CS648A : Randomized Algorithms
CSE, IIT Kanpur

Assignment 3

Deadline : 6:00 PM, 4th October 2018

Important Guidelines:

1. It is only through the assignments that one learns the most in any course on algorithms. You are advised to refrain from searching for a solution on the net or from a notebook or from other fellow students. **If you are cheating the instructor, you are cheating yourself first**. The onus of learning from a course lies first on you and then on the quality of teaching of the instructor. So act wisely while working on this assignment.

2. Handwritten submissions will not be accepted. You must type your solution using any word processor (For example : Latex, Microsoft Word, Google Doc). You will have to upload the soft copy on moodle before the deadline. You will also have to submit its printed copy before the deadline.

3. The answer of each questions must be formal, complete, and to the point. Do not waste your time writing intuition or idea. There will be penalty if you provide any such unnecessary details.

4. This assignment consists of 3 problems and carries a total of 150 marks.

5. **Hints:**
   Hints to Problem 1 and Problem 2 are given at the last page. Please refrain from the temptation to see them. This class consists of outstanding students who have the potential to do ground breaking research in years to come. Consider these problems as an opportunity to test your perseverance in addition to your analytical creative skills. So ponder over the problems for hours and see the hints only after you have given your best. After seeing the hints, try to make sure you internalize them fully and solve the problem completely.

1 Estimating all-pairs distances exactly

(*marks = 50*)
This problem will test your skills of *random sampling*.
Consider an undirected unweighted graph G on n vertices. For simplicity, assume that G is connected. We are also given a partial distance matrix M : For a pair of vertices i, j the entry $M[i, j]$ stores exact distance if i and j are separated by distance $\leq n/100$, otherwise M stores a symbol # indicating that distance between vertex i and vertex j is greater than $n/100$. Unfortunately, there are $\Theta(n^2)$ # entries in M, i.e., for $\Theta(n^2)$ pairs of vertices, the distance is not known. Design a Monte Carlo algorithm to compute exact distance matrix for G in $O(n^2 \log n)$ time. All entries of the distance matrix have to be correct with probability exceeding $1 - 1/n^2$.

2 Rumour Spreading

(*marks = 50*)
This problem will test your skills of *partitioning an experiment into stages*.
Recall the rumour spreading problem briefly discussed in the class - there is a town of population n and initially only one person knows the rumour. Every day each person who knows the rumour, calls another person selected randomly uniformly from the population and informs him about the rumour. Show that the expected number of days till everyone receives the rumour is $O(\log n)$.

3 Approximate Ham-Sandwich Cut

(*marks = 50*)
There is a set P of $2n$ points in a plane such that no three of them are collinear. Out of the $2n$ points, n points are colored red, and the remaining n points are colored blue. A line L is called a ham-sandwich cut if it simultaneously bisects the red points as well as the blue points, that is, there are at most $n/2$ red (as well as blue) points on each of the two sides of the line.

There is a deterministic $O(n)$ time algorithm for this problem which uses point line duality concept and is quite non-trivial. For all practical purposes, even a slightly weaker version of the ham-sandwich cut, defined below, also works equally well.

A line L is said to be $(1 + \epsilon)$-approximate ham-sandwich cut if the number of red (as well as blue) points on each side of the line L is at most $(1 + \epsilon)n/2$.

Design an $O(n)$ expected time Las Vegas algorithm that computes an $(1 + 1/2)$-approximate ham-sandwich cut of P. You must fully analyse the algorithm.

Hints for some questions

Problem 1: This problem is an application of *"finding a sample with a desired property"*. Note that a BFS tree rooted at a vertex may implicitly store exact distance between potentially many pairs of vertices.

Problem 2: Partition the entire experiment into the following stages and then calculate the expected number of days spent in each stage. Let X be the number of persons who know the rumour at any given time.

- $X < c \log n$.
- $c \log n < X < n/2$.
- $n/2 < X < n$.

Make use of the above hint to show that the number of days for rumor spreading will be $O(\log n)$. Showing it for 1st and the 3rd stage requires elementary tools we discussed in the class. Hint for showing it for the 2nd stage is given below. Please look at it only after you have tried your best.

Hints for 2nd stage:
Establish the following fact and make right use of this fact.
If ν is the number of persons knowing the rumor at the end of some day during 2nd stage, then for some constant $c > 1$ the number of new persons knowing the rumor at the end of the following day will be $c\nu$ with high probability.

Problem 3: (a) Try to realize that there are only $\binom{n}{2}$ distinct cuts.

(b) Use (a) to design a trivial deterministic algorithm that computes exact ham-sandwich cut in polynomial time. Try to get a resonably small exponent that you can. It is easy.

Think hard on the randomized algorithm using the two hints given above before seeing the hints on the next page.

(c) Use random sampling (inspiration from 1/2-approximate median problem we discussed in the first class).

(d) Do rigorous analysis.

(e) Convert from Monte-Carlo to Las-Vegas algorithm.