Designing Rapid Transit Networks from the Results of a Survey

L.F. Escudero [a], S. Muñoz [b,*]

[a] *Departamento de Estadística e Investigación Operativa, Universidad Rey Juan Carlos,*
28933 Móstoles (Madrid), Spain

E-mail: laureano.escudero@urjc.es

[b] *Departamento de Estadística e Investigación Operativa, Facultad de Ciencias Matemáticas,*
Universidad Complutense de Madrid, 28040 Madrid, Spain

E-mail: smunoz@estad.ucm.es

Abstract

In this work we present a two-stage approach for designing rapid transit networks. It is based on another approach that we described elsewhere. In the first stage, the stations and links to be constructed are selected by solving an integer linear programming model that maximizes an estimation of the number of trips through the rapid transit network. In the second stage, a set of lines is generated by utilizing a greedy heuristic procedure that, taking into consideration the transfers that should be made by the users to arrive at their destinations, attempts to maximize a more accurate estimation for the number of trips. This new estimation is done by means of a modification of the well-known Floyd-Warshall algorithm. The main contributions are a novel way of computing the expected number of trips by making use of the results from a survey amongst the potential users of the rapid transit network, as well as the contemplation of the possibility of linking certain pairs of station locations by more than one line. Some computational experiments on several randomly generated instances are also reported.

1 Introduction

As the population concentration increases in urban areas, it gets necessary to either develop new transportation systems or to improve and/or expand the existing ones. There are so many factors to be taken into consideration to tackle these problems, that the resulting mathematical programming models would be too complex to be solved in an exact way. Consequently, it is required to resort to simplifications and heuristic procedures.

---

*Corresponding author

Several authors have dealt with these problems, focussing them mainly on two opposite aims: to achieve a high service quality with affordable operating costs, or to reduce as much as possible the operating costs while maintaining a certain service quality level. Herein we focus on the first aim; some works focussing on the second aim are Claessens et al. (1998), Bussieck et al. (2004) and Goossens et al. (2004).

In order to illustrate the great diversity of the approaches that can be found in the literature, some works are outlined below. A point in common for all of them is the consideration of a static origin-destination (O-D) matrix which contains the demand for each O-D pair of station locations.

Mandl (1980) proposes a heuristic algorithm for improving an existing public transportation network of streets or rails in an urban area, in case of absence of capacity constraints for utilizing the network. The algorithm starts with a feasible set of lines (i.e., a set of lines such that all station locations belong to at least one line and all pairs of station locations are mutually reachable), and iteratively searches for new feasible sets of lines which lead to a reduction of the total average transportation cost of the passengers. The transportation cost is defined as a weighted sum of the waiting, travel and transfer costs, which can be interpreted as time, and it is assumed that each passenger utilizes a path that minimizes his or her average transportation cost. The total average transportation cost is estimated from a given O-D matrix. The vehicle assignment problem is also briefly discussed.

Given a set of potential bus station locations and a set of potential links between them, and given a symmetric O-D matrix, Baaj and Mahmassani (1991) describes an Artificial Intelligence-based approach for determining a set of lines and their associated frequencies attempting to reduce the number of passengers that require any transfer to arrive at their destinations, so that both the percentages of passengers that require no transfer and of those that require a maximum of two transfers, are greater or equal to certain prespecified values. The approach consists of three major components: a line generation design algorithm, an analysis procedure, and a line improvement algorithm. It is assumed that each passenger utilizes a path that involves the fewest possible number of transfers.

Given a set of potential lines for a railway system, Bussieck et al. (1997) provides an integer linear programming model for selecting a subset of these lines and determining their frequencies, with the goal of maximizing the number of passengers that require no transfer to arrive at their destinations. It assumes the symmetry of the O-D matrix as well as that each passenger utilizes a shortest path (with respect to some measure) between his or her origin and destination.

Guan et al. (2006) proposes a 0-1 linear programming model for simultaneous optimization of transit line configuration and passenger line assignment. Starting from a given set of station locations and links between them, its aim is to select the lines to be constructed and to assign a path in the resulting network to the passengers of each O-D pair of station locations, in such a way that the union of the selected lines contains all the given links, and that a weighted sum

of the total length of the lines, the total number of lines used by the passengers and the total distance covered by the passengers is minimized. For this purpose, a pool of potential lines to be selected and a pool of potential paths to be assigned to the passengers of each O-D pair are considered, and each one of these potential lines is assigned a prefixed frequency. Moreover, it is assumed that the cost of operating any line is linearly proportional to the length of that line, that each passenger utilizes a path that minimizes his or her expected travel time, and that finding the smallest number of transfers for the path assigned to the passengers of each O-D pair is equivalent to finding the smallest number of lines that those passengers should use. Neither the waiting time for the passengers nor the effect of passenger crowding are taken into consideration.

Marín (2007) states the extended rapid transit network design problem and provides a 0-1 linear programming model for solving it. Given a set of potential station locations and a set of potential links between them, this problem basically consists in selecting which stations and links to construct without exceeding the available budget, and determining an upper bounded number of noncircular lines from them, to maximize the expected total number of trips through the rapid transit network, which is computed from a given O-D matrix and a given private transportation cost for each O-D pair of station locations. It is assumed that each user will utilize the rapid transit network if and only if there is any path in this network between his or her origin and destination such that its length is less or equal to the corresponding transportation cost in the private transit network. Similar models are considered in Laporte et al. (2010a, 2010b).

Marín and García-Ródenas (2009) presents a nonlinear programming model for locating the infrastructure of a rapid transit network without exceeding the available budget. Two alternative objective functions are proposed, namely, the expected total number of trips through the rapid transit network (to be maximized), and the expected total transportation cost through an existing private transit network (to be minimized). Both of them are defined from a given O-D matrix, a given private transportation cost for each O-D pair of station locations, and the Logit function, which is approximated by a piecewise linear function. As a consequence of this approximation, the initial nonlinear model results in an integer linear programming model. Among the considered assumptions to simplify the model are that there is no waiting time for the users and there are no capacity constraints for utilizing the network. The model also includes some constraints that avoid the definition of circular lines, and others that attempt to minimize the number of lines to be constructed. The potential users' behavior is modeled by means of the Logit function, instead of considering the all-or-nothing model from Marín (2007).

Escudero and Muñoz (2009a) provides a two-stage approach for solving a modification of the extended rapid transit network design problem to allow the definition of circular lines, and shows that it outperforms the solving of a modification of the model given in Marín (2007) to adapt it to this new problem. In the first stage of the proposed approach, an integer linear programming model is solved for selecting the stations and links to be constructed without exceeding the available

budget, so that the expected total number of trips through the rapid transit network is maximized (without loss of generality, it is assumed that whichever two station locations are linked by one line at most). In the second stage the line design problem is solved by assigning each selected link to exactly one line, in such a way that the number of lines that go to each selected station location is as small as possible (no upper bound for the number of lines is required).

Escudero and Muñoz (2009b) proposes some improvements on the approach stated in Escudero and Muñoz (2009a). On one side, it introduces several modifications in the model considered in the first stage to obtain a connected rapid transit network. On other side, it presents a greedy heuristic procedure which is a modification of the algorithm proposed for solving the line design problem of the second stage. This new procedure attempts to minimize an estimation of the total number of transfers that should be made by the users to arrive at their destinations, without increasing the number of lines going to each selected station location.

In this work we tackle the problem of designing a rapid transit network, i.e., determining the stations and links to be constructed, as well as the set of lines. The two-stage approach that we present is based on the one given in Escudero and Muñoz (2009b), and it is structured in the same manner.

In the first stage, an integer linear programming problem is solved for selecting the stations to be constructed and the links between them, considering a budget for the total construction cost. All of the station locations are assumed to be known, but we distinguish between key and non-key stations: the key stations have to be compulsorily constructed and they may belong to more than one line, whereas the non-key stations are always located on some link joining two key stations, and they are constructed if and only if that link is constructed.

Obviously, the solution obtained will strongly depend on the objective function considered; thus, an appropriate choice of the objective function is crucial for getting a successful rapid transit network. Among the different objective functions considered in the works outlined above, the one more directly related to the service quality is the expected total number of trips through the rapid transit network, since the higher the service quality, the greater the total number of trips. Therefore, we are considering it as the objective function (to be maximized).

This type of objective function has already been considered in Marín (2007), Marín and García-Ródenas (2009), Escudero and Muñoz (2009a, 2009b) and Laporte et al. (2010a, 2010b), and its value has been computed by considering a unique transportation cost for each O-D pair of station locations in an alternative transit network. This way of computing the objective function value is not very accurate, since the users of each O-D pair can actually utilize distinct means of transportation and distinct routes for arriving at their destinations, hence it does not seem adequate to consider the same alternative transportation cost for all of them. Instead, we propose to perform a survey in order to collect certain data which will make it possible to consider each potential

user's behavior individually and, as a consequence, to compute the objective function value more accurately.

In the second stage of the presented approach, the line design problem is solved by means of a generalization of the greedy heuristic procedure given in Escudero and Muñoz (2009b) to allow pairs of station locations linked by more than one line, attempting to maximize the total number of trips through the rapid transit network. For this purpose, we introduce a modification of the well-known Floyd-Warshall algorithm to determine the shortest routes for each O-D pair of key station locations, and we define an expanded network that will make it possible to consider the transfer times for the users.

We are not addressing the problem of determining the headways for the lines, since these headways must vary over time, depending on different factors such as the available budget for the operating costs, whether we are in a peak hour or in an off-peak hour, in a working day or in a holiday, in a working week or in Christmas, Easter or summer holidays, whether a mass event (e.g., a concert, a conference, a demonstration, a football game, ...) is going to be held, etc. Consequently, the headways will have to be set in subsequent stages, by taking into consideration the circumstances at those moments as well as the trade-off between service quality and operating costs.

The remainder of the paper is organized as follows: Section 2 states the basic notation and assumptions that we consider. Section 3 proposes an integer linear programming model for selecting the stations and links to be constructed. Section 4 presents a modification of the Floyd-Warshall algorithm to determine a shortest chain between each pair of nodes of a graph with nonnegative length edges. Section 5 shows how to calculate the maximum possible expected number of trips taken on the rapid transit network between each pair of key station locations, assuming that no transfers are required, as well as that the capacity of the rapid transit network is enough to hold all those trips. Section 6 provides an algorithm for designing a set of lines for the rapid transit network which generalizes the greedy heuristic algorithm given in Escudero and Muñoz (2009b), to allow pairs of station locations linked by more than one line. Section 7 shows how to calculate in a more accurate way the maximum possible expected number of trips obtained in Section 5, by means of an expansion of the network considered therein that allows to take into account the transfer times. Section 8 proposes a greedy heuristic procedure for determining a line design for the rapid transit network, attempting to maximize the expected total number of trips through the rapid transit network. Section 9 reports some computational experience on three randomly generated example cases; the results show that the approach presented in this paper can significatively increase the expected total number of trips through the rapid transit network obtained by utilizing the procedure proposed in Escudero and Muñoz (2009b). Finally, Section 10 draws some conclusions and future research from this work.

2 Basic notation and assumptions

Let us consider two types of stations: key stations and non-key stations. The key stations will be located on the busiest zones of the area covered by the rapid transit network, which are assumed to be known (see in Laporte et al. (2007) a procedure for selecting such key station locations). We also assume that the potential links between the key station locations are known, and that some other stations, called non-key stations, can be located on those links, in such a way that each non-key station will be constructed if and only if the link on which it lies is constructed.

The key station locations will be represented as the nodes of a graph, and the potential links between them as the edges of that graph (it is not necessary to represent the non-key station locations). Thus, we are implicitly assuming that, for each pair of distinct key station locations that can be linked, the route followed by the users for going from one of the locations to the other one will be the same as for going from the second location to the first one, but in the opposite direction. Although this assumption is usually satisfied for rail networks, it can be violated for street networks containing one-way streets, but the approach presented below will remain valid for this case by considering arcs instead of edges and by modifying it accordingly.

Let $V = \{1,\ldots,n\}$ be the set of key station locations, let E be the set of (nonordered) pairs of key station locations that can potentially be linked, i.e.,

$$E = \{\{i,j\} \in V \times V \mid i \neq j \text{ and it is possible to link } i \text{ and } j\},$$

and let $m = |E|$.

Let us consider the simple graph $G = (V, E)$. Without loss of generality, whenever we refer to an edge $\{i,j\} \in E$ it will be assumed that $i < j$. We also assume that G is connected.

For each $i \in V$, let a_i be the cost of constructing a key station at i, and let $\Gamma(i)$ be the set of key station locations that can be linked to i (notice that $\Gamma(i)$ is the set of nodes adjacent to i in G and $|\Gamma(i)|$ is the degree of i in G).

For each $\{i,j\} \in E$, let d_{ij} be the length of link $\{i,j\}$ (expressed in kilometers), let s_{ij} be the number of non-key station locations on link $\{i,j\}$, and let c_{ij} be the cost of linking i and j (including the cost of constructing the corresponding non-key stations).

If there were λ lines going to a key station location i or linking two key station locations i and j, then the associated construction costs would be λa_i and λc_{ij}, respectively, since it is assumed that we construct as many stations at i and as many links between i and j as the number of lines involved. Although in the first stage of the proposed approach we implicitly consider that whichever two station locations are linked by one line at most (see Section 3), in the second stage we shall allow pairs of station locations linked by more than one line (see Sections 6, 7 and 8).

Let b be the available budget for constructing the rapid transit network, and let $\bar{v}$ be the average velocity of the network's vehicles (expressed in kilometers per hour).

For each $i \in V$, let $\bar{t}(i)$ be the average time required for going between the entrance of the key station located at i and its boarding and alighting platform (expressed in minutes).

Let $\bar{t}_a$ be the average interarrival time (i.e., the time difference between two consecutive arrivals) of the vehicles at each station (expressed in minutes), let $\bar{t}_s$ be the average dwell time (i.e., the time spent for boarding and alighting of passengers) of the vehicles at each station (expressed in minutes), and let $\bar{t}_r$ be the average time for making a transfer (expressed in minutes). Since in real-life rapid transit networks most of the time there are no vehicles at the stations, it will be assumed that $\bar{t}_s < \frac{\bar{t}_a}{2}$. If we had appropriate a priori information, instead of considering a unique value for $\bar{t}_r$, we could make it depend on the key station locations, i.e., we could consider the average time for making a transfer at each $i \in V$.

Let $W = \{(i,j) \in V \times V \mid i < j\}$, and let us denote $w = (e_w, e'_w) \quad \forall w \in W$ (W is understood as the set of all distinct pairs of key station locations).

In order to assess the behavior pattern of the potential users, we propose to survey a sample of people who, a priori, are willing to utilize the rapid transit network. Let Θ be the set of surveyed people.

For each $\theta \in \Theta$ and for each $w \in W$, let $\alpha_w(\theta)$ be the number of trips in a working week that the surveyee θ plans to take between e_w and e'_w (in any direction) during the hours of operation of the rapid transit network.

Let $W_\theta = \{w \in W \mid \alpha_w(\theta) > 0\} \quad \forall \theta \in \Theta$ (notice that W_θ is the set of distinct pairs of key station locations between which the surveyee θ plans to travel, without taking into consideration the direction of the related trips).

For each $\theta \in \Theta$ and for each $w \in W_\theta$, let $\tau_w(\theta)$ be the maximum number of minutes that the surveyee θ is willing to spend for travelling between e_w and e'_w (in any direction). We are considering $\bar{t}(e_w) + \frac{\bar{t}_a}{2} + \frac{60}{\bar{v}} d_{e_w e'_w} + \bar{t}(e'_w)$ as the minimum possible average time that the trip between e_w and e'_w will take on the rapid transit network, where $d_{e_w e'_w}$ is the Euclidean distance (expressed in kilometers) between e_w and e'_w if $\{e_w, e'_w\} \notin E$ (this minimum will be reached if $\{e_w, e'_w\} \in E$, $s_{e_w e'_w} = 0$ and link $\{e_w, e'_w\}$ is selected to be constructed). Therefore, without loss of generality it will be assumed that $\tau_w(\theta) \geq \bar{t}(e_w) + \frac{\bar{t}_a}{2} + \frac{60}{\bar{v}} d_{e_w e'_w} + \bar{t}(e'_w)$ (if $\tau_w(\theta) < \bar{t}(e_w) + \frac{\bar{t}_a}{2} + \frac{60}{\bar{v}} d_{e_w e'_w} + \bar{t}(e'_w)$, then the surveyee θ will not utilize the rapid transit network for travelling between e_w and e'_w, and, consequently, we shall set $\alpha_w(\theta) = 0$, hence $w \notin W_\theta$ and the value of $\tau_w(\theta)$ will not be considered).

In order to determine an initial setting for the headways in a subsequent stage, the surveyees could also be asked about the starting time and the direction of their trips.

Let $W' = \bigcup_{\theta \in \Theta} W_\theta$ (notice that W' is the set of distinct pairs of key station locations between which the surveyees plan to travel, without taking into consideration the direction of the related trips). It is expected that $W' = W$, since the cardinality of Θ should be large enough for the survey results to be reliable.

For each $w \in W'$, let $\Theta_w = \{\theta \in \Theta \mid \alpha_w(\theta) > 0\}$, $q(w) \in \mathbb{N}$, $u_w^1 = \min\{\tau_w(\theta) \mid \theta \in \Theta_w\}$, $u_w^2, \ldots, u_w^{q(w)} \in \mathbb{R}$ such that $u_w^1 < u_w^2 < \ldots < u_w^{q(w)} \leq \max\{\tau_w(\theta) \mid \theta \in \Theta_w\}$, and $g_w^k = \sum_{\theta \in \Theta_w^k} \alpha_w(\theta)$ $\forall k \in \{1, \ldots, q(w)\}$, where $\Theta_w^k = \{\theta \in \Theta_w \mid \tau_w(\theta) \geq u_w^k\}$ (the value of g_w^k can be interpreted as the expected number of weekly trips taken on the rapid transit network by the surveyees that plan to travel between e_w and e_w', assuming that the fastest route for taking them takes u_w^k minutes, as well as that the capacity of the rapid transit network is enough to hold all those trips). It will also be assumed that the values of $q(w)$ and $\{u_w^k\}_{k \in \{2, \ldots, q(w)\}}$ have been set in such a way that $g_w^1 > g_w^2 > \ldots > g_w^{q(w)}$ (notice that $g_w^{q(w)} > 0$). The idea behind the consideration of $\{u_w^k\}_{k \in \{1, \ldots, q(w)\}}$ is to group the values $\{\tau_w(\theta)\}_{\theta \in \Theta_w}$ into the intervals $[u_w^1, u_w^2), \ldots, [u_w^{q(w)-1}, u_w^{q(w)}), [u_w^{q(w)}, +\infty)$, and to utilize these intervals for estimating the number of weekly trips that the surveyees in Θ_w will take on the rapid transit network.

3   Station and link location

In this section we present an integer linear programming model for selecting the stations to be constructed and the links between them, so that the resulting rapid transit network is connected and its construction cost does not exceed the available budget. This model is based on the one proposed in Escudero and Muñoz (2009b), for the particular case where at least one key station has to be constructed at each location; see also Escudero and Muñoz (2009a).

We attach more importance to linking two station locations by one line, than not linking them in exchange of linking some two other station locations by more than one line. Thus, in this first stage of the approach we implicitly assume that whichever two station locations are linked by one line at most, whereas in the second stage we shall check whether, without eliminating the already selected links, it is possible and advisable to have pairs of station locations linked by more than one line (see Sections 6, 7 and 8).

Each feasible solution to the model will define a route for travelling between each pair of key station locations. These routes are understood as preliminary routes for taking the trips on the rapid transit network (it will be attempted to improve them in a subsequent phase of the approach; see Section 5).

The optimization criterion is the maximization of an estimation of the number of weekly trips that the surveyees will take on the rapid transit network, which is equivalent to maximizing an estimation of the gross profits, assuming that the users have to buy a ticket per trip and that there is a unique fare for the tickets. In order to do this estimation, we consider the average time for the preliminary routes for taking the trips demanded by the surveyees (since the lines have not still been defined at this stage, these average times are calculated assuming that no transfers are required).

We define the following variables:

$$x_{ij} = \begin{cases} 1 & \text{if } i \text{ and } j \text{ are linked} \\ 0 & \text{otherwise} \end{cases} \quad \forall \{i,j\} \in E$$

$$\gamma_i = \begin{cases} 1 & \text{if } \sum_{j\in\Gamma(i),j>i} x_{ij} + \sum_{j\in\Gamma(i),j<i} x_{ji} \text{ is odd} \\ 0 & \text{otherwise} \end{cases} \quad \forall i \in V$$

$$\Delta_i \in \{0,\dots,r(i)\} \quad \forall i \in V$$

$$f_{ij}^w = \begin{cases} 1 & \text{if the preliminary route for travelling between} \\ & \quad e_w \text{ and } e'_w \text{ utilizes edge } \{i,j\} \\ 0 & \text{otherwise} \end{cases} \quad \forall w \in W, \forall \{i,j\} \in E$$

$$\varepsilon_i^w = \begin{cases} 1 & \text{if the preliminary route for travelling between} \\ & \quad e_w \text{ and } e'_w \text{ passes through } i \\ 0 & \text{otherwise} \end{cases} \quad \forall w \in W, \forall i \in V \setminus \{e_w, e'_w\}$$

$$p_w^k \in \{0,1\} \quad \forall w \in W', \forall k \in \{1,\dots,q(w)\},$$

where $r(i) = \begin{cases} \frac{|\Gamma(i)|}{2} & \text{if } |\Gamma(i)| \text{ is even} \\ \frac{|\Gamma(i)|-1}{2} & \text{if } |\Gamma(i)| \text{ is odd} \end{cases}$, $\Delta_i = \frac{\sum_{j\in\Gamma(i),j>i} x_{ij} + \sum_{j\in\Gamma(i),j<i} x_{ji} - \gamma_i}{2}$, and a necessary condition for p_w^k to take the value 1 is that the average time for travelling between e_w and e'_w by following the associated preliminary route is less or equal to u_w^k.

We propose the following model:

$$\text{Maximize} \quad z = \sum_{w\in W'} \sum_{k=1}^{q(w)} g_w^k p_w^k$$

subject to:

$$\sum_{j\in\Gamma(i),j>i} x_{ij} + \sum_{j\in\Gamma(i),j<i} x_{ji} = 2\Delta_i + \gamma_i \quad \forall i \in V \tag{1}$$

$$\sum_{i\in V} a_i(\Delta_i + \gamma_i) + \sum_{\{i,j\}\in E} c_{ij} x_{ij} \leq b \tag{2}$$

$$f_{ij}^w \leq x_{ij} \quad \forall w \in W, \forall \{i,j\} \in E \tag{3}$$

$$\sum_{j\in\Gamma(i),j>i} f_{ij}^w + \sum_{j\in\Gamma(i),j<i} f_{ji}^w = \begin{cases} 1 & \text{if } i \in \{e_w, e'_w\} \\ 2\varepsilon_i^w & \text{otherwise} \end{cases} \quad \forall w \in W, \forall i \in V \tag{4}$$

9

$$\sum_{k=1}^{q(w)} p_w^k \leq 1 \quad \forall w \in W' \tag{5}$$

$$\bar{t}'(w) + \sum_{\{i,j\} \in E} \bar{t}'_{ij} f_{ij}^w - u_w^k \leq M_w^k (1 - p_w^k) \quad \forall w \in W', \forall k \in \{1, \dots, q(w)\} \tag{6}$$

$$x_{ij} \in \{0,1\} \quad \forall \{i,j\} \in E$$

$$\gamma_i \in \{0,1\} \quad \forall i \in V$$

$$\Delta_i \in \{0, \dots, r(i)\} \quad \forall i \in V$$

$$f_{ij}^w \in \{0,1\} \quad \forall w \in W, \forall \{i,j\} \in E$$

$$\varepsilon_i^w \in \{0,1\} \quad \forall w \in W, \forall i \in V \setminus \{e_w, e'_w\}$$

$$p_w^k \in \{0,1\} \quad \forall w \in W', \forall k \in \{1, \dots, q(w)\},$$

where $\bar{t}'(w) = \bar{t}(e_w) + \frac{\bar{t}_a}{2} - \bar{t}_s + \bar{t}(e'_w)$, $\bar{t}'_{ij} = \frac{60}{\bar{v}} d_{ij} + \bar{t}_s(s_{ij} + 1)$, and M_w^k is an upper bound for the value of $\bar{t}'(w) + \sum_{\{i,j\} \in E} \bar{t}'_{ij} f_{ij}^w - u_w^k$ over the feasible region of the above model relaxation resulting from eliminating the constraints (5) and (6), and the variables $\{p_w^k\}_{w \in W', k \in \{1, \dots, q(w)\}}$.

Constraints (1) and (2) impose the budget constraint. Constraints (3) and (4) define the preliminary routes and guarantee that any feasible solution to this model will give rise to a connected rapid transit network. Constraints (5) impose that, for each $w \in W'$, at most one of the variables $\{p_w^k\}_{k \in \{1, \dots, q(w)\}}$ can be equal to 1. Constraints (6) impose that, for each $w \in W'$ and for each $k \in \{1, \dots, q(w)\}$, if $p_w^k = 1$ then $\bar{t}'(w) + \sum_{\{i,j\} \in E} \bar{t}'_{ij} f_{ij}^w \leq u_w^q$ (we are considering $\bar{t}'(w) + \sum_{\{i,j\} \in E} \bar{t}'_{ij} f_{ij}^w$ as the average time for travelling between e_w and e'_w by following the associated preliminary route, since $\sum_{\{i,j\} \in E} \bar{t}'_{ij} f_{ij}^w - \bar{t}_s$ is the average time for travelling between the boarding and alighting platforms at e_w and e'_w for this preliminary route, assuming that no transfers are required).

For each $w \in W'$, in the objective function we are considering $\sum_{k=1}^{q(w)} g_w^k p_w^k$ as the estimation for the number of weekly trips taken on the rapid transit network by the surveyees that plan to travel between e_w and e'_w. Since we are dealing with a maximization problem, each optimal solution to the above model will satisfy the following properties (the first one will also be satisfied by any feasible solution to the model):

(1) For each $w \in W'$ such that $\bar{t}'(w) + \sum_{\{i,j\} \in E} \bar{t}'_{ij} f_{ij}^w > u_w^k \quad \forall k \in \{1, \dots, q(w)\}$, it will be $p_w^k = 0 \quad \forall k \in \{1, \dots, q(w)\}$.

(2) For each $w \in W'$ such that $\exists k \in \{1, \dots, q(w)\}$ with $\bar{t}'(w) + \sum_{\{i,j\} \in E} \bar{t}'_{ij} f_{ij}^w \leq u_w^k$, it will be $p_w^{k^*} = 1$ and $p_w^k = 0 \quad \forall k \in \{1, \dots, q(w)\} \setminus \{k^*\}$, where $k^* = \min\{k \in \{1, \dots, q(w)\} \mid \bar{t}'(w) + \sum_{\{i,j\} \in E} \bar{t}'_{ij} f_{ij}^w \leq u_w^k\}$.

10

For the particular case where $q(w) = 1 \quad \forall w \in W'$, the above objective function is the same as in the model from Escudero and Muñoz (2009b), i.e., the all-or-nothing model for the potential users' behavior from Marín (2007) is considered.

The greater the values of $\{q(w)\}_{w \in W'}$, the more accurate the estimation of the number of weekly trips that the surveyees will take on the rapid transit network (assuming that no transfers are required) but also the more variables and constraints in the above model.

We have considered some other objective functions that do not require either the constraints (5) and (6) or the variables $\{p_w^k\}_{w \in W', k \in \{1, \ldots, q(w)\}}$, but the global computational results came out worse than for the above objective function.

In order to illustrate the application of the first stage of our proposed approach, a small-size instance is provided in the following example:

Example 1. *Consider the graph $G = (V, E)$, where $V = \{1, 2, 3, 4, 5, 6\}$ and $E = \{\{1, 2\}, \{1, 4\},$ $\{1, 5\}, \{2, 3\}, \{2, 6\}, \{4, 5\}, \{5, 6\}\}$ (see Figure 1). Then $\Gamma(1) = \{2, 4, 5\}$, $\Gamma(2) = \{1, 3, 6\}$, $\Gamma(3) = \{2\}$, $\Gamma(4) = \{1, 5\}$, $\Gamma(5) = \{1, 4, 6\}$, $\Gamma(6) = \{2, 5\}$ and $W = \{(1, 2), (1, 3), (1, 4), (1, 5),$ $(1, 6), (2, 3), (2, 4), (2, 5), (2, 6), (3, 4), (3, 5), (3, 6), (4, 5), (4, 6), (5, 6)\}$.*
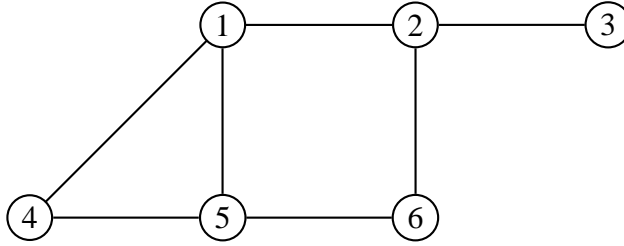
Figure 1: Graphic representation of $G = (V, E)$

Let $a_i = 30 \quad \forall i \in V$, $d_{12} = 1.6$, $d_{14} = 2.1$, $d_{15} = 1$, $d_{23} = 1.1$, $d_{26} = 1.5$, $d_{45} = 1.2$, $d_{56} = 0.9$, $s_{ij} = 0 \quad \forall \{i, j\} \in E$, $c_{ij} = 45 d_{ij} \quad \forall \{i, j\} \in E$, $b = 645$, $\bar{v} = 60$, $\bar{t}(i) = 2 \quad \forall i \in V$, $\bar{t}_a = 4$, $\bar{t}_s = 0.5$ and $\bar{t}_r = 3$ ($\{a_i\}_{i \in V}$, $\{c_{ij}\}_{\{i, j\} \in E}$ and b are expressed in millions of euros). Then $\bar{t}'_{ij} = d_{ij} + 0.5 \quad \forall \{i, j\} \in E$.

Let $\Theta = \{\theta_1, \ldots, \theta_5\}$, $W_{\theta_1} = \{(1, 2), (1, 6), (2, 4), (2, 6)\}$, $W_{\theta_2} = \{(1, 4), (1, 5), (4, 5)\}$, $W_{\theta_3} = \{(1, 6), (2, 5), (3, 5)\}$, $W_{\theta_4} = \{(1, 3), (1, 6), (2, 3), (3, 5), (3, 6), (5, 6)\}$ and $W_{\theta_5} = \{(1, 6), (2, 4), (3, 4), (3, 6), (4, 6)\}$. Therefore, we have $W' = W$ and $\bar{t}'(w) = 5.5 \quad \forall w \in W'$.

For each $\theta \in \Theta$ and for each $w \in W_\theta$, Table 1 shows the values we have considered for $\alpha_w(\theta)$ and $\tau_w(\theta)$:

Table 1: Values of $\{\alpha_w(\theta)\}_{\theta \in \Theta, w \in W_\theta}$ and $\{\tau_w(\theta)\}_{\theta \in \Theta, w \in W_\theta}$

θ	w	$\alpha_w(\theta)$	$\tau_w(\theta)$
θ_1	$(1,2)$	1	8
	$(1,6)$	2	12
	$(2,4)$	5	10
	$(2,6)$	1	9
θ_2	$(1,4)$	1	10
	$(1,5)$	1	7
	$(4,5)$	1	8
θ_3	$(1,6)$	1	13
	$(2,5)$	1	9
	$(3,5)$	8	10
θ_4	$(1,3)$	1	12
	$(1,6)$	10	8
	$(2,3)$	1	8
	$(3,5)$	1	12
	$(3,6)$	1	16
	$(5,6)$	1	8
θ_5	$(1,6)$	5	9
	$(2,4)$	1	12
	$(3,4)$	1	15
	$(3,6)$	12	9
	$(4,6)$	1	9

For each $w \in W'$, Table 2 shows the values we have considered for $q(w)$, $\{u_w^k\}_{k \in \{1,...,q(w)\}}$, $\{g_w^k\}_{k \in \{1,...,q(w)\}}$ and $\{M_w^k\}_{k \in \{1,...,q(w)\}}$ (the procedure we have utilized for obtaining them will be detailed in Section 9):

Table 2: Values of $\{q(w)\}_{w \in W'}$, $\{u_w^k\}_{w \in W', k \in \{1,...,q(w)\}}$, $\{g_w^k\}_{w \in W', k \in \{1,...,q(w)\}}$ and $\{M_w^k\}_{w \in W', k \in \{1,...,q(w)\}}$

w	$q(w)$	$\{u_w^k\}_{k \in \{1,...,q(w)\}}$	$\{g_w^k\}_{k \in \{1,...,q(w)\}}$	$\{M_w^k\}_{k \in \{1,...,q(w)\}}$
$(1,2)$	1	8	1	7.5
$(1,3)$	1	12	1	3.5
$(1,4)$	1	10	1	5.5
$(1,5)$	1	7	1	8.5
$(1,6)$	2	8, 9	18, 8	7.5, 6.5
$(2,3)$	1	8	1	7.5
$(2,4)$	1	10	6	5.5
$(2,5)$	1	9	1	6.5
$(2,6)$	1	9	1	6.5
$(3,4)$	1	15	1	0.5
$(3,5)$	1	10	9	5.5
$(3,6)$	2	9, 16	13, 1	6.5, −0.5
$(4,5)$	1	8	1	7.5
$(4,6)$	1	9	1	6.5
$(5,6)$	1	8	1	7.5

Upon solving the above model for this particular instance, we obtain that all links except $\{1,4\}$ are constructed, the construction cost of the rapid transit network is 568.5 million euros, and the optimal value of the objective function is 20.

4 Modification of the Floyd-Warshall algorithm

Let $G' = (V', E')$ be a simple graph, and let $d'_{ij} \geq 0 \quad \forall \{i, j\} \in E'$ and $d'_{ij} = +\infty \quad \forall i, j \in V'$ such that $i \neq j$ and $\{i, j\} \notin E'$ (for each $\{i, j\} \in E'$, d'_{ij} is understood as the length of edge $\{i, j\}$). Without loss of generality, let us assume that $V' = \{1, \ldots, n'\}$.

Below we present a modification of the well-known Floyd-Warshall algorithm (see e.g. Korte and Vygen (2008)) to determine a shortest chain in G' between each pair of nodes of V'. It takes advantage of the undirectedness of G' as well as of the nonnegative lengths of the edges in E', and it will be utilized in subsequent phases of the approach (see Sections 5 and 7).

Algorithm 1.

Step 1. Set $h_{ij} = d'_{ij}, h_{ji} = h_{ij}, p_{ij} = i$ and $p_{ji} = j \quad \forall i, j \in V'$ with $i < j$, and set $k = 1$.

Step 2. Set $h_{ij} = h_{ik} + h_{kj}, h_{ji} = h_{ij}, p_{ij} = p_{kj}$ and $p_{ji} = p_{ki} \quad \forall i, j \in V' \setminus \{k\}$ such that $i < j$ and $h_{ik} + h_{kj} < h_{ij}$.

Step 3. If $k = n'$, STOP; otherwise, set $k = k + 1$ and go to Step 2.

5 Determination of the shortest routes without considering transfers

We are interested in determining a route of minimum average time for each trip that will be taken on the rapid transit network by the surveyees. However, since a line design for the rapid transit network is not still available, we are forced to assume that no transfers will be required in those routes (later, in Section 7, transfers will be taken into consideration).

Let $(\overline{x}_{ij})_{\{i,j\} \in E}$, $(\overline{\gamma}_i)_{i \in V}$, $(\overline{\Delta}_i)_{i \in V}$, $(\overline{f}^w_{ij})_{w \in W, \{i,j\} \in E}$, $(\overline{\varepsilon}^w_i)_{w \in W, i \in V \setminus \{e_w, e'_w\}}$, $(\overline{p}^k_w)_{w \in W', k \in \{1,\dots,q(w)\}}$ be an optimal solution to the model stated in Section 3 (or an incumbent solution if the model has not been solved to optimality), and let $\overline{E} = \{\{i,j\} \in E \mid \overline{x}_{ij} = 1\}$.

The rest of the proposed approach would be valid for any subset $\overline{E}$ of E such that $(V, \overline{E})$ is a connected graph and the construction cost of its associated rapid transit network does not exceed the available budget. Consequently, it could also be employed for redesigning the lines of existing rapid transit networks.

Let us consider the partial graph $\overline{G} = (V, \overline{E})$ of G. For each $i \in V$, let $\overline{\Gamma}(i)$ be the set of nodes adjacent to i in $\overline{G}$. Let $\overline{s} = \sum_{\{i,j\} \in \overline{E}} s_{ij}$ (notice that $\overline{s}$ is the total number of non-key station locations in $\overline{G}$).

It is worth noting that, for each $w \in W$, the preliminary route defined by $\{\overline{f}^w_{ij}\}_{\{i,j\} \in \overline{E}}$ is not necessarily a chain of minimum length in $\overline{G}$ joining e_w and e'_w, considering $\overline{t}'_{ij}$ as the length of each edge $\{i, j\} \in \overline{E}$. In order to determine such shortest chains, let us apply Algorithm 1 taking $G' = \overline{G}$ and $d'_{ij} = \overline{t}'_{ij} \quad \forall \{i, j\} \in \overline{E}$.

Given the output $\{h_{ij}\}_{i,j \in V, i \neq j}, \{p_{ij}\}_{i,j \in V, i \neq j}$ of Algorithm 1, let $\overline{\lambda}_w = \overline{t}'(w) + h_{e_w e'_w} \quad \forall w \in W'$ (notice that $\overline{\lambda}_w$ is the minimum average time that the trip between e_w and e'_w will take on the rapid transit network, assuming that no transfers are required), and let $\{\hat{f}^w_{ij}\}_{w \in W', \{i,j\} \in \overline{E}}$ and $\{\hat{\varepsilon}^w_i\}_{w \in W', i \in V \setminus \{e_w, e'_w\}}$ be, respectively, the values of the variables $\{f^w_{ij}\}_{w \in W', \{i,j\} \in \overline{E}}$ and $\{\varepsilon^w_i\}_{w \in W', i \in V \setminus \{e_w, e'_w\}}$ that define the routes identified by Algorithm 1, i.e.,

14

$$\hat{f}_{ij}^w = \begin{cases} 1 & \text{if the shortest chain in } \overline{G} \text{ joining } e_w \text{ and } e'_w \\ & \text{determined by Algorithm 1 contains edge } \{i,j\} \qquad \forall w \in W', \forall \{i,j\} \in \overline{E} \\ 0 & \text{otherwise} \end{cases}$$

$$\hat{\varepsilon}_i^w = \begin{cases} 1 & \text{if the shortest chain in } \overline{G} \text{ joining } e_w \text{ and } e'_w \\ & \text{determined by Algorithm 1 passes through } i \qquad \forall w \in W', \forall i \in V \setminus \{e_w, e'_w\} \\ 0 & \text{otherwise} \end{cases}$$

(these values can easily be obtained from $\{p_{ij}\}_{i,j \in V, i \neq j}$; in particular, for each $w \in W'$ such that $h_{e_w e'_w} = \sum_{\{i,j\} \in \overline{E}} \overline{t}'_{ij} \overline{f}_{ij}^w$, we can set $\hat{f}_{ij}^w = \overline{f}_{ij}^w \quad \forall \{i,j\} \in \overline{E}$ and $\hat{\varepsilon}_i^w = \overline{\varepsilon}_i^w \quad \forall i \in V \setminus \{e_w, e'_w\}$).

For each $w \in W'$, let $\overline{g}_w = \sum_{\theta \in \overline{\Theta}_w} \alpha_w(\theta)$, where $\overline{\Theta}_w = \{\theta \in \Theta_w \mid \tau_w(\theta) \geq \overline{\lambda}_w\}$ (notice that $\overline{g}_w$ is the maximum possible expected number of weekly trips taken on the rapid transit network by the surveyees that plan to travel between e_w and e'_w, and it will be reached if the surveyees follow the route defined by $\{\hat{f}_{ij}^w\}_{\{i,j\} \in \overline{E}}$, no transfers are required in that route and the capacity of the rapid transit network is enough to hold all those trips). Let $\overline{z} = \sum_{w \in W'} \overline{g}_w$.

Example 2. *Let us continue solving the instance given in Example 1. We get that $\overline{G} = (V, \overline{E})$, where $\overline{E} = \{\{1,2\}, \{1,5\}, \{2,3\}, \{2,6\}, \{4,5\}, \{5,6\}\}$ (see Figure 2). Thus, $\overline{\Gamma}(1) = \{2,5\}$, $\overline{\Gamma}(2) = \{1,3,6\}$, $\overline{\Gamma}(3) = \{2\}$, $\overline{\Gamma}(4) = \{5\}$, $\overline{\Gamma}(5) = \{1,4,6\}$ and $\overline{\Gamma}(6) = \{2,5\}$ (notice that $\overline{s} = 0$).*
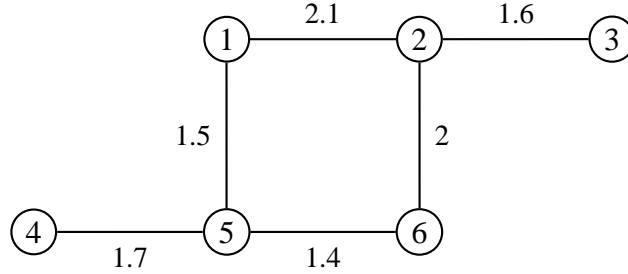
Figure 2: Graphic representation of $\overline{G} = (V, \overline{E})$ and values of $\{\overline{t}'_{ij}\}_{\{i,j\} \in \overline{E}}$

For each $w \in W'$, Table 3 shows the shortest chain in $\overline{G}$ joining e_w and e'_w determined by Algorithm 1 and the values of $\overline{\lambda}_w$ and $\overline{g}_w$ (notice that $\overline{z} = 22$):

Table 3: Shortest chain between e_w and e'_w $\forall w \in W'$ and values of $\{\overline{\lambda}_w\}_{w \in W'}$ and $\{\overline{g}_w\}_{w \in W'}$

w	Shortest chain between e_w and e'_w	$\overline{\lambda}_w$	$\overline{g}_w$
$(1,2)$	$\{\{1,2\}\}$	7.6	1
$(1,3)$	$\{\{1,2\},\{2,3\}\}$	9.2	1
$(1,4)$	$\{\{1,5\},\{5,4\}\}$	8.7	1
$(1,5)$	$\{\{1,5\}\}$	7	1
$(1,6)$	$\{\{1,5\},\{5,6\}\}$	8.4	8
$(2,3)$	$\{\{2,3\}\}$	7.1	1
$(2,4)$	$\{\{2,6\},\{6,5\},\{5,4\}\}$	10.6	1
$(2,5)$	$\{\{2,6\},\{6,5\}\}$	8.9	1
$(2,6)$	$\{\{2,6\}\}$	7.5	1
$(3,4)$	$\{\{3,2\},\{2,6\},\{6,5\},\{5,4\}\}$	12.2	1
$(3,5)$	$\{\{3,2\},\{2,6\},\{6,5\}\}$	10.5	1
$(3,6)$	$\{\{3,2\},\{2,6\}\}$	9.1	1
$(4,5)$	$\{\{4,5\}\}$	7.2	1
$(4,6)$	$\{\{4,5\},\{5,6\}\}$	8.6	1
$(5,6)$	$\{\{5,6\}\}$	6.9	1

6 Greedy heuristic procedure for minimizing the number of transfers given the set of links to be constructed

In this section we provide a generalization of the greedy heuristic algorithm for designing a set of lines presented in Escudero and Muñoz (2009b), to allow pairs of station locations linked by more than one line (the algorithm presented in Escudero and Muñoz (2009a) could also be generalized analogously). Each link to be constructed will be assigned to exactly one line, attempting to minimize an estimation of the number of weekly transfers that should be made by the surveyees to arrive at their destinations. Although the new algorithm can fail in designing a set of lines for a nonsimple graph, it will be applied at least to graph $\overline{G}$ (see Section 8), hence obtaining a line design for the rapid transit network will be guaranteed.

Let $\overline{G}' = (V, \overline{E}')$ be a graph such that the edges in $\overline{E}'$ are the same as in $\overline{E}$, but their multiplicities can be greater than one. When we refer to an edge $\{i,j\} \in \overline{E}'$, it will can be either $i < j$ or $i > j$.

For each $i \in V$, let $d(i)$ be the degree of i in $\overline{G}'$, and, for each $\{i,j\} \in \overline{E}'$, let m_{ij} be the multiplicity of $\{i,j\}$ in $\overline{G}'$.

The proposed approach can require designing a set of lines for different sets of edges $\overline{E}'$. We shall impose that each one of these sets $\overline{E}'$, jointly with the set of lines obtained for it, satisfy the two following conditions:

(1) For each $i \in V$, the number of lines that go to i is $\frac{d(i)}{2}$ if $d(i)$ is even, or $\frac{d(i)+1}{2}$ if $d(i)$ is odd.

(2) $\sum_{i \in V, d(i) \text{ even}} a_i \frac{d(i)}{2} + \sum_{i \in V, d(i) \text{ odd}} a_i \frac{d(i)+1}{2} + \sum_{\{i,j\} \in \overline{E}} c_{ij} m_{ij} \leq b$.

Condition (1) is imposed in order to reduce as much as possible the construction cost of the rapid transit network defined by $\overline{G}'$ (notice that a necessary condition for (1) to be satisfied is that $\min\{d(i), d(j)\} \geq 2m_{ij} - 1 \quad \forall\{i,j\} \in \overline{E}$). Condition (2) imposes that the construction cost of the rapid transit network defined by $\overline{G}'$ does not exceed the available budget.

For obvious reasons, lines containing two or more equal links will not be allowed.

It will be assumed that the surveyees follow the routes defined by $\{\hat{f}_{ij}^w\}_{w \in W', \{i,j\} \in \overline{E}}$.

Let $\overline{W} = \{w \in W' \mid \overline{g}_w > 0\}$. For simplicity of notation, we define $\hat{f}_{ji}^w = \hat{f}_{ij}^w \quad \forall w \in \overline{W}, \forall\{i,j\} \in \overline{E}$.

Let $\overline{W}_i = \{w \in \overline{W} \mid i \notin \{e_w, e_w'\}, \hat{\varepsilon}_i^w = 1\} \quad \forall i \in V$, let $t_j(i) = \sum_{w \in \overline{W}_i, \hat{f}_{ij}^w = 1} \overline{g}_w \quad \forall i \in V, \forall j \in \overline{\Gamma}(i)$, and let $t_k(i,j) = \sum_{w \in \overline{W}_j, \hat{f}_{ij}^w + \hat{f}_{jk}^w = 1} \overline{g}_w \quad \forall i \in V, \forall j \in \overline{\Gamma}(i), \forall k \in \overline{\Gamma}(j) \setminus \{i\}$ (notice that $\overline{W}_i$ is the set of distinct pairs of key station locations such that the surveyees that plan to travel between them pass through location i, $t_j(i)$ is the maximum possible expected number of weekly transfers at location i made by the surveyees that utilize link $\{i,j\}$, provided that i is an endpoint of the lines that link i and j, and $t_k(i,j)$ is the maximum possible expected number of weekly transfers at location j made by the surveyees that utilize one and only one of the links $\{i,j\}$ and $\{j,k\}$, provided that these links belong to the same line).

Algorithm 2 below is a greedy heuristic procedure for designing a set of lines for $\overline{G}'$. It generalizes Algorithm 2 in Escudero and Muñoz (2009b) and can be outlined as follows: Starting from a node with odd degree, or, in its absence, with positive even degree, other nodes are chosen sequentially through edges in $\overline{E}'$ attending to certain criteria, until a node is reached which either has previously been visited or has no incident edges (once an edge has been considered, it is eliminated from $\overline{E}'$). In the first case, we define a circular line and check whether it can still be possible to design a set of lines such that condition (1) is satisfied; if so, the above procedure is carried on from the last reached node which is an endpoint of an edge that has been eliminated from $\overline{E}'$ but has not yet been assigned to a line, if such a node exists. In the second case, we define a noncircular line. This approach is repeated until we either detect that it cannot be possible to design a set of lines such that condition (1) is satisfied or get $\overline{E}' = \emptyset$ (see Escudero and Muñoz (2009b, 2009a) for more details).

If $\overline{E}' = \overline{E}$, Algorithm 2 proceeds in the same way as Algorithm 2 in Escudero and Muñoz (2009b); therefore, it will obtain a line design. Otherwise, Algorithm 2 can fail in obtaining a line design.

In order to store the sequence of nodes chosen at each iteration, a nonnegative integer value $p(i)$ is associated to each node $i \in V$, in such a way that a positive value $p(i)$ means that node i has been reached from node $p(i)$ (for the starting node i_0 we define $p(i_0) = i_0$). A counter l for the number of lines that are being defined is also considered. These lines are denoted by L_l.

Algorithm 2.

Step 1. Set $p(i) = 0 \quad \forall i \in V$ and $l = 0$.

Step 2. If $\overline{E}' = \emptyset$, STOP.

Step 3. If $d(i)$ is even $\forall i \in V$, choose $i_0 \in V$ such that $d(i_0) > 0$, set $j = \arg\max\{m_{i_0 j'} \mid j' \in \overline{\Gamma}(i_0)\}$ and go to Step 5.

Step 4. Choose $i_0 \in V$ such that $d(i_0)$ is odd and set $j = \arg\max\{m_{i_0 j'} \mid j' \in \overline{\Gamma}(i_0)\}$. If $d(i_0) > 2m_{i_0 j}$, set $j = \arg\min\{t_{j'}(i_0) \mid j' \in \overline{\Gamma}(i_0)\}$.

Step 5. Set $l = l + 1$, $L_l = \emptyset$, $p(i_0) = i_0$ and $i = i_0$.

Step 6. Set $\overline{E}' = \overline{E}' \setminus \{\{i, j\}\}$, $d(i) = d(i) - 1$, $d(j) = d(j) - 1$, $m_{ij} = m_{ij} - 1$ and $m_{ji} = m_{ij}$. If $m_{ij} = 0$, set $\overline{\Gamma}(i) = \overline{\Gamma}(i) \setminus \{j\}$ and $\overline{\Gamma}(j) = \overline{\Gamma}(j) \setminus \{i\}$.

Step 7. If $p(j) > 0$, set $j_0 = j$ and go to Step 11.

Step 8. If $d(j) = 0$, set $j_0 = i_0$ and go to Step 11.

Step 9. Set $k = \arg\max\{m_{jk'} \mid k' \in \overline{\Gamma}(j) \setminus \{i\}\}$. If $d(j) < 2m_{jk}$, set $p(j) = i$, $i = j$, $j = k$ and go to Step 6.

Step 10. Set $k = \arg\min\{t_{k'}(i, j) \mid k' \in \overline{\Gamma}(j) \setminus \{i\}\}$. If $d(j)$ is even and $t_k(i, j) > t_i(j)$, set $j_0 = i_0$; otherwise, set $p(j) = i$, $i = j$, $j = k$ and go to Step 6.

Step 11. Set $L_l = L_l \cup \{\{j, i\}\}$. If $i \neq j_0$, set $j = i$, $i = p(i)$, $p(j) = 0$ and repeat Step 11.

Step 12. If $i = i_0$, set $p(i_0) = 0$ and go to Step 2.

Step 13. If $d(i) < 2m_{p(i),i}$, STOP (no line design has been obtained); otherwise, set $l = l + 1$, $L_l = \emptyset$, $j = i$, $i = p(i)$, $p(j) = 0$ and go to Step 8.

REMARK 1. *If $d(i_0)$ is even, then Step 8 can be skipped, since it will always be $d(j) > 0$.*

We have set $j = \arg\max\{m_{i_0 j'} \mid j' \in \overline{\Gamma}(i_0)\}$ in Step 3, since, if $d(i_0) \leq 2m_{i_0 j}$ and a circular line were defined containing two links of the form $\{i_0, j_1\}$ and $\{j_2, i_0\}$, where $j_1, j_2 \in \overline{\Gamma}(i_0) \setminus \{j\}$ and $j_1 \neq j_2$, then condition (1) would not be satisfied.

We have initially set $j = \arg\max\{m_{i_0 j'} \mid j' \in \overline{\Gamma}(i_0)\}$ in Step 4, since, if $d(i_0) < 2m_{i_0 j}$ and a line were defined containing a link of the form $\{i_0, j_1\}$, where $j_1 \in \overline{\Gamma}(i_0) \setminus \{j\}$, and not containing link $\{i_0, j\}$, then condition (1) would not be satisfied.

We have initially set $k = \arg\max\{m_{jk'} \mid k' \in \overline{\Gamma}(j) \setminus \{i\}\}$ in Step 9 (below it will be shown that $\overline{\Gamma}(j) \setminus \{i\} \neq \emptyset$, hence k is correctly defined), since, if $d(j) < 2m_{jk}$ and a line were defined containing link $\{i, j\}$ and not containing link $\{j, k\}$, then condition (1) would not be satisfied.

Given that $\min\{d(i), d(j)\} \geq 2m_{ij} - 1 \quad \forall \{i, j\} \in \overline{E}$, it can easily be deduced from Steps 3, 4, 9 and 13 that each time a line L_l is set to the empty set, we have that $\min\{d^l(i), d^l(j)\} \geq 2m^l_{ij} - 1 \quad \forall \{i, j\} \in \overline{E}$, where $d^l(i)$ and $d^l(j)$ are, respectively, the degree of i and j in $\overline{G}'_l$, m^l_{ij} is the multiplicity of $\{i, j\}$ in $\overline{G}'_l$, and $\overline{G}'_l = (V, \overline{E}' \setminus \bigcup_{l'=1}^{l-1} L_{l'})$. As a consequence, it will always be $\overline{\Gamma}(j) \setminus \{i\} \neq \emptyset$ in Step 9.

Example 3. *Let us continue solving the instance given in Example 1 (see also Example 2). We get that $\overline{W} = W'$, $\overline{W}_1 = \emptyset$, $\overline{W}_2 = \{(1,3), (3,4), (3,5), (3,6)\}$, $\overline{W}_3 = \emptyset$, $\overline{W}_4 = \emptyset$, $\overline{W}_5 = \{(1,4), (1,6), (2,4), (3,4), (4,6)\}$ and $\overline{W}_6 = \{(2,4), (2,5), (3,4), (3,5)\}$.*

Below Algorithm 2 is applied considering three different sets of edges $\overline{E}'$ (these sets will also be considered in Sections 7 and 8).

- *Let $\overline{E}' = \overline{E}$. Then $d(1) = 2$, $d(2) = 3$, $d(3) = 1$, $d(4) = 1$, $d(5) = 3$, $d(6) = 2$ and $m_{12} = m_{21} = m_{15} = m_{51} = m_{23} = m_{32} = m_{26} = m_{62} = m_{45} = m_{54} = m_{56} = m_{65} = 1$.*

  Algorithm 2 proceeds as follows:

  Step 1. $p(1) = p(2) = p(3) = p(4) = p(5) = p(6) = 0$, $l = 0$

  Step 4. $i_0 = 2$, $j = 1$, $t_1(2) = 1$, $t_3(2) = 4$, $t_6(2) = 3$, $j = 1$

  Step 5. $l = 1$, $L_1 = \emptyset$, $p(2) = 2$, $i = 2$

  Step 6. $\overline{E}' = \{\{1,5\}, \{2,3\}, \{2,6\}, \{4,5\}, \{5,6\}\}$, $d(2) = 2$, $d(1) = 1$, $m_{21} = 0$, $m_{12} = 0$, $\overline{\Gamma}(2) = \{3, 6\}$, $\overline{\Gamma}(1) = \{5\}$

  Step 9. $k = 5$, $p(1) = 2$, $i = 1$, $j = 5$

  Step 6. $\overline{E}' = \{\{2,3\}, \{2,6\}, \{4,5\}, \{5,6\}\}$, $d(1) = 0$, $d(5) = 2$, $m_{15} = 0$, $m_{51} = 0$, $\overline{\Gamma}(1) = \emptyset$, $\overline{\Gamma}(5) = \{4, 6\}$

  Step 9. $k = 4$

Step 10. $t_4(1,5) = 11$, $t_6(1,5) = 4$, $k = 6$, $t_1(5) = 9$, $p(5) = 1$, $i = 5$, $j = 6$

Step 6. $\overline{E}' = \{\{2,3\},\{2,6\},\{4,5\}\}$, $d(5) = 1$, $d(6) = 1$, $m_{56} = 0$, $m_{65} = 0$, $\overline{\Gamma}(5) = \{4\}$, $\overline{\Gamma}(6) = \{2\}$

Step 9. $k = 2$, $p(6) = 5$, $i = 6$, $j = 2$

Step 6. $\overline{E}' = \{\{2,3\},\{4,5\}\}$, $d(6) = 0$, $d(2) = 1$, $m_{62} = 0$, $m_{26} = 0$, $\overline{\Gamma}(6) = \emptyset$, $\overline{\Gamma}(2) = \{3\}$

Step 7. $j_0 = 2$

Step 11. $L_1 = \{\{2,6\}\}$, $j = 6$, $i = 5$, $p(6) = 0$

Step 11. $L_1 = \{\{2,6\},\{6,5\}\}$, $j = 5$, $i = 1$, $p(5) = 0$

Step 11. $L_1 = \{\{2,6\},\{6,5\},\{5,1\}\}$, $j = 1$, $i = 2$, $p(1) = 0$

Step 11. $L_1 = \{\{2,6\},\{6,5\},\{5,1\},\{1,2\}\}$

Step 12. $p(2) = 0$

Step 4. $i_0 = 2$, $j = 3$

Step 5. $l = 2$, $L_2 = \emptyset$, $p(2) = 2$, $i = 2$

Step 6. $\overline{E}' = \{\{4,5\}\}$, $d(2) = 0$, $d(3) = 0$, $m_{23} = 0$, $m_{32} = 0$, $\overline{\Gamma}(2) = \emptyset$, $\overline{\Gamma}(3) = \emptyset$

Step 8. $j_0 = 2$

Step 11. $L_2 = \{\{3,2\}\}$

Step 12. $p(2) = 0$

Step 4. $i_0 = 4$, $j = 5$

Step 5. $l = 3$, $L_3 = \emptyset$, $p(4) = 4$, $i = 4$

Step 6. $\overline{E}' = \emptyset$, $d(4) = 0$, $d(5) = 0$, $m_{45} = 0$, $m_{54} = 0$, $\overline{\Gamma}(4) = \emptyset$, $\overline{\Gamma}(5) = \emptyset$

Step 8. $j_0 = 4$

Step 11. $L_3 = \{\{5,4\}\}$

Step 12. $p(4) = 0$

Thus, one circular line $L_1 = \{\{2,6\},\{6,5\},\{5,1\},\{1,2\}\}$, and two noncircular lines $L_2 = \{\{3,2\}\}$ and $L_3 = \{\{5,4\}\}$ have been defined.

- *Let $\overline{E}' = \overline{E} \cup \{\{1,5\}\}$. Then $d(1) = 3$, $d(2) = 3$, $d(3) = 1$, $d(4) = 1$, $d(5) = 4$, $d(6) = 2$, $m_{12} = m_{21} = m_{23} = m_{32} = m_{26} = m_{62} = m_{45} = m_{54} = m_{56} = m_{65} = 1$ and $m_{15} = m_{51} = 2$ (notice that condition (2) is satisfied).*

Algorithm 2 proceeds as follows:

Step 1. $p(1) = p(2) = p(3) = p(4) = p(5) = p(6) = 0, l = 0$

Step 4. $i_0 = 1, j = 5$

Step 5. $l = 1, L_1 = \emptyset, p(1) = 1, i = 1$

Step 6. $\overline{E}' = \{\{1,2\}, \{1,5\}, \{2,3\}, \{2,6\}, \{4,5\}, \{5,6\}\}, d(1) = 2, d(5) = 3, m_{15} = 1, m_{51} = 1$

Step 9. $k = 4$

Step 10. $t_4(1,5) = 11, t_6(1,5) = 4, k = 6, p(5) = 1, i = 5, j = 6$

Step 6. $\overline{E}' = \{\{1,2\}, \{1,5\}, \{2,3\}, \{2,6\}, \{4,5\}\}, d(5) = 2, d(6) = 1, m_{56} = 0, m_{65} = 0, \overline{\Gamma}(5) = \{1,4\}, \overline{\Gamma}(6) = \{2\}$

Step 9. $k = 2, p(6) = 5, i = 6, j = 2$

Step 6. $\overline{E}' = \{\{1,2\}, \{1,5\}, \{2,3\}, \{4,5\}\}, d(6) = 0, d(2) = 2, m_{62} = 0, m_{26} = 0, \overline{\Gamma}(6) = \emptyset, \overline{\Gamma}(2) = \{1,3\}$

Step 9. $k = 1$

Step 10. $t_1(6,2) = 4, t_3(6,2) = 1, k = 3, t_6(2) = 3, p(2) = 6, i = 2, j = 3$

Step 6. $\overline{E}' = \{\{1,2\}, \{1,5\}, \{4,5\}\}, d(2) = 1, d(3) = 0, m_{23} = 0, m_{32} = 0, \overline{\Gamma}(2) = \{1\}, \overline{\Gamma}(3) = \emptyset$

Step 8. $j_0 = 1$

Step 11. $L_1 = \{\{3,2\}\}, j = 2, i = 6, p(2) = 0$

Step 11. $L_1 = \{\{3,2\}, \{2,6\}\}, j = 6, i = 5, p(6) = 0$

Step 11. $L_1 = \{\{3,2\}, \{2,6\}, \{6,5\}\}, j = 5, i = 1, p(5) = 0$

Step 11. $L_1 = \{\{3,2\}, \{2,6\}, \{6,5\}, \{5,1\}\}$

Step 12. $p(1) = 0$

Step 4. $i_0 = 2, j = 1$

Step 5. $l = 2, L_2 = \emptyset, p(2) = 2, i = 2$

Step 6. $\overline{E}' = \{\{1,5\}, \{4,5\}\}, d(2) = 0, d(1) = 1, m_{21} = 0, m_{12} = 0, \overline{\Gamma}(2) = \emptyset, \overline{\Gamma}(1) = \{5\}$

Step 9. $k = 5, p(1) = 2, i = 1, j = 5$

Step 6. $\overline{E}' = \{\{4,5\}\}, d(1) = 0, d(5) = 1, m_{15} = 0, m_{51} = 0, \overline{\Gamma}(1) = \emptyset, \overline{\Gamma}(5) = \{4\}$

Step 9. $k = 4, p(5) = 1, i = 5, j = 4$

Step 6. $\overline{E}' = \emptyset, d(5) = 0, d(4) = 0, m_{54} = 0, m_{45} = 0, \overline{\Gamma}(5) = \emptyset, \overline{\Gamma}(4) = \emptyset$

Step 8. $j_0 = 2$

Step 11. $L_2 = \{\{4,5\}\}$, $j = 5$, $i = 1$, $p(5) = 0$

Step 11. $L_2 = \{\{4,5\},\{5,1\}\}$, $j = 1$, $i = 2$, $p(1) = 0$

Step 11. $L_2 = \{\{4,5\},\{5,1\},\{1,2\}\}$

Step 12. $p(2) = 0$

Thus, two noncircular lines $L_1 = \{\{3,2\},\{2,6\},\{6,5\},\{5,1\}\}$ and $L_2 = \{\{4,5\},\{5,1\},\{1,2\}\}$ have been defined.

- *Let $\overline{E}' = \overline{E} \cup \{\{5,6\}\}$. Then $d(1) = 2$, $d(2) = 3$, $d(3) = 1$, $d(4) = 1$, $d(5) = 4$, $d(6) = 3$, $m_{12} = m_{21} = m_{15} = m_{51} = m_{23} = m_{32} = m_{26} = m_{62} = m_{45} = m_{54} = 1$ and $m_{56} = m_{65} = 2$ (notice that condition (2) is satisfied).*

Algorithm 2 proceeds as follows:

Step 1. $p(1) = p(2) = p(3) = p(4) = p(5) = p(6) = 0$, $l = 0$

Step 4. $i_0 = 2$, $j = 1$, $t_1(2) = 1$, $t_3(2) = 4$, $t_6(2) = 3$, $j = 1$

Step 5. $l = 1$, $L_1 = \emptyset$, $p(2) = 2$, $i = 2$

Step 6. $\overline{E}' = \{\{1,5\},\{2,3\},\{2,6\},\{4,5\},\{5,6\},\{5,6\}\}$, $d(2) = 2$, $d(1) = 1$, $m_{21} = 0$, $m_{12} = 0$, $\overline{\Gamma}(2) = \{3,6\}$, $\overline{\Gamma}(1) = \{5\}$

Step 9. $k = 5$, $p(1) = 2$, $i = 1$, $j = 5$

Step 6. $\overline{E}' = \{\{2,3\},\{2,6\},\{4,5\},\{5,6\},\{5,6\}\}$, $d(1) = 0$, $d(5) = 3$, $m_{15} = 0$, $m_{51} = 0$, $\overline{\Gamma}(1) = \emptyset$, $\overline{\Gamma}(5) = \{4,6\}$

Step 9. $k = 6$, $p(5) = 1$, $i = 5$, $j = 6$

Step 6. $\overline{E}' = \{\{2,3\},\{2,6\},\{4,5\},\{5,6\}\}$, $d(5) = 2$, $d(6) = 2$, $m_{56} = 1$, $m_{65} = 1$

Step 9. $k = 2$

Step 10. $t_2(5,6) = 0$, $k = 2$, $t_5(6) = 4$, $p(6) = 5$, $i = 6$, $j = 2$

Step 6. $\overline{E}' = \{\{2,3\},\{4,5\},\{5,6\}\}$, $d(6) = 1$, $d(2) = 1$, $m_{62} = 0$, $m_{26} = 0$, $\overline{\Gamma}(6) = \{5\}$, $\overline{\Gamma}(2) = \{3\}$

Step 7. $j_0 = 2$

Step 11. $L_1 = \{\{2,6\}\}$, $j = 6$, $i = 5$, $p(6) = 0$

Step 11. $L_1 = \{\{2,6\},\{6,5\}\}$, $j = 5$, $i = 1$, $p(5) = 0$

Step 11. $L_1 = \{\{2,6\},\{6,5\},\{5,1\}\}$, $j = 1$, $i = 2$, $p(1) = 0$

Step 11. $L_1 = \{\{2,6\},\{6,5\},\{5,1\},\{1,2\}\}$

Step 12. $p(2) = 0$

Step 4. $i_0 = 2$, $j = 3$

Step 5. $l = 2$, $L_2 = \emptyset$, $p(2) = 2$, $i = 2$

Step 6. $\overline{E}' = \{\{4,5\},\{5,6\}\}$, $d(2) = 0$, $d(3) = 0$, $m_{23} = 0$, $m_{32} = 0$, $\overline{\Gamma}(2) = \emptyset$, $\overline{\Gamma}(3) = \emptyset$

Step 8. $j_0 = 2$

Step 11. $L_2 = \{\{3,2\}\}$

Step 12. $p(2) = 0$

Step 4. $i_0 = 4$, $j = 5$

Step 5. $l = 3$, $L_3 = \emptyset$, $p(4) = 4$, $i = 4$

Step 6. $\overline{E}' = \{\{5,6\}\}$, $d(4) = 0$, $d(5) = 1$, $m_{45} = 0$, $m_{54} = 0$, $\overline{\Gamma}(4) = \emptyset$, $\overline{\Gamma}(5) = \{6\}$

Step 9. $k = 6$, $p(5) = 4$, $i = 5$, $j = 6$

Step 6. $\overline{E}' = \emptyset$, $d(5) = 0$, $d(6) = 0$, $m_{56} = 0$, $m_{65} = 0$, $\overline{\Gamma}(5) = \emptyset$, $\overline{\Gamma}(6) = \emptyset$

Step 8. $j_0 = 4$

Step 11. $L_3 = \{\{6,5\}\}$, $j = 5$, $i = 4$, $p(5) = 0$

Step 11. $L_3 = \{\{6,5\},\{5,4\}\}$

Step 12. $p(4) = 0$

Thus, one circular line $L_1 = \{\{2,6\},\{6,5\},\{5,1\},\{1,2\}\}$, and two noncircular lines $L_2 = \{\{3,2\}\}$ and $L_3 = \{\{6,5\},\{5,4\}\}$ have been defined.

7 Determination of the shortest routes considering transfers

The shortest chains determined in Section 5 allowed us to calculate the minimum average time for each trip taken on the rapid transit network by the surveyees, as well as the maximum possible expected number of weekly trips taken by the surveyees, assuming that no transfers were required. However, once a line design for the rapid transit network is available, these values can be recalculated in a more accurate way by taking into consideration the transfer times. For this purpose, we require to expand the network considered in Section 5 (a similar expanded network was defined in Mandl (1980)).

Let $L_1,\ldots,L_{\hat{l}}$ be a set of lines for $\overline{G}'$ (see Section 6), and, for each $i \in V$, let $L(i)$ be the set of indices $l \in \{1,\ldots,\hat{l}\}$ such that line L_l goes to i.

Let us consider the simple graph $\tilde{G} = (\tilde{V}, \tilde{E})$, where $\tilde{V} = \{i\text{-}l \mid i \in V, l \in L(i)\}$, $\tilde{E} = \tilde{E}' \cup \tilde{E}''$, $\tilde{E}' = \{\{i\text{-}l, j\text{-}l\} \mid l \in \{1,\ldots,\hat{l}\}, \{i,j\} \in L_l\}$ and $\tilde{E}'' = \{\{i\text{-}l, i\text{-}l'\} \mid i \in V, l \in L(i), l' \in L(i) \setminus \{l\}\}$ (the edges in $\tilde{E}'$ are the same as in $\overline{E}'$, but now the indices l of their endpoints indicate the lines L_l to which they belong; each edge $\{i\text{-}l, i\text{-}l'\} \in \tilde{E}''$ indicates a transfer at i between lines L_l and $L_{l'}$).

Given two nodes $i\text{-}l, j\text{-}l' \in \tilde{V}$, we say that $i\text{-}l < j\text{-}l'$ either if $i < j$ or if $i = j$ and $l < l'$. Without loss of generality, whenever we refer to an edge $\{i\text{-}l, j\text{-}l'\} \in \tilde{E}$ it will be assumed that $i\text{-}l < j\text{-}l'$.

Let $\tilde{t}_{i\text{-}l, j\text{-}l} = \bar{t}'_{ij} \quad \forall \{i\text{-}l, j\text{-}l\} \in \tilde{E}'$ and $\tilde{t}_{i\text{-}l, i\text{-}l'} = \bar{t}_r + \frac{\bar{t}_a}{2} - \bar{t}_s \quad \forall \{i\text{-}l, i\text{-}l'\} \in \tilde{E}''$ (notice that $\bar{t}_r + \frac{\bar{t}_a}{2} - \bar{t}_s > 0$, since it was assumed that $\bar{t}_s < \frac{\bar{t}_a}{2}$), and let us apply Algorithm 1 taking $G' = \tilde{G}$ and $d'_{i\text{-}l, j\text{-}l'} = \tilde{t}_{i\text{-}l, j\text{-}l'} \quad \forall \{i\text{-}l, j\text{-}l'\} \in \tilde{E}$.

Given the output $\{h_{i\text{-}l, j\text{-}l'}\}_{i\text{-}l, j\text{-}l' \in \tilde{V}, i\text{-}l \neq j\text{-}l'}$, $\{p_{i\text{-}l, j\text{-}l'}\}_{i\text{-}l, j\text{-}l' \in \tilde{V}, i\text{-}l \neq j\text{-}l'}$ of Algorithm 1, let $\tilde{\lambda}_w = \bar{t}'(w) + h_{e_w\text{-}\tilde{l}_w, e'_w\text{-}\tilde{l}'_w} \quad \forall w \in W'$, where $\tilde{l}_w \in L(e_w)$, $\tilde{l}'_w \in L(e'_w)$ and $h_{e_w\text{-}\tilde{l}_w, e'_w\text{-}\tilde{l}'_w} = \min\{h_{e_w\text{-}l, e'_w\text{-}l'} \mid l \in L(e_w), l' \in L(e'_w)\}$ (notice that $\tilde{\lambda}_w$ is the minimum average time that the trip between e_w and e'_w will take on the rapid transit network).

For each $w \in W'$, let $\tilde{g}_w = \sum_{\theta \in \tilde{\Theta}_w} \alpha_w(\theta)$, where $\tilde{\Theta}_w = \{\theta \in \Theta_w \mid \tau_w(\theta) \geq \tilde{\lambda}_w\}$ (notice that $\tilde{g}_w$ is the maximum possible expected number of weekly trips taken on the rapid transit network by the surveyees that plan to travel between e_w and e'_w, and it will be reached if the surveyees follow a route that takes an average time of $\tilde{\lambda}_w$ minutes, which can easily be obtained from $\{p_{e_w\text{-}\tilde{l}_w, j\text{-}l'}\}_{j\text{-}l' \in \tilde{V}, e_w\text{-}\tilde{l}_w \neq j\text{-}l'}$, and the capacity of the rapid transit network is enough to hold all those trips). Let $\tilde{z} = \sum_{w \in W'} \tilde{g}_w$.

Example 4. *Let us continue solving the instance given in Example 1 (see also Examples 2 and 3). We have that $\bar{t}_r + \frac{\bar{t}_a}{2} - \bar{t}_s = 4.5$.*

- *Let $\overline{E}' = \overline{E}$. The lines defined by Algorithm 2 were $L_1 = \{\{2,6\}, \{6,5\}, \{5,1\}, \{1,2\}\}$, $L_2 = \{\{3,2\}\}$ and $L_3 = \{\{5,4\}\}$. Consequently, we get that $\tilde{G} = (\tilde{V}, \tilde{E})$, where $\tilde{V} = \{1\text{-}1, 2\text{-}1, 2\text{-}2, 3\text{-}2, 4\text{-}3, 5\text{-}1, 5\text{-}3, 6\text{-}1\}$ and $\tilde{E} = \{\{1\text{-}1, 2\text{-}1\}, \{1\text{-}1, 5\text{-}1\}, \{2\text{-}1, 2\text{-}2\}, \{2\text{-}1, 6\text{-}1\}, \{2\text{-}2, 3\text{-}2\}, \{4\text{-}3, 5\text{-}3\}, \{5\text{-}1, 5\text{-}3\}, \{5\text{-}1, 6\text{-}1\}\}$ (see Figure 3).*
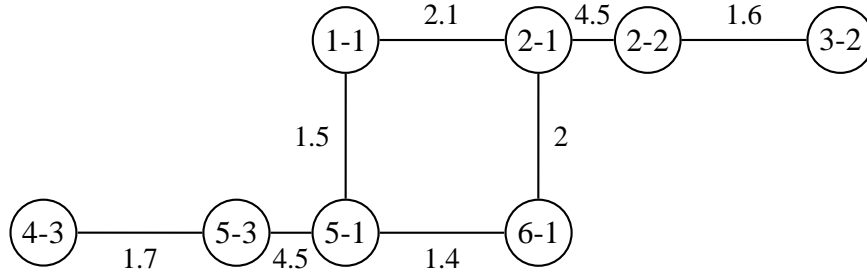
Figure 3: Graphic representation of $\tilde{G} = (\tilde{V}, \tilde{E})$ for $\overline{E}' = \overline{E}$, and values of $\{\tilde{t}_{i\text{-}l, j\text{-}l'}\}_{\{i\text{-}l, j\text{-}l'\} \in \tilde{E}}$

For each $w \in W'$, Table 4 shows the shortest chain in $\tilde{G}$ joining $e_w\text{-}\tilde{l}_w$ and $e'_w\text{-}\tilde{l}'_w$ determined by Algorithm 1 and the values of $\tilde{\lambda}_w$ and $\tilde{g}_w$ (notice that $\tilde{z} = 16$):

Table 4: Shortest chain between e_w-$\tilde{l}_w$ and e'_w-$\tilde{l}'_w$ $\quad \forall w \in W'$ for $\overline{E}' = \overline{E}$, and values of $\{\tilde{\lambda}_w\}_{w \in W'}$ and $\{\tilde{g}_w\}_{w \in W'}$

w	Shortest chain between e_w-$\tilde{l}_w$ and e'_w-$\tilde{l}'_w$	$\tilde{\lambda}_w$	$\tilde{g}_w$
$(1,2)$	$\{\{1\text{-}1,2\text{-}1\}\}$	7.6	1
$(1,3)$	$\{\{1\text{-}1,2\text{-}1\},\{2\text{-}1,2\text{-}2\},\{2\text{-}2,3\text{-}2\}\}$	13.7	0
$(1,4)$	$\{\{1\text{-}1,5\text{-}1\},\{5\text{-}1,5\text{-}3\},\{5\text{-}3,4\text{-}3\}\}$	13.2	0
$(1,5)$	$\{\{1\text{-}1,5\text{-}1\}\}$	7	1
$(1,6)$	$\{\{1\text{-}1,5\text{-}1\},\{5\text{-}1,6\text{-}1\}\}$	8.4	8
$(2,3)$	$\{\{2\text{-}2,3\text{-}2\}\}$	7.1	1
$(2,4)$	$\{\{2\text{-}1,6\text{-}1\},\{6\text{-}1,5\text{-}1\},\{5\text{-}1,5\text{-}3\},\{5\text{-}3,4\text{-}3\}\}$	15.1	0
$(2,5)$	$\{\{2\text{-}1,6\text{-}1\},\{6\text{-}1,5\text{-}1\}\}$	8.9	1
$(2,6)$	$\{\{2\text{-}1,6\text{-}1\}\}$	7.5	1
$(3,4)$	$\{\{3\text{-}2,2\text{-}2\},\{2\text{-}2,2\text{-}1\},\{2\text{-}1,6\text{-}1\},\{6\text{-}1,5\text{-}1\},\{5\text{-}1,5\text{-}3\},\{5\text{-}3,4\text{-}3\}\}$	21.2	0
$(3,5)$	$\{\{3\text{-}2,2\text{-}2\},\{2\text{-}2,2\text{-}1\},\{2\text{-}1,6\text{-}1\},\{6\text{-}1,5\text{-}1\}\}$	15	0
$(3,6)$	$\{\{3\text{-}2,2\text{-}2\},\{2\text{-}2,2\text{-}1\},\{2\text{-}1,6\text{-}1\}\}$	13.6	1
$(4,5)$	$\{\{4\text{-}3,5\text{-}3\}\}$	7.2	1
$(4,6)$	$\{\{4\text{-}3,5\text{-}3\},\{5\text{-}3,5\text{-}1\},\{5\text{-}1,6\text{-}1\}\}$	13.1	0
$(5,6)$	$\{\{5\text{-}1,6\text{-}1\}\}$	6.9	1

- *Let $\overline{E}' = \overline{E} \cup \{\{1,5\}\}$. The lines defined by Algorithm 2 were $L_1 = \{\{3,2\},\{2,6\},\{6,5\}, \{5,1\}\}$ and $L_2 = \{\{4,5\},\{5,1\},\{1,2\}\}$. Consequently, we get that $\tilde{G} = (\tilde{V},\tilde{E})$, where $\tilde{V} = \{1\text{-}1,1\text{-}2,2\text{-}1,2\text{-}2,3\text{-}1,4\text{-}2,5\text{-}1,5\text{-}2,6\text{-}1\}$ and $\tilde{E} = \{\{1\text{-}1,1\text{-}2\},\{1\text{-}1,5\text{-}1\},\{1\text{-}2,2\text{-}2\}, \{1\text{-}2,5\text{-}2\},\{2\text{-}1,2\text{-}2\},\{2\text{-}1,3\text{-}1\},\{2\text{-}1,6\text{-}1\},\{4\text{-}2,5\text{-}2\},\{5\text{-}1,5\text{-}2\},\{5\text{-}1,6\text{-}1\}\}$ (see Figure 4).*
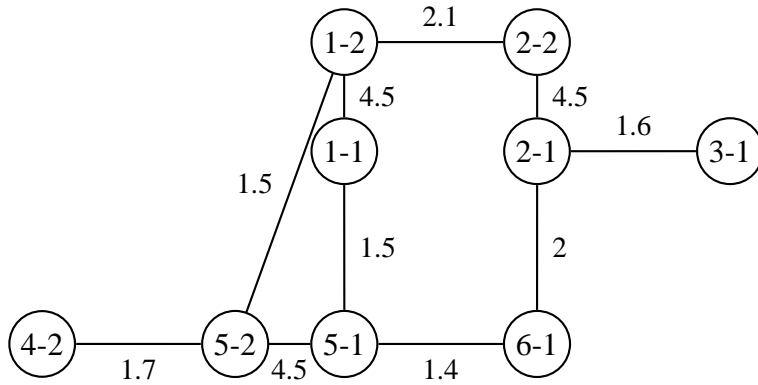

Figure 4: Graphic representation of $\tilde{G} = (\tilde{V},\tilde{E})$ for $\overline{E}' = \overline{E} \cup \{\{1,5\}\}$, and values of $\{\tilde{t}_{i\text{-}l,j\text{-}l'}\}_{\{i\text{-}l,j\text{-}l'\}\in \tilde{E}}$

For each $w \in W'$, Table 5 shows the shortest chain in $\tilde{G}$ joining e_w-$\tilde{l}_w$ and e'_w-$\tilde{l}'_w$ determined by Algorithm 1 and the values of $\tilde{\lambda}_w$ and $\tilde{g}_w$ (notice that $\tilde{z} = 20$):

Table 5: Shortest chain between e_w-$\tilde{l}_w$ and e'_w-$\tilde{l}'_w$ $\quad \forall w \in W'$ for $\overline{E}' = \overline{E} \cup \{\{1,5\}\}$, and values of $\{\tilde{\lambda}_w\}_{w \in W'}$ and $\{\tilde{g}_w\}_{w \in W'}$

w	Shortest chain between e_w-$\tilde{l}_w$ and e'_w-$\tilde{l}'_w$	$\tilde{\lambda}_w$	$\tilde{g}_w$
(1,2)	{{1-2,2-2}}	7.6	1
(1,3)	{{1-1,5-1},{5-1,6-1},{6-1,2-1},{2-1,3-1}}	12	1
(1,4)	{{1-2,5-2},{5-2,4-2}}	8.7	1
(1,5)	{{1-1,5-1}}	7	1
(1,6)	{{1-1,5-1},{5-1,6-1}}	8.4	8
(2,3)	{{2-1,3-1}}	7.1	1
(2,4)	{{2-2,1-2},{1-2,5-2},{5-2,4-2}}	10.8	1
(2,5)	{{2-1,6-1},{6-1,5-1}}	8.9	1
(2,6)	{{2-1,6-1}}	7.5	1
(3,4)	{{3-1,2-1},{2-1,6-1},{6-1,5-1},{5-1,5-2},{5-2,4-2}}	16.7	0
(3,5)	{{3-1,2-1},{2-1,6-1},{6-1,5-1}}	10.5	1
(3,6)	{{3-1,2-1},{2-1,6-1}}	9.1	1
(4,5)	{{4-2,5-2}}	7.2	1
(4,6)	{{4-2,5-2},{5-2,5-1},{5-1,6-1}}	13.1	0
(5,6)	{{5-1,6-1}}	6.9	1

- *Let $\overline{E}' = \overline{E} \cup \{\{5,6\}\}$. The lines defined by Algorithm 2 were $L_1 = \{\{2,6\},\{6,5\},\{5,1\}, \{1,2\}\}$, $L_2 = \{\{3,2\}\}$ and $L_3 = \{\{6,5\},\{5,4\}\}$. Consequently, we get that $\tilde{G} = (\tilde{V}, \tilde{E})$, where $\tilde{V} = \{$1-1, 2-1, 2-2, 3-2, 4-3, 5-1, 5-3, 6-1, 6-3$\}$ and $\tilde{E} = \{\{$1-1,2-1$\}, \{$1-1,5-1$\}, \{$2-1,2-2$\}, \{$2-1,6-1$\}, \{$2-2,3-2$\}, \{$4-3,5-3$\}, \{$5-1,5-3$\}, \{$5-1,6-1$\}, \{$5-3,6-3$\}, \{$6-1,6-3$\}\}$ (see Figure 5).*
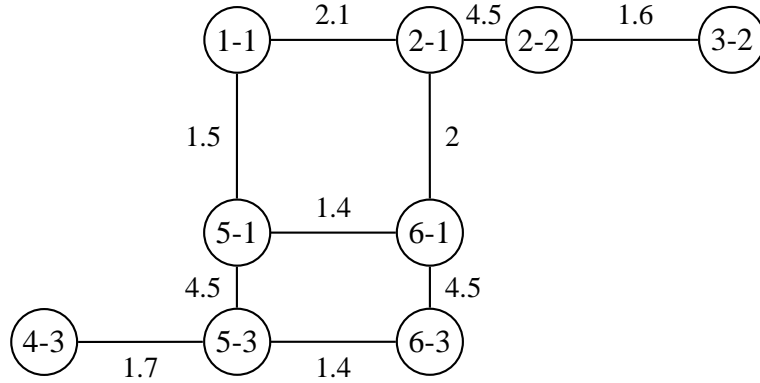

Figure 5: Graphic representation of $\tilde{G} = (\tilde{V}, \tilde{E})$ for $\overline{E}' = \overline{E} \cup \{\{5,6\}\}$, and values of $\{\tilde{t}_{i\text{-}l,j\text{-}l'}\}_{\{i\text{-}l,j\text{-}l'\} \in \tilde{E}}$

For each $w \in W'$, Table 6 shows the shortest chain in $\tilde{G}$ joining e_w-$\tilde{l}_w$ and e'_w-$\tilde{l}'_w$ determined by Algorithm 1 and the values of $\tilde{\lambda}_w$ and $\tilde{g}_w$ (notice that $\tilde{z} = 17$):

Table 6: Shortest chain between e_w-$\tilde{l}_w$ and e'_w-$\tilde{l}'_w$ $\quad \forall w \in W'$ for $\overline{E}' = \overline{E} \cup \{\{5,6\}\}$, and values of $\{\tilde{\lambda}_w\}_{w \in W'}$ and $\{\tilde{g}_w\}_{w \in W'}$

w	Shortest chain between e_w-$\tilde{l}_w$ and e'_w-$\tilde{l}'_w$	$\tilde{\lambda}_w$	$\tilde{g}_w$
$(1,2)$	$\{\{1\text{-}1, 2\text{-}1\}\}$	7.6	1
$(1,3)$	$\{\{1\text{-}1, 2\text{-}1\}, \{2\text{-}1, 2\text{-}2\}, \{2\text{-}2, 3\text{-}2\}\}$	13.7	0
$(1,4)$	$\{\{1\text{-}1, 5\text{-}1\}, \{5\text{-}1, 5\text{-}3\}, \{5\text{-}3, 4\text{-}3\}\}$	13.2	0
$(1,5)$	$\{\{1\text{-}1, 5\text{-}1\}\}$	7	1
$(1,6)$	$\{\{1\text{-}1, 5\text{-}1\}, \{5\text{-}1, 6\text{-}1\}\}$	8.4	8
$(2,3)$	$\{\{2\text{-}2, 3\text{-}2\}\}$	7.1	1
$(2,4)$	$\{\{2\text{-}1, 6\text{-}1\}, \{6\text{-}1, 5\text{-}1\}, \{5\text{-}1, 5\text{-}3\}, \{5\text{-}3, 4\text{-}3\}\}$	15.1	0
$(2,5)$	$\{\{2\text{-}1, 6\text{-}1\}, \{6\text{-}1, 5\text{-}1\}\}$	8.9	1
$(2,6)$	$\{\{2\text{-}1, 6\text{-}1\}\}$	7.5	1
$(3,4)$	$\{\{3\text{-}2, 2\text{-}2\}, \{2\text{-}2, 2\text{-}1\}, \{2\text{-}1, 6\text{-}1\}, \{6\text{-}1, 5\text{-}1\}, \{5\text{-}1, 5\text{-}3\}, \{5\text{-}3, 4\text{-}3\}\}$	21.2	0
$(3,5)$	$\{\{3\text{-}2, 2\text{-}2\}, \{2\text{-}2, 2\text{-}1\}, \{2\text{-}1, 6\text{-}1\}, \{6\text{-}1, 5\text{-}1\}\}$	15	0
$(3,6)$	$\{\{3\text{-}2, 2\text{-}2\}, \{2\text{-}2, 2\text{-}1\}, \{2\text{-}1, 6\text{-}1\}\}$	13.6	1
$(4,5)$	$\{\{4\text{-}3, 5\text{-}3\}\}$	7.2	1
$(4,6)$	$\{\{4\text{-}3, 5\text{-}3\}, \{5\text{-}3, 6\text{-}3\}\}$	8.6	1
$(5,6)$	$\{\{5\text{-}1, 6\text{-}1\}\}$	6.9	1

8 Greedy heuristic procedure for maximizing the total number of trips

In this section we propose a greedy heuristic algorithm for determining a line design for the rapid transit network, allowing pairs of station locations linked by more than one line and attempting to maximize an estimation of the maximum possible number of weekly trips taken on the rapid transit network by the surveyees. Since it can be expected that the smaller the number of transfers that should be made by the surveyees to arrive at their destinations, the higher the number of trips they would take on the rapid transit network, the proposed algorithm will be based on Algorithm 2.

In Algorithm 3 below, the set of edges representing the links of the rapid transit network to be constructed is denoted by E^*, the best set of lines for the rapid transit network obtained by Algorithm 2 is denoted by $L_1^*, \ldots, L_{\hat{l}^*}^*$, and the maximum possible expected number of weekly trips taken on the rapid transit network by the surveyees is denoted by $\tilde{z}^*$ and defined as the value of $\tilde{z}$ corresponding to $L_1^*, \ldots, L_{\hat{l}^*}^*$ (see Section 7).

We initially set $E^* = \overline{E}$ and iteratively add edges to E^* in such a way that conditions (1) and (2) stated in Section 6 are satisfied. Moreover, only the edges that are duplicated in an iteration will be candidates to be subsequently triplicated; only the edges that are triplicated will be candidates to be quadruplicated, and so forth.

At each iteration we consider the set $\hat{E}$ of edges that can potentially be added to the current set E^*. Let $G^* = (V, E^*)$. In order to determine $\hat{E}$, we compute the degree $d^*(i)$ in G^* of each node $i \in V$, the multiplicity m_{ij}^* in G^* of certain edges $\{i,j\} \in \overline{E}$, the remainder b^* of the budget for constructing the rapid transit network defined by G^*, and the construction cost $\hat{c}_{ij}$ of adding to E^* certain edges $\{i,j\} \in \overline{E}$, whose value is given as follows:

$$\hat{c}_{ij} = \begin{cases} c_{ij} & \text{if } d^*(i) \text{ and } d^*(j) \text{ are odd} \\ a_j + c_{ij} & \text{if } d^*(i) \text{ is odd and } d^*(j) \text{ is even} \\ a_i + c_{ij} & \text{if } d^*(i) \text{ is even and } d^*(j) \text{ is odd} \\ a_i + a_j + c_{ij} & \text{if } d^*(i) \text{ and } d^*(j) \text{ are even} \end{cases}$$

Algorithm 3.

Step 1. Apply Algorithm 2 taking $\overline{E}' = \overline{E}$, denote the obtained set of lines by $L_1, \ldots, L_{\hat{l}}$, compute the value of $\tilde{z}$ for $L_1, \ldots, L_{\hat{l}}$, and set $E^* = \overline{E}$, $\hat{l}^* = \hat{l}$, $L_l^* = L_l \quad \forall l \in \{1, \ldots, \hat{l}\}$, $\tilde{z}^* = \tilde{z}$, $d^*(i) = |\overline{\Gamma}(i)| \quad \forall i \in V$ and $b^* = b - \sum_{i \in V} a_i (\overline{\Delta}_i + \overline{\gamma}_i) - \sum_{\{i,j\} \in \overline{E}} c_{ij}$.

Step 2. Compute the value of $\hat{c}_{ij} \quad \forall \{i,j\} \in \overline{E}$ such that $\min\{d^*(i), d^*(j)\} \geq 2$, and set $\hat{E} = \{\{i,j\} \in \overline{E} \mid \min\{d^*(i), d^*(j)\} \geq 2, \hat{c}_{ij} \leq b^*\}$. If $\hat{E} = \emptyset$, STOP; otherwise, set $m_{ij}^* = 1 \quad \forall \{i,j\} \in \hat{E}$.

Step 3. Set $E_0^* = E^*$.

Step 4. Denote $\hat{E}$ by $\{\{i_1, j_1\}, \ldots, \{i_{\hat{m}}, j_{\hat{m}}\}\}$ and set $\hat{c}_0 = 0$ and $k = 1$.

Step 5. Apply Algorithm 2 taking $\overline{E}' = E^* \cup \{\{i_k, j_k\}\}$. If no line design is obtained, go to Step 7. Denote the obtained set of lines by $L_1, \ldots, L_{\hat{l}}$, and compute the value of $\tilde{z}$ for $L_1, \ldots, L_{\hat{l}}$.

Step 6. If $\tilde{z} > \tilde{z}^*$, set $\hat{l}^* = \hat{l}$, $L_l^* = L_l \quad \forall l \in \{1, \ldots, \hat{l}\}$, $\tilde{z}^* = \tilde{z}$, $k^* = k$ and $\hat{c}_0 = \hat{c}_{i_k j_k}$; otherwise, if $\tilde{z} = \tilde{z}^*$ and $\hat{c}_{i_k j_k} < \hat{c}_0$, set $\hat{l}^* = \hat{l}$, $L_l^* = L_l \quad \forall l \in \{1, \ldots, \hat{l}\}$, $k^* = k$ and $\hat{c}_0 = \hat{c}_{i_k j_k}$.

Step 7. If $k < \hat{m}$, set $k = k + 1$ and go to Step 5. If $\hat{c}_0 = 0$, go to Step 10.

Step 8. Set $E^* = E^* \cup \{\{i_{k^*}, j_{k^*}\}\}$, $d^*(i_{k^*}) = d^*(i_{k^*}) + 1$, $d^*(j_{k^*}) = d^*(j_{k^*}) + 1$, $m_{i_{k^*} j_{k^*}}^* = m_{i_{k^*} j_{k^*}}^* + 1$ and $b^* = b^* - \hat{c}_0$.

Step 9. Compute the value of $\hat{c}_{ij} \quad \forall \{i,j\} \in \hat{E} \setminus \{\{i_{k^*}, j_{k^*}\}\}$ and set $\hat{E} = \{\{i,j\} \in \hat{E} \setminus \{\{i_{k^*}, j_{k^*}\}\} \mid \hat{c}_{ij} \leq b^*\}$. If $\hat{E} \neq \emptyset$, go to Step 4.

28

Step 10. Compute the value of $\hat{c}_{ij}$ $\quad \forall \{i,j\} \in E^* \setminus E_0^*$ such that $\min\{d^*(i), d^*(j)\} \geq 2m_{ij}^*$, and set $\hat{E} = \{\{i,j\} \in E^* \setminus E_0^* \mid \min\{d^*(i), d^*(j)\} \geq 2m_{ij}^*, \hat{c}_{ij} \leq b^*\}$. If $\hat{E} = \emptyset$, STOP; otherwise, go to Step 3.

REMARK 2. *In Step 9 it suffices to compute the value of $\hat{c}_{ij}$ if $i = i_{k^*}$ or $i = j_{k^*}$ or $j = i_{k^*}$ or $j = j_{k^*}$, since, otherwise, it remains the same as for the previous considered set E^*; furthermore, $\min\{d^*(i), d^*(j)\} \geq 2m_{ij}^* \quad \forall\{i,j\} \in \hat{E} \setminus \{\{i_{k^*}, j_{k^*}\}\}$.*

Although it is not likely, the final line design obtained by Algorithm 3 could comprise some lines which are contained in other lines, i.e., there could exist $l, l' \in \{1, \ldots, \hat{l}^*\}$ such that $L_l^* \subseteq L_{l'}^*$. In that case, eliminating line L_l^* would not modify the value of $\tilde{z}^*$ and would increase the value of b^*; nevertheless, if the capacity of the resulting rapid transit network were not enough to hold all the demanded trips between the endpoints of L_l^*, it could be preferable not to eliminate it.

Given that we have utilized a heuristic procedure, the final line design obtained by Algorithm 3 could be improved. Therefore, it should be analyzed in detail, searching for potential slight modifications which could result in an increase on the maximum possible expected number of weekly trips taken on the rapid transit network by the surveyees.

Example 5. *Let us continue solving the instance given in Example 1 (see also Examples 2, 3 and 4).*

Algorithm 3 proceeds as follows:

Step 1. $\overline{E}' = \overline{E}$, $L_1 = \{\{2,6\}, \{6,5\}, \{5,1\}, \{1,2\}\}$, $L_2 = \{\{3,2\}\}$, $L_3 = \{\{5,4\}\}$, $\tilde{z} = 16$, $E^* = \{\{1,2\}, \{1,5\}, \{2,3\}, \{2,6\}, \{4,5\}, \{5,6\}\}$, $\hat{l}^* = 3$, $L_1^* = \{\{2,6\}, \{6,5\}, \{5,1\}, \{1,2\}\}$, $L_2^* = \{\{3,2\}\}$, $L_3^* = \{\{5,4\}\}$, $\tilde{z}^* = 16$, $d^*(1) = 2$, $d^*(2) = 3$, $d^*(3) = 1$, $d^*(4) = 1$, $d^*(5) = 3$, $d^*(6) = 2$, $b^* = 76.5$

Step 2. $\hat{c}_{12} = 102$, $\hat{c}_{15} = 75$, $\hat{c}_{26} = 97.5$, $\hat{c}_{56} = 70.5$, $\hat{E} = \{\{1,5\}, \{5,6\}\}$, $m_{15}^* = 1$, $m_{56}^* = 1$

Step 3. $E_0^* = \{\{1,2\}, \{1,5\}, \{2,3\}, \{2,6\}, \{4,5\}, \{5,6\}\}$

Step 4. $\hat{m} = 2$, $\{i_1, j_1\} = \{1,5\}$, $\{i_2, j_2\} = \{5,6\}$, $\hat{c}_0 = 0$, $k = 1$

Step 5. $\overline{E}' = \overline{E} \cup \{\{1,5\}\}$, $L_1 = \{\{3,2\}, \{2,6\}, \{6,5\}, \{5,1\}\}$, $L_2 = \{\{4,5\}, \{5,1\}, \{1,2\}\}$, $\tilde{z} = 20$

Step 6. $\hat{l}^* = 2$, $L_1^* = \{\{3,2\}, \{2,6\}, \{6,5\}, \{5,1\}\}$, $L_2^* = \{\{4,5\}, \{5,1\}, \{1,2\}\}$, $\tilde{z}^* = 20$, $k^* = 1$, $\hat{c}_0 = 75$

Step 7. $k = 2$

Step 5. $\overline{E}' = \overline{E} \cup \{\{5,6\}\}$, $L_1 = \{\{2,6\}, \{6,5\}, \{5,1\}, \{1,2\}\}$, $L_2 = \{\{3,2\}\}$, $L_3 = \{\{6,5\}, \{5,4\}\}$, $\tilde{z} = 17$

Step 8. $E^* = \{\{1,2\}, \{1,5\}, \{1,5\}, \{2,3\}, \{2,6\}, \{4,5\}, \{5,6\}\}$, $d^*(1) = 3$, $d^*(5) = 4$, $m_{15}^* = 2$, $b^* = 1.5$

Step 9. $\hat{c}_{56} = 100.5$, $\hat{E} = \emptyset$

Step 10. $\hat{E} = \emptyset$

Thus, the rapid transit network to be constructed is defined by the graph $G^* = (V, E^*)$, where $E^* = \{\{1,2\}, \{1,5\}, \{1,5\}, \{2,3\}, \{2,6\}, \{4,5\}, \{5,6\}\}$, there are two noncircular lines $L_1^* = \{\{3,2\}, \{2,6\}, \{6,5\}, \{5,1\}\}$ and $L_2^* = \{\{4,5\}, \{5,1\}, \{1,2\}\}$, the construction cost of the rapid transit network is 643.5 million euros, and the maximum possible expected number of weekly trips taken on the rapid transit network by the surveyees is 20.

9 Computational experience

We have randomly generated three example cases $C1$, $C2$ and $C3$ by using the following procedure (see Section 2):

The n key station locations have been randomly generated from a continuous uniform distribution on a square of given side length ρ (expressed in kilometers), in such a way that whichever two of them are at least two kilometers apart (considering the Euclidean distance). The pairs of key station locations that can potentially be linked are the ones with the m shortest Euclidean distances between them.

We have set the parameter values having a metro system in mind. For $C1$ we have taken $\rho = 10$, $n = 15$, $m = 35$, $b = 6000$ and $|\Theta| = 1500$, for $C2$ we have taken $\rho = 10$, $n = 20$, $m = 45$, $b = 8000$ and $|\Theta| = 2000$, and for $C3$ we have taken $\rho = 15$, $n = 25$, $m = 60$, $b = 13000$ and $|\Theta| = 3000$ (b is expressed in millions of euros). For all the example cases we have set the values for the rest of the parameters as follows: $a_i = 30 \quad \forall i \in V$ (expressed in millions of euros), d_{ij} is the Euclidean distance between i and $j \quad \forall (i, j) \in W$ (expressed in kilometers), $s_{ij} = \lceil d_{ij} \rceil - 1 \quad \forall \{i, j\} \in E$, $c_{ij} = 45d_{ij} + 30s_{ij} \quad \forall \{i, j\} \in E$ (expressed in millions of euros), $\overline{v} = 55$, $\overline{t}(i) = 3 \quad \forall i \in V$, $\overline{t}_a = 5$, $\overline{t}_s = 0.6$ and $\overline{t}_r = 4$.

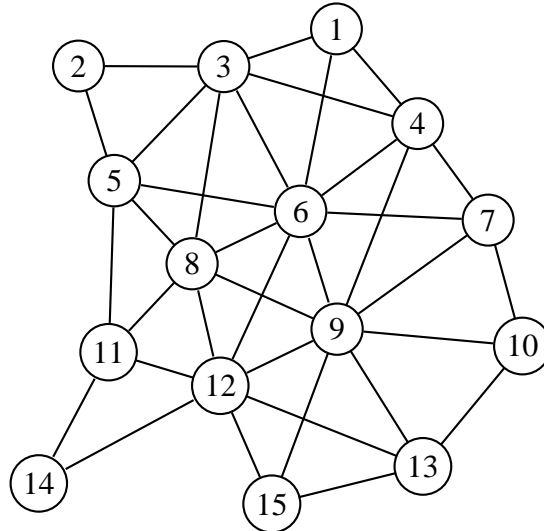Figures 6, 7 and 8 show the underlying graphs of $C1$, $C2$ and $C3$, respectively.

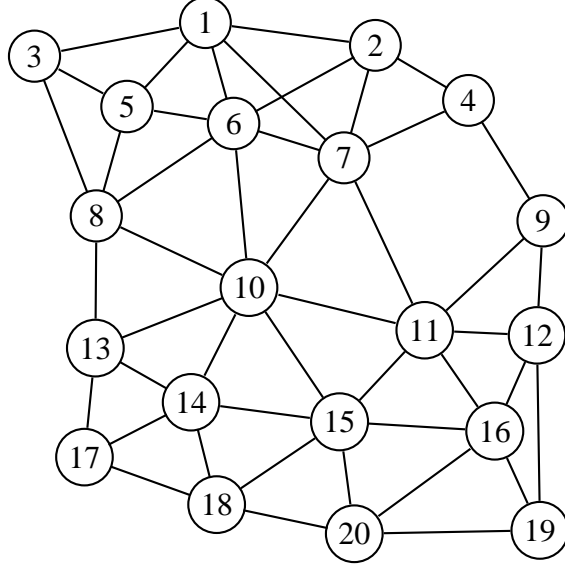Figure 6: Graphic representation of the underlying graph of $C1$

Figure 7: Graphic representation of the underlying graph of *C*2

The survey answers for each $\theta \in \Theta$ have been generated as follows:

Let β_θ be the number of distinct trips that the surveyee θ plans to take on the rapid transit network, without taking into consideration their direction. We generate the value of β_θ in such a way that $P(\beta_\theta = k) = 0.3 \quad \forall k \in \{1,2\}$, $P(\beta_\theta = 3) = 0.2$ and $P(\beta_\theta = k) = 0.1 \quad \forall k \in \{4,5\}$.

Let us denote $W = \{w_1, \ldots, w_{\frac{n(n-1)}{2}}\}$ and $W_\theta = \{w_1(\theta), \ldots, w_{\beta_\theta}(\theta)\}$. For each $k \in \{1, \ldots, \beta_\theta\}$, we randomly generate a value φ_k from a discrete uniform distribution on $\{1, \ldots, \frac{n(n-1)}{2}\} \backslash \bigcup_{k'=1}^{k-1} \{\varphi_{k'}\}$ and we set $w_k(\theta) = w_{\varphi_k}$.

Let η_θ be the number of weekly trips that the surveyee θ plans to take on the rapid transit network. We generate the value of η_θ in such a way that $P(\eta_\theta = k) = \frac{1}{2(17-\beta_\theta)} \quad \forall k \in \{\beta_\theta, \ldots, 9\} \cup \{15, \ldots, 21\}$ and $P(\eta_\theta = k) = 0.1 \quad \forall k \in \{10, \ldots, 14\}$.

For each $k \in \{1, \ldots, \beta_\theta - 1\}$, we generate the value of $\alpha_{w_k(\theta)}(\theta)$ from a discrete uniform distribution on $\{1, \ldots, \eta_\theta - \sum_{k'=1}^{k-1} \alpha_{w_{k'}(\theta)}(\theta) - \beta_\theta + k\}$ and we set $\alpha_{w_{\beta_\theta}(\theta)}(\theta) = \eta_\theta - \sum_{k'=1}^{\beta_\theta - 1} \alpha_{w_{k'}(\theta)}(\theta)$ and $\alpha_w(\theta) = 0 \quad \forall w \in W \backslash W_\theta$.

For each $w \in W_\theta$, we generate the value of $\tau_w(\theta)$ from a discrete uniform distribution on $\left\{ \left\lceil \bar{t}(e_w) + \frac{\bar{t}_a}{2} + \frac{60}{\bar{v}} d_{e_w e'_w} + \bar{t}(e'_w) \right\rceil, \ldots, \left\lceil \left(\bar{t}(e_w) + \frac{\bar{t}_a}{2} + \frac{60}{\bar{v}} d_{e_w e'_w} + \bar{t}(e'_w) \right) \left(1 + \frac{3}{2} \sqrt{\frac{\bar{d}}{d_{e_w e'_w}}} \right) \right\rceil \right\}$, where $\bar{d} = \max\{d_{e_w e'_w} \mid w \in W\}$.

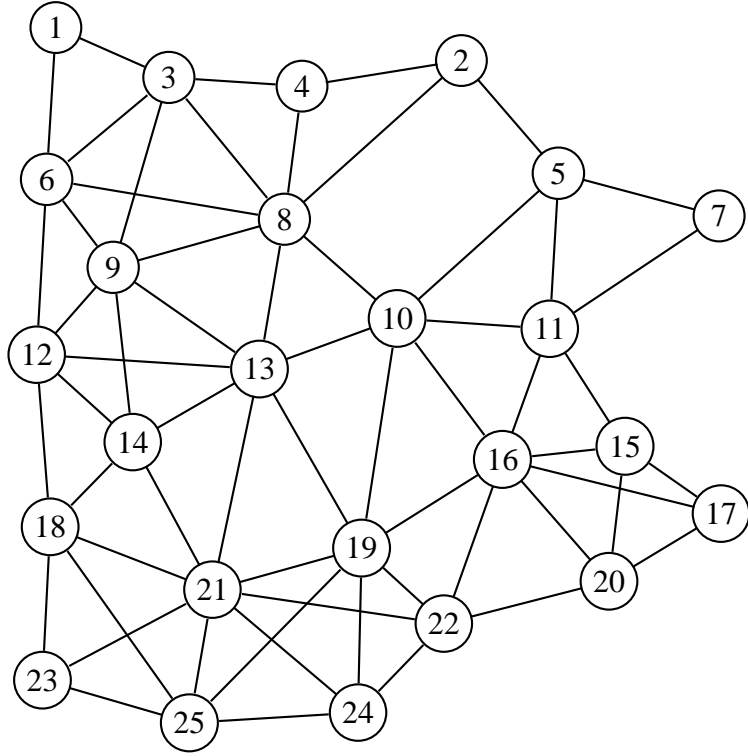For all the example cases it has been obtained that $W' = W$.

Figure 8: Graphic representation of the underlying graph of *C*3

Let us consider an upper bound $\overline{q} \in \mathbb{N}$ for the values of $\{q(w)\}_{w \in W'}$, and let $\mu = \frac{\max\{g_w^1 | w \in W'\}}{\overline{q}}$ (notice that $g_w^1 = \sum_{\theta \in \Theta_w} \alpha_w(\theta) \quad \forall w \in W'$). In the computational results reported below each example case will be solved considering several values for $\overline{q}$.

For each $w \in W'$, the procedure we utilize for computing the values of $q(w)$, $\{u_w^k\}_{k \in \{2,...,q(w)\}}$ and $\{g_w^k\}_{k \in \{2,...,q(w)\}}$ is given by Algorithm 4; for simplicity of notation, without loss of generality we assume that $\Theta_w = \{\theta_1, \ldots, \theta_{o_w}\}$ and $\tau_w(\theta_1) \leq \ldots \leq \tau_w(\theta_{o_w})$ (notice that $u_w^1 = \tau_w(\theta_1)$).

Algorithm 4.

Step 1. Set $i = 1$ and $k = 1$.

Step 2. If $g_w^k \leq \mu$, set $q(w) = k$ and STOP; otherwise, set $j = i$.

Step 3. Set $j = \max\{j' \in \{j, \ldots, o_w\} \mid \tau_w(\theta_{j'}) = \tau_w(\theta_j)\}$.

Step 4. If $j = o_w$, set $q(w) = k$ and STOP.

Step 5. If $\sum_{j'=i}^{j} \alpha_w(\theta_{j'}) \geq \mu$, set $k = k+1$, $u_w^k = \tau_w(\theta_{j+1})$, $g_w^k = g_w^{k-1} - \sum_{j'=i}^{j} \alpha_w(\theta_{j'})$, $i = j+1$ and go to Step 2; otherwise, set $j = j+1$ and go to Step 3.

Given that any preliminary route for travelling between two key station locations will utilize at most $n-1$ edges in E (see Section 3), we have taken $M_w^k = \bar{t}'(w) + \bar{t}'_{n-1} - u_w^k \quad \forall w \in W'$, $\forall k \in \{1, \ldots, q(w)\}$, where $\bar{t}'_{n-1}$ is the sum of the $n-1$ greatest values in $\{\bar{t}'_{ij}\}_{\{i,j\} \in E}$ (we have considered some other settings for the values of $\{M_w^k\}_{w \in W', k \in \{1, \ldots, q(w)\}}$, but the best general computational results have been obtained with these ones).

In Steps 1 and 5 of Algorithm 3 we have taken as the set of lines $L_1, \ldots, L_{\hat{l}}$ the one with the greatest value of $\tilde{z}$ obtained by repeatedly applying Algorithm 2 for two minutes, choosing the nodes i_0, j and k in its Steps 3, 4, 9 and 10 randomly and uniformly distributed over the set of all of their possible values.

For all the example cases Algorithm 3 has duplicated some links, but it has not managed to triplicate any of them.

The implementation platform has been Microsoft Visual C++ 2005, CPLEX v12.1 and Pentium 4, 3.00 GHz, 1.00 Gb RAM.

In order to solve the model stated in Section 3, we have run the CPLEX mixed integer optimizer by using the default rules except that the relative and absolute optimality tolerances have been set to zero and, in the branching process, the priorities for the variables $\{x_{ij}\}_{\{i,j\} \in E}$ and $\{p_w^k\}_{w \in W', k \in \{1, \ldots, q(w)\}}$ have been set to 1 and 2, respectively (we have considered many other settings for the priority values, but the best general computational results have been obtained with these ones).

Tables 7, 8 and 9 show, respectively, the computational results obtained for $C1$, $C2$ and $C3$ by considering several values for $\overline{q}$.

The columns headed "z^*", "*Nodes*", "*M. time*", "$\overline{s}$" and "$\overline{z}$" give, respectively, the optimal value of the objective function of the model stated in Section 3, the number of branch-and-cut nodes evaluated for solving that model, the related CPU time expressed in seconds, and the values of $\overline{s}$ and $\overline{z}$ for the optimal solution obtained (the CPU time required for computing the value of $\overline{z}$ by using Algorithm 1 has been less than 0.04 seconds for all the considered instances).

The columns headed "$\tilde{z}_0^*$", "$\tilde{z}^*$", "*Dup.*", "b^*" and "*T. time*" give, respectively, the greatest value of $\tilde{z}$ obtained in Step 1 of Algorithm 3 (i.e., without allowing pairs of station locations linked by more than one line), the value of $\tilde{z}^*$, the number of duplicated links in the best line design obtained, the related value of b^* expressed in millions of euros, and the total required CPU time expressed in seconds (including the time for solving the model stated in Section 3 and for computing the value of $\overline{z}$).

Table 7: Computational results for $C1$

$\overline{q}$	z^*	Nodes	M. time	$\overline{s}$	$\overline{z}$	$\tilde{z}_0^*$	$\tilde{z}^*$	Dup.	b^*	T. time
1	4094	8	13	41	16357	14809	14872	1	27.4728	373
2	6810	110	11	41	16302	14891	15037	1	12.8641	2052
3	10047	289	11	40	16343	15025	15172	1	21.2142	1211
4	11947	254	9	41	16299	14738	14738	0	112.198	129
5	12891	310	10	41	16302	14891	15037	1	12.8641	2050
6	13604	352	12	40	16284	14946	15042	1	27.9963	1692
7	14106	188	8	40	16327	14570	14685	1	9.01539	1328
8	14456	710	50	42	16407	14692	14692	0	62.2447	170
9	14556	794	43	42	16362	14651	14651	0	73.8903	163
10	14725	1671	183	41	16357	14809	14872	1	27.4728	543
11	14927	2919	250	40	16345	14990	15180	1	8.07191	2290
12	15071	1221	81	41	16347	14873	14873	0	113.373	201
13	15123	1986	173	41	16347	14873	14873	0	113.373	293
14	15183	698	64	42	16395	14780	14780	0	69.6482	184
15	15253	1596	141	42	16395	14780	14780	0	69.6482	261
16	15407	2525	200	41	16347	14873	14873	0	113.373	320
17	15418	1511	127	41	16347	14873	14873	0	113.373	247
18	15418	1511	127	41	16347	14873	14873	0	113.373	247
19	15555	2869	195	42	16439	15036	15036	0	5.47454	315
20	15603	2717	198	42	16439	15036	15036	0	5.47454	318

For $C1$, the greatest value of $\tilde{z}^*$ is 15180, and it is reached for $\overline{q} = 11$. Links $\{3,4\}$, $\{4,9\}$, $\{5,11\}$, $\{6,7\}$, $\{9,15\}$, $\{12,13\}$ and $\{12,14\}$ are not constructed, and link $\{5,8\}$ is duplicated. There are 55 distinct station locations, and the best set of lines obtained is $L_1^* = \{\{8,3\},\{3,6\},\{6,9\},\{9,13\}\}$, $L_2^* = \{\{3,5\},\{5,6\}\}$, $L_3^* = \{\{1,6\},\{6,8\},\{8,11\},\{11,14\}\}$, $L_4^* = \{\{5,8\},\{8,9\},\{9,10\}\}$, $L_5^* = \{\{7,9\},\{9,12\},\{12,11\}\}$, $L_6^* = \{\{4,6\},\{6,12\}\}$ and $L_7^* = \{\{1,3\},\{3,2\},\{2,5\},\{5,8\},\{8,12\},\{12,15\},\{15,13\},\{13,10\},\{10,7\},\{7,4\},\{4,1\}\}$.

For $C2$, the greatest value of $\tilde{z}^*$ is 19296, and it is reached for $\overline{q} \in \{9,10,11,12,13,16\}$. Links $\{1,2\}$, $\{1,3\}$, $\{9,11\}$ and $\{16,20\}$ are not constructed, and no link is duplicated. There are 75 distinct station locations, and the best set of lines obtained is $L_1^* = \{\{1,6\},\{6,10\},\{10,15\},\{15,20\}\}$, $L_2^* = \{\{3,5\},\{5,6\},\{6,7\},\{7,11\},\{11,16\},\{16,19\},\{19,20\},\{20,18\},\{18,17\},\{17,13\},\{13,10\},\{10,8\},\{8,3\}\}$, $L_3^* = \{\{4,2\},\{2,7\},\{7,10\},\{10,14\},\{14,18\},\{18,15\},\{15,16\},\{16,12\},\{12,19\}\}$, $L_4^* = \{\{10,11\}\}$, $L_5^* = \{\{2,6\},\{6,8\}\}$, $L_6^* = \{\{1,5\},\{5,8\},\{8,13\},\{13,14\},\{14,15\},\{15,11\},\{11,12\},\{12,9\},\{9,4\},\{4,7\},\{7,1\}\}$ and $L_7^* = \{\{14,17\}\}$.

Table 8: Computational results for *C*2

$\overline{q}$	z^*	Nodes	M. time	$\overline{s}$	$\overline{z}$	$\tilde{z}_0^*$	$\tilde{z}^*$	Dup.	b^*	T. time
1	4283	715	248	55	20912	19050	19106	1	0.496186	728
2	6929	1421	178	54	21090	19093	19228	1	8.73638	2458
3	11294	1464	204	55	21093	18890	18890	0	74.2944	324
4	13846	552	124	55	21187	18835	18835	0	54.7169	244
5	15581	1024	177	55	21164	18743	18743	0	95.5954	297
6	16713	1272	213	55	21180	18889	18889	0	22.5063	334
7	17461	684	125	55	21168	18835	18866	1	2.93431	605
8	18009	574	97	54	21135	18841	19074	1	11.5397	3097
9	18326	1781	312	55	21175	19296	19296	0	2.98694	432
10	18530	862	138	55	21175	19296	19296	0	2.98694	258
11	18938	887	133	55	21175	19296	19296	0	2.98694	253
12	19178	1336	154	55	21175	19296	19296	0	2.98694	274
13	19362	813	115	55	21175	19296	19296	0	2.98694	235
14	19447	1547	316	55	21172	19253	19253	0	12.398	436
15	19565	2279	319	55	21189	19256	19256	0	33.036	439
16	19681	5202	1458	55	21175	19296	19296	0	2.98694	1578
17	19734	2156	593	55	21189	19256	19256	0	33.036	713
18	19734	2156	591	55	21189	19256	19256	0	33.036	711
19	19910	3900	1991	55	21189	19256	19256	0	33.036	2111
20	19910	3900	1985	55	21189	19256	19256	0	33.036	2105

For *C*3, the greatest value of $\tilde{z}^*$ is 28870, and it is reached for $\overline{q} = 3$. Links $\{3,9\}$, $\{6,8\}$, $\{9,12\}$, $\{10,19\}$, $\{12,13\}$, $\{13,21\}$, $\{16,17\}$ and $\{21,24\}$ are not constructed, and links $\{3,4\}$, $\{6,9\}$, $\{8,10\}$, $\{12,18\}$, $\{13,14\}$ and $\{13,19\}$ are duplicated. There are 111 distinct station locations, and the best set of lines obtained is $L_1^* = \{\{2,8\},\{8,9\}\}$, $L_2^* = \{\{19,25\}\}$, $L_3^* = \{\{13,19\}, \{19,24\}\}$, $L_4^* = \{\{15,16\}, \{16,20\}, \{20,15\}\}$, $L_5^* = \{\{1,3\}, \{3,8\}, \{8,10\}, \{10,16\}, \{16,22\}, \{22,24\}, \{24,25\}, \{25,21\}, \{21,14\}, \{14,9\}, \{9,6\}, \{6,1\}\}$, $L_6^* = \{\{2,4\}, \{4,3\}, \{3,6\}, \{6,12\}, \{12,18\}, \{18,25\}, \{25,23\}, \{23,21\}, \{21,19\}, \{19,16\}, \{16,11\}, \{11,5\}, \{5,2\}\}$, $L_7^* = \{\{3,4\}, \{4,8\}, \{8,13\}, \{13,14\}, \{14,18\}, \{18,23\}\}$, $L_8^* = \{\{8,10\}, \{10,5\}, \{5,7\}, \{7,11\}\}$, $L_9^* = \{\{6,9\}, \{9,13\}, \{13,19\}, \{19,22\}\}$ and $L_{10}^* = \{\{10,11\}, \{11,15\}, \{15,17\}, \{17,20\}, \{20,22\}, \{22,21\}, \{21,18\}, \{18,12\}, \{12,14\}, \{14,13\}, \{13,10\}\}$.

Table 9: Computational results for $C3$

$\bar{q}$	z^*	Nodes	M. time	$\bar{s}$	$\bar{z}$	$\tilde{z}_0^*$	$\tilde{z}^*$	Dup.	b^*	T. time
1	5370	941	605	96	31514	28586	28715	1	28.2064	2645
2	9108	1547	651	97	31407	28272	28272	0	105.962	771
3	15669	3979	1090	86	31330	28266	28870	6	53.718	36850
4	19872	3966	700	96	31457	28250	28627	1	8.48011	2620
5	22507	2549	529	97	31542	28291	28291	0	31.001	650
6	24399	1551	552	96	31480	28507	28507	0	127.765	792
7	25464	1412	465	96	31527	28427	28427	0	136.933	705
8	26260	1706	861	97	31550	28273	28273	0	4.28986	981
9	26790	1071	529	97	31490	28621	28621	0	33.6817	649
10	27332	2507	613	95	31495	28585	28706	1	64.9178	5773
11	27618	757	492	95	31519	28581	28717	1	110.181	5652
12	28172	1222	422	95	31488	28286	28674	1	92.4354	5942
13	28575	1450	848	97	31548	28665	28665	0	78.1124	968
14	28692	1836	469	95	31534	28605	28825	1	12.9417	4069
15	28919	10320	5048	95	31550	28617	28770	1	91.2356	9608
16	29106	1448	456	95	31550	28617	28770	1	91.2356	5016
17	29308	846	301	97	31545	28562	28562	0	54.1833	421
18	29308	846	298	97	31545	28562	28562	0	54.1833	418
19	29458	1610	461	96	31545	28783	28783	0	116.128	581
20	29458	1610	461	96	31545	28783	28783	0	116.128	582

The shape of the part of line L_6^* joining locations 18 and 21 is not usual in real-life rapid transit networks. In order to avoid it, we can consider the following modifications to $L_1^*, \ldots, L_{10}^*$ (notice that these modifications do not affect the construction cost of the rapid transit network):

- Let $L_j = L_j^*$ $\forall j \in \{1,2,3,4,7,8,9,10\}$, $L_5 = \{\{1,3\}, \{3,8\}, \{8,10\}, \{10,16\}, \{16,22\}, \{22,24\}, \{24,25\}, \{25,23\}, \{23,21\}, \{21,14\}, \{14,9\}, \{9,6\}, \{6,1\}\}$ and $L_6 = \{\{2,4\}, \{4,3\}, \{3,6\}, \{6,12\}, \{12,18\}, \{18,25\}, \{25,21\}, \{21,19\}, \{19,16\}, \{16,11\}, \{11,5\}, \{5,2\}\}$. Then the value of $\tilde{z}$ corresponding to $L_1, \ldots, L_{10}$ is 28677, which is less than 28870, hence this set of lines is worse than $L_1^*, \ldots, L_{10}^*$.

- Let $L_j = L_j^*$ $\forall j \in \{1,2,3,4,5,8,9,10\}$, $L_6 = \{\{2,4\}, \{4,3\}, \{3,6\}, \{6,12\}, \{12,18\}, \{18,23\}, \{23,21\}, \{21,19\}, \{19,16\}, \{16,11\}, \{11,5\}, \{5,2\}\}$ and $L_7 = \{\{3,4\}, \{4,8\}, \{8,13\}, \{13,14\}, \{14,18\}, \{18,25\}, \{25,23\}\}$. Then the value of $\tilde{z}$ corresponding to $L_1, \ldots, L_{10}$ is 28830, which is less than 28870, hence this set of lines is worse than $L_1^*, \ldots, L_{10}^*$.

- Let $L_j = L_j^*$ $\forall j \in \{1,3,4,5,8,9,10\}$, $L_2 = \{\{19,25\}, \{25,23\}\}$, $L_6 = \{\{2,4\}, \{4,3\},$ $\{3,6\}, \{6,12\}, \{12,18\}, \{18,23\}, \{23,21\}, \{21,19\}, \{19,16\}, \{16,11\}, \{11,5\}, \{5,2\}\}$ and $L_7 = \{\{3,4\}, \{4,8\}, \{8,13\}, \{13,14\}, \{14,18\}, \{18,25\}\}$. Then the value of $\tilde{z}$ corresponding to $L_1, \ldots, L_{10}$ is 28830, which is less than 28870, hence this set of lines is worse than $L_1^*, \ldots, L_{10}^*$.

- Let $L_j = L_j^*$ $\forall j \in \{1,2,3,4,7,8,9\}$, $L_5 = \{\{1,3\}, \{3,8\}, \{8,10\}, \{10,16\}, \{16,22\},$ $\{22,24\}, \{24,25\}, \{25,23\}, \{23,21\}, \{21,14\}, \{14,9\}, \{9,6\}, \{6,1\}\}$, $L_6 = \{\{2,4\},$ $\{4,3\}, \{3,6\}, \{6,12\}, \{12,18\}, \{18,21\}, \{21,19\}, \{19,16\}, \{16,11\}, \{11,5\}, \{5,2\}\}$ and $L_{10} = \{\{10,11\}, \{11,15\}, \{15,17\}, \{17,20\}, \{20,22\}, \{22,21\}, \{21,25\}, \{25,18\}, \{18,12\},$ $\{12,14\}, \{14,13\}, \{13,10\}\}$. Then the value of $\tilde{z}$ corresponding to $L_1, \ldots, L_{10}$ is 28553, which is less than 28870, hence this set of lines is worse than $L_1^*, \ldots, L_{10}^*$.

- Let $L_j = L_j^*$ $\forall j \in \{1,2,3,4,5,8,9\}$, $L_6 = \{\{2,4\}, \{4,3\}, \{3,6\}, \{6,12\}, \{12,18\}, \{18,21\},$ $\{21,19\}, \{19,16\}, \{16,11\}, \{11,5\}, \{5,2\}\}$, $L_7 = \{\{3,4\}, \{4,8\}, \{8,13\}, \{13,14\}, \{14,18\},$ $\{18,25\}, \{25,23\}\}$ and $L_{10} = \{\{10,11\}, \{11,15\}, \{15,17\}, \{17,20\}, \{20,22\}, \{22,21\},$ $\{21,23\}, \{23,18\}, \{18,12\}, \{12,14\}, \{14,13\}, \{13,10\}\}$. Then the value of $\tilde{z}$ corresponding to $L_1, \ldots, L_{10}$ is 28667, which is less than 28870, hence this set of lines is worse than $L_1^*, \ldots, L_{10}^*$.

- Let $L_j = L_j^*$ $\forall j \in \{1,3,4,5,8,9\}$, $L_2 = \{\{19,25\}, \{25,23\}\}$, $L_6 = \{\{2,4\}, \{4,3\}, \{3,6\},$ $\{6,12\}, \{12,18\}, \{18,21\}, \{21,19\}, \{19,16\}, \{16,11\}, \{11,5\}, \{5,2\}\}$, $L_7 = \{\{3,4\}, \{4,8\},$ $\{8,13\}, \{13,14\}, \{14,18\}, \{18,25\}\}$ and $L_{10} = \{\{10,11\}, \{11,15\}, \{15,17\}, \{17,20\},$ $\{20,22\}, \{22,21\}, \{21,23\}, \{23,18\}, \{18,12\}, \{12,14\}, \{14,13\}, \{13,10\}\}$. Then the value of $\tilde{z}$ corresponding to $L_1, \ldots, L_{10}$ is 28684, which is less than 28870, hence this set of lines is worse than $L_1^*, \ldots, L_{10}^*$.

If we had considered the all-or-nothing model for the potential users' behavior from Marín (2007) and we had assumed that whichever two station locations are linked by one line at most, as proposed in Escudero and Muñoz (2009b), then the maximum possible expected number of weekly trips taken on the rapid transit network by the surveyees would have been given by the value of $\tilde{z}_0^*$ for $\overline{q} = 1$. Consequently, the line designs for $C1$, $C2$ and $C3$ obtained by the approach presented in this paper have produced, respectively, an increase of 371, 246 and 284 trips over the line designs that would have been obtained by means of the procedure proposed in Escudero and Muñoz (2009b) (notice that these increases refer to the number of weekly trips taken on the rapid transit network by the surveyees; therefore, the increases corresponding to the number of annual trips taken on the rapid transit network by all the users are expected to be much higher).

It is worth noting that, despite the tested instances being of considerably larger size than the ones tested in other works in the literature (see e.g. Guan et al. (2006), Marín (2007) and Laporte et al. (2007, 2010a, 2010b)), we have managed to deal with all of them within a reasonable computational effort.

Obviously, the ideal situation would be to be able to know a priori the value of $\overline{q}$ that will give rise to the best rapid transit network, but this is not possible. Thus, we propose to impose a limit for the overall CPU time for designing the rapid transit network, and apply successively our approach considering different values for $\overline{q}$ until reaching this time limit. If the values of $\overline{z}$ and $\overline{s}$ for a certain value of $\overline{q}$ are the same as for some other value of $\overline{q}$ previously considered, then it is very likely to obtain the same line design as that of the previous value of $\overline{q}$; consequently, in this case we propose not to apply Algorithm 3, but to go on to another value of $\overline{q}$. The amount of time for repeatedly applying Algorithm 2 in Steps 1 and 5 of Algorithm 3 will be set depending on the instance size.

10   Conclusions and future research

We have presented a two-stage approach for designing rapid transit networks which is based on another approach that we described elsewhere.

Whereas most of the procedures that can be found in the literature compute their objective function values by means of a given static O-D matrix, we have proposed to perform a survey amongst the potential users of the rapid transit network. The survey results make it possible to consider each potential user's behavior individually, which allows to compute our objective function value (i.e., the expected number of trips through the rapid transit network) in a more accurate way.

The model to be solved in the first stage for selecting the stations and links to be constructed without exceeding the available budget does not take into account the transfer times for the users, since no line design is available at this stage. These transfer times are considered in the second stage, where a greedy heuristic procedure is applied for generating a set of lines for the rapid transit network, allowing pairs of station locations linked by more than one line. This procedure could also be used for redesigning the lines of existing rapid transit networks.

The computational results have shown that, in a relatively small time, our approach can handle instances of larger size than other procedures taken from the literature, as well as obtain better line designs than the approach that we described elsewhere.

In order to deal with large-size instances, we are working on preprocessing techniques for solving the model proposed for the first stage of the approach. We are also working on the problem of determining the headways for the lines.

Acknowledgements

References

Baaj, M.H., Mahmassani, H.S., 1991. An AI-based approach for transit route system planning and design. Journal of Advanced Transportation 25 (2), 187-209.

Bussieck, M.R., Kreuzer, P., Zimmermann, U.T., 1997. Optimal lines for railway systems. European Journal of Operational Research 96 (1), 54-63.

Bussieck, M.R., Lindner, T., Lübbecke, M.E., 2004. A fast algorithm for near cost optimal line plans. Mathematical Methods of Operations Research 59 (2), 205-220.

Claessens, M.T., van Dijk, N.M., Zwaneveld, P.J., 1998. Cost optimal allocation of rail passenger lines. European Journal of Operational Research 110 (3), 474-489.

Escudero, L.F., Muñoz, S., 2009a. An approach for solving a modification of the extended rapid transit network design problem. Top 17 (2), 320-334.

Escudero, L.F., Muñoz, S., 2009b. An approach for designing connected rapid transit networks considering transfers. Technical Reports on Statistics and Decision Sciences, TR09/01, Universidad Rey Juan Carlos, Spain.

Goossens, J.-W., van Hoesel, S., Kroon, L., 2004. A branch-and-cut approach for solving railway line-planning problems. Transportation Science 38 (3), 379-393.

Guan, J.F., Yang, H., Wirasinghe, S.C., 2006. Simultaneous optimization of transit line configuration and passenger line assignment. Transportation Research Part B 40 (10), 885-902.

Korte, B., Vygen, J., 2008. Combinatorial Optimization. Theory and Algorithms, 4th ed., Springer, Berlin.

Laporte, G., Marín, Á., Mesa, J.A., Ortega, F.A., 2007. An integrated methodology for the rapid transit network design problem. In: Geraets, F., Kroon, L., Schoebel, A., Wagner, D., Zaroliagis, C.D. (Eds.), Algorithmic methods for railway optimization. Lecture Notes in Computer Science 4359, 187-199.

Laporte, G., Marín, Á., Mesa, J.A., Perea, F., 2010a. Designing robust rapid transit networks with alternative routes. Journal of Advanced Transportation (in press).

Laporte, G., Mesa, J.A., Perea, F., 2010b. A game theoretic framework for the robust railway transit network design problem. Transportation Research Part B 44 (4), 447-459.

Mandl, C.E., 1980. Evaluation and optimization of urban public transportation networks. European Journal of Operational Research 5 (6), 396-404.

Marín, Á., 2007. An extension to rapid transit network design problem. Top 15 (2), 231-241.

Marín, Á., García-Ródenas, R., 2009. Location of infrastructure in urban railway networks. Computers and Operations Research 36 (5), 1461-1477.