



INSTITUTO SUPERIOR TÉCNICO
Universidade Técnica de Lisboa

A module for solving point location problems in a GIS environment

Isabel Cristina Bilro Batista

MSc in Territorial Engineering, IST

October 2010

Abstract

Problems concerning the location of facilities are frequent in territorial management and planning at various scales, from the inspection on the ideal combined location for a network of hospitals at the national level to the best location of shops or waste collecting bins within a municipality. Mathematical models can be used to represent such spatial problems in the planning context, together with a wide range of methodologies to solve them, but sometimes due to unawareness, unavailability or operative difficulties, this field has not been explored to the full extent of its applicability. On the other hand, Geographic Information Systems (GIS) are widely used programs that enable multiple types of analysis for an extensive range of spatial problems, but, generically, do not have the ability to provide solutions for location problems.

In this work a module to solve point facility location problems operating on a GIS package is presented. The issues on the integration of the models and solution methodologies for location problems in a GIS are focused. This module has the particularity of being usable not just by experts on location problems but by all GIS users.

In a first section, the various models for point location problems considered in the work are presented. The used heuristic methodologies for solving the referred problems are also discussed: some are collected from the literature without changes, while others were improved from known algorithms. A questionnaire was developed and included in the module, allowing users to identify the type of problem they are interested in. In the last part, an application example displays a complete session with the implemented module.

1. Introduction

In the last years, changes in the Portuguese legislation for territorial planning have promoted new planning approaches, where the dissemination of information and capabilities of interaction with the citizens favour the possibility of a larger public participation (Abrantes and Estanqueiro, 2008). These changes are related with the integration of information systems, such as Geographical Information Systems (GIS), in planning tasks and promoting the access to information. The most important capability of these systems is the spatial analysis of the territory, which nowadays has turned to be essential for planners.

The location of educational or health equipments, shops or recycling bins in a territory are examples of location problems that could be addressed by GIS, as intrinsically defined by spatial constraints. In planning, these are common problems and sometimes the decision favouring a location among a set of alternatives is not justified with mathematical models, just because the planner is not aware of the existence of these models or because there are difficulties in its use. The mathematical modelling of problems for which the solution is the location of some facilities with a few constraints (covering, costs, or other types) is known as “Location Science” and these problems are known as facility location problems. Dedicated optimization software is used to solve these problems, but the software requires modelling them in terms of variables and constraints. The solution methods are said to be exact if the optimal solution is found, or, mainly in large combinatorial problems, described as heuristic if the output solution isn’t guaranteed as optimal but it is the best solution in the analysed set of alternatives in the solution space.

If a user intends to use this type of computational solutions, he has to know the structure of his problem, the methodologies to solve it and the impact of the solutions in his decisions. Computational solutions are not integrated, so the user has to export data between distinct platforms. This is called a “loose coupling” approach (Church 2002).

It would be interesting to use the geographical component of the problem, the functionalities of a GIS software and the various solution methodologies to solve facility location problems in the same platform without using loose coupling approaches.

In the second section of this study, some facility location problems will be described. In section three, solutions methods for facility location problems will be presented. In the fourth section, a GIS-based module for solving facility location problems is shown. Finally, in the fifth section, conclusions on the study and suggestions for its future developments are presented.

2. Location Problems

2.1 Models for location problems

There are multiple models of facility location problems. The most common are called *coverage*, *centre* and *median* problems. All these can be modelled by a graph (network) for which the nodes are the possible locations for the facilities and edges represent links, costs or distances between the nodes. One variant of median problems is the *fixed charge facility location problem*. In this section all these four types of problems are presented, starting with the coverage model.

Coverage distance is the maximum distance between a facility and a node that the same facility can serve. This requires the definition of a service to be (or not) applied between two locations: if the distance between the demand node and the facility is less than or equal to a fixed value called coverage distance, the demand is said to be “covered” by the facility.

The objective of coverage problems is to maximize the covered demand with a minimum number of facilities. There are two groups of coverage problems: *maximum covering problems* and *set covering problems*.

The objective of maximum covering location models is to find the set of facilities that maximizes the covered demand, for a fixed number of facilities. The objective of set covering location models is to find the minimum number of facilities that coverage the total demand.

Centre problems use the notion of coverage distance too. However, in these problems the coverage distance is an unknown value and the objective is to minimize such distance, so that all demand nodes will be covered by at least one facility. There are two types of centre problems: *absolute centre problems* and *vertex centre problems*. The difference between these is that in the first problems the facilities can be located anywhere in the network (including along the edges) and in the second ones the facility can only be located in the nodes of the graph that models the problem.

In median problems, the increase of distance leads to the decrease of quality. The notion of coverage distance is not applicable in this type of problems, because it is assumed that all the facilities can cover all demand nodes but with different costs. The objective of median problems is then to find the location of a set of facilities that minimizes the total cost, given a fixed number of facilities.

Finally, fixed charge facility location problems are identical to median problems. The main difference is that candidate sites have different location costs, instead of assuming that all candidate sites have the same cost. The objective of these problems is to find the number of facilities and its locations in order to minimize the total cost. There are two different problem families: the *uncapacitated fixed charge facility location problems* and *capacitated fixed charge facility location problems*. As the name suggests, in the first ones candidate sites have a limit of service they can provide, while in the second case each facility could provide an unlimited capacity to serve the demand.

2.2 Greedy algorithms

For some of the presented location problems, exact solutions are accessible only through algorithms with high computational complexity (in time or memory space required to analyse the solution space). So, for a large sample of nodes, it wouldn't be possible to solve the problem in real time. Heuristic algorithms might not be able to calculate the optimal solution but present the best solution from the explored set of solutions.

Greedy algorithms are iterative heuristics that, in every iteration, change the solution under construction to reflect an improvement on its quality by choosing the best option at the moment (for instance, if it is a coverage problem, the best case is the node that covers more demand which remained uncovered at the current iteration), without calculating the impacts of that step in the next iterations. Two greedy methodologies are the *add* and *drop* algorithms: the add algorithm starts with an empty solution set and adds a single point at each iteration; drop algorithms are the opposite, starting with all the possible locations for the facilities selected in the solution set, while the objective at each iteration is to remove a single element; in both, each step is taken guided by the least raise of cost or the largest saving, until a final solution is reached using some predefined stopping criteria, such as reaching a fixed number of facilities or not being able to find better solutions than the current best found so far.

Substitution algorithms are heuristics too. These algorithms are useful when a set of facility locations is defined and conduct a search with the objective of reducing the solution cost. This is done by

searching for improvements on the solution quality, i.e., by experimenting the replacement of subsets of the selected elements with subsets of those which are unselected.

The methods implemented in this work are the greedy (add and drop) and substitution algorithms.

2.3 Facility location in GIS

GIS functionalities are very interesting for facility location if both GIS and location modules were integrated. Both GIS practitioners and users of facility location software can benefit with this integration. With GIS, facility location software users can calculate the distance between all nodes, combine layers with different coordinates and scales or different databases, insert new data, and other useful operations (Church 2002). In turn, GIS users could easily conceive location problems in the planning context, and explore the available spatial data in new interesting applications.

In ArcGIS™, one of the most widespread desktop GIS packages, there are spatial analysis algorithms involving some sort of optimisation - such as the calculation of least-cost paths -, but effective solutions for the presented location problems are rare. For instance, Business Analyst is a solution developed by ESRI to assist planners and market analysts which in several useful tasks involving spatial analysis (calculation of service areas, drive-time analysis, similarity finding and other spatial-economical analysis) but does not present a solution for the optimisation problems that might occur in the planning context. The only operating module solving a diverse set of location problems was LoLA (Library of Location Algorithms), a solution which integrates the location modules in the ArcView GIS™ platform. The objective of LoLA is to group a set of algorithms to solve location problems and give a quick and easy solution (Drezner and Hamacher 2002). LoLA is composed by three different interfaces: the first is based on the five position classification scheme of Hamacher and Nickel (1998); the second integrates the first interface and a command line; and the third is based in a programming interface where users can edit specific algorithms. However, this solution requires a previous knowledge of both facility location modelling and solution methodologies. Furthermore, LoLA is only configured to work as an extension of the now discontinued ArcView GIS™ 3.x, software.

3. A GIS Module for Facility Location Problems

3.1 Generic description and implemented algorithms

The main objective of developing a GIS module to solve facility location problems is to aggregate GIS and facility location in a unique solution that can be very useful to all GIS and facility location users.

This module was developed in ArcGIS™, using the Visual Basic for Applications (VBA) programming language that this software provides for macro development. Only heuristic algorithms were used to solve the considered facility location problems. The selection of these algorithms was based in their simplicity, utility and good performance. Some of the algorithms were presented by Daskin (1995) or Sridharan (1995). The remaining methods were adapted from solutions presented by Daskin (1995) or developed based on algorithms in the same reference. Eleven different algorithms were integrated in GIS module, one for each of the considered models: three for coverage problems, five for centre problems, one for the p-median problem and two for fixed charge facility location problems. Table 1 presents the name of the algorithms included in the developed module and their main features.

Table 1: Algorithms in the developed location module

Algorithm	Main features
Maximum coverage with P elements	Maximum Coverage – Covers the maximum demand with a number (P) of facilities predefined
Maximum coverage with %	Maximum Coverage – Covers a predefined proportion of demand locating the minimum number of facilities
Set covering	Set Covering – Covers all the demand while locating the minimum number of facilities
Capacitated fixed charge	Fixed Charge Facility Location – Calculates the number of facilities that minimizes the total cost, admitting a limited capacity for each node
Uncapacitated fixed charge	Fixed Charge Facility Location – Calculates the number of facilities that minimizes the total cost, with no considerations about capacity limits
P-Median	Median – Locates a predefined number of facilities while minimizing the total cost, using the median model
Unweighted P-Centre	Centre – Locates (only in nodes) a predefined number of facilities considering that all nodes have the same weight
Weighted P-Centre	Centre – Locates (only in nodes) a predefined number of facilities, considering that nodes may have different weights
Unweighted absolute 1-Centre	Centre – Locates a single facility anywhere in the network, considering that all nodes have the same weight
Weighted absolute 1-Centre	Centre – Locates a single facility anywhere in the network, considering that nodes may have different weights
Absolute 2-Centre	Centre – Locates two facilities anywhere in the network, considering that all nodes have the same weight

As an example, one of the implemented algorithms (for the *capacitated fixed charge* problem) is illustrated in Figure 1 and briefly described.

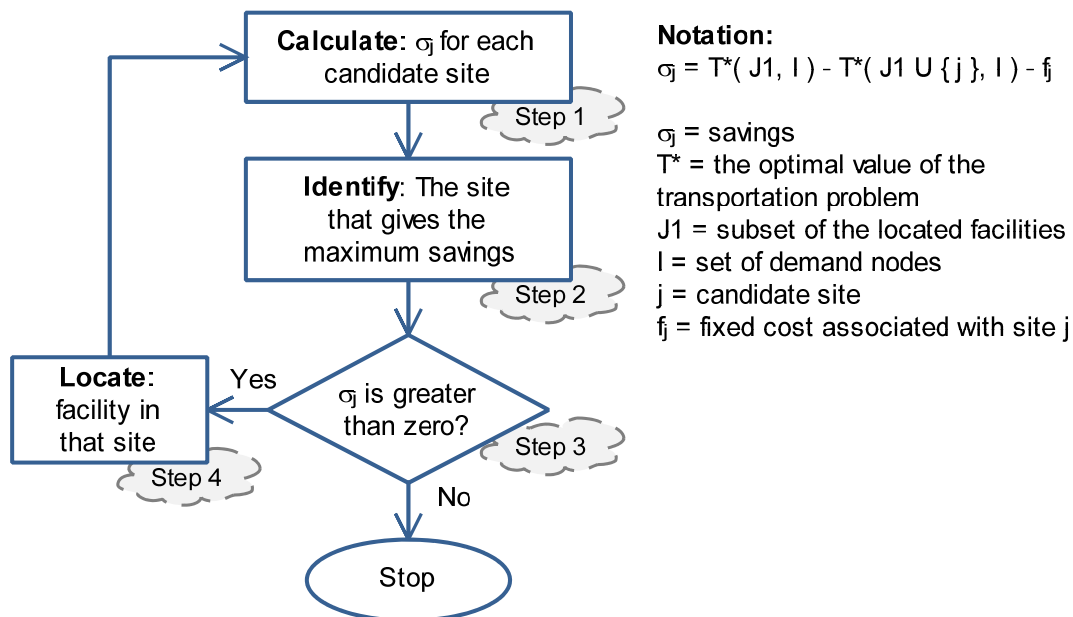


Figure 1: Algorithm for the *capacitated fixed charge* problem (Adapted from Sridharan 1995)

The solution method starts by calculating, for every candidate site σ_j the values of the savings produced if a facility is located there. The location that presents the highest value of savings is memorized in the next step. If that value is positive, a facility is located at the corresponding site. If the savings are not positive, the algorithm halts, since it is no longer possible to further reduce the total cost by activating new facilities. For the savings calculation it is necessary to solve a transportation problem.

3.2 Questionnaire

To achieve the main objective of the module it is necessary that everyone can explore the module being or not familiarized with facility location. For that, an interactive questionnaire was developed. With it, a user can easily identify the model that applies to his problem, enabling the system to select the algorithm for its solution. This is a tree-shaped schema, where each question depends on the previous answers. Figure 2 presents a diagram with the sequence of questions.

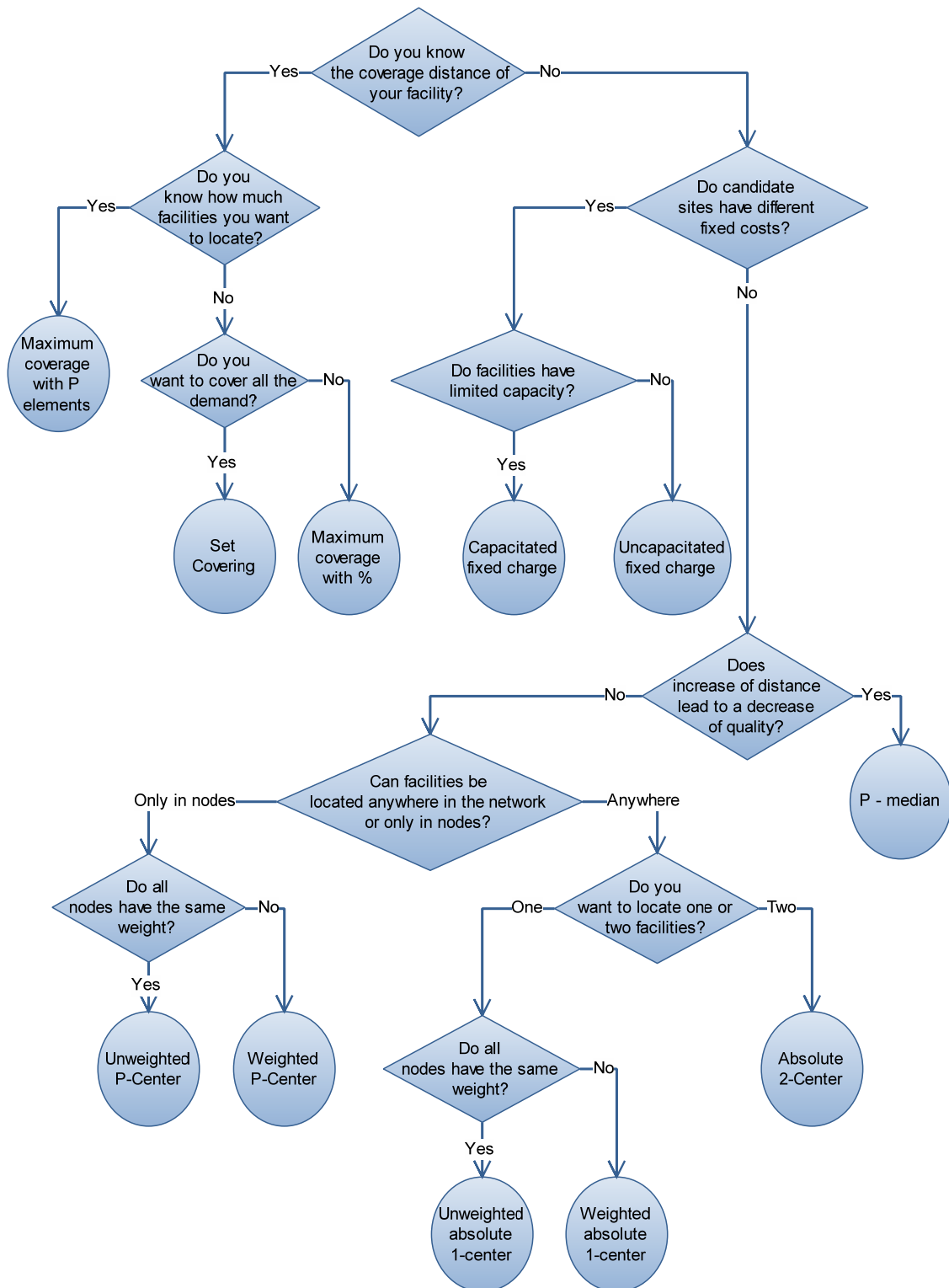


Figure 2: Question diagram for model identification

When the module is initialised, the user can state if he knows or not what algorithm he wants to use. If he knows it, he can choose the algorithm to his problem from the set of algorithms. If not, he has to answer to the questions sequentially. As the user is answering the questions, the window expands and the next question appears. When an answer leads to an algorithm solution, the data entry window displays the model name for the problem under analysis.

The algorithm is executed and the solution appears. In the solution window, the users still has the possibility to save the solution data in a text file.

An example of the developed module usage is presented on Figure 3, which displays a sequence of questions and answers. Supposing that the user does not know anything about facility location problems, he must use this module. In this example, the objective is to locate three facilities in the Timor-Leste territory (Direcção Nacional de Estatística de Timor-Leste 2004).

The screenshot shows a window titled "Module for Facility Location Problems" with a close button (X) in the top right corner. Below the title bar, there is a section labeled "Answer to the following questions:" followed by a "Help ->" button. The window contains five questions, each with two radio button options:

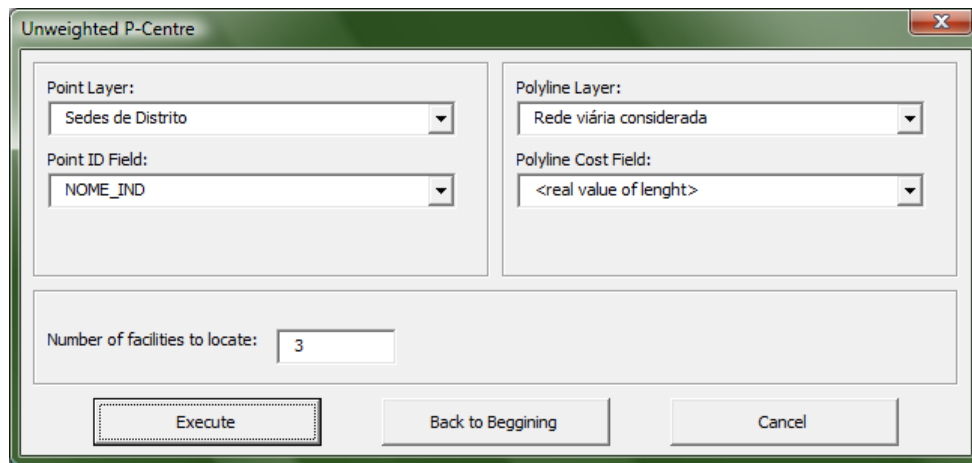
- Question 1: "Do you know the coverage distance of your facility?" with options "Yes" and "No". The "No" option is selected.
- Question 2: "Do candidate sites have different fixed costs?" with options "Yes" and "No". The "No" option is selected.
- Question 3: "Does increase of distance lead to a decrease of quality?" with options "Yes" and "No". The "No" option is selected.
- Question 4: "Can facilities be located anywhere in the network or only in nodes?" with options "Anywhere" and "Only in nodes". The "Only in nodes" option is selected.
- Question 5: "Do all nodes have the same weight?" with options "Yes" and "No". The "Yes" option is selected.

At the bottom right of the window is a "Cancel" button.

Figure 3: Example of a sequence of questions and answers

In this example, the first answer ("No") excludes the coverage problems, the second excludes the fixed charge facility location problems and the third excludes median problems, so to the user is considering the centre model for his problem. The next question concerns the alternatives of locating the facility anywhere in the network or only in the nodes, and this last option was selected. When asked about the weight of nodes, the user answered that all nodes have the same weight. The combination of these answers leads the unweighted P-centre model.

Afterwards, a window to insert data appears, so the user can parameterise the problem using the layers in the GIS window and their attributes. In the example, the user chose to locate three facilities.

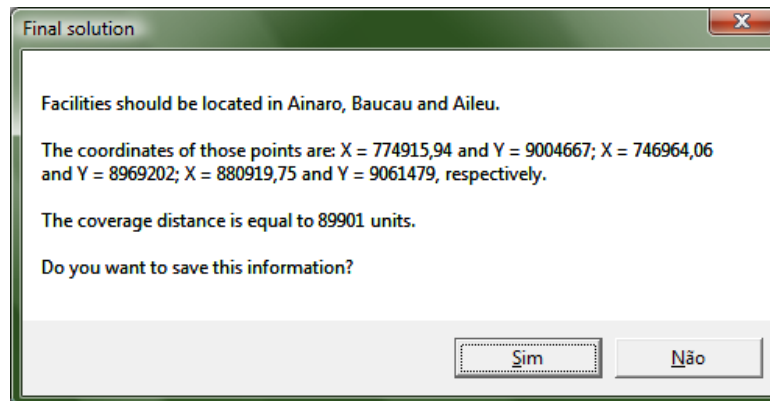


The dialog box titled "Unweighted P-Centre" contains the following fields and buttons:

- Point Layer:** A dropdown menu with "Sedes de Distrito" selected.
- Point ID Field:** A dropdown menu with "NOME_IND" selected.
- Polyline Layer:** A dropdown menu with "Rede viária considerada" selected.
- Polyline Cost Field:** A dropdown menu with "<real value of lenght>" selected.
- Number of facilities to locate:** A text input field containing the number "3".
- Buttons:** "Execute", "Back to Beggining", and "Cancel".

Figure 4: Inserting data for the example

When the “Execute” button is pressed, the algorithm that corresponds to the identified model is invoked and a solution is produced. This solution is displayed in a window reporting the coordinates of the located facilities and the calculated coverage distance, as shown in Figure 5. This information can also be saved in a text file.



The dialog box titled "Final solution" displays the following text:

Facilities should be located in Ainaro, Baucau and Aileu.

The coordinates of those points are: X = 774915,94 and Y = 9004667; X = 746964,06 and Y = 8969202; X = 880919,75 and Y = 9061479, respectively.

The coverage distance is equal to 89901 units.

Do you want to save this information?

Buttons: "Sim" (Yes) and "Não" (No).

Figure 5: Example of a report displayed on the solution window

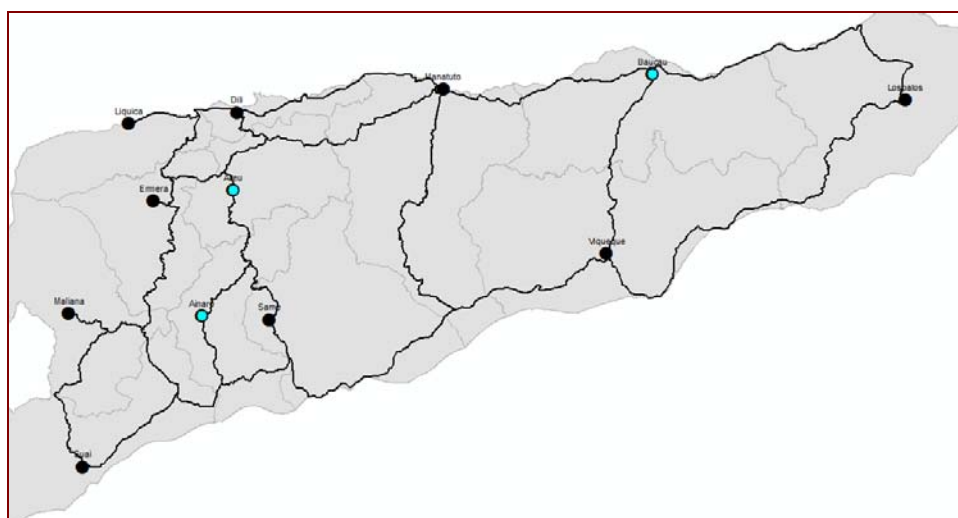


Figure 6: Displaying a solution

When this process ends, the nodes where facilities should be located in accordance with the solution are selected. Figure 6 presents the selected solution for the described example.

4. Conclusions and Future Developments

This work presented a module that enables the production of solutions for point facility location problems in a GIS environment. This module can easily be used by people that are familiarized or not with facility location. The interactive questionnaire is an innovative solution and permits a different interaction with the user, guiding him along the location process.

The developed module leads the users identifying the model for their problems through a questionnaire with a maximum of six questions. This module has advantages for all GIS users that have a facility location problem, namely:

- Users who don't know the models but need a tool to find a solution or to justify their decision
- Users who know the problems and want to use GIS functionalities.

There are some aspects that can be improved in future developments. The first one involves the diversification and further exploration of both models and solution methodologies which apply in the point facility location context. The most usual problems are covered by the developed module but there are many others which might be included. The second one involves the migration of the VBA solution to a Visual Basic solution, so that the module could be used as a GIS extension. The third one implicates the improvement of solution display, for example, creating more complete and informative tables and reports. A fourth development involves testing the capacity for solving problems with real large sets of data as well as exploring and finding the most adequate solution methodologies given the size, complexity and spatial properties involved in a problem.

References

- Abrantes, P.; Estanqueiro, R. (2008) SIG, Análise Espacial e Ordenamento de Territórios Metropolitanos. Sociedade e território, November, pp. 114-124.
- Church, Richard L. (2002) Geographical information systems and location science. Computers & Operations Research, Vol. 29 (6), pp. 541-562.
- Daskin, Mark S. (1995) Network and Discrete Location - Models, Algorithms, and Applications. Wiley, New York
- Direcção Nacional de Estatística de Timor-Leste (2004) The 2004 Population and Housing Census of Timor-Leste. Timor-Leste. In: <http://dne.mof.gov.tl/census/tables/index.htm>
- Drezner, Zvi; Hamacher, Horst W. (2002) Facility Location - Application and Theory. Springer, Berlin.
- ESRI®. 2009. ESRI Business Analyst. Redlands, California. In: <http://www.esri.com/software/bao/index.html>
- Hamacher, Horst W.; Nickel, Stefan (1998) Classification of location models. Location Science. Vol. 6 (1-4), pp. 229-242.
- Sridharan, R. (1995) The capacitated plant location problem. European Journal of Operational Research, Vol. 87 (2), pp. 203–213