

Comprehensive Exam – Algorithms and Data Structures
Fall, 2006

General directions: Answer 3 of the following 4 questions. Each problem is worth 10 points.

Before giving more detailed descriptions of algorithms, describe in a brief (1-3 sentence) paragraph the main ideas and techniques. Complete pseudo-code is not necessary as long as you clearly specify how the algorithm works. You may use without proof any well-known algorithms and lower bounds for standard problems, as long as you state precisely and correctly the known result.

Give at least an informal proof for all answers; for algorithms, this should include some convincing argument of correctness, and a time analysis. The time complexity of the algorithms you present will be given some weight in your score, as well as your correctness proofs and run-time analysis. For example, a correct $O(n^2)$ algorithm might be worth only 8 or 9 points if there is an $O(n \log n)$ algorithm. The exact weight of efficiency versus correctness is given after the specific problems, if appropriate.

I. Minimum spanning tree of a very sparse graph. Given a weighted graph with n nodes and $n + 10$ edges, show how to find a minimum spanning tree in $O(n)$ time.

II. Clustering. You want to divide n objects into two groups. For each pair of objects $1 \leq i, j \leq n$, you have a similarity value $s(i, j) \geq 0$; ideally, you do not want to split similar objects between groups. Show how to efficiently partition the objects into two nonempty sets $T_1, T_2 \subset \{1, 2, \dots, n\}$, so as to minimize

$$\sum_{i \in T_1, j \in T_2} s(i, j).$$

[5 points correct polynomial time solution, 5 points running time]

III. Optimal drug design. You are working with a protein, represented by a string x of length n over symbols $S = \{\sigma_1, \sigma_2, \dots, \sigma_k\}$ (these symbols correspond to amino acids). You want to design a new protein of the same length (i.e. to choose a string $z \in S^n$) which “binds” well with x . Information about binding comes in the form of an affinity matrix A : for each $1 \leq i, j \leq k$, the affinity $A(\sigma_i, \sigma_j)$ is a measure of how well amino acids σ_i and σ_j interact. Show how to efficiently find the string $z \in S^n$ for which the overall affinity

$$\sum_{i=1}^n A(x_i, z_i) + \sum_{i=2}^n A(z_{i-1}, z_i)$$

is maximized.

[5 points correct polynomial time solution, 5 points running time.]

IV. 2-cover of a tree. In a rooted tree, a node is said to *2-cover* itself, its children, and its grandchildren (to put it differently, a node is *2-covered* by itself, by its parent, and by its grandparent). Given a rooted binary tree $T = (V, E)$, show how to efficiently find a *minimal* set of nodes $S \subset V$ such that every node of V is 2-covered by some element of S .

[5 points correct polynomial time solution, 5 points running time]