

PFLOCK Report

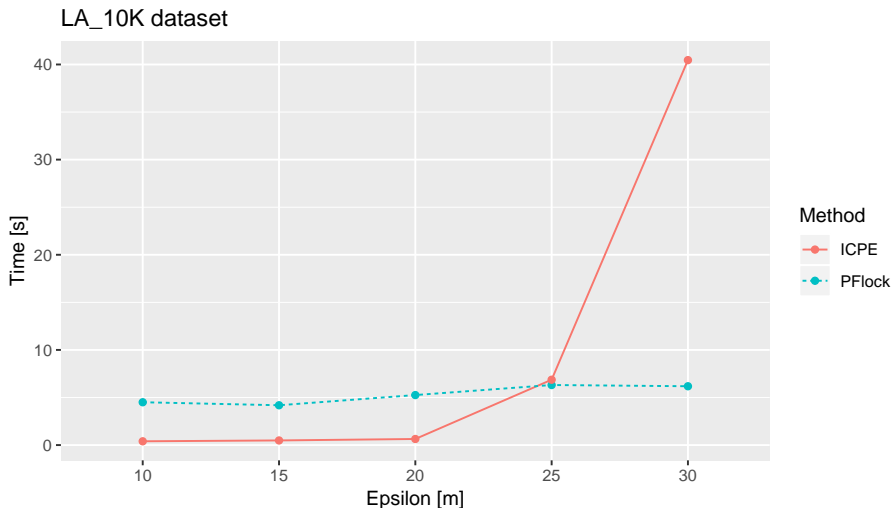
Andres Calderon

University of California, Riverside

September 20, 2019

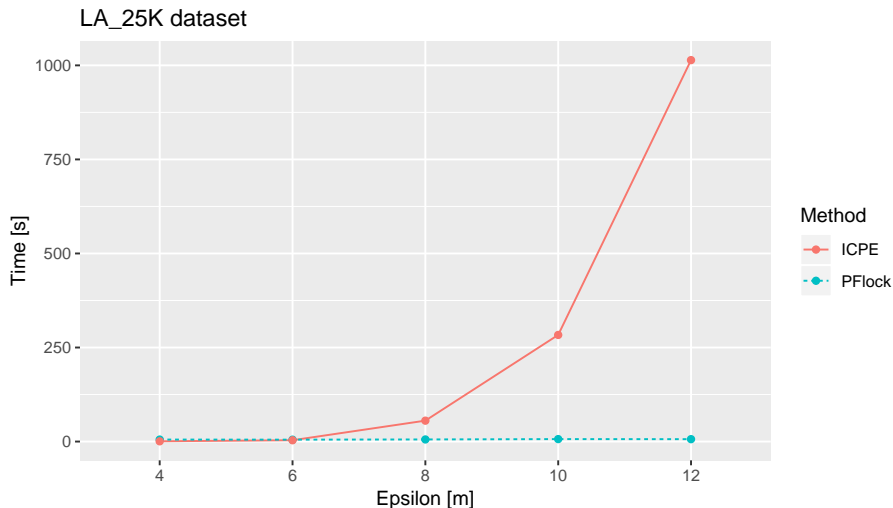
LA_10K Maximal disks finding (ICPE vs PFlock)

$\mu = 3$.



LA_25K Maximal disks finding (ICPE vs PFlock)

$\mu = 3$.



Some remarks (doubts)

- ▶ ICPE proposes **Range Join**: The main purpose is to remove noise so DBScan would deal with only core and reachable points. I perform this task through a self distance-join using GeoSpark. Is it a valid approach or I must follow the steps to build a Range Join?
- ▶ ICPE approach goes just until the cluster finding. Up to this point, the approach is very fast. The performance decreases after that, when it finds maximal disks following the BFE algorithm.

Some remarks (doubts)

- ▶ ICPE claims it does not need to distribute the DBScan algorithm, so I neither distribute the BFE algorithm. Both are sequential runnings. It should be interesting to use the DBScan algorithm as a partitioner.
- ▶ Datasets from Chen et al. has duplicates and long jumps between their time instaces. The Brinkhoff dataset goes from 22M points to 9M after removing duplicates.