Dr. Eick

Group Homework Credit Task Group K

**Hierarchical Clustering**

*To be presented on Tuesday, November 1*

**Hierarchical Clustering**

A dataset consisting of object A, B, C, D, E and F with the following distance matrix is given:

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| distance | A | B | C | D | E | F |
| A | 0 | 2 | 4 | 5 | 9 | 11 |
| B |  | 0 | 3 | 6 | 15 | 12 |
| C |  |  | 0 | 7 | 10 | 8 |
| D |  |  |  | 0 | 1 | 13 |
| E |  |  |  |  | 0 | 14 |
| F |  |  |  |  |  | 0 |

a) Assume single[[1]](#footnote-1) link hierarchical clustering is applied to the dataset! What dendrogram will be returned?

b) Does the clustering result change of we use Max/Complete Link[[2]](#footnote-2) instead?

c) How does hierarchical clustering differ from more classical clustering algorithms, such as K-Means and DBSCAN?

1. When assessing the distance between clusters the minimum distance is used. [↑](#footnote-ref-1)
2. When assessing the distance between clusters the maximum distance is used; that is, the distance of the pair of objects one in one cluster and one in the other cluster that are furthest apart is used. [↑](#footnote-ref-2)