CPSC 320: Intermediate Algorithm Design and Analysis
Assignment #6, due Friday, June 17th, 2016 by 2:15pm in Room x235, Box 32

[10] 1. Consider the ϵ-Heavy-Hitters algorithm that we discussed in class, for streams *without* deletions. (For simplicity, assume that the items in the stream are integers.) In class we said that the algorithm processes an entire stream of length n in $O(n \cdot k) = O(n/\epsilon)$ time and $O(1/\epsilon)$ space.

[5] a. Slightly modify the algorithm so that it can process the stream of length n in $O(n)$ worst-case time, while still using $O(1/\epsilon)$ space.

Hints:

- What steps of the algorithm make it slow?
- What techniques can improve the runtime? Maybe better data structures? Maybe better analysis techniques?

[5] b. Implement your new algorithm in your favorite programming language. Download the dataset at `http://www.cs.ubc.ca/~nickhar/S16/Data.txt`, which has one item per line of the file. Use your algorithm to find a set of nine (or fewer) elements containing all $\frac{1}{10}$-heavy-hitters. Your output is allowed to have false positives — it can output some items that are not $\frac{1}{10}$-heavy-hitters. Hand in the code for your program, and the list of nine (or fewer) elements that it outputs.

[9] 2. This problem is about reductions. The textbook uses the notation $Y \leq_P X$ to denote that problem Y has a polynomial-time reduction to problem X. In other words, an arbitrary instance of problem Y can be solved using polynomially many calls to a subroutine that solves X, together with polynomially many additional computations.

Consider the following two problems, both of which we discussed in class:

- A: The Interval Scheduling Problem
- B: The Weighted Interval Scheduling Problem
- C: The Shortest Path Problem Problem
- D: The Longest Path Problem Problem

Remark: The Longest Path Problem is not discussed in the textbook, but it was mentioned in the lectures. It is an NP-complete problem.

[3] a. Is it true that $A \leq_P B$? Explain.

[3] b. Is it true that $B \leq_P A$? Explain.

[3] c. Is it true that $D \leq_P C$? Explain.

[10] 3. Consider the following computational problem.

- The CS department needs a committee to select the department's head. The committee cannot include people who have "conflicts of interest" with each other. The input consists of (a) the desired committee size, (b) a list of all the professors, and (c) a list of all pairs of professors that are conflicted. The goal is to determine whether there's a conflict-free committee of that size.

Remarks: For this question, it is important to understand the definition of NP (in Section 8.3), and the definitions of some of the standard NP-complete problems (Vertex Cover, Set Cover, Independent Set, Hamiltonian Path, etc.)

[5] a. Prove that this problem is in NP.

[5] b. Prove that this problem is NP-complete.

[1] 4. (Bonus) How long did it take you to complete this assignment (not including any time you spent revising your notes before starting)?