Last Name: First Name: WPI ID:

Artificial Intelligence

Mock Midterm 1 (Max: 150 points)

Note: This is a closed book exam and no cheat sheets are allowed.

Problem 1 (Total: 30 points)

(a) (15 points) Define in your own words the following terms: state, state space, search tree, search node, goal, action, transition model, and branching factor.

(b) (15 points) Consider an 8-puzzle example:

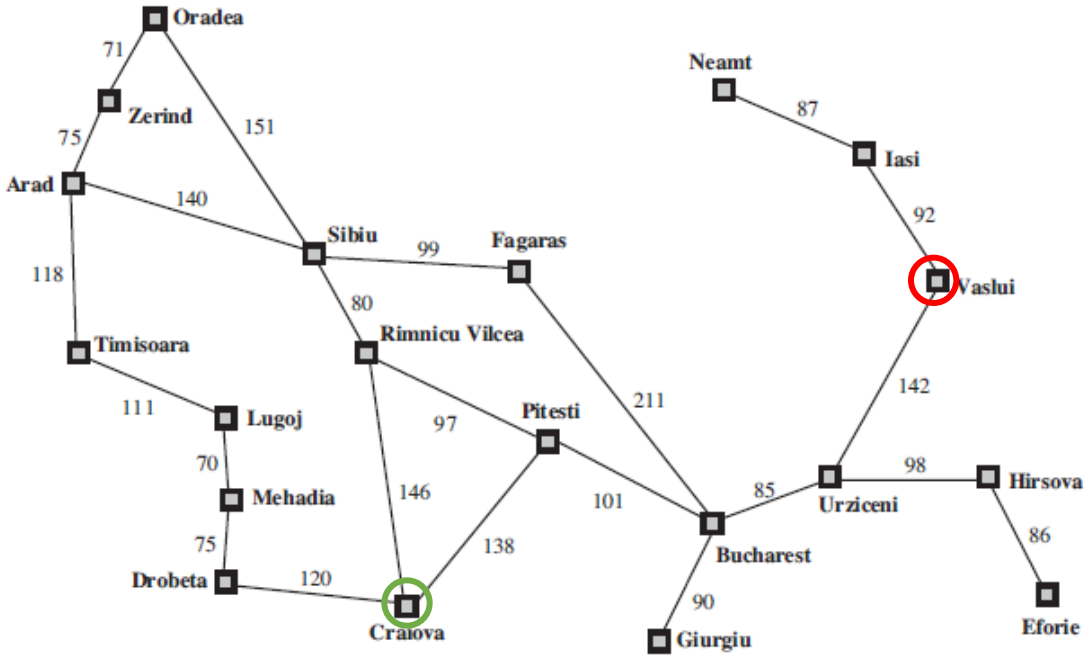
Which of the following is an admissible heuristic and why?

1. $h(n) =$ Number of tiles in wrong position in state n.

2. $h(n) = 0$.

3. $h(n) = max\ (2, h*(n))$, where $h*(n)$ is the true minimal cost to goal from n.

Problem 2 (Total: 30 points)



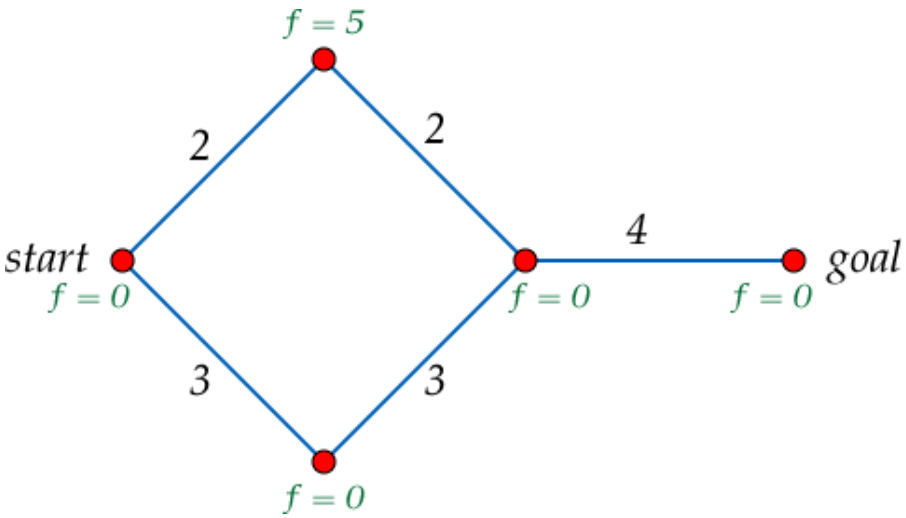
(a) (15 points) Consider the Traveling Salesperson Problem (TSP) which is formulated as , "Given a list of cities, C, and the distances between each pair of cities, $D(i,j)$, what is the shortest possible route that visits each city exactly once and returns to the origin city?"

If two cities are not directly connected by a road, $D(i,j)$ is infinite. It is impractical to enumerate all possible routes, so one must devise better algorithms to solve TSP. TSP is typically defined as a single agent search problem in which optimal solutions correspond to TSP solutions. Formulate the search problem by defining States, Initial State, Successor Function, Goal Test, and Cost Function.

(b) (15 points) Travelling in Hungary: Build a search tree and provide step-by-step breadth-first search for finding a path from Craiova to Vaslui.

Problem 3 (Total: 30 points)

(a) (15 points) Perform a step-by-step A* for the above graph, with heuristics $f(i)$ provided for each node i. Determine if $f()$ is an admissible heuristic.



(b) (15 points) Did it find the optimal path? Explain.

Problem 4 (Total: 30 points)

Consider vectors x1 = [2,5]', x2 = [6,4]', x3 = [5,3]', x4 = [2,2]', x5 = [1,4]'.

(a) (12 points) Assume the dissimilarity measure between a data point *x* and a class *C* to be: $\min_{y \in C} d(x, y)$, where $d(x, y)$ is Euclidean distance. Obtain the clustering using BSAS algorithm with thresholds Θ = 2.5 and q = 3.

(b) (12 points) Obtain the clustering using MBSAS algorithm with thresholds Θ = 2.5 and q = 3.

(c) (6 points) Define a problem of missing data in the feature vectors and propose at least three solutions.

Problem 5 (Total: 30 points)

Consider a dataset defined through the following **dissimilarity** matrix *D*:

	$x1$	$x2$	$x3$	$x4$	$x5$
$x1$	0	9	1	10	3
$x2$		0	11	1	7
$x3$			0	10	4
$x4$				0	7
$x5$					0

(a) (15 points) Obtain the first two levels of agglomerative hierarchical clustering, **K1** and **K2**, using the **single link algorithm**. Level **K0** is: ({x1}, {x2}, {x3}, {x4}, {x5}, {x6}). List explicitly the clustering at each level and the new dissimilarity matrix. Provide necessary explanation.
Hint: The distance of newly formed cluster C_q from the remaining clusters C_s is defined as:

$$d(C_q, C_s) = min\{d(C_i, C_s), d(C_j, C_s)\}$$

(b) (15 points) Elucidate all the differences between single link and complete link algorithms for agglomerative clustering. Also, explain the differences in the dendrograms each algorithm constructs.