Chapter 2
p-Median Problems

Alfredo Marín and Mercedes Pelegrín

Abstract One of the basic problems in the field of discrete location is the p-median problem. In this chapter we present and analyze several versions of the problem, but we can roughly define it as the choice of p facilities, among a set of n candidates, that minimize the cost of supplying a finite set of users. The p chosen facilities are usually called *medians*. Since the nature of the problem is combinatorial, integer programming is the common framework in which the problem is studied. Hence different formulations and their polyhedral properties constitute the kernel of this chapter. The study of the problem on a graph and heuristic procedures are treated in separate sections. Necessarily and unfortunately, we have to overlook many important references and results in the literature in the interest of legibility. Extensions of the problem, also of great interest, are covered in subsequent chapters and therefore are also ignored here. A companion problem of unquestionable importance, the Simple Plant Location Problem, is one of the main subjects of Chap. 4. Consequently, we have paid only little attention to it in our discussion.

2.1 Introduction

Discrete location problems consist of choosing a subset of locations, among a finite set of candidates, in which to establish facilities and then using these to satisfy the demand of a finite set of users. The choice of the locations must be made to minimize the sum of the fixed facility costs and of the cost of supplying the demand from the facilities.

Within this general framework, various problems can be identified as discrete location problems, most of which are studied and analyzed in this book. In this chapter we deal with a problem in the family of *median* problems. This term, in

A. Marín (✉) · M. Pelegrín
Universidad de Murcia, Facultad de Matemáticas, Campus de Espinardo, Murcia, Spain
e-mail: amarin@um.es; mariamercedes.pelegrin@um.es

G. Laporte et al. (eds.), *Location Science*,
https://doi.org/10.1007/978-3-030-32177-2_2

contrast with others like *center* and *equity*, refers to the definition of the cost to be minimized. When we speak about *median* (or *minisum*) problems we mean that the objective to be minimized depends in equal measure on the costs associated with each of the users.

The letter p in the term *p-median* refers to the number of locations to be chosen among the candidates, which is fixed beforehand. In other words, in the p-median problem a fixed number of p locations, usually called medians, must be chosen from the set of candidate facilities. Alternatively, it can be considered that p is the maximum number of locations that can be chosen. The cost to be minimized is calculated as the sum of the *allocation costs* of users to the medians. Let then $I = \{1, \ldots, m\}$ be the set of potential facilities and $J = \{1, \ldots, n\}$ the set of users to be supplied. The unit costs of supplying users from candidate facilities are arranged in a matrix $C = (c_{ij})$. We assume that *supplying costs* satisfy $c_{ij} \geq 0\ \forall i \in I, j \in J$. The demand of a user $j \in J$ is denoted with $d_j > 0$; then, the allocation cost of j to a median $i \in I$ is given by $d_j c_{ij}$. In order to obtain the lowest overall cost, each user will be assigned to the median with minimum allocation cost.

Now we can formally define the p-median problem as follows. Suppose a matrix $C = (c_{ij})$ with non-negative entries, m rows denoted by $I = \{1, \ldots, m\}$ and called candidates facilities, and n columns denoted by $J = \{1, \ldots, n\}$ and called users. Given an n-dimensional vector (d_j) with positive entries and given $p \in \mathbb{Z}$, $1 \leq p \leq m - 1$, choose a subset $P \subseteq I$ of p rows of C in such a way that the *total cost* defined by $\sum_{j \in J} \min_{i \in P}\{d_j c_{ij}\}$ is minimized.

Figure 2.1 shows several examples of optimal solutions to p-median problems. Here $I = J$ is given by the same set of $n = 30$ points on the plane. Costs c_{ij} are given by the Euclidean distances between points and demands are assumed to be equal to one. In Fig. 2.1a we have taken $p = 2$ and drawn the best choice of 2 facilities (represented with squares) and the allocation of the 30 points to the corresponding closest facility. Different optimal solutions for $p = 3$, 4 and 5 are given also in Fig. 2.1b, c, and d, respectively.

Note that the kernel of the problem is the choice of the p facilities among the m candidates (a purely combinatorial subject, with $\binom{m}{p}$ possible solutions). Customers allocation to the facilities is trivially carried out by choosing, for each user $j \in J$, the facility in P with minimum allocation cost.

The p-median problem is strongly related with a problem that will be studied in Chap. 4, the Simple Plant Location Problem (SPLP)—also called Uncapacitated Facility Location Problem. In the SPLP, the number of facilities is not fixed *a priori*. Instead, a cost associated to each of the candidates is given, usually represented by $f_i \geq 0\ \forall i \in I$. Then, given $C = (c_{ij})$ with non-negative entries, (d_j) with positive entries, and given the vector of non-negative costs $f = (f_i)$, SPLP aims to choose a subset $P \subseteq I$ of rows of C in such a way that $\sum_{i \in P} f_i + \sum_{j \in J} \min_{i \in P}\{d_j c_{ij}\}$ is minimized. SPLP is also a *minisum* problem, with a trade-off between costs associated to the facilities and allocation costs.

Despite its apparent simplicity, the p-median problem is NP-hard (Kariv and Hakimi 1979). Its origins can be traced back to Hakimi (1964, 1965), where the problem was defined on a graph, and ReVelle and Swain (1970), where an

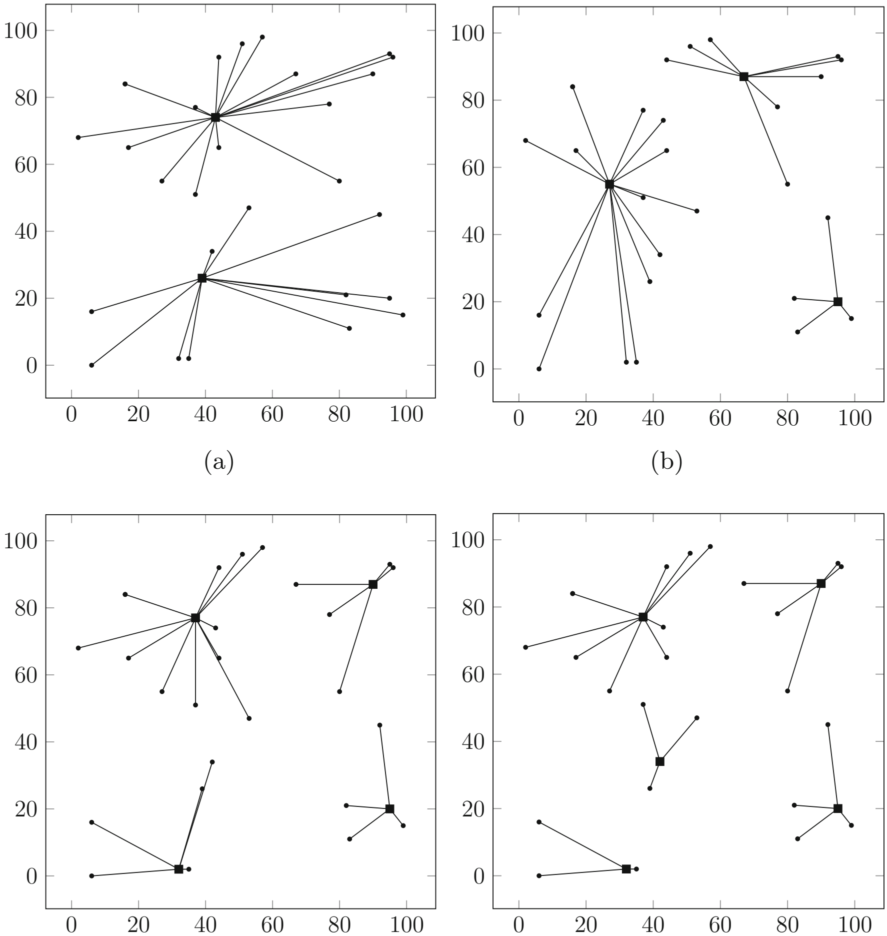

Fig. 2.1 Optimal solutions to the same instance of the p-median problem for different values of p. (**a**) $p = 2$. (**b**) $p = 3$. (**c**) $p = 4$. (**d**) $p = 5$

integer linear programming (ILP) formulation was proposed, inspired in Balinski (1965). Other related seminal papers are Hua et al. (1962), Kuehn and Hamburger (1963) and Manne (1964). Given its combinatorial nature, (mixed) integer linear programming (Nemhauser and Wolsey 1988; Wolsey 1998) has usually been the approach used to formulate and optimally solve the problem. The literature on the p-median problem is vast and it is not our aim to give an exhaustive list of papers. We focus our attention on recent results and suggest consulting Mirchandani (1990) and Reese (2006) as additional information sources.

We have organized the rest of the chapter as follows. In Sect. 2.2 several non-immediate applications, that show a wide range of possibilities of use, are presented.

In Sect. 2.3 the first integer linear programming formulations of the problem are introduced and analyzed. Section 2.4 deals with some of the most interesting available solution methods. Valid inequalities and facets for the polyhedra defined by the linear relaxations of different formulations are described in Sect. 2.5. We have included in a separate Sect. 2.6 the formulations and polyhedral results that arise when the p-median problem is solved on a (possibly non-complete) directed graph. Since solving large instances of the p-median problem is a difficult task, the literature on heuristic approaches is vast, and we try to give an idea of this vastness in Sect. 2.7, before closing the chapter with some final considerations.

2.2 Applications

In this section we present some applications of the p-median model taken from the literature. To emphasize its wide range of possibilities, we have selected applications outside the field of location of warehouses, plants, shelters or other kind of facilities, which is the natural interpretation of our problem.

Clustering was one of the first applications of the p-median problem. In the paper by Vinod (1969) it is said that *a large number of objects, persons, variables, symbols, etc. have been often to be grouped into a smaller number of mutually exclusive groups so that members within a group are similar to each other in some sense*. There is a limited number of groups, each of them having a distinguished member called centroid. The fitness of the partition depends on the average similarity of each object with the centroid of its group. The similarity between two pairs can be calculated from the input data and would correspond with costs $(d_j c_{ij})$ in our problem. The number of groups or clusters would be p and the centroids would be our medians.

Another application of the p-median problem, as presented in Vigneron et al. (2000), is the optimal placement of cache proxies in a computer network (see also Li et al. 1998). Nodes in a rooted tree network request a service that follows the path from the node to the root. When a proxy, located at a node of the tree, is found along this path, it satisfies the request. The location of p proxies in the nodes of the network in such a way that the sum of the distances from the nodes to the closest proxy in the corresponding path is minimized can be seen as a p-median problem. Vigneron et al. (2000) developed an algorithm to solve it on this special tree network topology.

We also include in this review of applications the so-called *Optimal Diversity Management Problem* (see Briant and Naddef 2004). Assume that a factory will manufacture a product that can, to some extent, be customized. For example, a car with t different improvements to be chosen or not by the users. The car becomes better and more expensive with each of these improvements, and then the users will not complain if they receive a car with more extras than required, at the same price. The factory cannot produce the 2^t different vehicles, so they decide to produce only p of the combinations and to deliver to each user the car with minimum cost among those that include all the extras the user asked for. In this p-median problem, p is the number of different versions of the product that the factory can produce and $I = J$ is the set of all possible combinations of extras. Medians are the versions of the product that will be finally produced, and a combination of extras will be assigned to the median that will replace it when serving user requests. Replacing user request j by the version of the product i has a cost $d_j c_{ij}$.

A similar application is to determine p times for public vehicle departures on a temporal line, aiming at maximizing the total satisfaction of users. This served as the base for addressing the *Transit Network Timetabling and Scheduling Problem* in Mesa et al. (2014). In a public transit line, each vehicle performs a number of line runs or expeditions that have to be located in time. Users of the transit corridor have to be allocated to the line run that better fits their preferences, while fulfilling some capacity requirements. The formulation in Mesa et al. (2014) is a more complex version of the classical p-median that includes additional constraints.

Finally, in Goldengorin et al. (2012) (see also AlBdaiwi et al. 2011) the *cell formation* problem is established and studied as a p-median problem. A set of machines and their dissimilarities $d_j c_{ij}$ are given. It can be considered, for example, that when two machines process almost the same set of parts, there is a small dissimilarity between them (and can take part or the same cell). The problem is then to find p machines that are best representatives of p manufacturing cells, that is to say, the sum over the cells of the dissimilarities between these representatives and all other machines belonging to the same the cell has to be minimum. The problem can be considered as a special p-median problem on a graph, as defined in Sect. 2.6 below.

2.3 Integer Programming Formulations for the *p*-Median Problem

The *classical* ILP formulation for the p-median problem is

$$\text{(F1) minimize} \sum_{i \in I} \sum_{j \in J} d_j c_{ij} x_{ij} \tag{2.1}$$

$$\text{subject to} \quad \sum_{i \in I} x_{ij} = 1 \quad \forall j \in J \tag{2.2}$$

$$x_{ij} \le y_i \quad \forall i \in I, j \in J \tag{2.3}$$

$$\sum_{i \in I} y_i = p \tag{2.4}$$

$$x_{ij} \in \{0, 1\} \quad \forall i \in I, j \in J \tag{2.5}$$

$$y_i \in \{0, 1\} \quad \forall i \in I. \tag{2.6}$$

Two sets of binary variables are used. On the one hand,

$$y_i = \begin{cases} 1 & \text{if candidate location } i \text{ is chosen as a median,} \\ 0 & \text{otherwise,} \end{cases} \quad \forall i \in I.$$

These variables are sometimes called *location variables*. Constraint (2.4) ensures that p candidate locations are chosen as facilities. Note that $y_i = 1$ when $i \in P$. On the other hand,

$$x_{ij} = \begin{cases} 1 & \text{if user } j \text{ is supplied from candidate facility } i, \\ 0 & \text{otherwise,} \end{cases} \quad \forall i \in I, j \in J.$$

The variables in this second set are sometimes called *allocation variables*. Constraints (2.2) guarantee that each user $j \in J$ is allocated to (supplied from) some candidate location $i \in I$. And constraints (2.3) prohibit allocations to candidate locations that were not chosen as facilities: when $y_i = 0$ (i.e., $i \notin P$), $x_{ij} = 0$ $\forall j \in J$, i.e., no user can be assigned to the location.

Allocation variables also serve to select the individual allocation costs that the solution entails and that are used to compute the total cost in linear combination (2.1).

Formulation (F1) contains $nm + m$ binary variables and $n + nm + 1$ linear constraints. A reduced formulation can be produced by replacing the set of nm constraints (2.3) by a set with only m constraints in the form

$$\sum_{j \in J} x_{ij} \leq n y_i \quad \forall i \in I. \tag{2.7}$$

Note that the effect of (2.7) when $y_i = 0$ is the same of (2.3), fixing to zero x_{ij} for all $j \in J$. In the case $y_i = 1$, the sum of n binary variables will be upperly bounded by n, thus producing no effect. We call (F2) formulation (F1) where constraints (2.3) have been replaced by (2.7).

Although formulation (F2) is more compact than formulation (F1), it has obvious disadvantages when a branch-and-bound procedure is used to solve the p-median problem, since summing up (2.3) for all $j \in J$, constraints (2.7) directly follow. This means that the polytope defined by the constraints of (F1) after relaxing the integrity of the variables, is included in the polytope analogously defined by the constraints of (F2). The consequence is that the lower bounds produced by (F1) will be better than those produced by (F2).

Several ways of reducing the size of (F1) without loss of quality in the formulation have been explored. First, it can be observed (see e.g. Church 2003) that a user will never be supplied from a facility if there are at least $m - p + 1$ candidates with strictly less associated supplying cost. To formalize this, we sort, for each user $j \in J$, the corresponding column in the cost matrix C to obtain $\hat{c}_{1j} \leq \hat{c}_{2j} \leq \cdots \leq \hat{c}_{mj}$. Then, some x-variables can be fixed to zero: $x_{ij} := 0$

$\forall i \in I : c_{ij} > \hat{c}_{m-p+1,j}$. Another possibility, see Church (2003), is to identify and match equivalent x-variables in the formulation. Consider two users $j_1 < j_2 \in J$, a candidate $i \in I$ and a scenario where $\Omega := \{\ell \in I : c_{\ell j_1} < c_{ij_1}\} = \{\ell \in I : c_{\ell j_2} < c_{ij_2}\}$. If, in an optimal solution, $x_{ij_1} = 1$, it follows that no candidate in Ω has been chosen as a facility, but i has been (since j_1 was assigned to i). Then, one facility to which j_2 is allocated with minimum cost is i as well. Consequently, $x_{ij_2} = 1$ is an optimal choice. On the other hand, $x_{ij_1} = 0$ means that either a candidate in Ω has been chosen as median or there are no medians in Ω and neither is i a median. In both cases, $x_{ij_2} = 0$. The conclusion is that x_{ij_1} and x_{ij_2} can be identified, and thus the size of the formulation can be reduced by replacing all x_{ij_2} with x_{ij_1}.

Following the same reasoning as in Cho et al. (1983a), we can handle formulation (F1) to rewrite constraints (2.2) and (2.3) in a different way. Note that, since (2.2) are equalities, the sums $\sum_{i \in I} x_{ij}\ \forall j \in J$ will be constant in any feasible solution to (F1). Hence using a large enough number, M, the alternative objective

$$\sum_{i\in I}\sum_{j\in J} d_j c_{ij} x_{ij} - \sum_{i\in I}\sum_{j\in J} M x_{ij} = \sum_{i\in I}\sum_{j\in J} \tilde{c}_{ij} x_{ij}$$

where $\tilde{c}_{ij} := d_j c_{ij} - M < 0\ \forall i \in I, j \in J$ can be utilized. The advantage of this function is that, since the coefficients are negative and we are minimizing, the x-variables will take value one in an optimal solution if they are not restricted by the constraints of the formulation. This means that constraints (2.2) can be relaxed to

$$\sum_{i\in I} x_{ij} \le 1 \qquad \forall j \in J. \tag{2.8}$$

Consider now a different set of binary variables

$$y'_i = \begin{cases} 1 & \text{if candidate location } i \text{ is } \textbf{not} \text{ chosen as a facility,} \\ 0 & \text{otherwise,} \end{cases} \quad \forall i \in I,$$

that is to say, $y'_i := 1 - y_i\ \forall i \in I$. Using this new set of variables, constraints (2.3) can be rewritten as

$$x_{ij} + y'_i \le 1 \qquad \forall i \in I, j \in J. \tag{2.9}$$

Both sets of constraints, (2.8) and (2.9), are defined as sums of binary variables upperly bounded by 1. These *set packing* constraints can be analyzed, see Cánovas et al. (2000, 2002, 2003), Cho et al. (1983a,b), and Cornuéjols and Thizy (1982), to produce a tighter formulation, using the so-called *intersection* (or *conflict*) graph, where each node is associated with a variable, and nodes are neighbors if they share at least one constraint. Since this analysis is the same as that carried out for the SPLP, we refer the reader to Chap. 4 for a detailed analysis. The reformulation of

(F1) by means of (2.8) and (2.9) still contains constraint (2.4), which enables us to perform the polyhedral analysis of the formulation in a different way, see Sect. 2.5.

A different relaxation of (F1) can be carried out, that of the integrity of the x-variables. Constraints (2.5) can be replaced by

$$x_{ij} \geq 0 \qquad \forall i \in I, j \in J. \tag{2.10}$$

To see this, observe that (2.2) and (2.10) imply $x_{ij} \in [0, 1]$ $\forall i \in I, j \in J$. Now, consider a set $P \subseteq I$ of p facilities and the sets $A_j := \{i \in P : c_{ij} = \min_{\ell \in P} c_{\ell j}\}$. It is obvious that in any optimal solution where P is the set of chosen facilities, $\sum_{i \in A_j} x_{ij} = 1$ holds for all $j \in J$. Since all variables in the last sum have the same cost, an equivalent integer solution can be trivially obtained by fixing one of them to one and the rest to zero. After relaxing (2.5)–(2.10), the meaning of the x-variables can be re-established as x_{ij} = *fraction of the demand of user j that is supplied by facility i*.

Consider now the version of the problem where $I = J$ and $c_{ii} = 0$ $\forall i \in I$. This case has some special characteristics that allow to reformulate the problem. Whenever $y_i = 1$, the minimum possible allocation cost for point i will be 0, obtained by allocating i to itself. Then $y_i = 1 \Rightarrow x_{ii} = 1$. Since $y_i = 0 \Rightarrow x_{ii} = 0$, both variables can be identified, and y_i can be replaced by x_{ii} in the formulation. The resulting reduced formulation is given by

$$\text{(F3) minimize} \sum_{i \in I} \sum_{\substack{j \in I: \\ i \neq j}} d_j c_{ij} x_{ij}$$

$$\text{subject to} \qquad (2.2)$$

$$x_{ij} \leq x_{ii} \qquad \forall i, j \in I : i \neq j \tag{2.11}$$

$$\sum_{i \in I} x_{ii} = p \tag{2.12}$$

$$x_{ij} \in \{0, 1\} \quad \forall i, j \in I : i \neq j \tag{2.13}$$

$$x_{ii} \in \{0, 1\} \quad \forall i \in I. \tag{2.14}$$

Again, constraints (2.13) can be relaxed to $x_{ij} \geq 0$ $\forall i, j \in I$: $i \neq j$.

Under the given conditions, constraints (2.7) can be slightly improved. Note that the existence of p users that are going to be self-allocated guarantees that no more than $n - p + 1$ users will be allocated to the same facility. Hence the constant in the right hand side of (2.7) can be modified to yield the tighter constraints

$$\sum_{\substack{j \in J: \\ j \neq i}} x_{ij} \leq (n - p) x_{ii} \quad \forall i \in I. \tag{2.15}$$

In what follows we still assume $c_{ii} = 0\ \forall i \in I$ and $c_{ij} \geq 0\ \forall i, j \in I\colon i \neq j$ to produce a formulation based on a completely different set of variables. The ideas we are going to present come from Cornuéjols et al. (1980), where they were applied to the SPLP. A preprocessing of the data is required before proceeding. It is necessary, for each $j \in J$, to sort the entries of the j-th column of the cost matrix C, removing the multiplicities: $0 = \bar{c}_{1j} < \bar{c}_{2j} < \cdots < \bar{c}_{G_j j} = \max_{i \in I} c_{ij}$. Since we do not know *a priori* how many different supplying costs there are in column j of C, we use G_j to denote this number. A new set of binary variables, sometimes called *cumulative variables*, is defined as

$$z_{kj} = \begin{cases} 1 \text{ if the supplying cost of user } j \text{ is at least } \bar{c}_{kj} \\ \quad \text{(no matter which facility it is allocated to),} & \forall j \in J, 2 \leq k \leq G_j. \\ 0 \text{ otherwise,} \end{cases}$$

Note that the variables z_{1j} have not been used, since by definition $z_{1j} = 1$ if the supplying cost of user j is at least $\bar{c}_{1j} = 0$, and this condition is always satisfied. Initially we will also use variables y_i, $\forall i \in I$, to keep track of the chosen facilities. Then consider a new formulation for the p-median problem given by

$$\text{(F4) minimize} \sum_{j \in J} \sum_{k=2}^{G_j} d_j (\bar{c}_{kj} - \bar{c}_{k-1,j}) z_{kj} \tag{2.16}$$

$$\text{subject to} \qquad (2.4),\ (2.6)$$

$$z_{kj} + \sum_{\substack{i \in I: \\ c_{ij} < \bar{c}_{kj}}} y_i \geq 1 \qquad \forall j \in J, 2 \leq k \leq G_j \tag{2.17}$$

$$z_{kj} \in \{0, 1\} \qquad \forall j \in J, 2 \leq k \leq G_j. \tag{2.18}$$

In formulation (F4) we keep constraints (2.4) and (2.6) to account for the number of facilities. The difference between (F4) and the previously introduced formulations is that in (F4) there is no information in the variables about the allocation of users to facilities, but there is about the smallest allocation costs of the users when only chosen facilities are considered. Let us analyze constraints (2.17). The term $\sum_{i \in I\colon c_{ij} < \bar{c}_{kj}} y_i$ takes value zero only when no candidate with supplying cost less than $\bar{c}_{kj}$ (the k-th supplying cost for user j) has been selected as a facility. It is clear, then, that z_{kj}, as defined, must take value 1. Since the coefficients in the objective function (2.16) are strictly positive, in an optimal solution all variables will take value 0 unless the corresponding constraint (2.17) force them to take value 1. For this reason, z-variables can be relaxed to be positive continuous variables and constraints (2.18) can be simply removed.

For a given user $j \in J$, the sets of candidates inside a *radius* $\bar{c}_{kj}$, $B_k := \{i \in I : c_{ij} < \bar{c}_{kj}\}$, satisfy the strict inclusion relations $B_2 \subsetneq B_3 \subsetneq \cdots \subsetneq B_{G_j}$. This implies that, in any optimal solution, $z_{2j} \geq z_{3j} \geq \cdots \geq z_{G_j j}$, that is to say, the appearance of vector $z_{\cdot j}$ will be $(1, \ldots, 1, 0, \ldots, 0)$. Assume the last 1 corresponds with variable z_{aj}. Then, in the objective function (2.16) the sum $\sum_{k=2}^{G_j} d_j(\bar{c}_{kj} - \bar{c}_{k-1,j})z_{kj}$ will be $\sum_{k=2}^{a} d_j(\bar{c}_{kj} - \bar{c}_{k-1,j})$. Taking into account that $\bar{c}_{1j} = 0$, the value of this telescopic sum will be $d_j\bar{c}_{aj}$, that is to say, the cost of allocating j to median a, as wished.

In Fig. 2.2 we see, using the same example as in Fig. 2.1d (where $I = J$, $d_j = 1$ $\forall j \in J$, and supplying costs are given by Euclidean distances between points), the effect of constraints (2.17) on user $j = 1$ assuming that the facilities of the optimal solution are given. Constraint (2.17), with $k = 2$, reads $z_{21} + y_1 \geq 1$. Since 1 is not a median, it follows that $z_{21} = 1$. Taking now $k = 3$, it reads $z_{31} + y_1 + y_2 \geq 1$, implying $z_{31} = 1$. Similarly, $z_{41} = z_{51} = 1$. Then, for $k = 6$, $z_{61} + y_1 + y_2 + y_3 + y_4 + y_5 \geq 1$ is satisfied since $y_5 = 1$. Due to the objective function, $z_{61} = z_{71} = \cdots = 0$, and that the cost of allocating point 1 to point 5 will be $(10.77-0)\cdot 1+(15.65-10.77)\cdot 1+(16.49-15.65)\cdot 1+(17.72-16.49)\cdot 1 = 16.49$, the distance between points 1 and 5.

A reduction in the size of (F4) can be made noting that constraints (2.17) when $k = 2$ read $z_{2j} + y_j \geq 1$ and these constraints are always satisfied as equalities by an optimal solution. Then y_i can be replaced with $1 - z_{2i}$ $\forall i \in I$.

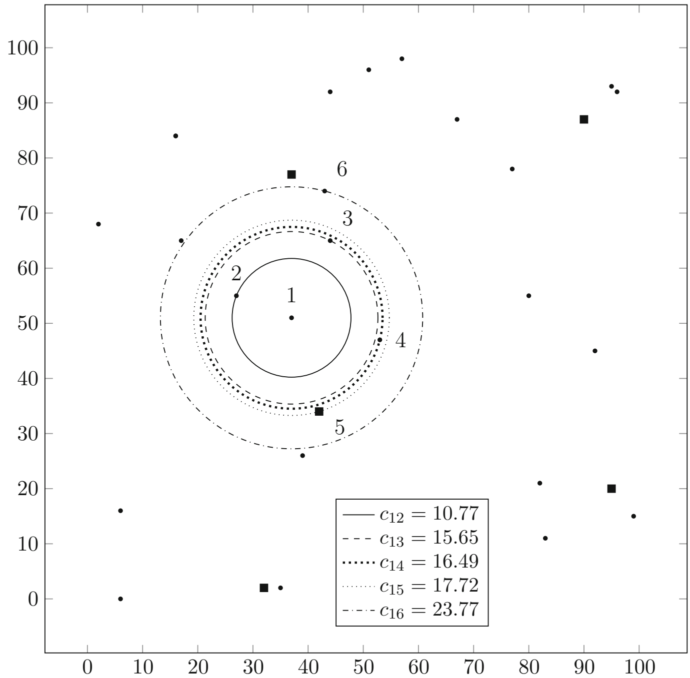

Fig. 2.2 Graphical representation of the role of the z-variables in formulation (F4) on the same example as in Fig. 2.1d

Regarding the size of (F4), observe that for each user $j \in J$, the number of z-variables in (F4) will be the number of different costs in the j-th column of C, minus one. Therefore, the total number of z-variables in the formulation will be in the set $\{0, \ldots, nm\}$. For each z-variable there is one constraint in family (2.17), thus the number of linear constraints will be in $\{1, \ldots, nm + 1\}$. In the worst case, when all costs in each column of C are distinct, the size of (F4) will be exactly the same as the size of (F1).

Although the size of (F4) can be smaller than the size of (F1), Cornuéjols et al. (1980) proved that both linear relaxations yield the same lower bound on the optimal value of the problem. There exist many works where formulations (F1)–(F3) have been used. However, references containing formulation (F4) are scarce, and almost limited to the study of the companion problem SPLP: Kolen (1983) used a version of formulation (F4) to solve the SPLP in polynomial time on a tree; Simão and Thizy (1989) studied the linear relaxation of a modification of (F4); (F4) for SPLP was also considered in Cornuéjols et al. (1990) and Kolen and Tamir (1990). Finally, Xu and Lowe (1993) compared the work of Simão and Thizy (1989) with a previous method in the literature to solve the SPLP.

2.4 Optimal Solution Procedures

Several exact algorithms for the p-median problem are available. We summarize some of them here, without intending to be exhaustive.

Galvão (1980) realized that solving the p-median problem within a branch-and-bound framework means solving many linear relaxations of subproblems of large size. He then devised a method to efficiently obtain good lower bounds instead of optimally solving the relaxed continuous subproblems. To this end, he considered formulation (F3), replaced the equality (2.2) by '$\geq$', relaxed constraints (2.13) and (2.14) and built the dual problem

$$
\begin{aligned}
\text{(F3D) maximize} \quad & p\sigma_{n+1} + \sum_{i \in I} \sigma_i \\
\text{subject to} \quad & \sigma_i + \sigma_{n+1} - \sum_{\substack{j \in I: \\ j \neq i}} \pi_{ij} \leq 0 \quad \forall i \in I \\
& \sigma_j - \pi_{ij} \leq d_j c_{ij} \qquad \forall i, j \in I \\
& \pi_{ij} \leq 0 \qquad \forall i, j \in I : i \neq j \\
& \sigma_i \geq 0 \qquad \forall i \in I \\
& \sigma_{n+1} \leq 0.
\end{aligned}
$$

Table 2.1 A summary of the computational experience on exact solution methods up to date

Authors	Year	Computer	n	t (s)
Galvão	1980	Unknown	30	879
Church	2003	Sun Ultra Sparc 10	372	879
Avella et al.	2007	Compaq EVO W4000 PC Pentium IV 1.8 GHz, 1 GB RAM	5535	394
García et al.	2011	Intel CORE 2 CPU 6600 2.4 GHz, 3 GB RAM	85,900	66,000

The last three columns stand for the maximum size and time in seconds of the instances tested but do not necessarily correspond to one same instance

Noticing then that, in any optimal solution to (F3D),

$$\sigma_{n+1} \leq \min_{i\in I}\{-\sigma_i + \sum_{\substack{j\in I:\\ j\neq i}} \pi_{ij}\} \text{ and } \pi_{ij} = -\max\{0, \sigma_j - d_j c_{ij}\}\ \forall i, j \in I : i \neq j,$$

he designed a two-phase method to calculate good feasible solutions of (F3D) in an attempt to increase the objective value. In the first phase the value of σ_{n+1} was maximized and then the values of σ_i, $i \in I$, were maximized without modifying σ_{n+1}. Then he embedded this procedure, which produces good lower bounds in a short time, into the branch-and-bound algorithm and obtained good computational results. Table 2.1 gives an insight about the evolution of the sizes of the instances that could be solved with each exact method. Note that the best lower bound that can be produced with this approach is the one provided by the linear relaxation of (F3).

The use of formulations (F1) and (F3) with aggregated but weaker constraints (2.7) or (2.15), combined with the inclusion of (2.3) as valid inequalities, has served as an alternative strategy in several papers. As an example, in Church (2003) a subset of constraints (2.3), those corresponding to the candidates with minimum supplying cost with respect to each user, is initially incorporated in formulation (F3). The combination of this strategy and the matching of equivalent x-variables (see Sect. 2.3) also produced good computational results (see Table 2.1).

Beltrán et al. (2006) approached the p-median problem from a similar point of view. They initially considered formulation (F1) and the Lagrangian relaxation of constraints (2.2) and (2.4) by means of unrestricted multipliers v_j, $\forall j \in J$ and v_0, respectively. An overview on Lagrangian relaxation can be consulted in Guignard (2003). The advantage of relaxing equality constraints is that any optimal solution to a relaxed subproblem that also satisfies the relaxed constraints is an optimal solution of the primal problem. The disadvantage of relaxing all these constraints is that the optimal value of the dual problem is the same as the optimal value of the linear relaxation of the problem. The authors found first a good set of Lagrangian multipliers and used them as a starting point for a second problem relaxed in a Lagrangian fashion. In this case they added constraints $\sum_{i\in I} x_{ij} \leq 1, \forall j \in J$, to the relaxed subproblem, which becomes more difficult to solve but can yield better lower bounds. The advantage of using the '$\leq$' version of the constraints is that all variables x_{ij} with non-negative coefficient $d_j c_{ij} + v_j$ in the relaxed subproblem can be fixed to zero. The subproblem is then easier to solve and can even be decomposed, since the non-removed variables could be grouped in subsets that do not relate each other. The final set of multipliers is then used as the starting point for a third and last relaxation obtained by adding one more constraint to the subproblem, namely $\sum_{i \in I} y_i \leq p$.

Avella et al. (2007) designed a branch-and-cut-and-price algorithm that was able to solve very large instances (see Table 2.1) of the p-median problem on a graph (see forthcoming formulation (F5)). Cuts were added based on new valid inequalities called $W - q$, *lifted odd hole* and *cycle inequalities*. Details of them are given in Sect. 2.6. Pricing was carried out by solving a master problem to optimality and using dual variables to price out the variables of the initial problem that were not considered in the master, adding new variables if necessary. The novelty of the approach was that constraints (2.20) were also relaxed and incorporated to the master problem when the corresponding column was. The authors also developed criteria to fix the values of some y-variables to zero when lower bounds calculated fixing y_i to one were greater than previously known upper bounds.

Finally, we summarize the solution method based on (F4) developed in García et al. (2011). Recall that, in (F4), given an optimal solution (y^*, z^*) and a fixed user $j \in J$, $z^*_{\cdot j}$ will have the shape $(1, \ldots, 1, 0, \ldots, 0)$. We have also a similar property of any optimal solution of the linear relaxation of (F4), $(\bar{y}, \bar{z})$: for all $j \in J$, $\bar{z}_{2j} \geq \bar{z}_{3j} \geq \ldots \geq \bar{z}_{G_j j}$. Therefore, if $\bar{z}_{aj} = 0$ for some a, then $\bar{z}_{kj} = 0$ for all $k > a$. Suppose we could know this optimal solution $(\bar{y}, \bar{z})$ beforehand. Since each z-variable only appears in one constraint, and the z-variables taking value zero have not been forced by the optimal values of the y-variables to take value 1, we could remove all variables and constraints associated with the null $\bar{z}$-values and the linear relaxation of this reduced formulation would provide us with the same optimal solution. Conversely, let us remove variables $z_{a+1,j}, \ldots, z_{G_j j}$, for a given $j \in J$, from the linear relaxation of (F4). If $\bar{z}_{aj} = 0$ in the optimal solution of the relaxed problem, this is done. Otherwise, if $\bar{z}_{aj} > 0$, it is possible that some of the removed variables had taken a positive value in the optimal solution. In this case, a has been wrongly selected and a larger value for it must be considered. The method proposed in García et al. (2011) then considered a first formulation with a very small set of z-variables and constraints, and added more variables and their corresponding constraints when needed. At every node of the branching tree, the final formulation of the predecessor node was used. The result was an exact branch-and-cut-and-price method that allowed the authors to solve the p-median problem with a drastically reduced formulation that required much fewer constraints and variables than formulations (F1)–(F4). This method performed extremely well on very large instances (see Table 2.1) with large values of p. Note that the larger the value of p, the smaller the allocation costs associated to the users and, consequently, the smaller the number of z-variables (and constraints) added to the initial reduced formulation.

2.5 Polyhedral Properties

In this section we present polyhedral properties of the formulations (F1) and (F3) or their modifications. It is worth mentioning that since the polyhedron of these p-median formulations is obtained from the polyhedron of the SPLP by adding only one constraint, all valid inequalities for the corresponding formulations of the SPLP are also valid for the p-median problem. Nevertheless, they do not usually define facets. In this section we focus on models that produce valid inequalities or facets for the p-median problem that are not necessarily valid for the SPLP. Basic knowledge on polyhedral theory is assumed in this section (we refer the interested reader to Nemhauser and Wolsey 1988)

A seminal paper in this field is de Farias (2001). The author considered a modified version of formulation (F1), with equalities (2.2) and (2.4) relaxed to inequalities of type '≤'. He proved that the polyhedron so defined is fully dimensional, and found a family of facets by taking a subset J' of J with cardinality at least $p+1$ and disjoint nonempty subsets of I named I_j, $j \in J'$, with $\cup_{j\in J'} I_j \subsetneq I$. He showed that the constraints

$$\sum_{j\in J'}\sum_{i\in I_j} x_{ij} + \sum_{i\notin \underset{\ell\in J'}{\cup} I_\ell}\ \sum_{j\in J'} x_{ij} \le p + (|J'| - p) \sum_{i\notin \underset{\ell\in J'}{\cup} I_\ell} y_i$$

are valid for the given formulation and define facets. We now present an example taken from de Farias (2001) with $n = 3$, $m = 4$, $p = 2$, $J' = J$, $I_j = \{j\}$, $j = 1, 2, 3$:

$$x_{11} + x_{22} + x_{33} + x_{41} + x_{42} + x_{43} \le 2 + y_4.$$

Note that $y_4 = 0$ implies $x_{41} + x_{42} + x_{43} = 0$ and then $x_{11} + x_{22} + x_{33} \le 2$ is valid since $p = 2$. On the other hand, in the case $y_4 = 1$, the inequality becomes trivial.

Consider now de Vries et al. (2003). Among different results on the polyhedral structure of the p-median problem, the authors generate a family of valid inequalities for (F3) of the form

$$\sum_{i\in R\cup S} x_{ii} - \frac{1}{r-p}\sum_{i\in R}\sum_{\substack{j\in R:\\ i\neq j}} x_{ij} - \frac{1}{r-p+1}\sum_{i\in S}\sum_{j\in R} x_{ij} \le p-1, \tag{2.19}$$

where R is a subset of $I = J$ of cardinality $r \ge p$, and S is a subset of $I \setminus R$. For example, take $m \ge 4$, $p = 2$, $R = \{1, 2, 3\}$ and $S = \{4\}$. The facet in family (2.19) would be

$$\begin{aligned} &2x_{11} + 2x_{22} + 2x_{33} + 2x_{44} \\ &\le 2x_{12} + 2x_{21} + 2x_{13} + 2x_{31} + 2x_{23} + 2x_{32} + x_{41} + x_{42} + x_{43} + 2. \end{aligned}$$

Observe that when all medians belong to the set $\{1, 2, 3\}$. To illustrate, assume that the two medians are 1 and 2. Then $2x_{11}+2x_{22}+2x_{33}+2x_{44} = 4$ and the inequality becomes $1 \leq x_{13} + x_{23}$, and it obviously holds. A second possibility is that the two medians are 1 and 4. Then it follows that $2 \leq 2x_{12} + 2x_{13} + x_{42} + x_{43}$. Since 2 and 3 must be supplied from 1 or 4, it also holds. Finally, if $x_{11} + x_{22} + x_{33} + x_{44} \leq 1$, the inequality becomes trivial. In de Vries et al. (2003) it is proven that inequalities (2.19) define facets when $r > p$, $S \neq \emptyset$ and $S \cup R \neq I$.

In Zhao and Posner (2011), a generalization of the family of facets (2.19) is developed. Here, a partition of I given by the sets $T_1,\ldots,T_r$, S and Q, with $r > p$ and $T_i \neq \emptyset, i = 1, \ldots, r$, $Q \neq \emptyset$, is required. Defining $T = \cup_{i=1}^{r} T_i$, $R \subseteq T \cup Q$ of cardinality r such that $|R \cap T_i| \leq 1, i = 1, \ldots, r$ and a bijection τ of R in the set $\{1, \ldots, r\}$, the new family of valid inequalities for (F3) is given by

$$\sum_{i \in T \cup S} x_{ii} - \frac{1}{r-p} \sum_{j \in R} \sum_{i \in T \setminus T_{\tau(j)}} x_{ij} - \frac{1}{r-p+1} \sum_{i \in S} \sum_{j \in R} x_{ij} \leq p-1.$$

These inequalities define facets when $2 \leq p < r$ and $|Q| = 1$ or $|(T \cup S) \setminus R| \geq 1$. The authors also devised a heuristic procedure to separate these inequalities.

Also observe that Cánovas et al. (2007) introduce *dominance constraints* in the shape of $x_{ij_1} \leq x_{ij_2}$ that can be incorporated to formulation (F3). These inequalities can be used whenever $\{\ell \in I : c_{\ell j_2} < c_{ij_2}\} \subseteq \{\ell \in I : c_{\ell j_1} < c_{ij_1}\}$. We present additional polyhedral material after introducing a new version of the problem, in the next section.

2.6 *p*-Median Problem on a Graph and Additional Polyhedral Results

Many authors consider and analyze a particular case of the p-median problem defined on a directed graph (V, A). The set of nodes, V, represents users and also candidate locations for facilities. The set of arcs A, is used to express the possible allocations of users to facilities. Self-allocation is implicitly assumed or, in other words, a node is either chosen as a median or it must be allocated to another node. Note that this is equivalent to fixing some variables x_{ij} to zero in formulation (F3): $x_{ij} = 0$ if $(i, j) \notin A$. The same effect can be achieved by taking c_{ij} large enough in the objective function of (F3). Nevertheless, knowing beforehand that some variables have been removed from the formulation has some advantages that several authors have exploited. We explicitly state the following formulation of the p-median problem on a directed graph (V, A):

$$\text{(F5) minimize} \quad \sum_{(i,j) \in A} d_j c_{ij} x_{ij}$$

$$\text{subject to } x_{ii} + \sum_{\substack{j \in V: \\ (j,i) \in A}} x_{ji} = 1 \ \forall i \in V$$

$$x_{ij} \leq x_{ii} \qquad \forall (i, j) \in A \tag{2.20}$$

$$\sum_{i \in V} x_{ii} = p$$

$$x_{ij} \in \{0, 1\} \qquad \forall (i, j) \in A \tag{2.21}$$

$$x_{ii} \in \{0, 1\} \qquad \forall i \in V.$$

A different version of this formulation is considered by Avella and Sassano (2001) who do not make use of the x_{ii} variables. Instead, they pay attention to the fact that $n - p$ nodes must be allocated by means of an arc (i.e., they are not self-allocated) and then each feasible solution will correspond to a set of $n - p$ arcs in A. They then propose the following formulation:

$$\text{(F6) minimize} \quad \sum_{(i,j) \in A} d_j c_{ij} x_{ij}$$

$$\text{subject to} \quad (2.21)$$

$$x_{ij} + \sum_{\substack{\ell \in V: \\ (\ell, i) \in A}} x_{\ell i} \leq 1 \;\; \forall (i, j) \in A \tag{2.22}$$

$$\sum_{(i,j) \in A} x_{ij} = n - p. \tag{2.23}$$

Avella and Sassano (2001) consider the case where A is a complete digraph and develop two families of inequalities. The first family, the so-called $W - 2$ inequalities, only makes use of constraints (2.22). They can then be used for the SPLP. The shape of these constraints is

$$\sum_{(i,j) \in A \cap [((W \times W) \setminus H) \cup (\bar{W} \times U)]} x_{ij} \leq |W| - 2, \tag{2.24}$$

where $W \subseteq V$ and $3 \leq |W| \leq n - p + 1$, H is a subset of arcs of A in $W \times W$ such that $\forall w \in W$ there is exactly one arc in H with origin in w, and U is the set of nodes of W that are not destinations of any arc of H. Inequalities (2.24) are facets whenever $|U| \leq \max\{1, |W| - 3\}$. We present here the example used in Avella and Sassano (2001) to illustrate this family. Consider the complete directed graph of eight nodes and the subgraph given in Fig. 2.3a. Here $|W| = 6$, $H = \{(1, 3), (3, 1), (2, 4), (4, 2), (5, 3), (6, 4)\}$ and $U = \{5, 6\}$. It produces the inequality in the family (2.24) in the shape of

$$x_{12} + x_{14} + x_{15} + x_{16} + x_{21} + x_{23} + x_{25} + x_{26} + x_{32} + x_{34} + x_{35} + x_{36} + x_{41} + x_{43} +$$
$$x_{45} + x_{46} + x_{51} + x_{52} + x_{54} + x_{56} + x_{61} + x_{62} + x_{63} + x_{65} + x_{75} + x_{76} + x_{85} + x_{86} \leq 4.$$

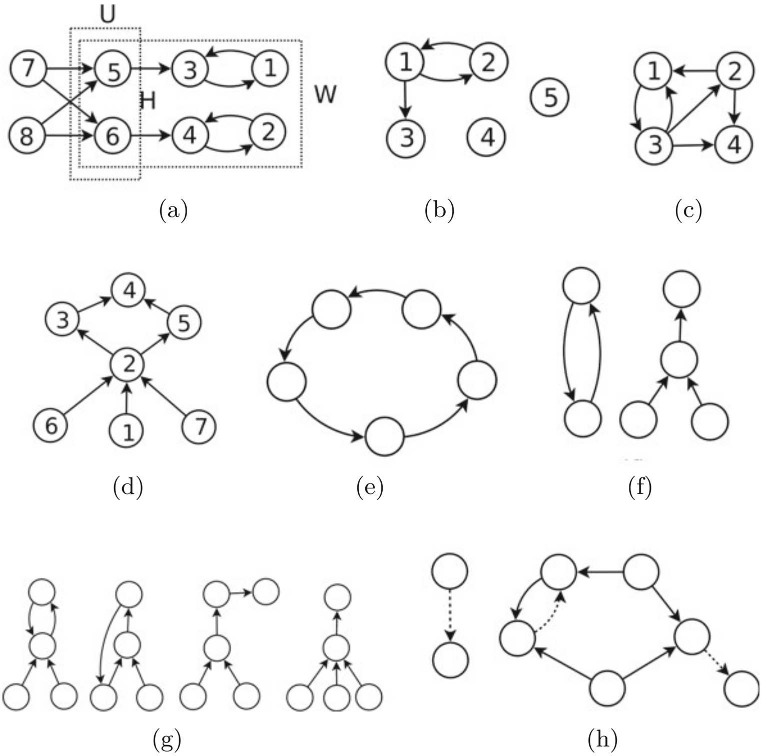

Fig. 2.3 Illustration of several inequalities and families of subgraphs. (**a**) $W-2$. (**b**) Cover. (**c**) $W-q$. (**d**) Odd hole inequalities. (**e**) $W-1$ and $|F|$ odd. (**f**) 2-cycle and Y-graph. (**g**) Graphs with Y-subgraphs. (**h**) Forbidden structure in Baïou and Barahona (2011)

Nodes 5 and 6 can be supplied or not from nodes not belonging to W. Take $x_{75} = x_{76} = 1$. Thus, the inequality becomes $x_{12}+x_{14}+x_{21}+x_{23}+x_{32}+x_{34}+x_{41}+x_{43} \leq 2$. Since no node in the set $\{1, 2, 3, 4\}$ can supply more than two other nodes in the set, it must be satisfied. Similar reasonings can be applied by taking other values of x_{75} and x_{76}.

The second family of inequalities in Avella and Sassano (2001), called *cover inequalities*, make use of constraint (2.23). They again consider A to be a complete digraph. Consider a set S of arcs and let $r(S)$ be the maximum number of arcs of S that can simultaneously take part in a solution for (F6). Let $F(S)$ be the collection of all subsets of A containing $r(S)$ arcs from S that form a solution for (F6). Choose at least one arc from each subset in $F(S)$ to create set $T(S)$. Then

$$\sum_{(i,j)\in S} x_{ij} - \sum_{(i,j)\in T(S)} x_{ij} \leq r(S) - 1$$

are valid inequalities for (F6). As an example, take the complete directed graph of five nodes, let S be the subset of arcs of Fig. 2.3b and $p = 2$. Then $r(S) = 2$ and $F(S) = \{\{(1,2),(1,3),(4,5)\}, \{(1,2),(1,3),(5,4)\}, \{(1,2),(1,3),(1,4)\}, \{(1,2),(1,3),(1,5)\}\}$. Taking $T(S) = \{(4,5),(5,4),(1,4),(1,5)\}$, the inequality produced is $x_{12} + x_{21} + x_{13} \leq 1 + x_{45} + x_{54} + x_{14} + x_{15}$. In the case $x_{45} = x_{54} = x_{14} = x_{15} = 0$, all nodes other than 1 should be assigned to node 1, but in this case $p \neq 2$. Otherwise, the sum in the left hand side is bounded by 2, the value of $r(S)$.

Regarding (F5), valid inequalities and characterizations of the polyhedron in some particular cases have been obtained by several authors. We present the main results below.

In Avella et al. (2007), the so-called $W - q$ inequalities were derived. We show an example of such inequalities based on the graph of Fig. 2.3c. Let W be the set of nodes $\{1,2,3,4\}$ and F the set of arcs $\{(2,1),(3,2),(1,3),(2,4),(3,4)\}$. Note that arc $(3,1)$ is not included in the set.

Consider the following valid inequalities:

$$\begin{array}{ll}
x_{21} & +x_{13} \leq 1, \\ \hline
x_{32} & +x_{21} \leq 1, \\
x_{32} & +x_{24} \leq 1, \\ \hline
x_{13} & +x_{32} \leq 1, \\
x_{13} & +x_{34} \leq 1, \\ \hline
x_{24} + x_{34} & \leq 1.
\end{array}$$

These valid inequalities are arranged in blocks and have been systematically built in the following way. Each block is devoted to one node $j \in W$. For each j, the sum of all variables corresponding to arcs of F that end in j is considered. Then, the sum is completed in several ways (one represented by each row of the block) by adding x_{jh} for all distinct h such that $(j,h) \in F$ (if any). In the example, this yields at most two inequalities for each block, since no more than two arcs of F leaves the same node. Note that this construction of the inequalities implies that every variable x_{ij} with $(i,j) \in F$ appears in two inequalities or in three when there are two arcs leaving node j. In order to complete those blocks that only have one inequality, we add a copy of $x_{24} + x_{34} \leq 1$ to the last block and $x_{21} \leq 1$ to the first one. Summing up the resulting set of eight inequalities, we obtain $3(x_{21} + x_{32} + x_{13} + x_{24} + x_{34}) \leq 8$. Dividing by 3 and rounding down the right-hand side, the following valid inequality in the family $W - 2$ is produced: $x_{21} + x_{32} + x_{13} + x_{24} + x_{34} \leq 2$. In the general case, consider a set of nodes $W \subseteq V$, an integer number $1 \leq q \leq |W| - 1$, and a set of arcs with both ends in W, $F \subset A \cap (W \times W)$, in such a way that no more than q arcs leave the same node. The valid inequality associated to W and F, in the family $W - q$, is then

$$\sum_{(i,j) \in F} x_{ij} \leq \left\lfloor \frac{q|W|}{q+1} \right\rfloor.$$

Avella et al. (2007) also studied odd-hole inequalities and lifted them. As an example, consider Fig. 2.3d. It is obvious that $x_{12} + x_{23} \leq 1$, $x_{23} + x_{34} \leq 1$, $x_{34} + x_{54} \leq 1$, $x_{54} + x_{25} \leq 1$ and $x_{25} + x_{12} \leq 1$. Summing up and rounding down, it follows that $x_{12} + x_{23} + x_{34} + x_{54} + x_{25} \leq 2$, named *odd-hole inequality* by the authors. Moreover, this kind of inequality can be lifted to $x_{12} + x_{23} + x_{34} + x_{54} + x_{25} + x_{62} + x_{72} \leq 2$ since arcs (1,2), (6,2) and (7,2) play the same role and only one of them can be taken in a feasible solution.

Baïou and Barahona (2008) consider the particular case of $W - q$ when $q = 1$ and $|F|$ is odd. This corresponds with oriented odd-cycles of k nodes C_k, like the one shown in Fig. 2.3e, that generate the inequalities

$$\sum_{(i,j)\in C_k} x_{ij} \leq \frac{k-1}{2}. \tag{2.25}$$

They prove that, when the graph does not contain either of the two subgraphs of Fig. 2.3f, the linear relaxation plus all the constraints in family (2.25) completely describe the polyhedron associated with formulation (F5). Graphs that do not contain these two structures are called *Y-free graphs*. They also describe a separation procedure for inequalities (2.25) through an auxiliary graph. Baïou and Barahona (2011) show that the family of graphs whose p-median polytope is integer (that is to say, the linear relaxation of formulation (F5) always produces an integer optimal solution) for all values of p are those containing none of any of the structures of Fig. 2.3g, nor any cycle of the type depicted in Fig. 2.3h. They also give additional polyhedral results in their recent paper, Baïou and Barahona (2016). Note that the structure of Fig. 2.3h is a cycle (continuous arcs) with an odd number of nodes with positive in-degree in the cycle; there are arcs (dotted) with origin in the nodes of in-degree two in the cycle and destination at nodes that either are not in the cycle or have out-degree other than two in the cycle; and there is an arc with its two nodes outside the cycle.

2.7 Heuristics

The literature on heuristics for p-median problems is vast. The account presented here does not pretend to be exhaustive and many interesting works on the topic may have been omitted. We invite the interested reader to consult other reviews for an overview of the problem from different perspectives. For instance, in Reese (2006) works are classified by solution method and are also listed by year; Mladenović et al. (2007) classify them into two classes, classical heuristics and metaheuristics, and describe the methods belonging to each group; Basu et al. (2015) focus on metaheuristics; finally, Irawan (2016) is devoted to aggregation methods, which reduce the number of demand points to obtain smaller problems.

2.7.1 Classical Heuristics

The first methodologies approaching p-median problems were heuristics. A simple one produces a feasible solution by starting from an empty set of medians and successively adding the candidate that yields the greatest decrease in the current solution value, until p candidates have been added to the set. This is known as the *greedy heuristic*. Even if Kuehn and Hamburger (1963) is usually cited as the earliest work on greedy heuristics for facility location, Cornuéjols et al. (1977) were the first to formally state the greedy heuristic for p-median problems. In the same vein, the *greedy drop* or simply *drop heuristic*, first devised by Feldman at al. (1966), starts with I as the initial set of medians and iteratively discards the candidate location whose closure produces the smallest increment of the objective function, until the initial set has been reduced to p candidates (see e.g. Whitaker 1981; Salhi and Atkinson 1995).

Other heuristics try to improve a given selection of p candidates. One of the oldest and most widely known of these heuristics allocates each user to the candidate in the initial selection with minimum supplying cost. By grouping users allocated to the same candidate, p neighborhoods are obtained. Then, a 1-median problem is solved for each neighborhood, yielding a new set of p (potentially) different medians. The process is iterated until the set of medians becomes steady. This heuristic is usually referred to as the *alternate heuristic*, and was first proposed by Maranzana (1964). Nevertheless, the idea was not new at the time and it is, in fact, a particular case of the k-means clustering, first conceived by Steinhaus (1957). Another heuristic of this type is the so-called *interchange heuristic* or *vertex substitution*, first proposed by Teitz and Bart (1968). The starting point is also a feasible set of p location candidates, and possible exchanges with the rest of the candidates are iteratively examined. A formal description of the interchange heuristic can be consulted in Whitaker (1983). The alternate and interchange heuristics have been compared empirically in several works. All of them conclude that the interchange heuristic finds better solutions but consumes more time (see e.g. Rushton and Kohler 1973; Rosing et al. 1979). This is probably why the alternate heuristic has received less attention and efforts have concentrated on improving the performance of the interchange heuristic. Countless attempts have been made in this direction, and here we mention some of them. Whitaker (1983) designed a variant of the interchange heuristic that uses a greedy initialization, called *fast interchange*; Densham and Rushton (1991) detailed specific speedup strategies and, later on, Densham and Rushton (1992) introduced GRIA (global regional interchange algorithm); Resende and Werneck (2003) presented an implementation of the fast interchange that performed especially well for large instances and reported speedups of up to three orders of magnitude over the original implementation of Whitaker. Finally, Lim and Ma (2013) introduced a parallel vertex substitution and reported speedups ranging from 10 to 57 times over the traditional algorithm.

2.7.2 Metaheuristics

The above-mentioned methods, together with dynamic programming, dual ascent and Lagrangean relaxation, can be considered as classical heuristics. These first heuristic approaches were followed by the development of metaheuristics in the 1990s. The list of works on metaheuristics for p-median problems is long. One can find well-known schemes, such as tabu search, variable neighborhood search, genetic algorithms, simulated annealing or neural networks, among others. As Mladenović et al. (2007) conclude in their review, empirical results show that metaheuristics represent an improvement in solution quality on large instances, where the performance of classical heuristics is poor. In the last decade the focus has been on solving larger and larger instances. Most effective algorithms usually combine features from different metaheuristics. In this section, we outline the most noteworthy attempts to produce scalable solution techniques. Table 2.2 summarizes some information on the accuracy and computational effort of these heuristics.

Resende and Werneck (2004) proposed a hybrid heuristic that has features of GRASP (greedy randomized adaptive search procedure), tabu search, scatter search and genetic algorithms. They empirically compared the procedure with six other methods and concluded that it was a valuable candidate for a general-purpose

Table 2.2 Summary of the available computational experience on metaheuristics

Authors	Year	Computer	n	t (s)	dev. (%)
Resende and Werneck	2004	SGI Challenge (196 MHz)	5934	8687	0.6
Hansen et al.	2009	Pentium 4 1800 MHz, 256 MB RAM	89,600	50,083	3.2
Avella et al.	2012	IntelCore 2Quad 2.6 GHz, 4 GB RAM, 64 bits	89,600	5779	54.7
Irawan and Salhi	2013	IntelCore i5-650 3.20 GHz, 4 GB RAM, 32 bits	89,600	4415	95.8
Irawan et al.	2014	IntelCore i5-6503.20 GHz, 4 GB RAM, 32 bits	89,600	3404	5.9
Salhi and Irawan	2015	IntelCore i5-650 3.20 GHz, 4.00 GB RAM, 32 bits	264,000	1,875,300	271.0
Janáček and Kvet	2016	IntelCore 2 Duo E6700 2.66 GHz, 3 GB RAM	3038	1102	9.7
Cebecauer and Buzna	2017	Brutus high-performance cluster of ETH Zurich	670,000	–[a]	4.0

The last three columns stand for the maximum size, time in seconds and deviations of the instances tested but do not necessarily correspond to one same instance. Deviations are calculated either with respect to the optimum or to the best objective known

[a]The authors set a time limit of several days and reported time efficiency with respect to the unaggregated problem

approach for the p-median problem. They used a varied testbed with instances of up to 5934 demand points and gave an account of the strengths and weaknesses of their approach, which they did not recommend for really large instances. Hansen et al. (2009) tackled the clustering problem as a large scale p-median model, using an approach based on the variable neighborhood search metaheuristic. They report better solutions in less time than with the state-of-the-art heuristics, even after upgrading these procedures with the same efficient strategies on instances of up to 89,600 nodes.

Avella et al. (2012) introduced a heuristic for large-scale instances that consists of three main components: subgradient column generation, a core heuristic, which computes an upper bound based on Lagrangean reduced costs, and an aggregation procedure that defines reduced size instances. They compared their approach with that of Resende and Werneck (2004) and Hansen et al. (2009) using the same testbed as these authors. They reported excellent results that have merited the recognition as state-of-the-art heuristic for years. Irawan and Salhi (2013) designed a hybrid heuristic for large-scale instances. The proposed approach was tested on the largest "BIRCH" instances of Hansen et al. (2009) (from 25,000 to 89,600 demand points). The authors claimed to have obtained better solutions than those of the algorithm by Avella et al. (2012), AV, and relatively similar to the ones of the algorithm by Hansen et al. (2009), HA. Nonetheless, improvement respect to AV in quality represents some decimals (in %) and they do not run AV nor HA, but take the times reported by Avella et al. (2012) and apply a transformation to estimate running times in their machine.

Irawan et al. (2014) presented a multiphase approach that incorporates aggregation, variable neighborhood search and an exact method. This heuristic proved to be faster than the one by Irawan and Salhi (2013) on the same testbed used in that previous work. This time, the algorithm is also compared with AV and HA, and times for these algorithms are again obtained by estimation. Regarding solution quality, the proposed heuristic compares with AV and HA in a similar way as that of Irawan and Salhi (2013). Salhi and Irawan (2015) introduced a data compression approach for very large facility location problems in the Euclidean space. They incorporated these techniques into two different methods for p-median problems, a multi-start and a reduced variable neighborhood search. After testing their approach, the authors concluded that it is very effective when applied to very large instances (up to 264,000 demand points in their experiments). Janáček and Kvet (2016) suggested an approximate approach based on the radius formulation (F4) and presented it as a compromise approach enabling a trade-off between accuracy and computational time. They compared their proposal with AV and the exact approach by García et al. (2011) on instances having up to 3038 demand points. Even though the results reported are not conclusive, their method seems to be a good candidate for some instances. Cebecauer and Buzna (2017) proposed the concept of adaptive aggregation that keeps the problem size in reasonable limits. They introduced a framework to approach facility location problems that iteratively adjusts the aggregation level during the solution process. They applied it to the p-median and compared its performance to the exact approach by García et al. (2011), obtaining promising results for benchmarks, which reach up to 670,000 demand points.

2.7.3 Approximation Heuristics

One of the drawbacks of many heuristics is that they do not provide any guarantee regarding the quality of the solution obtained. Since the p-median is a core problem in location, it is not surprising to find works that focus on guaranteeing good-quality approximations, even these days. One of the first works concerned with approximate solutions quality is Cornuéjols et al. (1977), who presented a worst-case analysis for relative errors of the Lagrangean relaxation, the greedy, the interchange and dynamic programming heuristics. Some of the heuristics mentioned above also provide a lower bound on the objective function, which gives an estimation of the quality of their solutions. When we have a precise assessment of the quality of the solution with respect to the optimum we speak about *approximation algorithms*. We define an α-approximation algorithm as a polynomial-time algorithm that computes a solution with cost at most α times that of an optimal solution. Most of the papers on approximation algorithms make some assumption regarding costs. When they are given by Euclidean distances, it is known that, for any $\epsilon > 0$, there exists a nearly linear-time $(1 + \epsilon)$-approximation algorithm, see Kolliopoulos and Rao (1999). When costs satisfy the triangle inequality, we speak about the *metric p*-median and the best current approximation factor is $2.675 + \epsilon$, obtained by Byrka et al. (2014). Moreover, Jain et al. (2002) proved that there is no α-approximation of the metric p-median with $\alpha < 1 + 2/e$, unless P = NP.

2.8 Conclusions

We have briefly presented different versions of the p-median problem, their formulations, solution methods, polyhedral properties and heuristic algorithms. We have focused on the basic models, without going into details about the properties of the Simple Plant Location Problem, a very similar problem that is well studied in Chap. 4. Neither have we paid attention to the many possible extensions of the problem, that make it more applicable and realistic, but which are covered in different chapters of this book (addition of a limit of capacity in the facilities, opening and closing facilities in different periods of time, stochastic demands, different objective functions and a long list of options). The p-median problem still receives considerable attention 50 years after its first appearance in the literature and is an exciting field of future research.

Acknowledgements This research has been partially supported by the research projects MTM2015-65915-R (MINECO, Spain), 19320/PI/14 (*Fundación Séneca*) and FUNDB-BVA/016/005 (*Fundación BBVA*), and the PhD grant FPU15/05883 (MECD, Spain).

References

AlBdaiwi BF, Ghosh D, Goldengorin B (2011) Data aggregation for p-median problems. J Combin Optim 21:348–363

Avella P, Sassano A (2001) On the p-median polytope. Math Program A 89:395–411

Avella P, Sassano A, Vasilyev I (2007) Computational study of large scale p-median problems. Math Program A 109:89–114

Avella P, Boccia M, Salerno S, Vasilyev I (2012) An aggregation heuristic for large scale p-median problem. Comput Oper Res 39:1625–1632

Baïou M, Barahona F (2008) On the p-median polytope on Y-free graphs. Discrete Optim 5:205–219

Baïou M, Barahona F (2011) On the linear relaxation of the p-median problem. Discrete Optim 8:344–375

Baïou M, Barahona F (2016) On the p-median polytope and the directed odd cycles inequalities: triangle-free oriented graphs. Discrete Optim 22:206–224

Balinski ML (1965) Integer programming: methods, uses, computation. Manag. Sci 12:253–313

Basu S, Sharma M, Ghosh PS (2015) Metaheuristic applications on discrete facility location problems: a survey. Opsearch 52:530–561

Beltrán C, Tadonki C, Vial JP (2006) Solving the p-median problem with a semi-Lagrangian relaxation. Comput Optim Appl 35:239–260

Briant O, Naddef, D (2004) The optimal diversity management problem. Oper Res 52:515–526

Byrka J, Pensyl T, Rybicki B, Srinivasan A, Trinh K (2014) An improved approximation for k-median, and positive correlation in budgeted optimization. In: Proceedings of 26th annual ACM-SIAM symposium on discrete algorithms, San Diego, pp 737–756

Cánovas L, Landete M, Marín A (2000) New facets for the set packing polytope. Oper Res Lett 27:153–161

Cánovas L, Landete M, Marín A (2002) On the facets of the simple plant location problem. Discrete Appl Math 124:27–53

Cánovas L, Landete M, Marín A (2003) Facet obtaining procedures for set packing problems. SIAM J Discrete Math 16:127–155

Cánovas L, García S, Labbé M, Marín A (2007) A strengthened formulation for the simple plant location problem with order. Oper Res Lett 35:141–150

Cebecauer M, Buzna L' (2017). A versatile adaptive aggregation framework for spatially large discrete location-allocation problems. Comput Ind Eng 111:364–380

Cho DC, Johnson EL, Padberg MW, Rao MR (1983a) On the uncapacitated plant location problem I: valid inequalities and facets. Math Oper Res 8:579–589

Cho DC, Padberg MW, Rao MR (1983b) On the uncapacitated plant location problem II: facets and lifting theorems. Math Oper Res 8:590–612

Church RL (2003) COBRA: a new formulation of the p-median location problem. Ann Oper Res 122:103–120

Cornuéjols GP, Thizy JM (1982) Some facets of the simple plant location polytope. Math Program 22:50–74

Cornuéjols G, Fisher ML, Nemhauser GL (1977) Location of bank accounts to optimize float: an analytical study of exact and approximate algorithms. Manage Sci 22:789–810

Cornuéjols G, Nemhauser GL, Wolsey LA (1980) A canonical representation of simple plant location problems and its applications. SIAM J Algebra Discrete Methods 1:261–272

Cornuéjols G, Nemhauser GL, Wolsey LA (1990) The uncapacitated facility location problem. In: Mirchandani PB, Francis RL (eds) Discrete location theory. Wiley, New York, pp 119–171

de Farias IR (2001) A family of facets for the uncapacitated p-median polytope. Oper Res Lett 28:161–167

Densham PJ, Rushton G (1991) Designing and implementing strategies for solving large location-allocation problems with heuristic methods. Technical report 91-10, National Center for Geographic Information and Analysis, Buffalo

Densham PJ, Rushton G (1992) A more efficient heuristic for solving large p-median problems. Pap Reg Sci 71:307–329

de Vries S, Posner ME, Vohra RV (2003) Polyhedral properties of the K-median problem on a tree. Math Program 110:261–285

Feldman E, Lehrer FA, Ray TL (1966) Warehouse locations under continuous economies of scale. Manag Sci 12:670–684

Galvão RD (1980) A dual-bounded algorithm for the p-median problem. Oper Res 28:1112–1121

García S, Labbé M, Marín A (2011) Solving large p-median problems with a radius formulation. INFORMS J Comput 22:546–556

Goldengorin B, Krushinsky D, Slomp J (2012) Flexible PMP approach for large-size cell formation. Oper Res 60:1157–1166

Guignard M (2003) Lagrangean relaxation. Top 11:151–200

Hakimi SL (1964) Optimum locations of switching centers and the absolute centers and medians of a graph. Oper Res 12:450–459

Hakimi SL (1965) Optimum distribution of switching centers and some graph-related theoretic problems. Oper Res 12:462–475

Hansen P, Brimberg J, Urošević D, Mladenović N (2009) Solving large p- median clustering problems by primal-dual variable neighborhood search. Data Min Knowl Disc 19:351–375

Hua LK et al. (1962) Application of mathematical methods to wheat harvesting. Chin Math 2(7791):8

Irawan, CA (2016) Aggregation and non aggregation techniques for large facility location problems-a survey. Yugosl. J. Oper. Res. 25(3):313–341

Irawan CA, Salhi S (2013) Solving large p-median problems by a multistage hybrid approach using demand points aggregation and variable neighbourhood search. J Global Optim 63(3):537–554

Irawan CA, Salhi S, Scaparra MP (2014) An adaptive multiphase approach for large unconditional and conditional p-median problems. Eur J Oper Res 237(2):590–605

Jain K, Mahdian M, Saberi A (2002) A new greedy approach for facility location problems. In: Proceedings of 34th annual ACM symposium on theory of computing, New York, pp 731–740

Janáček J, Kvet M (2016) Sequential approximate approach to the p-median problem. Comput Ind Eng 94:83–92

Kariv O, Hakimi SL (1979) An algorithmic approach to network location problems. II: the p-medians. SIAM J Appl Math 37:539–560

Kolen A (1983) Solving covering problems and the uncapacitated plant location problem on trees. Eur J Oper Res 12:266–278

Kolen A, Tamir A (1990) Covering problems. In: Mirchandani PB, Francis RL (eds) Discrete location theory. Wiley, New York, pp 263–304

Kolliopoulos SG, Rao S (1999) A nearly linear-time approximation scheme for the Euclidean k-median problem. In: Nešetřil J (ed) Algorithms - ESA'99. Lecture notes in computer science, vol 1643. Springer, Berlin, pp 378–389

Kuehn AA, Hamburger MJ (1963) A heuristic program for locating warehouses. Manag Sci 11:643–666

Li, B, Deng X, Golin MJ, Sohraby K (1998) On the optimal placement of web proxies in the internet: the linear topology. In: High performance networking. Springer, Boston, pp 485–495

Lim GJ, Ma L (2013) GPU-based parallel vertex substitution algorithm for the p-median problem. Comput Ind Eng 64(1):381–388

Manne A (1964) Plant location under economics of scale – decentralization and computation. Manag Sci 11:213–235

Maranzana FE (1964) On the location of supply points to minimize transport costs. Oper Res Quart 15:261–270

Mesa JA, Ortega FA, Pozo MA (2014) Locating optimal timetables and vehicle schedules in a transit line. Ann Oper Res 222(1):439–455

Mirchandani P (1990) The p-median problem and generalizations. In: Mirchandani PB, Francis RL (eds) Discrete location theory. Wiley, New York, pp 55–117

Mladenović N, Brimberg J, Hansen P, Moreno-Pérez JA (2007) The p-median problem: a survey of metaheuristic approaches. Eur J Oper Res 179:927–939

Nemhauser GL, Wolsey LA (1988) Integer and combinatorial optimization. Wiley, New York

Reese J (2006) Solution methods for the p-median problem: an annotated bibliography. Networks 48:125–142

Resende MGC, Werneck RF (2003) On the implementation of a swap-based local search procedure for the p-median problem. In: Proceedings of 5th workshop on algorithm engineering and experiments, Philadelphia, pp 119–127

Resende MGC, Werneck RF (2004) A hybrid heuristic for the p-median problem. J Heuristics 10:59–88

ReVelle CS, Swain R (1970) Central facilities location. Geogr Anal 2:30–42

Rosing KE, Hillsman EL, Rosing-Vogelaar H (1979) A note comparing optimal and heuristic solutions to the p-median problem. Geogr Anal 11:86–89

Rushton G, Kohler JA (1973) ALLOC: heuristic solutions to multifacility location problems on a graph. In: Rushton G, Goodchild M, Ostresh L (eds) Computer programs for location-allocation problems. University of Iowa, Department of Geography. Monograph, vol 6, pp 163–187

Salhi S, Atkinson RA (1995) Subdrop: a modified drop heuristic for location problems. Locat Sci 2:267–273

Salhi S, Irawan CA (2015) A quadtree-based allocation method for a class of large discrete Euclidean location problems. Comput Oper Res 55:23–35

Simão HP, Thizy JM (1989) A dual simplex algorithm for the canonical representation of the uncapacitated facility location problem. Oper Res Lett 8:279–286

Steinhaus H (1957) Sur la division des corps matériels en parties. Bull Acad Polon Sci 4(12):801–804

Teitz MB, Bart P (1968) Heuristic methods for estimating the generalized vertex median of a weighted graph. Oper Res 16:955–961

Vigneron A, Gao L, Golin MJ, Italiano GF, Li B (2000) An algorithm for finding a k-median in a directed tree. Inf Process Lett 74:81–88

Vinod HD (1969) Integer programing and the theory of grouping. J Am Stat Assoc 64:506–519

Whitaker RA (1981) A tight bound drop exchange algorithm for solving the p-median problem. Environ Plan A 12:669–680

Whitaker RA (1983) A fast algorithm for the greedy interchange of large-scale clustering and median location problems. INFOR Inf Syst Oper Res 21:95–108

Wolsey LA (1998) Integer programming. Wiley, New York

Xu N, Lowe TJ (1993) On the equivalence of dual method for two location problems. Transp Sci 27:194–199

Zhao W, Posner ME (2011) A large class of facets for the K-median polytope. Math Program 128:171–203