Name _____

Student Number _____

HAND IN
answers recorded
on question paper

BISHOP'S UNIVERSITY

DEPARTMENT OF COMPUTER SCIENCE

CS 467/567

SAMPLE FIRST EXAMINATION

29 February 1901

Instructor: Stefan D. Bruda

Instructions

- This examination is *80 minutes in length* and is *open book*. You are allowed to use any kind of printed documentation. Electronic devices are not permitted. You are *not* allowed to share material with your colleagues. *Any violation of these rules will result in the complete forfeiture of the examination.*
- There is no accident that the total number of marks add up to the length of the test in minutes. The number of marks awarded for each question should give you an estimate on how much time you are supposed to spend answering the question.
- *To obtain full marks provide all the pertinent details*. This being said, do not give unnecessarily long answers. In principle, all your answers should fit in the space provided for this purpose. If you need more space, use the back of the pages or attach extra sheets of paper. However, if your answer is not (completely) contained in the respective space, clearly mention within this space where I can find it.
- Make sure that your name and student number appear on top of each sheet which is not securely stapled to the booklet (just in case). This also applies to any sheet which you detach from the booklet.
- The number of marks for each question appears in square brackets right after the question number. If a question has sub-questions, then the number of marks for each sub-question is also provided.

When you are instructed to do so, turn the page to begin the test.

1		10	/	10				
2		10	/	10				
3	a,b,c	30	/	30				
4	a,b	15	/	15				
5		10	/	10				
6		5	/	5				
	Total:	80	/	80	=	20	/	20

1. [10] Consider a domain D of *individuals*. A *predicate* is a named relation over D^k for some $k \geq 0$. A *literal* is either a predicate instance whose arguments are either individuals or *variables*, or the negation of a predicate. A *clause* is a disjunction of literals. A *first-order logic formula* is a conjunction of clauses. Here is an example of first-order logic formula, where variables are capitalized and individuals are written in lower caps:

$$(\overline{\text{dog}(fido)} \vee \text{dog}(X) \vee \text{smelly}(fido)) \wedge (\overline{\text{dog}(fido)} \vee \text{smelly}(X) \vee \text{smelly}(fido))$$

An *interpretation* maps each variable X to a value in D and each predicate $p(x_1, \ldots, x_k)$ to either True or False.

The *first-order satisfiability problem* (FOL-SAT) is stated as follows: Given a first-order logic formula, determine whether there exists an interpretation that makes the formula true (case in which the formula is deemed satisfiable).

Find a polynomial reduction from SAT to FOL-SAT. Justify your answer formally.

ANSWER:

A SAT formula ϕ is already a first-order logic formula ψ with no variables and containing only predicates with zero arguments. An interpretation makes ϕ true iff it makes ψ true (there are no variables to map in ψ). The identity function is therefore a reduction from SAT to FOL-SAT.

---

2. [10] Let $\mathbb{A}$ and $\mathbb{B}$ be two problems. Let τ be a reduction from $\mathbb{A}$ to $\mathbb{B}$ (meaning that $w \in \mathbb{A}$ iff $\tau(w) \in \mathbb{B}$), where $\tau(w)$ for some string w of length n is computed as follows:

   (a) Find two numbers p and q such that $1 < p < q < n$ and $p \times q = n$
   (b) If such numbers exist then return the first p symbols of w followed by the last q symbols of w
   (c) If such numbers do not exist then return w unchanged

   Suppose that $\mathbb{A}$ is NP-complete. What can you say about $\mathbb{B}$? Explain.

ANSWER:

The reduction τ is computable in nondeterministic linear time, but we have no way of knowing whether it can be computed in deterministic polynomial time. Indeed, two nondeterministic guesses take place, and it is likely that such guesses cannot be computed in deterministic polynomial time. If this is the case, then we cannot say anything about $\mathbb{B}$. If on the other hand nondeterministic guesses could be computed in deterministic polynomial time, then τ would become a polynomial reduction and so $\mathbb{B}$ would become NP-complete (but then P would be the same as NP s well).

---

3. [30] The SCHEDULING WITH DEPENDENCIES problem is stated as follows: *We are given a set of jobs $J = \{1, 2, \ldots, n\}$ with running times $t_i \geq 0$, $1 \leq i \leq n$, respectively. We are also given a directed acyclic dependency graph $G = (J, E)$ such that $(i, j) \in E$ iff job i must be completed before job j starts. Given $m > 0$, find a schedule for all the jobs in J on m identical machines such that the overall running time is minimized.*

   Consider the following candidate for an approximation algorithm for SCHEDULING WITH DEPENDENCIES, called *list scheduling*: Whenever a machine is idle, we schedule the first available job (with no predecessors in G) and then we delete from J the vertex j and from E all the edges (j, u). We repeat this until no more jobs are available.

   (a) [5] Show that list scheduling runs in polynomial time.

   ANSWER:

   The body of the loop runs in $O(n^2)$ time, since in the worst case we need to check all the edges in G to find a job with no predecessor. Note that such a job always exists since the graph is acyclic, and we never add edges to it to it cannot become cyclic. Therefore we can always find one vertex to remove and so the loop iterates n times, for a total $O(n^3)$ running time.

   (b) [5] Let $p = (p_1, p_2, \ldots, p_k)$ be a path in the graph G. Define the running time of p as $T_p = \sum_{i=1}^{k} t_{p_k}$ (the sum of the running times of the jobs along the path). Furthermore, let T^* be the running time of the optimal schedule.
   Prove that T^* is larger than T_p for any path p in G.

   ANSWER:

   Assume that there exists a path p such that $T^* \leq T_p$. Then there must be two jobs p_i and p_j along p such that p_i precedes p_j in p and the execution of p_i and p_j overlap. This is however impossible, since p_j depends on p_{j-1}, which in turn depends on p_{j-2} and so on until p_i, so p_j can only execute after the completion of p_i.

(c) [20] Show that list scheduling is a 2-approximation scheme for SCHEDULING WITH DEADLINES

In doing so, let x be the job that completes last and let p_x be a path that starts with a job with no predecessors and end with x. At any moment in time the machines either execute a job from p_x or they do not. They are never idle when not executing a job from p_x however, for if they were then whatever job from p_x is available will be executed. Split the time units into two sets A and B, where A are the time units in which a job from p_x executes and B the other time units. Establish suitable upper bounds for both $|A|$ and $|B|$ and you should be done.

ANSWER:

It is easy to see that the algorithm produces a valid schedule. Indeed, we always schedule jobs with no precedents, and we never run new jobs until the ones already scheduled complete.

With the above notation note that $|A| \leq T^*$ according to Question 3b. The remaining time is dedicated to the remaining jobs and so $|B| = (\sum_{j \in J \setminus p_x} t_j)/m$ (where by abuse of notation we use $J \setminus p_x$ to denote all the jobs in J that do not appear in p_x); indeed, as explained above no machine is idle in the time units in B and given that x executes last all processes much complete by the time x completes. Finally, note that $(\sum_{j \in J} t_j)/m \leq T^*$; indeed, $(\sum_{j \in J} t_j)/m$ is the time it takes for all the jobs in J to run on m machines in the absence of any constraints, which is an obvious lower bound for T^*.

Putting everything together we have the following upper bound on the running time T produced by list scheduling: $T = |A| + |B| \leq T^* + |B| = T^* + (\sum_{j \in J \setminus p_x} t_j)/m \leq T^* + (\sum_{j \in J} t_j)/m \leq 2T^*$ and so list scheduling is a 2-approximation algorithm.

---

4. [15] Suppose we want to formulate HAMILTONIAN CYCLE as a linear programming problem. Note that this is a decision rather than an optimization problem.

   (a) [5] What is the objective function for such a formulation? Explain.

ANSWER:

We do not maximize anything, but only looks for the existence of a solution instead.

The objective function does not play any role and so can be set to any constant such as

0.

   (b) [10] Define the following property of Hamiltonian cycles as linear constraints: Given a graph $G = (V, E)$, any vertex $v \in V$ has exactly 2 incident edges belonging to the Hamiltonian cycle of G.
   Hint. You may want to use the variables b_e, $e \in E$ such that $b_e = 1$ if e belongs to the Hamiltonian cycle and $b_e = 0$ otherwise.

ANSWER:

Let $e \sim v$ stands for the edge e being incident to v that is, $e \sim v$ iff $e = (v, u)$ or

$e = (u, v)$. Then the property can be represented by the following $|V|$ constraints:

$$\sum_{e \in E, e \sim v} b_e = 2 \text{ for all } v \in V$$

---

5. [10] Formulate the following problem as a linear programming problem:
Given a graph $G = (V, E)$ and two designated vertices $s, t \in V$, find the maximum number of edge-disjoint paths from s to t (that is, no two paths have a common edge common edges).
Hint. Can you formulate this problem as a maximum flow instance?

ANSWER:

Transform G into a maximum flow network by setting the edge capacity to 1 for every edge. Then a flow $f(s,t) = k$ means that there must be k different paths from s to t, for indeed each such a path is a sequence of edges and so has a flow capacity of 1. These paths must further be edge disjoint; indeed, if two paths share an edge then the sum of flows along the two paths is 1 rather than 2 (both flows must traverse the common edge, which cannot accommodate a flow larger than 1).

Given the consideration above we can formulate our problem as follows: maximize $\sum_{v \in V} f_{sv} - \sum_{v \in V} f_{vs}$ subject to the constraints $0 \leq f_{uv} \leq 1$ for all $u, v \in V$ (edge capacity) and $\sum_{v \in V} f_{vu} = \sum_{v \in V} f_{uv}$ for all $u \in V \setminus \{s, t\}$ (flow conservation).

---

6. [5] The polar angle of a point p is the angle between the segment $\overrightarrow{(0,0)p}$ and the x axis. Explain how to use the cross product to compare the polar angles of two points p_1 and p_2.

ANSWER:

If the polar angle of p_2 is larger than the polar angle of p_1 then $\overrightarrow{(0,0)p_2}$ is counterclockwise from $\overrightarrow{(0,0)p_1}$ and so $p_1 \times p_2 < 0$. By a similar argument $p_1 \times p_2 > 0$ if the polar angle of p_1 is larger than the polar angle of p_2. Obviously $p_1 \times p_2 = 0$ if the two points have the same polar angle.
