St. Francis Xavier University
Department of Computer Science

CSCI 355: Algorithm Design and Analysis
Assignment 4
Due April 4, 2024 at 1:30pm

Assignment Regulations.

- This assignment must be completed individually.
- Please include your full name and email address on your submission.
- You may either handwrite or typeset your submission. If your submission is handwritten, please ensure that the handwriting is neat and legible.

[7 marks] 1. A biotechnology company is attempting to create a generic drug that is as similar as possible to the more expensive name-brand version. To this end, the company has identified a target sequence A corresponding to the makeup of the name-brand drug, and they want to produce a sequence that is as close to A as possible, made up of shorter sequences from a library L. The target sequence A has length m, while each sequence in the library L has length $k \leq n$ for some $n < m$.

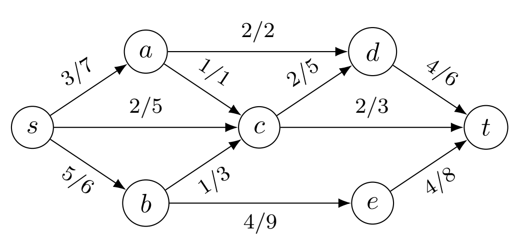
The company manufactures generic drugs by taking sequences from the library L (allowing repetitions) and *concatenating* them. The concatenation of two sequences $X = x_1 x_2 \cdots x_i$ and $Y = y_1 y_2 \cdots y_j$ is the sequence $XY = x_1 x_2 \cdots x_i y_1 y_2 \cdots y_j$; in other terms, it is the sequence produced by "gluing" the end of X to the beginning of Y.

Thus, the company's goal is to find some set of sequences $\{B_i\}$, where $B_i \in L$ for all $1 \leq i \leq \ell$, such that $B = B_1 B_2 \cdots B_\ell$ and the *sequence alignment cost* between A and B is as small as possible.

(a) Let $\text{OPT}(j)$ denote the alignment cost of the optimal solution for the subsequence $A[1..j]$. What is an appropriate formulation of $\text{OPT}(j)$ for this problem?

(b) Design an algorithm to solve this problem using the dynamic programming paradigm.
(You do not have to provide an analysis of your algorithm.)

[7 marks] 2. Consider the following flow network.



(a) What is the value of the flow shown in the network? Is this a max flow? If it is, explain why. If it is not, find a max flow.

(b) Find a min cut in the flow network, and give the capacity of your min cut.

[6 marks] 3. In lecture, we defined the decision problems INDEPENDENT-SET and VERTEX-COVER, which both ask some question about the vertices of a graph G. Given some integer k, INDEPENDENT-SET asks whether there exists a subset of k or more vertices of G such that no two vertices are adjacent, while VERTEX-COVER asks whether there exists a subset of k or fewer vertices of G such that each edge of G is incident to at least one vertex in the subset.

As you might expect, these two problems aren't so different. Given a graph $G = (V, E)$ with n vertices and an integer k, prove that some subset $S \subseteq V$ is an independent set of size k if and only if $V \setminus S$ is a vertex cover of size $n - k$.

[5 marks] 4. Choose your favourite topic from the course, and write a multiple-choice style question with one correct answer and 3–4 plausible-but-incorrect answers that tests a concept or notion related to that topic.

For inspiration, consider the multiple-choice style questions you saw on the midterm and practice midterm exams.