ISSN 1870-4069

The P-Median as a Problem of Clustering

María Beatriz Bernábe-Loranca[1], Carmen Cerón-Garnica[1],
Hugo Rodríguez-Cortes[2], Rogelio González-Velázquez[1]

[1] Benemérita Universidad Autónoma de Puebla,
México.

[2] Instituto Politécnico Nacional,
Centro de Investigación y de Estudios Avanzados,
México

{maria.bernabe, carmen.ceron,
rogelio.gonzalez}@correo.buap.mx,
hrodriguez@cinvestav.mx

Abstract. If a problem of Area Design can be modeled under conditions of classic Partitioning, then the problem is often called Territorial Clustering (TC) since the computational solution uses algorithmic techniques of portioning clustering. Under these characteristics, Partitioning can be seen as a methodology of support to solve problems of P-median and territory design type. Partitioning in the solution of territorial problems consists in to cluster small geographic areas called basic unity in a given number of bigger groups named territories. Such description requires of a mathematical model for its expression, and one of the useful definitions in the model is the definition of discrete partition. Is in this point where the following work is situated: it is proposed to show that the definition of Zones Design (ZD) and partitioning share some properties in their definition, both in the restrictions and in objective function. Likewise, the transformation of the P-median problem model of integer binary problem to a combinatorial optimization model is posed.

Keywords: Territory design, integer-binary model, combinatorial model, P- median.

1 Introduction

The problems of territorial clustering have applications in the determination of political and scholar districts, installation of social and emergency services, commercial territories, etc. In various works, geographic criteria are used as adequacy measures of solutions, for example, that a territory is geographically compact, and it is often used as compactness measure the sum of distances between the basic unities and the centroid to which they are assigned, thus modeling the problem as the P-median.

The modeling of clustering problems adjusted as the P-median motivate the study of instances of big scale, as an example, the problems of P-median defined in graphs $G = (V,A)$ with $|A| >= 360,000$ are difficult to solve with commercial software intended for problems of Mixed Integer Programming, then additional strategies are needed such as the metaheuristics to solve the problem.

However, the objective in this document is centered in to present the conditions of equivalence between the P-median and the clustering problem. On the other hand, the transformation of the model of the P-median is proposed from an integer-binary to a combinatorial.

2 Preliminaries

Definition 1. Let the initial set of unities of area be $X = \{x_1, x_2, \ldots, x_n\}$ where the *ith* unity of area is x_i. The number of zones is denoted by k and Z_i is the set of all unities of area that belong to the zone Z_i [1]. Then:

$$Z_i \neq \emptyset \quad \text{for i= } 1 \ldots k, \tag{1}$$

$$Z_i \cap Z_j = \emptyset \ \text{ for i} \neq \text{j y}, \tag{2}$$

$$\cup_{i=1}^{k} Z_i = X. \tag{3}$$

Therefore, (1), (2) and (3) constitute the set of "equivalent restrictions" both in the clustering and in zones design.

It is stablished that these criteria comply indirectly with at least the geometric compactness, although in several computational solved problems it has been graphically observed that compactness and connectedness are satisfied.

In other cases, it has been identified in maps that the contiguity is also fulfilled, but this visual observation is valid only for some applications such as the P-median and it is not guaranteed that the contiguity is reached, even some authors have already warned that the inclusion of contiguity represents a more complex problem [2].

Zones design, territorial design, design of territory in a pragmatic sense, share the same semantic, also the terms of zoning or regionalization are [3-5].

Now, considering the definition 1 and observing the following definition 2:

Definition 2. Let a partition of a set A of n elements in k pairs, be a family of k non-empty disjoint subsets of A, such that their union is the same set A, thus $A_1, \ldots, A_k$, are the subsets of the partition, then it is fulfilled:

2.1 Problem of Territorial Clustering as a Problem of the P-Median

An application of Territorial Design (TD) that can be employed as methodology when criteria of compactness and connectedness are treated, are the location-allocation models. When talking about model, it refers to the mathematical representation of the problem and it is translated to a methodology always that points of the model are respected and adapted to a series of algorithmic instructions.

This leads to the election of adequate techniques that answer to the model, such that candidate algorithms are identified for its solution and consequently, evaluate them to give answer to the model.

However, any variant in the model that does not change the meaning, implies a strategy reasonably distinct, but that leads to the same solution, it is said that the model of the basic P-median is the same in all the literature, although the methodology and proposal of algorithm that are suggested for its solution is different but searching the same result.

For example, when different metaheuristics are used to answer the P-median in the combinatorial model, the solutions are not exactly the same since they are approximated, but always trying that such approximations get close to the optimum, or in the best of the cases to be the exact value. To value the implementations, the instances of the P-median test are found on the website OR-Library [6].

The P-median can be stablished in terms of graphs as follows:

Let G = (V, E) be a graph not aimed where V is the set of n vertices and E is the set of edges with a weight associated that can be the distance between the vertices $d_{ij} = d(v_i, v_j)$ for all $i, j = 1, \dots, n$ according to certain metrics, then with the distances a symmetric matrix is formed, then $V_p \subseteq V$ must be found such that $\left|V_p\right| = p$, where p can be either variable or fixed, and that the sum of the shortest distances of the vertices in $\left\{V - V_p\right\}$ to its nearest vertex in V_p is reduced to the minimum.

Particularly, the P-median is useful in these conditions independently of the operations research approach (integer-binary) or the combinatorial problem (tackled with metaheuristics).

The problem of the P-median has been widely studied in literature. In [7] can be found an excellent revision of methodologies to solve the P-median of approximated form and exact or through graphs.

Geographically, P-median is observed to partition the territory, then it can be said that just for this characteristic, the P-media belongs to problems of Territorial Design, which have diverse applications such as the identification of political districts, social services installation, commercial territories, location-allocation, etc. [3-5].

In lots of works, geographic criteria of adequacy measures of the solutions are utilized. The criteria commonly used are compactness, connectedness and contiguity. According to Kalcsis, a territory is geographically compact if it has a round approximated form and it is not distorted, but there is no rigorous definition of the concept [3].

3 Analysis of the Clustering Problem as One of Optimization

Let the clustering be a problem of partitioning grouping, then given a set of n objects denoted by $X = \{x_1, x_2, \dots, x_n\}$ in that $x_i \in R^D$ let k be a positive integer known a priori, the clustering problem consists in finding a partition:

$P = \{C_1, C_2, \dots, C_k\}$ of X, being C_j a conglomerate (group) conformed by similar objects, satisfying an objective function $f\colon R^D \to R$, and the conditions:

$$C_i \cap C_j = \emptyset \text{ for } i \neq j \text{ and } \bigcup_{j=1}^{k} C_j = X$$

To measure the similarity between two objects x_a and x_b, it is used a function of distance denoted by $d(x_a, x_b)$, being the Euclidean distance the most popular to measure the similarity. Thus, the distance between two different elements:

$$x_i = (x_{i1}, \dots, x_{iD}) \quad and \quad x_j = (x_{j1}, \dots, x_{jD})$$

is $d(x_i, x_j) = \sqrt{\sum_{l-1}^{D}(x_{il} - x_{jl})^2}$.

The objects of a conglomerate are similar when the distances between them is minimal; this allows to formulate the objective function$f$, as:

$$\sum_{j=1}^{k} \sum_{x_i \in C_j} d(x_i, \overline{x_j}). \tag{1}$$

That is to say, it is desired to minimize (1); where $\overline{x_j}$, known as representative element of the conglomerate (group), is the mean of the elements of the conglomerate:

$$C_j, \overline{x_j} = \frac{1}{|C_j|} \sum x_i \in C_j. \tag{2}$$

And it corresponds to the center of the conglomerate. Under these characteristics, clustering is a problem of combinatorial optimization, and has been demonstrated that is NP-difficult.

3.1 P-Median Classic Approach

The P-median consists in, given a set of points (or location of consumers) and a matrix of distances (or costs) between all and each of points, choosing p points (or location of installations) with the purpose of minimizing the sum of each of the distances of all points to the “nearest chosen point”.

In 1970 ReVelle and Swain presented the first formulation of integer programming for the p-median problem cited by Church in 2003 [8-12]. In general, the p-median problem can be expressed mathematically as a problem of discrete optimization.

First, the matrix of distances is denoted as d_{ij}, that expresses the distance between the potential points of location i and the points of demand j. The binary variable x_{ij} corresponds to the allocation or not allocation of the demand points j to the installation i. The binary variable y_i indicates that an installation is stablished in the point i or not. The approach as an integer binary problem is the following form:

Let $$x_{ij} = \begin{cases} 1 & \text{If the point } j \text{ is assigned to the point } i \\ 0 & \text{in other case;} \end{cases}.$$

and $$y_i = \begin{cases} 1 & \text{If in the point } i \text{ is located in an installation} \\ 0 & \text{in other case;} \end{cases}$$

$$min \quad Z = \sum_{i=1}^{k} \sum_{j=1}^{n} d_{ij} x_{ij}, \tag{a}$$

subject to $$\sum_{i=1}^{k} x_{ij} = 1 \quad \forall j = 1,2,\ldots,n; \tag{b}$$

$$\sum_{j=1}^{n} x_{ij} \leq n y_i \quad \forall i = 1,2,\ldots,k; \tag{c}$$

$$\sum_{i=1}^{k} y_i = p, \tag{d}$$

where k is the number of potential vertices where a median, generally$k = n$, can be localized, p is the fixed number of required medians. The equation (a) is the objective function that minimizes the distance of the system, the restriction (b) stablishes that each demand point can only be allocated in an installation, the restriction (c) stablishes that the assignation of demand points to each one of the installations or medians and finally the restriction (d) guarantees that between k potential points of location that choose exactly p.

Test. We define, $u_{jr} = \begin{cases} 1 & if\ \ r\ is\ assigned\ to\ \ j \\ 0 & in\ other\ case \end{cases}$

Since as each point belongs to a solely group, it is satisfied $\sum_{j=1}^{k} u_{jr} = 1$

Moreover, $\sum_{x_i \in C_j} d(x_i, \overline{x_j}) = \sum_{r=1}^{n} u_{jr} d(x_i, \overline{x_j})$

Then $\sum_{j=1}^{k} \sum_{x_i \in C_j} d(x_i, \overline{x_j}) = \sum_{j=1}^{k} \sum_{r=1}^{n} u_{jr} d(x_i, \overline{x_j})$

If the problem is interpreted as potential points of location j and the demand points r. The binary variable u_{jr} corresponds to the allocation of demand points r to the installation j or not. The variable $\overline{x_j}$ can be interpreted as a installation that is stablished in the group C_j:

$$y_j = \begin{cases} 1 & in\ the\ installation\ \ \overline{x_j} \\ 0 & in\ other\ case \end{cases}$$

Then it is satisfied:

$$\sum_{j=1}^{k} y_j = k \quad and \quad \sum_{r=1}^{n} u_{jr} = n y_j$$

In conclusion, all hypothesis of the model of the P-median and vice versa are fulfilled.

4 Transformation of the Model of the P-Median Problem of Integer Binary Programming to Combinatorial Optimization

Is of interest to show the transformation of the model of integer-binary programming of the P-median to one of combinatorial optimization.

To implement approximation algorithms, in the search of solutions to the NP-hard problems such as the P-Median is, it must be approached as a combinatorial optimization problem [13].

The problem of the P-median is correct in the solution of problems of localization, of territorial design, set partitioning, of cluster design, among others and it is necessary to construct methods of search of solutions such as the metaheuristics [11].

The applications of the localization problems are responsible of solving the location of one or several installations, in this way optimizing one or various objectives, such is the case of transport cost, client service, market partition, etc. The study of localization problems involves a lot of fields of knowledge as operations research, administration science, industrial engineering, computer science, urban planning and different related areas. Among the applications to the design of logistic networks, there is territorial design, location of warehouses, production plants, assembly plants, hospitals, fire stations, police stations, schools, etc.

To show the transformation of the mathematical model of the problem de la P-median as one of programming integer binary to one of combinatorial optimization, the following nomenclature is given:

PPM= problem of the p-median,

PIBP=problem of integer binary programming,

PCO= problem of combinatorial optimization.

First, it is important an introduction to the model of combinatorial optimization of P-median. Let be the problem of the P-Median (PPM) in the following way:

Given a set of n vertices of a graph in the plane denoted by$V = \{1, 2, \dots, n\}$, $|V| = n,$ the objective is to find a subset of vertices $L \subseteq V,$ with $|\mathrm{L}| = p$ of potential locations of the medians such that the total mean distance of the design is the minimum. It is said that this approach of is of type Problem of Combinatorial Optimization (PCO) since the space of feasibility Ω of PPM are all the subsets L of cardinality p of a set of cardinality $n,$ with $p < n,$ this is:

$$|\Omega| = \frac{n!}{p!\,(n-p)!} = \binom{n}{p}$$

In general, a instance for the PPM is denoted by PPM(n, D, p), where n is the cardinality of the set V, p is the cardinality of the obtained subsets y $D = (d_{ij})$ is the matrix of distances among all pair of vertices of V.

Second, let be the problem of the P-median seen as an integer-binary model, then the mathematical model of the PIBP of the PPM is developed as follows:

The PPM about a graph can be expressed mathematically as an integer-binary programming problem (PIBP) of the form [8]:

$$\text{Let } x_{ij} = \begin{cases} 1 & \text{if the vertex } j \text{ is assigned to the vertex } i \\ 0 & \text{in any other case} \end{cases}$$

$$\text{Sea } y_i = \begin{cases} 1 & \text{if the vertex } i \text{ is a median} \\ 0 & \text{in any other case} \end{cases}$$

$$Min\ Z = \sum_{i=1}^{k} \sum_{j=1}^{n} d_{ij} x_{ij}, \tag{1}$$

$$\text{subject to} \sum_{i=1}^{k} x_{ij} = 1 \quad \forall\, j = 1, \dots, n \quad , \tag{2}$$

$$\sum_{j=1}^{n} x_{ij} \leq n\, y_i \quad \forall\, i = 1, \dots, k \quad , \tag{3}$$

$$\sum_{i=1}^{k} y_i = p \ . \tag{4}$$

Table 1. Binary solutions.

Binary Solution	Median	Group
$x_{61}=1, x_{62}=1, x_{65}=1, x_{69}=1, x_{611}=1$	6	6, 1, 2, 5, 9, 11
$x_{73}=1, x_{74}=1, x_{78}=1, x_{710}=1$	7	7, 3, 4, 8, 10
$x_{1512}=1, x_{1514}=1, x_{1517}=1$	15	15, 12, 14, 17
$x_{1316}=1$	13	13, 16

Table 2. Results for the OR-Library P-Median instances (Part 1).

Instance	VNS		SA		VNS-BIO	
	Cost		Cost	T	Cost	T
1	5819	103	6209	28	5884	0.55
2	4341	180	4646	70	4601	7.872
3	4467	180	4785	47	4870	7.793
4	3380	250	3693	93	3744	2.877
5	1664	360	1820	119	1771	2.749
6	7917	330	8349	49	8236	2.427
7	5952	540	6446	104	6415	3.1
8	5204	1003	5536	174	5507	2.647
9	3385	1860	3626	286	3432	2.134
10	1700	1980	1850	907	1844	2.662
11	7803	720	8346	149	7968	9.762
12	7200	900	7717	298	7485	9.615
13	5126	1860	5475	692	5444	9.877
14	3823	1440	3992	71	3907	10.403
15	2464	240	2558	1721	2482	10.354
16	8423	300	8958	220	8736	21.732
17	7651	540	8197	322	7905	20.896
18	5821	2700	6038	505	5935	22.986
19	3747	2760	3881	2036	3780	24.238
20	2647	2040	2755	1909	2707	24.865
21	9557	240	10231	102	9794	43.587
22	9433	300	9802	170	9744	38.277
23	5645	3600	5941	839	5827	44.257
24	3974	600	4065	1440	3988	46.955
25	2726	720	2852	240	2810	51.757

26	10312	60	10869	14	10458	64.285
27	9065	120	9511	25	9159	80.152
28	5664	480	5799	141	5768	75.77
29	4114	780	4176	70	4132	86.531
30	2960	1320	3058	47	2996	122.48
31	10528	70	11157	93	10739	105.61
32	10383	120	10818	119	10606	34.87
33	6007	720	6166	49	6102	43.789
34	4193	1500	4286	104	4191	53.045
35	11037	120	11698	174	11180	47.757
36	9994	180	11544	250	11154	60.261
37	6460	1620	6715	50	6602	62.387
38	11725	180	12252	10	11678	68.799
39	10570	300	11017	25	10599	62.698
40	6632	2460	6803	40	6664	72.65
32	10383	120	10818	119	10606	34.87
33	6007	720	6166	49	6102	43.789
34	4193	1500	4286	104	4191	53.045
35	11037	120	11698	174	11180	47.757
36	9994	180	11544	250	11154	60.261
37	6460	1620	6715	50	6602	62.387
38	11725	180	12252	10	11678	68.799
39	10570	300	11017	25	10599	62.698
40	6632	2460	6803	40	6664	72.65

The equation (1) is the objective function with variables of decision x_{ij} and y_i binary. The number of potential vertices where a median can be located is k and generally $k = n$. The fixed number of required medians is p. The restrictions (2) guarantee that every vertex has associated a median. The restrictions (3) indicate the distribution of the vertices to the medians. The (4) determines the number of medians.

4.1 The Proceeding for the Transformation through the Analysis of a Case

The PPM as problem of integer binary programming (PIBP) has a space of feasibility of exponential type 2^n and as a POC is $\binom{n}{p}$

Then a transformation is $T: 2^n \to \binom{n}{p}$

Considering a case of a graph of $n = 17$ vertices, that is to say, $V = \{1, 2, \ldots, 17\}$ illustrated in Fig. 1, whose solution was determined by the PIBP model that is observed in Fig. 2. The results with a cost C determined by the objective function of the equation (1) are represented in the following figures.

The solution is shown in Table 1 and the binary solution determines the medians in $y_6 = 1, y_7 = 1, y_{15} = 1, y_{13} = 1$.

The subsequent combinatorial solution obtained with an enumerative algorithm of Partitioning, are the “centroids” $L = \{6, 7, 13, 15\}$. Said solution has a cost C obtained of the sum of the sums of the distances of the “medians”, now expressed as centroids to their associated vertices, that is to say:

C = d(6,1)+ d(6,2)+d(6,5)+ d(6,9)+ d(6,11)+ d(7,3)+ d(7,4)+ d(7,8)+ d(7,10)+ d(15,12)+ d(15,14)+ d(15,17)+ d(13,16).

The test instance for the combinatorial problem of the P-Median $(17, D_{17x1}\ , 4)$ has shown that the equivalency between the solutions of PIBP and PCO for the PPM.

5 Application

Since the P-median is a problem of NP-hard, we have resolved the P-median with different approaches and strategies. The most recent result is a hybrid metaheuristic (Hybrid VNS/TABU), which consists in a smart combination of Variable Neighborhood Search (VNS) and Tabu Search (TS) that collects the most important approximations [15]. In this strand, previous articles that helped to find solutions like those presented in VNS-TS, can be seen in [15, 16].

In this section, we have chosen data from OR-Library [17] and we have tested four algorithms that have given good results for geographical data: 1.-Simulated Annealing (SA), 2.-Variable Neighborhood Search (VNS), 3.-Bioinspired Variable Neighborhood Search (VNS-BIO) and a 4.-Tabu Search-VNS Hybrid (H-TS-VNS).

However, the partitioning method PAM (Partitioning Around Medoids), that is modeled like the P-median, attained similar results along with H-TS-VNS, but better results than the other metaheuristics for the OR-Library instances, in a favorable computing time.

Nevertheless, for bigger instances that represent real states in Mexico, H-TS-VNS has surpassed PAM in time and quality in all instances. We expose the behavior of these five different algorithms for the test matrices from OR-Library and real geographical data from Mexico.

Furthermore, we made an analysis with the goal of explaining the quality of the results obtained to conclude that PAM behaves with efficiency for the OR-Library instances yet is overcome by the hybrid when applied to real instances.

On the other hand, we have tested the two best algorithms (PAM and H-TS-VNS) with geographic data from Jalisco, Queretaro and Nuevo León. At this point, as we said before, their performance was different than the OR-Library tests. The algorithm that attains the best results is H-TS-VNS.

The following tables contains the tests for the OR-Library instances. The nomenclature is Cost=objective cost and T=time (seconds). The details of each one of the 40 instances can be seen in [17], they go from 100 to 900 objects and the values of P from 5 to 200.

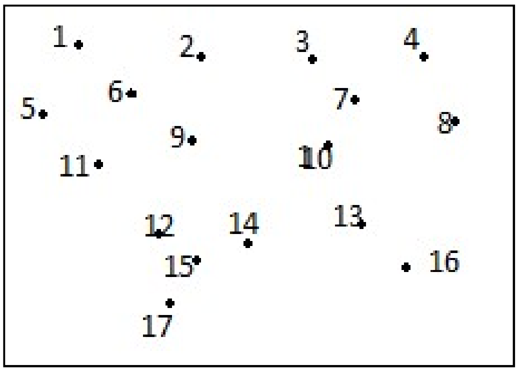

Fig. 1. Graph of 17 vertices.

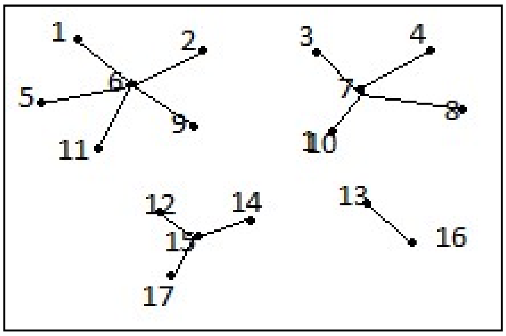

Fig. 2. Case of $n = 17$ with $p = 4$.

Table 3. Results for the OR-Library P-Median instances (Part 2).

Instance	PAM		H-VNS-TS	
	Cost	**T**	**Cost**	**T**
1	**5819**	0	**5819**	2.556
2	4105	0	**4093**	1.672
3	**4250**	0	**4250**	1.604
4	3046	1	3041	5.703
5	**1355**	1	1394	5.928
6	**7824**	0	**7824**	49.28
7	5645	1	**5631**	21.744
8	4457	2	4451	19.764
9	2753	8	2804	31.729
10	1263	14	1318	25.288
11	**7696**	0	**7696**	145.137
12	**6634**	1	**6634**	63.67
13	**4374**	20	4388	48.169

14	2974	56	3091	37.845
15	1738	82	1858	48.857
16	**8162**	1	**8162**	222.629
17	**6999**	2	**6999**	97.449
18	4811	67	4840	25.538
19	2859	296	2927	29.422
20	1805	600	1882	36.45
21	**9138**	0	**9138**	164.141
22	8669	4	8579	58.606
23	**4619**	160	4664	58.606
24	2965	938	3093	127.046
25	1844	1608	1937	132.722
26	**9917**	2	**9917**	389.316
27	**8307**	9	**8307**	68.365
28	4515	605	4551	35.594
29	3039	2101	3181	66.12
30	2009	2208	2119	105.318
31	**10086**	2	**10086**	479.083
32	**9301**	8	**9310**	109.158
33	4703	1495	4735	47.558
34	3026	4685	3168	119.309
35	**10400**	2	**10400**	413.429
36	**9934**	10	**9934**	141.098
37	5064	2092	5278	68.316
38	**11060**	8	**11060**	86.544
39	**9423**	13	**9423**	99.102
40	5138	5076	5214	76.852

The values in bold in both tables represent the tests that returned the best-known value (probably the global optimum) until today.

We can see that PAM achieves the best results, but its computing time increases as the problem size increases. The worst algorithm was SA.

Even though H-TS-VNS attains good results for the instances in Table 4 it's not a guarantee that it will be the same for the real geographical data of our interest. For this reason, we selected PAM and H-TS-VNS to test them with data from three states of Mexico: Jalisco, Querétaro and Nuevo León (3484, 814 and 2416). Our results are in Table 4.

For Table 4, we executed 9 instances, 3 for each map using 12, 24 and 48 P's (medians) for each. The instances 1, 2 and 3 are for the map of Querétaro that has 814 objects, instances 4, 5 and 6 are for Nuevo León with 2416 objects and 7, 8 and 9 are the instances for Jalisco, which has 3484 objects.

Table 4. Results for the geographical data.

Inst.	**H-TS-VNS**		**PAM**	
	Cost	**Time**	**Cost**	**Time**
1	**50.7595**	00:00:56	51.59754	00:02:03
2	**33.9236**	00:00:35	33.93228	00:15:33
3	23.7555	00:00:36	**23.338**	00:25:45
4	**210.9751**	00:08:05	211.2497	00:15:38
5	**139.5007**	00:04:21	140.1448	01:49:55
6	**94.5667**	00:02:15	**Didn't finish after 5 hours**	
7	**529.9198**	00:15:01	531.7125	00:24:28
8	**371.3131**	00:09:58	**Didn't finish after 5 hours**	
9	**243.5297**	00:05:51	**Didn't finish after 5 hours**	

We see in Table 4 that H-TS-VNS surpasses PAM in all the instances except for instance 3 (Querétaro with P = 48), however there's a big difference between the execution times. A peculiar aspect is that H-TS-VNS was run with the same input parameters for all the instances: nit = 1000, nit2 = 500, ip = 20 and tt = p/2 (see section 2.2.1.2 for more details) but its execution time considerably decreases.

This is because of our modified swap method where the candidate list is formed by the objects assigned to the selected medoid, therefore when the P is bigger the objects are more evenly distributed, generating smaller candidate lists for each medoid.

For example, for the map of Jalisco with P = 12 the algorithm finished in 15 minutes and for P = 48, in almost 6 minutes. Another aspect to consider is that the greedy method to generate the initial solution was only used for the map of Querétaro, because it required several minutes to generate the solution for the other maps, for this reason we used a randomly generated solution for Jalisco and Nuevo León but as we can see this did not negatively affect the quality of the final solution, as we still obtained better results than PAM.

These maps were generated with a custom Geographic Information Software developed in Java with the aid of GeoTools library, available for free in [1].

6. Conclusions

We observed that PAM surpassed the quality of all the other algorithms in the P-median instances from OR-Library. Even when we tried to give H-TS-VNS more time to find a solution, in many cases didn't manage to match the quality of PAM, not even do better than PAM in the instances where neither matched the best known result, however, as we saw in table 4 H-TS-VNS worked better than PAM, in quality and time, when working with geographical data, this is an issue that we need to examine in more detail to find the reasons for this behavior and make the necessary changes to achieve a consistent behavior with different kinds of data. For this we need to do more testing and debugging with our H-TS-VNS algorithm.

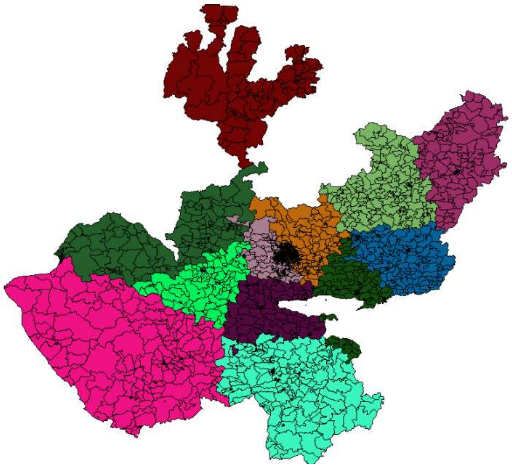

Fig. 3. Map of Jalisco with P = 12 returned by PAM. Cost: 531.7125.

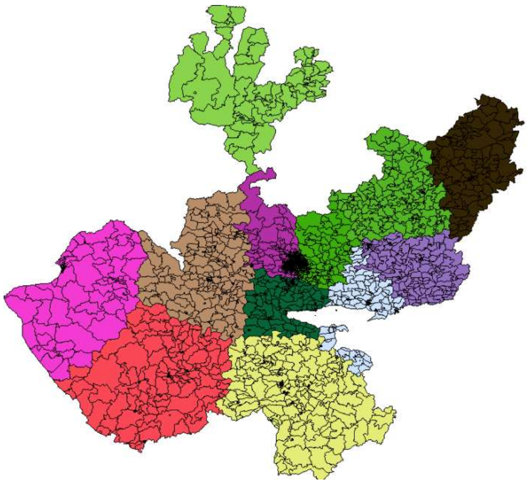

Fig. 4. Map of Jalisco with P = 12 returned by H-TS-VNS. Cost: 529.9198

The contribution of this article is to clarify that moreover than owning an equivalence the definition of zonas designed presented by Bação [19] and the classic definition of partitioning, it is also necessary to precise that the graphic solution generated by the model of the P-median answers to the fact that the P-median is a tool to cluster territories in problems of zones designs and in turn, the partitioning is a computational tool to solve DT problems.

Certainly, hierarchical clustering can be chosen instead of partitioning, situation that the authors will solve later. On the other hand, considering that the P-median model is integer-binary, for small instances, the commercial software solves the problem generating exact solutions.

However, for big problems, is insufficient the capacity of commercial software, then the inclusion of approximated methods is necessary to grant solutions close to the optimum. In this point, with an example, a proposal of transformation of the P-median model from an integer binary model to a combinatorial one has been presented, the challenge is to generalize the transformation.

References

1. Shirabe, T.: A model of contiguity for spatial unit allocation. Geographical Analysis, vol. 37, no. 1, pp. 2–16 (2005) doi: 10.1111/j.1538-4632.2005.00605.x
2. Macmillan, W.: Optimization modelling in GIS framework: the problem of political redistricting. In: Fotheringham, S., Rogerson, P. (eds.) Spatial analysis and GIS, pp. 221–246 (1994)
3. Kalcsics, J., Nickel, S., Schröder, M.: Towards a unified territorial design Approach-Applications, algorithms and GIS integration. Top, vol. 13, no. 1, pp. 1–56 (2005) doi: 10.10 07/bf02578982
4. Hess, S. W., Samuels, S. A.: Experiences with a sales districting model: Criteria and implementation. Management Science, Institute for Operations Research and the Management Sciences, vol. 18, no. 4-part-ii., pp. P-41-P-54 (1971) doi: 10.1287/ mnsc.18.4.p41
5. Zoltners, A. A., Sinha, P.: Sales territory alignment: A review and model. Management Science, vol. 29, no. 11, pp. 1237–1256 (1983) doi: 10.1287/mnsc.29.11.1237
6. Beasley, J. E.: Welcome to OR-Library (2022) www.brunel.ac.uk/~mastjjb/jeb/info.html
7. Reese, J.: Solution methods for thep-median problem: An annotated bibliography. Networks, vol. 48, no. 3, pp. 125–142 (2006) doi: 10.1002/net.20128
8. Church, R. L.: COBRA: A new formulation of the classic p-median location problem. Annals of Operations Research, vol. 122, no. 1-4, pp. 103–120 (2003) doi: 10.1023/ a:1026142406234
9. Church, R. L.: BEAMR: An exact and approximate model for the p-median problem. Computers and Operations Research, vol. 35, no. 2, pp. 417–426 (2008) doi: 10.1016/ j.cor.2006.03.006
10. Drezner, Z.: Facility location: A survey of applications and methods. Springer (1995)
11. Daskin, M. S.: Network and discrete location: Models, algorithms, and applications, Second Edition, Wiley (2013) doi: 10.1002/9781118537015
12. ReVelle, C. S., Swain, R. W.: Central facilities location. Geographical analysis, vol. 2, no. 1, pp. 30–42 (2010) doi: 10.1111/j.1538-4632.1970.tb00142.x
13. Kariv, O., Hakimi, S. L.: An algorithmic approach to network location problems. II: The p-medians. SIAM Journal of Applied Mathematics, vol. 37, no. 3, pp. 539–560 (1979)
14. Romero-Montoya, M., Granillo, E., González-Velázquez, R., Bernabé-Loranca, M. B., Estrada-Analco, M.: A hybrid VNS/TABU search algorithm for solution the p-median problem. International Journal of Combinatorial Optimization Problems and Informatics, vol. 11, no. 2, pp. 67–74 (2020) ijcopi.org/ojs/article/view/137
15. Bernábe-Loranca, M. B., González-Velázquez, R., Granillo-Martinez, E., Romero-Montoya, M., Barrera-Cámara, R. A.: P-median problem: A real case application. Advances in Intelligent Systems and Computing. Springer International Publishing, pp. 182–192 (2020) doi: 10.1007/978-3-030-49342-4_18
16. Bernábe-Loranca, M. B., Estrada-Analco, M., González-Velázquez, R., Martíne-Guzman, G., Ruiz-Vanoye: Location-allocation problem: A methodology with VNS metaheuristic. Advances in Intelligent Systems and Computing. Springer International Publishing, pp. 1015–1024 (2019) doi: 10.1007/978-3-030-16660-1_99

17. Beasley, J. E: OR-Library (2014) people.brunel.ac.uk/~mastjjb/jeb/orlib/pmedinfo.html
18. GeoTools: GIS utilities library for Java (2014) www.geotools.org/
19. Bação, F., Lobo, V., Painho, M: Applying genetic algorithms to zone design. Soft Computing, vol. 9, no. 5, pp. 341–348 (2004) doi: 10.1007/s00500-004-0413-4