

# Genome 540 Discussion

Conor Camplisson

January 26<sup>th</sup>, 2023

# Outline

- Homework 4 overview
- Homework 3 & 4 questions

# Outline

- Homework 4 overview
- Homework 3 & 4 questions

# Homework 4 Overview

## Part one: max weight path through a WDAG

- Write a program to find the max weight path
- Convert a WDAG diagram to a text representation (by hand)
- Determine the max weight path using your program
  - Both: unconstrained, constrained start/stop vertices

## Part two: GC-rich genomic sub-sequence

- Write a program to represent a genome as a WDAG, export .txt
  - GC vs. AT scoring scheme
- Determine the max weight path using your program
  - GC-rich sub-sequence, lookup feature in .gbff file

# Homework 4 Overview

## Part one: max weight path through a WDAG

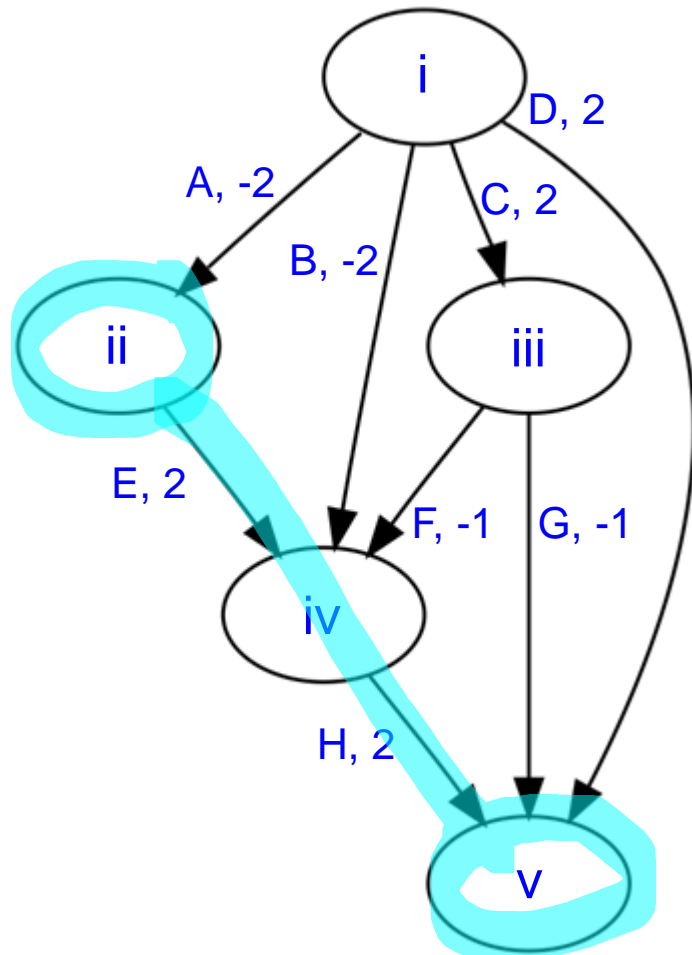
- Write a program to find the max weight path
- Convert a WDAG diagram to a text representation (by hand)
- Determine the max weight path using your program
  - Both: unconstrained, constrained start/stop vertices

## Part two: GC-rich genomic sub-sequence

- Write a program to represent a genome as a WDAG, export .txt
  - GC vs. AT scoring scheme
- Determine the max weight path using your program
  - GC-rich sub-sequence, lookup feature in .gbff file

# Homework 4 Overview

Example WDAG



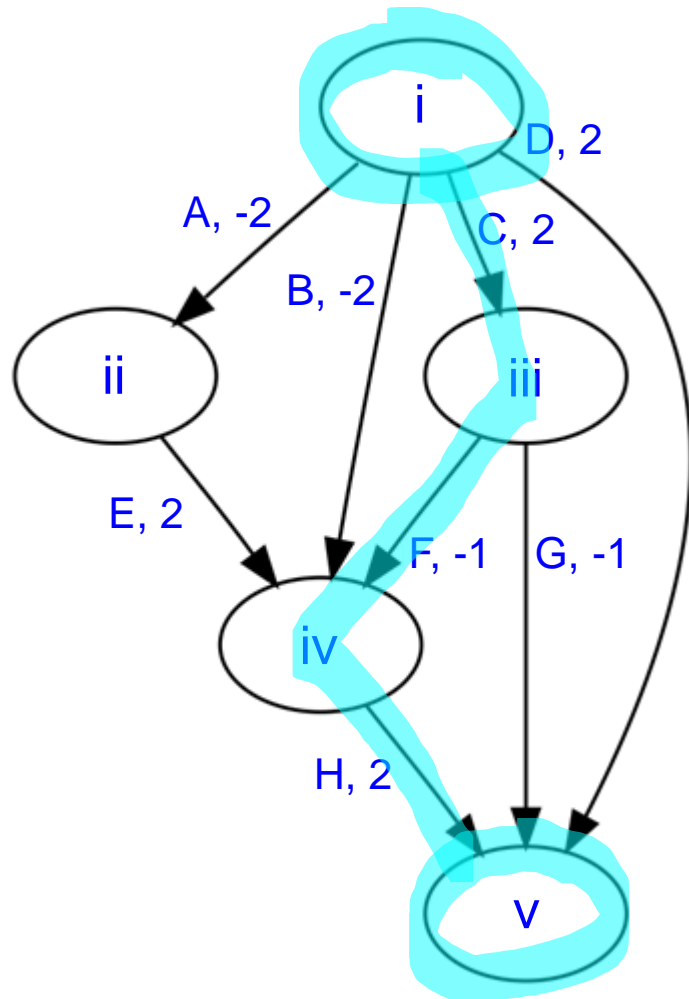
wdag\_unconstrained.txt

```
V i
V ii
V iii
V iv
V v
E A i ii -2
E B i iv -2
E C i iii 2
E D i v 2
E E ii iv 2
E F iii iv -1
E G iii v -1
E H iv v 2
```

Score: 4.0

# Homework 4 Overview

Example WDAG



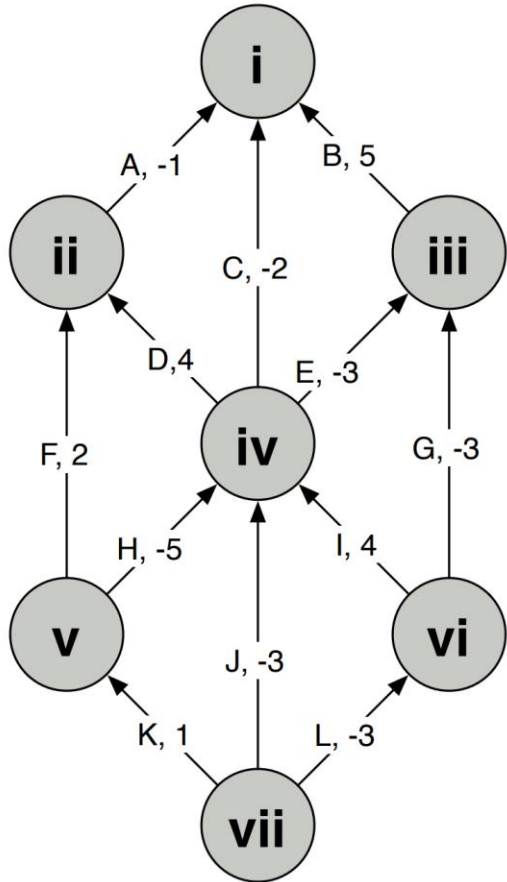
wdag\_constrained.txt

```
V i START ←
V ii
V iii
V iv
V v END ←
E A i ii -2
E B i iv -2
E C i iii 2
E D i v 2
E E ii iv 2
E F iii iv -1
E G iii v -1
E H iv v 2
```

Score: 3.0

# Homework 4 Overview

## Test WDAG

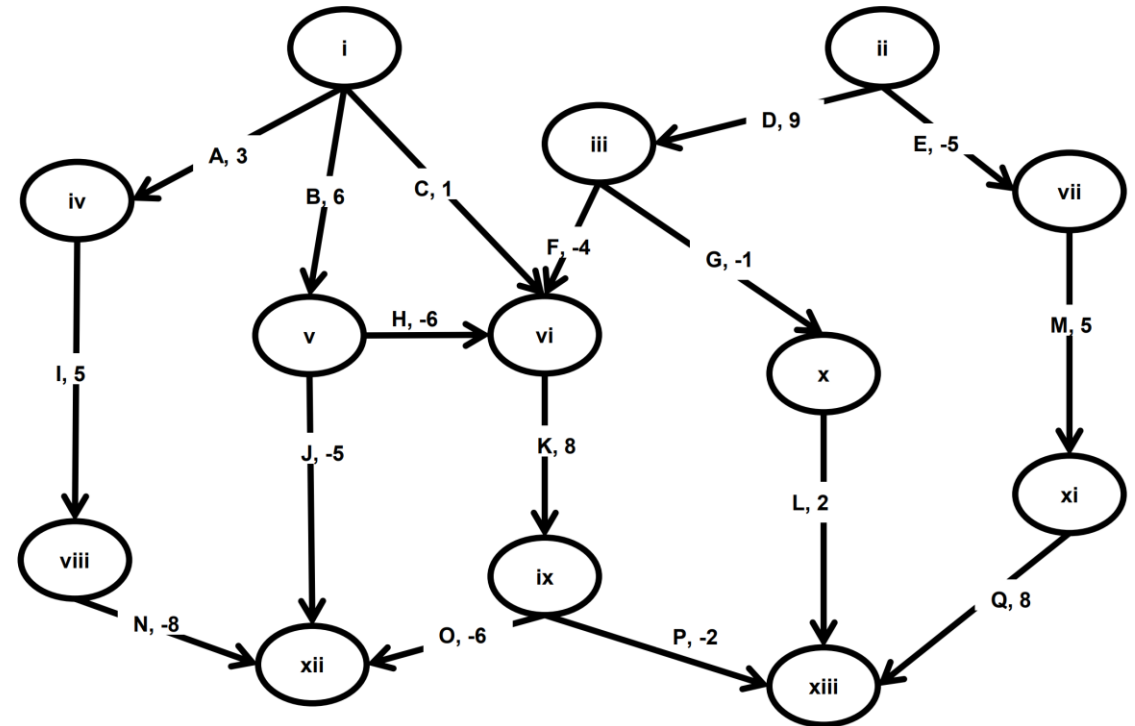


Assignment: GS 540 HW4  
Name: Conor Camplisson  
Email: concamp@uw.edu  
Language: C++/Python  
Runtime: 0m17.545s

Part 1  
Score: 8  
Begin: vi  
End: ii  
Path: ID

Part 2  
Score: 4  
Begin: vii  
End: i  
Path: LIDA

## Homework WDAG





# Homework 4 Overview

## Part one: max weight path through a WDAG

- Write a program to find the max weight path
- Convert a WDAG diagram to a text representation (by hand)
- Determine the max weight path using your program
  - Both: unconstrained, constrained start/stop vertices

## Part two: GC-rich genomic sub-sequence

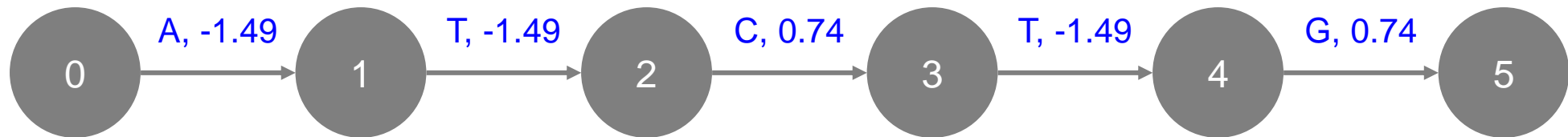
- Write a program to represent a genome as a WDAG, export .txt
  - GC vs. AT scoring scheme
- Determine the max weight path using your program
  - GC-rich sub-sequence, lookup feature in .gbff file

# Homework 4 Overview

Example sequence:

5'-ATCTG-3'

WDAG representation:



# Homework 4 Overview

genome.fa

5'-ATCTG-3'

soring\_scheme.txt

```
A -1.49
C 0.74
G 0.74
T -1.49
N 0
```

Program 2



```
V 0
V 1
V 2
V 3
V 4
V 5
E A 0 1 -1.49
E T 1 2 -1.49
E C 2 3 0.74
E T 3 4 -1.49
E G 4 5 0.74
```

Program 1



# Homework 4 Overview

## Test Genome

Part 3  
Fasta: CP003508.fna  
Non-alphabetic characters: 0  
>gi|400273702|gb|CP003508.1| Mycoplasma gallisepticum NC96\_1596-4-2P, complete genome  
\*=986257  
A=337443  
C=156212  
G=155909  
T=336693  
N=0  
Score: 11.07  
Begin: 344420  
End: 344444  
Path: GGCGGCGGCCCTGGCGATGGCCG  
Description: This sequence lies within the HFMG96NCA\_2038 gene (encodes a hypothetical protein).

## Homework Genome

*S. pyogenes*



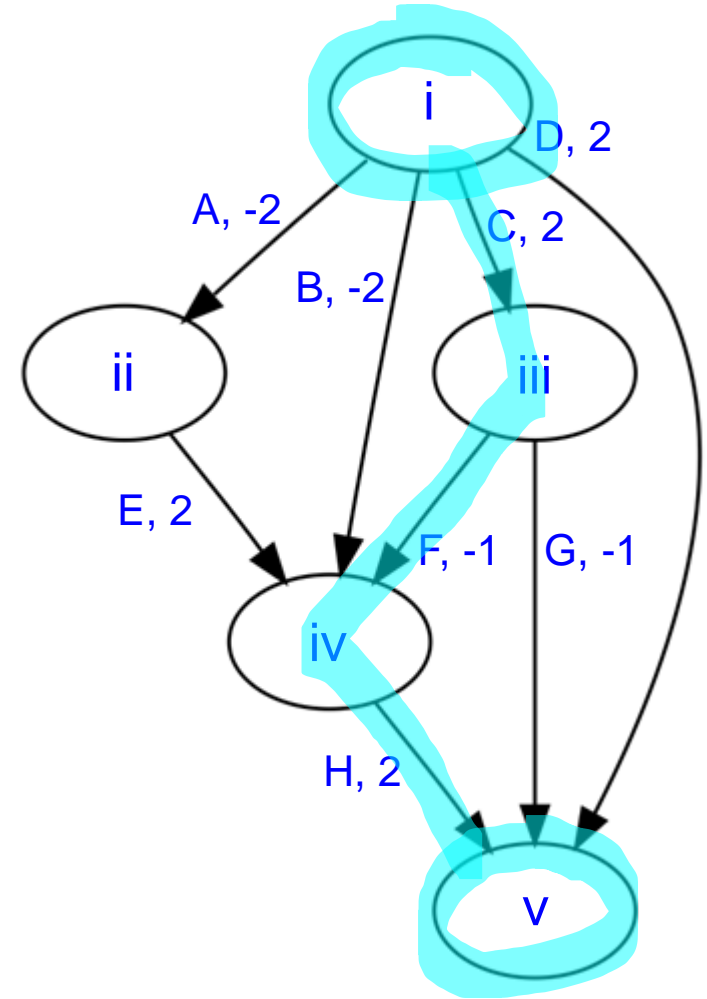
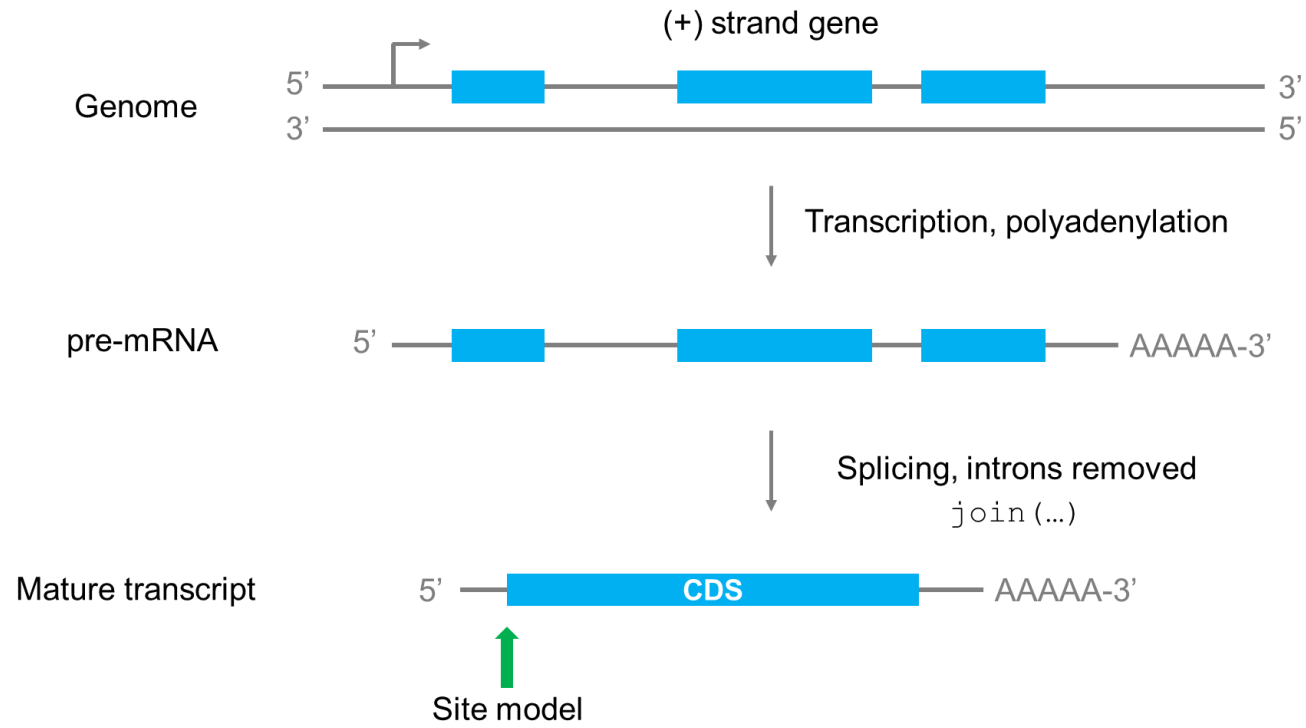
[Home](#) > [Genome Editing](#) > [Products](#) > Cas9 Nuclease, *S. pyogenes*

Cas9 Nuclease, *S. pyogenes*    

# Outline

- Homework 4 overview
- Homework 3 & 4 questions

# Homework 3 & 4 Questions ?



# Reminders

- Homework 3 due this Sunday Jan. 29, 11:59 pm
  - Single text file, compressed with `gzip`
  - name in the file: `camp_lissson_hw3.txt.gz`
- Homework 4 due next Sunday Feb. 5, 11:59 pm

