

Exercise Sheet 5 for Computational Biology (Part 1), SS 15

Hand In: Until Tuesday, 07.07.2015, 10:00 am, email to `r_muelle@cs...`, hand-in box in stairwell 48-6 or in lecture.

Problem 13

4 points

Show that the *Disjoint Double Digest Problem* is \mathcal{NP} -complete. This problem is like the *Double Digest Problem* from lecture, but here we enforce that the enzymes have disjoint restriction sites. Formally, the (decision) problem becomes:

Given multisets A , B and C over \mathbb{N} with $S = \sum_{a \in A} a = \sum_{b \in B} b = \sum_{c \in C} c$, are there permutations π_A and π_B of the elements in A and B , respectively, such that

$$C = \text{Dist}(\text{Pos}(\pi_A) \cup \text{Pos}(\pi_B))$$

and additionally

$$\text{Pos}(\pi_A) \cap \text{Pos}(\pi_B) = \{0, S\}?$$

Hint: A solution can be found in the following paper, which should be accessible from the university network, however, the description is very brief. If you follow their lines of proof you are expected to develop the arguments in detail.

Cieliebak, M., Eidenbenz, S., & Woeginger, G. J. (2003). Double Digest Revisited: Complexity and Approximability in the Presence of Noisy Data. COCOON 2003 Proceedings (pp. 519–527), Springer. http://dx.doi.org/10.1007/3-540-45071-8_52

Problem 14

4 points

Consider the backtracking algorithm for the partial digest problem (PDP) from lecture. How does an input multiset A look like that leads to an exponential running time of the algorithm?

Describe a family of inputs (i.e. for infinitely many sizes n) and prove that the backtracking algorithm needs exponential time.

Problem 15

4 points

Given the set of STS probes $\{\alpha, \beta, \gamma, \delta, \epsilon, \zeta, \eta, \theta\}$ and

$$A = (A_{i,j}) = \begin{array}{c|cccccccc} & \alpha & \beta & \gamma & \delta & \epsilon & \zeta & \eta & \theta \\ \hline \text{I} & 1 & 0 & 0 & 0 & 1 & 0 & 0 & 1 \\ \text{II} & 0 & 0 & 1 & 1 & 0 & 1 & 0 & 0 \\ \text{III} & 1 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \\ \text{IV} & 1 & 1 & 0 & 0 & 1 & 0 & 0 & 1 \\ \text{V} & 0 & 0 & 1 & 0 & 0 & 0 & 1 & 0 \end{array}$$

the matrix assigning the occurrences of the probes to the fragments $\{\text{I}, \text{II}, \text{III}, \text{IV}, \text{V}\}$. $A_{i,j}$, $i \in \{\text{I}, \text{II}, \dots, \text{V}\}$, $j \in \{\alpha, \beta, \dots, \theta\}$, is equal to 1 iff probe j occurs in fragment i .

Use the algorithm from the lecture to determine the PQ tree representation of the permutations of the columns which transform the matrix to consecutive-ones form.

Start with the universal tree for $\{\alpha, \beta, \gamma, \delta, \epsilon, \zeta, \eta, \theta\}$ and show the reduction of all five restrictions. Draw all intermediate PQ trees following the convention introduced in the lecture and name the rule applied to transform one tree to the next. How many permutations are represented by the resulting tree?