

## Algorithms for Spatial and Spatio-Temporal Data Mining

### Project

NWO-BSIK/BRICKS programme, FOCUS project no. 642.065.503 (GADGET).

### Objective of GADGET

The Objective of GADGET (Geometric Algorithm Design for Geographic EnvironmenTs) is to develop efficient geometric algorithms for advanced geographical analysis techniques and geographic data mining on spatio-temporal data. The geographic context should be apparent in the problems that are solved algorithmically.

### Members

Dr. Marc van Kreveld (coordinator).  
 Dr. Jun Luo (researcher).  
 Dr. Magdalene G. Borgelt (researcher).  
 Department of Information & Computer Science,  
 Utrecht University.  
 marc, ljroger, magdalene@cs.uu.nl

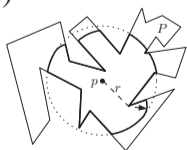
### Geodesic distance clusters

In clustering, geodesic distance rather than classical euclidean distance captures distance better when geographic obstacles are present. For example, for tortoises (they dwell on land), natural obstacles exist like rivers that should be taken into consideration.

### Problem statement

- Input:** A set  $S$  of  $m$  points in the plane.  
 A distance  $r > 0$ .  
 An  $n$  vertex simple polygon  $P$  containing all points of  $S$ .
- Output:** All  $m$  boundaries of geodesic discs centered at the points of  $S$ , with radius  $r$ .

Each geodesic disc centered at a point  $p \in S$  is the set of points inside  $P$  that have a geodesic distance of at most  $r$  from  $p$  (the geodesic distance is the length of the shortest polygonal path connection that lies inside  $P$ .)



The arrangement induced by the boundaries of the geodesic discs partitions the polygon into regions where fixed subsets of  $S$  are closest. This aids in clustering.

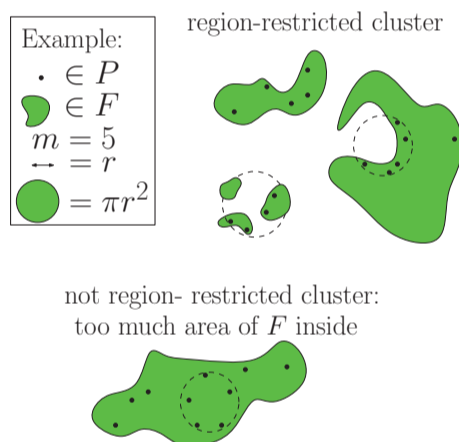
### Result

An output sensitive algorithm that runs in  $O((n + (kn)^{\frac{2}{3}} + k) \log^c n)$  time, for some constant  $c$  and output size  $k$ , ( $k = \Omega(m)$ ) is given.

[M. G. Borgelt, M. van Kreveld, and J. Luo: Computing Geodesic Circles in a Simple Polygon. To appear in Proc. 23rd European Workshop on Computational Geometry, March 2007]

### Region-restricted clusters

The concept of region-restricted clustering captures the idea that when detecting clusters of points in a geographic space, there may be regions where the points can occur and regions where they cannot occur. For example, sea birds can have nests on islands, but not in the sea in between.



### Problem statement

- Input:** A set  $P$  of  $n$  points in the plane.  
 A distance  $r > 0$ .  
 A minimum subset size  $m$ .  
 A set  $F$  of simple polygons (inclusion regions) containing all points of  $P$ , with  $f$  edges in total.
- Output:** All subsets  $P' \subset P$  of size at least  $m$  contained in a circle of radius  $r' \leq 2r$ , where the circle may not contain more total area of  $F$  than  $\pi \cdot r^2$ .

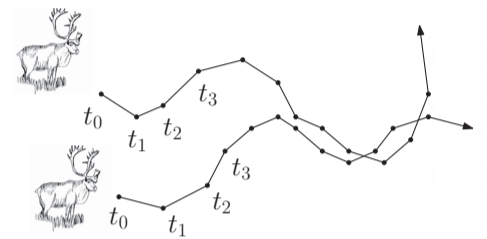
### Result

Region-restricted clusters can be found in  $O(nm \log^2 n + (nm + nf) \log^2 f)$  time, and radius  $(1 + \epsilon)$  approximations in  $O(n \log n + n/\epsilon^2 + f \log^2 f + (n \log^2 f)/(\epsilon^2))$  time.

[J. Gudmundsson, M. van Kreveld, and G. Narasimhan: Region-restricted clustering for geographic data mining, In proc. of the 14th European Symp. on Algorithms, September 2006]

### Flocks and meetings

Two pattern types studied in spatio-temporal data mining are flocking and meeting patterns in trajectory data. Trajectory data may come from animals with GPS collars, for example caribou, mountain goats, leopards, and so on. To examine their social behavior, trajectories can be analyzed to find out if subgroups meet, or travel together for a duration of time.



- Flock:** A subset of at least  $m$  entities is in each others proximity (circle of radius  $r$  with continuously moving location) during a time interval.
- Meet:** A subset of at least  $m$  entities is in each others proximity (circle of radius  $r$  with fixed location) during a time interval.

### Problem statement

Compute the subset and location of the longest duration flock and meet.

### Result

The algorithmic idea is to view trajectories in 3D space-time and fit skew (flock) or straight (meet) cylinders around enough trajectories.

Results for  $n$  trajectories with  $\tau$  time steps each:

- Flock:**  
 Fixed subsets of  $m$  entities: NP-hard.  
 Varying subset of  $m$  entities:  $O(n^3 \tau \log n)$ .
- Meet:**  
 Fixed/Varying subsets of  $m$  entities:  
 $O(n^4 \tau^2 \log n + n^2 \tau^3)$ .

[J. Gudmundsson, M. van Kreveld: Computing longest duration flocks in trajectory data, In Proc. of the 14th ACM Symposium on Advances in GIS, November 2006]