**Utrecht University**
**Faculty of Science**
**Department of Information and Computing Sciences**

**Final Exam Algorithms for Decision Support, Monday November 7, 2016, 13.30-16.30 hr.**

- Switch off your smart phone, PDA and any other mobile device and put it far away.
- You do not need a calculator.
- This exam consists of **4** questions
- Answers may be provided in either Dutch or English.
- All your answers should be clearly written down and provide a clear explanation. Unreadable or unclear answers may be judged as false.
- Please write down your name and student number on every exam paper that you hand in.
- The maximum score is indicated at each of the questions.

*Good luck, veel succes !*

**Question 1: Simulation of an emergency department**

*(a: 1 pt., b: 1 pt., c: 1 pt., total: 3 points.)*

We consider the emergency department of a medium-sized hospital. Patients arrive according to a Poisson process with an average of 5 per hour. When a patient arrives, he/she is immediately assigned to one of the categories A, B, or C (this is called triage). Category A consists of patients in a life threatening situation who most urgently need medical help. Morever, patients of category B need medical help more urgently than patients of category C. In the department, there is one team of a docter and a nurse, denoted by $DN_1$, for the treatment of patients of category A, and one team of a doctor and a nurse, denoted by $DN_2$, for the treatment of the other patients. If a patient arrives and the team for his/her category is free, treatment is started immediately after triage. $DN_1$ treats the patients in category A in order of arrival. $DN_2$ treats the patients in categories B and C in order of arrival, but all patients in B have priority over all patients in C.

On average, the fraction of patients of categories A, B, and C, equals 0.25, 0.25 and 0.5 respectively. The time required for the treatment of a patient from category A, B, or C is assumed to follow a Gamma distribution with $\alpha = 6$ and $\beta$ equal to $\beta_A, \beta_B, \beta_C$, respectively.

The hospital wants to perform a simulation study of the emergency department to find the maximum waiting time of a patient.

(a) Describe three *different* possible actions to validate the simulation model of the emergency department. The descriptions of the actions should be as specific as possible and include which aspects of the model are adressed. A general remark like: "discussion with an expert" will not result in any credit points. Different for example means that "action X for category A" and "same action X for category B" do not count as two different actions.

(b) A careful study of the treatment times at the emergency department revealed that $\beta_A = 10$, i.e. the treatment time of a patient of category A follows a Gamma(6, 10) distribution. Describe how can we generate these treatment times in a program written in an imperative programming language like Java or C♯ without using *any* specific random generation libraries or functions.
**Note:** You do not have to give a program, but just a description or pseudo-code.

(c) Describe an experimental set-up for the production runs of the simulation study of the emergency department with the maximum waiting time of a patient as performance measure. The set-up should use the fact that the emergency department is working continuously (24 hours per day, 7 days a week). Include an accurate description of the contents of your output results.

**Question 2: Tutoring lessons**

*(a: 0.75 pts., b: 0.75 pts., c: 0.75 pts., d: 0.75 pts. total: 3 points.)*

To earn more money, a student X decides to offer tutoring in mathematics for highschool students. It turns out that there are more than enough interested highschool students. X decides to take a systematic approach in planning his work. He divides his available time in $T$ intervals of one hour. In each hour interval, he can do one tutoring lesson. Let $p_{jt}$ denote the amount student $j$ is willing to pay for a lesson during interval $t$. Clearly, X wants to maximize the amount he earns and does not have to teach to all students.

(a) Suppose each student wants to take one lesson. Give an integer linear programming formulation for this situation. Give a description of the decision variables, objective and constraints.

(b) Show that the constraint matrix of part (a) is Totally Unimodular.

(c) Next, we consider the situation where students want to take more lessons. X asks every interested high-school student $j$ ($j = 1, \ldots, n$), to indicate

- during which of the $T$ intervals he/she can take a lesson; $a_{jt}$ equals 1 if student $j$ can take a lesson during interval $t$, and 0 otherwise.
- how many lessons he/she at most wishes to take, denoted by $q_j$.

Formulate the problem as an integer linear programming problem. Give a description of the decision variables, objective and constraints.

(d) Suppose a number of high-school students want to use the lessons as a crash course, which implies that student $j$ wants to take all the $q_j$ lessons in $q_j$ consecutive hours. If this is impossible he does not want to take any lessons at all. X reserves a full two weeks (except the hours between 2.00 AM and 10.00 AM) for these crash courses. Use $b_{jt}$ to indicate whether student $j$ can follow the crash course when starting with interval $t$ ($b_{jt} = 1$ if this is possible, and $b_{jt} = 0$ otherwise). It is possible to interrupt the crash course from 2.00 AM to 10.00 AM. We still have that $p_{jt}$ indicates the amount that $j$ is willing to pay for having a lesson in interval $t$. X does not have to teach all students, but wants to earn as much as possible.
Give an integer linear programming formulation for planning the crash courses. Give a description of the decision variables, objective and constraints.

**Question 3: Distribution network**
*(a: 1 pt., b: 1.25 pts., c: 0.75 pts., total: 3 points.)*
In this question you may use the fact that is it known that the following problems are $\mathcal{N}P$-complete: PARTITION, SUBSET SUM, CLIQUE, INDEPENDENT SET,VERTEX COVER, HAMILTONIAN PATH, HAMILTONIAN CYCLE, TRAVELLING SALESMAN PROBLEM.

(a) Let UCF be the Uncapacitated Facility Location problem. We are given $n$ possible locations of depots and $m$ customers. Each customer has to be served by one depot; the cost for serving customer $j$ ($j = 1, \ldots, n$) by depot $i$ are given by $c_{ij}$. A depot can only serve customers when it is open. Opening depot $i$ ($i = 1, \ldots, n$) has fixed cost $F_i$. Let $Q$ be a given nonnegative number. The question is if there exists a feasible solution, i.e. set of depots that are open and assignment of customers to depots, with cost at most $Q$. Show that UCF is $\mathcal{N}P$-complete.

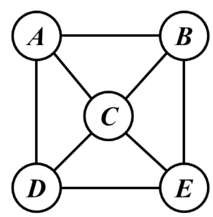
(b) We consider the problem DIRECT TRANSPORTATION. A truck has to pick-up goods at $n$ different locations $p_i$ ($i = 1, \ldots, n$). The truck starts at the depot $s$. When it has picked up goods at a specific location $p_i$ the goods have to be transported directly to delivery location $d_i$. After delivery the truck drives to another pick-up location. When it has visited all locations, it returns to the depot. The distances $w(u, v)$ between all the different locations are given. For a given number $W$ the questions if there is a route for the truck with total length at most $W$. Show that DIRECT TRANSPORTATION is $\mathcal{N}P$-complete.

(c) Describe two different **construction** heuristics for the DIRECT TRANSPORTATION problem. You do not have to analyze their worst-case behavior.

**Question 4: Structures for Bayesian Networks**
*(a: 0.25 pts., b: 0.25 pts., c: 0.5 pts., total: 1 point.)*

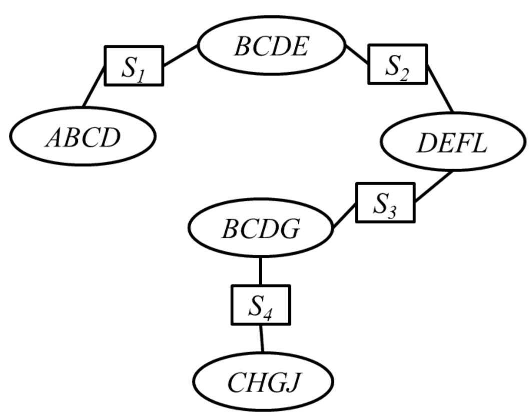
The construction of a *junction tree*, exploited as secondary structure for probabilistic inference in Bayesian networks, consists of the following steps: 1) moralisation of the DAG, 2) triangulation of the moral graph, 3) identifying cliques, 4) identifying a junction tree from the junction graph that organises the cliques. This question addresses two of these steps.

(a) Consider the following undirected graph $G$:




Is graph $G$ triangulated? If so, explain why. If not, explain which egde(s) can be added to triangulate $G$.

Consider the following spanning tree $T$ of a junction graph for some Bayesian network defined over variables $A, B, C, D, E, F, G, H, J, L$. All oval nodes represent cliques of variables and the rectangular nodes represent the clique separators.




(b) Specify the variables in each of the separators $S_1, S_2, S_3$, and $S_4$. Explain your answers.

(c) Does tree $T$ represent a junction tree? Clearly explain your answer.