OPTICS

Renato R. Maaliw III, *DIT*
College of Engineering
Southern Luzon State University
Lucban, Quezon, Philippines

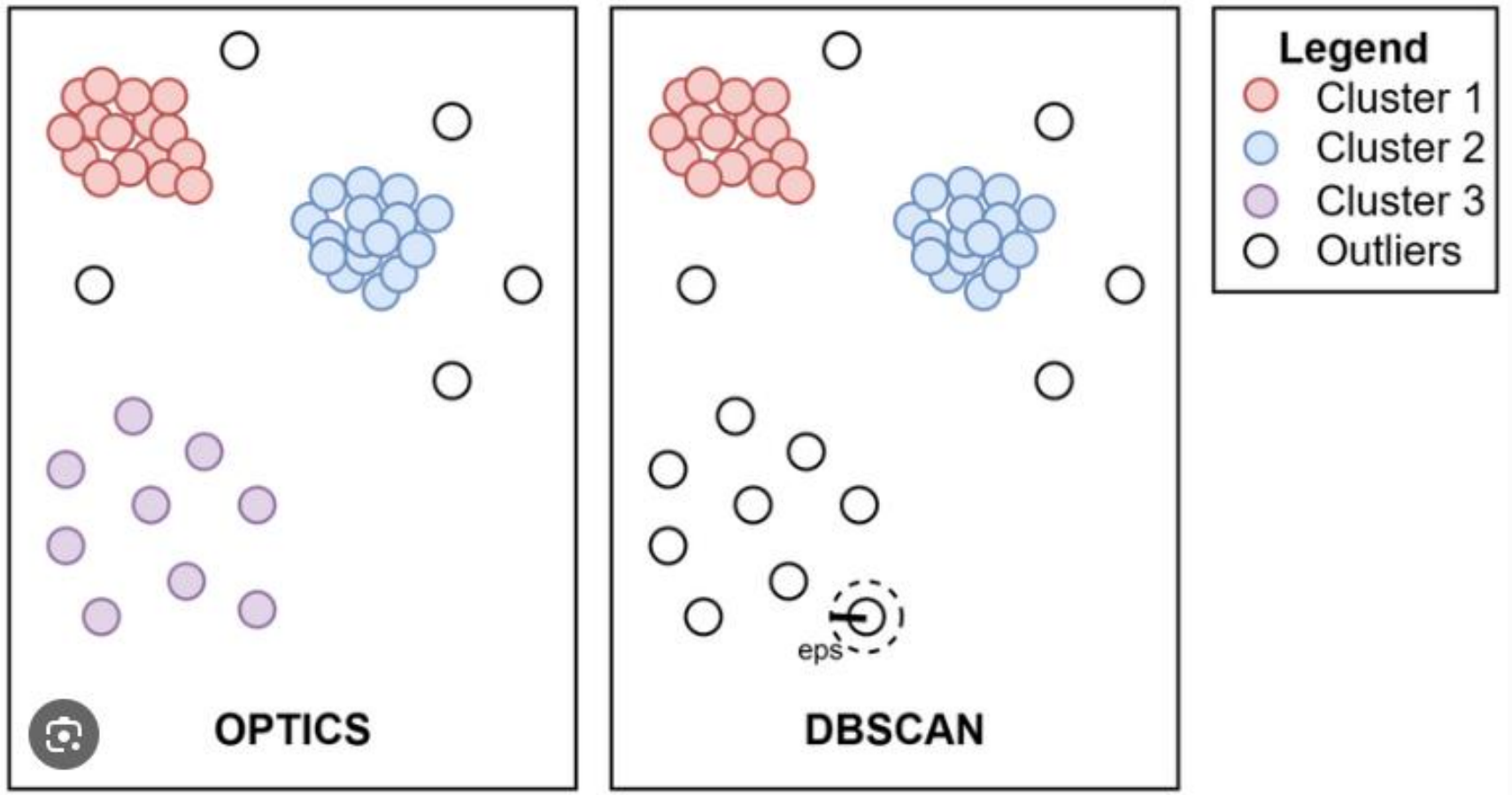
Ordering Points To Identify the Clustering Structure

- is a density-based clustering algorithm that is an extension of DBSCAN
- designed to address some of the limitations of DBSCAN by providing a richer and more robust view of the data's cluster structure

Core Concept

- it does not explicitly produce a clustering like DBSCAN, it creates an augmented ordering of database representing its density-based structure

Cognate/Professional Electives



OPTICS is ideal for:

- Datasets with varying density clusters
- Understanding hierarchical density-based structures
- Applications where DBSCAN would struggle due to a single global epsilon value

Core Components:

- **Core Distance:** Minimum radius required to include at least minPts neighbors around a point
- **Reachability Distance:** Distance between two points p and q defined as:
- **Ordering:** OPTICS builds an ordering of data points based on their reachability distances which allows for easier visualization and cluster extraction

Key Differences Between OPTICS and Others

 vs. DBSCAN

OPTICS avoids the need for a global epsilon by building a reachability plot.

Better for datasets with **clusters of varying density**.

DBSCAN gives a hard clustering; OPTICS gives a **cluster ordering** that can be used flexibly.

◈ vs. KMeans

KMeans assumes spherical clusters and requires specifying k.

OPTICS detects arbitrary shapes and automatically infers the number of clusters.

KMeans is distance-based, while OPTICS is density-based.

 vs. Hierarchical Clustering

Hierarchical methods build a dendrogram but don’t consider **density**.

OPTICS provides **density-aware hierarchical structures**.

Hierarchical methods are **greedy**, while OPTICS is more **global** in approach.

 vs. Gaussian Mixture Models

GMM assumes data is generated from a **mixture of Gaussians**.

OPTICS is **non-parametric** and makes **no assumptions** about distribution.

GMM fails with non-Gaussian, irregularly shaped clusters.

vs. Spectral Clustering

Spectral clustering relies on the eigenvalues of a similarity matrix, which is computationally expensive.

Spectral clustering works well for complex boundaries, but doesn't handle density variation or noise well.

OPTICS is more **robust to noise and density fluctuations**.

Thank you very much for listening.