CS648A : Randomized Algorithms
Semester II, 2021-22, CSE, IIT Kanpur

Theoretical Assignment 2

Deadline : 11:55 PM, 27th March 2022.

Most Important guidelines

- It is only through the assignments that one learns the most about the algorithms and data structures. You are advised to refrain from searching for a solution on the net or from a notebook or from other fellow students. Remember - **Before cheating the instructor, you are cheating yourself**. The onus of learning from a course lies first on you. So act wisely while working on this assignment.
- Refrain from collaborating with the students of other groups. If any evidence is found that confirms copying, the penalty will be very harsh. Refer to the website at the link: https://cse.iitk.ac.in/pages/AntiCheatingPolicy.html regarding the departmental policy on cheating.

General guidelines

1. This assignment is to be done in groups of 2 students. You have to form groups on your own. You are strongly advised not to work alone.
2. The assignment consists of 3 problems. Each problem carries 50 marks.
3. **Naming the file**:
   The submission file has to be given a name that reflects the information about the type of the assignment, the number of the assignment, and the roll numbers of the 2 students of the group. If you are submitting the solution of Theoretical Assignment x, you should name the file as **Theor_x_Rollnumber1_Rollnumber2.pdf**.
4. **Each student of a group** has to upload the same submission file separately. Be careful during the submission of an assignment. Once submitted, it can not be re-submitted.
5. Deadline is strict. Make sure you upload the assignment well in time to avoid last minute rush.

1 Estimating all-pairs distances exactly

(*marks = 50*)
Note: 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 $1 - 1/n^2$.

Application of the Method of Bounded Difference

(*marks= 50*)
Note: This problem will test your skills of Chebyshev's inequality and the method of bounded difference.

Recall the random graph model $G(n, p)$ defined in the programming assignment. Let X denote the number of triangles in $G(n, 1/2)$.

1. Calculate the variance of X. Then use Chebyshev's inequality to derive a bound on $\mathbf{P}[|X - \mathbf{E}[X]| > \mathbf{E}[X]/2]$.
2. Use method of bounded difference to derive the best possible bound on $\mathbf{P}[|X - \mathbf{E}[X]| > \mathbf{E}[X]/2]$.
3. What inference do you draw from (1) and (2).

Reducing diameter of a connected graph by adding $\Theta(n)$ edges

(*marks= 50*)
Note: This problem will test your skills of partitioning an experiment into stages.

Given an undirected connected graph on n vertices. Further, each vertex of the graph adds an edge to every other vertex with probability c/n independently for $c > 0$. Show that, for some large enough constant c, the diameter of this graph is $O(\log n)$ with high probability.