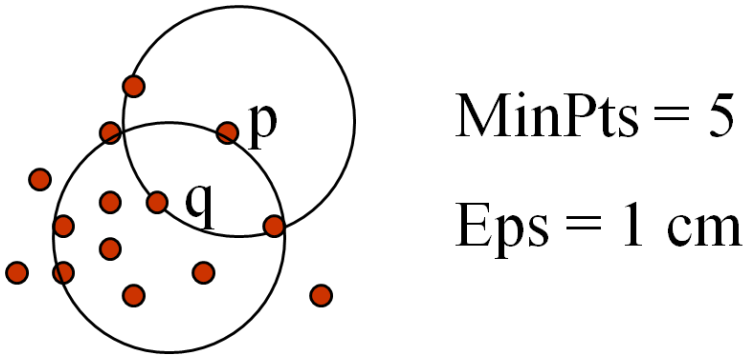
Density-Based Clustering

- **Clustering based on density** (local cluster criterion), such as density-connected points
  - Major features:
  - Discover clusters of arbitrary shape
  - Handle noise
  - One scan
  - Need density parameters as termination condition

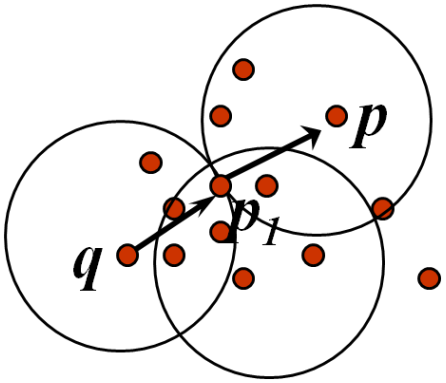
- **Representative algorithms:**
  - DBSCAN (Ester et al., 1996)
  - DENCLUE (Hinneburg & Keim, 1998)

- **Clustering Parameters:**
  - ***Eps****:* Maximum radius of neighborhood
  - ***MinPts****:* Minimum number of points in an Eps-neighborhood of a point
  - $N_{Eps}(p) = \{q \in D \mid dist(p,q) <= Eps\}$

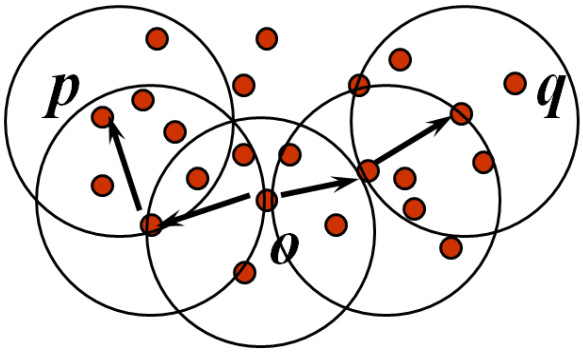
- **Directly density-reachable:**
  - A point ***p*** is directly density-reachable from a point ***q*** wrt. ***Eps***, ***MinPts*** iff
    - 1) ***p*** belongs to $N_{Eps}(q)$
    - 2) ***q*** is a core point:
      - $|N_{Eps}(q)| >= MinPts$



- **Density-reachable:**
  - A point *p* is density-reachable from a point *q* wrt. *Eps*, *MinPts* if there is a chain of points $p_1, \ldots, p_n$, $p_1 = q$, $p_n = p$ such that p_{i+1} is directly density-reachable from p_i

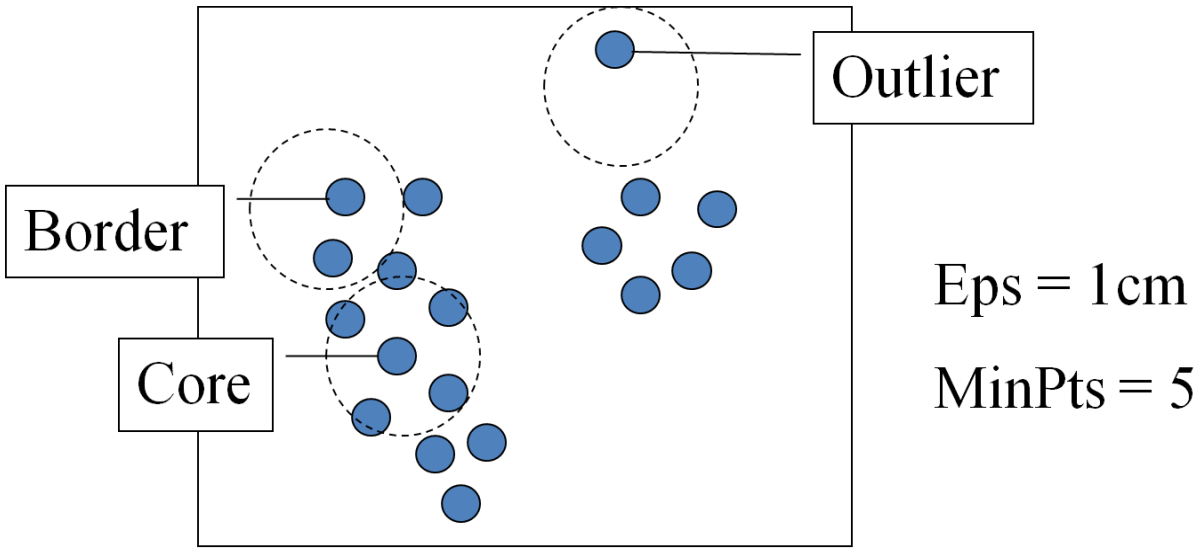


- **Density-connected**
  - A point *p* is density-connected to a point *q* wrt. *Eps*, *MinPts* if there is a point *o* such that both, *p* and *q* are density-reachable from *o* wrt. *Eps* and *MinPts*.



- Relies on a *density-based* notion of cluster: A *cluster* is defined as a maximal set of density-connected points

- Discovers clusters of arbitrary shape in spatial databases with noise



- **DBSCAN: The Algorithm**
  - Arbitrarily select a point ***p***
  - Retrieve all points density-reachable from ***p*** wrt ***Eps*** and ***MinPts***.
  - If ***p*** is a core point, a cluster is formed.
  - If ***p*** is a border point, no points are density-reachable from ***p*** and DBSCAN visits the next point of the database.
  - Continue the process until all of the points have been processed.