Complex Networks
Problem Sheet 6

$**$ **Please hand in solutions to Question 5 on this sheet.** $**$

1. (a) A *weighted graph* G has a weight $a_{ij} > 0$ associated with each edge $(i, j) \in E$. If the graph is undirected, we take $a_{ji} = a_{ij}$. The weighted degree of node i is defined as $d_i = \sum_{j \in V} a_{ij}$, and the matrix D_G is taken to be $diag(d_i)$. The weighted adjacency matrix A_G has elements a_{ij}, and the weighted Laplacian is defined as $L_G = D_G - A_G$. Now show that, for any $\mathbf{x} \in \mathbb{R}^{|V|}$,

$$\mathbf{x}^T L_G \mathbf{x} = \sum_{(i,j) \in E} a_{ij}(x_i - x_j)^2.$$

   Explain why L_G is a positive semi-definite matrix.

   (b) Consider the graph with weighted adjacency matrix

$$A_G = \begin{pmatrix} 0 & 1 & 2 \\ 1 & 0 & 1 \\ 2 & 1 & 0 \end{pmatrix}.$$

   Compute $\mathbf{x}^T L_G \mathbf{x}$ and $\sum_{(i,j) \in E} a_{ij}(x_i - x_j)^2$ and verify that they are equal. Also compute the eigenvalues of the Laplacian and verify that they are non-negative.

2. A Markov chain with rate matrix Q and invariant distribution π is said to be reversible if $\pi_i q_{ij} = \pi_j q_{ji}$ for all i and j in the state space. Consider a reversible irreducible Markov chain on a finite state space S, with $|S| = n$. Define an inner product on $\mathbb{C}^n$ by setting

$$\langle \mathbf{x}, \mathbf{y} \rangle = \sum_{i \in S} \pi_i x_i y_i^*,$$

where we use the superscript $*$ to denote the complex conjugate of a number.

   (a) Check that the definition above is a valid inner product, i.e., that the following hold for all $\mathbf{x}, \mathbf{y}, \mathbf{z} \in \mathbb{C}^n$ and all $c \in \mathbb{C}$:

$$\begin{aligned} &\langle \mathbf{x}, \mathbf{y} \rangle = \langle \mathbf{y}, \mathbf{x} \rangle^*, \\ &\langle \mathbf{x} + \mathbf{y}, \mathbf{z} \rangle = \langle \mathbf{x}, \mathbf{z} \rangle + \langle \mathbf{y}, \mathbf{z} \rangle, \quad \langle c\mathbf{x}, \mathbf{y} \rangle = c\langle \mathbf{x}, \mathbf{y} \rangle, \\ &\langle \mathbf{x}, \mathbf{x} \rangle \geq 0, \text{ with equality only if } \mathbf{x} = \mathbf{0}. \end{aligned}$$

   You may use the fact that $\pi_i > 0$ for all $i \in S$, which holds because the Markov chain is irreducible.

(b) Show that the rate matrix Q is "self-adjoint" for the above inner product, i.e., show that

$$\langle Q\mathbf{x}, \mathbf{y} \rangle = \langle \mathbf{x}, Q\mathbf{y} \rangle,$$

for all $\mathbf{x}$ and $\mathbf{y} \in \mathbb{C}^n$.

(c) Use the answer to the last part to show that all eigenvalues of Q are real. The proof follows along the same lines as the proof in lectures that all eigenvalues of a symmetric matrix are real. You may assume that there is an eigenvector associated with each eigenvalue.

3. Let $\mathbf{p}$ and $\mathbf{q}$ be discrete probability distributions on a finite set $\Omega = \{1, 2, \ldots, n\}$. Recall that the total variation distance between $\mathbf{p}$ and $\mathbf{q}$ is defined as

$$d_{TV}(\mathbf{p}, \mathbf{q}) = \max_{S \subseteq \Omega} |p(S) - q(S)|,$$

where $p(S) = \sum_{i \in S} p_i$ and $q(S)$ is defined similarly.

(a) Show that $d_{TV}(p, q) = \frac{1}{2} \|\mathbf{p} - \mathbf{q}\|_1$, where $\|\mathbf{p} - \mathbf{q}\|_1$ is defined as $\sum_{i=1}^n |p_i - q_i|$. (*Hint.* Show that the maximum in the definition of total variation distance is attained by the set $S = \{i : p_i \geq q_i\}$.)

(b) Using the answer to the last part or otherwise, show that $\|\mathbf{p} - \mathbf{q}\|_1 \leq 2$ for all probability distributions $\mathbf{p}$ and $\mathbf{q}$. Give an example where equality holds.

(c) Show that $\|\mathbf{p} - \mathbf{q}\|_1 \leq \sqrt{n} \|\mathbf{p} - \mathbf{q}\|_2$, where $\|\mathbf{x}\|_2$ is defined as $\sqrt{x_1^2 + \ldots + x_n^2}$. (*Hint.* Use the Cauchy-Schwarz inequality.)

(d) Show that $\|\mathbf{p} - \mathbf{q}\|_2 \leq \sqrt{2}$ for any probability distributions $\mathbf{p}$ and $\mathbf{q}$. Give an example where equality holds.

4. Compute the total variation distance between the two distributions in each of the following examples:

(a) Binomial($2, \frac{1}{2}$) and uniform on $\{0, 1, 2\}$.

(b) Binomial($2, \frac{1}{2}$) and Poisson(1).

(c) Exponential(1) and Uniform[0,1].

(d) Exponential(1) and Exponential(2).

5. 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 continuous time random walk on this graph generated as follows: there are $n - 1$ independent Poisson processes, $\{N_e(t), t \geq 0\}$, one on each edge of the graph. Each of these Poisson processes has rate 1. If the Poisson process $N_e(\cdot)$ has an increment at time t and the walker is at the vertex on one end of this edge, it moves to the vertex at the other end. An equivalent description is that, if the walker is at a leaf, it moves to the hub at rate 1; if it is at the hub, it moves at rate $n - 1$, choosing a leaf uniformly at random to move to.

(a) The position X_t of the random walk evolves as a continuous time Markov chain on the set of vertices. Write down the transition rate matrix of this Markov chain. How is it related to the Laplacian matrix of the graph S_n?

(b) Show that the uniform distribution on all nodes is an invariant distribution for the Markov chain in part (a), and hence that it is the unique invariant distribution.

(c) The conductance of a graph G is defined as

$$\Phi(G) = \min_{S \subset V : S \neq \emptyset} \frac{|E(S, S^c)|}{\frac{1}{n}|S| \cdot |S^c|},$$

where $E(S, S^c)$ denotes the set of all edges consisting of one vertex in S and the other in its complement S^c, and the minimum is taken over all subsets of the vertex set V other than the empty set and the set of all vertices.

Compute the conductance of the star graph.

(d) Obtain a lower bound on λ_2, the second smallest eigenvalue of the Laplacian of S_n, using Cheeger's inequality and the answer to part (c).

(e) It is known that the total variation distance between the distribution of the random walk position at time t, which we denote $p(t)$, and the invariant distribution π is bounded as follows:

$$d_{TV}(p(t), \pi) := \frac{1}{2} \sum_{i=1}^{n} |p_i(t) - \pi_i| \leq \sqrt{n} e^{-\lambda_2 t}.$$

Let $\epsilon > 0$ be a given constant. Using the lower bound on λ_2 computed in part (d), find the smallest value of t for which you can guarantee that $d_{TV}(p(t), \pi) \leq \epsilon$.