Name: ____

CSE 4502/5717 Big Data Analytics

Exam III; December 5, 2024

Note: You are supposed to give proofs to the time and processor bounds of your algorithms. Read the questions carefully before attempting to solve them.

1. (16 points) Input is a database DB with n transactions from a set $I = \{i_1, i_2, \ldots, i_d\}$ of items. Input also is a threshold $minSupport$ for the minimum support. It is known that each item in I occurs in at most c transactions, where c is a constant. Present an $O(d)$ time algorithm to identify all the frequent items in DB.

2. (17 points) Let DB be a database with n transactions. Let d be the number of possible items. X is a specific k-itemset. The problem is to compute an estimate on $q = \sigma(X)$, i.e., the number of transactions in which X occurs. It is known that $\sigma(X) \geq \sqrt{n}$. Present an $O(k\sqrt{n}\log n)$ time algorithm to come up with an estimate μ such that $0.9q \leq \mu \leq 1.1q$. Show that μ will have this property with a high probability (i.e., a probability of $\geq (1 - n^{-\alpha})$, for any fixed α). Assume that each transaction is given as a bit array as discussed in class.

3. (17 points) Input is a database DB with n transactions from a set $I = \{i_1, i_2, \ldots, i_d\}$ of items. Input also is a threshold *minSupport* for the minimum support. We are required to identify all the frequent k-itemsets, where k is a constant. Present a parallel algorithm for this problem that runs in $O(\log n)$ time. You can use up to $\frac{nd^k}{\log n}$ CREW PRAM processors. Assume that each transaction is given as a bit array.

4. (17 points) Input are k polynomials $f_1(x), f_2(x), \ldots, f_k(x)$ with degrees $d_1, d_2, \ldots, d_k$, respectively, with $\sum_{i=1}^{k} d_i = n$. Present an $O(n \log n \log k)$ time algorithm to compute $\Pi_{i=1}^{k} f_i(x)$.

5. (17 points) A and B are sets of integers in the range $[0, 5n]$. Define $C_i = \{(x, y) : x \in A, y \in B, \& x + y = i\}$, for $i = 0, 1, \ldots, 10n$. Present an $O(n \log n)$ time algorithm to compute $|C_i|$, for $i = 0, 1, \ldots, 10n$.

6. (16 points) Present an $O(n \log^2 n)$ time algorithm to compute all the roots of a given degree-n polynomial $f(x)$. Assume the following: 1) The roots of $f(x)$ are integers in the range $[1, cn]$ where c is a constant; 2) The polynomial is given in coefficients form. (Recall that a is a root of $f(x)$ if $f(a) = 0$.)