18.S096: Homework Problem Set 2

Topics in Mathematics of Data Science (Fall 2015)

Afonso S. Bandeira

Due on October 20, 2015

Problem 2.1 *Given a graph $G = (V, E, W)$ consider the random walk on V with transition probabilities*

$$M_{ij} = \text{Prob}\{X(t+1) = j | X(t) = i\} = \frac{w_{ij}}{\deg(i)}.$$

Partition the vertex set as $V = V_+ \cup V_- \cup V_*$. Suppose that every node in V_* is connected to at least a node in either V_+ or V_-. Given a node $i \in V$ let $g(i)$ be the probability that a random walker starting at i reaches a node in V_+ before reaching one in V_-, i.e.:

$$g(i) = \text{Prob}\left\{ \inf_{t \geq 0:\ X(t) \in V_+} t < \inf_{t \geq 0:\ X(t) \in V_-} t \,|X(0) = i \right\}.$$

Note that if $i \in V_+$ then $g(i) = 1$ and, if $i \in V_-$, then $g(i) = 0$. What is the value of g in V_*? How would you compute it?

Problem 2.2 *For a graph G let $h(G)$ denote its Cheeger constant and $\lambda_2(\mathcal{L}_G)$ the second smallest eigenvalue of its normalized graph Laplacian. Recall that Cheeger inequality guarantees that*

$$\frac{1}{2}\lambda_2(\mathcal{L}_G) \leq h_G \leq \sqrt{2\lambda_2(\mathcal{L}_G)}.$$

This exercise shows that this inequality is tight (at least up to constants).

1. *Construct a family of graphs for which $\lambda_2(\mathcal{L}_G) \to 0$ and for which there exists a constant $C > 0$ for which, for every G in the family,*

$$h_G \leq C\lambda_2(\mathcal{L}_G)$$

2. *Construct a family of graphs for which* $\lambda_2(\mathcal{L}_G) \to 0$ *and for which there exists a constant* $c > 0$ *for which, for every* G *in the family*

$$h_G \geq c\sqrt{\lambda_2(\mathcal{L}_G)}$$

Problem 2.3 *Given a graph* G *show that the dimension of the nullspace of* L_G *corresponds to the number of connected components of* G*.*

Problem 2.4 *Given a connected unweighted graph* $G = (V, E)$*, its diameter is equal to*

$$\operatorname{diam}(G) = \max_{u,v \in V} \min_{\text{path } p \text{ from } u \text{ to } v} \text{length of } p.$$

Show that

$$\operatorname{diam}(G) \geq \frac{1}{\operatorname{vol}(G)\lambda_2(\mathcal{L}_G)}.$$

Problem 2.5

1. *Prove the Courant Fisher Theorem: Given a symmetric matrix* $A \in \mathbb{R}^{n \times n}$ *with eigenvalues* $\lambda_1 \leq \cdots \leq \lambda_n$*, for* $k \leq n$*,*

$$\lambda_k(A) = \min_{U:\, \dim(U)=k} \left[\max_{x \in U} \frac{x^T A x}{x^T x}\right].$$

2. *Show also that:*

$$\lambda_2(A) = \max_{y \in \mathbb{R}^n} \left[\min_{x \in \mathbb{R}^n:\, x \perp y} \frac{x^T A x}{x^T x}\right].$$

Problem 2.6 *Given a set of points* $x_1, \ldots, x_n \in \mathbb{R}^p$ *and a partition of them in* k *clusters* $S_1, \ldots, S_k$ *recall the* k*-means objective*

$$\min_{S_1,\ldots,S_k} \min_{\mu_1,\ldots,\mu_k} \sum_{l=1}^{k} \sum_{i \in S_i} \|x_i - \mu_l\|^2.$$

Show that this is equivalent to

$$\min_{S_1,\ldots,S_k} \sum_{l=1}^{k} \frac{1}{|S_l|} \sum_{i,j \in S_l} \|x_i - x_j\|^2.$$