

Oberseminar Theoretische Informatik

Wintersemester 2008/2009

Sepp Hartung

Fixed-parameter Tractability and Data Reduction for Hierarchical Clustering

Mo, 02.02.2009 um 14 Uhr (c.t.) im SR 3319 (Ernst-Abbe-Platz 2, 3. Stock).

We consider the NP-hard HIERARCHICAL CLUSTERING problem. It occurs in many practical applications like statistics, computational biology, and data mining. Given a node set X , a nonnegative integer $M > 0$ and a pairwise distance function d on X , the problem is to arrange the elements of X as leaves in a rooted tree such that all leaves have a distance $M + 1$ to the root and the “correction effort” $\sum_{i,j \in X} |d_c(i,j) - d(i,j)|$ is less than a constant k for the induced ultrametric d_c . Thereby, the distance $d_c(i,j)$ between two nodes in X is defined by the height of the least common ancestor of i and j in the rooted tree. We perform parameterized complexity studies, leading to a search tree algorithm with a running time of $O(3^k + |X|^3)$. Moreover, we will present data reductions rules which shrink a given HIERARCHICAL CLUSTERING instance to a $O(k^2)$ -node kernel. Finally, we present some advanced observations on the deeper problem structure, which can be used to decompose an instance into independent sub-instances.

Homepage:

<http://theinfl.informatik.uni-jena.de/teaching/ws0809/oberseminar-ws0809>