THE AUSTRALIAN NATIONAL UNIVERSITY

Second Semester 2008

COMP1140
(Data Structures and Algorithms)

Writing Period: 3 hours duration

Study Period: 15 minutes duration

Permitted Materials: None

Answer ALL Questions

All your answers must be written in the boxes provided in this booklet. You may be provided with scrap paper for working, but it must **not** *be used to write final answers.*

There is additional space at the end of the booklet in case the boxes provided are insufficient. Label such overflow boxes with the question number.

Do not remove this booklet from the examination room.

Name:

Student Number:

Official use only:

Q1 (15)	Q2 (20)	Q3 (25)	Q4 (20)	Q5 (20)	Total (100)

QUESTION 1 [15 marks]

(a) Suppose arrays A and B are both sorted and both contain n elements. Give an $O(\log n)$ algorithm to find the median of $A \cup B$.

QUESTION 1(a)	**[6 marks]**

(b) Given an integer sequence $10, 4, 18, 23, 27, 15, 28, 27, 12, 9, 11, 16$, which is stored in an array $A[1..12]$.

(i) Construct a binary min-heap for the numbers in the sequence.

QUESTION 1(b)(i)	**[4 marks]**

(ii) Show that there are at most $\lceil n/2^{h+1} \rceil$ nodes of height h in any n-element heap.

QUESTION 1(b)(ii)	**[5 marks]**

QUESTION 1(b)(ii) (continued)

QUESTION 2 [20 marks]

(a) In how many ways can a committee of 5 be formed from 6 men and 4 women candidates if

(i) there are no restrictions,

QUESTION 2(a)(i) **[2 marks]**

(ii) there must be more women than men?

QUESTION 2(a)(ii) **[2 marks]**

(b) How many 8-bit bytes (strings of 0s and 1s) contain

(i) exactly three 1s,

QUESTION 2(b)(i)	**[2 marks]**

(ii) at most six 1s?

QUESTION 2(b)(ii)	**[2 marks]**

(c) In how many ways can eight identical balls be distributed into three containers so that

(i) no container is empty,

QUESTION 2(c)(i)	**[2 marks]**

(ii) the third container contains an even number of balls?

QUESTION 2(c)(ii)	**[2 marks]**

(d) What's the probability that

(i) there's at least one 6 out of three throws of a fair 6-sided die,

QUESTION 2(d)(i)	**[2 marks]**

(ii) two cards dealt randomly from a standard 4-suit, 52-card deck are of the same rank (i.e., a pair)?

QUESTION 2(d)(ii)	**[2 marks]**

(e) In a given population $1/100$ of the people are X-positive. A test for X comes back positive with probability 0.8 when administered on a person who is X-positive, and with probability 0.1 otherwise. What's the probability that

(i) a test on a random person comes back positive,

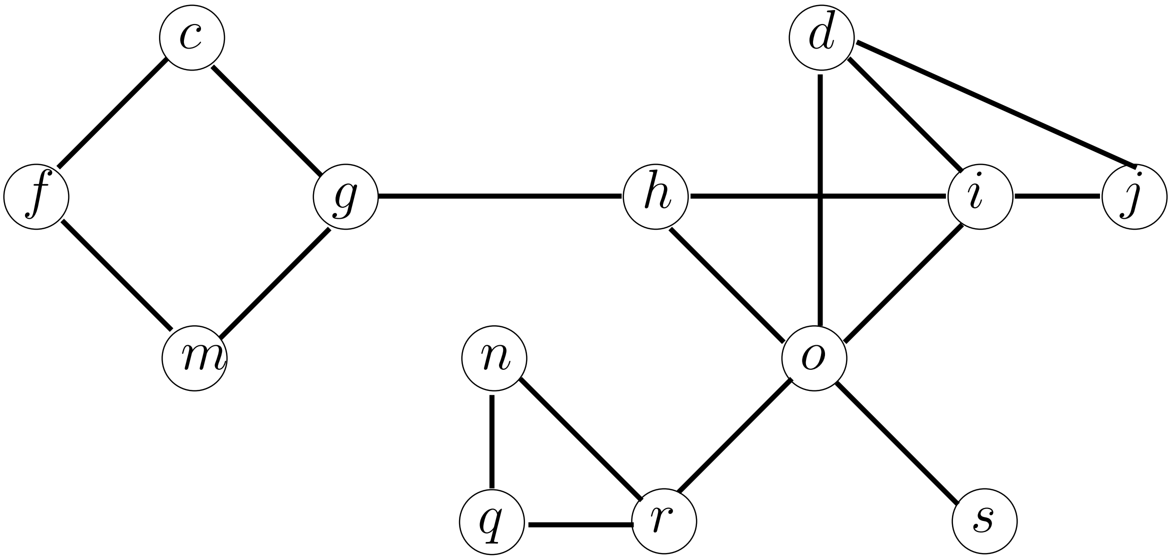
QUESTION 2(e)(i)	**[2 marks]**

(ii) a random person is X-positive given that their test comes back positive?

QUESTION 2(e)(ii) **[2 marks]**

QUESTION 3 [25 marks]

Parts (a) and (b) of this question refer to the following graph G:



(a) Apply BFS on the graph G to find all the neighbors v of vertex h such that $d(h, v) = 2$. That is, find all the vertices whose distance from h is 2.

(i) Explain when and why the algorithm may be terminated before BFS is completed;

QUESTION 3(a)(i)	**[3 marks]**

(ii) Show the partial BFS tree as generated up to the termination point in (a);

QUESTION 3(a)(ii)	**[2 marks]**

(iii) List all the vertices v in G such that $d(h, v) = 2$.

QUESTION 3(a)(iii) **[2 marks]**

(b) Apply DFS on G, starting at vertex h.

(i) Show the DFS tree;

QUESTION 3(b)(i) **[3 marks]**

(ii) For each vertex v, indicate its discovery time $d[v]$ and its lowpoint $L[v]$;

QUESTION 3(b)(ii) **[4 marks]**

(iii) Give a characterization of separating vertices in terms of discovery time and low-point, and list all separating vertices of G.

QUESTION 3(b)(iii) **[4 marks]**

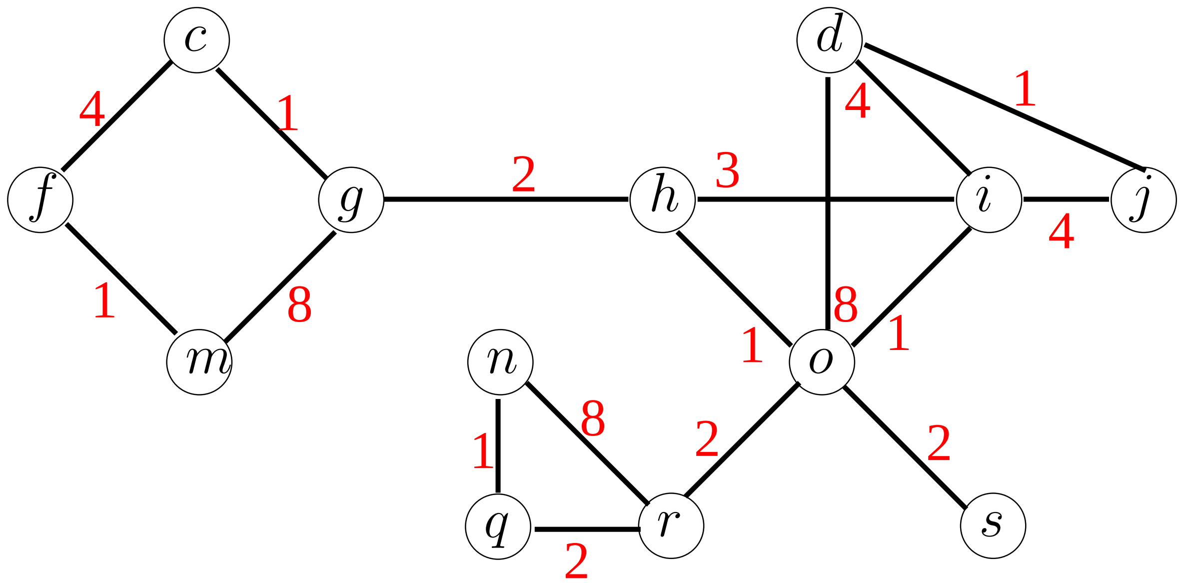
(iv) Give a method to find the biconnected components of a graph. Apply it to find the biconnected components of the subgraph of G with vertex set $\{c, f, g, h, m\}$: you will perform DFS on this subgraph, starting at vertex h.

Reminders:

A *separating vertex* of a graph G is a vertex whose removal disconnects G. For a graph G, let $d[u]$ be the DFS discovery time of u for each vertex u in G. For a given vertex v, the *lowpoint* of v is the least number $d[u]$ of a vertex u which can be reached from v via a possibly empty path consisting of zero or more tree edges moving away from the root followed by at most one back edge. A *biconnected component* of a graph G is a maximal set of vertices C such that $\forall u, v \in C$ there exists two vertex-disjoint paths in C from u to v. For a graph G, let $d[u]$ be the DFS discovery time of u for each vertex u in G.

QUESTION 3(b)(iv) **[3 marks]**

(c) Show the minimum spanning tree found by Prim's algorithm applied to the weighted graph below, starting at vertex h.



(i) Show the order in which the edges are added to the minimum spanning tree as the algorithm progresses;

QUESTION 3(c)(i)	**[2 marks]**

(ii) For each edge added, and for each vertex v, show the current distance of v to the minimum spanning tree.

QUESTION 3(c)(ii)	**[2 marks]**

QUESTION 4 [20 marks]

(a) Let $*$ and $|$ be two binary connectives defined below:

x	y	$x * y$	$x\|y$
0	0	1	0
0	1	0	1
1	0	0	1
1	1	1	0

Show that the set $\{*, |, \vee\}$ is a complete set of connectives. Recall that a set of connectives C is complete if every boolean function can be represented as a formula built using only propositional variables and the connectives in C. You can use the fact that $\{\neg, \vee\}$ is complete.

QUESTION 4(a)	**[4 marks]**

(b) Let G be the formula

$$(x_1 \wedge x_2) \vee (x_1 \wedge x_3) \vee (\neg x_4 \wedge x_5) \vee (x_4 \wedge x_6).$$

Assume that the variables are ordered according to their indices, that is, $x_i < x_j$ if and only if $i < j$.

(i) Give the reduced BDD for G.

QUESTION 4(b)(i)	**[4 marks]**

(ii) Give the reduced BDD for $\neg G$.

QUESTION 4(b)(ii)	**[1 mark]**

(iii) In general, given a reduced BDD of F, how do you compute the reduced BDD of $\neg F$, without going through the reduction steps?

QUESTION 4(b)(iii)	**[1 mark]**

(c) In most SAT solvers, it is common to perform a pre-processing phase on input clauses, to remove trivial clauses and reduce the number of variables. One such pre-processing step is defined as follows:

> Let Δ be the input set of clauses. If there exists a clause $\alpha \in \Delta$ such that there exists a literal $p \in \alpha$ such that neither p nor the NNF of $\neg p$ appear in $\Delta \setminus \{\alpha\}$, then return $\Delta \setminus \{\alpha\}$. Otherwise, return Δ.

Let Γ be a set obtained by one application of the above pre-processing step to Δ. Prove that Γ is satisfiable if and only if Δ is satisfiable.

QUESTION 4(c)	**[5 marks]**

(d) Call a resolution rule a *two-resolution* if one of the clauses to be resolved is a set containing at most two literals. Just as unit resolution, two-resolution is not complete. Prove this by giving a set of clauses, which is unsatisfiable, but which cannot be refuted using two-resolutions alone.

QUESTION 4(d) **[5 marks]**

QUESTION 5 [20 marks]

The government of Tasmanistan wants to set up a new health insurance database and employs you to design an appropriate database system. Here are some key facts about the health care system TMCare in Tasmanistan:

- TMCare provides unlimited health cover to all citizens.
- All citizens pay 10% of their income to TMCare.
- For every visit to a GP a patient has to pay a fee of $10. GPs can refer patients to specialists for no additional charge.
- Every non-referred visit to a specialist costs a patient a fee of $100.
- TMCare pays GPs and specialists the difference between the actual costs and the fee paid by the patient.

The database system you design needs to store the following information:

- citizens: Name, DOB, suburb, income, accumulated actual TMCare cost
- doctors: Name, DOB, suburb, expertise
- GP visits: patient, doctor, referral to specialist, fee paid, actual cost of visit
- specialist visits: patient, doctor, referred by, fee paid, actual cost of visit

(a) TMCare is interested in the following queries. Formulate them in relation algebra. You can define additional operators if necessary by explaining what they do.

(i) What is the total income of all citizens per suburb, and the total income of all doctors per suburb?

QUESTION 5(a)(i)	**[1 mark]**

(ii) What is the difference between the total actual costs TMCare has to pay and the total fees paid by the citizens (percentage of income plus fees per visit)?

QUESTION 5(a)(ii) **[2 marks]**

(iii) For doctors, what area of expertise gets the most patients and what area of expertise earns most?

QUESTION 5(a)(iii) **[2 marks]**

(iv) What is the total cost to TMCare each GP causes (own costs plus costs of specialist visits the GP refers)?

QUESTION 5(a)(iv) **[2 marks]**

(v) List all specialist visits where the referring GP is in a different suburb than the specialist

QUESTION 5(a)(v) **[2 marks]**

(b) Suggest what attributes, if any, could serve as keys for your relations from question 1.

QUESTION 5(b) **[2 marks]**

(c) Make a change to the database scheme so that at least one query is faster while the runtime for the other queries does not increase. Please justify briefly.

QUESTION 5(c) **[2 marks]**

(d) What attribute(s) could be added/changed so that the information contained in the above database scheme would be practically more useful? Please justify briefly.

QUESTION 5(d)	**[2 marks]**

(e) What are the keys for the relations mentioned above?

QUESTION 5(e)	**[1 mark]**

(f) What is best suited as primary index for the relations mentioned above taking into account the given queries? Please justify briefly.

QUESTION 5(f)	**[1 mark]**

(g) Prove or disprove the following equality:

$$((R \bowtie_{A=B} T) \bowtie_{C=D} S) \equiv (R \bowtie_{A=B} (T \bowtie_{C=D} S))$$

where A is an attribute of R, B, C are different attributes of T, and D is an attribute of S.

QUESTION 5(g) **[3 marks]**

Additional answers. Clearly indicate the corresponding question and part.

Additional answers. Clearly indicate the corresponding question and part.

Additional answers. Clearly indicate the corresponding question and part.

Additional answers. Clearly indicate the corresponding question and part.

Additional answers. Clearly indicate the corresponding question and part.