CSC2419 W17: Homework Assignment

Due: April 4, beginning of class

April 2, 2017

Guidelines:

- Your assignment solution must be submitted as a *typed* PDF document.
- Submit this assignment at the beginning of the last class, on April 4.
- This is an individual assignment, i.e. you are not supposed to consult other students. You may consult the materials linked to and posted on the course website: Dwork and Roth's monograph, and Vadhan's tutorial notes, our lecture notes.
- Unless stated otherwise, you should justify all your answers using rigorous arguments.

Question 1. (20 marks) Consider a data universe $\mathcal{X} = [N] = \{1, \dots, N\}$ (i.e. each database row is just an integer between 1 and N), and the set $\mathcal{Q} = \{q_1, \dots, q_N\}$ of counting queries defined for any $j \in \mathcal{X}$ by

$$q_i(j) = \begin{cases} 1 & j \le i \\ 0 & \text{otherwise} \end{cases}.$$

Assume that $n \ge \frac{C(1+\log N)^c}{\alpha\varepsilon}$, for C and c large enough absolute constants which are independent of N, n, α, and ε. Describe an ε-differentially private mechanism that, on input a database $x \in \mathcal{X}^n$ (with n as above), outputs answers $y_1, \dots, y_N$ such that, with probability at least $1/2$, $\max_{i=1}^N |y_i - q_i(x)| \le \alpha$. Here, as usual, we define $q_i(x) = \frac{1}{n}\sum_{j=1}^n q_i(x_j)$.

Question 2. (20 marks) We consider a privacy model in which the universe $\mathcal{X}$ is $\binom{[N]}{2}$, i.e. the edges of the complete graph on the vertices $\{1, \dots, N\}$. In other words, a database here is simply a (multi-)graph $G = ([N], E)$. For this question we will assume all databases are simple graphs, i.e. there are no multiedges. We define two databases/graphs $G = ([N], E)$ and $G' = ([N], E')$ to be neighboring if the sets E and E' have symmetric difference of size at most 1. This is a suitable model when analyzing a social network in which friendships are considered private.

We will consider a private algorithm for the minimum cut problem in this model. For a graph $G = ([N], E)$ and a subset of the vertices $S \subseteq [N]$, $S \ne \emptyset, [N]$, let $E(S, \bar{S})$ be the set of edges with exactly one endpoint in S. The size of the minimum cut of G is $\mathrm{OPT}(G) = \min\{|E(S, \bar{S})| : \emptyset \subset S \subset [N]\}$. Assume that $\mathrm{OPT}(G) \ge C(1 + \log N)/\varepsilon$ for a large enough constant C independent of N and ε. Show that a suitable application of the exponential mechanism on input $G = ([N], E)$ outputs a set S, $\emptyset \subset S \subset [N]$, such that, with probability at least $1/2$, $|E(S, \bar{S})| \le \mathrm{OPT}(G) + O(\log(N)/\varepsilon)$. You should set the parameters of the mechanism so that it is ε-differentially private.

NOTE: Because there are exponentially many cuts of G, the standard analysis of the exponential mechanism will not be sufficient. You can use the following theorem, due to Karger: in any connected undirected graph G with N vertices, for any integer $\alpha \ge 1$, the number of cuts $(S, \bar{S})$ such that $|E(S, \bar{S})| \le \alpha \mathrm{OPT}(G)$ is at most $N^{2\alpha}$.

Question 3. (20 marks) Suppose that $\mathcal{M}$ is an algorithm that, on input $x \in \{0,1\}^n$ outputs $x' \in \{0,1\}^n$ such that, with probability at least $2/3$, $\frac{|\{i : x_i \ne x'_i\}|}{n} \le \alpha$. Show that for all sufficiently small ε and δ there exists a constant $c(\varepsilon, \delta) > 0$ such that if $\alpha < c(\varepsilon, \delta)$, then $\mathcal{M}$ is *not* (ε, δ)-differentially private. Give an explicit value for $c(\varepsilon, \delta)$.