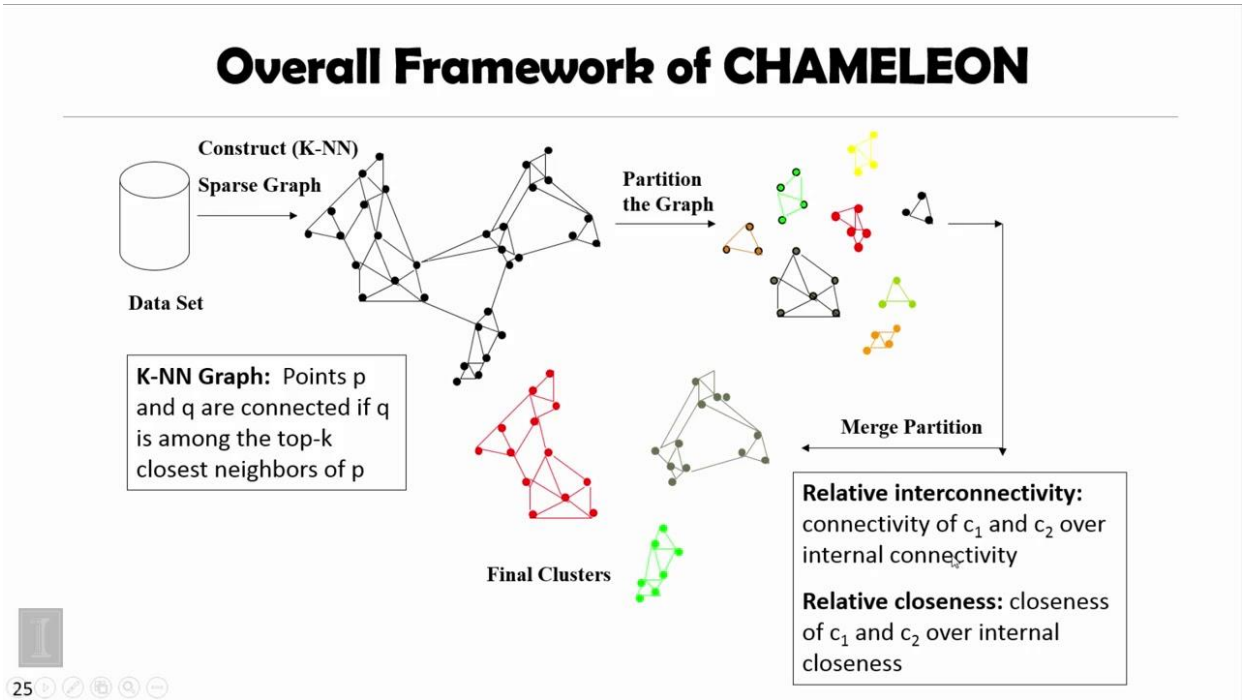
CSE454 Data Mining

Tuba TOPRAK

161044116

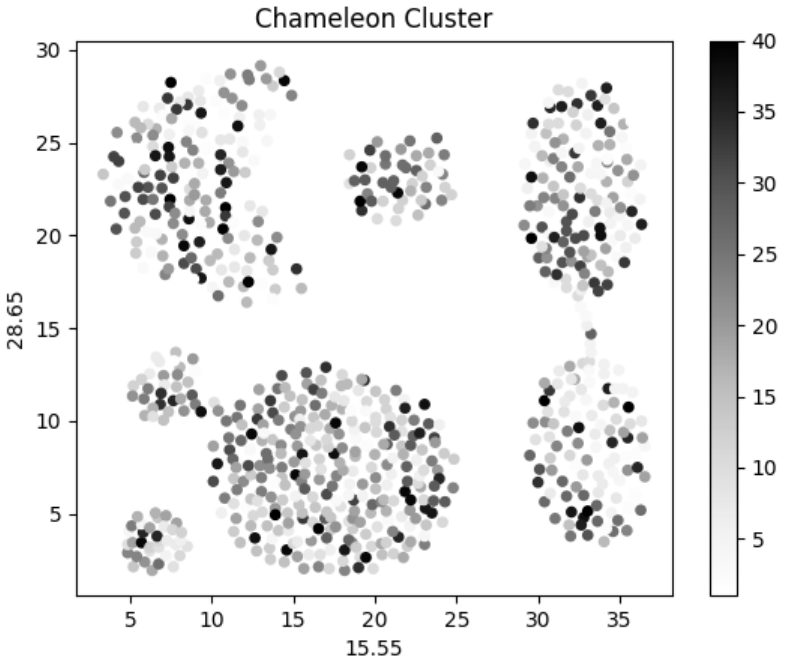
Homework 1

3-) I used the well-known Aggregation dataset for clustering.

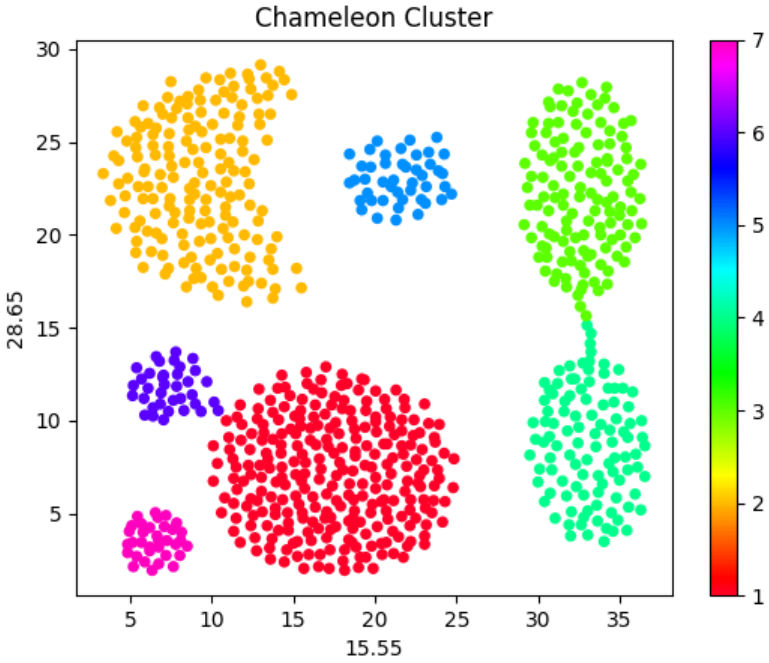


Chameleon working logic is as you see above.

The dataset we have before clustering is black as below.



After clustering, as you can see, 7 clusters were separated.



Advantages:

- This algorithm is proven to find clusters of diverse shapes, densities, and sizes in two-dimensional space
- incremental algorithm
- CHEMELEON is an efficient algorithm that uses a dynamic model to obtain clusters of arbitrary shapes and arbitrary densities

Disadvantages:

- Time complexity of CHAMELEON algorithm in high dimensions is O (n2).
- CHAMELEON is known for low dimensional spaces, and was not applied to high dimensions

5-)

Hierarchical

Birch -- Time complexity O(n)

Cure -- Time complexity O(n^2logn)

Rock -- Time complexity O(n^2logn)

Chameleon --Time complexity O(n^2)

Partitioning

K-means -- Time complexity O(n k d)

K-medoids -- Time complexity O(k(n-k)^2)

Density-based

DBSCAN -- Time complexity O(nlogn)

OPTICS -- Time complexity O(nlogn)

Grid-based

STING -- Time complexity O(n)

CLIQUE -- Time complexity O(n+d^2)