Attraction and Avoidance Detection from Movements

Zhenhui Li, Bolin Ding, Fei Wu, Tobias Kin Hou Lei, Roland Kays, and Margaret Crofoot. VLDB’14

Movement data contain valuable information

Movement patterns from human/animal reveal behaviors.

Understanding movement in road network helps us do better routing.

Advancement in positioning technology enables easy collection of large-scale movement data.

GPS sensor

Time	Longitude	latitude
10:00:00	38.3	40.2
10:05:00	38.5	40.1
10:10:00	38.9	40.0
…	…	…
21:00:00	40.2	38.2

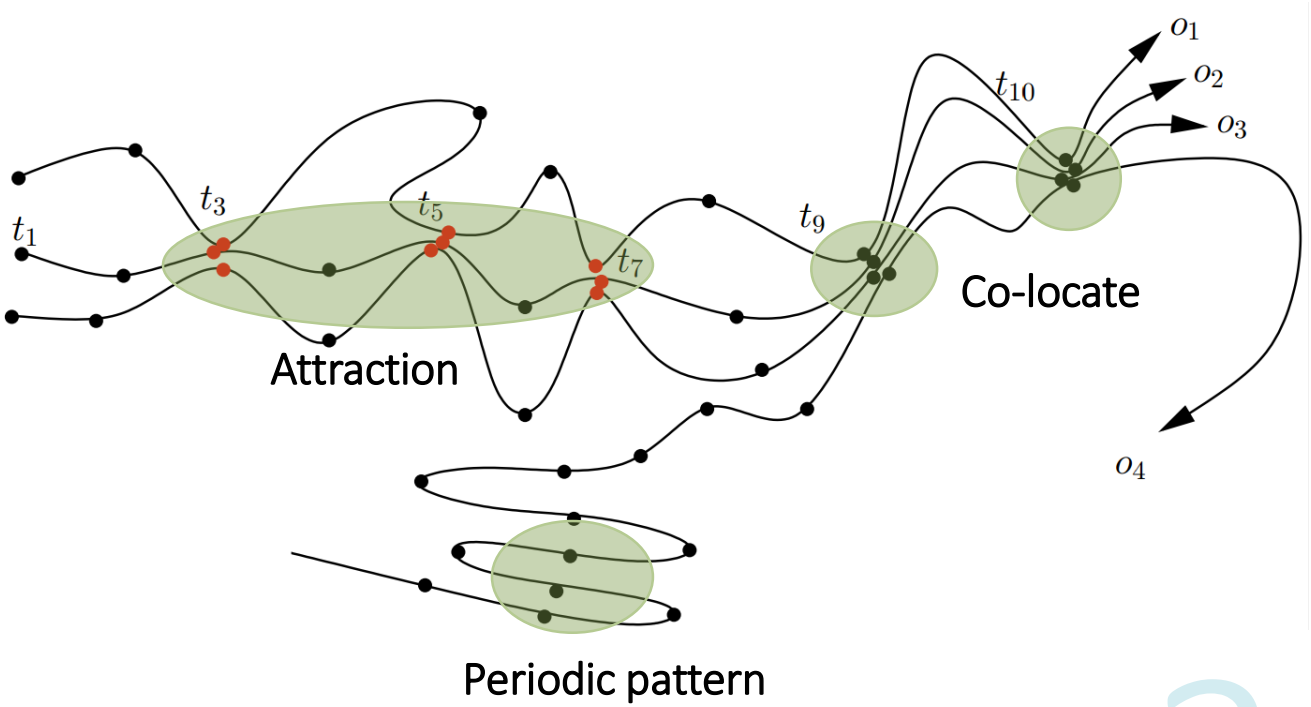
Patterns from movement data

Different kinds of patterns exist in movement data.

- Spatial pattern: co-locate
- Temporal pattern: periodic
- Spatiotemporal pattern: attraction

Attraction and avoidance relationship

- Attraction
  - *Human groups (e.g., colleague)*
  - *Bird migration*
- Avoidance
  - *Criminals avoids polices*
  - *Predators and prays*



Previous work on attraction relationship

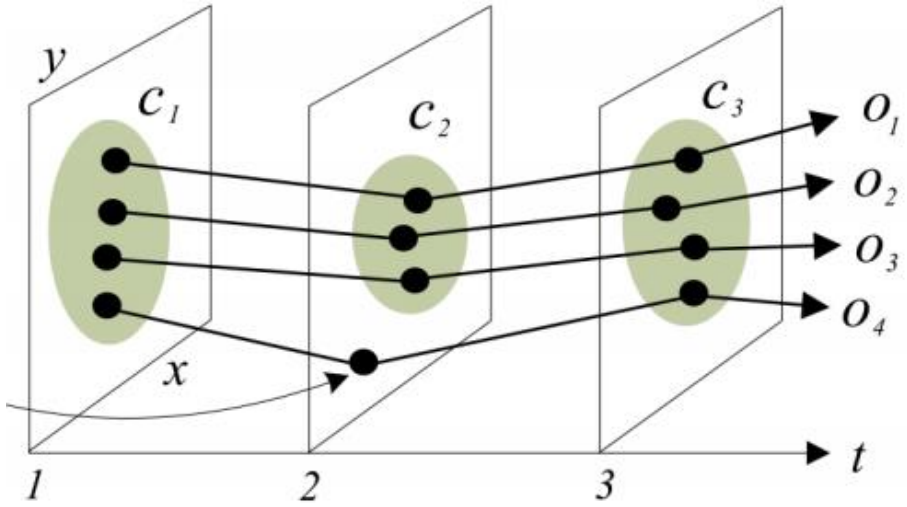
Previous work only study attraction

Moving objects clustering methods use the *meeting frequency* to indicate the strength of attraction.

- *Meeting frequency* is the number of timestamps objects being together at the same time.
- Moving clusters[Kalnis,SSTD'05], flock[Gudmundsson,GIS'06], convoy[Jeung,VLDB'08], swarm[Li,VLDB'10]

Assumption:

The higher the *meeting frequency* is,
the stronger the attraction relationship is.



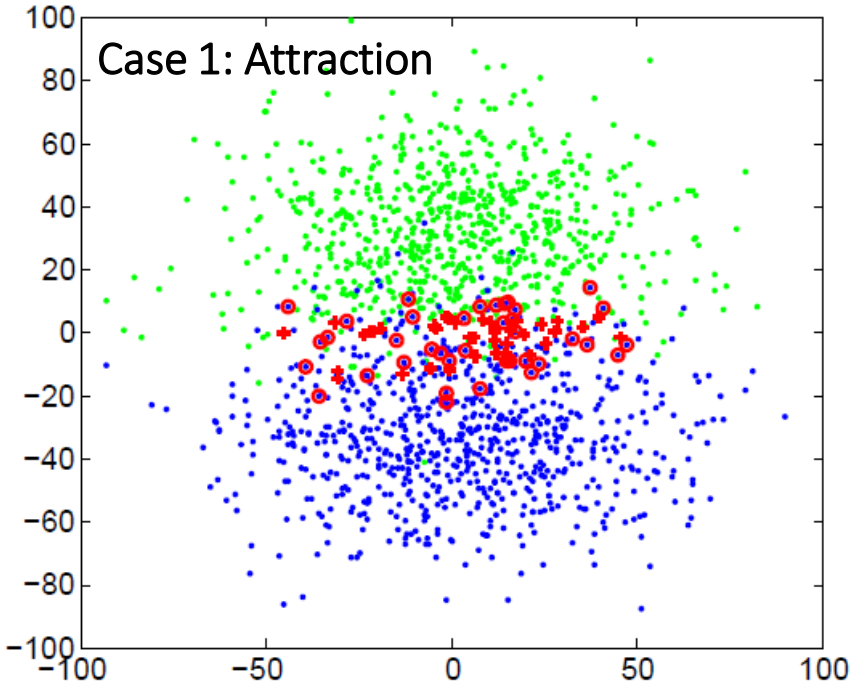
Assumption of previous work: Meeting frequency = Attraction

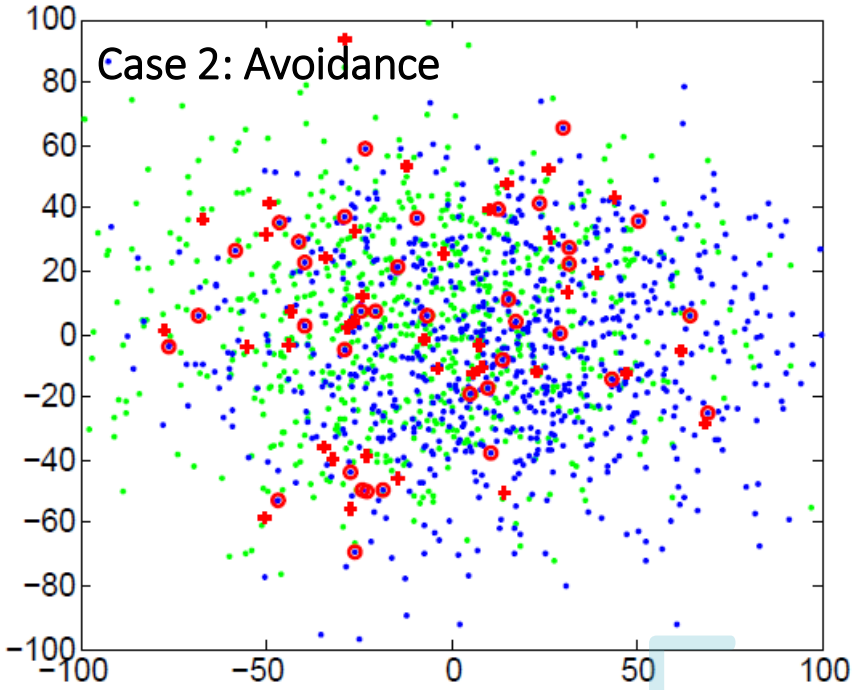
Locations of two pairs of moving objects.

Meeting places are marked in red.

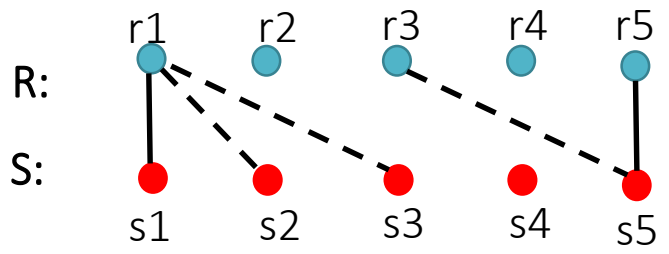
Objects meet 40 times in case 1.

Objects also meet 40 times in case 2.





Compute expected meeting frequency



The probability of r1 to meet with any point in S is $\frac{3}{5}$.

The probability of r3 to meet with any point in s is $\frac{1}{5}$.

$$freq_{expect}(R,S) = \frac{\#\ of\ all\ edges}{n}$$

Attraction: $freq_{actual} > freq_{expected}$

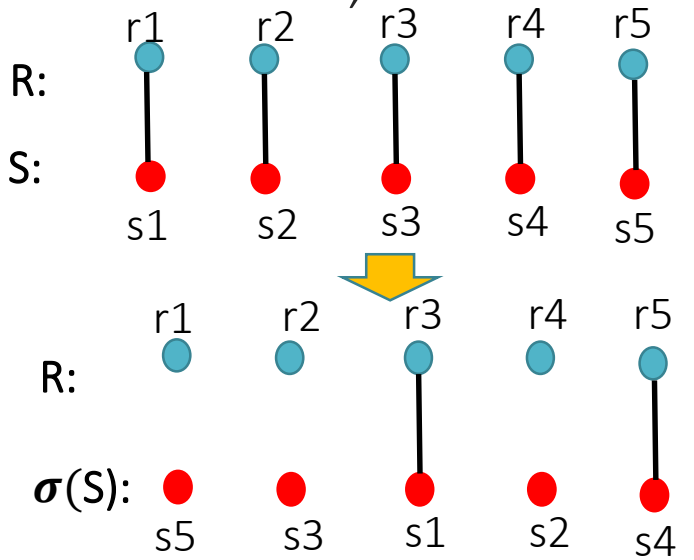
Avoidance: $freq_{actual} < freq_{expected}$

Expected meeting frequency does not tell the ***degree!***

People are more interested in significant relationship.

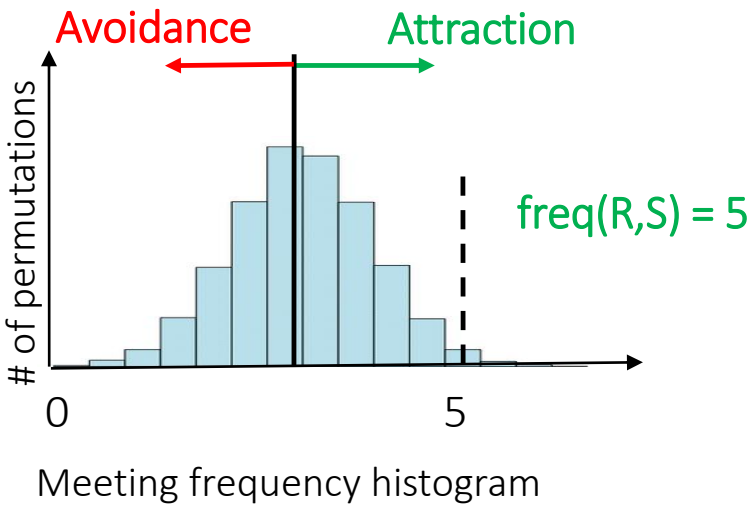
Compute the degree of attraction/avoidance through permutation test

Given R and S, each time we randomly permute the order of S.



$freq(R,S) = 5$

$freq(R,\sigma(S)) = 2$



$$Sig_{attract} = \Pr[freq(R,S) > freq(R,\sigma(S))]$$

$$Sig_{avoid} = \Pr[freq(R,S) < freq(R,\sigma(S))]$$

Using Monte Carlo sampling to approximate significant value

The total number of permutations is factorial.

Monte Carlo sampling: Sample N permutations.

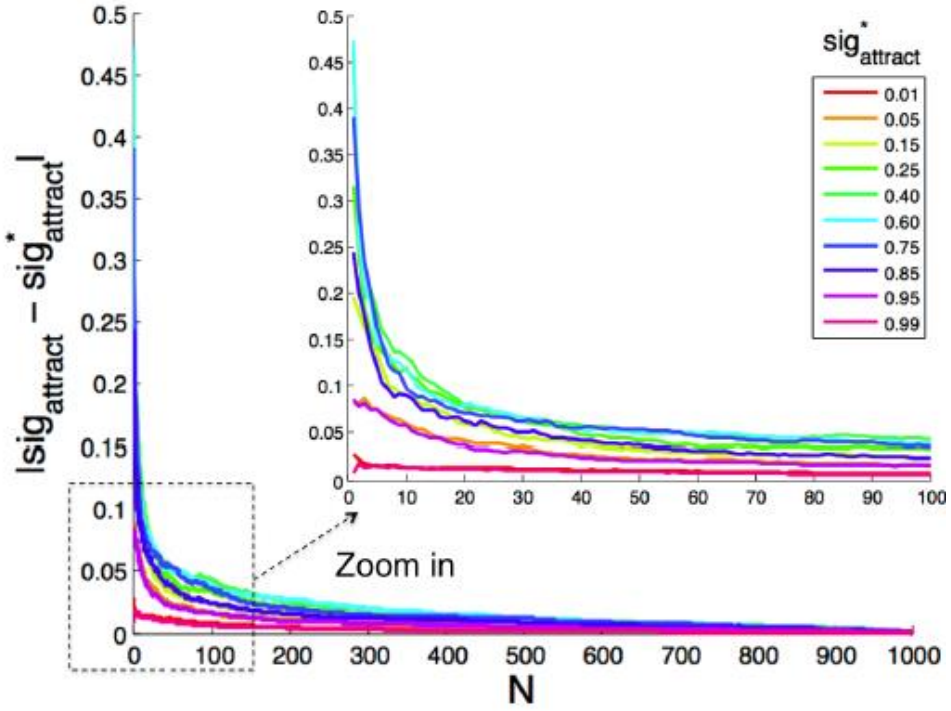
When $N \geq \frac{4}{\epsilon^2 \rho} \ln(\frac{2}{\delta})$, we guarantee to have $(1-\epsilon)\rho \leq \hat{\rho} \leq (1+\epsilon)\rho$ with probability $1-\delta$.

ρ is the exact significant value

ϵ is error

δ is a probability

The significant value converges quickly.



Pruning and threshold query

Pruning rule 1: Eliminate non-overlapping points in two trajectories.

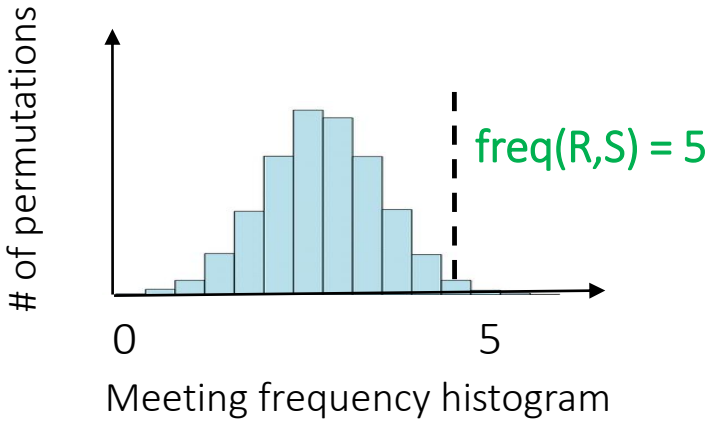
Eliminate pairs that have zeros probability to meet (their territories do not overlap).

Pruning rule 2: Early termination.

We only need to know whether permutated meeting frequency is larger/smaller than the actual meeting frequency.

Threshold query: In practice, people are often more interesting in significant attraction/avoidance with regard to some object.

I want to know which animals are significantly attracted ($Sig_{attract} > 0.95$) by this lion.

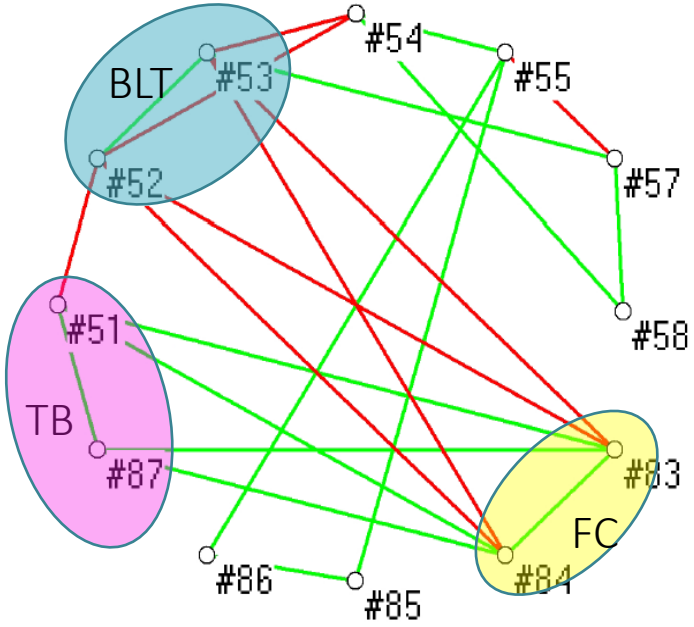


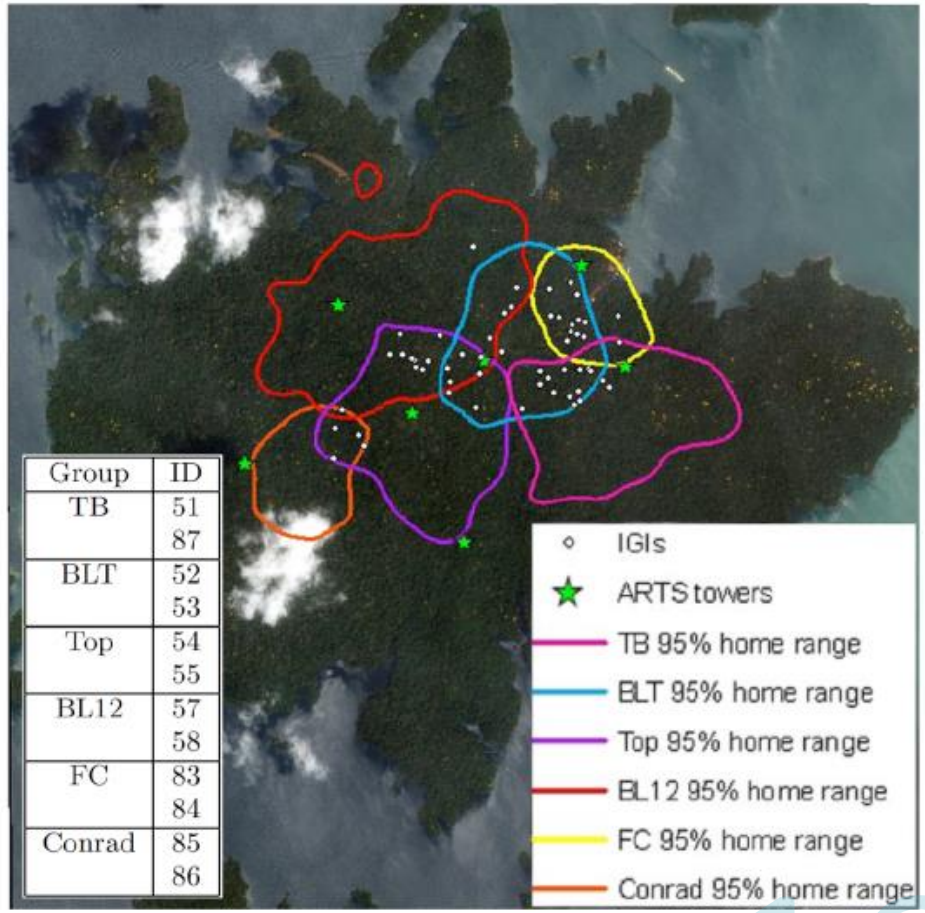
Experiment on the Monkey dataset

A network is constructed by using our measure

Green link: having $Sig_{attraction}$ value > 0.95

Red link: having $Sig_{avoidance}$ value > 0.95



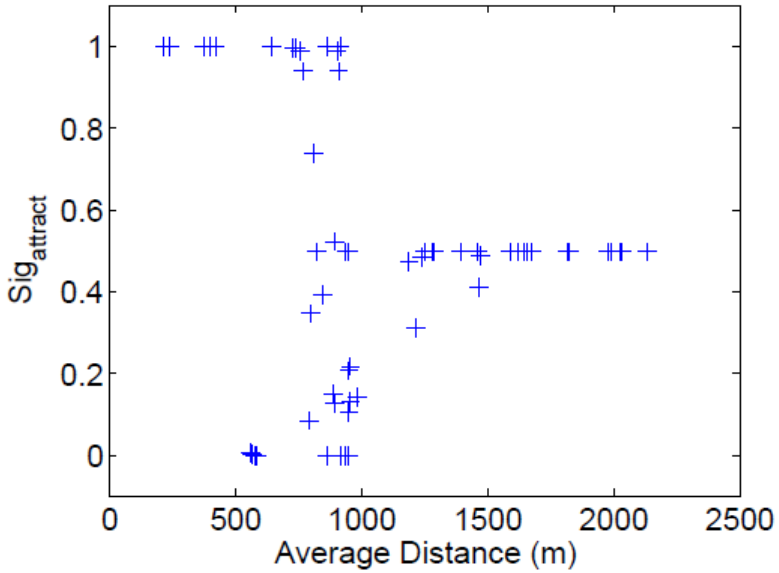


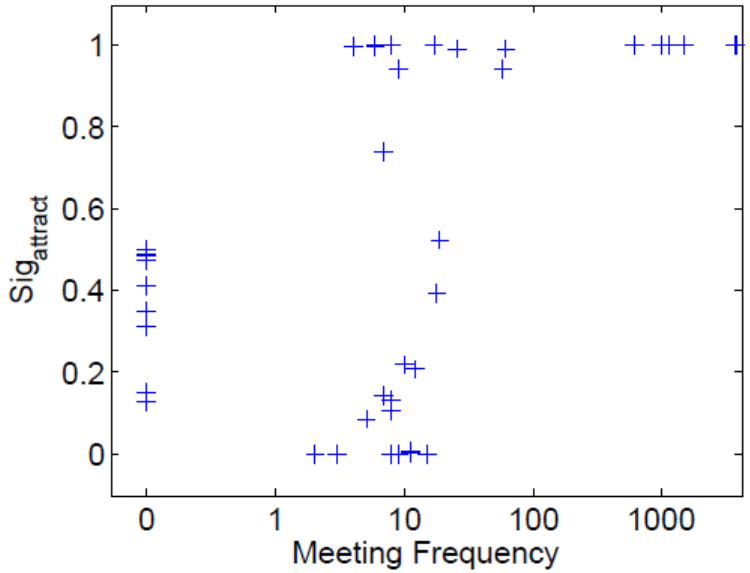
Group	ID
TB	51 87
BLT	52 53
Top	54 55
BL12	57 58
FC	83 84
Conrad	85 86

13 fights reported between group BLT and FC.

Experiment on the Monkey dataset

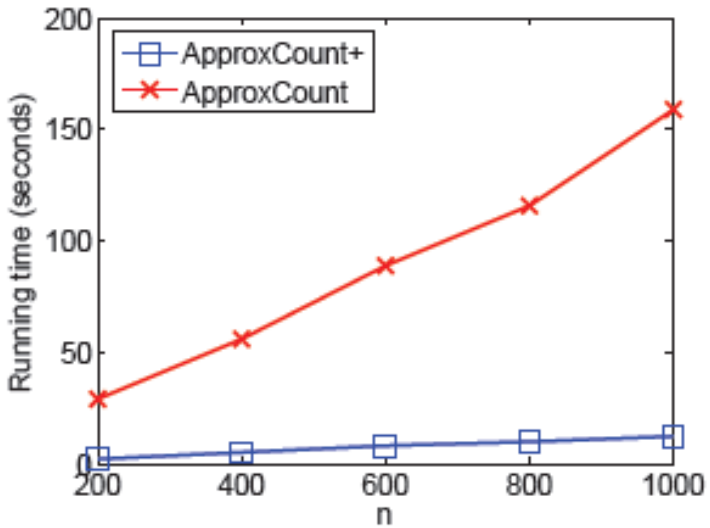
Compare with average distance and meeting frequency



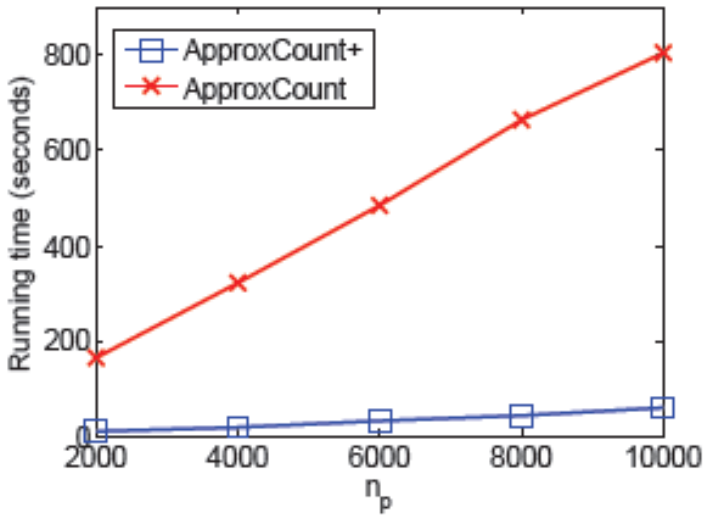


Experiment on efficiency of pruning

Efficiency study on a synthetic dataset



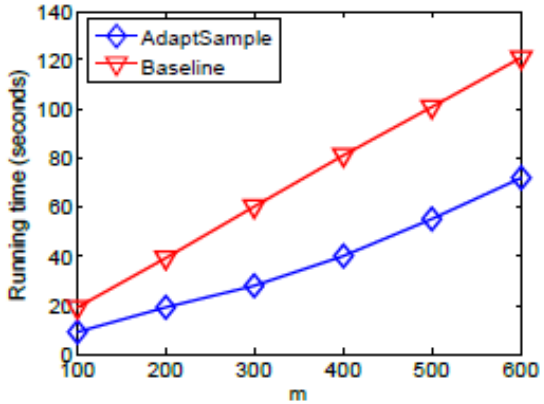
(a) Running time w.r.t. trajectory length n



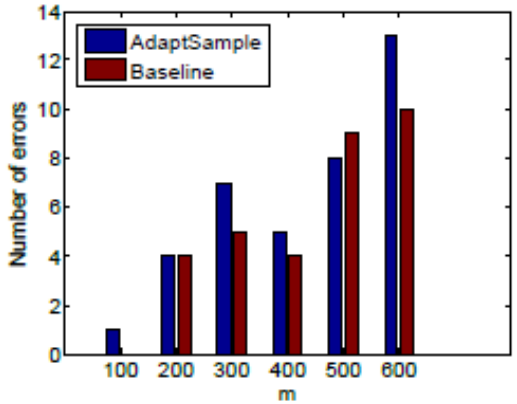
(b) Running time w.r.t. number of object pairs n_p

Experiment on efficiency of query processing

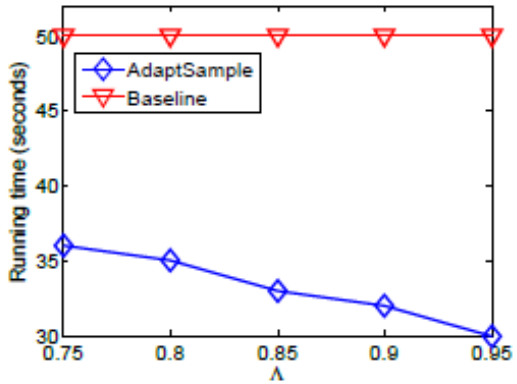
Efficiency study on threshold query on the Monkey dataset.



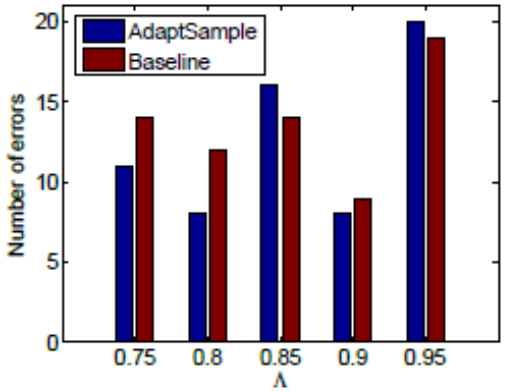
Running time w.r.t. number of objects m



Number of errors w.r.t. number of objects m



Running time w.r.t. threshold Λ



Number of errors w.r.t. query threshold Λ

Summary

Contributions:

- Our work is the first one finding the **avoidance** relationship from movements.
- Our measure quantifies the **degree** of relationship through permutation test.
- We develop approximation algorithm to make the computation **feasible** and give a theoretic **bound**.
- We further develop **pruning rules** to speed up the computation and approximation algorithm for fast **threshold query processing**.

Future work:

- Random permutation might generate impossible trajectory.

Experiment on reality mining dataset

	Group	Affiliation	Friend
Significance Value	0.6221	0.6538	0.7001
Meeting Frequency	0.3104	0.3371	0.5053
Dynamic Time Warping	0.1967	0.2627	0.5576