

Grid-Based Clustering Methods

- ❑ Grid-Based Clustering: Explore multi-resolution grid data structure in clustering
 - ❑ Partition the data space into a finite number of cells to form a grid structure
 - ❑ Find clusters (dense regions) from the cells in the grid structure
- ❑ Features and challenges of a typical grid-based algorithm
 - ❑ Efficiency and scalability: # of cells \ll # of data points
 - ❑ Uniformity: Uniform, hard to handle highly irregular data distributions
 - ❑ Locality: Limited by predefined cell sizes, borders, and the density threshold
 - ❑ Curse of dimensionality: Hard to cluster high-dimensional data
- ❑ Methods to be introduced
 - ❑ **STING** (a Statistical Information Grid approach) (Wang, Yang and Muntz, VLDB'97)
 - ❑ **CLIQUE** (Agrawal, Gehrke, Gunopulos, and Raghavan, SIGMOD'98)
 - ❑ Both grid-based and subspace clustering