15-859NN Spectral Graph Theory, Spring 2021

Homework 4 **Due: Friday May 7th in class**

Gary Miller

Instructions. Collaboration is permitted in groups of size at most three. You must write the names of your collaborators on your solutions and must write your solutions alone.

Question	Points	Score
1	25	
2	25	
3	25	
4	25	
Total:	100	

(25) 1. **Low Stretch Trees**

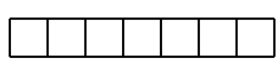
One of the most fundamental types of graphs is the spanning tree. The most well known class of trees is the maximum/minimum weight spanning tree. In the case when all edges of the graph have the same weight then all spanning trees have the same weight. The goal of this problem is to define another class of spanning trees, namely, ones called **low stretch spanning trees**. Recall that in a tree there is a unique path connecting any two nodes and we define the length of this path to be the number of edges on this path.

Let $G = (V, E)$ be a connected, unweighted and undirected graph with n vertices and m edges and $T = (V, E')$ a spanning tree of G. If $e = (v, w)$ is an edge of G we define the **stretch** of e in T to be the length of the path in T from v to w, denoted by $Str_T(e)$. The **stretch of T in G** is:

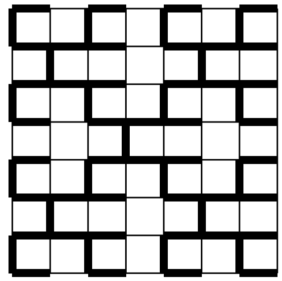
$$Str(T, G) = \sum_{e \in E} Str_T(e)$$

The average stretch is $Str(T, G)/m$

1. Construct two spanning trees T_1 and T_2 for the 2 by n mesh graph with average stretch $\Theta(1)$ and $\Theta(n)$, respectively. The 2 by 8 mesh graph is shown below.



2. Show that the recursive "C" construction for the $\sqrt{n}$ by $\sqrt{n}$ mesh has average stretch $O(\log n)$. Here you may assume that $\sqrt{n}$ is a power of 2. Below is an example of the tree for the 8 by 8 square mesh.



Hint: First bound the diameter if the recursive "C" tree.

(25) 2. **Recurrent Random Spanning Trees**

In class in order to prove Markov Chain Tree Theorem we need to consider a random walk over rooted (convergent) spanning trees in a strongly connected direct graph G. In particular, we defined a random walk on trees given a random walk on the vertices of G. In particular, Given a CST T_c rooted at c in one step we can go to the tree $T_a = (T_c \cup (c, a)) \setminus (a, b)$ where (c, a) is an edge out-edge of c and (a, b) is an edge in T_c, the **Last Visited** tree.

Show that the last visit walk has a single recurrent class. Namely, given a CST T there is walk where T is it's last visit tree.

I see how to construct a path of length $O(|V|^2)$. Can you do better or find an example requiring $\Omega(|V|^2)$ length path?

(25) 3. **Leverage Scores and Resistors**

Recall that much of this class has focused on the theorems regrading effective resistance of graphs. The goal of this problem is to determine if we can generalized these theorems to arbitrary matrices. Let B be the edge-by-vertex matrix of a connected graph G with a diagonal conductance matrix C then the Laplacian of G is $B^T CB$. If b_i is the ith row of B corresponding to the ith edge e_i of G then the effective resistance from one end of e_i to the other is $b_i(B^T CB)^+ b_i^T$. If we define $\bar{B} = C^{1/2} B$ then the resistance is $b_i(\bar{B}^T \bar{B})^+ b_i^T$. This motivates to following definition. Let $A^{m \times n}$ be a matrix of rank n.

Definition 1. *The* **Leverage Score** $\sigma(A, a)$ *where a is a column vector of size n is $a^T(A^T A)^+ a$ where $^+$ is the pseudoinverse. If a_i is the ith row of A the $\sigma_i(A) = \sigma(A, a_i^T)$.*

The goal of this problem is to determine what if any of the properties of effective resistance carry over the leverage scores.

1. Show that the leverage score of a nonzero row vector with itself is one.

2. Show that the column space $Col(A)$ of A and the left null space $Null_L(A)$ of A form an orthogonal bases of R^m. We will think of these vectors in $f \in \mathbb{R}^m$ as the flows. In the case when $A = \sqrt{C}B$ what kind of flows are the $Col(A)$ and the $Null_L(A)$?
3. We next prove a generalization of Foster's Theorem. Show that the sum of the row leverage scores of A is $rank(A)$. In this problem assume that A is not of full rank.
4. We next prove a generalization of Thomson's Principle. Suppose that x is a solution to the system $A^T Ax = b$ were b is in the column space of A^T. Show that the flow $f = Ax$ is the unique minimum energy flow such that $A^T f = b$. We define the energy of f to be $f^T f$.
5. We next prove a generalization of Rayleigh's Monotonicity Law. If we increase a row of A by scaling it by $1 + c$ for $c > 0$ or add a new row then no leverage score except the changed one will increase.
6. We next prove a generalization of Spielman-Srivastava Graph Sparsification Theorem. We say that $A^T A \approx_\epsilon B^T B$ if $(1 - \epsilon)A^T A \preccurlyeq B^T B \preccurlyeq (1 + \epsilon)A^T A$. Prove that that there exist a matrix QB where B is a subset of m' rows of A and Q is a nonnegative diagonal matrix where $m' = cn \log n$ for some constant c and $A^T A \approx_\epsilon B^T Q^T QB$.
7. Prove a variant of the fact that conductors add when placed in parallel. In particular prove a relationship between $\sigma(a, a)$, $\sigma(A, a)$, and $\sigma(\bar{A}, a)$ where $\bar{A}$ is the matrix A with row a appended.

   Hint: Consider the Sherman-Morrison formula.

   Can you find a more general formula?

Research questions:

1. We also showed that the effective resistance forms a metric over the **vertices**. Thus in our generalization we should be looking for a metric on the columns of A. Suppose we define the score between two columns as $D_{ij} = \chi_{ij}^T (A^T A)^{-1} \chi_{ij}$ where χ_{ij} is the column vector with a 1 and -1 at i and j respectively. Does our score say anything interesting about the relationship of two columns? Is D_{ij} a metric on the columns of A.
2. Is there theory of random walks for leverage scores, either on the columns or rows of A?
3. If there in such a theory as randoms walks does commute time make sense and is related to leverage score?

(25) 4. **Hardy Inequality and Path Embeddings**

In this problem we will prove one of the famous inequalities of Hardy using path embedding method from class.

Let $a_1, \ldots, a_n \ldots$ be a sequence of nonnegative real numbers. then.

Theorem 0.1 (Hardy).

$$\sum_{k=1}^{\infty}\left(\frac{1}{k}\sum_{i=1}^{k}a_i\right)^2 \leq 4\sum_{k=1}^{\infty}a_k^2$$

We will prove the Hardy inequality in stages.

1. Show that if we prove a finite version of Hardy we will get a proof of Hardy. In particular show that

$$\sum_{k=1}^{n}\left(\frac{1}{k}\sum_{i=1}^{k}a_i\right)^2 \leq (4+\delta(n))\sum_{k=1}^{n}a_k^2 \tag{1}$$

,where $\delta(n)$ goes to zero with n, implies theorem 0.1.

2. Let P_n be the path graph, with vertices $\{V_0,\dots,V_n\}$ and unit conductances. The weight of vertex V_i is $m_i = 1/i^2$ for $i \geq 1$. We consider the Dirichlet boundary problem of minimizing the Rayleigh quotient:

$$Ray = \min_{f(0)=0}\frac{\sum_{i=1}^{n}(f(i)-f(i-1))^2}{\sum_{i=1}^{n}m_i f^2(i)}$$

Show that proving that $1/4 \leq Ray$ implies equation 1.

3. At this point in the proof will use the path embedding lemma from class. We generalize the proof to handle non-uniform node weights. Let S_n be the star graph with n leaves. Let leaf vertex L_i have mass m_i and the edge from the center node C to L_i have conductance m_i.

   Show that all Rayleigh quotients that set the value at C to zero have value 1.

4. We next consider a path embedding of S_n into P_n where the center vertex of S_n is mapped to V_0. We first consider the allocation of conductance in proportion to the mass m_i for all path using an edge. Using this allocation show the effective resistance from V_n to V_0 is

$$\sum_{i=1}^{n} i m_i.$$

per unit of mass. Use this bound to show that $\frac{1}{\ln n} = \Omega(Ray)$.

5. In order to improve our lower bound we will need a more careful allocation of conductances. Suppose we allocate with a priority given to longer path. Suppose P_i from V_i to V_0 is given a priority p_i. Thus is given conductance propositional to $p_i m_i$. Using this allocation show the effective resistance from V_n to V_0 is:

$$\frac{\sum_{i=1}^{n} i p_i m_i.}{p_n}$$

per unit of mass. While the effective resistance from V_k to V_0 is:

$$\frac{\sum_{i=1}^{k} i p_i m_i + k\sum_{i=k}^{n} p_i m_i}{p_k} \tag{2}$$

6. Use equation 2 to show that $\frac{1}{4+o(1)} \leq Ray$.
   Hint: Equations 2 takes on a maximum value for k about $n/2$.