Prof. Dr. Alfred Benedikt Brendel
Chair of Business Information Systems, esp. Intelligent Systems and Services

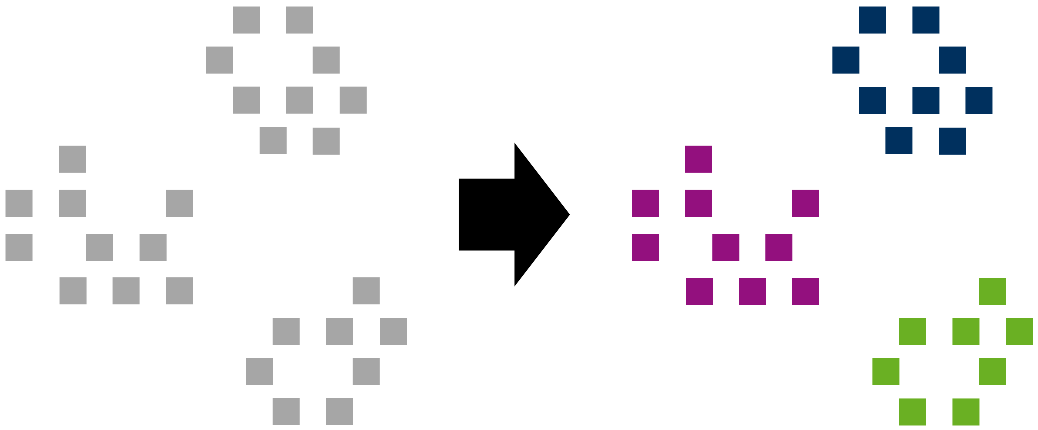
# Data Science: Advanced Analytics
## Segmentation Methods

Dresden // 26.04.2023
Sommersemester 2023

# Recap: Aims of the cluster analysis

Segmentation or cluster analysis is used to group objects and/or characteristics into classes or groups so that

- between the elements of the same classes the greatest possible similarity,
- between the elements of different classes the greatest possible diversity is achieved



# Clustering Methods

**Partitional Clustering**

K-Means & K-Medoids
For n objects or data tuples, a partitioning method constructs k partitions of the data, where each partition represents a cluster k <= n.

**Hierarchical Clustering**
Agglomerative, Divisive
Grouping data objects into a tree (dendrogram) of clusters

**Density-Based Clustering**
DBSCAN
Grouping data tuples along density-connected points.

**Grid-Based Clustering**

STING, WaveCluster & CLIQUE
Method uses a multi-resolution grid data structure. E.g. spatial area is divided into rectangular cells (STING)

**Others**

Model-Based Clustering, Fuzzy, Evolutionary, Simulated Annealing

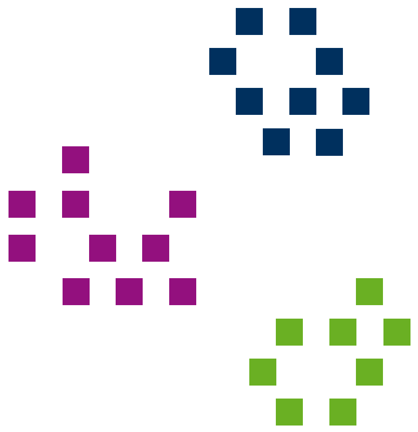
**Prof. Dr. Alfred Benedikt Brendel**
Chair of Business Information Systems, esp. Intelligent Systems and Services

# Partional Clustering Methods
for Segmentation

# Partional Clustering Methods

Partitioning methods are segmentation methods which divide the set of **objects N** on the basis of a fixed number of **classes s** in such a way that the computed segmentation or **partition K** minimizes a given **quality index b(K)**:

$$\min_{\in \wp(\wp(N))} \left\{ b(\boldsymbol{K}) : \quad = \{K_1, K\ , K_s\}, \quad \bigcup_{i=1}^{s} K_i = N, \quad K_i \cap K_j = \varnothing \right\}$$



# The exchange principle

(1) Choose start partition $\boldsymbol{K}^0 = \{K_1^0, \ldots, K_s^0\}$ (start heuristic).

Choose $b(\boldsymbol{K}^0)$.

(2) Search Object(s), so that a transfer reduces *b*.

(3) Change Object(s) from the current to the best new class.

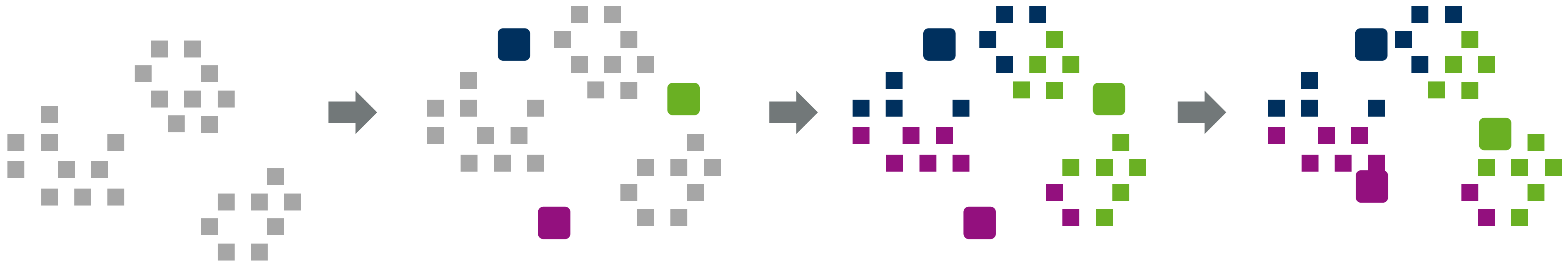
(4) Repeat (2) and (3) until no other change is possible.
$\Rightarrow$ **local optimum** found

- The procedures break off after a finite number of steps.
- The procedures usually reach only a **suboptimum** (global optima are usually reached only if several objects can be exchanged simultaneously, taking into account all exchange possibilities).
- The result depends i.a. on the selected start partition (use several start partitions).

# Error Minimization Algorithms
e.g. K-means

Partitions the data into K clusters represented by their centers or means.

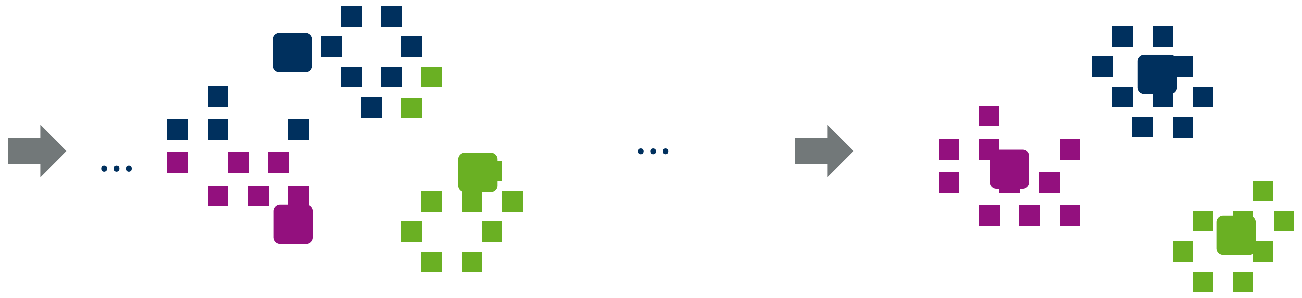
**Example: 2-dimensional, K=3**



a) initialize with K cluster centers (random or heuristic)

b) Assign objects to the closest cluster center

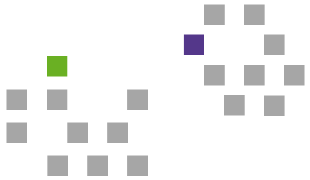
c) Update cluster centers based on the assignment



# Error Minimization Algorithms
## e.g. K-medoids or PAM (partition around medoids)

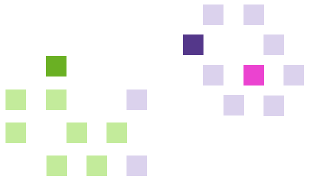
Find representative objects (medoids) in clusters

**Example: 2-dimensional, K=2**

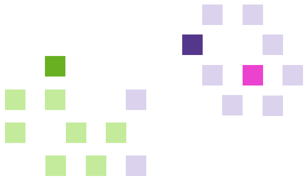


Random choose K objects as initial medoids

Assign remaining objects to nearest medoids



Randomly select a non-medoid object



Calculate the „total costs“ (quality index) of a swap to the new medoid

…

repeat until convergence criterion

CHAIR OF BUSINESS INFORMATION SYSTEMS
ESP. INTELLIGENT SYSTEMS AND SERVICES

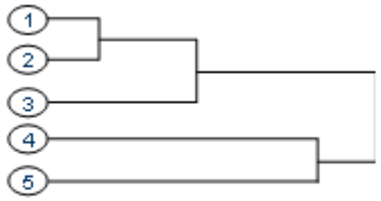
**Prof. Dr. Alfred Benedikt Brendel**
Chair of Business Information Systems, esp. Intelligent Systems and Services

# Hierarchical Clustering Methods
## for Segmentation

# Hierarchical Clustering Methods

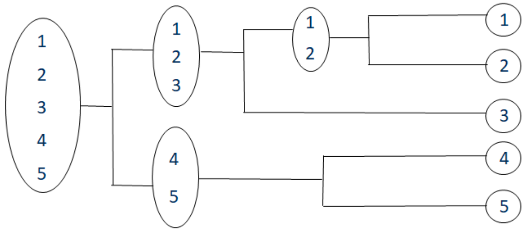
Segmentation methods that construct a sequence of partitions on the basis of a set of **objects N**.

**Agglomerative Clustering**
Each object initially represents a cluster of its own. Then clusters are successively merged until the desired cluster structure is obtained.




- Starting point are n = |N| one-element classes.
- Successive transition to coarser decompositions
- Termination as soon as given criterion is fulfilled
- Low computation times, good practical suitability

**Diversive Clustering**
All objects initially belong to one cluster. Then the cluster is divided into sub-clusters, which are successively divided into their own sub-clusters. This process continues until the desired cluster structure is obtained.



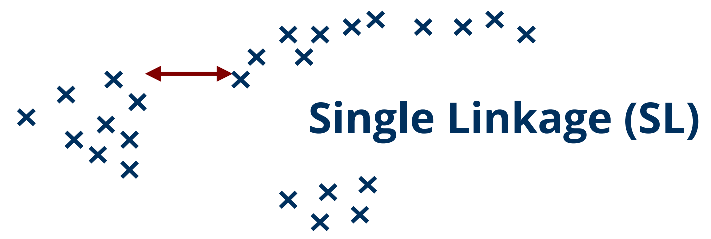

- Starting point is the class of all objects.
- Successive transition to finer decompositions
- Termination as soon as given criterion is fulfilled

> Kaufman, L. and Rousseeuw, P.J. (1990) Partitioning around Medoids (Program PAM). In: Kaufman, L. and Rousseeuw, P.J., Eds., Finding Groups in Data: An Introduction to Cluster Analysis, John Wiley & Sons, Inc., Hoboken, 68-125.

# Similarity measures

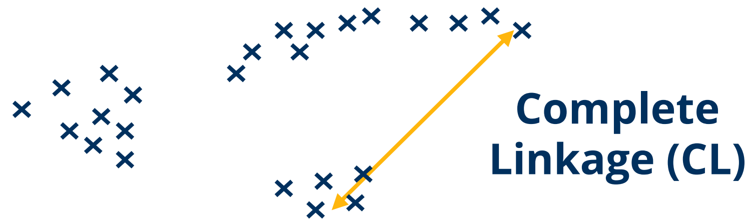
are based on the different **recalculation** of the interclass differences:

**Single Linkage**
Nearest Neighbour




$$v(K,L) = \min_{i \in K, j \in L} d(i,j)$$

**Complete Linkage**
Furthest Neighbour




$$v(K,L) = \max_{i \in K, j \in L} d(i,j)$$

**Average Linkage**
Group Average


$$v(K,L) = \frac{1}{|K| \cdot |L|} \sum_{\substack{i \in K \\ j \in L}} d(i,j)$$

# Single Linkage
## An Example

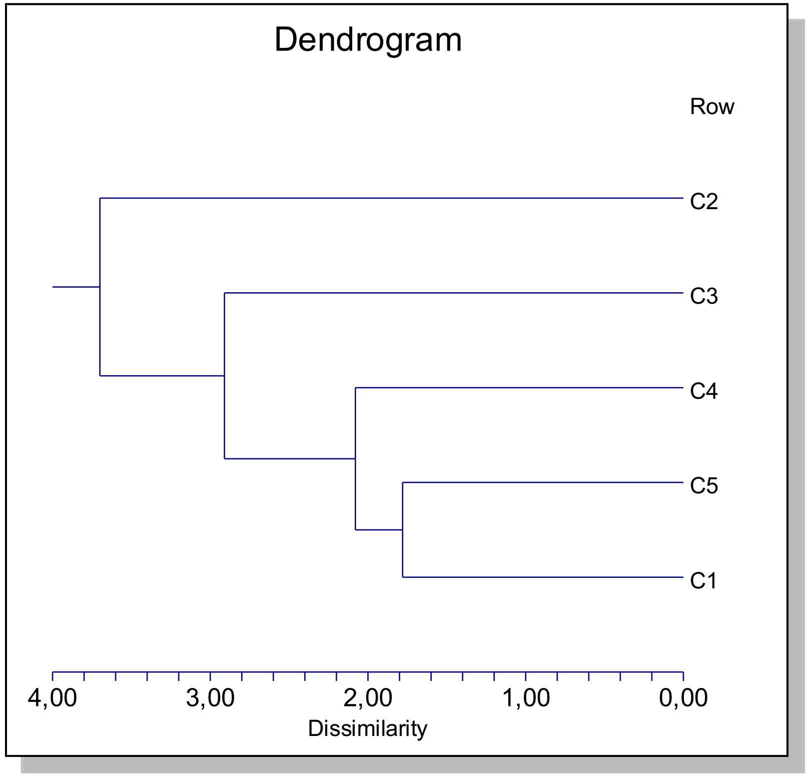
| D | 2 | 3 | 4 | 5 |
|---|---|---|---|---|
| 1 | 4,48 | 2,91 | 2,08 | **1,78** |
| 2 | | 5,07 | 3,92 | 3,70 |
| 3 | | | 4,33 | 4,03 |
| 4 | | | | 2,63 |

First fusion of objects 1 and 5 (min $d_{ij}$ = 1.78).

$\Rightarrow$ $K^1$ = {{1,5}, {2}, {3}, {4}}

Fusion levels :
1.78 2.08 2.91 3.70

What we are looking for is a hierarchy with v(i,j) = d(i,j) and $v(K,L) = \min_{i \in K, j \in L} d(i,j)$



# Group Average Linkage

## An Example

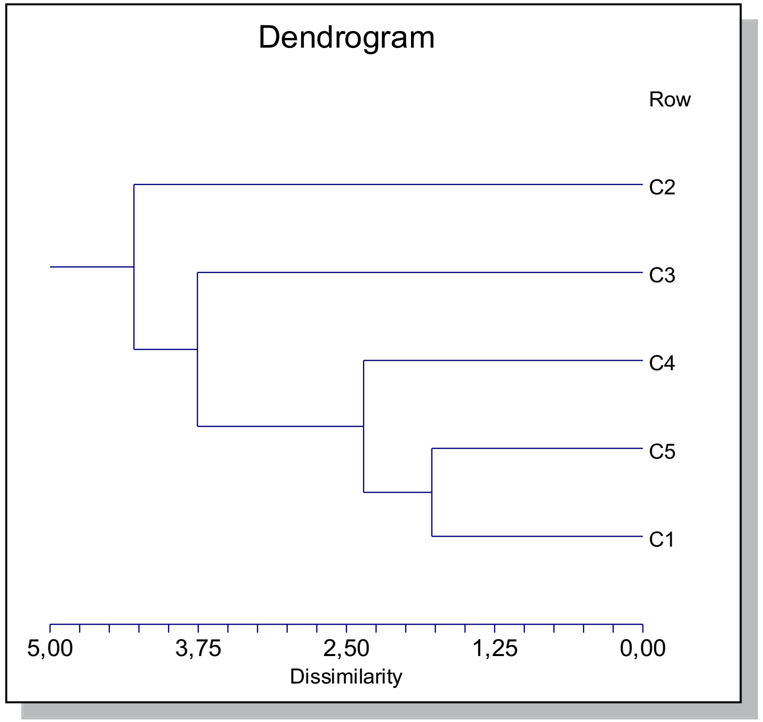
| D | 2 | 3 | 4 | 5 |
|---|---|---|---|---|
| 1 | 4,48 | 2,91 | 2,08 | **1,78** |
| 2 | | 5,07 | 3,92 | 3,70 |
| 3 | | | 4,33 | 4,03 |
| 4 | | | | 2,63 |

First fusion of objects 1 and 5 (min $d_{ij}$ = 1.78).

$\Rightarrow K^1 = \{\{1,5\}, \{2\}, \{3\}, \{4\}\}$

Fusion levels :
1.78 2.36 3.76 4.29

What we are looking for is a hierarchy with v(i,j) = d(i,j) and $v(K,L) = \frac{1}{|K| \cdot |L|} \sum_{\substack{i \in K \\ j \in L}} d(i,j)$



# Complete Linkage

## An Example

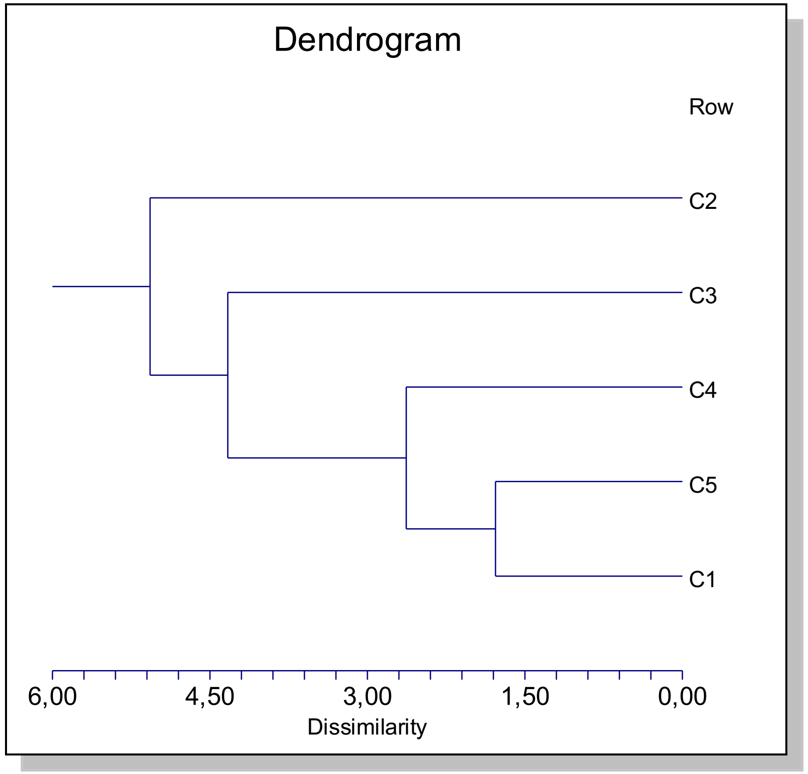
| D | 2 | 3 | 4 | 5 |
|---|---|---|---|---|
| 1 | 4,48 | 2,91 | 2,08 | **1,78** |
| 2 | | 5,07 | 3,92 | 3,70 |
| 3 | | | 4,33 | 4,03 |
| 4 | | | | 2,63 |

First fusion of objects 1 and 5 (min $d_{ij}$ = 1.78).

$\Rightarrow K^1 = \{\{1,5\}, \{2\}, \{3\}, \{4\}\}$

Fusion levels :
1.78 2.63 4.33 5.07

What we are looking for is a hierarchy with v(i,j) = d(i,j) and $v(K,L) = \max_{i \in K, j \in L} d(i,j)$



# Interpretation of a Dendrogram

**Abrupt changes** in the value of the quality criterion allow conclusions to be drawn about the appropriate **number of classes (EB).**

**Similar objects** are merged early, dissimilar objects later; **outliers** are assigned to a large cluster only at the end.

The class structure is
- **stable**, if different methods lead to similar results,
- **intensive**, if classes of comparable size are successively merged and
- **weak**, if successively only neighboring single objects are added.

# Assessment of a hierarchy

To answer the question, which of the determined hierarchies "best" reproduces the given **distance matrix D**, one first calculates the (ultrametric) **distance matrix D***, which can be uniquely determined from the dendrograph, according to

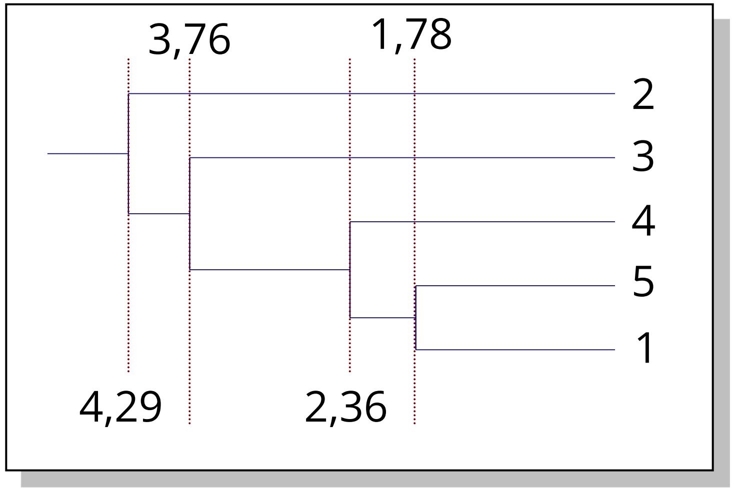
$$d^*(i,j) = \min_{\substack{i \in K,\ j \in L \\ K,L \in \mathcal{K}}} v(K,L)$$

and compares this matrix suitably with the original distance matrix D.

# Distance matrix D* of the average linkage

## An Example

| D | 2 | 3 | 4 | 5 |
|---|---|---|---|---|
| 1 | 4,48 | 2,91 | 2,08 | **1,78** |
| 2 | | 5,07 | 3,92 | 3,70 |
| 3 | | | 4,33 | 4,03 |
| 4 | | | | 2,63 |



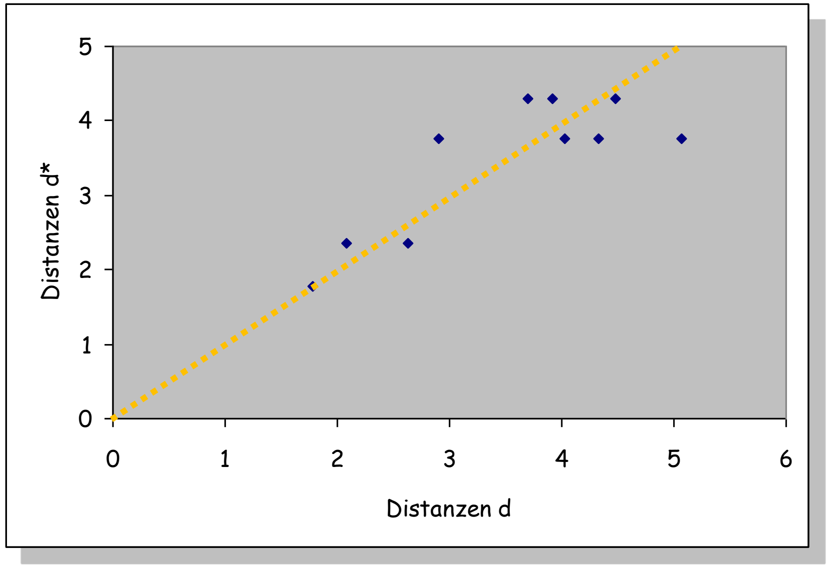

Calculation of the dissimilarity using the **dendrogram**

| D* | 2 | 3 | 4 | 5 |
|---|---|---|---|---|
| 1 | 4,29 | 3,76 | 2,36 | **1,78** |
| 2 | | 4,29 | 4,29 | 4,29 |
| 3 | | | 3,76 | 3,76 |
| 4 | | | | 2,36 |

# The Shepard-Diagram

The simplest way to evaluate the different distance matrices D and D* is the so-called Shepard diagram, in which the true **distances d** and the **calculated distances d*** are compared in a **coordinate system**.

**Example**: Average Linkage



# Variance-Accounted-For-Criteria

The **VAF criterion** can be calculated to assess the loss of information in procedures that explicitly use distances:

$$\mathrm{VAF} = 1 - \frac{\sum_{i=2}^{n}\sum_{j=1}^{i-1}\left(d(i,j) - d^*(i,j)\right)^2}{\sum_{i=2}^{n}\sum_{j=1}^{i-1}\left(d(i,j) - \bar{d}\right)^2} \quad \text{mit} \quad \bar{d} = \frac{2}{n(n-1)}\sum_{i=2}^{n}\sum_{j=1}^{i-1} d(i,j)$$

For procedures that use variance to measure heterogenity, the VAF cannot be meaningfully interpreted.

A value close to 1 is desired.

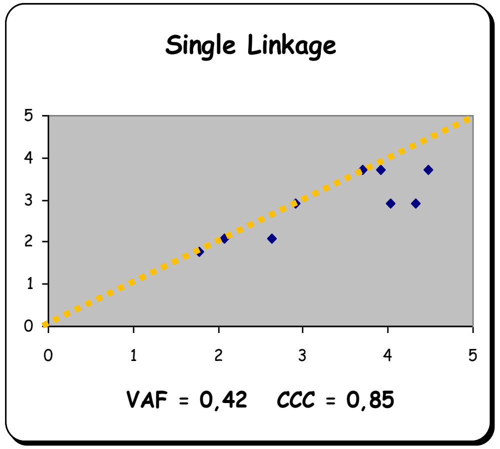
# Cophenetic correlation coefficient

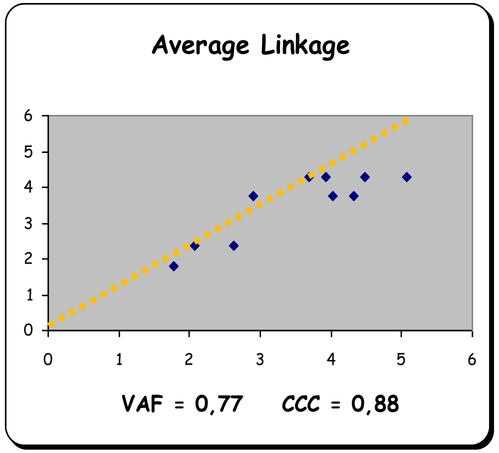
To assesses the existence of a linear relationship between the true distances d and the calculated distances d* according to:

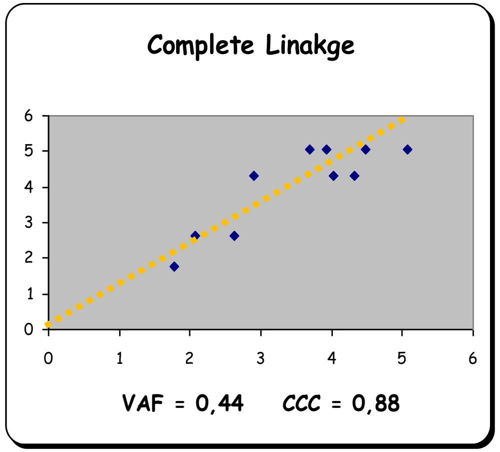
$$CCC = \frac{\sum_{i=2}^{n}\sum_{j=1}^{i-1}\left(d(i,j)-\bar{d}\right)\left(d^*(i,j)-\bar{d}^*\right)}{\sqrt{\sum_{i=2}^{n}\sum_{j=1}^{i-1}\left(d(i,j)-\bar{d}\right)^2\sum_{i=2}^{n}\sum_{j=1}^{i-1}\left(d^*(i,j)-\bar{d}^*\right)^2}} \quad \text{mit} \quad \begin{aligned}\bar{d} &= \frac{2}{n(n-1)}\sum_{i=2}^{n}\sum_{j=1}^{i-1}d(i,j)\\ \bar{d}^* &= \frac{2}{n(n-1)}\sum_{i=2}^{n}\sum_{j=1}^{i-1}d^*(i,j)\end{aligned}$$

Values close to 1 indicate a small loss of information.

# Example: Assessment of a hierarchy




It can be seen that the **Average Linkage** solution is judged the best.

TECHNISCHE UNIVERSITÄT DRESDEN

CHAIR OF BUSINESS INFORMATION SYSTEMS
ESP. INTELLIGENT SYSTEMS AND SERVICES

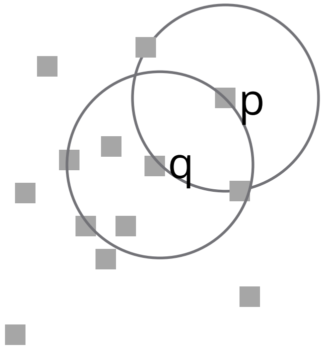
DRESDEN concept

**Prof. Dr. Alfred Benedikt Brendel**
Chair of Business Information Systems, esp. Intelligent Systems and Services

# Density Based Clustering Methods
for Segmentation

# Density Based Clustering Methods

Designed for discovering clusters of arbitrary shape. It is also used to handle noise in the data clusters




ε … maximum radius of the neighborhood

MinPts … minimum number of points in an ε-neighborhood

„p is directly *density-reachable* from q“

**Idea:** continue growing a given cluster as long as the density (number of objects or data points) in the neighborhood exceeds some threshold. The neighborhood within a radius ε has to contain at least a minimum number of objects. It needs density parameters as a termination condition.

> Ester, M., Kriegel, H.P., Sander, J. and Xu, X. (1996) A Density-Based Algorithm for Discovering Clusters in Large Spatial Databases with Noise. Proceedings of the 2nd International Conference on Knowledge Discovery and Data mining, 226-231

**Prof. Dr. Alfred Benedikt Brendel**
Chair of Business Information Systems, esp. Intelligent Systems and Services

# Thank you for your attention