COMS20010 — Mock Exam

Short answer questions

1. (5 marks) (Short question.) Consider the **unoptimised** version of the union-find data structure on a 1000-element set, in which a sequence of n operations takes $\Theta(n \log n)$ time in the worst case. Suppose this data structure is freshly-initialised, so that which each element is currently in its own (singleton) set. What is the least number of `Union` operations that could result in a component of the data structure having depth 4?

   A. 15.

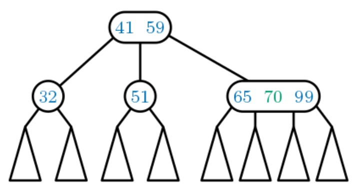
   B. 16.

   C. 31.

   D. 32.

   E. None of the above.

2. (5 marks) (Short question.) Consider the following 2-3-4 tree. What will the root contain and how many children will it have after we insert 72 into the tree?



   A. The root will contain $41, 59$ and will have three children.

   B. The root will contain $41, 59, 99$ and will have four children.

   C. The root will contain $59, 70$ and will have three children.

   D. The root will contain $41, 59, 65$ and will have three children.

   E. The root will contain $41, 59, 70$ and will have four children.

3. (5 marks) (Short question.) Mark each of the following statements true or false.

   (a) (1 mark) $40^n \in o(20^n)$.

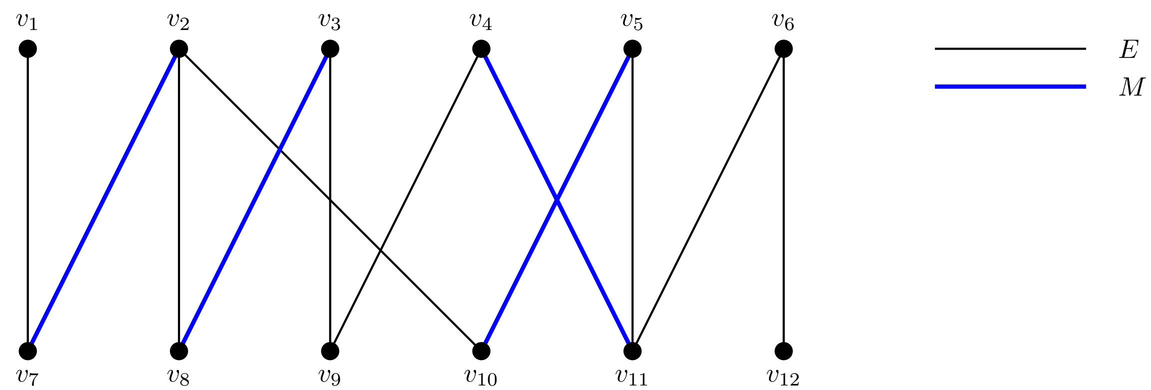
   (b) (1 mark) $40^n \in O(20^n)$.

   (c) (1 mark) $40^n \in \Theta(20^n)$.

   (d) (1 mark) $40^n \in \Omega(20^n)$.

   (e) (1 mark) $40^n \in \omega(20^n)$.

4. (5 marks) (Short question.) Consider the bipartite graph $G = (V, E)$ and indicated matching M below.



Which of the following is **not** an augmenting path?

A. $v_1v_7v_2v_8v_3v_9$.

B. $v_9v_4v_{11}v_5v_{10}v_2v_7$.

C. v_6v_{12}.

D. $v_9v_4v_{11}v_6$.

E. Either they are all augmenting paths, or more than one of them is not an augmenting path.

5. (5 marks) (Short question.) Which of the following logical formulae is **not** in conjunctive normal form (CNF)?

A. $(a \vee b \vee c) \wedge (c \vee \neg d)$.

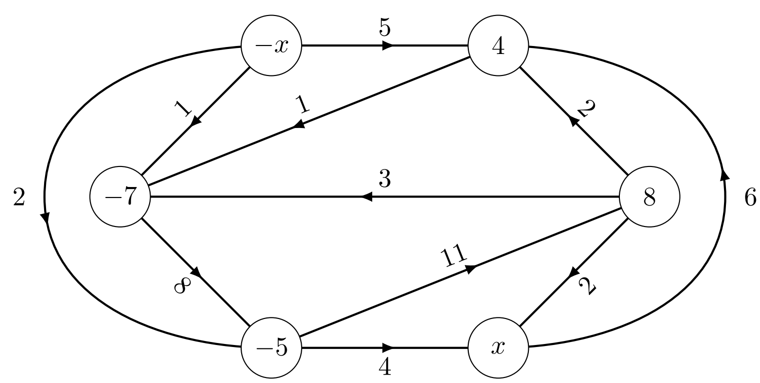
B. $a \wedge b$.

C. $a \vee b$.

D. $(a \vee \neg b) \wedge \neg(c \vee d)$.

E. Either they are all in CNF, or more than one of them is not in CNF.

6. (5 marks) (Short question.) Consider the following circulation network.



What is the largest integer value of x such that the network has a valid circulation?

A. 5.

B. 6.

C. 7.

D. There's no valid circulation for any integer value of x.

E. None of the above.

7. (5 marks) (Short question.) Consider an instance of weighted interval scheduling with interval set $\mathcal{R} = [(1,3),(2,5),(4,9),(0,9),(7,10),(8,10),(11,13),(9,14),(12,15)]$ and weight function w given by

$$\begin{aligned} &w(1,3)=6, \quad &&w(2,5)=6, \quad &&w(4,9)=4, \quad &&w(0,9)=10, \quad &&w(7,10)=3,\\ &w(8,10)=3, \quad &&w(11,13)=8, \quad &&w(9,14)=9, \quad &&w(12,15)=4. \end{aligned}$$

What is the maximum possible weight of a compatible set?

A. 19.

B. 20.

C. 21.

D. 22.

E. None of the above.

8. (5 marks) (Short question.) Mark the following five statements true or false.

(a) (1 mark) In all directed graphs $G=(V,E)$, we have $|E| = \sum_{v \in V} d^+(v)$.

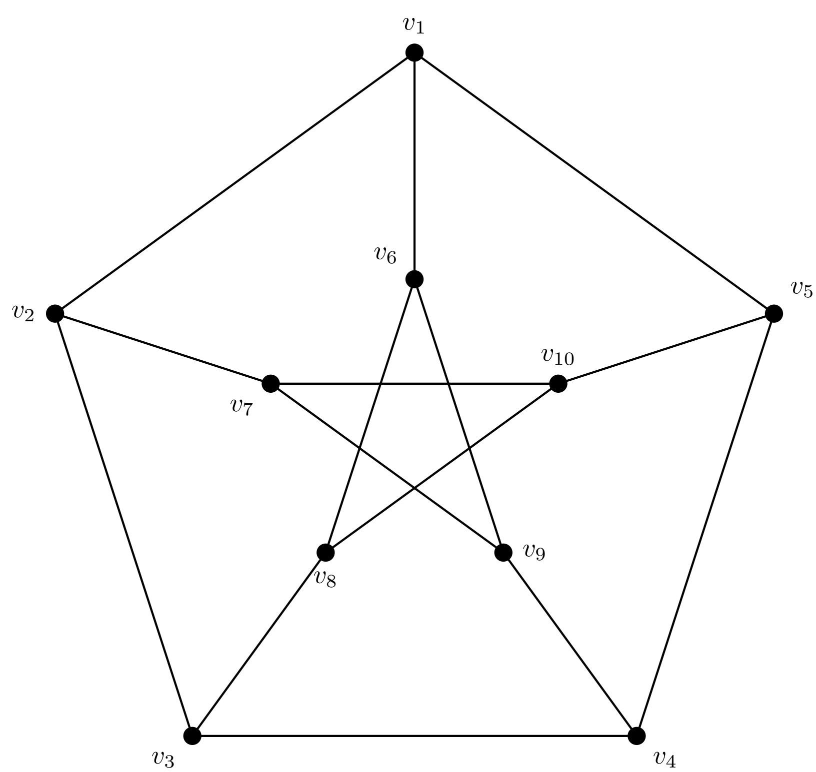
(b) (1 mark) All forests are connected.

(c) (1 mark) All induced subgraphs of trees are trees.

(d) (1 mark) Any (unrooted) tree with at least two vertices has at least two leaves.

(e) (1 mark) Any tree with n vertices has $n-1$ edges.

9. (5 marks) (Short question.) Consider the graph below.



Are each of the following sets vertex covers or not?

(a) (1 mark) $\{v_i : i \in [10]\}$;

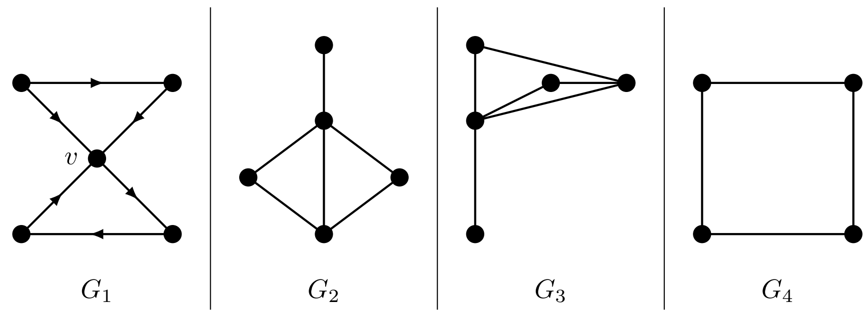
(b) (1 mark) $\{v_1, v_3, v_7, v_{10}\}$.

(c) (1 mark) $\{v_i : i \in [5]\}$;

(d) (1 mark) $\{v_1, v_3, v_4, v_7, v_8, v_{10}\}$.

(e) (1 mark) $\{v_2, v_4, v_5, v_6, v_7, v_8\}$.

10. (5 marks) (Short question.) We define directed and undirected graphs G_1, G_2, G_3 and G_4 as in the picture below.



Mark each statement true or false.

(a) (1 mark) The in-degree of v in G_1 is 3.

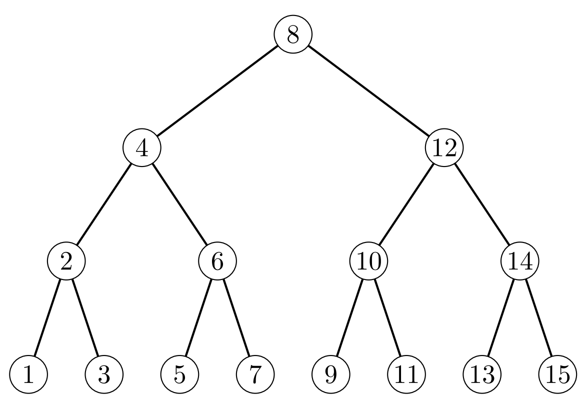
(b) (1 mark) G_1 is weakly connected.

(c) (1 mark) G_2 and G_3 are isomorphic.

(d) (1 mark) Some induced subgraph of G_3 is isomorphic to G_4.

(e) (1 mark) G_2 and G_4 each contain an Euler walk.

11. (5 marks) (Medium question.) Consider the effects of deleting 4 from the 2-3-4 tree below.



After deleting 4:

(a) (1 mark) How many 2-nodes are in the tree?

(b) (1 mark) How many 3-nodes are in the tree?

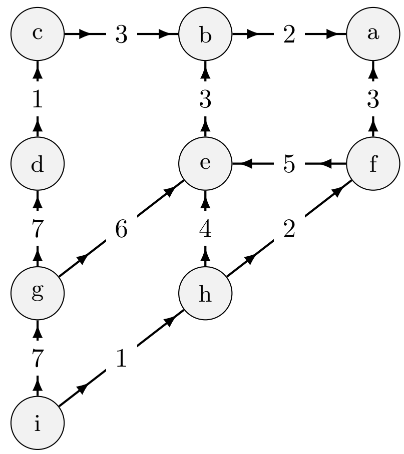
(c) (1 mark) How many 4-nodes are in the tree?

(d) (1 mark) How many nodes are in the bottom layer of the tree?

(e) (1 mark) How many total fuse, split and transfer operations occurred in the process of deleting 4?

This question is multiple choice. The available options for each part are 0, 1, 2, 3, 7, 8, 9, and "none of the above".

12. (5 marks) (Medium question.) Suppose Dijkstra's algorithm is run on the following graph starting at vertex i. List the vertices in the order the algorithm performs final processing on them (i.e. the order in which their distance from i is finalised).



A. i, h, b, g, e, a, d, f, c

B. i, h, f, e, b, a, g, d, c

C. i, h, f, e, a, g, b, c, d

D. i, h, g, e, a, b, f, d, c

E. i, h, f, e, a, g, b, d, c

13. (10 marks) (Medium question.) Suppose that x, y, and z are strings. We say that z is a "shuffle" of x and y if z can be obtained by mixing the characters from x and y in a way that preserves the left-to-right ordering of the characters from x and the characters from y. For example, "LOhamLbuCrgerAT" is a shuffle of "hamburger" and "LOLCAT". Given three strings x, y, and z, you wish to tell whether or not z is a shuffle of x and y.

(a) (8 marks) Fill in the blanks in the incomplete dynamic programming algorithm below. The return value of SHUFFLED(x, y, z) should be `True` if z is a shuffle of x and y, and `False` if it is not.

begin
- **Let** $n \leftarrow \text{length}(x)$.
- **Let** $m \leftarrow \text{length}(y)$.
- **Let** $r \leftarrow \text{length}(z)$.
- **if** $n = 0$ **then**
  - **If** $y == z$ **return** ____, **else return** ____.
- **else if** $m = 0$ **then**
  - **If** $x == z$ **return** ____, **else return** ____.
- **else**
  - **Let** $x^- \leftarrow x[1, \ldots, n-1]$.
  - **Let** $y^- \leftarrow y[1, \ldots, m-1]$.
  - **Let** $z^- \leftarrow z[1, \ldots, r-1]$.
  - **Return** $((z[r-1] == x[n-1]) \wedge \text{SHUFFLED}(\_\_, \_\_, z_1)) \vee ((z[r-1] == y[m-1]) \wedge \text{SHUFFLED}(\_\_, \_\_, z_1))$

Each blank should contain `True`, `False`, x, y, z, x_1, y_1 or z_1. Some terms may appear more than once or not at all.

(b) (2 marks) What is the running time of SHUFFLED, if properly memoised? Choose the strongest bound that applies.

A. $O(m+n)$
B. $O(n^2)$.
C. $O(mn)$
D. $O(mnr)$.
E. None of the above.

14. (5 marks) (Medium question.) Mark the following five statements true or false.
    (a) (1 mark) The problem of finding a minimum vertex cover in a given input graph is in NP.
    (b) (1 mark) Every Co-NP-complete problem is the complement of an NP-complete problem (under Karp reductions).
    (c) (1 mark) There is a Karp reduction from every problem in NP to 3-SAT.
    (d) (1 mark) Every decision problem with a polynomial-time algorithm is in NP.
    (e) (1 mark) If a problem is NP-complete under Karp reductions, then it is NP-complete under Cook reductions.

Long answer questions

15. (10 marks) (a) (5 marks) (Short question.) Give an example of an undirected graph with a closed Euler walk but no Hamilton cycle. You do not need to justify your answer.
    (b) (5 marks) (Short question.) Give an example of an undirected graph with a Hamilton cycle but no closed Euler walk. You do not need to justify your answer.

16. (5 marks) (Short question.) The *Bacon number* of an actor is defined as follows. Kevin Bacon has a Bacon number of zero. The Bacon number of another actor is the shortest chain of co-stars linking them to Kevin Bacon. For example, Elvis Presley was in Change of Habit with Edward Asner, and Edward Asner was in JFK with Kevin Bacon, so Elvis Presley has a Bacon number of at most 2. Elvis was never in any movies with Kevin directly, so he has a Bacon number of exactly 2. Given API access to IMDB, **briefly summarise** an efficient algorithm to output the Bacon number of a given actor by reducing to a problem solved in the course. You do not need to prove your algorithm works.

17. (10 marks) (Short question.) You are running a cattery. You have a large number of cats, and you wish to feed them a diet which meets all their nutritional needs while spending as little money as possible. You are given a list of store-brand foods and supplements, along with their prices and nutritional content per kilo. Express this as a linear programming problem by filling in the blanks below. You may assume all nutritional needs are lower bounds (e.g. "needs at least 200g protein per day").

    The linear programming problem will be of the form

$$\begin{aligned} c_1x_1 + \cdots + c_tx_t &\to \min, \\ A\vec{x} &\geq \vec{b}, \\ \vec{x} &\geq \vec{0}. \end{aligned}$$

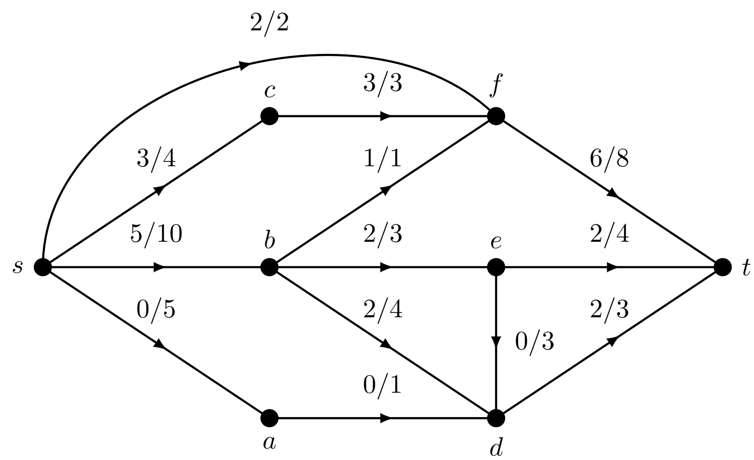
    Here $x_1, \ldots, x_t$ represent ____, $c_1, \ldots, c_t$ are given by ____, the rows of A are given by ____, and $\vec{b}$ is given by ____. This problem ____ an optimal solution.

    Each blank will be filled by one of the following:

    - the costs of each food or supplement per kilo;
    - the nutritional content of each food per kilo;
    - the amounts of each food to give the cats in kilos per day;

- the total amount of each nutrient the cats require per day;
- always has;
- may or may not have;
- never has.

18. (5 marks) (Short question.) Give a list of augmenting paths from s to t in the below flow network, with respect to the given flow.



19. (a) (10 marks) (Medium question.) Consider the following greedy algorithm to find a large independent set in a graph.

Algorithm: GreedyIS(G, k)

Input : A graph G.
Output: An independent set of G.

begin
  Let `output` $\leftarrow \emptyset$.
  foreach $v \in V(G)$ **do**
    if `output` $\cup \{v\}$ *is an independent set* **then**
      `output` $\leftarrow$ `output` $\cup \{v\}$.
  Return `output`.

Fix an integer $\Delta > 0$. If G has n vertices and maximum degree Δ, then how large an independent set is GreedyIS **guaranteed** to output? Your answer should include both an upper bound and a lower bound, and a brief justification of each.

(b) (5 marks) (Medium question.) Now consider the following refinement of GreedyIS.

Algorithm: BetterGreedyIS(G, k)

Input : A graph G.
Output: An independent set of G.

begin
  Sort $V(G)$ in increasing order of degree. Write $V(G) = \{v_1, \ldots, v_n\}$, with $d(v_1) \leq \cdots \leq d(v_n)$.
  Let `output` $\leftarrow \emptyset$.
  for $i = 1$ *to* n **do**
    if `output` $\cup \{v_i\}$ *is an independent set* **then**
      `output` $\leftarrow$ `output` $\cup \{v_i\}$.
  Return `output`.

Prove that BetterGreedyIS may still output an independent set of size $O(1)$ even if G contains an independent set of size $\Omega(n)$.

20. (5 marks) (Medium question.) Let (G, w) be a connected weighted graph with non-negative weights, and let T be result of running Kruskal's algorithm on G. Let $w'(x, y) = \log(w(x, y) + 1)$ for all edges $\{x, y\} \in E(G)$. Is T a minimum spanning tree of (G, w')? Briefly explain your answer.

21. (5 marks) (Medium question.) The Travelling Salesman Problem (TSP) is defined as follows. We are given a list of n cities and, for each unordered pair $\{i, j\}$ of distinct cities, the cost $c(i, j)$ of travelling between i and j. (These costs can be arbitrary positive integers.) We are also given an integer k. We must output `Yes` if there is a way to travel to each city **exactly once**, then return back to the starting point, with total cost at most k. We call this a *round trip*. Otherwise, we must output `No`.

    As an example, the input may be the set of cities $\{\text{Amsterdam}, \text{Baghdad}, \text{Cairo}, \text{Dublin}\}$, the cost function

    $$c(\text{Amsterdam}, \text{Baghdad}) = 175, \quad c(\text{Amsterdam}, \text{Cairo}) = 95, \quad c(\text{Amsterdam}, \text{Dublin}) = 24,$$
    $$c(\text{Baghdad}, \text{Cairo}) = 140, \quad c(\text{Baghdad}, \text{Dublin}) = 250, \quad c(\text{Cairo}, \text{Dublin}) = 122,$$

    and the integer $k = 500$. The round trip from Amsterdam to Baghdad to Cairo to Dublin and back to Amsterdam costs $175 + 140 + 122 + 24 = 461$. So there is a round trip with cost at most 500, and the desired output is `Yes`.

    Prove that the Travelling Salesman Problem is NP-complete (under Karp reductions). You may use the fact that the problem HC of deciding whether or not a graph contains a Hamilton cycle is NP-complete.

22. (5 marks) (Long question.) Consider the following problem, which we call Approx-SAT. The input is a logical formula F in conjunctive normal form. The desired output is `Yes` if there is an assignment of truth values to variables which satisfies at least two thirds of F's OR clauses, and `No` otherwise. Prove that Approx-SAT is NP-complete under Karp reductions.

23. (5 marks) (Long question.) Let $G = (V, E)$ be a bipartite graph with bipartition (A, B). Let a and b be positive integers, and suppose that every vertex in A has degree a and every vertex in B has degree b. What is the average degree $\frac{1}{|V|} \sum_{v \in V} d(v)$ of G's vertices, in terms of a and b? (Note that your answer should **not** depend on $|A|$ or $|B|$.)

24. (5 marks) (Long question.) Let $G = (V, E)$ be a graph. A *dominating set* in G is a subset $X \subseteq V$ such that for all $v \in V \setminus X$, we have $\{x, v\} \in E$ for some $x \in X$. In other words, every vertex of G is either contained in X or joined to X by an edge (or both). *DS* is the problem which asks: Given a graph G and an integer k, does G contain a dominating set of size at most k? Give a Karp reduction from VC to DS and briefly explain why it works. (**Hint:** Among other things, you will need to insert a vertex into the middle of each of G's edges.)

25. (5 marks) (Long question.) You are given two sets $\{x_1, \ldots, x_n\}$ and $\{y_1, \ldots, y_n\}$ of real values, and one list $e_1, \ldots, e_n \geq 0$ of tolerances. You wish to determine whether or not there is a map $f\colon \{y_1, \ldots, y_n\} \rightarrow$

$\{x_1, \ldots, x_n\}$ such that each y_i is mapped to exactly one x_j, and if y_i is mapped to x_j then $x_j - e_j \leq y_{i_j} \leq x_j + e_j$. In other words, you wish to know whether $\{x_1, \ldots, x_n\}$ and $\{y_1, \ldots, y_n\}$ are the same set up to additive error given by $e_1, \ldots, e_n$. Sketch an $O(n^2)$-time greedy algorithm for this problem, and prove using an exchange argument that it works.

[This one is a bit harder than anything in the exam — sorry... Still good practice though.]