INDIAN STATISTICAL INSTITUTE

End Semestral Examination

M. Tech (CS), 2025-2026 (Semester – I)

Algorithms for Big Data

Date: 27.11.2025 | Maximum Marks: 100 | Duration: 3.0 Hours

Note: Answer as much as you can but the maximum you can score is 100.

You should try to write a proof sketch for the correctness of your algorithms.

(Q1) (a) Let σ be a permutation of $\{1, 2, \ldots, n\}$ that is arriving in a streaming fashion. But σ has two numbers missing. Can you design an efficient streaming algorithm that detects the two numbers?

(b) We have a stream $\sigma = <a_1, \ldots, a_m>$, with each $a_i \in [n]$, and this implicitly defines a frequency vector $f = (f_1, \ldots, f_n)$. Note that $f_1 + \ldots + f_n = m$. The FREQUENT problem, with parameter k, is the set $\{j : f_j > m/k\}$. Design an efficient streaming algorithm to solve the FREQUENT problem exactly. How many passes does your algorithm make?

(c) A one-pass algorithm for the FREQUENCY-ESTIMATION problem that achieves an accuracy parameter of ϵ must use $\Omega(\min\{m, n, \epsilon^{-1}\})$ space. In particular, in order to get exact answers, i.e., $\epsilon = 0$, it must use $\Omega(\min\{m, n\})$ space. The FREQUENCY-ESTIMATION problem is to estimate the frequency value of each $a_i \in [n]$ within $1 \pm \epsilon$ accuracy.

[6+6+8=20]

(Q2) (a) Define F_2, the second frequency moment.

(b) Describe, with proper analysis on the quality of the estimate, the AMS algorithm for F_2 estimation.

(c) Show how the median-of-means technique can be applied to obtain an (ϵ, δ) algorithm. Derive the space usage in this case.

[2+10+8=20]

(Q3) (i) Define a t-spanner of an undirected graph G.

(ii) Design and analyze a streaming algorithm for computing a t-spanner of G.

(iii) Design a streaming algorithm to find if G is bipartite.

[2+10+8=20]

(Q4) (i) Show that every n-point metric space (X, d_X) admits an isometric embedding into ℓ_∞^n.

(ii) Given an n-point metric space (X, d_X), design a polynomial-time reduction from determining whether (X, d_X) admits an isometric embedding into ℓ_2^n to the problem of checking whether an $(n-1) \times (n-1)$ matrix is positive semi-definite.

[10 + 10 = 20]