# Path based algorithms for metro network design

Gilbert Laporte[1,2] and Marta M. B. Pascoal[3,4*]

[1]Interuniversity Research Centre on Enterprise Networks, Logistics and Transportation (CIRRELT)
E-mail: Gilbert.Laporte@cirrelt.ca

[2]Canada Research Chair in Distribution Management, HEC, Montréal, 3000 chemin de la
Côte-Sainte-Catherine, Montréal, Canada H3T 2A7

[3]Institute for Systems Engineering and Computers – Coimbra (INESCC)

[4]Department of Mathematics, University of Coimbra, Apartado 3008, EC Santa Cruz, 3001-501 Coimbra,
Portugal
Phone: +351 239 791150, Fax: +351 239 832568
E-mail: marta@mat.uc.pt

**April 16, 2015**

**Abstract** This paper proposes a practical methodology for the problem of designing a metro configuration under two criteria: population coverage and construction cost. It is assumed that a set of corridors defining a rough a priori geometric configuration is provided by the planners. The proposed algorithm consists of fine tuning the location of single alignments within each corridor. This is achieved by means of a bicriteria methodology that generates sets of non-dominated paths. These alignments are then combined to form a metro network by solving a bicriteria integer linear program. Extensive computational experiments confirm the efficiency of the proposed methodology.

**Keywords:** metro network design, path based algorithm, bicriteria optimization.

## 1 Introduction

In several cities, metro systems provide a desirable alternative to private transportation. They reduce traffic congestion, improve mobility and contribute to the protection of the environment. Designing a metro network is a complex problem in which several criteria must be considered, technical difficulties are important and the costs are very high. These problems also involve many different decision makers such as city planners, engineers, interest groups and politicians. For a recent overview of the metro design literature, see [11].

Because we are dealing with mobility issues, one of the goals when designing a metro system is to serve as many people as possible. It has been shown that people are willing to walk to a metro station provided that the walking distance to the metro station is not too large. Vuchic [24] states that this distance should not exceed 400 meters (5 minutes). In metro location, two objective functions are typically considered: maximizing population coverage [5, 8, 16, 19, 20, 18] and maximizing origin-destination trip coverage [10, 15]. In most papers, construction cost is considered as a constraint. Here we jointly maximize population coverage and minimize cost in a bicriteria function. Typical constraints for the location of a metro line are the distance between

---

*Corresponding author

stations (either consecutive stations in a line or between any two of them) and the total number of stations to be built.

The planners often have a rough idea about the general shape of the metro system to be constructed and about the major destinations it should serve, such as universities, hospitals, sport facilities, commercial centers and tourist attractions. Therefore, rather than looking at a representation of the whole city, broad corridors are defined within which the metro lines should be located. Several studies on real metro system network topologies [12, 13, 14] indicate that certain configurations are more effective than others. Thus Laporte et al. [12] have compared a number of measures on various generic metro configurations (see Figure 1): 1) network complexity (number of edges divided by number of nodes); 2) connectivity (ratio of the number of edges compared to the maximal number of edges that could exist in a planar graph for a given number of nodes); 3) directness (proportion of origin-destination paths that can be traveled without transfers); 4) passenger/ network effectiveness (ratio of travel time with transfers to travel time disregarding transfers). These measures indicate that star configurations are the least complex, but far worse than triangle and cartwheel configurations in terms of connectivity, directness and passenger/network effectiveness. With respect to the same criteria, triangle configurations are slightly worse than cartwheels. In this paper we will restrict our attention to these three basic configurations. While stars are relatively inefficient, they are often preferred as initial configurations for new systems (e.g., Montreal, Minsk). With time these systems evolve and triangles are often the next step (e.g., the Minsk metro extension) and are used by some cities (e.g., Prague). Cartwheel-like configurations are common in some complex and mature metro systems such as those found in London or Moscow.
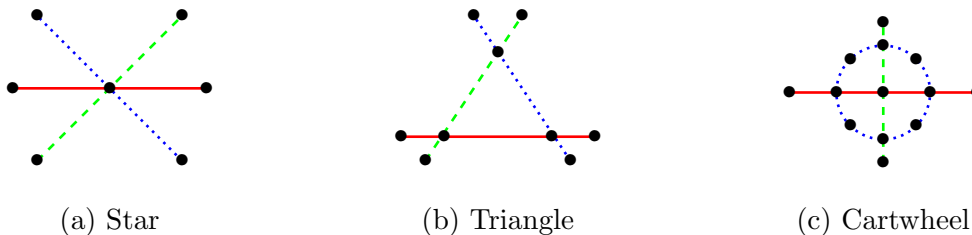


(a) Star  (b) Triangle  (c) Cartwheel

Figure 1: Metro configurations

In this paper we work with the modular approach first suggested by Bruno and Laporte [2]. The idea consists of first setting broad corridors whose combination defines a desired configuration such as a star, a triangle or a cartwheel. (In Bruno and Laporte [2] an interactive system allows planners to make a selection among a menu of predefined configurations and to define their own.) The system constructs an alignment within a given corridor, i.e., a metro line defined by the location and sequence of its stations. This alignment minimizes a single cost. Because we consider two criteria, several non-dominated alignments are constructed within each corridor. The final metro configuration is then obtained by selecting one alignment per corridor through the solution of an integer linear program. With respect to Bruno and Laporte, we consider two criteria instead of one, and we generate several alignments for each corridor instead of only one.

The scientific contribution of this paper is algorithmic. We propose a modular heuristic capable

of constructing a complex metro network by decomposing it into components and solving each component separately, thus considerably reducing the complexity of the construction process. The remainder of this article is organized as follows. In the next section we present a mathematical model for the problem of locating a metro line and its stations. Section 3 is dedicated to the development of algorithms for the location of a metro line within a corridor, considering one or two criteria. In Section 4 integer formulations are introduced in order to optimally combine solutions of the problem in Section 3 according to classical metro system configurations with several corridors. In Section 5 we present and discuss results of extensive computational experiments. Conclusions are drawn in Section 6.

## 2 Mathematical model for the location of a metro line

We first present a model for locating a metro line in a general network representing a city, or part of it. We then explain how to obtain a network that represents a corridor within a city.

We consider an area in which a metro line is to be located as a network $G = (\mathcal{N}, \mathcal{E})$, where $\mathcal{N}$ is the set of nodes, the possible station locations, and $\mathcal{E}$ is the set of edges, the links between stations. A metro line is represented by a sequence of stations connected by edges. A metro line corresponds to an undirected path of the form $p = \langle v_1, v_2, \ldots, v_\ell \rangle$, where $v_i \in \mathcal{N}$ for $i = 1, \ldots, \ell$. For simplicity, we write $v_i \in p$ if the node $v_i$ appears on path $p$. Each node $i$ of the network is associated with the catchment area population of a station located there, $d_{ij}$. With each arc $(i,j)$ we associate with cost $c_{ij}$ of constructing a link between stations $i$ and $j$, and the Euclidian distance $l_{ij}$ between the nodes $i$ and $j$.

Given an initial and a terminal nodes, respectively $s$ and $t$, that correspond to ending stations, the problem of locating a metro line linking $s$ to $t$ is defined as

$$\text{maximize} \quad d(p) = \sum_{i \in p} d_i \tag{1}$$

$$\text{minimize} \quad c(p) = \sum_{(i,j) \in p} c_{ij} \tag{2}$$

$$\text{subject to} \quad \underline{l}(p) = \min\{l_{ij} : (i,j) \in p\} \geq m \tag{3}$$

$$\bar{l}(p) = \max\{l_{ij} : (i,j) \in p\} \leq M \tag{4}$$

$$h(p) \leq W \tag{5}$$

$$l_{ij} \geq D, \ \forall i,j \in p \tag{6}$$

$$p \text{ a path between } s \text{ and } t. \tag{7}$$

Possible objective functions for this problem are $d$, providing a measure for the population coverage along the line associated with that path, and $c$, a measure for the construction cost along the same metro line. In terms of a metro line, the functions $\underline{l}$ and $\bar{l}$ represent the minimum and the maximum distances between two consecutive stations. Our goal is to locate a metro line that maximizes the population coverage defined by (1) (in the next section we will also include as second objective the minimization of the construction cost defined by (2)) over the set of all feasible paths.

Constraint (3) sets a lower bound on the distance between consecutive stops, with $m$ a given constant. In addition, for accessibility reasons two consecutive stations on a line should not be too far from each other, which is represented by (4), with $M$ another constant. Typical values for $m$ and $M$ are around 1 km and 2 km, respectively [24]. We also consider an upper bound on the number of stops allowed in a metro line, given by (5), with $h$ a function that assigns to each path its number of nodes and $W$ a predefined upper bound. This constraint is necessary since it is difficult to operate a metro line with too many stations.



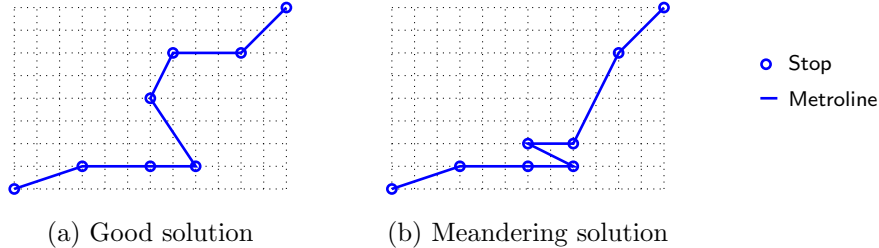(a) Good solution       (b) Meandering solution

Figure 2: Meandering along a metro line

It is also important to avoid meandering lines, as shown in Figure 2. This means that, not only should the distance between consecutive stations be taken into account, but also the distance between any two stations along the line. The minimum distance between stations is predefined and is denoted as $D$. This constraint is represented by (6) in the formulation above.

A feasible metro line is illustrated in Figure 3, considering that the network nodes coincide with the mesh intersections and omitting constraint (5). In this figure, the parameters are $m = 1.5$, $M = 3.5$ and $D = 1.8$. The shaded ring centered at point $A$ defines the area containing the points reachable from $A$ and satisfying constraints (3) and (4). The remaining shaded circles represent the forbidden areas needed to enforce the meandering constraint, assuming that when point $A$ is chosen only the leftmost metro line location is known.
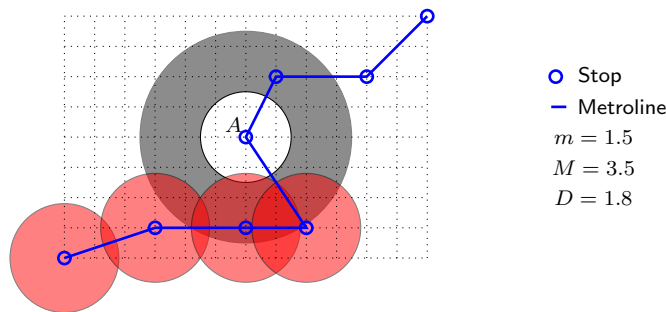


Figure 3: Feasible reachable area from point $A$ (mesh size of 1 unit)

Typically one has access to a representation of the city as a uniform grid, together with the population distribution for the grid points and the construction cost for each link between two neighboring grid points. Let $p_k$ be the population associated with a point $k$ in the grid. This population is typically that of a small neighborhood surrounding $k$. We assume that people living near the station can still walk to it, in a proportion that decreases with the walking distance. Let

4

$w_{ik}$ be the walking distance between points $i$ and $k$ of the grid, and let $\theta_d$ be the proportion of the population located $d$ units away from the station, likely to use it. Typical values, used also in our experiments, are $\theta_0 = \theta_1 = 1.00$, $\theta_2 = 0.50$, $\theta_3 = 0.25$, and $\theta_d = 0.00$, for $d \geq 4$. As in [6], the catchment area population of a station located at point $i$ is defined by

$$d_i = \sum_{d=0}^{3} \sum_{k:w_{ik}=d} \theta_d p_k, \tag{8}$$

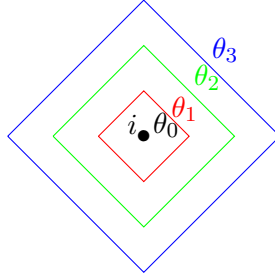and is graphically depicted in Figure 4 for the case of a Manhattan walking distance.



Figure 4: Catchment area of a metro station

Another important factor for the location of a transportation network is the cost of constructing the metro line. The value of $c_{ij}$ is either given (when $i$ and $j$ are associated with neighbor points in the grid) or can be calculated as an average of the construction cost of the areas traversed by arc $(i, j)$.

As mentioned earlier, the search space for the location of one metro line is usually a rough corridor identified by the planners. We assume this corridor is defined by the initial and the terminal point, points $A$ and $B$, which stand for the line ends and by a parameter $\Delta$, denoting the corridor width, all given by the planner. The corridor includes the regions located within a radius $\Delta/2$ from $A$ and $B$, and the region along the straight line that connects $A$ and $B$ (Figure 5). The network that represents the corridor corresponding to these data is obtained as follows. The network nodes in the ending areas are equally distant grid points. A dense mesh of points is considered around the points $A$ and $B$, corresponding to the circles in Figure 5. A usually sparser mesh is used in the rectangular corridor. We use a tighter mesh in the ending areas of the metro line to allow more flexibility when selecting the stops location around those sites considered as crucial.
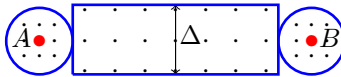


Figure 5: Corridor with two ending areas

Because the arcs in a path represent lines connecting consecutive metro line stations, these can link any two nodes as long as they respect the minimum and the maximum station separation dis-

tances between consecutive metro line stations, given by constraints (3) and (4), and thus Lemma 1 provides a feasibility condition for the edges in a metro line.

**Lemma 1.** *If $(i, j) \in p$, for some path $p$ satisfying (3) and (4), then $m \leq l_{ij} \leq M$.*

It can then be concluded that only edges $(i, j)$ satisfying $m \leq l_{ij} \leq M$ have to be considered. In the following, it is assumed that the network representing a corridor is restricted to such edges. Additionally, arc and node parameters are given, or calculated, as explained earlier.

Finally, given that a metro line is a path, a source and a destination nodes are associated with each of the ending areas of the corridor. That is done by taking as the source a dummy node $s$, linked to all the nodes in one of the ending areas. The sink is another dummy node $t$, and all the nodes in the other ending area are linked to $t$. The parameters of the arcs that link dummy nodes and other nodes are set to 0.

# 3 Location of a metro line along a corridor

In this section we present formulations for the problem of locating one metro line along a given corridor under the conditions introduced in Section 2. We first analyze a single criterion version of the problem, involving the population coverage maximization, and we then consider the construction cost as a second criterion.

If meandering is disregarded completely, then locating a metro line along a given corridor when maximizing $d$ can be formulated as a weight constrained shortest path problem and, given that $W$ is usually small, the problem can be solved in polynomial time by a method similar to that proposed in [17]. If two objective functions are to be optimized, namely $d$ and $c$, then another method in [17] can still serve as the basis for an algorithm. However, this algorithm is no longer polynomial given that it computes non-dominated solutions. In a general case the meandering constraint can raise some difficulties, given that it forces to compare each candidate station and all the preceding stations in that metro line. This task is more demanding as the line becomes longer. Besides, the corridors usually set by the planners are not too wide, when compared to $l$ and $M$, and in practice we want the metro line to avoid turns that make it go backward and forward along the corridor. Therefore, instead of comparing the locations of a new candidate station and all the preceding stations in that metro line, the algorithm limits these stations to two and checks whether the distance between the new location and the two previous stations is feasible with respect to meandering. Let $\xi_i$ denote the node preceding node $i$ in a metro line. Then the arc $(i, j)$ can be used only if

$$l_{\xi_i j} \geq D. \tag{9}$$

Given initial and terminal nodes, $s$ and $t$, it is possible to find a set of feasible and non-dominated solutions using a labeling process and constructing a search tree in which each node $x$, corresponding to a path from $s$ to $i$, is associated with a label $L_x$. Other labeling algorithms for bicriteria or multicriteria shortest path problems can be found in [1, 3, 9, 21, 22]. Given a node $i \in \mathcal{N}$, this partial solution can be extended by looking at its neighbors $j \in \mathcal{N}$ (such that $(i, j) \in \mathcal{E}$). Then, a new label $L_z$ is created, corresponding to the same path but with a new stop at the station

associated with node $j$, if the meandering constraint (9) is satisfied. The label of node $z$ is of the form $L_z = [\pi_z^d, \pi_z^c, \pi_z^h, \xi_z, \beta_z]$, where

- $\pi_z^d$ denotes the path coverage, updated with $\pi_z^d = \pi_x^d + d_j$;

- $\pi_z^c$ denotes its construction cost, updated as $\pi_z^c = \pi_x^c + c_{ij}$;

- $\pi_z^h$ is its number of stops, updated as $\pi_z^h = \pi_x^h + 1$;

- $\xi_z$ is the index of the tree node preceding $z$ in the path, $\xi_z = x$, and

- $\beta_z$ is the network node that corresponds to $z$, i.e., $\beta_z = j$.

It is possible that several labels, corresponding to different nodes in the search tree, will be associated with one network node. In the single-objective case where $d$ is maximized, it is not necessary to take into account the components $\pi_z^c$ related with the construction cost.

Checking label dominance can be achieved by comparing the coverage and the construction cost values for pairs with at most as many stations, and discarding dominated labels by comparing the coverage and the construction cost for pairs with at least the same number of stations.

**Definition 1.** *Let $L_{z'}$ and $L_{z''}$ be two labels associated with node $z$ such that $\pi_{z'}^h \leq \pi_{z''}^h$. Then $L_{z'}$ dominates $L_{z''}$ if*

1. *when $\pi_{z'}^h = \pi_{z''}^h$, then $\pi_{z'}^d \geq \pi_{z''}^d$, $\pi_{z'}^c \leq \pi_{z''}^c$, and at least one of the inequalities is strict, or*

2. *when $\pi_{z'}^h < \pi_{z''}^h$, then $\pi_{z'}^d \geq \pi_{z''}^d$ and $\pi_{z'}^c \leq \pi_{z''}^c$.*

Algorithm 1 summarizes the method just presented for the case where the metro line coverage is maximized and the metro line construction cost is minimized.

---

**Algorithm 1:** Metro line location along a corridor

```
/* Lx = [πˣᵈ, πˣᶜ, πˣʰ, βₓ, ξₓ] is the label associated with node x                    */
```

1   $nX \leftarrow 1$; $L_{nX} \leftarrow [0, 0, 1, s, -]$
2   $X \leftarrow \{1\}$
3   **while** $X \neq \emptyset$ **do**
4      $x \leftarrow$ node in $X$; $X \leftarrow X - \{x\}$
5      **if** $\pi_x^h < W$ **then**
6         $i \leftarrow \beta_x$
7         **for** $j$ such that $(i, j) \in \mathcal{E}$ **and** $l_{\xi_x j} \geq D$ **do**
8            **if** $[\pi_x^d + d_j, \pi_x^c + c_{ij}, \pi_x^h + 1, x, j]$ is not dominated by another label of $j$ **then**
9              $nX \leftarrow nX + 1$; $L_{nX} \leftarrow [\pi_x^d + d_j, \pi_x^c + c_{ij}, \pi_x^h + 1, x, j]$
10              $X \leftarrow X \cup \{nX\}$
11              Delete labels of nodes associated with $j$ in $X$ that are dominated by $L_{nX}$

12 **for** $k = 1, \ldots, W$ **do**
13     **return** *paths associated with labels of the form* $[\pi_x^d, \pi_x^c, k, \beta_x, t]$

---

---

**Procedure** $L_x$ is dominated by a label of $j$ such that $\beta_x = j$

/* $L_x = [\pi_x^d, \pi_x^c, \pi_x^h, \beta_x, \xi_x]$ */

**1 for** *any node $y$ such that $\beta_y = j$* ***and*** *$\pi_y^h \leq \pi_x^h$* **do**

**2**    **if** $\pi_y^d \geq \pi_x^d$ ***and*** $\pi_x^c \leq \pi_x^c$ ***and*** $(\pi_x^d, \pi_x^c) \neq (\pi_y^d, \pi_y^c)$ **then**

**3**      **return** $L_x$ *is dominated*

**4 return** $L_x$ *is not dominated*

---

This method can be seen as a labeling approach based on the assumption that the problem is a weight constrained shortest path problem. Therefore, the method works like other multi-objective constrained shortest path procedures [3, 9].

When considering $d$ as the single criterion, the dominance relation can be simplified according to Definition 2.

**Definition 2.** *Given $L_{z'}$ and $L_{z''}$, labels associated with node $z$ such that $\pi_{z'}^h \leq \pi_{z''}^h$, it is said that $L_{z'}$ dominates $L_{z''}$ if*

1. *when $\pi_{z'}^h = \pi_{z''}^h$, it holds that $\pi_{z'}^d > \pi_{z''}^d$, or*

2. *when $\pi_{z'}^h < \pi_{z''}^h$, it holds that $\pi_{z'}^d \geq \pi_{z''}^d$.*

In this case a similar but simpler version of Algorithm 1 can be derived, where Definition 1 is replaced by Definition 2 in terms of the dominance relation. Its presentation is omitted.

The output of Algorithm 1 is a set of feasible labels that are non-dominated with respect to $d$ and $c$. If only $d$ is considered, then the selected solution should maximize $d$ among all other feasible solutions, yielding a unique metro line. Otherwise, if both $d$ and $c$ are considered, then to each feasible number of stations is associated a set of non-dominated solutions, which are incomparable without the specification of extra conditions or preferences.

# 4 Location of metro lines along a multi-corridor configuration

The algorithms introduced in the previous section aim at locating a metro line between the two ending areas of a given corridor. This section is dedicated to the location of metro lines in configurations with more than one corridor, and focuses on cases known to be the most efficient [12]: the star, triangle and cartwheel configurations.

The configuration of the metro to be constructed is assumed to be known and to be made up of several corridors. The problem of computing a suitable location of the metro lines in that configuration, with respect to path coverage and construction cost, is tackled in two steps. In the first step the metro configuration is split into several corridors, so that their ending areas correspond either to terminal areas or to transfer areas between different lines. Algorithm 1 or its single criterion version, is then applied to the pair of ending areas in each corridor, in order to obtain a set of non-dominated metro lines, or an optimal metro line, respectively. In the second step these solutions are combined with the goal of completing the desired configuration in order to

optimize the objective function(s). We next focus on the case where only the maximization of $d$ is considered, and the methodology is then extended to two criteria.

## 4.1 Single criterion case

The methods just described can be applied to the location of a metro line within a corridor to create solutions starting at one end of the corridor and ending at the other. When such solutions are given, assembling metro lines in the different corridors of a configuration amounts to determining the location chosen for the station in each transfer area of the corridors.

In a multicorridor configuration the two ending areas of the corridors forming the metro configuration can either be used as terminals or as transfer areas. A transfer area is shared by several corridors, and a station within that area corresponds to a transfer station between lines, which means that its location affects the objective value associated with all lines ending in that area. A terminal area can be treated, as explained earlier, by linking every node in that area to a single dummy node. In this case it is useful to add the dummy node at the end of the line, given that the single criterion version of Algorithm 1 can be adapted to compute the metro lines that maximize coverage population from any node in the other ending area of the corridor to the dummy node, and this will provide the best metro line ending at a node of the first terminal area. The following result shows that the same sequence of nodes in one direction is also defined in the other direction, with the same objective values, and therefore the solution does not depend on the definition of the starting and ending areas.

**Lemma 2.** *Let $T_1$ and $T_2$ be the sets of nodes of two ending areas of a corridor. Let $p$ be a feasible metro line from $i \in T_1$ to $j \in T_2$ along that corridor, then there is $q$ a feasible metro line from $j \in T_2$ to $i \in T_1$ along the same corridor such that $d(p) = d(q)$.*

*Proof.* It is sufficient to consider the path $q$, given by reversing the sequence of arcs in $p$, given that the network representing the corridor is undirected. This means that if $(x, y)$ is a feasible link in $\mathcal{E}$, then $(y, x)$ is also a feasible link between the points $y$ and $x$. Then $h(q) = h(p) \leq W$, therefore $q$ is feasible, and $d(p) = d(q)$. □

Furthermore, because there are no constraints on $T_2$, it is only necessary to know the best option ending in $j$, for every $j \in T_1$. It is therefore sufficient to compute the best path from any $j \in T_1$ to the dummy node, as shown in Lemma 3.

**Lemma 3.** *Let $T_1$ and $T_2$ be the sets of nodes of two ending areas of a corridor, and $t$ be a dummy node linked to all nodes in $T_2$. Let $p^* = p' \diamond \langle j, t \rangle$, with $j \in T_2$, be the best feasible metro line from a node $i \in T_1$ to $t$ along that corridor. Then, $p'$ is the best feasible metro line from $i$ to any $j \in T_2$ to $i \in T_1$ along that corridor.*

*Proof.* Let $i \in T_1$ and $p^*$ be the best feasible line from $i$ to the dummy node $t$ associated with $T_2$. Then $p^* = p' \diamond \langle j, t \rangle$, where $\diamond$ represents the concatenation of two paths, with $j \in T_2$ and $p'$ the best feasible line from $i$ to any node in $T_2$. Assume that $q'$ is the best feasible line from $i$ to $k \in T_2$, and $q^* = q' \diamond \langle k, t \rangle$. Then $q^*$ is feasible and $d(q') > d(p')$, thus $d(q^*) = d(q') + 0 > d(p^*)$, which contradicts the optimality of $p^*$. □

For corridors ending in two transfer areas, we set one of them as the start area of that line, and the other as its terminal area, and we then compute the best path from any node in the starting area to any node in the terminal area.



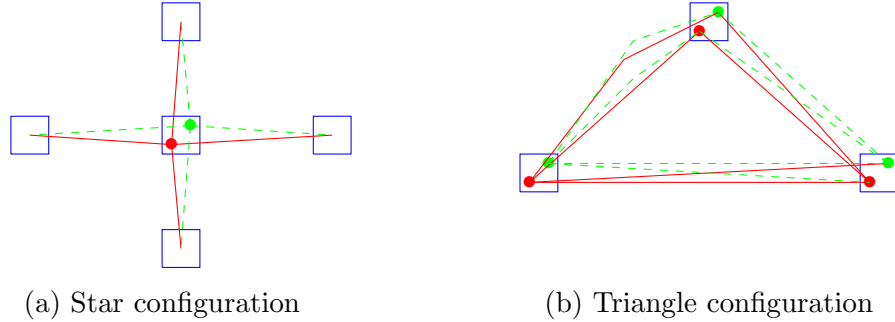(a) Star configuration           (b) Triangle configuration

Figure 6: Single criterion candidate metro lines

Figure 6 illustrates how the computed solutions for all the corridors look like in some simple star and triangle configurations with two possible station locations in each transfer area. Note that the star configuration has a single transfer area and four terminal areas, whereas there are three transfer areas and no terminal area in this triangle configuration. The number of solutions for the metro system location depends on the number of transfer areas in the configuration, and this number grows fast. There are two possible solutions for the metro system location in the configuration in Figure 6a and eight ($2^3$) in the configuration in Figure 6b.



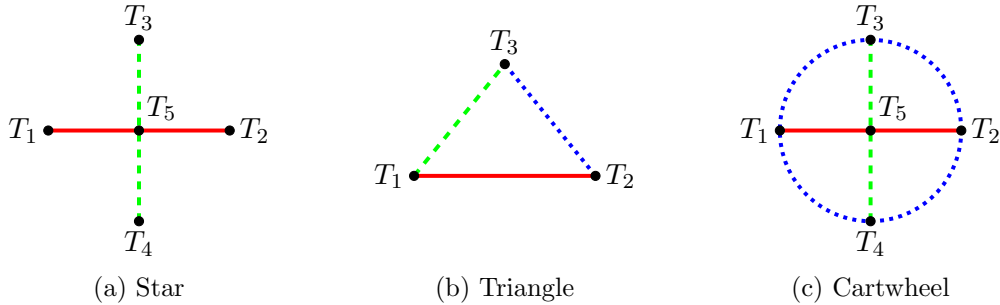(a) Star           (b) Triangle           (c) Cartwheel

Figure 7: Metro networks with the star, the triangle and the cartwheel configurations

Algorithms 2, 3, and 4, summarize the calculations of the metro lines in each corridor, which precede the determination of the best metro line configuration, for the simplified star, triangle and cartwheel schemes depicted in Figure 7. The sets of nodes in the transfer areas are denoted by $T_k$. The computed metro lines between $T_k$ and $T_{k'}$ are stored in set $S_{kk'}$, and the objective value, that is the population coverage, of one such metro line linking nodes $i \in T_k$ and $j \in T_{k'}$ is stored in $d_{ij}^{kk'}$.

When all partial solutions are known, i.e., the best metro lines between transfer arcs in each corridor have been determined, the next step consists of determining a set of lines forming the specified configuration and maximizing $d$ among all similar configurations. As mentioned before, this corresponds to determining the location of the station used as the transfer between lines in the

---

**Algorithm 2:** Star configuration

---

**1 for** $i \in T_5$ **do**

**2**     Compute the best metro line from $i$ to $t_k$, the terminal node in $T_k$, $k = 1, 2, 3, 4$

**3**     Insert the obtained solutions in sets $S_{5k}$, $k = 1, 2, 3, 4$

**4**     Set $d^{5k}_{it_k}$ to be the population coverage of the metro line from $i \in T_5$ to $t_k$, $k = 1, 2, 3, 4$

---

---

**Algorithm 3:** Triangle configuration

---

**1 for** $i \in T_1$ **do**

**2**     Compute the best metro line from $i$ to every node in $T_2 \cup T_3$

**3**     Insert the obtained solutions in sets $S_{1k}$, $k = 2, 3$

**4**     Set $d^{1k}_{ij}$ to be the population coverage of the metro line from $i \in T_1$ to $j \in T_k$, $k = 2, 3$

**5 for** $i \in T_2$ **do**

**6**     Compute the best metro line from $i$ to every node in $T_3$

**7**     Insert the obtained solutions in set $S_{23}$

**8**     Set $d^{23}_{ij}$ to be the population coverage of the metro line from $i \in T_2$ to $j \in T_3$

---

transfer areas. Let $T_k, T_{k'}$ be two linked transfer areas, and let $x^{kk'}_{ij}$ be a variable such that

$$
x^{kk'}_{ij} = \begin{cases} 1, & \text{if and only if the metro line between } T_k, T_{k'} \text{ links nodes } i, j \\ 0, & \text{otherwise} \end{cases} ,
$$

for $i \in T_k$ and $j \in T_{k'}$. In order to determine the best location for the transfer stations, two types of constraints are imposed:

1. One that ensures that only one line is chosen as the link of two transfer areas. Given $T_k$ and $T_{k'}$ the two transfer areas, we have

   (a) $\sum_{i \in T_k} \sum_{j \in T_{k'}} x^{kk'}_{ij} = 1$, in general, and

   (b) $\sum_{i \in T_k} x^{kk'}_{it_{k'}} = 1$, if $T_{k'}$ is a terminal area associated with the dummy node $t_{k'}$.

2. Another to ensure that a unique transfer station is chosen in each transfer area,

   (a) $\sum_{j \in T_{k'}} x^{kk'}_{ij} = \sum_{j \in T_{k''}} x^{kk''}_{ij}$, $i \in T_k$, if $T_k$ is the origin of the metro lines between $T_k$ and $T_{k'}$ and between $T_k$ and $T_{k''}$,

   (b) $\sum_{j \in T_{k'}} x^{kk'}_{ij} = \sum_{j \in T_{k''}} x^{k''k}_{ji}$, $i \in T_k$, if $T_k$ is the origin of the metro line between $T_k$ and $T_{k'}$ and the destination of the metro line between $T_k$ and $T_{k''}$,

   (c) $\sum_{j \in T_{k'}} x^{k'k}_{ji} = \sum_{j \in T_{k''}} x^{k''k}_{ji}$, $i \in T_k$, if $T_k$ is the destination of the metro lines between $T_k$ and $T_{k'}$ and between $T_k$ and $T_{k''}$,

   (d) $\sum_{j \in T_{k'}} x^{k'k}_{ji} = \sum_{j \in T_{k''}} x^{kk''}_{ij}$, $i \in T_k$, if $T_k$ is the destination of the metro line between $T_k$ and $T_{k'}$ and the origin of the metro line between $T_k$ and $T_{k''}$.

This problem can then be formulated as an integer linear program, with the goal of maximizing $d$. This program contains as many constraints of type 1 as the number of corridors of the configuration,

---
**Algorithm 4:** Cartwheel configuration
---

**1 for** $i \in T_5$ **do**

2     Compute the best metro line from $i$ to every node in $T_1 \cup T_2 \cup T_3 \cup T_4$

3     Insert the solutions in sets $S_{5k}$, $k = 1, 2, 3, 4$

4     Set $d_{ij}^{5k}$ to be the population coverage of the metro line from $i \in T_5$ to $j \in T_k$, $k = 1, 2, 3, 4$

**5 for** $i \in T_1$ **do**

6     Compute the best metro line from $i$ to every node in $T_3 \cup T_4$

7     Insert the solutions in sets $S_{1k}$, $k = 3, 4$

8     Set $d_{ij}^{1k}$ to be the population coverage of the metro line from $i \in T_1$ to $j \in T_k$, $k = 3, 4$

**9 for** $i \in T_2$ **do**

10    Compute the best metro line from $i$ to every node in $T_3 \cup T_4$

11    Insert the solutions in sets $S_{2k}$, $k = 3, 4$

12    Set $d_{ij}^{2k}$ to be the population coverage of the metro line from $i \in T_2$ to $j \in T_k$, $k = 3, 4$

---

and the number of constraints of type 2 depends on the number of transfer areas, times the number of nodes each one contains.

Maintaining the previous notation and keeping in mind the simple configurations shown in Figure 7, $T_k$ will represent the set of nodes in a transfer area of the configuration and $t_k$ will be the dummy terminal node associated with $T_k$ in the case of the star configuration. Assuming that Algorithms 2, 3 or 4 were applied in the preceding step, $d_{ij}^{kk'}$ will denote the population coverage of the computed metro line to link nodes $i \in T_k$ and $j \in T_{k'}$.

**Star configuration** In the case of a star configuration (Figure 7a) the only effective transfer area is $T_5$, whereas $T_1, T_2, T_3, T_4$ are used as terminal areas, associated with the terminal dummy nodes $t_1, t_2, t_3, t_4$, respectively:

$$\text{maximize} \quad \sum_{k=1}^{4} \sum_{i \in T_5} d_{it_k}^{5k} x_{it_k}^{5k} \tag{10}$$

$$\text{subject to} \quad \sum_{i \in T_5} x_{it_k}^{5k} = 1, \ k = 1, 2, 3, 4 \tag{11}$$

$$x_{it_1}^{51} = x_{it_2}^{52} = x_{it_3}^{53} = x_{it_4}^{54}, \ i \in T_5 \tag{12}$$

$$x_{it_k}^{5k} \in \mathbb{Z}_0^+, \ i \in T_5, \ k = 1, 2, 3, 4. \tag{13}$$

**Triangle configuration**  The triangle configuration in Figure 7b has three transfer areas, $T_1, T_2, T_3$, and no terminal areas:

$$\text{maximize} \quad \sum_{i \in T_1} \sum_{j \in T_2} d_{ij}^{12} x_{ij}^{12} + \sum_{i \in T_1} \sum_{j \in T_3} d_{ij}^{13} x_{ij}^{13} + \sum_{i \in T_2} \sum_{j \in T_3} d_{ij}^{23} x_{ij}^{23} \tag{14}$$

$$\text{subject to} \quad \sum_{i \in T_1} \sum_{j \in T_k} x_{ij}^{1k} = 1, \ k = 2,3 \tag{15}$$

$$\sum_{i \in T_2} \sum_{j \in T_3} x_{ij}^{23} = 1 \tag{16}$$

$$\sum_{j \in T_2} x_{ij}^{12} = \sum_{j \in T_3} x_{ij}^{13}, \ i \in T_1 \tag{17}$$

$$\sum_{i \in T_1} x_{ij}^{12} = \sum_{k \in T_3} x_{jk}^{23}, \ j \in T_2 \tag{18}$$

$$\sum_{i \in T_1} x_{ij}^{13} = \sum_{i \in T_2} x_{ij}^{23}, \ j \in T_3 \tag{19}$$

$$x_{ij}^{1k} \in \mathbb{Z}_0^+, \ i \in T_1, \ j \in T_k, \ k = 2,3 \tag{20}$$

$$x_{ij}^{23} \in \mathbb{Z}_0^+, \ i \in T_2, \ j \in T_3. \tag{21}$$

**Cartwheel configuration**  The cartwheel configuration in Figure 7c has five transfer areas, $T_1, T_2, T_3, T_4, T_5$, and no terminal areas:

$$\text{maximize} \quad \sum_{k=1}^{4} \sum_{i \in T_5} \sum_{j \in T_k} d_{ij}^{5k} x_{ij}^{5k} + \sum_{k=3,4} \sum_{i \in T_1} \sum_{j \in T_k} d_{ij}^{1k} x_{ij}^{1k} + \sum_{k=3,4} \sum_{i \in T_2} \sum_{j \in T_k} d_{ij}^{2k} x_{ij}^{2k} \tag{22}$$

$$\text{subject to} \quad \sum_{i \in T_5} \sum_{j \in T_k} x_{ij}^{5k} = 1, \ k = 1,2,3,4 \tag{23}$$

$$\sum_{i \in T_k} \sum_{j \in T_{k'}} x_{ij}^{kk'} = 1, \ k = 1,2, \ k' = 3,4 \tag{24}$$

$$\sum_{i \in T_5} x_{ij}^{5k} = \sum_{l \in T_3} x_{jl}^{k3} = \sum_{l \in T_4} x_{jl}^{k4}, \ j \in T_k, \ k = 1,2 \tag{25}$$

$$\sum_{i \in T_5} x_{ij}^{5k} = \sum_{i \in T_1} x_{ij}^{1k} = \sum_{i \in T_2} x_{ij}^{2k}, \ j \in T_k, \ k = 3,4 \tag{26}$$

$$x_{ij}^{5k} \in \mathbb{Z}_0^+, \ i \in T_5, \ j \in T_k, \ k = 1,2,3,4 \tag{27}$$

$$x_{ij}^{kk'} \in \mathbb{Z}_0^+, \ i \in T_k, \ j \in T_{k'}, \ k = 1,2, \ k' = 3,4 \tag{28}$$

$$x_{ij}^{kk'} \in \mathbb{Z}_0^+, \ i \in T_k, \ j \in T_{k'}, \ k = 3,4, \ k' = 1,2. \tag{29}$$

Algorithm 5 summarizes the general procedure for locating the metro lines in a configuration with several corridors when maximizing population coverage.

---

**Algorithm 5:** Single criteria metro lines location in a multi-corridor configuration

**1 for** *each corridor with k points in each ending area* **do**
**2** | Compute the best metro line between each of the $k^2$ pairs of points
**3** Solve the corresponding integer linear program

---

## 4.2 Bicriteria case

When considering both the population coverage and the construction cost as the objective functions, and assuming that these may be conflicting, the concept of optimality is replaced by that of non-dominance (or efficiency). Following Definitions 1 and 2, a solution, in this case a configuration, is said to dominate another one if it is never worse than the latter and it is strictly better for at least one of the criteria, and it is said to be non-dominated if there is no other solution that dominates it. The image of a non-dominated solution, that is, the pair of its objective values, is said to be efficient. Thus, rather than looking for an optimal solution, if $d$ and $c$ are considered, the goal is to compute the set of non-dominated solutions. In such a case, the output of Algorithm 1 is a set of solutions, that is, a set of metro lines, between each pair of nodes at different ends of each corridor. These solutions are non-dominated with respect to others with the same initial and terminal stations. Figure 8 depicts such a situation for simplified star and triangle configurations with two possible station locations in each transfer area.



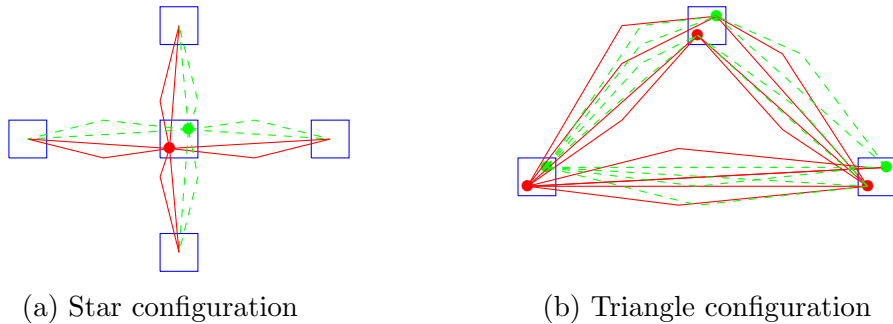(a) Star configuration    (b) Triangle configuration

Figure 8: Bicriteria candidate metro lines

Here $d$ is to be maximized and $c$ is to be minimized. We wish to combine the set of non-dominated solutions with respect to these criteria for each corridor, in a way that one transfer station is located in each transfer area. To this end, we reduce the problem to the case analyzed in Section 4.1 but with the goal of optimizing both $d$ and $c$. This is done by selecting one solution out of the set of non-dominated solutions for each pair of transfer stations. This choice can be made after discussion with the planners. The set of solutions resulting from this procedure is similar to that obtained in the first step of the single criterion version. As a result, the set of constraints of the integer linear program discussed previously still applies, now with two criteria. The integer linear programs to be solved in order to determine the "best" locations for transfer stations are the same as those already presented, but include the minimization of the construction cost as the second criterion. For this reason the problem formulations are not repeated here.

The procedure described above outputs a set of metro line configurations defined by the location of the stations at the ending and transfer areas. These solutions do not dominate each other. However, some may be dominated by solutions that have not been selected. With the aim of diversifying the obtained solutions, the planner can previously decide how many times such a set of solutions should be computed. This parameter is denoted by $T$. The method finishes with the comparison of the solutions found, in order to discard those that are dominated. The pseudo-code

14

is summarized in Algorithm 6.

---

**Algorithm 6:** Bicriteria metro lines location in a multi-corridor configuration

---

**1** **for** *each corridor with $k$ points in each ending area* **do**
**2** $\quad$ Compute the non-dominated metro lines between each of the $k^2$ pairs of points

**3** **repeat**
**4** $\quad$ **for** *each corridor* **do**
**5** $\quad\quad$ **for** *each pair of points in the ending areas of the corridor* **do**
**6** $\quad\quad\quad$ Select a non-dominated path at random

**7** $\quad$ Solve the corresponding bicriteria integer linear program
**8** **until** $T$ *times*
**9** Identify the non-dominated metro line networks amongst the $T$ obtained sets of non-dominated networks

---

Algorithm 6 is the bicriteria version of Algorithm 5 when maximizing population coverage and minimizing construction cost for locating metro lines in a configuration with several corridors. A common technique for solving bicriteria integer linear problems consists of iteratively transforming them into a sequence of single criterion problems. The related methods are known as scalarization methods. Some of these are the weighted-sum method, the $\epsilon$-constraint method and the weighted Chebyshev method. In the first method, the objective functions are linear combinations of the original criteria, and usually the linear combination coefficients reflect the relative weight of each criterion with respect to the others. The solutions calculated by this method are called supported. Varying the weights allows for the determination of all supported efficient solutions. Nevertheless, it is not possible to compute the other type of efficient solutions, called unsupported, using only weighted-sums. In the $\epsilon$-constraint method only one of the criteria is optimized, while the others are added to the problem as inequalities which force them not to be greater than a given parameter $\epsilon > 0$. All efficient solutions can be determined with this method. However, choosing a suitable $\epsilon$ is not easy. The Chebyshev method is a particular case of scalarization by distance, in which a weighted Chebyshev distance to a reference point is minimized. Like the previous method, this approach can determine all types of efficient solutions. Nevertheless, there may be some weakly efficient solutions, that is, feasible solutions such that there is at least another feasible solution with better values for some of the objective functions and equal values for the other objective functions. The determination of such dominated solutions can be prevented by augmenting the new objective function with the addition of an $l_1$-term weighted by a parameter $\rho > 0$. Tuning the parameter $\rho$ in order to avoid computing weakly dominated solutions and still allow the determination of all non-dominated solutions can be challenging. Details on these methods and different variations of methods can be found in [7, 23].

## 5 Computational experiments

The algorithms just introduced were coded in C, and CPLEX 12.4 was used to solve the integer linear programs. In this section, the results obtained by these codes for artificial instances are

discussed. All tests were performed on a Dual Core AMD at 2.7 GHz, with 4Gb of RAM, over Suse 10.3.

In the algorithm implementations the bicriteria binary problems were solved by a weighted-sum method inspired by the choice of weights applied in the non-inferior set estimation (NISE) method, presented in [4]. This method calculates supported non-dominated solutions by solving a sequence of weighted-sum problems and updating their parameters according to the solutions that have been found. The method starts by calculating the lexicographic best solution with respect to each criterion, and then takes pairs of consecutive solutions and, for each, tries to calculate a new one by considering a single criterion weighted-sum problem. Several features of the method have been assessed. First the method was tested for different configurations, namely star, triangle and cartwheel configurations. The following tests were performed for triangle configurations but major differences are not expected for other configurations. Afterwards the effect of the corridor width and of the population density on the type of obtained metro line locations is analyzed. Finally, the resolutions of single and bicriteria problems are compared.

In order to obtain realistic network instances, square grid graphs of the form $[1, \rho] \times [1, \rho]$ were generated, with $\rho \in \{20, 40\}$. The nodes in the network are the intersection between horizontal and vertical lines along the grid, which contains $n = \rho^2$ nodes. Unless otherwise stated, the instances were obtained from random $\rho \times \rho$ grids, for $D = \Delta = 5$, $m = 1$, $M = 3$, and with integer costs uniformly generated in $[1, 1000]$. The population density was set to $p_{ij} = 1$, then for $\lfloor n/100 \rfloor$ times:

1. Select a point and generate a centered square with side of length $2\alpha$, $\alpha \in [1, \rho/10]$.

2. Select $\beta \in [1, 3]$.

3. For $(i, j)$ in the centered square, $p_{ij} \leftarrow p_{ij} + \beta$.

The considered proportion of the population living $i$ units away from the station was $\theta_0 = \theta_1 = 1.00$, $\theta_2 = 0.50$, $\theta_3 = 0.25$, and $\theta_i = 0.00$, for $i \geq 4$.

**Configurations** We first analyze the metro line locations obtained for different network configurations, with star, triangle and cartwheel shapes. Tables 1, 2 and 3 summarize the average values found for 10 instances on this set of random tests by star, triangle and cartwheel configurations, respectively, concerning the running times and the number of solutions for each instance. The tables indicate the partial running times for computing the metro line location along each corridor L$ij$, which links the transfer areas $T_i$ and $T_j$, and the total running times. The maximum number of metro stations depended on $\rho$ and on the configuration.

Table 1: Average results for the location of a metro line with a star configuration on $\rho \times \rho$ grids

| W | CPU time (s) | | | | | | # sol. | W | CPU time (s) | | | | | | # sol. |
|---|------|------|------|------|------|-------|-------|---|------|------|------|------|------|-------|-------|
| | L51 | L52 | L53 | L54 | IP | Total | | | L51 | L52 | L53 | L54 | IP | Total | |
| 3 | 0.001 | 0.002 | 0.001 | 0.001 | 0.569 | 0.574 | 6.7 | 6 | 0.007 | 0.007 | 0.007 | 0.007 | 0.593 | 0.621 | 7.0 |
| 4 | 0.002 | 0.002 | 0.002 | 0.002 | 0.886 | 0.894 | 10.2 | 7 | 0.012 | 0.011 | 0.012 | 0.011 | 0.886 | 0.933 | 10.1 |
| 5 | 0.003 | 0.002 | 0.003 | 0.002 | 0.982 | 0.992 | 11.1 | 8 | 0.017 | 0.017 | 0.016 | 0.016 | 0.741 | 0.806 | 8.2 |

(a) $\rho = 20$        (b) $\rho = 40$

The problem of locating the metro lines in star configurations was the easiest to solve, whereas cartwheel configurations were the most difficult ones to deal with, given that they have more transfer areas than the other configurations and thus lead to larger integer programs. The difficulty of the problem also depends on the value of $W$. For all configurations solving the integer program was the most time consuming phase of the method. Nevertheless, the CPU times for finding the metro line location ranged between 0.355 and 2.434 seconds for $20 \times 20$ grids, and between 0.585 and 10.856 seconds for $40 \times 40$ grids.

Table 2: Average results for the location of a metro line with a triangle configuration on $\rho \times \rho$ grids

| $W$ | CPU time (s) | | | | | # sol. |
|-----|-------|-------|-------|-------|-------|--------|
| | L12 | L13 | L23 | IP | Total | |
| 5 | 0.002 | 0.002 | 0.002 | 0.350 | 0.355 | 2.2 |
| 6 | 0.003 | 0.002 | 0.003 | 0.763 | 0.770 | 4.3 |
| 7 | 0.004 | 0.003 | 0.003 | 1.463 | 1.472 | 5.2 |
| 8 | 0.004 | 0.003 | 0.003 | 2.425 | 2.434 | 6.6 |

(a) $\rho = 20$

| $W$ | CPU time (s) | | | | | # sol. |
|-----|-------|-------|-------|-------|-------|--------|
| | L12 | L13 | L23 | IP | Total | |
| 10 | 0.014 | 0.014 | 0.015 | 0.542 | 0.585 | 3.8 |
| 11 | 0.017 | 0.015 | 0.018 | 1.266 | 1.316 | 5.4 |
| 12 | 0.017 | 0.017 | 0.019 | 2.007 | 2.059 | 5.4 |

(b) $\rho = 40$

Table 3: Average results for the location of a metro line with a cartwheel configuration on $40 \times 40$ grids

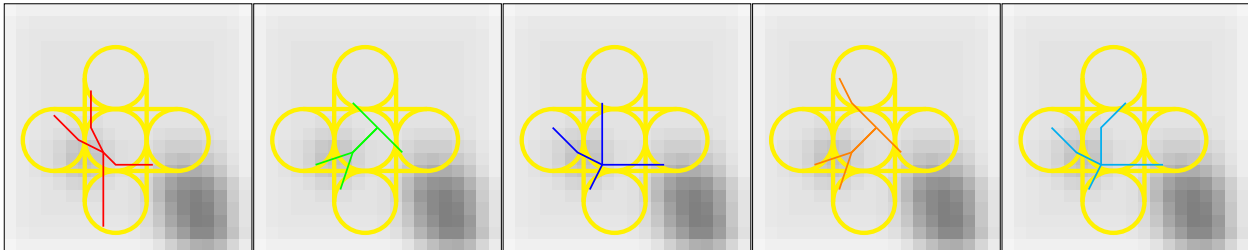| $W$ | CPU time (s) | | | | | | | | | | # sol. |
|-----|-------|-------|-------|-------|-------|-------|-------|-------|--------|--------|--------|
| | L51 | L52 | L53 | L54 | L13 | L14 | L23 | L24 | IP | Total | |
| 8 | 0.031 | 0.031 | 0.031 | 0.030 | 0.021 | 0.022 | 0.018 | 0.018 | 4.339 | 4.540 | 3.9 |
| 9 | 0.037 | 0.037 | 0.037 | 0.036 | 0.035 | 0.035 | 0.030 | 0.030 | 8.313 | 8.591 | 6.9 |
| 10 | 0.048 | 0.047 | 0.047 | 0.047 | 0.056 | 0.056 | 0.054 | 0.055 | 10.445 | 10.856 | 7.9 |



Figure 9: Metro line locations for a star configuration on a $20 \times 20$ grid, $W = 3$

The solutions obtained for particular instances in this set are represented in Figures 9 and 10 for $\rho = 20$, and in Figures 11 to 13 for $\rho = 40$. In the pictures the population density is depicted using different shades of grey. The darker the grade of the grid the higher the population density in the corresponding location. The yellow lines define the corridors on each grid. The non-dominated solution approximations with respect to population coverage and construction cost, obtained for those instances, namely the metro lines between transfer areas, are represented by a different color. The objective values of each solution are shown in Table 4, for the star and the triangle configurations on a $20 \times 20$ grid, and in Table 5 for all three configurations on a $40 \times 40$ grid.

It can be observed that some of the metro lines are the same in different solutions, like Line 23 in the two rightmost pictures of Figure 10, or Line 51 in the three rightmost pictures of Figure 13.
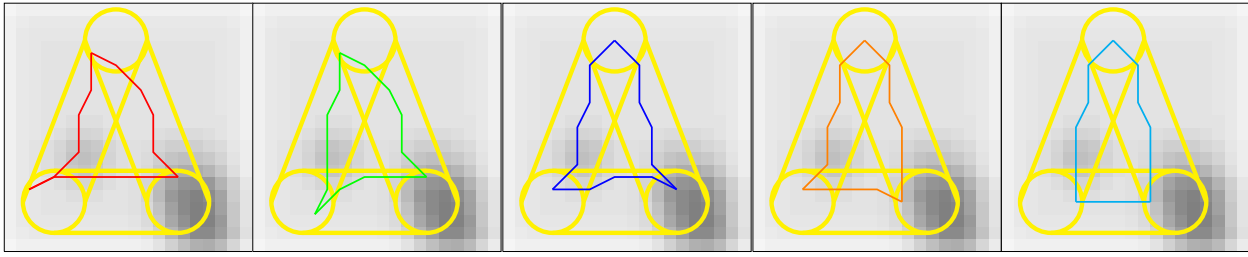
Figure 10: Metro line locations for a triangle configuration on a $20 \times 20$ grid, $W = 6$

Table 4: Objective values, $(d, c)$, of the metro line locations in Figures 10 and 9

| Sol. | L51 | L52 | L53 | L54 | Total |
|------|-----|-----|-----|-----|-------|
| 1 | (56, 77) | (53, 74) | (61, 104) | (58, 133) | (228, 388) |
| 2 | (52, 56) | (48, 11) | (52, 39) | (48, 2) | (200, 108) |
| 3 | (52, 56) | (48, 11) | (47, 18) | (48, 2) | (195, 87) |
| 4 | (51, 6) | (24, 3) | (55, 21) | (55, 27) | (185, 57) |
| 5 | (51, 6) | (24, 3) | (29, 2) | (55, 27) | (159, 38) |

(a) Star, $W = 3$

| Sol. | L12 | L13 | L23 | Total |
|------|-----|-----|-----|-------|
| 1 | (93, 197) | (104, 189) | (85, 306) | (282, 692) |
| 2 | (78, 109) | (113, 235) | (85, 306) | (276, 650) |
| 3 | (80, 151) | (105, 229) | (72, 148) | (257, 528) |
| 4 | (64, 69) | (105, 229) | (72, 150) | (241, 448) |
| 5 | (43, 101) | (108, 183) | (72, 150) | (223, 434) |

(b) Triangle, $W = 6$

However, the solutions found by the methods vary significantly given that they take into account both population coverage and cost. Additionally, the covered area is wider when $\rho = 40$ than when $\rho = 20$. For this reason, when considering the same configuration, the number $W$ of metro stations to construct is larger in the first case.
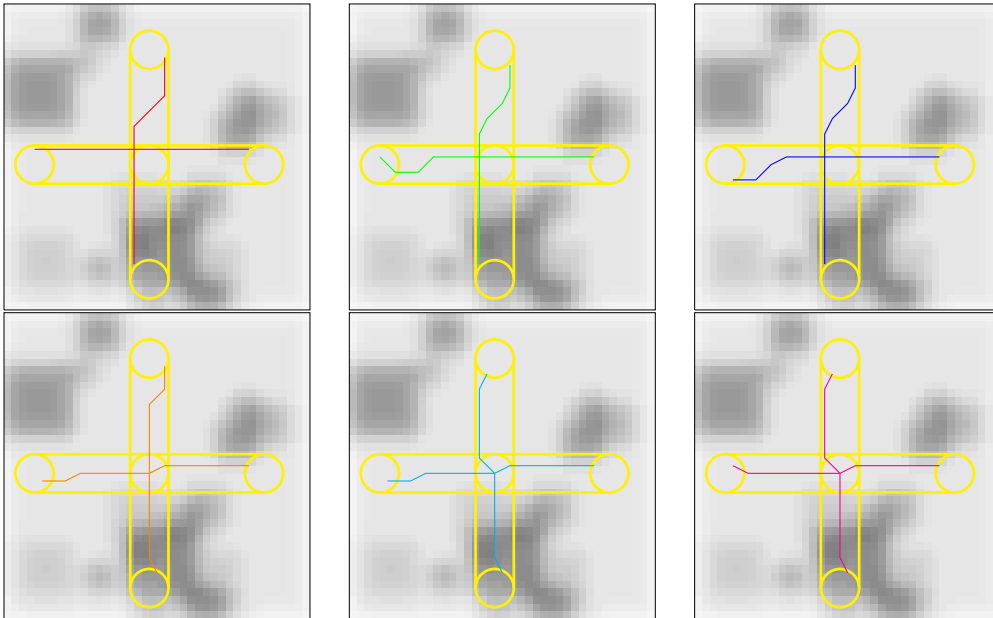


Figure 11: Metro line locations for a star configuration on a $40 \times 40$ grid, $W = 6$

**One and two criteria** The solutions depicted in Figures 9 to 13 and in Tables 4 and 5 are sorted from left to right from the solution that maximizes the population coverage, $d$, to the solution that minimizes the construction cost, $c$. Because these solutions do not dominated one another, both $d$
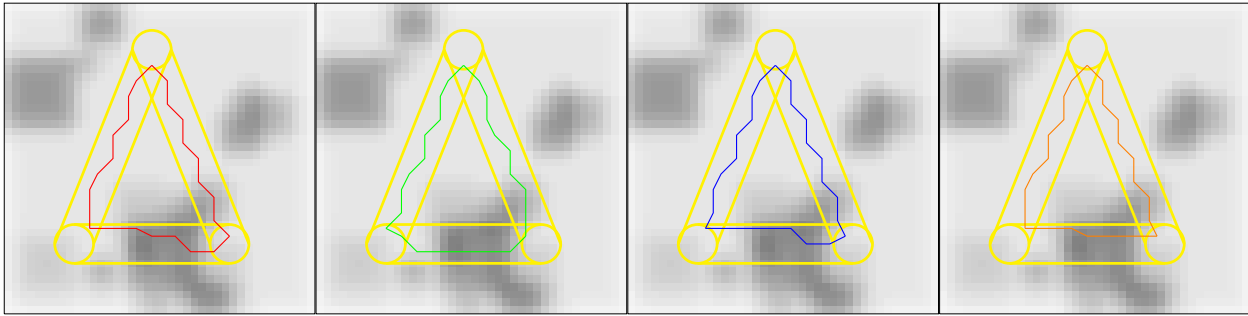
Figure 12: Metro line locations for a triangle configuration on a $40 \times 40$ grid, $W = 10$
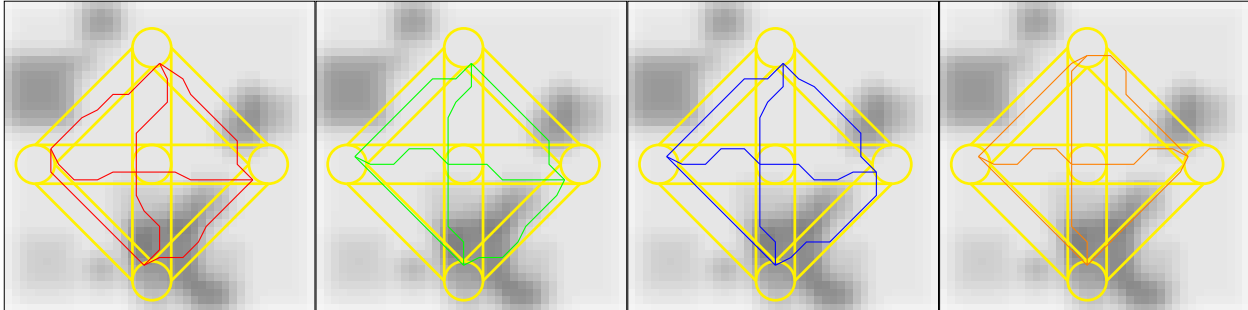


Figure 13: Metro line locations for a cartwheel configuration on a $40 \times 40$ grid, $W = 8$

and $c$ decrease from left to right. The best metro line location found with respect to criterion $d$ is the first solution, whereas it is the last with respect to $c$. For all three configurations metro line locations with intermediate $d$ and $c$ values were obtained, besides the best for each criterion.
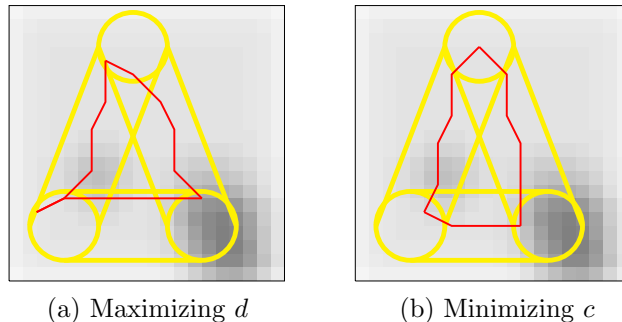


(a) Maximizing $d$        (b) Minimizing $c$

Figure 14: Metro line locations for a triangle configuration on a $20 \times 20$ grid, $W = 6$

It should be recalled that the output of the proposed method is a set of solutions that do not dominate each other, but they may be non-dominated only for the restricted set of solutions selected after all the non-dominated lines between transfer areas have been computed. For instance, the single criterion problem of maximizing $d$ (minimizing $c$) was solved for the triangle configuration and the grids in Figures 10 and 12. The solutions for these problems are shown on Figure 14 and Table 6. In some of the cases the single criterion solutions was also obtained by the bicriteria algorithm (namely, Solution 1 when $\rho = 20$, and Solution 2 when $\rho = 40$), but this was not always the case. In fact, the optimum with respect to $d$ when $\rho = 20$ and the best with respect to $c$ when

19

Table 5: Objective values, $(d, c)$, of the metro line locations in Figures 11, 12 and 13

| Sol. | L51 | L52 | L53 | L54 | Total |
|---|---|---|---|---|---|
| 1 | (248, 320) | (73, 353) | (127, 258) | (79, 320) | (527, 1251) |
| 2 | (246, 251) | (74, 190) | (119, 112) | (80, 296) | (519, 849) |
| 3 | (232, 80) | (74, 190) | (119, 112) | (80, 296) | (505, 678) |
| 4 | (223, 137) | (72, 75) | (107, 217) | (72, 101) | (474, 530) |
| 5 | (223, 137) | (72, 75) | ( 86, 152) | (72, 101) | (453, 465) |
| 6 | (217, 127) | (72, 75) | ( 86, 152) | (72, 101) | (447, 455) |

(a) Star, $W = 6$

| Sol. | L12 | L13 | L23 | Total |
|---|---|---|---|---|
| 1 | (141, 321) | (241, 289) | (157, 389) | (539, 999) |
| 2 | (123, 206) | (242, 302) | (168, 425) | (533, 933) |
| 3 | (111, 169) | (241, 300) | (155, 384) | (507, 853) |
| 4 | (90, 149) | (241, 289) | (155, 384) | (486, 822) |

(b) Triangle, $W = 10$

| Sol. | Line 51 | Line 52 | Line 53 | Line 54 | Line 13 | Line 14 | Line 23 | Line 24 | Total |
|---|---|---|---|---|---|---|---|---|---|
| 1 | (237, 160) | (87, 164) | (139, 244) | (78, 162) | (266, 332) | (222, 386) | (170, 232) | (116, 237) | (1315, 1917) |
| 2 | (236, 139) | (87, 143) | (103, 178) | (80, 133) | (252, 132) | (223, 321) | (155, 323) | (127, 164) | (1263, 1533) |
| 3 | (236, 139) | (87, 162) | (103, 178) | (80, 133) | (252, 132) | (223, 321) | (166, 255) | (109, 169) | (1256, 1489) |
| 4 | (236, 139) | (87, 151) | (88, 174) | (80, 133) | (244, 109) | (223, 321) | (172, 123) | (96, 252) | (1226, 1402) |

(c) Cartwheel, $W = 8$
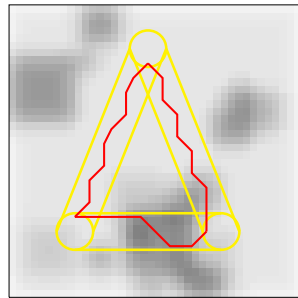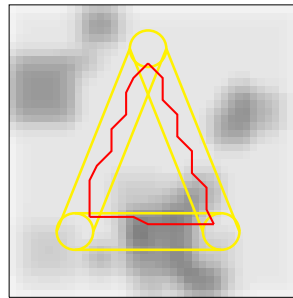


(a) Maximizing $d$    (b) Minimizing $c$

Figure 15: Metro line locations for a triangle configuration on a $40 \times 40$ grid, $W = 10$

$\rho = 40$ were also computed by the proposed method. However, when $\rho = 20$, the optimum for $c$ dominates Solution 5 of Table 10, whereas when $\rho = 40$ the optimum for $d$ does not appear in the list of solutions found by the algorithm.

Table 6: Objective values, $(d, c)$, of the metro line locations in Figures 14 and 15

| Sol. | L12 | L13 | L23 | Total |
|---|---|---|---|---|
| $d^*$ | (93, 197) | (104, 189) | (85, 306) | (282, 692) |
| $c^*$ | (59, 76) | (108, 154) | (72, 150) | (239, 380) |

(a) $\rho = 20$, $W = 6$

| Sol. | L12 | L13 | L23 | Total |
|---|---|---|---|---|
| $d^*$ | (138, 402) | (243, 474) | (162, 390) | (543, 1266) |
| $c^*$ | (90, 149) | (241, 289) | (155, 384) | (486, 822) |

(b) $\rho = 40$, $W = 10$

**Corridor widths** Still using the same grids and the same coverage and cost distributions, the previous problem was solved for triangle configurations and corridors with different widths. Increasing this width when $\rho = 20$ leads to an extensive overlap between the three areas where metro lines are to be located. Instead, the case $\rho = 40$ was considered together with the width $\Delta = 6$, for lines with at most $W = 10, 11$ metro stations. The results can be compared in Figures 12, 16 and 17, and in Tables 5 and 7. In these cases using wider corridors allowed to improve the population coverage and the construction cost of the solutions. Besides, a larger variability of the metro line
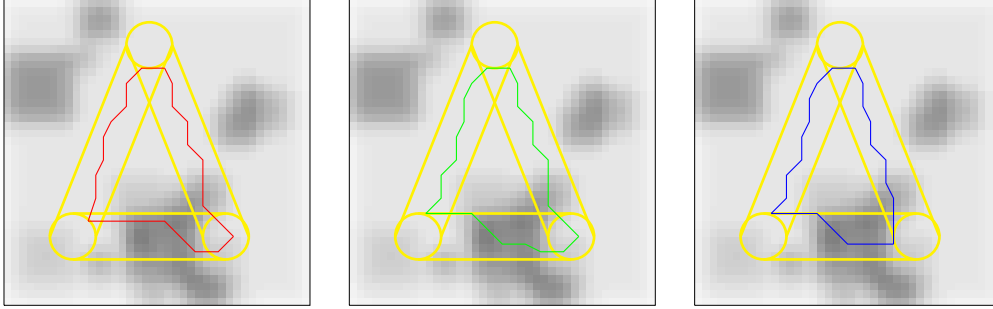
within the corridor can be observed.



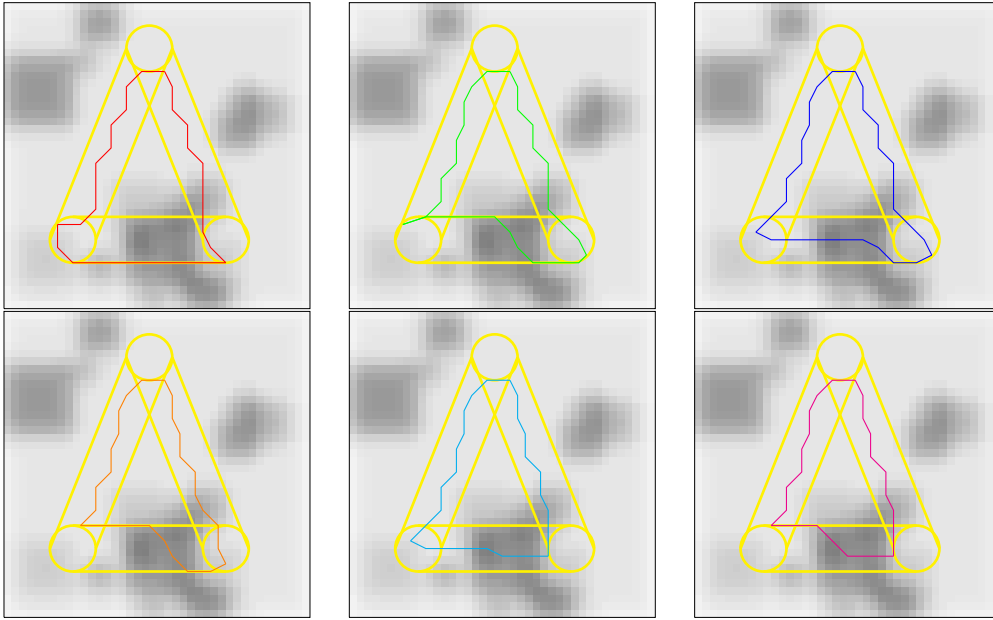Figure 16: Metro line locations for a triangle configuration on a $40 \times 40$ grid, $\Delta = 6$, $W = 10$



Figure 17: Metro line locations for a triangle configuration on a $40 \times 40$ grid, $\Delta = 6$, $W = 11$

**Population density** Additional tests were conducted to compare the metro lines obtained for different, and more realistic, population distributions in the grid. With that goal a single criterion problem was solved aiming only at maximizing the population coverage. The instances represented:

- an area with a uniform population density,

- an area with a population distribution closer to that of a city, with a dense downtown in the center and sparse away from the middle,

- a city with a dense downtown and with a few dense areas away from the center, like suburban areas.

The graphs were $20 \times 20$ grids, with the parameters $D = \Delta = 5$, $m = 1$, and $M = 3$. The three cases considered the following coverage values $d_i$, for any grid node $i$,

Table 7: Objective values, $(d,c)$, of the metro line locations in Figures 16 and 17

| Sol. | Line 12 | Line 13 | Line 23 | Total |
| --- | --- | --- | --- | --- |
| 1 | (173, 238) | (266, 472) | (190, 290) | (629, 1000) |
| 2 | (191, 189) | (247, 301) | (190, 290) | (628, 780) |
| 3 | (115, 65) | (247, 301) | (196, 238) | (558, 604) |

(a) $W = 10$

| Sol. | Line 12 | Line 13 | Line 23 | Total |
| --- | --- | --- | --- | --- |
| 1 | (235, 389) | (275, 394) | (247, 328) | (757, 1111) |
| 2 | (259, 285) | (268, 390) | (226, 305) | (753, 980) |
| 3 | (236, 201) | (260, 311) | (232, 294) | (728, 806) |
| 4 | (199, 166) | (247, 301) | (227, 201) | (673, 668) |
| 5 | (128, 68) | (260, 311) | (196, 238) | (584, 617) |
| 6 | (115, 65) | (247, 301) | (196, 238) | (558, 604) |

(b) $W = 11$

- uniform instances with $d_i = 1$ – Figure 18a,

- uniform instances where $d_i = 1$, and with poles with a higher coverage where $d_i \in \{10, 20, 30\}$ – Figures 18b and 18c.

The pictures show that the design of the metro line is affected by the denser neighborhoods in the grid. The metro lines in the uniform case are almost symmetric, Figure 18a. The scheme obtained when the center is denser than the outskirts is shrunk, that is, compared to the previous solution the lines of this one are closer to the dense center, Figure 18b. Finally, the configuration of the line changes again when there are a dense center and dense suburbs, Figure 18c. The red line on the left is closer to the dark rectangle outside the center, and that slightly affected the shape of the remaining lines.
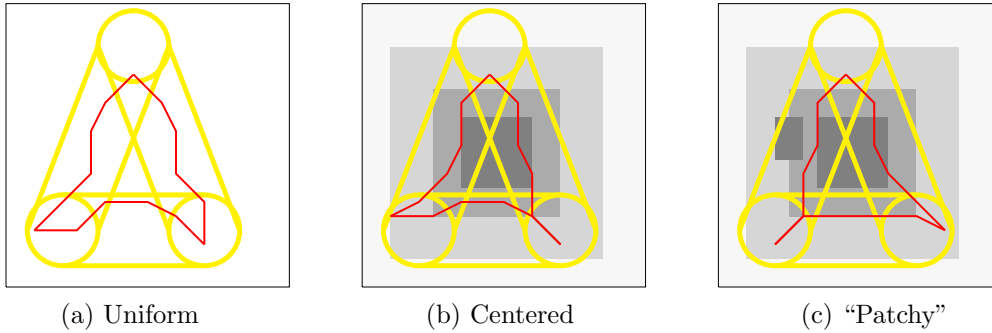


(a) Uniform     (b) Centered     (c) "Patchy"

Figure 18: Metro line locations for a triangle configurations on $20 \times 20$ grids, when maximizing the population coverage, $W = 6$

Table 8: Objective values, $(d,c)$, of the metro line locations in Figure 18

| Sol. | L12 | L13 | L23 | Total |
| --- | --- | --- | --- | --- |
| (a) | (6, 326) | (6, 310) | (6, 353) | (18, 989) |
| (b) | (100, 365) | (100, 245) | (110, 277) | (310, 887) |
| (c) | (110, 361) | (110, 347) | (100, 373) | (320, 1081) |

# 6 Conclusions

In this paper we have proposed a methodology to design and locate a metro line following a given configuration. It is assumed that the planners provide an idea of the corridors where the

metro line should be located. The method consists of locating one metro line within each of the provided corridors and afterwards selecting the way to combine them, that is, the way to select the transfer station locations such that the objective functions are optimized. The metro lines are single alignments computed as constrained paths. The second step consists of combining the set of alignments obtained on the first phase (or part of it), in order to choose the best network design. This is achieved by means of an integer linear program. Firstly only one criterion is considered, either the population coverage or the construction cost. Secondly both criteria are considered simultaneously. The one and the two criteria methods have been tested empirically for a number of instances with star, triangle and cartwheel configurations. The average CPU times for finding solutions varied between 0.355 and 10.856 seconds, and depended on the size of the problem and on the desired configuration. Star configurations were the easiest problems to solve, and cartwheel the most difficult.

# References

[1] J. Brumbaugh-Smith and D. Shier. An empirical investigation of some bicriterion shortest path algorithms. *European Journal of Operational Research*, 43:216–224, 1989.

[2] G. Bruno and G. Laporte. An interactive decision support system for the design of rapid public transit networks. *INFOR*, 40:111–118, 2002.

[3] J. Clímaco and M. M. B. Pascoal. Multicriteria path and tree problems: discussion on exact algorithms and applications. *International Transactions in Operational Research*, 19:63–98, 2012.

[4] J. Cohon. *Multiobjective Programming and Planning*. Academic Press, New York, 1978.

[5] K.M. Curtin and S. Biba. The transit route arc-node service maximization problem. *European Journal of Operational Research*, 208:46–56, 2011.

[6] H. Dufourd, M. Gendreau and G. Laporte. Locating a transit line using tabu search. *Location Science*, 4:1–19, 1996.

[7] M. Ehrgott and X. Gandibleux. *Multiple Criteria Optimization. State of the Art Annotated Bibliographic Surveys*. Kluwer Academic, Dordrecht, 2002.

[8] L.F. Escudero and S. Muoz. An approach for solving a modification of the extended rapid transit network design problem. *TOP*, 17:320–334, 2009.

[9] F. Guerriero and R. Musmanno. Label correcting methods to solve multicriteria shortest path problems. *Journal of Optimization Theory and Applications*, 111:589–613, 2001.

[10] G. Gutiérrez-Jarpa, C. Obreque, G. Laporte, and V. Marianov. Rapid transit network design for optimal cost and origin-destination demand capture. *Computers & Operations Research*, 40:3000–3009, 2013.

[11] G. Laporte and J.A. Mesa. The design of rapid transit networks. In G. Laporte, S. Nickel, and F. Saldanha da Gama, editors, *Location Science.* Springer, Berlin Heidelberg, 241–255, 2015.

[12] G. Laporte, J. A. Mesa, and F. A. Ortega. Assessing topological configurations for rapid transit networks. *Studies in Locational Analysis*, 7:105–121, 1994.

[13] G. Laporte, J. A. Mesa, and F. A. Ortega. Assessing the efficiency of rapid transit configurations. *TOP*, 5:95–104, 1997.

[14] G. Laporte, J. A. Mesa, F. A. Ortega, and F. Perea. Planning rapid transit networks. *Socio-Economic Planning Sciences*, 45(3):95–104, 2011.

[15] G. Laporte, J. A. Mesa, F. A. Ortega, and I. Sevillano. Maximizing trip coverage in the location of a single rapid transit alignment. *Annals of Operations Research*, 136:49–63, 2005.

[16] G. Laporte, A. Marín, J.A. Mesa, and F. Perea. Designing robust rapid transit networks with alternative routes. *Journal of Advanced Transportation*, 45:54–65, 2011.

[17] G. Laporte and M. M. B. Pascoal. Minimum cost path problems with relays. *Computers & Operations Research*, 38(1):165–173, 2011.

[18] T.C. Matisziw, A.T. Murray, and C. Kim. Strategic route extension in transit networks. *European Journal of Operational Research*, 171:661–673, 2006.

[19] A. Marín. An extension to rapid transit network design problem. *TOP*, 15:231–241, 2007.

[20] A. Marín and R. García Ródenas. Location of infrastructure in urban railway networks. *Computers & Operations Research*, 36:1461–1477, 2009.

[21] A. Raith and M. Ehrgott, A comparison of solution strategies for biobjective shortest path problems. *Computers & Operations Research*, 36:1299–1331, 2009.

[22] A. Skriver and K. Andersen. A label correcting approach for solving bicriterion shortest-path problems. *Computers & Operations Research*, 27:507–524, 2000.

[23] R. Steuer. *Multiple Criteria Optimization: Theory, Computation and Applicattion.* Wiley, New York, 1986.

[24] V. R. Vuchic. *Urban Transit Operations, Planning and Economics.* Wiley, Hoboken, New Jersey, 2005.