

# Initialization of K-Means

- ❑ Different initializations may generate rather different clustering results (some could be far from optimal)
- ❑ Original proposal (MacQueen'67): Select  $K$  seeds randomly
  - ❑ Need to run the algorithm multiple times using different seeds
- ❑ There are many methods proposed for better initialization of  $k$  seeds
  - ❑ ***K-Means++*** (Arthur & Vassilvitskii'07):
    - ❑ The first centroid is selected at random
    - ❑ The next centroid selected is the one that is farthest from the currently selected (selection is based on a weighted probability score)
    - ❑ The selection continues until  $K$  centroids are obtained

