Candidate Number

G6017

THE UNIVERSITY OF SUSSEX

BSc SECOND YEAR EXAMINATION
January 2016 (A1)

PROGRAM ANALYSIS

Assessment Period: January 2016 (A1)

DO NOT TURN OVER UNTIL INSTRUCTED TO BY THE CHIEF INVIGILATOR

Candidates should answer TWO questions out of THREE. If all three questions are attempted only the first two answers will be marked.

The time allowed is ONE AND A HALF hours.

Each question is worth 50 marks.

At the end of the examination the question paper and/or answer book, used or unused, will be collected from you before you leave the examination room.

1. (a) Consider the following algorithm.

Algorithm Alg1(X):

Input: a list of reals $X = (x_1, \ldots, x_n)$ where $n \geq 1$
Output: a real

```
1:  t ← 0
2:  p ← 1
3:  for i ← 1 to n do
4:      t ← t + x_i
5:      j ← i
6:      while j ≤ n and x_j ≠ 0 do
7:          p ← p × x_j
8:          j ← j + 1
9:  return p ÷ t
```


i. When *Alg1* is given an input of length n, how many times is line 4 executed? Justify your answer, and consider whether the answer depends on the values within the sequence X. [5 marks]

ii. When *Alg1* is given an input of length n, how many times is line 7 executed? Justify your answer, and consider whether the answer depends on the values in X. [5 marks]

iii. What can you say about the asymptotic running time of *Alg1*? [5 marks]

(b) Suppose that you have been told that *AlgorithmA* has an asymptotic running time that is $\Omega(n^3)$.

i. Precisely what do you know about *AlgorithmA*? [10 marks]

ii. Is it possible that the asymptotic running time of *AlgorithmA* is also $\theta(n^4)$? [10 marks]

(c) Given a list of n reals $X = (x_1, \ldots, x_n)$, let the mean $\bar{x}$, and standard deviation s of the values in X be given by the following formula.

$$\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i$$

$$s = \sqrt{\frac{1}{n-1}\sum_{i=1}^{n}(x_i - \bar{x})^2}$$

i. Complete the algorithm in pseudo-code that solves the problem of calculating the standard deviation of a list X of n reals.

Algorithm *StandardDeviation*(X) : real

Input: a list of reals $X = (x_1, \ldots, x_n)$ where $n \geq 1$
Output: the standard deviation of the values in X

 [10 marks]

ii. Discuss the asymptotic running time of your algorithm. [5 marks]

2. You are a supervisor at a children's summer camp which runs a variety of activities each day, some of which run in parallel. Each evening the children sign up for a selection of the activities on offer the following day. Thus, each morning you know which activities are running that day, how many children are taking part, and when each activity will start and end.

   Your pay for each day is the total of what you earn from each of the activities you supervise. You receive £10 for each activity irrespective of how long it is, or how many children have signed up.

   Fortunately, you have the first opportunity to select the activities you will supervise each day. Your only constraint is that you can only look after one group of children at a time. The lengths of activities can vary considerably, so making a good selection of activities can make a substantial difference to your income. This question concerns the problem of determining the best choice of activities for you to supervise each day.

   (a) Describe a greedy method that could be applied to this problem, but which is not guaranteed to always produce an optimal solution. Give an example of a problem instance where the greedy method you describe would not produce an optimal solution. [10 marks]

   (b) Describe a greedy method that could be applied to this problem that is guaranteed to always produce an optimal solution. [5 marks]

   (c) Discuss the asymptotic running time of an algorithm that is based on the greedy method that you have described in answer to Question 2b. [10 marks]

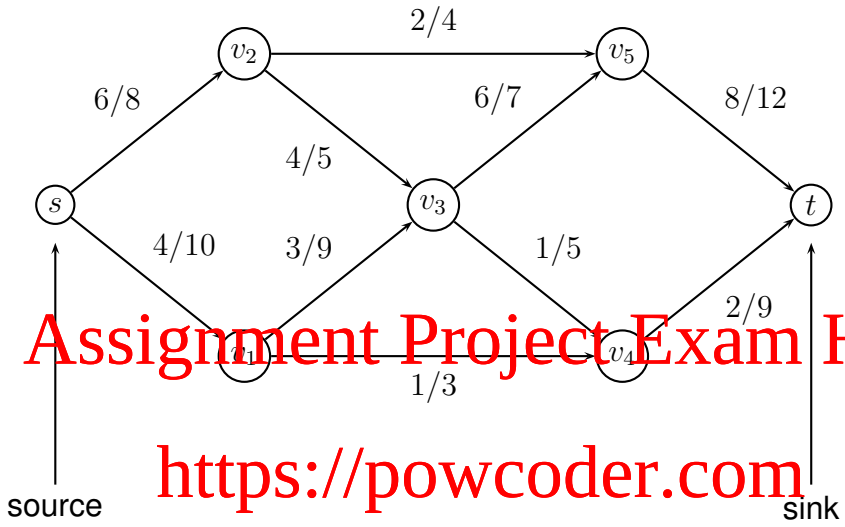
   (d) Suppose that we changed the scenario, so that the amount you are paid for running an activity is determined by the size of the group and the length of the session. In particular, the value in pounds for an activity is the number of hours that the activity lasts multiplied by the number of children taking part. So for a two hour session involving eight children you would be paid £16.

   Show that the greedy method that you used in the algorithm given in answer to Question 2b would *not* produce an optimal solution to this revised problem. [10 marks]

   (e) Describe in words an algorithm that would find the optimal solution to the problem that arises in the revised scenario described in Question 2d. [10 marks]

   (f) Discuss the asymptotic running time of the algorithm that you have given in answer to Question 2e. [5 marks]

3. Consider the following network and flow rate. Note that an edge labelled x/y indicates the rate of flow along the edge is x units of flow, and that the capacity of the edge is y units of flow. For example, the edge from s to v_1 has a capacity of 10 and a flow rate of 4.



(a) Describe a simple method that can be used to determine the amount of flow that runs from the source to the sink in a network, and use this method to determine the flow shown in the above network. [5 marks]

(b) Draw the residual graph that is produced from the above network and flow rate. Make sure that you clearly distinguish between forward and backwards edges. [10 marks]

(c) Consider the path $(s, v_1, v_3, v_2, v_5, t)$ in the residual graph you have given in answer to Question 3b. This path includes one backward edge which runs from v_3 to v_2. Explain why backward edges are included in residual graphs. [5 marks]

(d) What is the significance of the bottleneck edge of a path in a residual graph? In the residual graph you have given in answer to Question 3b what is the bottleneck edge of the path $(s, v_1, v_3, v_2, v_5, t)$? [5 marks]

(e) Show the network and flow that results from augmenting the flow of the network using the path $(s, v_1, v_3, v_2, v_5, t)$ in the residual graph you have given in answer to Quesiton 3b. [10 marks]

(f) Residual graphs can have more than one source to sink path. Discuss whether the decision as to which order these paths are chosen for flow augmentation can have an impact either on the maximum flow rate that is ultimately established, or on the length of time that the algorithm takes to run. [10 marks]

(g) Explain why an algorithm that solves the network flow problem might be useful for graph clustering. Graph clustering involves grouping the vertices in a weighted graph, where weights on edges correspond to a measure of how closely vertices are associated. The goal is to find a way to cluster the vertices into groups of vertices such that vertices in the same group are closely associated. [5 marks]