

## Suggestions to solutions: Lab 4

### Task 1 – Time complexity

Time complexity:  $O(n^2)$ . For each element in the list the algorithm runs through the remaining elements.

### Task 2 – Approximation algorithms 1

The approximation ratio  $R$  must be larger than  $5/2$ . The approximation ratio for a minimization algorithm is the maximum of the ratio  $A(\pi)/OPT(\pi)$  over all  $\pi$ .  $5/2$  is the maximum observed value.

### Task 3 – Motif search

MEME is a randomization algorithm.

Since all the yeast genes are regulated by MCM1, it is reasonable to assume that they all contain one occurrence of the MCM1-motif, so “One per sequence”.

The known motif is of width ten, so it seems like a good idea to search for motifs of the same length or longer. E.g. 10-20. The nucleotides flanking the core motif often carry some signal.

Looking at the sequence logos, it is apparent that Motif 1 is indeed consistent with the consensus motif CCNNNWWRRGG.

Looking at the “Block Diagrams”, it is obvious that the motif can occur anywhere in the promoters. There is no pattern (which is often the case).

The closest motif found by STAMP is indeed a motif experimentally verified to be bound by MCM1.

Using negative sequences in MEME allows the program to find motifs that not only occur in the input (positive) sequences, but also do not occur in the negative sequences. Thus “jalapeño” would not be selected as a special word in the ten pages, because the negative pages would reveal that this word in fact occurs on these pages as well.

The shuffle-option in MEME search for motifs in the input sequences, but with the letters in these sequences rearranged/shuffled. Obviously, any motif and corresponding E-values found here should be considered non-significant. So we are pleased to notice that the E-values here are nowhere near the E-values we got using the real sequences.