

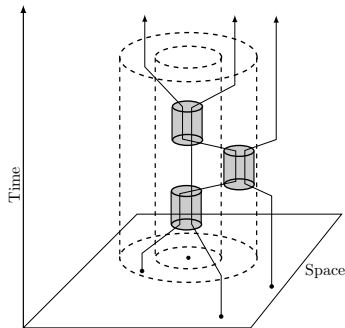
# SCALABLE PROCESSING OF MOVING FLOCK PATTERNS

Andres Calderon · [acald013@ucr.edu](mailto:acald013@ucr.edu)

University of California, Riverside

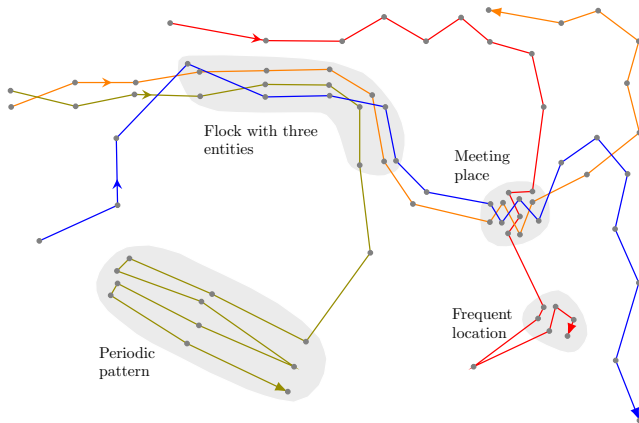
October 31, 2024

- A spatial trajectory is a trace in time generated by a moving entity in a geographical space.
- i.e.  $p_1 \rightarrow p_2 \rightarrow \dots \rightarrow p_n$
- A trajectory is stored as a time-ordered sequence of points,  $p_i = (x, y, t)$  (spatial coordinate + time stamp).



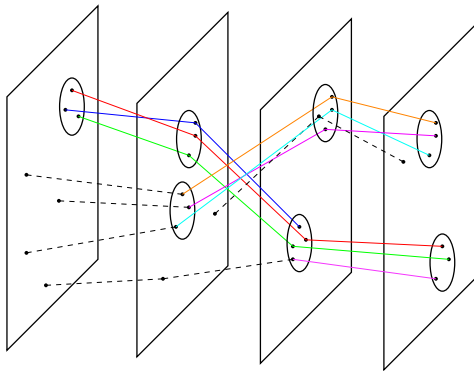
(Shoval, 2017)

# MOVEMENT PATTERNS



(Gudmundsson, et al. 2008)

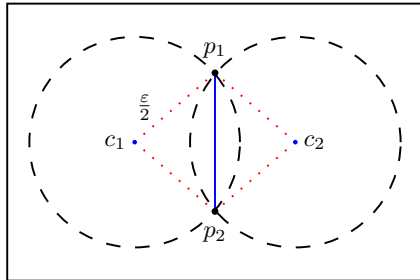
- i.e. convoys, moving clusters, swarms, gatherings, **flocks**, ...



- $\varepsilon$ : Diameter of the circle which contains all the objects.
- $\mu$ : Minimum number of objects.
- $\delta$ : Minimum time interval the objects travel 'together'.

# BASIC FLOCK EVALUATION ALGORITHM

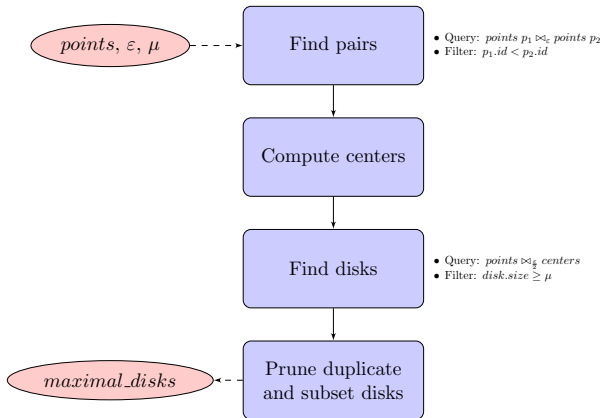
- Vieira, et al. 2009.
- The first polynomial-time solution for determining disk locations.
- Under fixed time duration it has polynomial time complexity  $O(\delta|\tau|^{(2\delta)+1})$



# BASIC FLOCK EVALUATION ALGORITHM

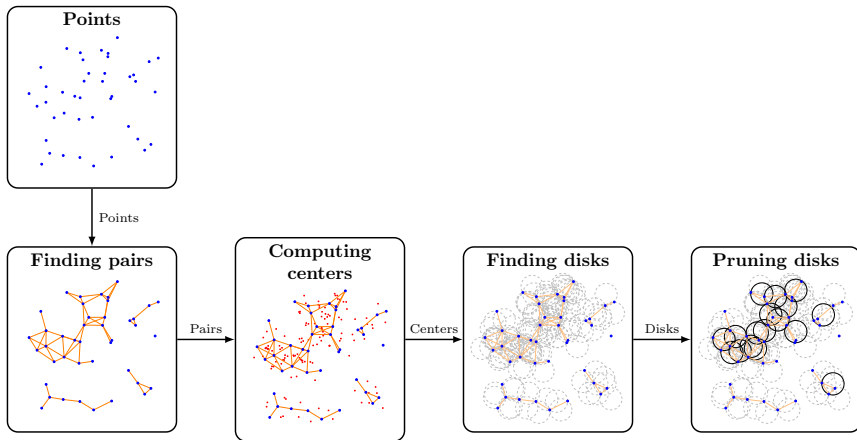
- Two main parts:
  - ▶ In the spatial domain it finds maximal disks at each time instant.
  - ▶ In the temporal domain it joins consecutive times to match set of maximal disks.

## ■ BFE overview...



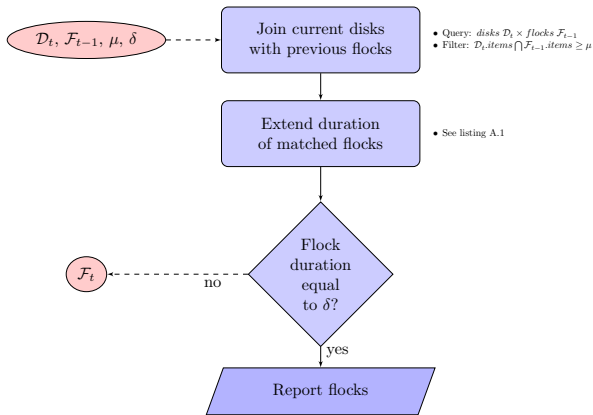
# ON THE SPATIAL DOMAIN

## ■ BFE overview...



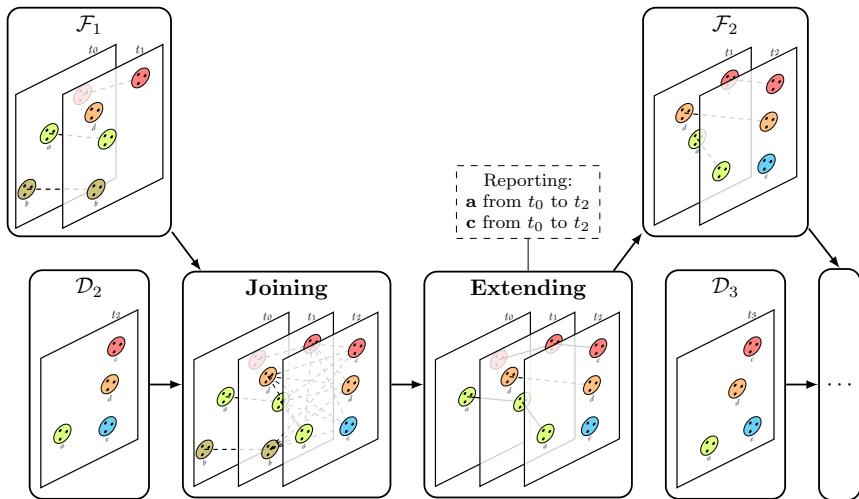


## ■ BFE overview...

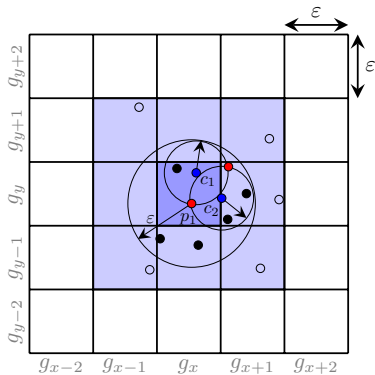


# ON THE TEMPORAL DOMAIN

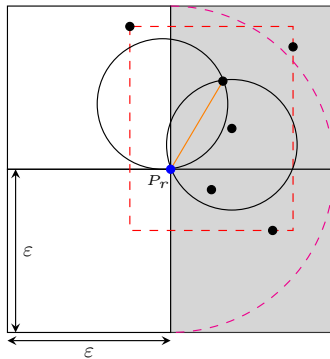
## ■ BFE overview...



# PSI ALGORITHM



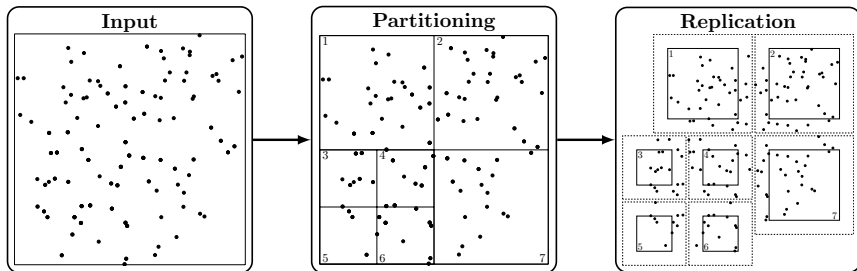
(Vieira, et al. 2009)



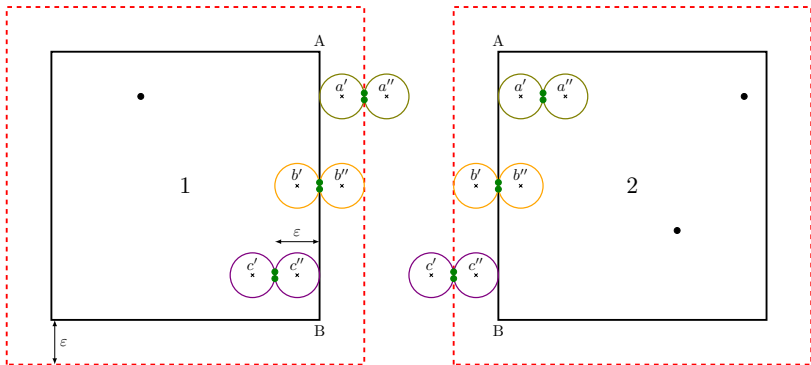
(Tanaka, et al. 2016)

- High complexity limits scalability.
- Large datasets with dense clusters of moving entities per time instant significantly impact performance.
- Specifically,
  - ▶ identifying maximal disks is hindered by the extensive number of candidates requiring pruning.
  - ▶ when parallelizing, we must address moving flocks that traverse contiguous partitions.
- We propose a parallel and scalable solution for both spatial and temporal domains.

## ■ Partitioning strategy...

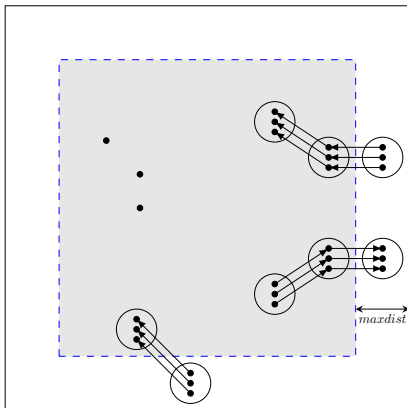


## ■ Handling duplication...

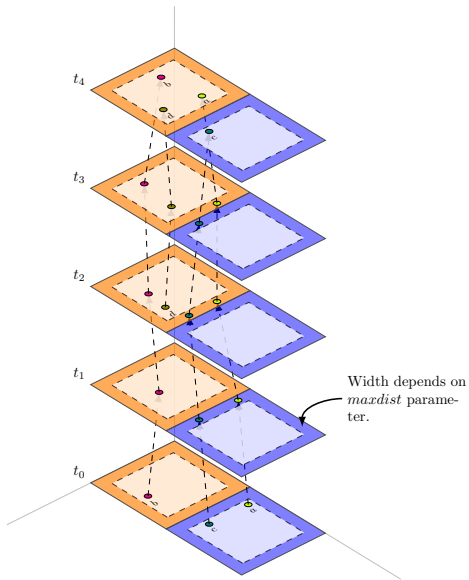


# ON THE TEMPORAL DOMAIN

- We introduce the *maxdist* parameter to define an area where we have to track **crossing partial flocks** (CPFs)...



# ON THE TEMPORAL DOMAIN

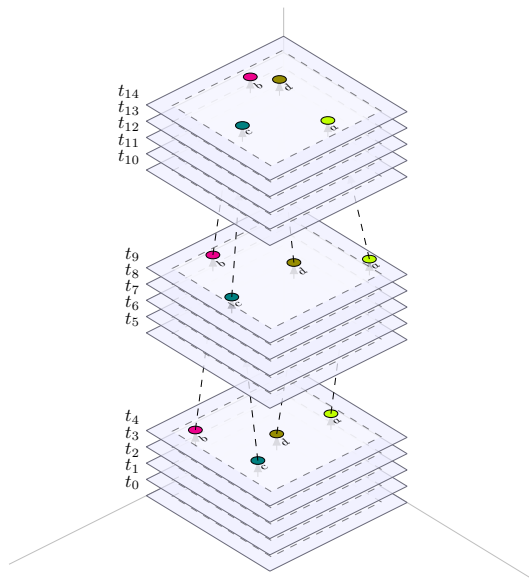


\* $a, b, c$  and  $d$  are flocks moving along time.



- Discovered flocks inside the safe area are ready to be reported.
- CPFs require post-processing. We propose four alternative:
  - ▶ Master
  - ▶ By-Level
  - ▶ Least Common Ancestor (LCA)
  - ▶ Cube-based

# ON THE TEMPORAL DOMAIN

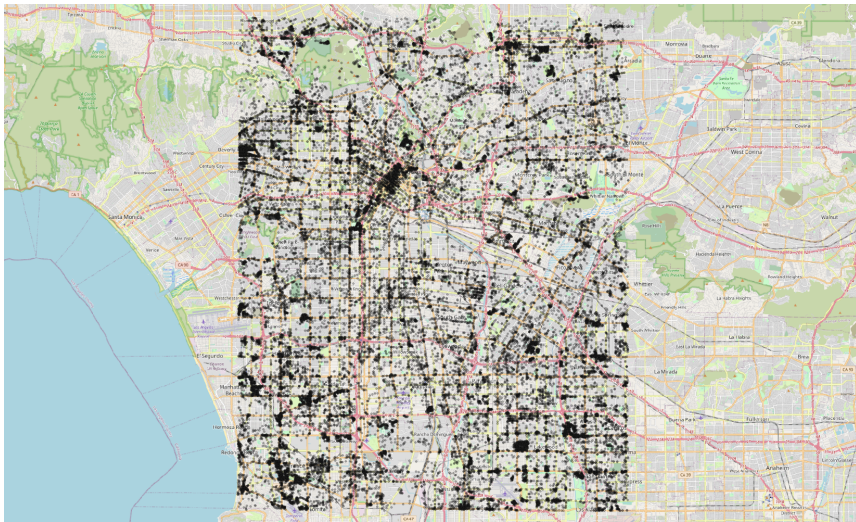


\*a,b,c and d are flocks moving along time.

# DATASETS

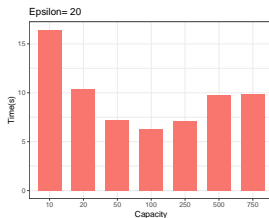
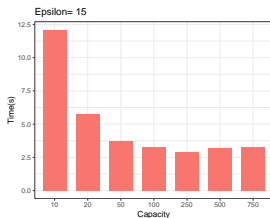
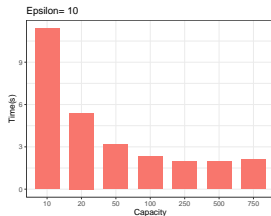
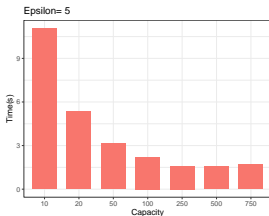
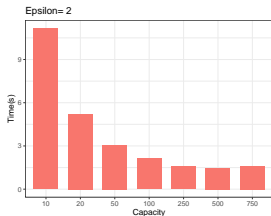
Dataset	Number of Trajectories	Total number of points	Maximum Duration (min)
Berlin10K	10000	97526	10
LA25K	25000	1495637	30
LA50K	50000	2993517	60

## ■ Synthetic datasets [LA50K]



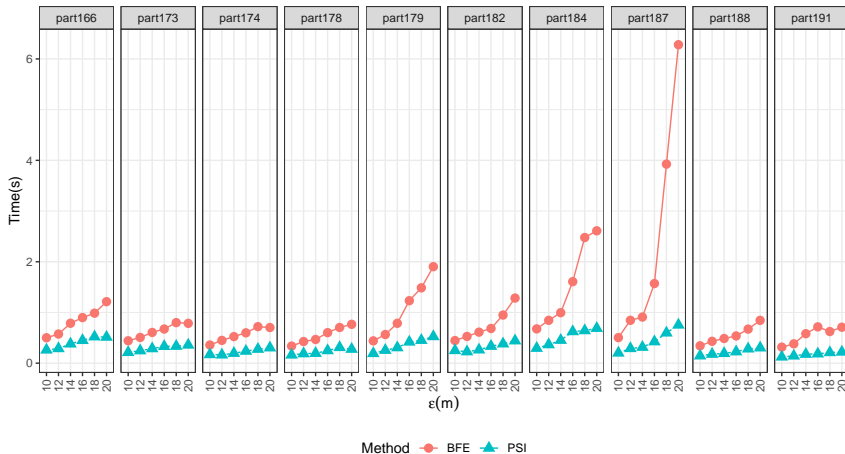
# EXPERIMENTS

## ■ Optimizing the number of partitions for Phase 1.



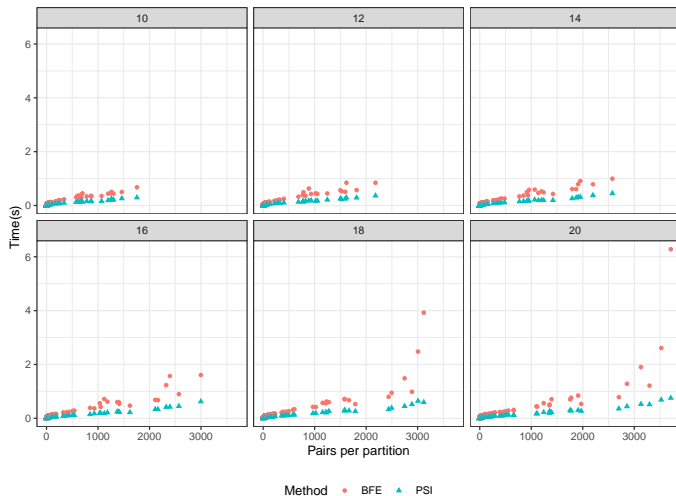
# EXPERIMENTS

- Analyzing most costly partitions.
  - Top 10 most costly partitions.



# EXPERIMENTS

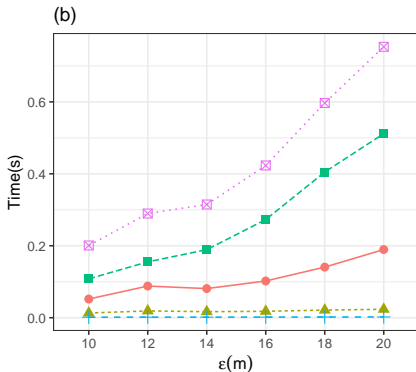
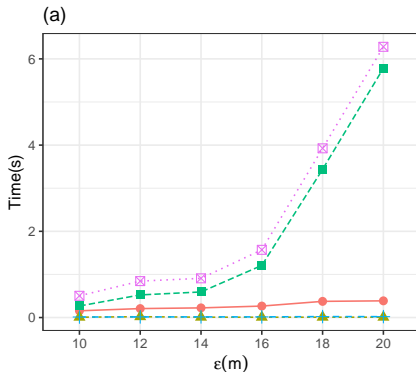
- Analyzing most costly partitions.
  - By Pairs density..



# EXPERIMENTS

## ■ Analyzing most costly partitions.

► By Stages in the most costly partition [(a) BFE (b) PSI].



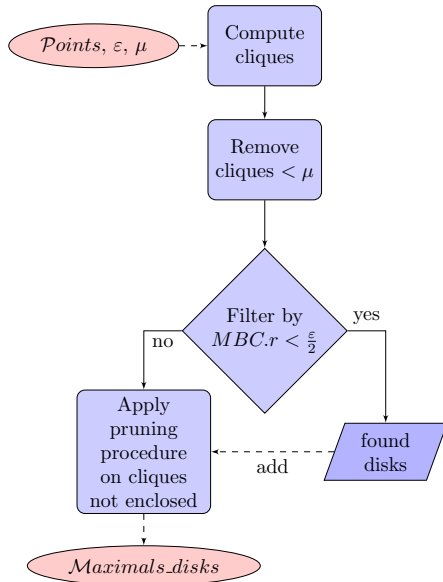
Stage —●— Candidates —▲— Centers —■— Maximals —+— Pairs —×— Total



# CAN WE REDUCE PRUNING TIME?

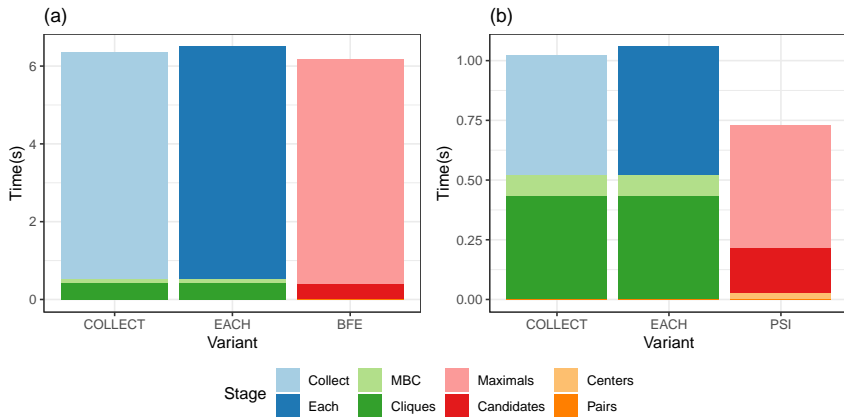
- **Maximal clique (MC):** Given an undirected graph, a MC is a subset of vertices, each directly connected to every other in the subset, that cannot be expanded by adding additional vertices.
- **Minimum Bounding Circle (MBC):** Given a set of points in Euclidean space, the MBC is the smallest circle that can enclose all the points.

# CAN WE REDUCE PRUNING TIME?



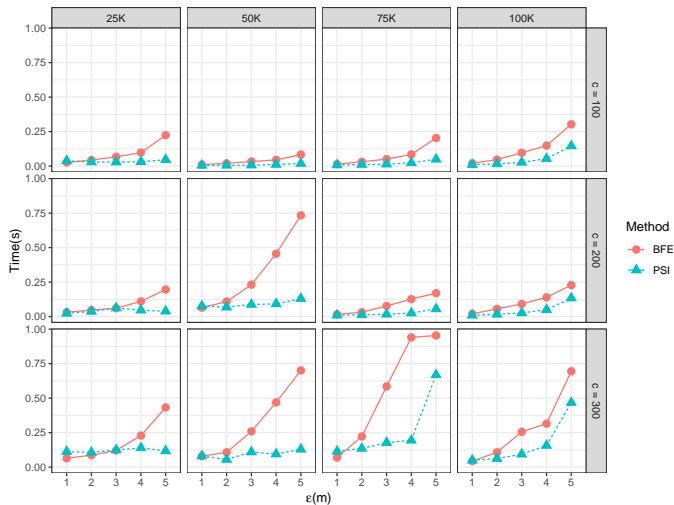
# CAN WE REDUCE PRUNING TIME?

- Phase 1 variants performance [(a) vs BFE (b) vs PSI].



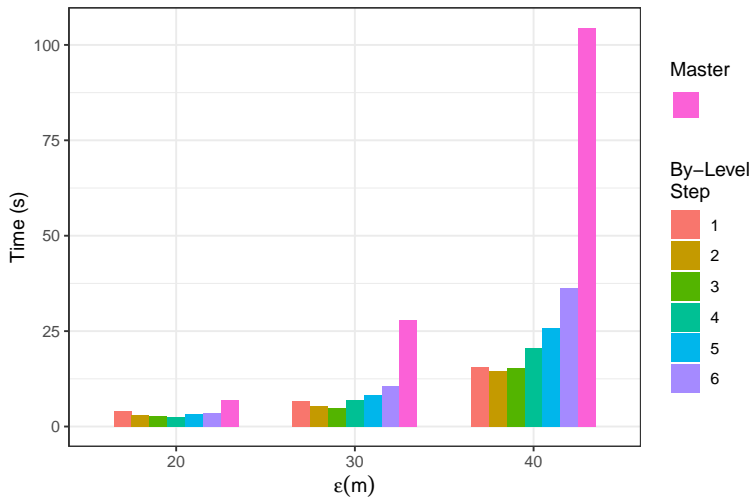
# EXPERIMENTS

## ■ Relative performance of Phase 1 using synthetic datasets.



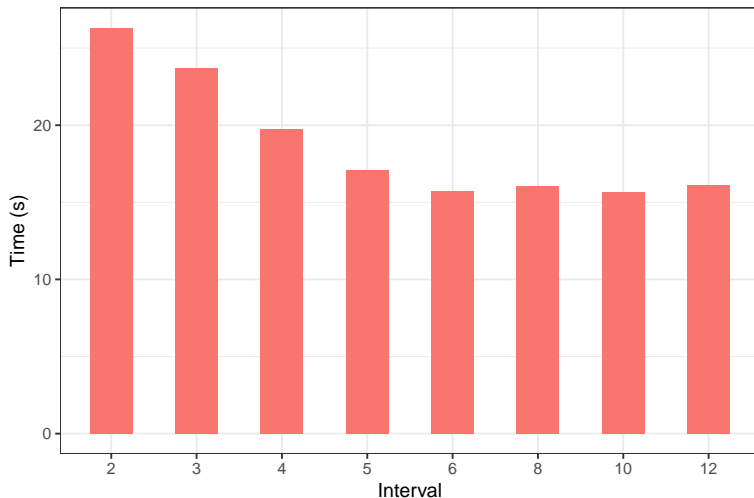
# EXPERIMENTS

- Finding best *step* value for By-Level alternative.



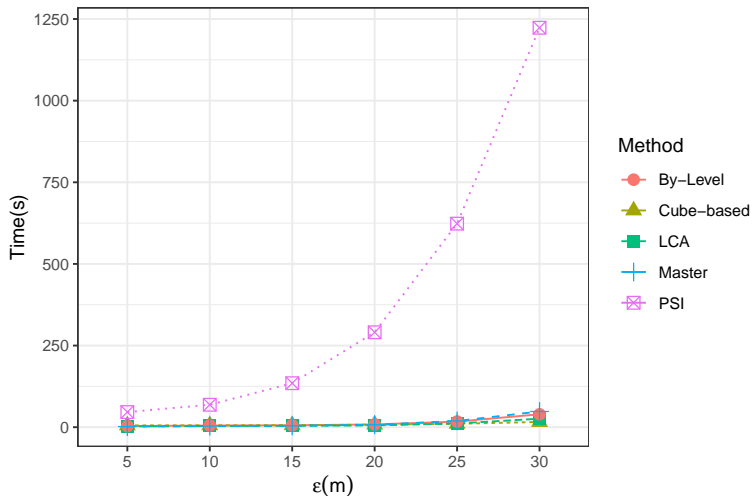
# EXPERIMENTS

- Finding best *interval* value for Cube-based alternative.



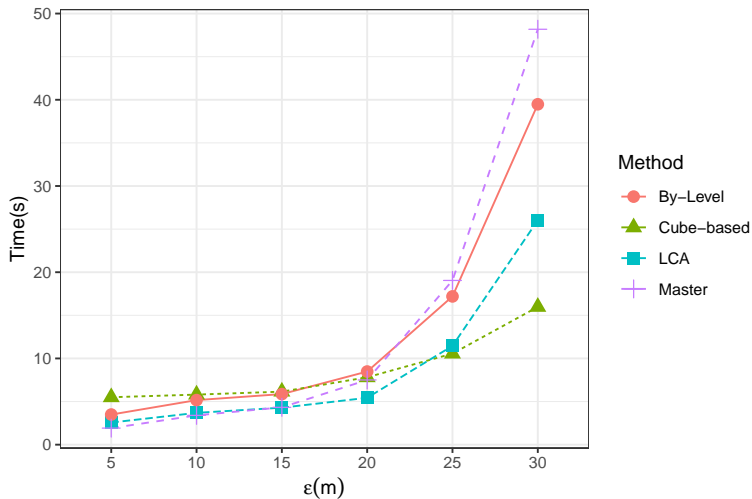
# EXPERIMENTS

## ■ Scalable alternatives vs standard PSI.



# EXPERIMENTS

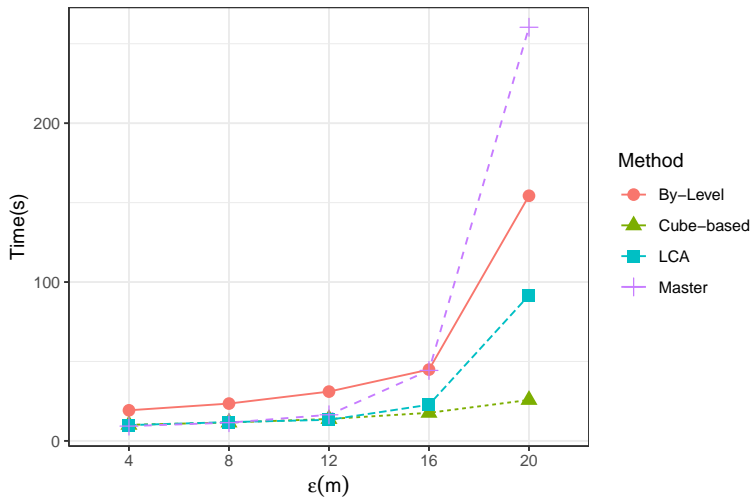
## ■ Scalable alternatives in LA25K dataset.





# EXPERIMENTS

## ■ Scalable alternatives in LA50K dataset.



Thank you!