Visualization and Analysis of Geographic Information: Representation Algorithms and Data Structures

João Valença
University of Coimbra
Coimbra, Portugal
valenca@student.dei.uc.pt

Luís Paquete
University of Coimbra
Coimbra, Portugal
paquete@dei.uc.pt

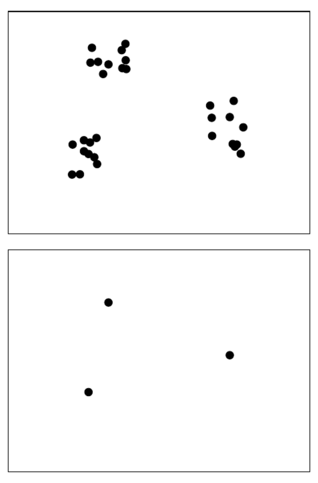
Carlos Caçador
Smartgeo Solutions
Lisboa, Portugal
carlos.cacador@smartgeo.pt

Pedro Reino
Smartgeo Solutions
Lisboa, Portugal
pedro.reino@smartgeo.pt

Introduction

The goal
Given a starting set of Geographic Points of interest, find a representative set of points for the region:

- Reduce visual information
- All points must be represented.
- Keep properties of the original set.
- Find the representative set in real-time.
- Allow for a panning movement over a larger region.



Minimum Coverage Subset

Given a set *N* of *n* points and a fixed number *k*; find a centroid subset *P* of size *k*, such that the element not in *P* farthest from its closest centroid is minimized.

$$\min_{\substack{P\subseteq N\\ |P|=k}} \max_{n\in N} \min_{p\epsilon P} \|p-n\|$$

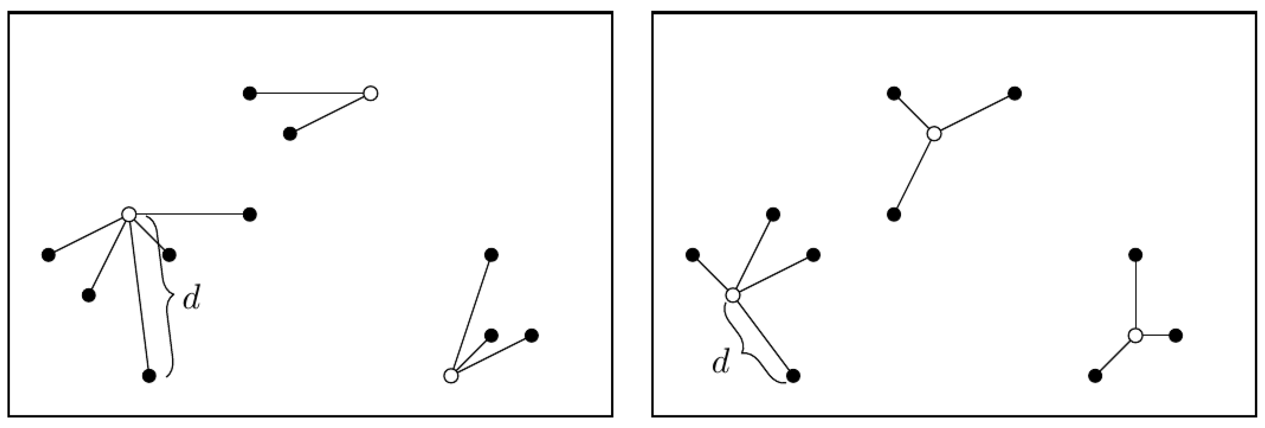


Figure 2: Suboptimal centroid selection (left) and optimal centroid selection (right)

This problem is known as Minimum Coverage Subset and is an example of a Facility Location *k*-center problem [1]. This can be solved through Integer Linear Programming [2] and Branch-and-Bound algorithms.

Naive Incremental Branch-and-Bound:
At each recursive step, decide whether a point is a centroid or not and branch accordingly. Update the objective function:

- a new centroid may change assignments and lowers the objective function
- a new non-centroid is assigned to closest centroid and raises the objective function

At each recursion step, calculate a lower bound for the best possible outcome for the solution of the current branch by assuming all remaining points as centroids.

Geometric Incremental Branch-and-Bound:
Use an incremental approach, but build and maintain a Delaunay Triangulation between the centroids.

Use a greedy routing algorithm [3] to quickly find the nearest centroid for new non-centroids. Use Hilbert Curves [4] to pre-sort the points and further accelerate the process by heuristically reducing the searching path for the routing algorithm.

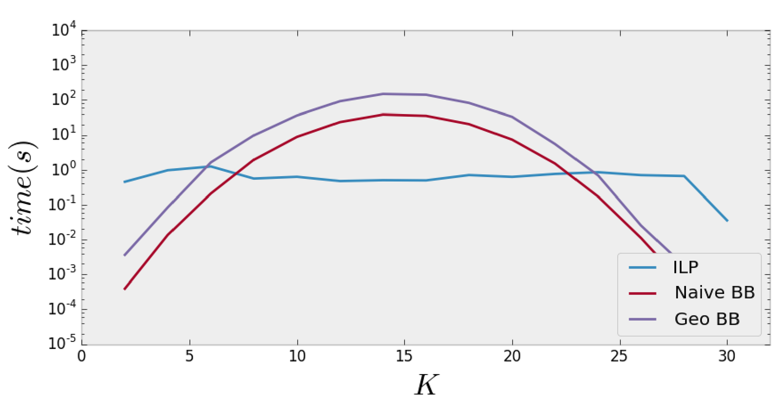


Figure 3: Effect of the value of *K*

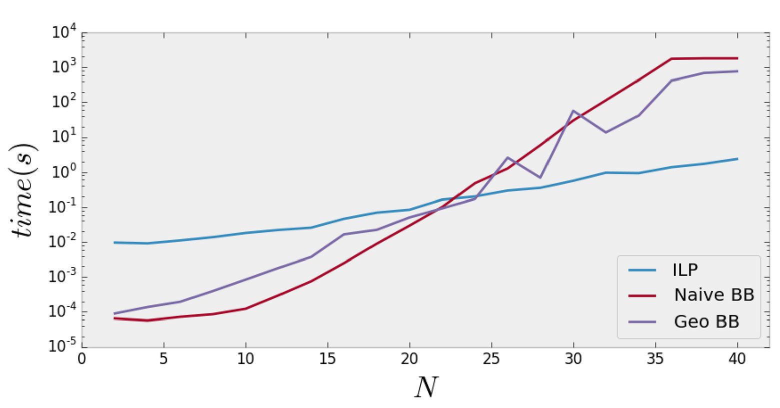


Figure 4: Effect of the value of *N*

Results: Both Branch and Bound implementations were too slow. The simple Integer Linear Programming performed better, but still unsuitable for real time use. It also requires previous information on the instance of the problem.

Fixed Minimum Distance

Given a set of *n* points and a minimum distance *d*. Find a subset of *k* points such that no distance between two points is smaller than *d* whilst minimizing *k*.

Optimal methods take too long to be used in real time so we use an approximation algorithm instead.

Using a k-d tree range search, we create a graph with edges connecting all pairs of points whose distance is smaller than *d*. This takes $O(n^2)$ time to perform [5].

Until the graph is empty, chose the point with the largest number of neighbors to be selected and remove it and all its neighbors from the graph. Removing all points can be cast as a set cover problem, and this approach can be performed in O(n log n) time [6].

This algorithm ensures that the selected set will have no more than *m* ln *n* elements, with *m* being the optimal number for total coverage.

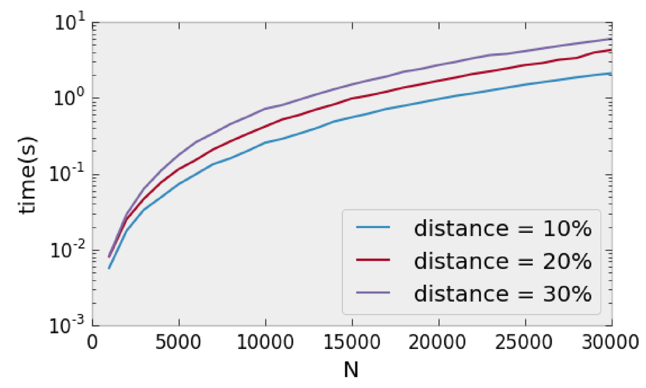


Figure 5: Effect of the value of *N* on cpu time

Results: This approach is fast enough to be used in real time. And depends on the established minimum distance as well as the density.

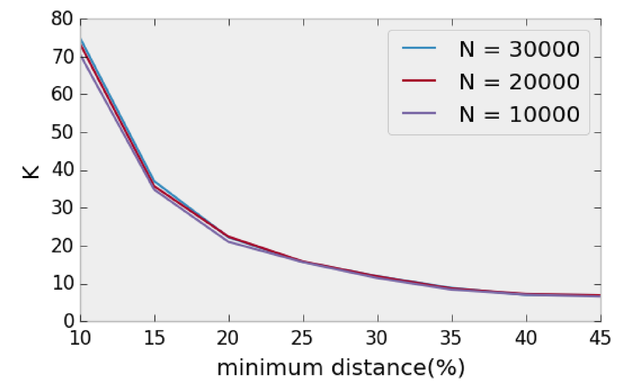


Figure 6: Effect of the point density on the value of *K*

Results: The points chosen do not vary with the density. The algorithm can be sped up with a faster/preprocessed range search

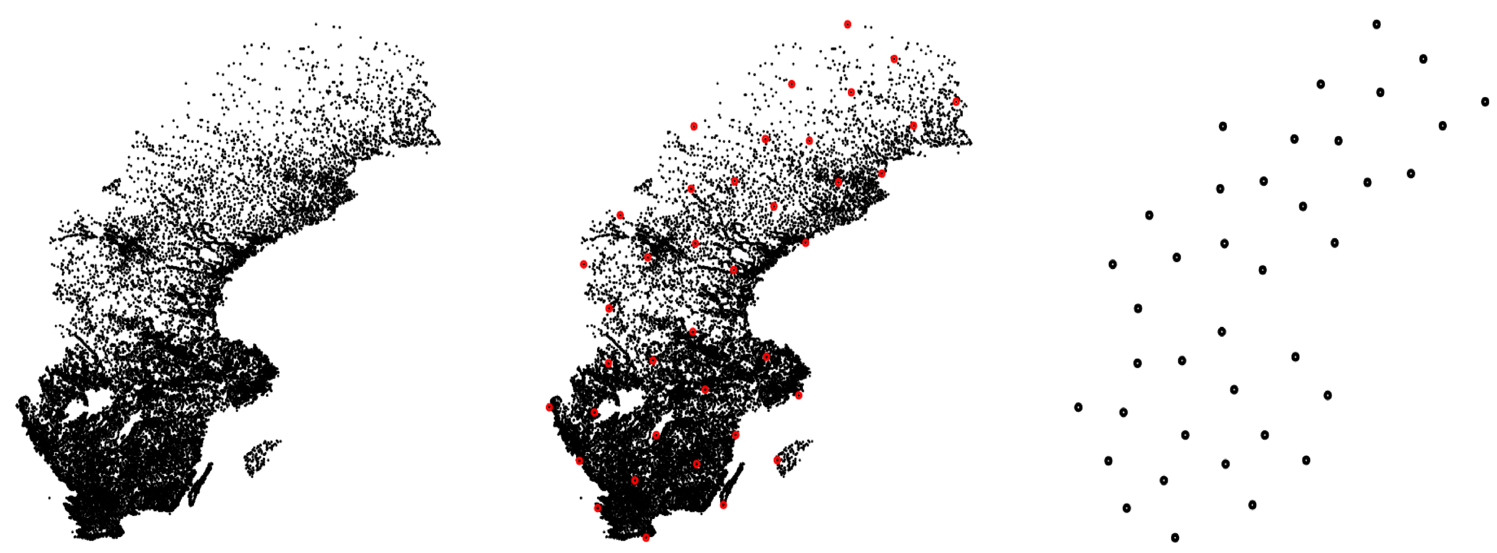


Figure 7: Representative subset of the geographic points of interest of Sweden

References:

[1] M. S. Daskin. *Network and discrete location: models, algorithms, and applications.* John Wiley & sons, 2011

[2] S. Elloumi, M. Labbé, and Y. Pochet. A new formulation and resolution method for the p-center problem. *INFORMS Journal on Computing*, 16(1):84-94 2004

[3] P. Bose and P. Morin. Online routing in triangulations. In *Algorithms and Computation*, pages 113-122. Springer 1999

[4] H. Sagan. *Space-Filling Curves*, volume 18. Springer-Verlag New York, 1994.

[5]

[6]

Acknowledgements:

This work is being developed within the work plan of a research grant for undergraduates, in the context of the R&D project QREN unit in co-promotion no.34164 - Smartgeo Portal ”Geographic Information Management System”. The project is co-financed by QREN and Innovation Agency within the scope of the technological development and research incentive system, by the Operational Program for Competitiveness and Internationalization (Compete).