Genome 540 Discussion

February 1st, 2024 Clifford Rostomily



Agenda

- Assignment 4
- Assignment 5

Assignment 4

Overview

Part 1: Write a program to find the highest-weight path in a directed acyclic graph using dynamic programming

Part 2: Run your program on a linked list created from DNA sequence

Program 1: Highest weight path

- 1. Convert graph to text file of **vertices** and **edges** by hand
- 2. Use dynamic programming to find the max weight path through the graph (Lectures 7/8)
 - a. Overall
 - b. With constraints (START/END)
- 3. Output
 - a. Path Score
 - b. The start/end vertex on the path
 - c. Labels for all the edges on path (in order)

V vii START V vi V v ... E A ii i -1 E B iii i 5

Example:

Part 2 Score: 4.0 Begin: vii End: i Path: LIDA

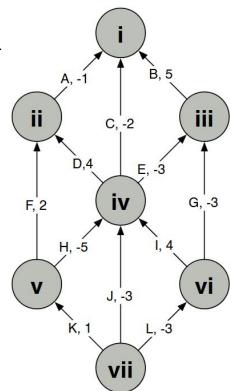
Part 1

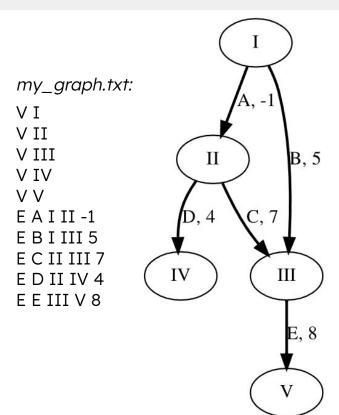
Score: 8.0

Begin: vi

Path: ID

End: ii



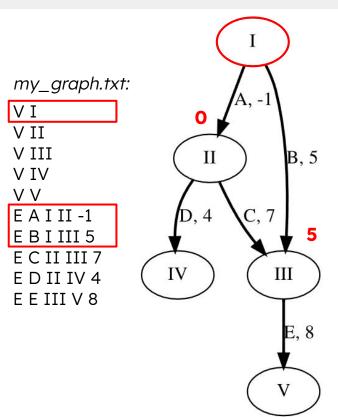


- Assume that graph file is depth ordered
- Vertex I has no parents so points to itself

Vertex	Í	II	III	IV	V
Highest Weight Parent	I	II	III	IV	V
w(v) (Vertex weight)	0	0	0	0	0

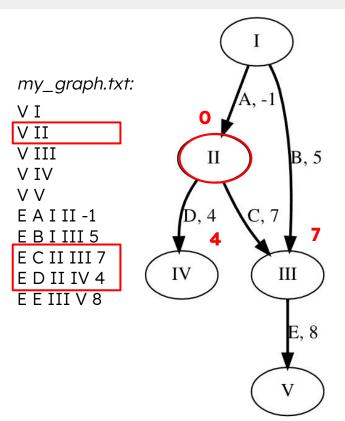
Best Path Start

Ι



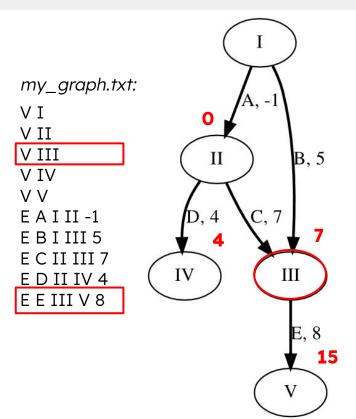
Vertex	1	II	III	IV	V
Highest Weight Parent	I	II	ı	IV	V
w(v) (Vertex weight)	0	0	5	0	0

Best Path Start III



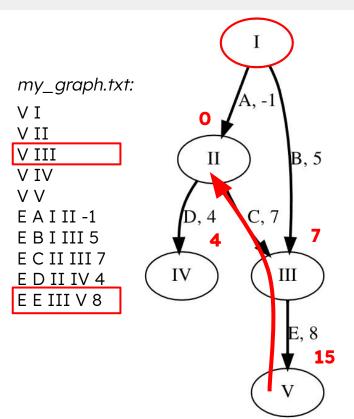
Vertex	I	II	Ш	IV	V
Highest Weight Parent	I	II	II	II	V
w(v) (Vertex weight)	0	0	7	4	0

Best Path Start III



Vertex	1	Ш	111	IV	V
Highest Weight Parent	I	II	II	II	III
w(v) (Vertex weight)	0	0	7	4	15

Best Path Start



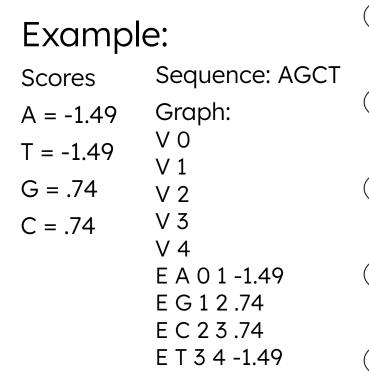
Vertex	1	II	111	IV	V
Highest Weight Parent	I	II	II	II	III
w(v) (Vertex weight)	0	0	7	4	15

Best Path Start V

 Now traceback to find highest weight path

Program 2: DNA Linked List

- 1. Create a linked list from a DNA sequence and a scoring scheme
 - a. Positions are vertices
 - b. Bases are edges
- 2. Run your program from part 1 on the graph



G. 0.74

C, 0.74

-1.49

Questions?

Program 1:

Use dynamic programming to find the highest weight path in an arbitrary WDAG

Program 2:

Make a linked list from a fasta and run program 1 on it

Homework 5

Overview

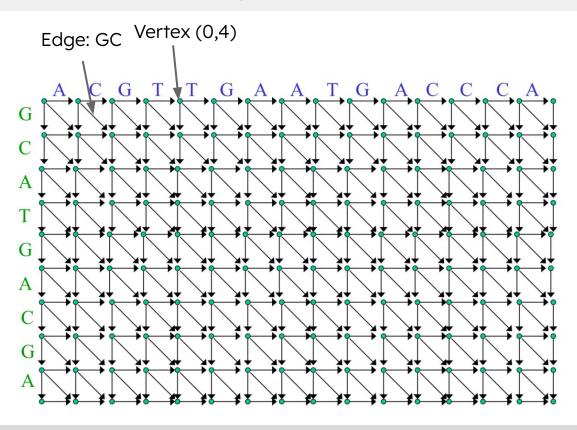
■ Part 1

 Build a weighted edit graph for 3 amino acid sequences of the insulin protein (human, frog, water buffalo) using the BLOSUM62 scoring matrix and save it as a text file

■ Part 2:

 Use your program from HW4 to find the max weight path through the edit graph

The edit graph for 2 sequences



Vertices

```
(0,0) (0,1) (0,2) ··· (0,N2)

(1,0) (1,1)

(2,0) ··

(N1,0) (N1,N2)
```

Edges are alignments

(0,0) (0,1) weight (_A)

(0,0) (1,0) weight (G_)

(0,0) (1,1) weight (GA)

Computing edge weights

BLOSUM62

What is the weight for edge DR?

What is the edge weight of _A?

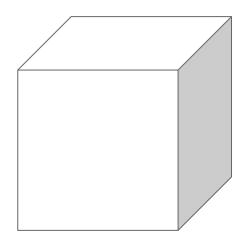
Computing edge weights

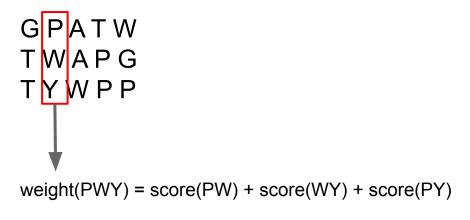
BLOSUM62

What is the weight for edge DR? -2

What is the edge weight of _A? -6 (Gap)

Now do it for 3 sequences





Computing edge weights

BLOSUM62

What is the weight for edge DR_?

Computing edge weights

BLOSUM62

What is the weight for edge DR_?

DR + R_ + D_ -2 + -6 + -6 = -14

Reminders

- HW4 due this Sunday, 11:59pm
- Please have your name in the filename of your homework assignment and match the template