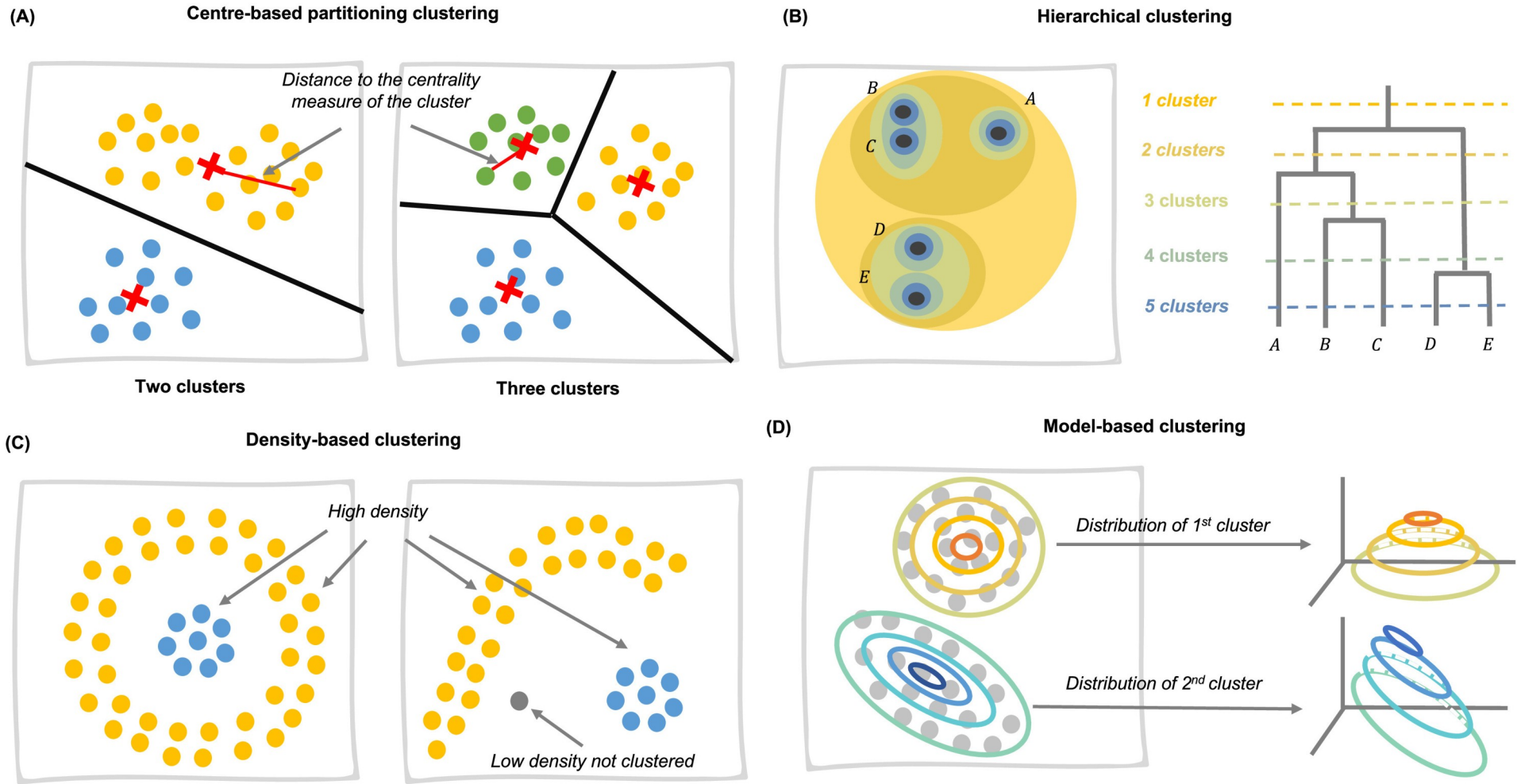
Tutorial 11

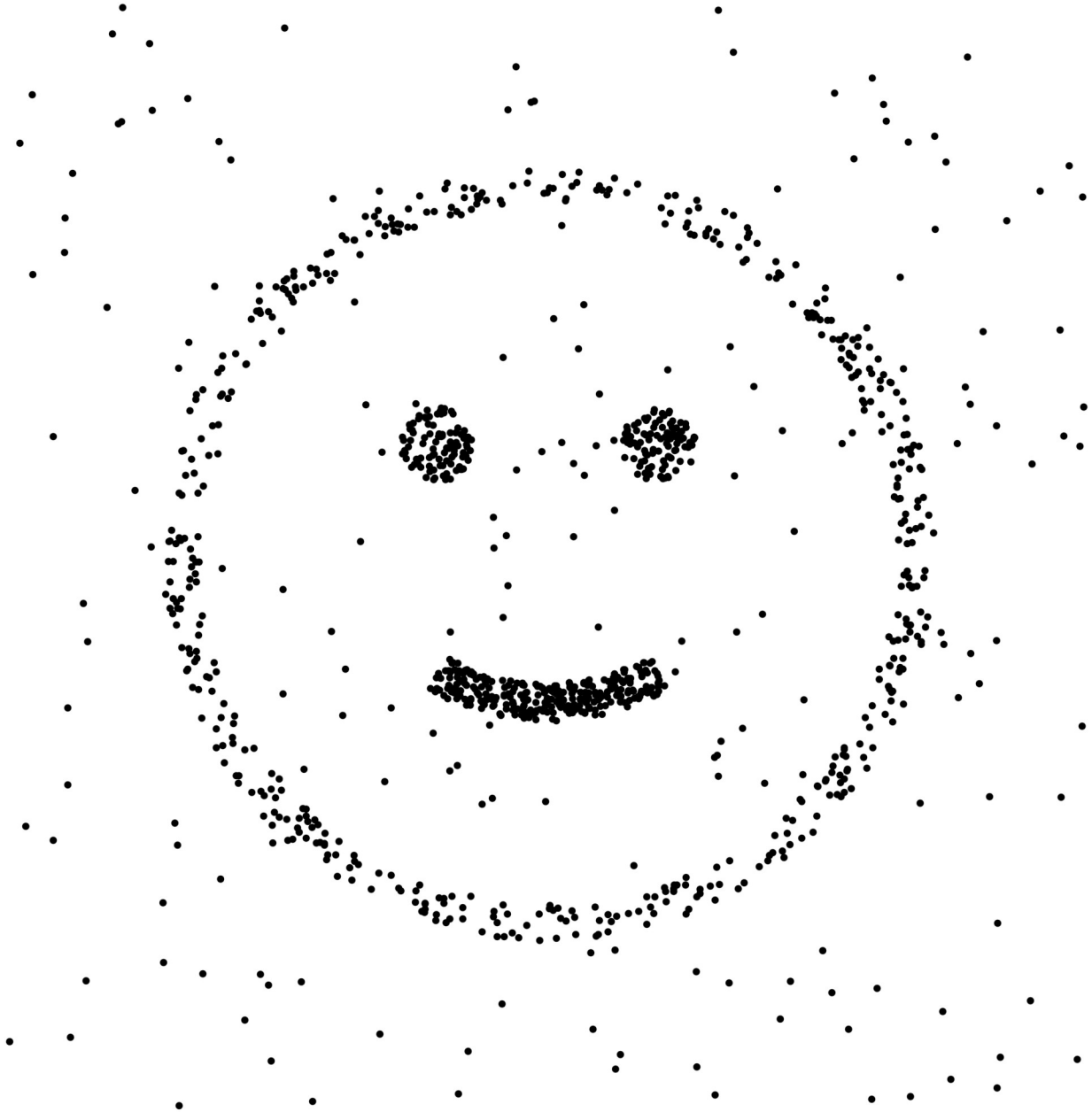
COMP90014 Algorithm for Bioinformatics

Semester 2, 2025

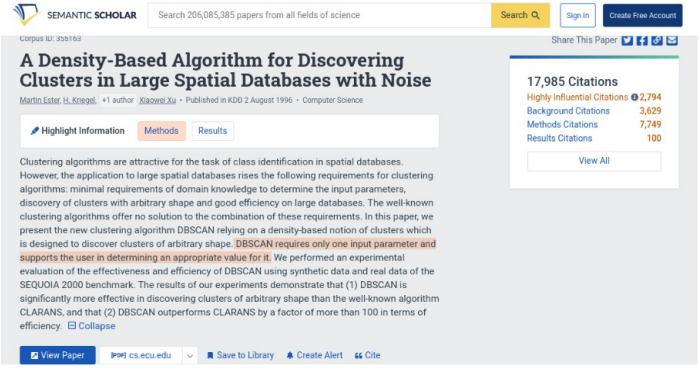
Main Types of Clustering Models

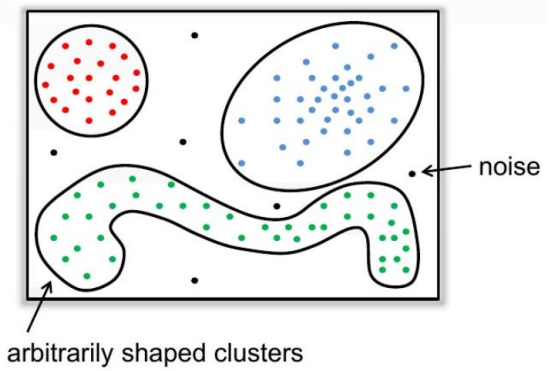


Gao, Caroline X., et al. "An overview of clustering methods with guidelines for application in mental health research." Psychiatry Research 327 (2023): 115265.



DBSCAN





- first density-based clustering algorithm
- one 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

Ester *et al.*, 1996. KDD-96 Proceedings

Definitions

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

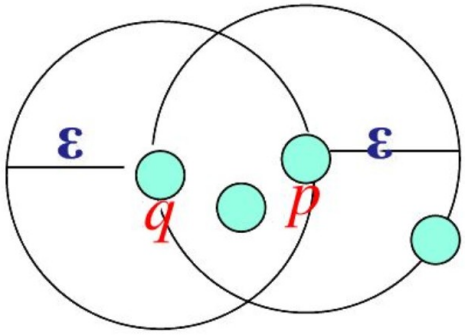
$$N_\varepsilon(p) : \{q \mid d(p,q) \leq \varepsilon\}$$

ε: **input parameter**.

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

minpts: **input parameter**.

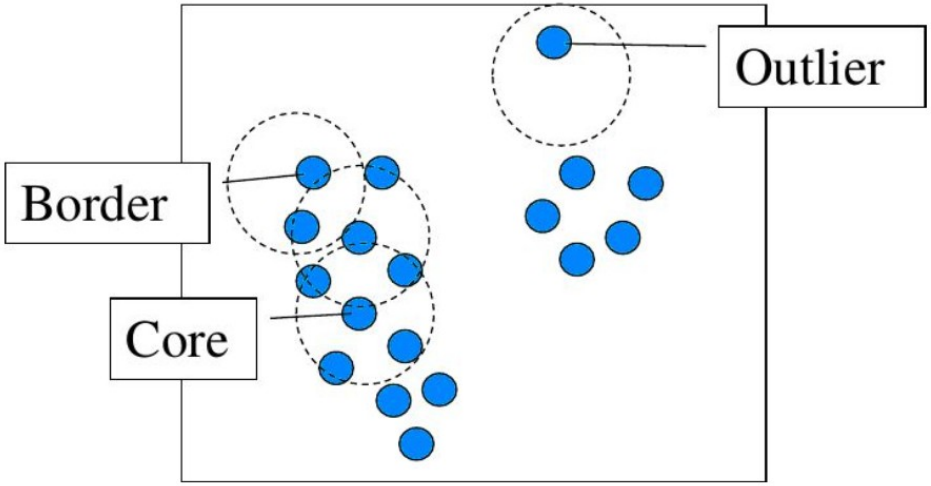
ε-neighbourhood of p and q: Density of p is "high" (minpts $= 4$);
Density of q is "low" (minpts $= 4$)



Shah, 2012 10.1109/NUICONE.2012.6493211

Point Types

- Core points:
  - Have >= 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