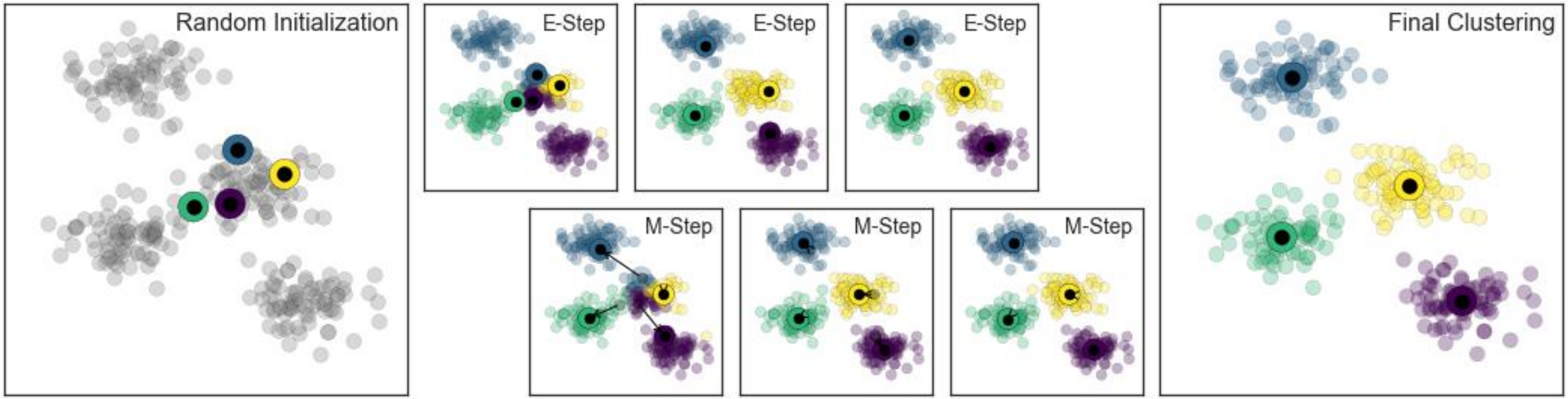
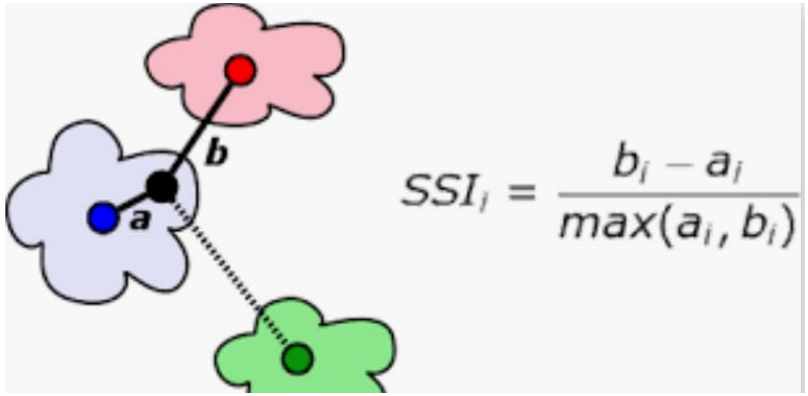
Clustering

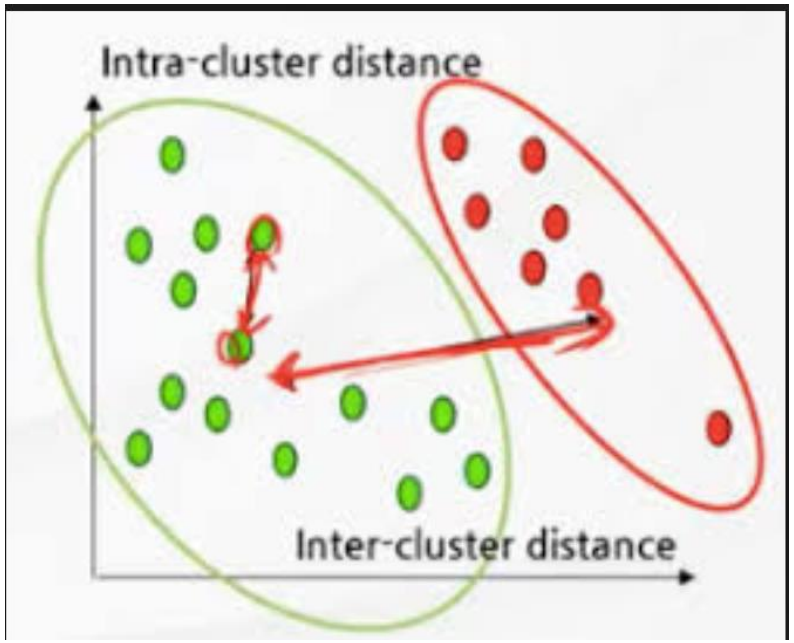
Kmeans


Random Initialization
E-Step
E-Step
E-Step
M-Step
M-Step
M-Step
Final Clustering

Silhouette

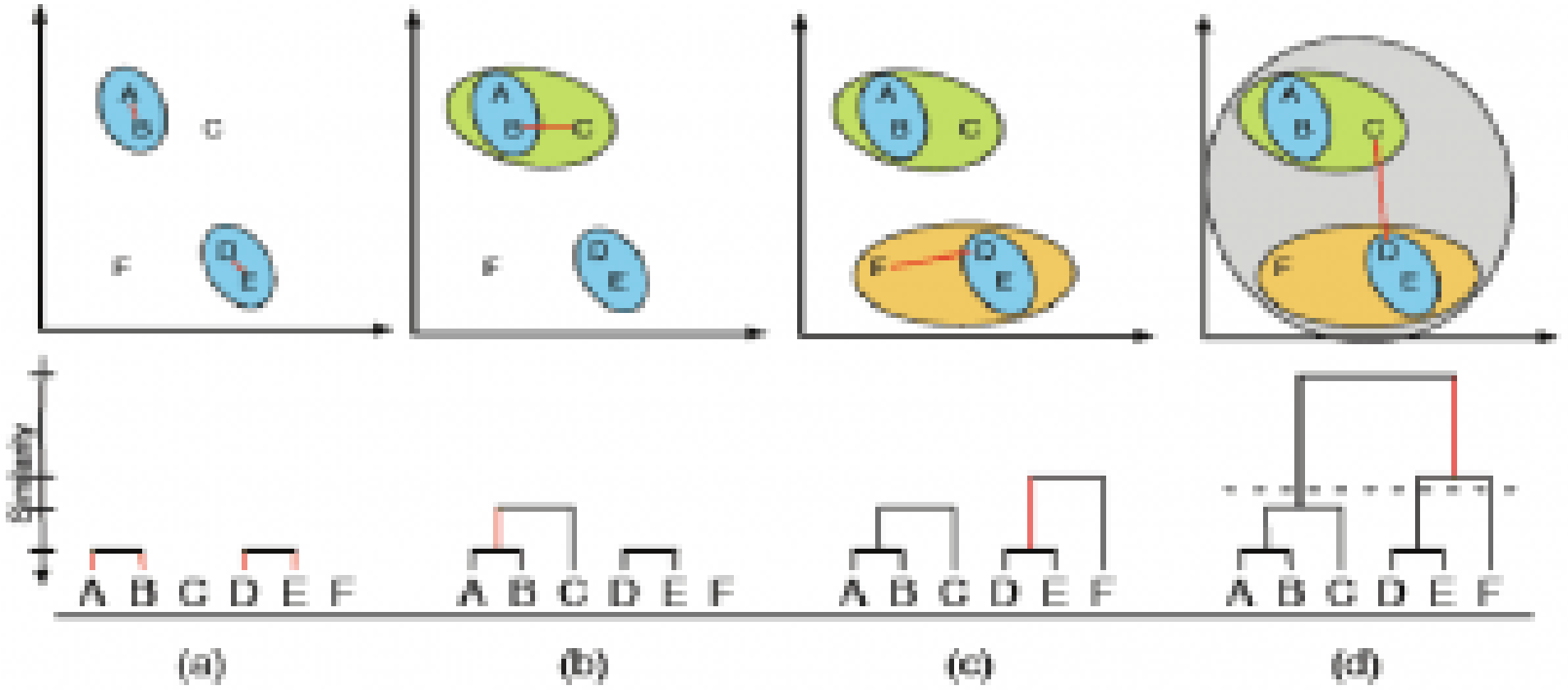
$$s = \frac{b - a}{\max(a, b)}$$





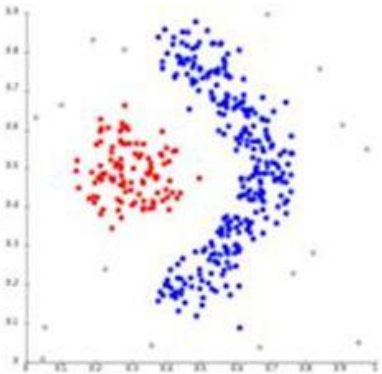
Hierarchical Clustering Algorithms

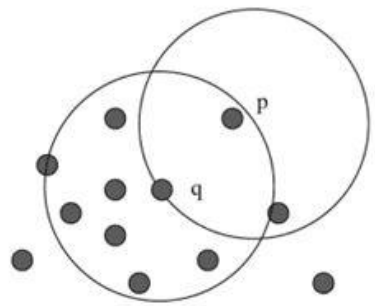
- Two main types of hierarchical clustering
  - **Agglomerative:**
    - Start with the points as individual clusters
    - At each step, merge the closest pair of clusters until only one cluster (or **k** clusters) left
  - **Divisive:**
    - Start with one, all-inclusive cluster
    - At each step, split a cluster until each cluster contains a point (or there are **k** clusters)
- Traditional hierarchical algorithms use a similarity or distance matrix
  - Merge or split one cluster at a time

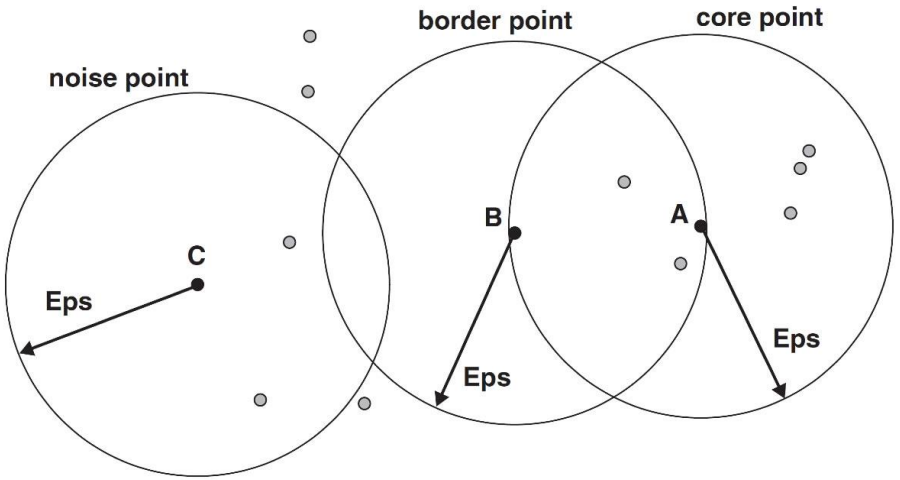

A
B
C
D
E
F
Similarity
A B C D E F
A B C D E F
A B C D E F
A B C D E F
(a)
(b)
(c)
(d)

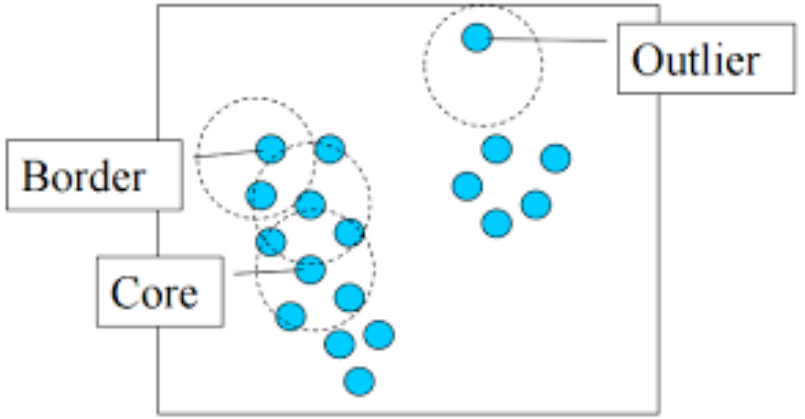
DBSCAN basic idea

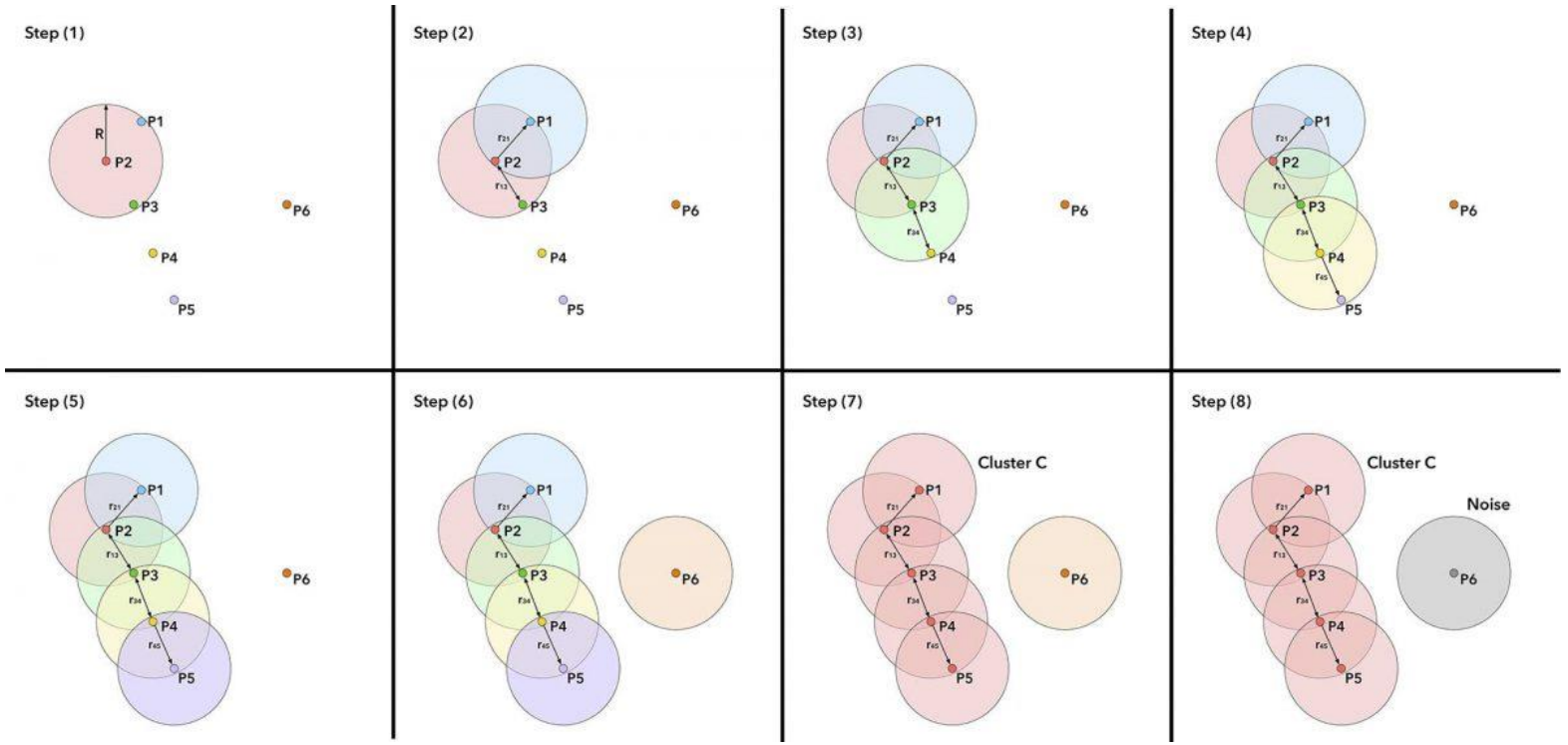
- Density-Based Spatial Clustering of Applications with Noise
  - *Munich, 1996*
  - *Derived from a human natural clustering approach*
- Input parameters
  - *The size of epsilon neighborhood – ε*
  - *Minimum points in cluster – MinPts*
- Neighborhood of a given radius *ε* has to contain at least a minimum number of points *MinPts*






noise point
border point
core point
C
Eps
B
Eps
A
Eps


Outlier
Border
Core


Step (1)
R
P1
P2
P3
P4
P5
P6
Step (2)
Step (3)
Step (4)
Step (5)
Step (6)
Step (7)
Cluster C
Step (8)
Cluster C
Noise