

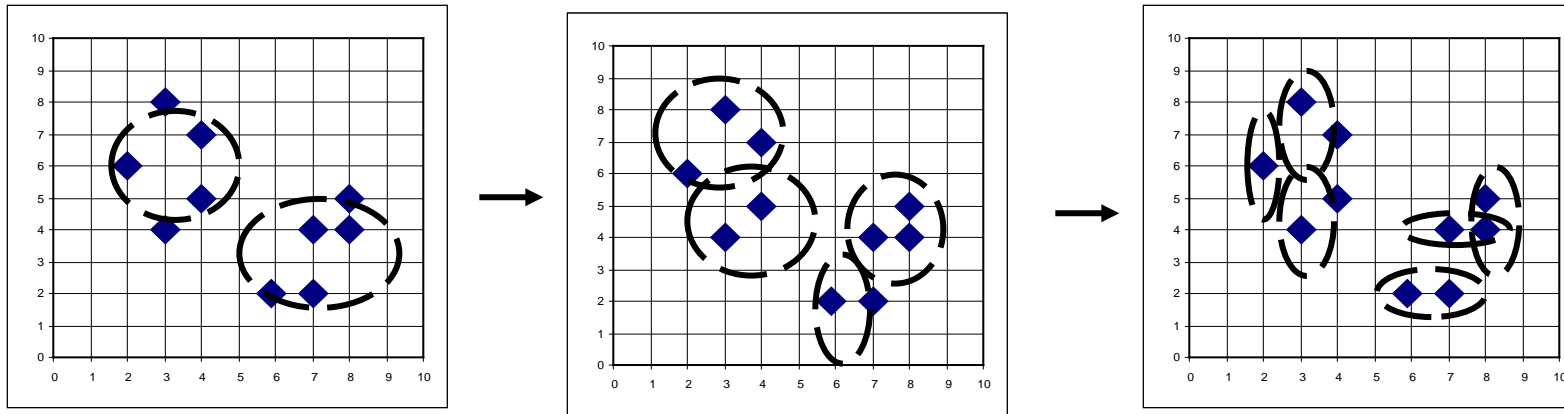
The background features a complex network of thin, light-colored lines forming a web-like structure. Scattered throughout are numerous small, colored dots in shades of green, blue, and orange. A prominent, thicker red line forms a large, irregular loop in the center. The overall aesthetic is technical and data-driven.

# **Divisive Clustering Algorithms**



# Divisive Clustering

- ❑ DIANA (Divisive Analysis) (Kaufmann and Rousseeuw, 1990)
  - ❑ Implemented in some statistical analysis packages, e.g., Splus
- ❑ Inverse order of AGNES: Eventually each node forms a cluster on its own



- ❑ Divisive clustering is a top-down approach
  - ❑ The process starts at the root with all the points as one cluster
  - ❑ It recursively splits the higher level clusters to build the dendrogram
  - ❑ Can be considered as a global approach
  - ❑ More efficient when compared with agglomerative clustering

# More on Algorithm Design for Divisive Clustering

---

- ❑ Choosing which cluster to split
  - ❑ Check the sums of squared errors of the clusters and choose the one with the largest value
- ❑ Splitting criterion: Determining how to split
  - ❑ One may use Ward's criterion to chase for greater reduction in the difference in the SSE criterion as a result of a split
  - ❑ For categorical data, Gini-index can be used
- ❑ Handling the noise
  - ❑ Use a threshold to determine the termination criterion (do not generate clusters that are too small because they contain mainly noises)