Complex Networks
Solutions 3

---

1. (a) Let $T(k)$ be the first time that at least k nodes are infected. Then $T(1) = 0$ by the assumption that there is a single initial infective. We start by writing a recursion for $T(k+1)$ in terms of $T(k)$. At time $T(k)$, there are exactly k infected nodes. Each makes new infection attempts according to independent unit rate Poisson processes, so the process of all infection attempts is Poisson(k), as the superposition of independent Poisson processes is Poisson with the sum of the rates.

   Now, each of these attempts targets an infected node with probability k/n (in which case nothing happens), and a healthy node with probability $(n-k)/n$, in which case one more node becomes infected. Thus, the process of times at which a healthy node is infected (until the first such event occurs) happens according to a Poisson process of rate $k(n-k)/n$, because the Bernoulli thinning of a Poisson process is Poisson. Thus, $T(k+1) - T(k)$ is exponentially distributed with parameter $k(n-k)/n$, hence with mean

$$\mathbb{E}[T(k+1) - T(k)] = \frac{n}{k(n-k)} = \frac{1}{k} + \frac{1}{n-k}.$$

   By the linearity of expectation and the fact that $T_1 = 0$, we get

$$\begin{aligned}
\mathbb{E}[T(\lceil\sqrt{n}\rceil)] &= \sum_{k=1}^{\lceil\sqrt{n}\rceil-1} \mathbb{E}[T(k+1) - T(k)] = \sum_{k=1}^{\lceil\sqrt{n}\rceil-1} \left(\frac{1}{k} + \frac{1}{n-k}\right) \\
&= \sum_{k=1}^{\lceil\sqrt{n}\rceil-1} \frac{1}{k} + \sum_{k=n-\lceil\sqrt{n}\rceil+1}^{n-1} \frac{1}{k} \\
&\approx \log\sqrt{n} + \frac{\sqrt{n}}{n} \approx \frac{1}{2}\log n.
\end{aligned}$$

   (b) By using the exact same recursion,

$$\begin{aligned}
\mathbb{E}[T(n/2)] &= \sum_{k=1}^{(n/2)-1} \mathbb{E}[T(k+1) - T(k)] = \sum_{k=1}^{(n/2)-1} \left(\frac{1}{k} + \frac{1}{n-k}\right) \\
&= \sum_{k=1}^{(n/2)-1} \frac{1}{k} + \sum_{k=(n/2)+1}^{n-1} \frac{1}{k} \\
&\approx \log n.
\end{aligned}$$

   Comparing this with the answer to the last part, $c = 2$ suffices.

   (c) By using the same recursion as above, but now with the initial condition $T(\sqrt{n}) = 0$, we get

$$\begin{aligned}\mathbb{E}[\tilde{T}(n)] &= \sum_{k=\sqrt{n}} n-1\mathbb{E}[T_{k+1}-T_k] = \sum_{k=\sqrt{n}}^{n-1}\Big(\frac{1}{k}+\frac{1}{n-k}\Big) \\ &\approx \int_{\sqrt{n}}^{n-1}\Big(\frac{1}{x}+\frac{1}{n-x}\Big)dx \\ &\approx \log\frac{n}{\sqrt{n}}+\log(n-\sqrt{n}) \approx \frac{3}{2}\log n.\end{aligned}$$

So $c = 2/3$ suffices.

2. Consider the following rumour spreading model on the same graph, with just the node s initially informed. Each node becomes active at the points of a Poisson process of rate n, with the processes corresponding to different nodes being mutually independent. When a node v becomes active, it picks a node w to talk to with probability $1/n$, and informs it if it isn't already informed. As seen in Problem 1(a), this is equivalent to the following description.

   When a node v becomes informed, it will need to sample iid Exp(1) random variables for every node w, and communicate at the time corresponding to the minimum of these random variables with that node for which the minimum is achieved. Thus, we may identify the edge lengths (v, w) with the Exp(1) random variables that are being sampled to decide who to speak to. Subsequently, after v has communicated with some w, we need to sample again from an Exp(1) distribution to determine subsequent communications, because the residual time to the next communication along an edge is again Exp(1) by the memoryless property of the Exp(1) distribution.

   Hence, the spread of the rumour is exactly modelled by computing minimum path lengths in the graph with random iid Exp(1) edge lengths. The first time that some node $v \neq s$ becomes informed of the rumour has exactly the distribution of the minimum path length $d_{s,v}$ from s to v in this edge-weighted graph. Hence, the maximum over v of the distance from s to v, $D_s = \max_{v\in V} d_{s,v}$ has exactly the same distribution, and therefore mean, as the rumour spreading time.

   We saw in lectures that the mean rumour spreading time in this model is $(2\log n)/n$ (rather, in lectures, each node was activated at rate 1 and the mean rumour spreading time was $2\log n$, but here the model runs n times faster). Hence,

$$\mathbb{E}[D_s] = \frac{2\log n}{n}.$$

3. (a) The value of r_{ij} for the equivalence to hold is $r_{ij} = \lambda_i p_{ij}$. The reason is that the times at which node i contacts node j are obtained as a Bernoulli thinning of the Poisson(λ_i) process of random times at which node i becomes active, with thinning probability p_{ij}, and we know that Bernoulli thinning of a Poisson process yields another Poisson process.

(b) Let S_k denote the random set of k nodes that know the rumour at time T_k. Then, using the edge-driven model of part (a), the time until a new node learns the rumour is given by the minimum of all the Exponential random variables corresponding to Poisson processes on the edges (i,j), where $i \in S_k$ and $j \in S_k^c$. This is because these are the times at which an informed node $i \in S_k$ contacts an uninformed node $j \in S_k^c$. Hence, using the fact that the minimum of independent Exponential random variables is Exponential with the sum of their rates, we get

$$T_{k+1} - T_k \sim \text{Exp}\Big(\sum_{i \in S_k, j \in S_k^c} r_{ij} \Big),$$

conditional on S_k, and so

$$\mathbb{E}[T_{k+1} - T_k | S_k] = \Big(\sum_{i \in S_k, j \in S_k^c} r_{ij} \Big)^{-1}.$$

However, this depends on the random set S_k, and it would be very difficult to compute the unconditional expectation by working out the probability of reaching each set S_k and averaging with respect to these probabilities. Instead, we see from the definition of conductance that, irrespective of the actual set S_k reached at time T_k, but using only the fact that $|S_k| = k$ and $|S_k^c| = n-k$, we have

$$\mathbb{E}[T_{k+1} - T_k | S_k] \le \frac{n}{k(n-k)\Psi(R)} = \frac{1}{\Psi(R)} \Big(\frac{1}{k} + \frac{1}{n-k} \Big), \tag{1}$$

which doesn't depend on S_k. Hence, the unconditional expectation $\mathbb{E}[T_{k+1} - T_k]$ is also bounded by the same quantity.

The stochastic domination result claimed in the problem follows from the fact that the mean of the exponential random variable $T_{k+1} - T_k$ is bounded as above, and the rate of the exponential is the reciprocal of its mean. Hence,

$$T_{k+1} - T_k \preceq \text{Exp}\Big(\frac{k(n-k)\Psi(R)}{n} \Big).$$

(c) Note that $T_1 = 0$, and so $T_n = \sum_{k=1}^{n-1} (T_{k+1} - T_k)$. Using the linearity of expectation, we have by equation (1) that

$$\mathbb{E}[T_n] = \sum_{k=1}^{n-1} \mathbb{E}[T_{k+1} - T_k] \le \frac{1}{\Psi(R)} \sum_{k=1}^{n-1} \Big(\frac{1}{k} + \frac{1}{n-k} \Big).$$

Using the fact stated in the problem, it is immediate that

$$\mathbb{E}[T_n] \le \frac{2}{\Psi(R)} (1 + \log n),$$

since $\sum_{k=1}^{n-1} \frac{1}{n-k} = \sum_{j=1}^{n-1} \frac{1}{j}$ by the change of variable $j = n-k$.

4. (a) Suppose only the hub node knows the rumour initially. Then, at time T_k, the hub and $k-1$ leaves know it, while $n-k$ leaves don't know it. Only the hub can inform these leaves. It contacts each of them at rate $1/n$, according to the problem statement, so the total rate at which some uninformed leaf becomes informed is $(n-k)/n$. Hence,

$$T_{k+1} - T_k \sim \text{Exp}\Big(\frac{n-k}{n}\Big) \text{ and so } \mathbb{E}[T_{k+1} - T_k] = \frac{n}{n-k},$$

exactly. Summing this over k between 1 and $n-1$, using the fact that $T_1 = 0$, and using the linearity of expectation, we get the exact estimate

$$\mathbb{E}[T_n] = \sum_{k=1}^{n-1} \mathbb{E}[T_{k+1} - T_k] = n \sum_{k=1}^{n-1} \frac{1}{n-k} = n \sum_{k=1}^{n-1} \frac{1}{k} \sim n \log n.$$

The last $\sim$ denotes the quantities on the two sides are asymptotically equivalent; their ratio tends to 1 as n tends to infinity. (Not to be confused with the $\sim$ in the previous equation, which was used to denote that two random variables have the same distribution. It should be clear from context which is meant.)

(b) We first translate the model to the setting of Problem 1(a) and compute the corresponding R matrix. It is easy to see that we get

$$r_{ij} = \begin{cases} 1/n, & \text{if } i \text{ is the hub and } j \text{ is a leaf} \\ 1, & \text{if } i \text{ is a leaf and } j \text{ is the hub} \\ 0, & \text{if } i \text{ and } j \text{ are both leaves.} \end{cases}$$

Now we compute $\Psi(R)$ for this matrix.

First consider a non-empty proper subset S of V that contains the hub. Say it also contains $k-1$ leaves for some k between 1 and $n-1$, so that $|S| = k$. For $i \in S$ and $j \in S^c$, either i is the hub and $r_{ij} = 1/n$ (there are $n-k$ terms like this), or i is a leaf and $r_{ij} = 0$ (there are $(k-1)(n-k)$ terms like this). Hence,

$$\frac{\sum_{i \in S, j \in S^c} r_{ij}}{\frac{1}{n}|S| \cdot |S^c|} = \frac{\frac{n-k}{n}}{\frac{1}{n}k(n-k)} = \frac{1}{k}.$$

For k between 1 and $n-1$, this is minimised at $k = n-1$, where it takes the value $1/(n-1)$.

Next consider a non-empty proper subset S of V that doesn't contain the hub, but contains k leaves, for some k between 1 and $n-1$. For $i \in S$ and $j \in S^c$, either j is the hub and $r_{ij} = 1$ (there are k terms like this), or j is a leaf and $r_{ij} = 0$ (there are $k(n-k-1)$ terms like this). Hence,

$$\frac{\sum_{i \in S, j \in S^c} r_{ij}}{\frac{1}{n}|S| \cdot |S^c|} = \frac{k}{\frac{1}{n}k(n-k)} = \frac{n}{n-k}.$$

For k between 1 and $n-1$, this is minimised at $k = 1$, where it takes the value $n/(n-1)$. Comparing this with $1/(n-1)$, the smaller quantity is $1/(n-1)$, and so $\Psi(R) = \frac{1}{n-1}$.

Hence, by the answer to Problem 1(c), we get the upper bound,

$$\mathbb{E}[T_n] \leq 2(n-1)(1+\log n).$$

The upper bound is roughly twice the exact answer for large n, which is quite good. For the smallest possible value of n, namely $n=2$, the upper bound is $2(1+\log 2) \approx 3.4$, while the exact answer is $2\sum_{k=1}^{1} 1/k = 2$.

(c) When a single node initially knows the rumour, the first rumour spreading event corresponds to this node informing the hub, which takes an Exp(1) time according to the problem description. In other words,

$$T_2 - T_1 \sim \text{Exp}(1) \text{ and } \mathbb{E}[T_2 - T_1] = 1.$$

Subsequent to this, all other rumour spreading events correspond to the hub informing another leaf. The analysis is exactly as in part (a); when the hub and $k-1$ leaves are informed, the time to inform one more leaf is Exp($\frac{n-k}{n}$)), which implies that $\mathbb{E}[T_{k+1} - T_k] = \frac{n}{n-k}$ for $k = 2, 3, \ldots, n-1$. Since $T_n = \sum_{k=1}^{n-1}(T_{k+1} - T_k)$, and $T_1 = 0$, we have by the linearity of expectation that

$$\begin{aligned} \mathbb{E}[T_n] &= \mathbb{E}[T_2 - T_1] + \sum_{k=2}^{n-1} \mathbb{E}[T_{k+1} - T_k] = 1 + \sum_{k=2}^{n-1} \frac{n}{n-k} \\ &= 1 + n\sum_{k=1}^{n-2} \frac{1}{k} \sim 1 + n\log n \sim n \log n. \end{aligned}$$

5. (a) At time T_k, a contiguous set of k nodes knows the rumour. They can only inform new nodes by either the leftmost of them communicating with its left neighbour, or the rightmost communicating with its right neighbour. From the problem description, each of these can happen at rate $\lambda p = 1 \cdot \frac{1}{2} = 1/2$, for a total rate of 1 for informing a new node. Hence, $T_{k+1} - T_k$ is exponentially distributed with mean 1, and $\mathbb{E}[T_{k+1} - T_k] = 1$ for all k. Also, $T_n = \sum_{k=1}^{n-1}(T_{k+1} - T_k)$ since $T_1 = 0$ (we start with a single informed node). Hence, by the linearity of expectation,

$$\mathbb{E}[T_n] = \sum_{k=1}^{n-1} \mathbb{E}[T_{k+1} - T_k] = n - 1.$$

(b) We compute the rates r_{ij} for communications along the directed edge (i,j) as in Problem 1. Since $\lambda_i = 1$ and $p_{ij} = 1/2$ whenever i and j are neighbours, we have $r_{ij} = 1/2$ if i and j are neighbours, and zero otherwise.

Consider a set S consisting of k contiguous nodes. Then, it is only the leftmost and rightmost nodes in this set that can communicate with nodes outside it. Thus,

$$\sum_{i \in S, j \in S^c} r_{ij} = \frac{1}{2} + \frac{1}{2} = 1,$$

where the two terms in the sum come from the edges from the leftmost node to its left neighbour and the rightmost node to its right neighbour (which could be the same node if $k = n-1$). Moreover, $|S| = k$ and $|S^c| = n-k$. Hence,

$$\frac{\sum_{i\in S, j\in S^c} r_{ij}}{\frac{1}{n}|S|\cdot|S^c|} = \frac{n}{k(n-k)},$$

which takes its minimum value of $4/n$ at $k = n/2$ (being a bit careless about whether n is odd or even). Note that the conductance $\Psi(R)$ is defined as the minimum of the LHS above over all subsets S of V other than the empty set and all of V. However, it should be intuitively obvious that this is the same as the minimum over sets of contiguous nodes because, for any k, a subset of k nodes which are not contiguous will have more edges going out of them than a contiguous subset, and hence the sum of r_{ij} over such edges will also be bigger (since $r_{ij} = 1/2$ is the same for every edge). Thus, we conclude that $\Psi(R) = 4/n$. Hence, using the bound from Problem 1(c),

$$\mathbb{E}[T_n] \leq \frac{2(1+\log n)}{\Psi(R)} = n\frac{1+\log n}{2}.$$

The bound is roughly a multiple $\log n/2$ of the exact value, which is rather less good than in the case of the star. On the other hand, the bound is generally applicable, whereas exact calculcation is generally only possible when there is a lot of symmetry, as in the examples we have considered.