CMPSCI 711 SPRING 2012: HOMEWORK 3
DUE 2:05 PM, APRIL 16TH

Rules: You may work in groups of at most four. Each group should submit (emailed or handed in at the start of class) one set of typed solutions (remember to include the names of everyone in the group). Cite all sources although work that uses published papers or material from related classes will receive no credit.

Question 1. *Consider the following algorithm that processes a length m stream of values in the range $[n]$. The algorithm maintains a set of at most k "monitored" elements $A \subset [n]$ and with each $a \in A$ we maintain a counter c_a. Initially $A = \emptyset$. When a new value b arrives, the algorithm updates A and the counters as follows:*

(1) *If $b \in A$: Increment c_b.*
(2) *If $b \notin A$:*
  (a) *If $|A| < k$: Add b to A and set $c_b = 1$*
  (b) *Else: Decrement c_a for all $a \in A$ and if c_a become 0, remove a from A*

At the end of the stream, estimate the frequency f_a of $a \in [n]$ to be

$$\tilde{f}_a = \begin{cases} c_a & \text{if } a \in A \\ 0 & \text{if } a \notin A \end{cases} .$$

Prove that $f_a - m/k \leq \tilde{f}_a \leq f_a$ for every $a \in [n]$.

Question 2. *Suppose a stream consists of a sequence of edges and their weights. Design a deterministic semi-streaming algorithm that finds the minimum spanning tree. Remember to prove correctness.*

Question 3. *Suppose a stream consists of a sequence of edges and their weights. Design a deterministic stream algorithm that constructs a data structure that can be used to $(1+\epsilon)(2t-1)$-approximate any distance in the graph.*

Question 4. *Suppose a stream consists of a sequence of edges and their weights. Assume that every weight is a power of 2. Consider the following matching algorithm: $M \leftarrow \emptyset$ and include each new edge e in M if its weight is strictly greater than the total weight of edges currently in M that share an endpoint with e (and remove such edges if necessary). Prove the best approximation guarantee you can for the algorithm.*

Question 5. *Recall that the Laplacian matrix of a graph G where the w_{ij} is the weight of edge (i,j) is the matrix $L \in \mathbb{R}^{n \times n}$ where $L_{ii} = \sum_j w_{ij}$ and $L_{ij} = -w_{ij}$. All other entries are zero. Design a semi-streaming algorithm based on the Laplacian matrix that constructs a data structure which allows you to $(1+\epsilon)$ approximate the size of any cut with high probability. Hint: Use ℓ_1 norm estimation.*