### Density-Based Clustering

Math 3210 By Fatine Bourkadi

### **Clustering Definition**

Clustering is the process of grouping a set of physical objects into classes of similar objects

It is similar to classification in that data are grouped. However, unlike classification, the groups are not predefined. Instead, the grouping is accomplished by finding similarities between data according to characteristics found in the actual data. (Dunham, 2003).

#### Where we use clustering?

**O**Business

Biology

**O**Statistics

Data Mining

#### **Clustering Algorithms**

Partitional clustering

Hierarchical clustering

Density-based clustering

Distribution-based clustering

Centroid-based clustering



## Density-based clustering definition

Is a set of density-connected objects that is maximal with respect to density-reachability. Every object not contained in any cluster is considered to be noise. That is, for each data point within a given cluster, the neighborhood of a given radius has to contain at least a minimum number of points. Such an algorithm can be used to filter out noise (outliers) and discover clusters of arbitrary shape.(Han, 2001)

#### Density-Based Clustering definition

ODefining density-based clustering requires new definitions.

#### Density-Based Clustering definition

- The neighborhood within a radius ε given object is called the ε-neighborhood of the object.
- If the *ɛ*-neighborhood of an object contains at least a minimum number, *MinPts*, of objects, then the object is called a core object.
- 3. Given a set of objects, D, we say that an object p is directly density-reachable from object q if p is within the *ɛ*-neighborhood of q, and q is a core object.

## Density-based clustering definition

- 4. An object p is **density-reachable** from object q with respect to  $\varepsilon$  and *MinPts* in a set of objects, D, if there is a chain of objects  $p \downarrow 1$ ,...,  $p \downarrow n = q$  and  $p \downarrow n = p$  such that  $p \downarrow i + 1$  is directly density-reachable from  $p \downarrow i$  with respect to  $\varepsilon$  and *MinPts*, for  $1 \le i \le n$ ,  $p \downarrow n \in D$ .
- 5. An object p is **density-connected** to object q with respect to  $\varepsilon$  and MinPts in a set of object, D, if there is an object  $o \in D$  such that both p and q are density-reachable from o with respect to  $\varepsilon$  and MinPts. (Han,2001)

# Density-based clustering definition



#### Figure 1. Density-reachability and connectivity

### Summary

Today we cover the following:
Clustering
Clustering applications
Clustering methods
Focusing on density-based clustering

#### References

Dunham, M. H. (2003). Data Mining Introductory and Advanced Topics. New Jersey: Pearson Education, Inc. http://en.wikipedia.org/w/index.php?title=Special %3ASearch&search=DENSITY-BASED+CLUSTERING. (n.d.).

http://en.wikipedia.org/wiki/DBSCAN. (n.d.).

Jiawei Han, Micheline Kamber. (2001). *Data Mining: Concepts and Techniques.* London, United Kingdom: Academic Press.

Micheal Ankerst, M. M.-P. (1999). OPTICS: Ordering Points To Identify the Clustering Structure. Philadelphia: Proc. ACM SIGMOD'99 Int. Conf.