

The background of the slide is a complex, abstract composition. It features a central white rectangular area containing the title. Surrounding this area are various geometric and data-related patterns. On the left, there's a vertical strip with a grid of small circles, some of which are highlighted in orange and red. To the right, there's a large, intricate network of thin, reddish-brown lines connecting numerous small green dots, resembling a complex graph or a spatial network. The overall color palette is muted, with earthy tones like beige, brown, and olive green, accented by the vibrant orange and red highlights.

# **Grid-Based Clustering Methods**



# 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:  $\# \text{ of cells} \ll \# \text{ of data points}$
  - 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