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

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

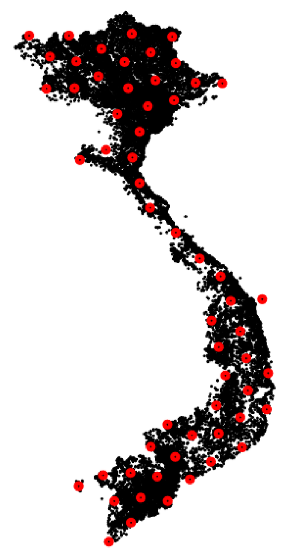
Luís Paquete
CISUC/DEI/University of Coimbra
Coimbra, Portugal
paquete@dei.uc.pt

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

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

Introduction

This work aims to design a real time algorithm to find a representative subset of points in a geographic region. It aims to reduce visual information so that it is not overwhelming for users and programs to easily manage and analyze.
The subset chosen must represent all points of the original set, as to keep some of its visual properties. It must also be efficient enough to be used in a real-time application using data from the Open Street Maps project.



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 maximum distance of a non-centroid to its closest centroid is minimized.

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

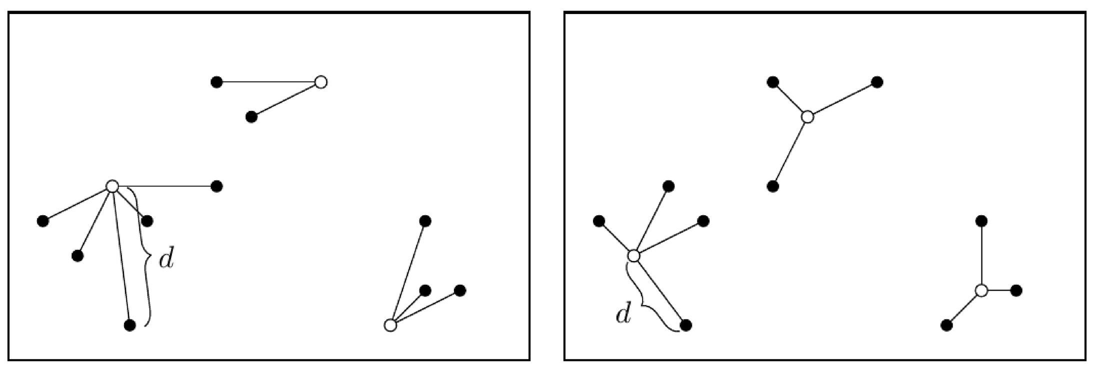


Figure 2: Suboptimal centroid selection (left) and optimal centroid selection (right)

This problem is known as Minimum Coverage Subset and is an example of a Facility Location k-center problem [1]. This can be formulated as Integer Linear Programming [2] and solved by 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 decreases the value of the objective function of the partial assignment.
- a new non-centroid is assigned to closest centroid and increases the value of the objective function of the partial assignment.

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.

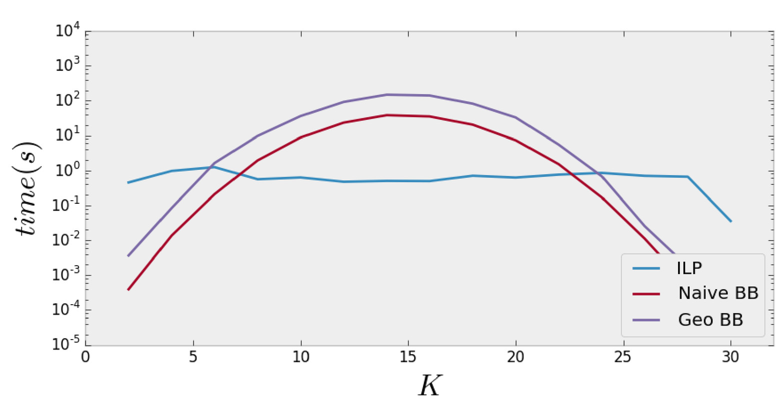


Figure 3: Effect of the value of K

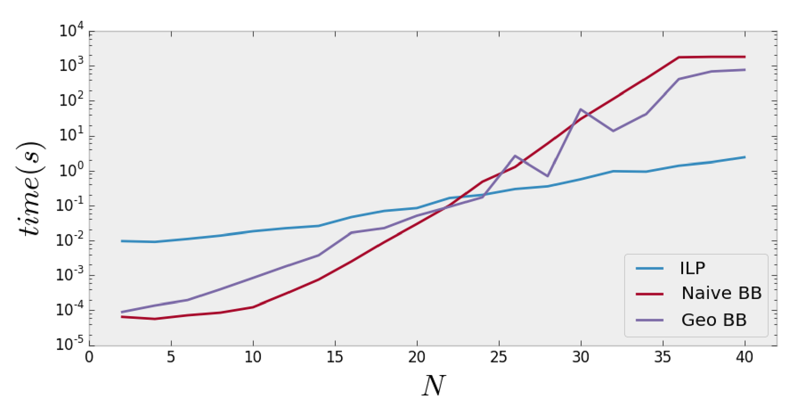


Figure 4: Effect of the value of N

Results:
Both Branch and Bound implementations were too slow. The solver for the Integer Linear Programming performed better, but it is 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 based on a greedy approach to the set cover problem 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].
While the graph still has uncovered points, select as a centroid the point whose number of uncovered neighbors is the largest and mark its neighbors as covered. This is a greedy approximation algorithm for the set cover problem, and can be performed in O(n log n) time [6].

This algorithm ensures that the selected set has no more than $m \ln n$ elements, with m being the optimal number for total coverage.

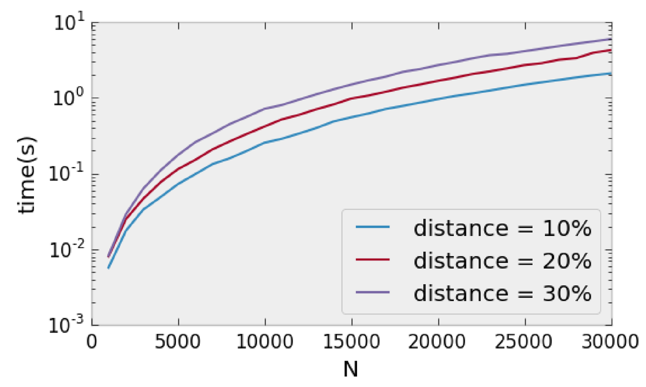


Figure 5: Effect of the value of N on *cpu* time

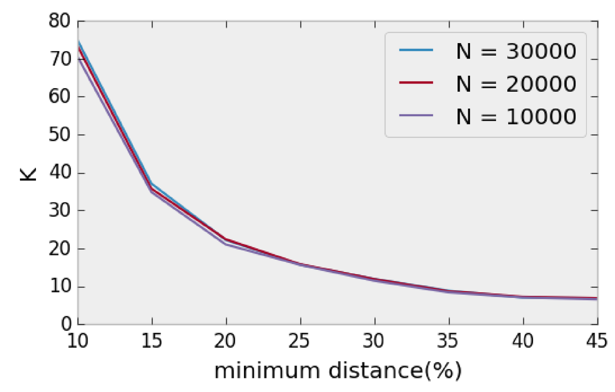


Figure 6: Effect of the point density on the value of K

Results:
This approach is fast enough to be used in real time and depends on the established minimum distance as well as the density. The number of points chosen does not vary with the density. The algorithm can be sped up with a faster/preprocessed range search.

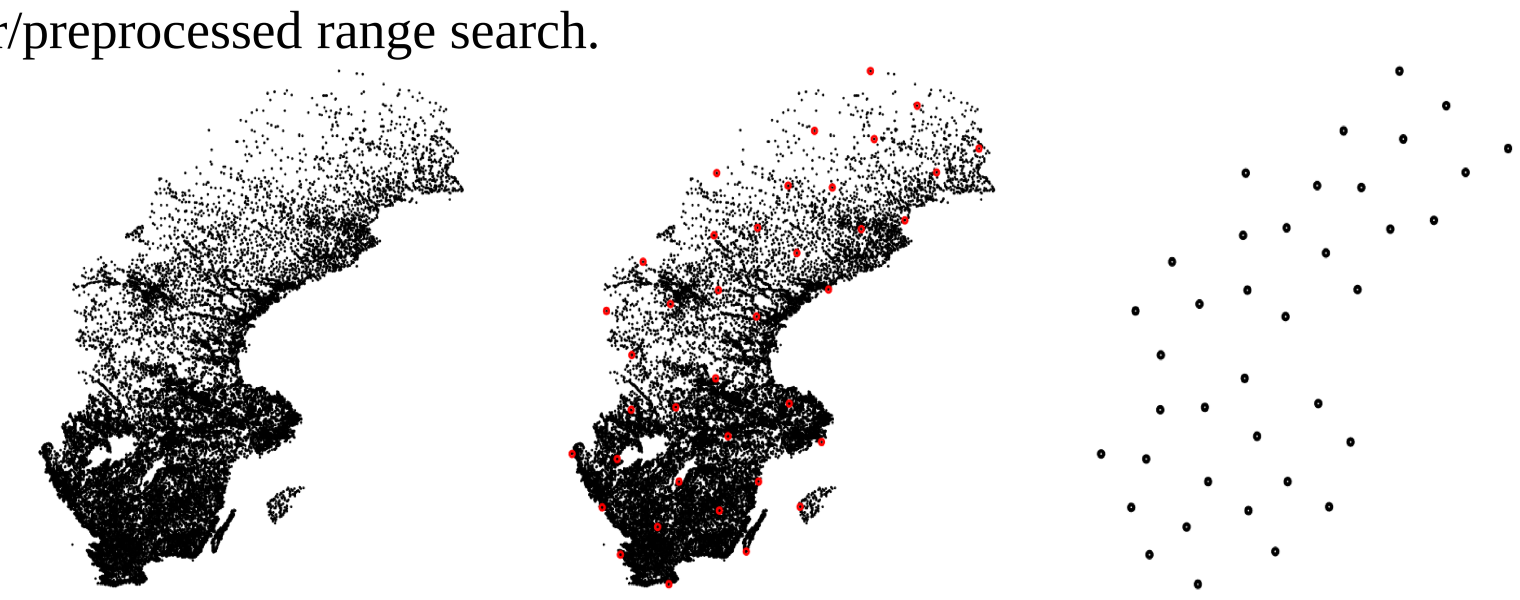


Figure 7: Representative subset of the geographic points of interest of Sweden

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] Lee, D. T., & Wong, C. K. (1977). Worst-case analysis for region and partial region searches in multidimensional binary search trees and balanced quad trees. *Acta Informatica*, *9*(1), 23-29.

[6] V. V. Vazirani. *Approximation algorithms*. Springer Science & Business Media, 2013.

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).