Complex Networks
Problem Sheet 3

$**$ **Please hand in solutions to question 4 on this sheet.** $**$

1. Let $G = (V, E)$ be the complete, undirected graph on n nodes. The simple epidemic on G is described as follows. There is an initial set $S \subseteq V$ of infected nodes, and all other nodes are healthy. Each infected node attempts to spread infection at the points of a unit rate Poisson process. The Poisson processes corresponding to distinct nodes are mutually independent.

   When a node attempts to spread infection, it chooses a node from V uniformly at random (including itself), independent of the Poisson processes of spreading times and past node choices of itself or of other nodes. If the chosen node is healthy, it becomes infected at this time. If it is already infected, nothing changes. Once a node becomes infected, it remains infected forever.

   In answering the questions below, think of n as large. You may replace sums by integrals, ignore terms of smaller order than the dominant term, and make any other reasonable approximations required, so long as you get the correct dominant term as a function of n.

   (a) Suppose a single node is initially infected. Compute the mean time until at least $\sqrt{n}$ nodes are infected. Call this $\mathbb{E}[T(\sqrt{n})]$.

   (b) What is the mean time until at least $n/2$ nodes are infected? Call this $\mathbb{E}[T(n/2)]$. What is the smallest constant c such that $\mathbb{E}[T(n/2)] \leq c\mathbb{E}[T(\sqrt{n})]$ for all n sufficiently large?
   Hint. Replacing sums by integrals and making other reasonable approximations may help you estimate c.

   (c) Next, suppose that $\sqrt{n}$ nodes are initially infected. (Assume that n is a perfect square.) What is the mean time until all n nodes are infected? Call this $\mathbb{E}[\tilde{T}(n)]$. What is the smallest constant c such that $\mathbb{E}[T(n/2)] \leq c\mathbb{E}[\tilde{T}(n)]$ for all n sufficiently large?

2. Consider the complete undirected graph $G = (V, E)$. Suppose each edge (v, w) has a random length drawn from an Exp(1) distribution, and that the lengths of different edges are mutually independent. Fix a node $s \in V$. For any other node $v \in V$, the distance from s to v, denoted $d(s, v)$, is defined as the minimum of the lengths of all paths between s and v. The length of a path is the sum of the lengths of the edges constituting the path. The distance $d(s, s)$ is defined to be zero. Finally, let $D_s = \max_{v \in V} d_{s,v}$ denote the maximum distance from s to any another node in the graph.

   Compute the mean of the random variable D_s (or a good bound on it) by reducing the problem to one you know how to solve. Explain your reasoning carefully.

3. Let $G = (V, E)$ be a *directed* graph on n nodes. Consider the following rumour-spreading model on G. There are n independent Poisson processes, $\{N_v(t), t \geq 0\}$, one associated with each node $v \in V$. The Poisson process at node v has rate λ_v. If there is an increment of the process $N_v(\cdot)$ at time t, then node v chooses ones of its neighbours w at random, with probability p_{vw}, which is the vw^{th} element of a stochastic matrix P. If node v knows the rumour at time t, then node w learns it at this time; if neither or both nodes know the rumour, there is no change.

   Let T_k be the first time that exactly k nodes know the rumour, and suppose that $T_1 = 0$, i.e., a single node knows the rumour to start with.

   (a) The above 'node-driven' model is equivalent to the following 'edge-driven' model. There are independent Poisson processes on the edges, with r_{ij} denoting the rate of the process on edge (i, j). Whenever there is an increment of the Poisson process on edge (i, j), node i informs node j of the rumour if i already knows it and j does not. If both or neither of i and j know the rumour, there is no change.

   Explicitly compute the value of r_{ij} for each (i, j), in terms of the rates λ_v and matrix P given above, so that the equivalence holds. You don't have to prove the equivalence.

   (b) For the model in part (a), let R denote the matrix with elements r_{ij}, and define its generalised conductance

   $$\Psi(R) = \min_{S \subset V : S \neq \emptyset} \frac{\sum_{i \in S, j \in S^c} r_{ij}}{\frac{1}{n}|S| \cdot |S^c|},$$

   where S^c denotes the complement of the subset S.

   Show that for each k between 1 and $n-1$, $T_{k+1} - T_k$ is stochastically dominated by an $Exp\big(\frac{1}{n}k(n-k)\Psi(R)\big)$ random variable. You may use the fact that if $\alpha \geq \beta$, then an $Exp(\alpha)$ random variable is stochastically dominated by an $Exp(\beta)$ random variable.

   (c) Use the answer to part (b) to obtain an upper bound on $\mathbb{E}[T_n]$, the time until all nodes learn the rumour. Specifically, show that

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

   You may use the fact that

   $$\sum_{k=1}^{n-1} \frac{1}{k} \leq 1 + \log(n).$$

   (All logarithms are natural.)

4. Let S_n be the star graph on n nodes consisting of a single hub node connected to each of $n - 1$ leaves; there are no edges between leaves. Consider the rumour-spreading model of Question 1, with $\lambda_v = 1$ for all v, $p_{ij} = 1/n$ if i is the hub and j is a leaf, $p_{ij} = 1$ if i is a leaf and j is the hub, and $p_{ij} = 0$ if i and j are both leaves.

   (a) Suppose that only the hub node knows the rumour at time 0. Compute $\mathbb{E}[T_{k+1} - T_k]$ exactly, and use this to compute $\mathbb{E}[T_n]$ exactly.

(b) Compute $\Psi(R)$, defined in Question 1(b), for the star graph with the rates and probabilities specified above, and the corresponding upper bound on $\mathbb{E}[T_n]$, and compare it with the exact answer.

(c) Repeat the exact analysis when only a single leaf node initially knows the rumour.

5. Let C_n be the cycle graph on n nodes numbered $\{1, 2, 3, \ldots, n\}$, where there are two directed edges out of each node i. These go to nodes $i-1$ and $i+1$ for $2 \leq i \leq n-1$. The edges out of node 1 go to nodes 2 and n, while the edges out of node n go to nodes $n-1$ and 1. Consider the rumour-spreading model of Question 1, with $\lambda_v = 1$ for all v and $p_{ij} = 1/2$ if for every $(i, j) \in E$.

   (a) Suppose that only a single node knows the rumour at time 0. Compute $\mathbb{E}[T_{k+1} - T_k]$ exactly, and use this to compute $\mathbb{E}[T_n]$ exactly. (*Hint.* Observe that, at any time, the set of nodes that knows the rumour has to be a contiguous set.)

   (b) Compute $\Psi(R)$ for the cycle graph with the rates and probabilities specified above, and the corresponding upper bound on $\mathbb{E}[T_n]$, and compare it with the exact answer.