Fundamental Algorithms for Bioinformatics	**Anno Acc. 2018-19**

Algorithm Design - Exercise Set 2

1. Provide a linear time algorithm that takes in input a weighted directed graph $G = (V, E)$, with costs on the edges $c(e)$ for $e \in E$ (cost may be negative), a node $t \in V$ and, for each $v \in V$ a value $d(v)$, and decides whether for each $v \in V$ it is true that $d(v)$ is the cost of the path of minimum cost among all the paths from from v to t.

2. You are given a weighted directed graph $G = (V, E)$, with costs on the edges $c(e)$ for $e \in E$ (cost may be negative), a node $t \in V$ and, for each $v \in V$ a value $d(v)$ which is the cost of a path of minimum cost from v to t. Provide an $O(|E| \log |V|)$ algorithm that given $G, d()$ and a node $t \neq t'$ computes, for each $v \in V$ the minimum cost of a path from v to t'. (*Hint*: It might be useful to consider a new cost function defined as follows: for edge $e = (v, w)$, let $\chi(e) = c(e) - d(v) + d(w)$. Is there a relation between costs of paths for the two different costs c and χ?)

3. Show that if we apply the algorithm seen in class to find the min-cost circulation for the network shown in the figure, there exist a sequence of choices of the cycle along which we improve that require 2×10^6 iteration to solve the problem.

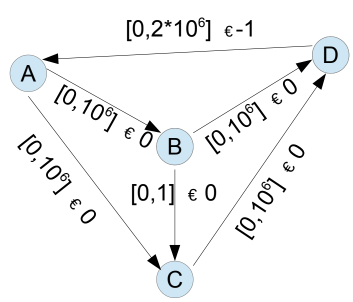

Figura 1: The Flow Network of 4

4. Assume that the following matrix has to be tested for the C1P. Using the algorithm seen in class provide
   (a) the overlap graph
   (b) the containment graph

   If the matrix has the C1P, provide 3 distict permutations of the columns that leave the 1s consecutive in each row (if there are fewer than 3 possible such permutations, provide them all). If the matrix does not have the C1P then apply the gap minimization algorithm considering only the last 5 columns.

$$A = \begin{bmatrix} 1 & 1 & 1 & 0 & 1 & 1 & 0 & 1 & 1 \\ 1 & 1 & 1 & 1 & 1 & 1 & 0 & 0 & 1 \\ 0 & 1 & 1 & 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 1 & 0 & 0 & 1 \\ 1 & 1 & 1 & 0 & 1 & 1 & 1 & 1 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \end{bmatrix}$$

5. A lab is seriously facing shortage of primary storage space for its files/data. The lab has to buying/rent secondary remote storage facilities. Preliminary analyses have indicated this as a better solution than in-house expansion of the primary storage. The different available options have both limitations in the maximum amount of space offered and costs related to the access of the information, once stored there. We are interested in determining the optimal policy to choose how to distribute the files not anymore fitting in-house on different secondary remote storage facilities in order to limit the cost incurred by taking into account the different expected usage rate of the different type of information we are relocating.

   Assume there are n remote facilities where to relocate the exceeding files. Let α_j be the maximum amount of information on the remote facility j and χ_j be the cost to access one unit of information from this facility. We assume that the information we need to store remotely is divided into m different categories, each one of which is accessed with some rate. Let γ_i be the amount of information units from the category i, and let ρ_i be the rate (how many times per unit time) a unit of information from category i will need to be retrieved. We aim at storing/distributing the information in the different remote storage places in order to minimize the overall expected cost of retrieval.

   (a) Model the problem of finding the optimal choice of information redistribution in the different remote locations as a min-cost circulation problem.

   (b) Prove that an optimal solution to the above problem is given by the greedy algorithm that distributes the information of the different categories by iteratively assigning as many information units from the category (among those still to be accommodated) of maximum retrieval rate to the remote location (among those that still have some available space) with the minimum access cost.