Design and Analysis of Algorithm (CS 412)
Instructor: Dr. Ayesha Enayet
Date: _____
CS 6th

SIS ID:_____ Name:_____

Note: Attempt all the questions. Use only blue or black pen.

1. **Can we use Prim's algorithm to find the spanning tree of the unconnected graph? Justify your answer. [Note: you must give a valid justification for your answer to score a point (No partial marking)] [1]**

   No. Prim's algorithm requires the graph to be connected because it relies on the fact that there is always a path from any vertex to any other vertex in the graph. But you can modify Prim's algorithm to find MST of each connected component, you can combine them to form a spanning forest.

2. **Compare the time complexity of Dijkstra's and BFS. Write down the worst-case time complexity of both. Identify at least one significant difference between the two. [1]**

   The time complexity of BFS is O(V+E). The time Complexity of Dijkstra's is O((V + E) log V) with a binary heap or Fibonacci heap. Dijkstra's algorithm is used to find the shortest path from a single source vertex to all other vertices in a weighted graph with non-negative edge weights. BFS is used to traverse or search a graph, typically to find the shortest path in an unweighted graph.

3. **Cluster analysis, also known as clustering, is a statistical technique used in data analysis to organize objects (such as data points, observations, or variables) into groups or clusters based on their similarity. The goal of cluster analysis is to partition the data so that objects within the same cluster are more like each other than they are to objects in different clusters.**

   Given a graph, G={V, E), where vertices represent the data points and edges represent the distance between the points in a 2D plane, devise a greedy algorithm to partition the points into clusters. The algorithm should ensure that points with a distance greater than a specified threshold value, t, belong to separate clusters. [1]

Apply Kruskal's Algorithm, in the resultant MST, remove the edges with a distance greater than t. The resultant graph will have multiple disconnected components (clusters).

I. Sort the edges
II. For each edge if the distance between the edges is greater than t, assign the vertices to the different cluster.

4. **Determine the Huffman code for each alphabet of a file based on their frequencies as provided in the following table.** [0.5]

Char	a	b	c	D	e	F
Frequency (Thousand)	34	23	78	2	9	33
Variable codeword	10	1111	0	11100	11101	110

a) **Calculate the number of bits required to represent the file using fixed length codeword.** [0.25]

537000

b) Calculate the number of bits required to represent the file using a variable length codeword (Huffman code)[0.25]

392000

5. You are given a set of investment options, each with an associated return and risk. Your goal is to invest a limited amount of capital to maximize your returns while effectively managing your risk. Your budget constraint limits the total amount you can invest.
Identify an appropriate approach to find the optimal investment strategy that maximizes returns while adhering to budget and risk constraints. [1]

Answer: The problem is similar to Fractional Knapsack and can be solved using Greedy Algorithm.

Char	a	b	c	D	e	F
Frequency (Thousand)	34	23	78	2	9	33
Variable codeword	10	1111	0	11100	11101	110