COT5405: Programming Assignment 2 (Fall 2017)

Instructions:

For all programming assignments, you shall write program code in one of the following languages: C, C++, Java, or Python (those that can be executed on the Eustis2 Unix machines). Note: it is your responsibility to know how to log in to the Unix machine in our Computer Science Division, how to code and run your programs on this machine.

Input:

Input for each programs is given below. It can be stored in the program; no input read-in needed.

Output/Results:

Your SSH client window should display on screen execution results at the end of the program run (complex output can be drawn by hand).

The ***screenshot image*** of your SSH client window containing the result printout should be captured and presented in the result section of your report.

Submission:

1. A project report containing the following:
    a. The purpose of the assignment and a brief introduction of the problems
    b. Method: language used, compilation instructions, format of input, platform specifics (i.e., Unix), algorithms to solve the problems, etc.
    c. Outputs of test input data
    d. Discussions/Remarks on running times, implementation problems, possible implementation improvements
2. A folder/directory for each of the three programs is created. Self-contained: The source code of each (along with any auxiliary files, e.g., input data) should be contained in its folder so that the code can be compiled and run from the folder.
3. Compressed all files and directories into one zip/rar file and submit it online by the due date.

Problem Description:

1. **(RNA Secondary Structure)** Given the following RNA molecule
   B = **AUGGCUACCGGUCGAUUGAGCGCCAAUGUAAUCAUU**
   find a secondary structure S that maximizes the number of base pairs.
   a. Show the value of the number of base pairs for the above molecule derived from the algorithm explained in '06dynamic-programming.ppt' slides.
   b. Show (or draw by hand) the RNA secondary structure graph, in the similar way as the graph on Page 31 of slides '06dynamic-programming.ppt'.

2. **(Sequence Alignment)**: Consider the following two strings:
   AGGCTATCACCTGACCTCCAGGCCGATGCCC
   TAGCTATCACGACCGCGGTCGATTTGCCCGAC
   For alignment, assume that each 'mismatch' penalty is α=2, each 'gap' penalty is δ=1. Please find the alignment with minimum penalty cost. Please refer to Page 40-45 in lecture slides '06dynamic-programming.ppt' to solve this problem.
   a. Show the final penalty cost of the alignment you have found.
   b. Show the alignment of these two strings, in the similar format as the alignment example on Page 42 of the lecture slides.

   Hint: Remember that you can know which alignment of the symbol x_i and y_j takes by checking which equation has been used on Page 44's equation set of the lecture slides.

3. **(Max Flow: Ford–Fulkerson algorithm)**: For the following directed graph (link capacity values are shown on the graph).
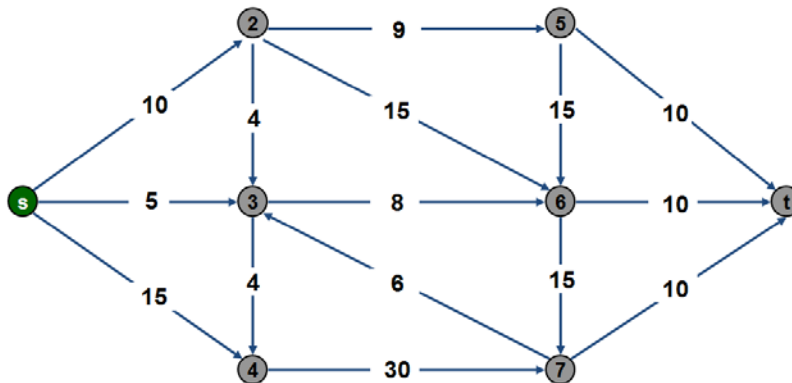
a. What is the maximum flow value from the source node s to the target node t?

b. Show the flow values on each link on this directed graph for the maximum flow scenario.