



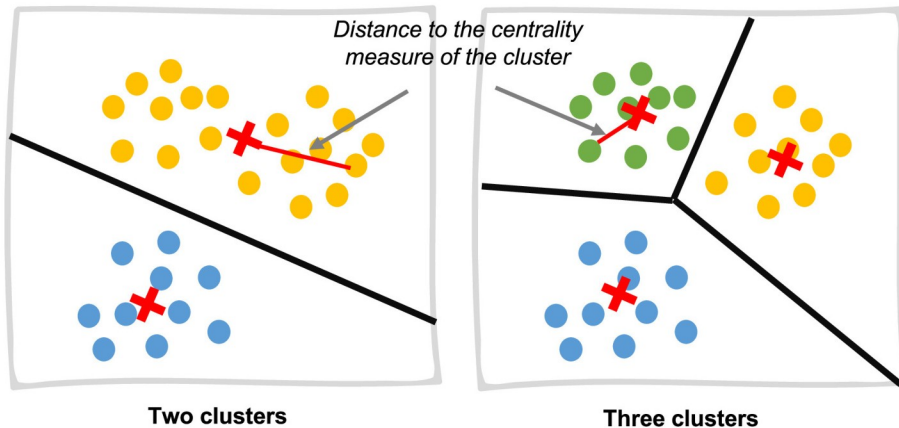
Tutorial 11

COMP90014 Algorithm for Bioinformatics

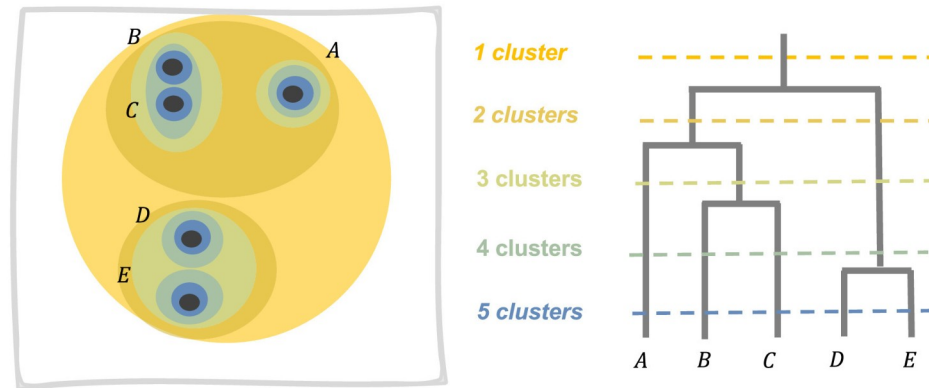
Semester 2, 2025

Main Types of Clustering Models

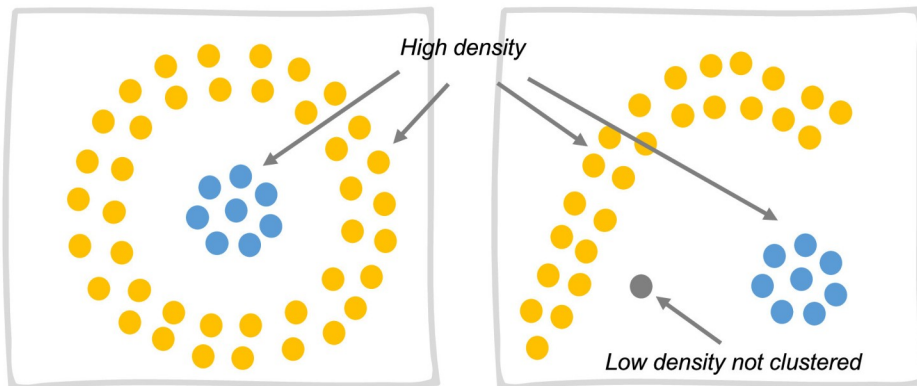
(A) Centre-based partitioning clustering



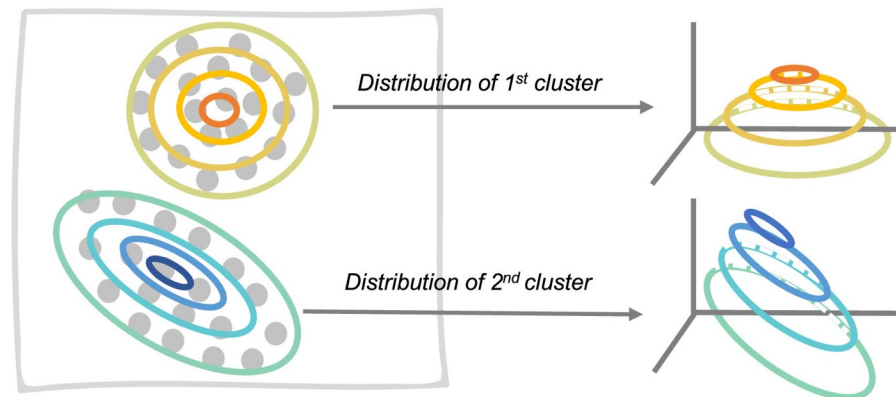
(B) Hierarchical clustering



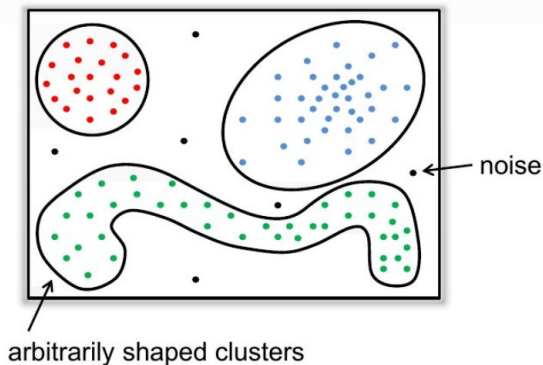
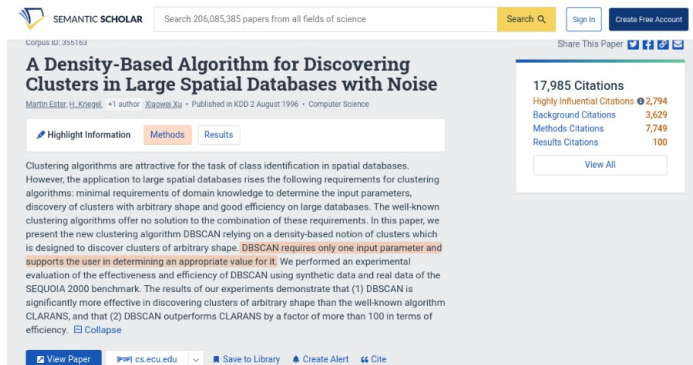
(C) Density-based clustering



(D) Model-based clustering



DBSCAN



- first density-based clustering algorithm
- one of the most widely used/cited clustering algorithms

Intuition:

- a cluster is a region of high density**
- noise points lie in regions of low density

We need to:

- define neighbourhood of a data point
- define high density

Definitions

ϵ -neighbourhood: objects within a radius ϵ of an object.

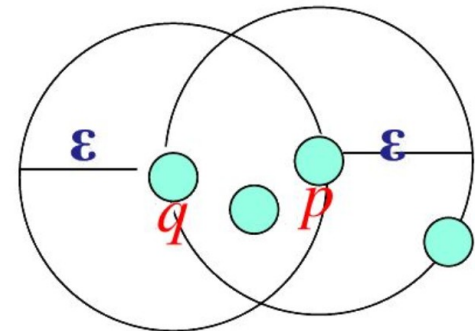
ϵ : input parameter.

High-density: ϵ -neighbourhood of an object contains at least minpts of objects.

minpts: input parameter.

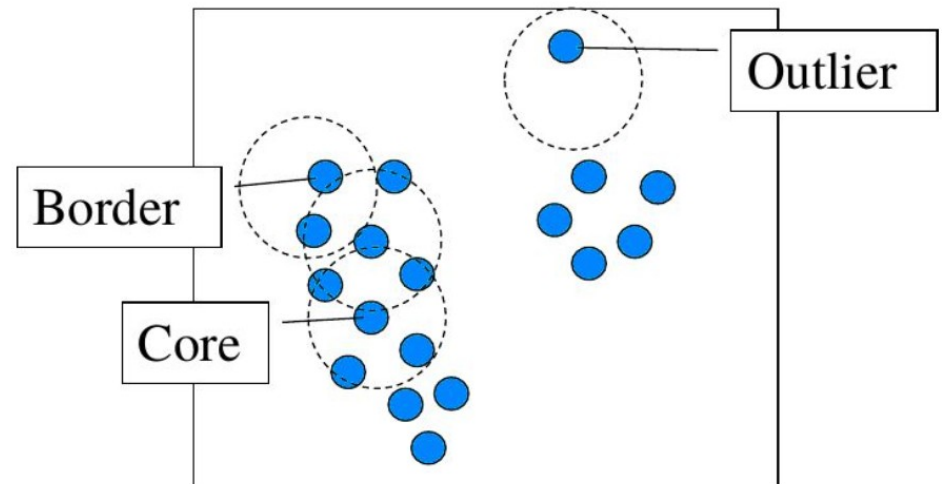
$$N_{\epsilon}(p) : \{q \mid d(p, q) \leq \epsilon\}$$

ϵ -neighbourhood of p and q : Density of p is “high” (minpts = 4);
Density of q is “low” (minpts = 4)



Point Types

- Core points:
 - Have $\geq \text{min_pts}$ points within radius ε
- Border points:
 - Neighbour to a core point
- Noise points:
 - Everything else



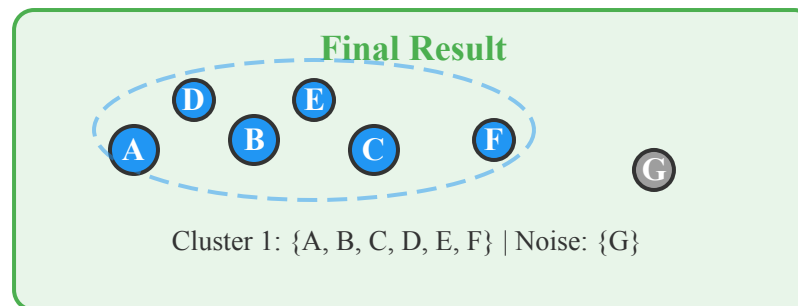
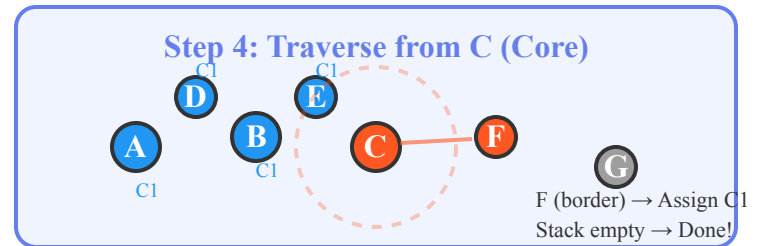
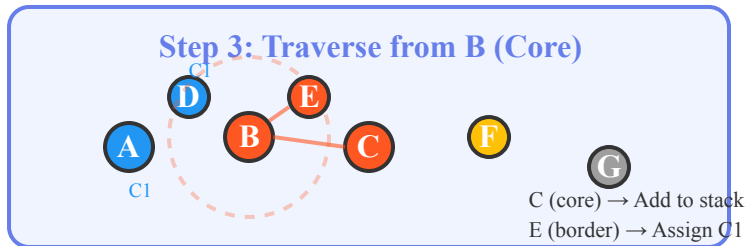
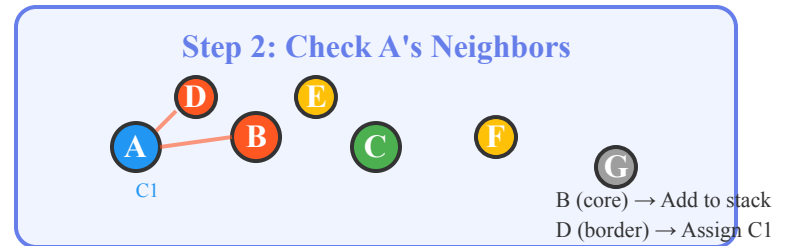
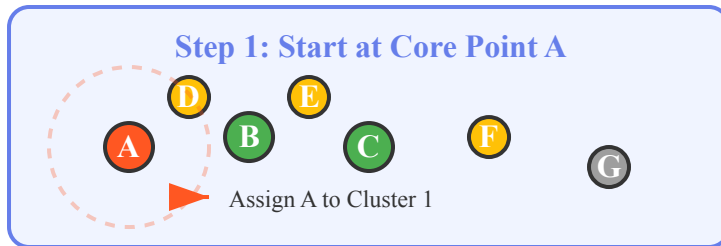
Algorithm Steps

1. Start with a core point

2. Get its neighbours

- Core point, recursively. (Expand the cluster)
- Border point, join cluster but don't expand
- Already assigned and noise – skip

DBSCAN Traverse Process



Legend: ● Processing ● Core (unvisited) ● Border (unvisited) ● Assigned to Cluster ● Noise