Sequence Bioinformatics
Assignment 03

Emil Paulitz, Raphael Olipitz

November 27, 2020

Task 1

There are $3 \cdot 4 \cdot \cdot = 48$ edges between nucleotides of different sequences: $4 \cdot 4$ between each sequence, so 3 times that for 3 sequences.

For each sequence, there are 11 possible paths to go through it with only entering once: do not visit this sequence (1), visit only one nucleotide (4), visit 2 nucleotides (3), 3 nucleotides (2), or all four nucleotides (1). Thus, there are $11 \cdot 11 \cdot 11$ possibilities altogether. You have to subtract $5 \cdot 5 \cdot 5$ paths, with which you do not use any directed edge. Also, for every sequence there are 6 possible paths with which you do take an directed edge, but you have to subtract these possibilities if you do not visit the other sequences: $3 \cdot 6$. But, if one visits all three sequences and takes at least one directed edge ($10 \cdot 10 \cdot 10 - 4 \cdot 4 \cdot 4$ possible paths), there are two possible combinations of sequences: 012 and 021, so one has to add this. Altogether:

$$11 \cdot 11 \cdot 11 - 5 \cdot 5 \cdot 5 - 3 \cdot 6 + 10 \cdot 10 \cdot 10 - 4 \cdot 4 \cdot 4 = 2124$$

Tasks 2-3

See jupyter notebook "Assignment03.ipynb"

Task 4

Result in "MSA.csv"

Task 5

Using the column with the best result, a graph representing the following alignment is returned by lp_solve:

```
CA-TG
CAGT-
-AGTT
```