



DEEP
LEARNING
INSTITUTE

Georgia
Tech



PRAIRIE VIEW
A&M UNIVERSITY

DLI Accelerated Data Science Teaching Kit

Lecture 15.4 - DBSCAN

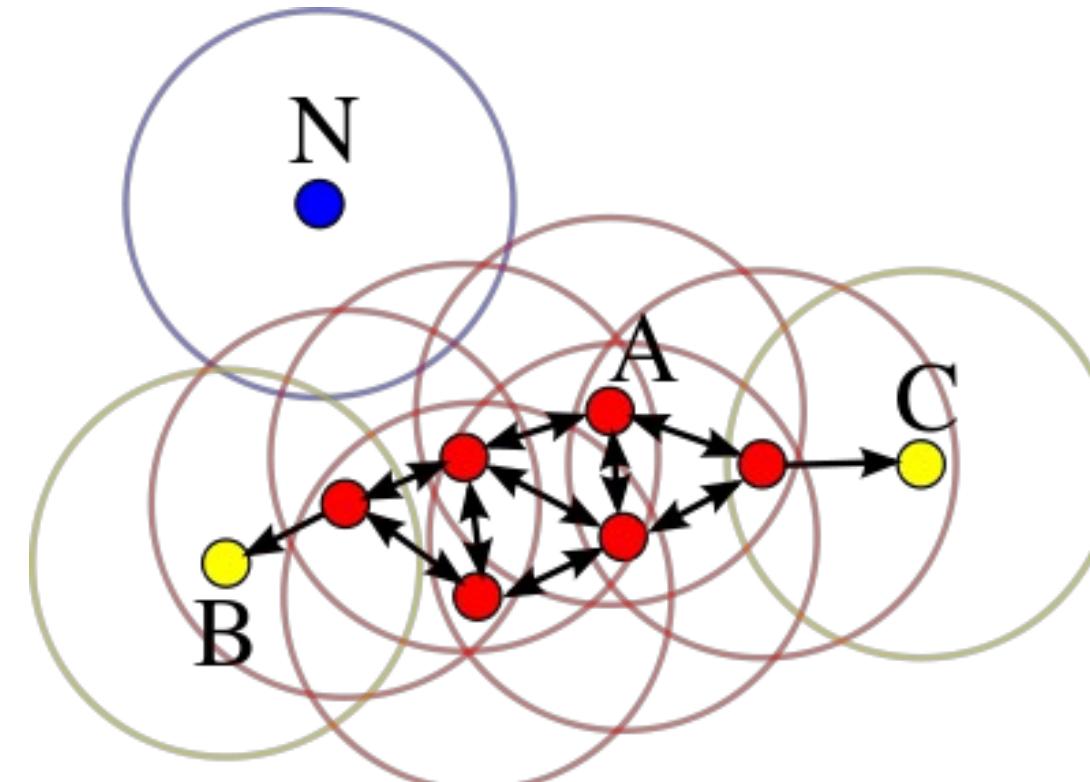


The Accelerated Data Science Teaching Kit is licensed by NVIDIA, Georgia Institute of Technology, and Prairie View A&M University under the [Creative Commons Attribution-NonCommercial 4.0 International License](#).

DBSCAN

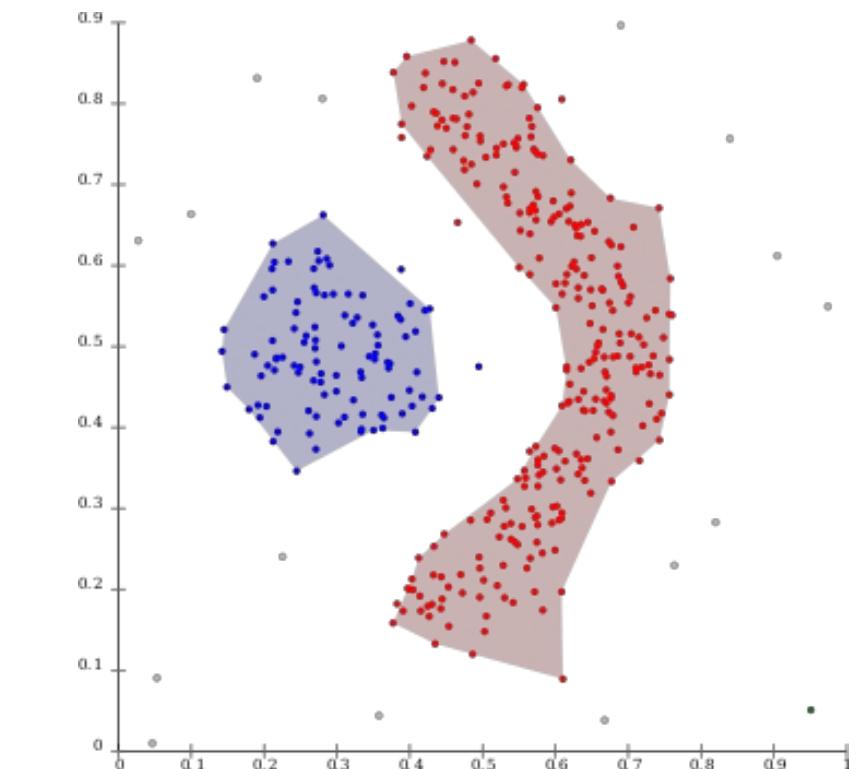
“Density-based spatial clustering with noise”

Received “**test-of-time award**” at KDD’14 — an extremely prestigious award.



Only need two parameters:

1. “radius” epsilon
2. minimum number of points (e.g., 4) to form a dense region

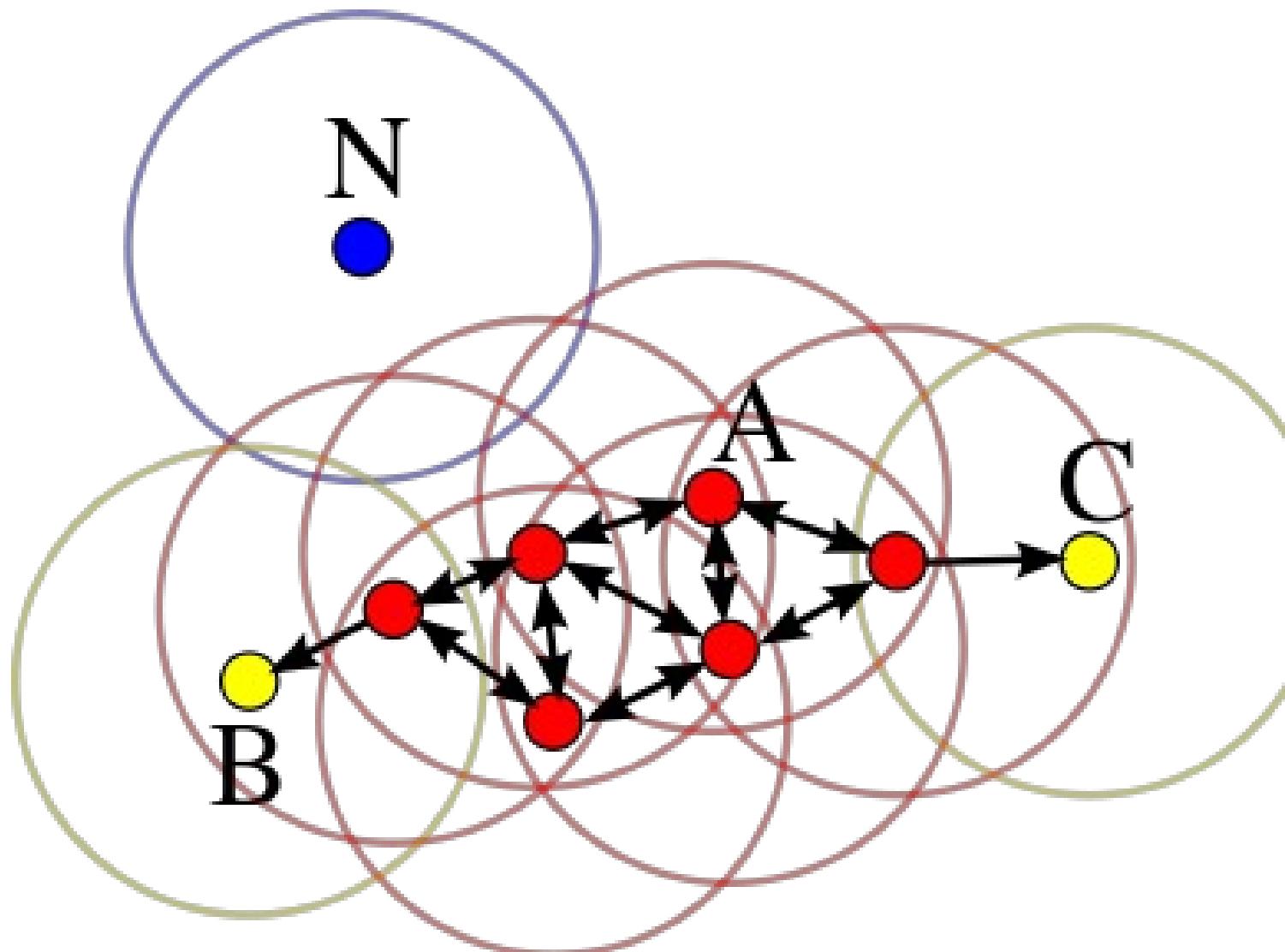


(Yellow “border points” are density-reachable from red “core points”, but not vice-versa.)

DBSCAN

Main ideas

- Group closely-packed “high-density” points (i.e., many nearby neighbors)
- Outliers in “low-density” not grouped (i.e., few or no nearby neighbors)



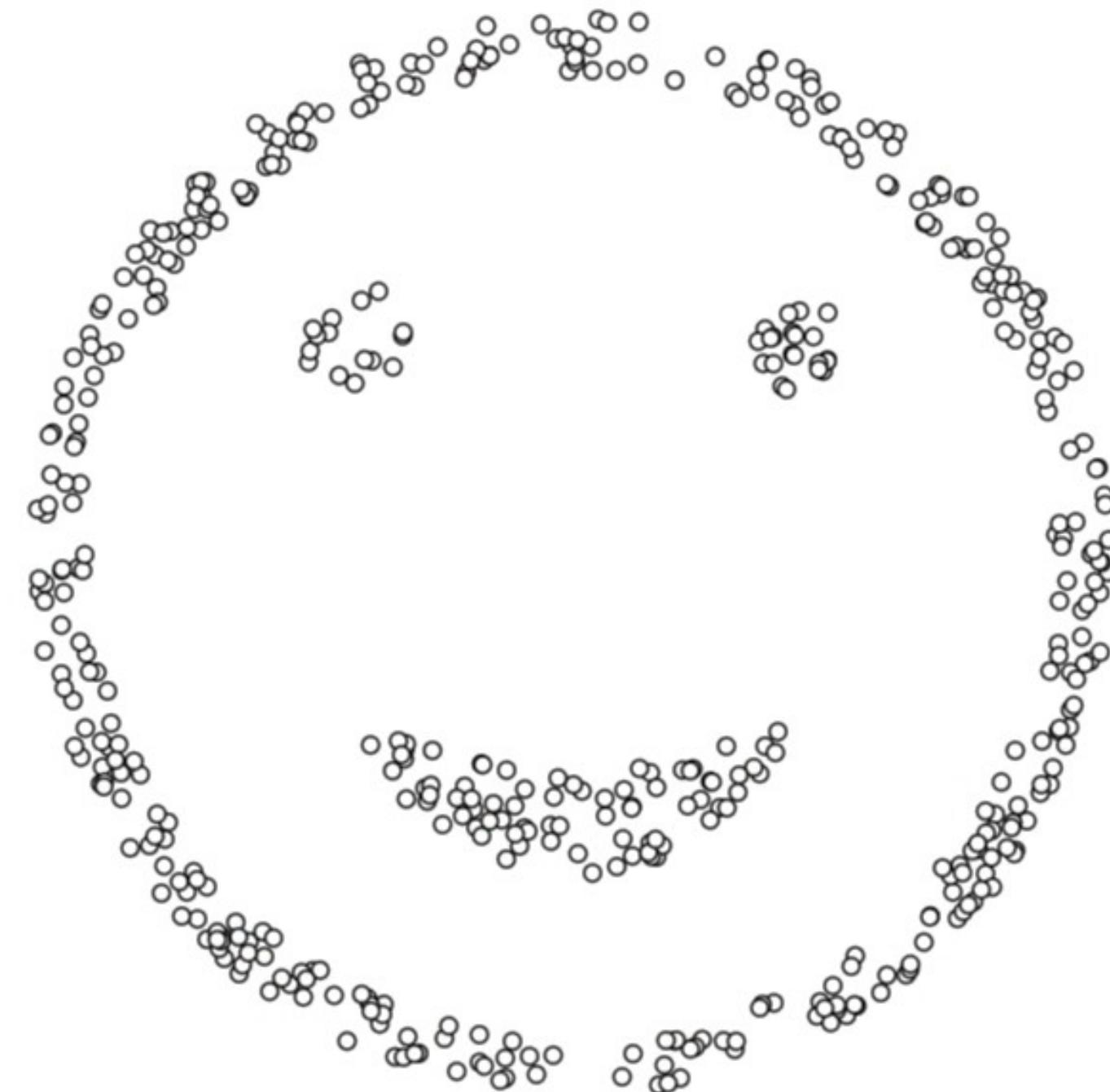
Red core points (each surrounded by 4 points, including self), reachable from each other, thus from a cluster

Yellow density-reachable points from **red** points, thus also join the cluster

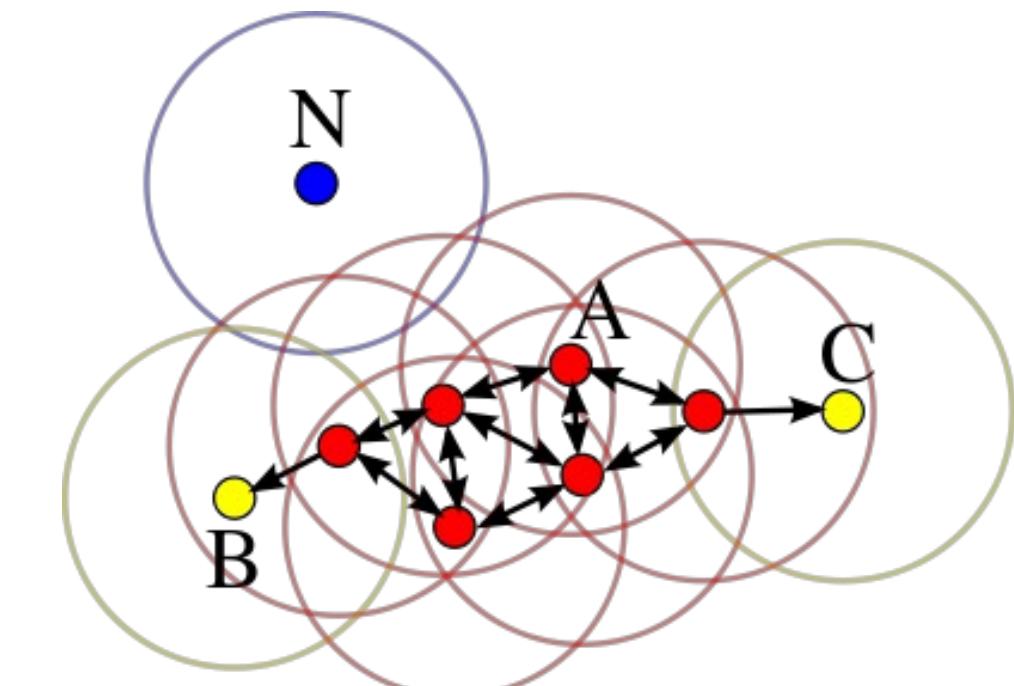
Blue outlier point not reachable from other points

Excellent Interactive DBSCAN Demo

<https://www.naftaliharris.com/blog/visualizing-dbscan-clustering/>



epsilon = 1.00
minPoints = 4



DLI Accelerated Data Science Teaching Kit

Thank You



DEEP
LEARNING
INSTITUTE

Georgia
Tech



PRAIRIE VIEW
A&M UNIVERSITY