Homework 4
Due Sun Dec 6, 5:00pm

Prof. Amit Chakrabarti
Computer Science Department
Dartmouth College

General Instructions: Please write concisely, but rigorously, and show your calculations explicitly, as we do in class. Each problem is worth 5 points, and only "nearly flawless" solutions will earn full credit.

Honor Principle (Different from earlier homeworks, please read carefully): For this homework, the only sources you are allowed to refer to are the class notes, the website and weblog for this course, and *published* textbooks. You may not refer to *any other* online material. You may not refer to research papers. As before, you are allowed to discuss the problems and exchange solution ideas with your classmates. But when you write up any solutions for submission, you must work alone. *If in doubt, ask the professor for clarification!*

Space lower bounds on data stream algorithms

For this homework, the only streaming model you need to think of is the vanilla (i.e., plain insertions) model. All of the lower bounds you are asked to prove apply even in this simple model. Obviously, as a result, they also apply to the more complicated turnstile models.

For each streaming problem, first think of the appropriate communication problem to reduce from. You have only a small catalogue of problems to try, so it should not be hard to find the right one.

Whenever randomization is allowed in a streaming algorithm, we tacitly assume that an error probability of $1/3$ is allowed.

1. Consider a graph stream describing an n-vertex graph G. Prove that $\Omega(n^2)$ space is required to determine, in one pass, whether or not G contains a triangle, even with randomization allowed.

2. Prove that computing F_2 exactly, in one pass with randomization allowed, requires $\Omega(\min\{m, n\})$ space. [Construct an appropriate "hard stream" of length m, with universe size n, where $m = \Theta(n)$ and show that $\Omega(n)$ space is required on this stream.]

3. Extend the above result to multiple passes, with randomization allowed. The lower bound for p passes should be $\Omega(\min\{m, n\}/p)$.

4. In the Δ-approximate median problem, we are asked to find an element x in a stream S of length m with the property that $|\mathrm{rank}(x, S) - \lfloor m/2 \rfloor| \le \Delta$. Prove that a one-pass randomized streaming algorithm for this problem requires $\Omega(\min\{m, n\}/(1 + \Delta))$ space. [For partial credit, prove this for the special case when $\Delta = 0$.]