(\log n)$ time with high probability.
Note: You don't have to formally prove the last point – however, convince yourself that this is easy to show using Markov's inequality.

Solution: As we discussed in class, observe that when i people are infected, then the number of edges between infected and non-infected nodes is at most id. Thus, in this state, the rate at which the number of infected people increase is at most $i\lambda d$, while the rate at which it decreases is i. This corresponds to the birth-death chain $Y(t)$ as described above – moreover, by the stochastic dominance property, we have that $\mathbf{P}[X(t) > k] \leq \mathbf{P}[Y(t) > k]\,\forall\, k$. Finally, based on its transitions, we have that $\mathbf{E}[Y(t)]$ obeys the differential equation:

$$\frac{d\mathbf{E}[Y(t)]}{dt} = (\lambda d - 1)\mathbf{E}[Y(t)],$$

which we can solve to get $\mathbf{E}[Y(t)] = Y(0)e^{(\lambda d-1)t}$. Now by Markov's inequality, we have:

$$\mathbf{P}[X(t) > 0] \leq \mathbf{P}[Y(t) \geq 1] \leq \mathbf{E}[Y(t)] = Y(0)e^{(\lambda d-1)t}$$

Note that $Y(0) = n_0 \leq n$. Now if $\lambda d = 1 - \epsilon < 1$, then for $t > 2\log n/\epsilon$, we have $\mathbf{P}[X(t) > 0] \leq n_0 n^{-2} \leq n^{-1}$.

Part (b)

Suppose now we have some budget to design a *viral advertising campaign*. To do so, we hire k people (for some fixed k, which is much smaller than n), who each make visits to random people at rate 1, and remind them to use the app. Argue that $X(t)$ is now stochastically dominated from above by a Markov chain $Y(t)$ with transitions

$$Y(t) \to Y(t) - 1 \text{ at rate } Y(t) \quad , \quad Y(t) \to Y(t) + 1 \text{ at rate } Y(t) \cdot \lambda \cdot d + k$$

Note though at $Y(t) = 0$, the above transition is not true, as we want it to be absorbing. However, we ignore this to get the following differential equation approximation for the expected trajectory:

$$\frac{d\mathbf{E}[Y(t)]}{dt} = (\lambda d - 1)\mathbf{E}[Y(t)] + k$$

Show that the solution to this equation is:

$$\mathbf{E}[Y(t)] = \left(\mathbf{E}[Y(0)] + \frac{k}{\lambda d - 1}\right)e^{(\lambda d-1)t} - \frac{k}{\lambda d - 1}$$

Using this solution, observe that for any fixed k, if $\lambda d < 1$, then the epidemic lifetime is still small. This suggests that *using a fixed number of marketing agents does not help make a brand go viral.*

Solution: In this case, when the number of infected nodes (people using the app) is i, then the rate at which the number of infected nodes decreases remains the same as before (i.e., at rate i). However, the rate at which people start using the app is now at most $\lambda id + k$ – the first term, as in the previous case, is due to spreading between infected and susceptible neighbors, while the second captures the fact that the k advertising agents are trying to convert people at rate 1. Thus, we now have that the dominating chain obeys:

$$\frac{d\mathbf{E}[Y(t)]}{dt} = (\lambda d - 1)\mathbf{E}[Y(t)] + k$$

One way to solve this is to take the given solution and verify it satisfies this equation. Alternately, we can multiply both sides by $e^{(-\lambda d-1)t}$, to get:

$$\begin{aligned}
&e^{-(\lambda d-1)t}\frac{d\mathbf{E}[Y(t)]}{dt} - (\lambda d-1)e^{-(\lambda d-1)t}\mathbf{E}[Y(t)] = e^{-(\lambda d-1)t}k \\
\Rightarrow &\frac{d(e^{-(\lambda d-1)t}\mathbf{E}[Y(t)])}{dt} = e^{-(\lambda d-1)t}k \\
\Rightarrow &\mathbf{E}[Y(t)]) = \frac{-k}{\lambda d-1} + ce^{(\lambda d-1)t}
\end{aligned}$$

Moreover, since $\mathbf{E}[Y(0)] = Y(0)$, we get:

$$\mathbf{E}[Y(t)] = \left(\mathbf{E}[Y(0)] + \frac{k}{\lambda d-1}\right)e^{(\lambda d-1)t} - \frac{k}{\lambda d-1}$$

Note that if $\lambda d = 1 - \epsilon$, then again we can use Markov's inequality to argue as before that after $t = O(\log n)$, we have $X(t) = 0$ w.h.p.

Part (c)

Suppose instead we design the following *referral scheme* – each user who is currently using the app is offered, at rate α, a reward for sending a promotion mail reminding one of their friends to start using the app. Moreover, assume that the social network graph has $\eta(n/2) = d/2$. Argue that $X(t)$ is now stochastically dominated from below by a Markov chain $Y(t)$ over the states $\{0, 1, \dots, n/2\}$, with transitions:

$$Y(t) \to Y(t) - 1 \text{ at rate } Y(t) \quad , \quad Y(t) \to Y(t) + 1 \text{ at rate } Y(t)\cdot(\lambda d/2 + \alpha)$$

This suggests the following differential equation for the expected trajectory (ignoring the upper bound on the state of $Y(t)$):

$$\frac{d\mathbf{E}[Y(t)]}{dt} = (\lambda d/2 + \alpha - 1)\mathbf{E}[Y(t)]$$

Using this as a guide, what value of α should you choose to ensure the app continues to be used by people for a long time?

Solution: To ensure a long-lasting infection, we want to stochastically dominate the proces $X(t)$ from below. In this case, when the number of people using the app is $i < n/2$, then the rate at which people start using the app is now at least $\lambda i d/2 + \alpha i$ – this follows from the fact that any set of i people have at least di neighbors, whom they infect at rate at most λ, and in addition, each person refers other people at rate α. Thus, we now have that the dominating chain obeys:

$$\frac{d\mathbf{E}[Y(t)]}{dt} = (\lambda d/2 + \alpha - 1)\mathbf{E}[Y(t)],$$

which we can solve to get $\mathbf{E}[Y(t)] = Y(0)e^{(\lambda d/2 - \alpha)t}$. Now, in order to ensure that the expected number of people using the app grows with time, we need to ensure that $\alpha > 1 - \lambda d/2$.

Problem 5: (The Preferential Attachment Process) (OPTIONAL)

In this question, we will use differential equation approximations to study the *preferential attachment* networks. This is a model of network growth that is used to explain the occurrence of *power-law degree distributions*, which are often seen in real networks. For a network with n nodes, the degree distribution $F(k)$ is the fraction of nodes with degree $\leq k$ – this is said to exhibit power-law-α if:

$$1 - F(k) \sim ck^{-\alpha}$$

In other words, the fraction of nodes with degree greater than k is proportional to $k^{-\alpha}$.

Part (a)

To understand why power-laws are unexpected, consider the $G(n,p)$ graph, with $p = \lambda/(n-1)$. Argue that for any node i, its degree $D_i \sim Binomial(n-1,p) \approx Poisson(\lambda)$ for large n. Now since the fraction of nodes with degree less than k is same as the probability that a random node has degree less than k, argue that for large n, $1 - F(k) = 1 - \sum_{i=0}^{k} e^{-\lambda}\lambda^i/i!$. Note that this is exponential in k (rather than power-law).

Part (b)

The preferential attachment model is a dynamic model for generating a graph, which results in a power-law degree distribution. We will study a model for forming a directed graph – however it can be used for undirected graphs in a similar way.

The model proceeds as follows: time is slotted and nodes arrive one-by-one in the order $\{1, 2, \ldots, n\}$. Each arriving node brings a directed edge – the node arriving at time t chooses the existing node to link to uniformly at random with probability $1-p$, and with probability p, it picks an existing node v with probability proportional to its in-degree $D_v(t)$. Note that at the end of slot t, the number of nodes in the system is t, and so is the number of directed edges (and hence the sum of in-degrees).

For node v, argue that its in-degree $D_v(t)$ has the following dynamics: $D_v(t) = 0$ if $t \leq v$, and for all $t > v$, $D_v(t)$ transitions as:

$$D_v(t) = D_v(t-1) + 1 \quad \text{with probability } \frac{1-p}{t} + \frac{pD_v(t)}{t}, \text{ else } D_v(t) = D_v(t-1)$$

This suggests that $\mathbf{E}[D_v(t)]$ has the following differential equation approximation:

$$\frac{d\mathbf{E}[D_v(t)]}{dt} \approx \frac{1-p}{t} + \frac{p\mathbf{E}[D_v(t)]}{t}$$

Part (c)

Using the boundary conditions $\mathbf{E}[D_v(v)] = 0$, solve the above equation to get (for $t > v$):

$$\mathbf{E}[D_v(t)] \approx \frac{(1-p)}{p}\left(\left(\frac{t}{v}\right)^p - 1\right),$$

Part (d)

An important thing to note from the above characterization is that the (expected) in-degree of a node depends on when it arrived (in contrast to the $G(n,p)$ setting, where all nodes were symmetric). Now after all n nodes arrive, we can approximate the degree distribution $F(k)$ as the fraction of nodes $v \in \{1, 2, \ldots, n\}$ for which $\mathbf{E}[D_v(t)] \leq k$. Let $\gamma = \frac{p}{(1-p)}$ – show that:

$$1 - F(k) = (\gamma k + 1)^{-\frac{1}{p}}$$

Thus, the resulting graph has a power-law degree distribution with $\alpha = 1/p$.