CS 798: Algorithmic Spectral Graph Theory **Last Updated:** Nov 21, 2015
Homework 2 **Due Date: Dec 14, 2015**

You are allowed to discuss with others but not allowed to use any references other than the course notes. Please list your collaborators for each question. This will not affect your marks. In any case, you must write your own solutions.

There are totally 95 marks, and the full mark is 80. This homework will be counted 30% of the course.

---

1. **Page Ranking**

   (10 marks) Suppose someone searches a keyword (like "car") and we would like to identify the webpages that are the most relevant for this keyword and the webpages that are the most reliable sources for this keyword (a page is a reliable source if it points to many most relevant pages). First we identify the pages with this keyword and ignore all other pages. Then we run the following ranking algorithm on the remaining pages. Each vertex corresponds to a remaining page, and there is a directed edge from page i to page j if there is a link from page i to page j. Call this directed graph $G = (V, E)$. For each vertex i, we have two values $s(i)$ and $r(i)$, where intentionally $r(i)$ represents how relevant is this page and $s(i)$ represents how reliable it is as a source (the larger the values the better). We start from some arbitrary initial values, say $s(i) = 1/|V|$ for all i, as we have no ideas at the beginning. At each step, we update s and r (where s and r are vectors of $s(i)$ and $r(i)$ values) as follows: First we update $r(i) = \sum_{j:ji \in E} s(j)$ for all i, as a page is more relevant if it is linked by many reliable sources. Then we update $s(i) = \sum_{j:ij \in E} r(j)$ for all i (using the just updated values $r(j)$), as a page is a more reliable source if it points to many relevant pages. To keep the values small, we let $R = \sum_{i=1}^{|V|} r(i)$ and $S = \sum_{i=1}^{|V|} s(i)$, and divide each $s(i)$ by S and divide each $r(i)$ by R. We repeat this step for many times to refine the values.

   Let $s, r \in \mathbb{R}^{|V|}$ be the vectors of the s and r values. Give a matrix formulation for computing s and r. Suppose G is weakly connected (when we ignore the direction of the edges the underlying undirected graph is connected) and there is a self-loop at each vertex. Prove that there is a unique limiting s and a unique limiting r for any initial s as long as $s \geq 0$ and $s \neq 0$.

   (You may use the Perror-Frobenius theorem which states that for any aperiodic irreducible matrix, there is a unique positive eigenvalue with maximum absolute value and the entries of the corresponding eigenvector are all positive.)

2. **Graph Partitioning by Random Walks**

   (15 marks) Let $G = (V, E)$ be an undirected d-regular graph and $S \subseteq V$. Let $p_t = W^t p_0$ where $W = A/d$ is the random walk matrix (A is the adjacency matrix) and p_0 be the initial distribution.

   (a) Prove that there is a probability distribution q on V such that if $p_0 = q$ then $\sum_{i \in S} p_t(i) \geq (1 - \phi(S))^t$.

   (Hint: you may use Q5 of homework 1.)

   (b) Use part (a) to prove that the random walk algorithm in L08 has the following performance guarantee: For $S \subseteq V$ and $\epsilon > 0$, there is a starting vertex $v \in S$ such that the random walk algorithm will return a set S' with $\phi(S') = O(\sqrt{\phi(S)/\epsilon})$ and $|S'| = O(|S|^{1+\epsilon})$.

3. **Hitting Time**

(10 marks) Consider a random walk on a graph $G = (V, E)$ that starts at a vertex $v \in V$, and stops when it reaches s or t. Let $p(v)$ be the probability that if the random walk starts at v then it reaches s before t. Establish a connection between these probabilities and some parameters of an appropriate electric flow problem.

4. **Effective Resistances and Spanning Trees**

(15 marks) Let $G = (V, E)$ be an undirected graph where each edge e has an integral weight w_e. The weight of a spanning tree T is defined as $w_T := \prod_{e \in T} w_e$. Let $p_T = w_T / \sum_{T'} w_{T'}$, where the sum is over all the spanning trees T' of G. Let T^* be a random spanning tree sampled from the distribution p.

(a) Prove that $\Pr(e \in T^*) = w_e R_{\text{eff}}(e)$ for any $e \in E$, where $R_{\text{eff}}(e)$ is the effective resistance of e when every edge e' has resistance $1/w_{e'}$.
(You may use the equation $\det(M + xx^T) = (1 + x^T M^{-1} x) \det(M)$, for any non-singular matrix $M \in \mathbb{R}^{n \times n}$ and $x \in \mathbb{R}^n$.)

(b) Prove that

$$\Pr(e \in T^* \mid f \in T^*) \leq \Pr(e \in T^*),$$

for any two edges $e, f \in E$. In words, conditioned on the event $f \in T^*$, the probability of the event $e \in T^*$ could not increase (i.e. the events $e \in T^*$ and $f \in T^*$ are negatively correlated).

5. **Multiplicative Update**

(10 marks) For the flow LP in L15, prove that if there is an oracle always returning a solution f with the following three properties:

- f satisfies the objective constraint, the non-negativity constraints and the flow conservation constraints;
- $\sum_{e \in E} w_e f_e \leq (1 + \epsilon) \sum_{e \in E} w_e$, where w is an arbitrarily given probability distribution;
- $0 \leq f_e \leq p$,

then the multiplicative weight update method for solving LP described in L14 could solve the flow LP in $O(p \log n / \epsilon^2)$ oracle calls.

6. **Flow Rounding**

(10 marks) Let $G = (V, E)$ be a directed graph and $s, t \in V$ and k a positive integer. A fractional s-t flow solution with value k is an assignment of each edge e to a fractional value $x(e) \in [0, 1]$, satisfying that

$$\sum_{e \in \delta^{\text{out}}(s)} x(e) = \sum_{e \in \delta^{\text{in}}(t)} x(e) = k$$

and

$$\sum_{e \in \delta^{\text{in}}(v)} x(e) = \sum_{e \in \delta^{\text{in}}(v)} x(e) \quad \forall v \in V - \{s, t\},$$

where $\delta^{\text{in}}(v)$ denotes the set of directed edges with v as the head (incoming edges of v), and $\delta^{\text{out}}(v)$ denotes the set of directed edges with v as the tail (outgoing edges of v). Given a fractional s-t flow solution with value k, design a randomized algorithm to return an integral s-t flow with value k in $O(|E| \text{polylog} |V|)$ time, i.e. x satisfies the above constraints and moreover $x(e) \in \{0, 1\}$ for all $e \in E$.

7. **2-lift**

   (10 marks) Prove the lemma in L20 that the spectrum of the adjacency matrix of $\hat{G}$ is equal to the disjoint union of the spectrum of the adjacency matrix of G and the spectrum of the signed matrix A_s.

8. **Tree polynomials**

   (a) (5 marks) Prove that the matching polynomial of a tree is equal to its characteristic polynomial.

   (b) (10 marks) Prove that the maximum eigenvalue of the adjacency matrix of a tree of maximum degree d is at most $2\sqrt{d-1}$.