Technische Universität München
Department of Informatics
Chair for Efficient Algorithms
Prof. Dr. Ernst W. Mayr/Dr. Jens Ernst
Johannes Nowak

Summer Term 2005
Exercise Sheet 10
July 5, 2005

Selected Topics in Computational Biology

Due: 12.07.2005 **after** *the lecture*

Exercise 1 (10 points)

Consider the algorithm for agglomerative bottom-up clustering of a set of n expression profiles given in the lecture. Show that the time complexity of the algorithm can be bounded by $O(n^2)$? For your analysis you can suppose that the distance function can be evaluated in constant time.
Hint: How can you determine the current minimum in each step efficiently?

Exercise 2 (10 points)

Consider again the algorithm for hierarchical profile clustering. Show that the following choices for the parameters $\alpha_{i_0}, \alpha_{i_1}$ and γ lead to the distance values for the newly created cluster $S_{i_{0,1}}$:

a) Single Linkage: For $\alpha_{i_0} = \alpha_{i_1} = 1/2$ and $\gamma = -1/2$ we have

$$d(S_{i_{0,1}}, S_k) = \min\{d(S_{i_0}, S_k), d(S_{i_1}, S_k)\}$$

.

b) Complete Linkage: For $\alpha_{i_0} = \alpha_{i_1} = 1/2$ and $\gamma = 1/2$ we have

$$d(S_{i_{0,1}}, S_k) = \max\{d(S_{i_0}, S_k), d(S_{i_1}, S_k)\}$$

Exercise 3 (10 points)

We compare two expression profiles X_i and X_j. We say X_i and X_j are *similar* if more than half of their components $\xi_{i,k}$ and $\xi_{j,k}$ for $1 \leq k \leq m$ have the same values. Suppose for each k we have $\xi_{i,k} = \xi_{j,k}$ with probability 1/3 and independently of each other. Use the Chernoff bound technique to give an upper bound on the probability that the two profiles are similar.