Dr. Eick

Group Homework Credit Task Group I

**Hierarchical Clustering**

*To presented on Thursday, October 24*

**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 | 9 | 1 | 15 | 2 | 11 |
| B |  | 0 | 8 | 6 | 5 | 12 |
| C |  |  | 0 | 7 | 10 | 4 |
| D |  |  |  | 0 | 3 | 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?

Group Homework Credit Task Group J

**PAM/K-Medoids**

*To presented on Thursday, October 24*

Assume you apply PAM/K-Medoids to a dataset X containing

7 objects for k=4. The distance matrix of the objects in

X is given on the left. Moreover, we assume that the initial

set of representatives is {D, E, F, G}. Run PAM for 1 iteration

indicating all computations PAM performs. What new

representative set will be obtained? Break ties by rolling a dice!

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
|  | **A** | **B** | **C** | **D** | **E** | **F** | **G** |
| **A** | 0 | 2 | 9 | 5 | 3 | 5 | 6 |
| **B** |  | 0 | 11 | 3 | 4 | 1 | 9 |
| **C** |  |  | 0 | 12 | 8 | 6 | 13 |
| **D** |  |  |  | 0 | 7 | 2 | 9 |
| **E** |  |  |  |  | 0 | 3 | 6 |
| **F** |  |  |  |  |  | 0 | 8 |
| **G** |  |  |  |  |  |  | 0 |

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)