Spring 2008 **CMSC 498K: Homework 4** Samir Khuller

Due in class: April 1.

If you cannot come up with algorithms that run in the required time, then provide (correct) slower algorithms for partial credit. Write your answers using *pseudo-code* in the same style as the textbook. These make the algorithm description precise, and easy to read (as opposed to code in C or some other language).

Please also provide a proof of correctness.

(1) Urn A has 5 white and 7 black balls. Urn B has 3 white and 12 black balls. We flip a fair coin. If the outcome is heads, then a ball from urn A is selected, whereas if the outcome is tails, then a ball from urn B is selected. Suppose that a white ball is selected. What is the probability that the coin landed tails?

(2) Suppose we have a universe U of elements. The *similarity* of two sets A, $B \subseteq U$ is defined by $s(A, B) = |A \cap B|/|A \cup B|$. Fix k random orderings $\sigma_1, \ldots, \sigma_k$ of the elements of U, and let the *sketch* of a set $S \subseteq U$ be the vector of size k whose ith component is the element of S that comes first in the ordering σ_i.

Given the sketches of two sets A and B, how can we estimate $s(A, B)$? How large must we make k to be confident that our estimate is fairly accurate? In other words, derive a high-probability bound for the error in the estimate in terms of k. (This is the same kind of bound as the $(1 - \epsilon)F_0 \leq \mathcal{F}_0 \leq (1 + \epsilon)F_0$ bound for the streaming heavy-hitters algorithm.)

(3) You are going to explore Mars, and your friend is going to explore Venus. Once you reach Mars you find life on Mars. You take a DNA sample from this species (think of this as an n bit binary string). Once you are there you would like to send a short "summary" of the string to your friend to check if the species he found in Venus has the same DNA or not. What protocol would you fix so that he can do the check, so that if the species are identical you can confirm this. You may have to allow for the possibility that the species are not actually the same, but you conclude otherwise.

(4) Design an efficient algorithm to check if a given graph has a K_4 (a complete graph on four vertices).