

Student ID: _____

Student Name: _____

CompSci 369: Computational Science

Midterm Test 2010

1. Short answer questions:

- (a) Give an example of a very high level programming language and explain why Java is not one of them.

(3 marks)

- (b) How many different (structurally) Steiner/Parsimony trees are there with $n = 3, 5$ and 6 distinct leaves?

(3 marks)

- (c) Name and informally describe three different string matching algorithm.

(3 marks)

(d) Explain the bisection method for numerically finding roots.

(3 marks)

(e) When is a memoized implementation of a dynamic program more suitable than an pure iterative implementation (and vise versa).

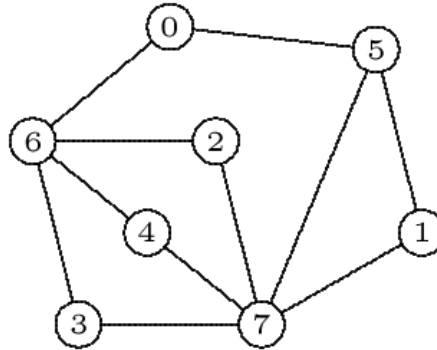
(3 marks)

2. Consider the following **Dominating Set Optimization Problem**.

Input: An undirected graph $G = (V, E)$.

Output: Minimum size of a dominating set $D \subseteq V$ such that each vertex $v \in V$ is either in the set D or has a neighbor in D .

- (a) What is the minimum dominating set size and how many different sets of that size exist for the following graph.



(3 marks)

- (b) Give a simple backtracking algorithm that solves this problem.

(4 marks)

- (c) Help complete an Integer Programming formulation that solves this problem.
Let $x_i = 1$ iff vertex i is in the Dominating Set of $G = (V, E)$.

What is the optimization objective? (Hint: what formula to minimize.)

(2 marks)

What are the optimization constraints? (Hint: one equation for each vertex.)

(2 marks)

- (d) Give a Local Search algorithm that approximates this problem. (Explain the neighborhood and gradient descent aspects similar to the vertex cover local search example given in lectures.)



(4 marks)

3. Computational Biology

- (a) State the two assumptions of the Jukes Cantor Model.

(1 marks)

- (b) Starting with 16 R bases (letters), and a mutation probability of $\frac{1}{4}$ (i.e. the probability of $R \rightarrow Y$ and $Y \rightarrow R$ is $\frac{1}{4}$ in each generation, for each base independently), what is the average number of R 's out of the 16 after 1 generation? And after 3 generations? Justify your answer.

(3 marks)

- (c) What is the “Principle of Optimality”?

(1 marks)

- (d) Explain why the BLOSUM50 matrix contains positive match scores for some entries not on the main diagonal.

(2 marks)

- (e) Align the sequences **ATT** and **TA** using the Needleman&Wunsch global alignment algorithm. Use a match score of **5**, mismatch score of **-3** and a linear gap penalty of **-2**. Clearly indicate the traceback arrows inside the table of scores.

(4 marks)

- (f) Use the UPGMA clustering method to build a phylogenetic tree for the taxa **A,B,C** from the distance matrix below. List the main steps explicitly.

	A	B	C
A	0		
B	2	0	
C	5	7	0



(4 marks)

— scratch paper; will not be marked —