# OPTICS: Ordering Points To Identify Clustering Structure

OPTICS: Ordering Points To Identify Clustering Structure

- OPTICS (Ankerst, Breunig, Kriegel, and Sander, SIGMOD'99)
  - DBSCAN is sensitive to parameter setting
  - An extension: finding clustering structure
- Observation: Given a *MinPts*, density-based clusters w.r.t. a higher density are completely contained in clusters w.r.t. to a lower density
- Idea: Higher density points should be processed first—find high-density clusters first
- OPTICS stores such a clustering order using two pieces of information:
  - *Core distance* and *reachability distance*

Reachability plot for a dataset

Reachability-distance

undefined

ε'

Cluster-order of the objects
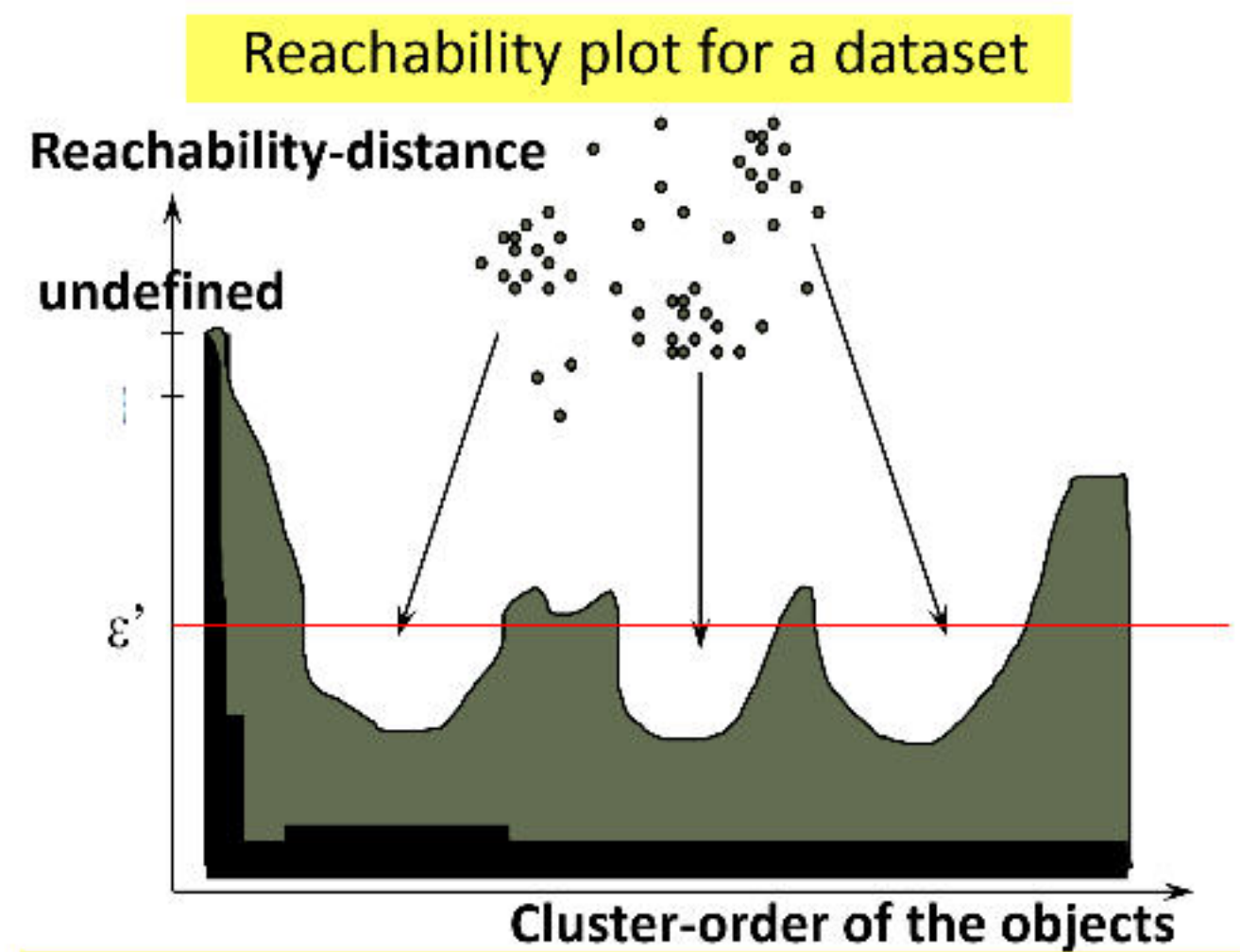
- Since points belonging to a cluster have a low reachability distance to their nearest neighbor, valleys correspond to clusters
- The deeper the valley, the denser the cluster

2

OPTICS: An Extension from DBSCAN

- **Core distance** of an object p: The smallest value ε such that the ε-neighborhood of p has at least *MinPts* objects

  Let $N_\varepsilon(p)$: ε-neighborhood of p

  ε is a distance value

  Core-distance$_{\varepsilon, MinPts}(p)$ = Undefined if card($N_\varepsilon(p)$) < *MinPts*

  MinPts-distance(p), otherwise

- **Reachability distance** of object p from core object q is the min. radius value that makes p density-reachable from q

  Reachability-distance$_{\varepsilon, MinPts}(p, q)$ =

  Undefined, *if q is not a core object*

  max(core-distance(q), distance (q, p)), *otherwise*

- Complexity: O($N \log N$) (if index-based)

  where N: # of points
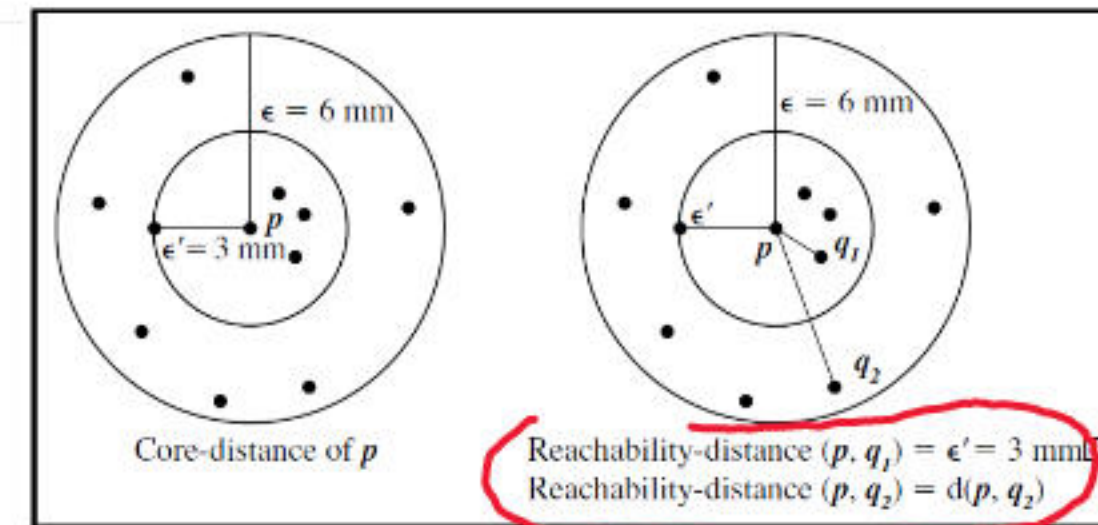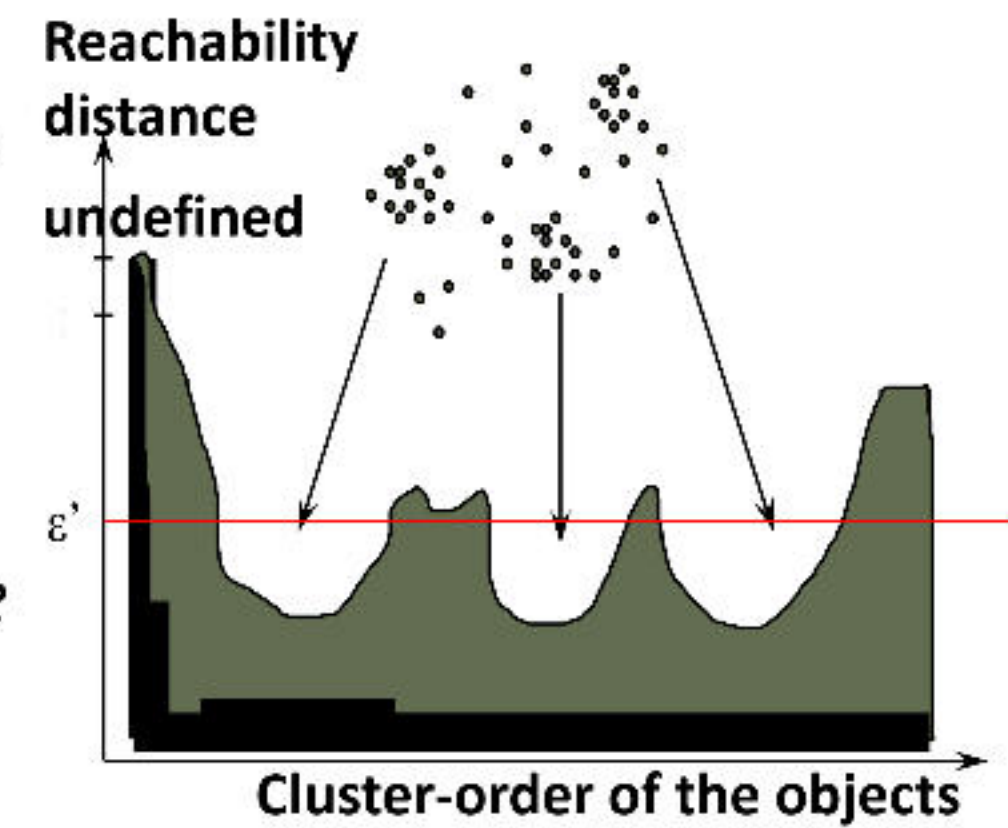

Core-distance of p Reachability-distance $(p, q_1) = \varepsilon' = 3$ mm
Reachability-distance $(p, q_2) = d(p, q_2)$

Figure 10.16: OPTICS terminology. Based on [ABKS99].

Reachability distance undefined

ε'

Cluster-order of the objects

3

OPTICS: Finding Hierarchically Nested Clustering Structures

- ❑ OPTICS produces a special cluster-ordering of the data points with respect to its density-based clustering structure
  - ❑ The cluster-ordering contains information equivalent to the density-based clusterings corresponding to a broad range of parameter settings
  - ❑ Good for both automatic and interactive cluster analysis—finding intrinsic, even hierarchically nested clustering structures
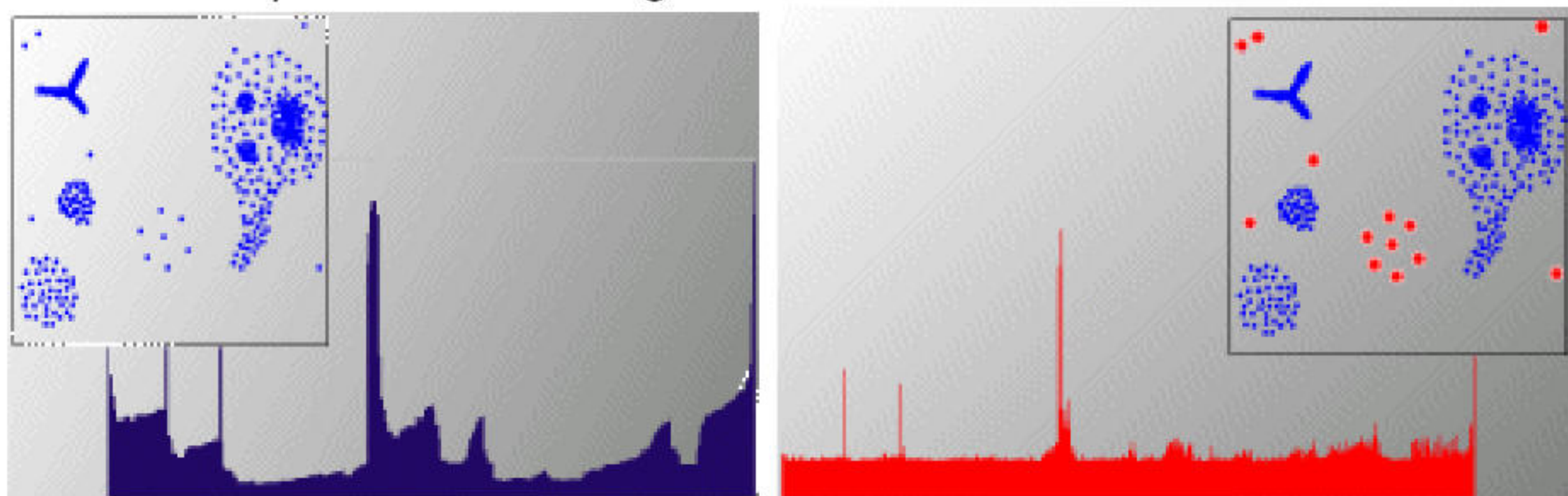
Finding nested clustering structures with different parameter settings

4