

20th EURO Working Group on Transportation Meeting, EWGT 2017, 4-6 September 2017,  
Budapest, Hungary

## Optimization based planning of Pedibus lines: an arc based approach

Federico Malucelli<sup>a</sup>, Maddalena Nonato<sup>b,\*</sup>, Emanuele Tresoldi<sup>a</sup>

<sup>a</sup>DEIB, Dipartimento di Elettronica, Informazione e Bioingegneria, Politecnico di Milano, P.zza Leonardo da Vinci 32, Milano 20133, Italy

<sup>b</sup>EnDiF, Dipartimento di Ingegneria, Università di Ferrara, Via Giuseppe Saragat 1, Ferrara 44122, Italy

---

### Abstract

Pedibus, also known as the *Walking School Bus*, is a popular system in Western countries aimed at increasing the percentage of children walking to school, reducing vehicular congestion at school gates, and legitimating walking as a mobility mode. In its simplest version, a Pedibus line is a sequence of stops starting from a child home, visiting a sequence of other children's home, and ending at the school. The service is usually run by volunteers, according to common sense based rules. This paper aims at providing optimization based methodological support to decision makers. The line design problem can be described as follows: given the school location, the children home addresses, and the distance between each pair of locations, we have to design a minimum number of lines rooted at the school so that each location belongs to one line and the distance from school to each location along the line is below a given threshold. The objective function is due to the need for adults supervising each line, whose limited availability may hamper the service long term viability. A secondary objective encourages line merging before destination. Heuristic solution approaches to the design of Pedibus lines have been proposed in the literature, considering Pedibus as a mere application of the school bus routing problem. We propose a new arc-based model tailored on the Pedibus features, i.e., allowing lines merging, which yields a constrained spanning tree network structure. Tests on real and realistic networks show that small and medium size instances are solved to optimality, while the weak linear relaxation of the proposed arc model prevents fast convergence so that largest instances with longest walking distances are solved heuristically. This work paves the way to further studies on path based models to speed up convergence to optimality and to encompass different Pedibus variants.

© 2017 The Authors. Published by Elsevier B.V.

Peer-review under responsibility of the scientific committee of the 20th EURO Working Group on Transportation Meeting.

*Keywords:* Pedibus lines design; Walking School Bus; Sustainable mobility;

---

### 1. Introduction and motivations

In modern societies most transfers are made by car or motorized vehicles in general, while active modes such as walking or biking have been progressively let down in favor of passive modes. Taking children to school makes not exception up to the point that cars gathering at school grounds at entry and exit time causes heavy traffic congestion while the vehicles' pollutant emissions make the environment particularly unhealthy, to children detriment. Increased

---

\* Corresponding author. Tel.: +39-0532-974994 ; fax: +39-0532-974870.

E-mail address: [maddalena.nonato@unife.it](mailto:maddalena.nonato@unife.it)

children obesity adds to the picture, making the need for a change a priority, see McDonald (2007). However, changing people behaviors requires to understand which factors affect the decision of families of whether walking children to school. Some obstacles that prevent such mode choice, such as safety concerns along the way or lack of parent's time for shepherding, can be lifted by a particular initiative set up by schools together with local communities. Its name is *Pedibus*, also known as the *Walking School Bus*, according to which children are picked up at home (or reach a nearby line stop) and walk to school with peers, turning a functional activity into a recreational one. Adults supervision is compulsory in most organizations (Japan is a notable exception) but parents are usually not able to provide it nor they are willing to take the responsibility for someone else children. In *Pedibus* this role is played by volunteers or social service workers. In most cases, *Pedibus* starts as a private initiatives of few families taking turns to shepherd children to become as structured as a route with meeting points, a timetable, and a regularly rotated schedule of trained volunteers. According to the guidelines provided by credited organizations such as *SafeRoutesToSchool* (SRTS) SRTS (2017), lines are set up in a greedy way and once designed can hardly accommodate new pupils without affecting service quality. The more *Pedibus* gains momentum the more the served community grows in size, and manual service management struggles to face supervisors shortages. As the service turns from experimental and episodic to systematic and new pupils cohorts enroll every school year, also good quality itineraries get harder to be devised. Indeed, routes duration is a critical issue, since having to walk for a distance much longer than the shortest path may discourage potential users and affect families morning schedules, see McDonald and Aalborg (2009). Optimization based decision support systems as the one here presented come into play to minimize the number of needed supervisors and plan the routes to avoid circuitous itineraries. In the rest of the paper we review the literature on this subject (Sect.2), present the combinatorial optimization problem underlying *Pedibus* lines planning, introduce an arc-based Mixed Integer Linear Programming (MILP) model (Sect.3), test it on random instances and apply it to realistic data for the city of Ferrara, Italy, discuss results (Sect.4) and sketch on going works (Sect.5).

## 2. *Pedibus* variants and state of the art

The scarce literature on this subject can be divided into application devoted and optimization oriented.

**Analysis of current practices.** Regarding practical experiences, a first body of research investigates the reasons why children living within walking distance from school are driven to it: the findings support the need for services like *Pedibus*. Indeed, traffic related safety concerns are often less significant than fear of dangerous strangers on route (see Mammen et al. (2012) and McDonald and Aalborg (2009)). Under the umbrella of SRTS several *Pedibus* related initiatives have been established in the U.S.A., mostly spontaneously and entirely based on parental support. These experiences led to a road map to set up a *Pedibus* service made of a list of common sense based recommendations (see Chillon et al. (2011)). In Italy where our study is conducted, a *Pedibus* service is active in at least 160 municipalities according to the official national web site *Pedibus-Italia* (2017) but these initiatives do not collect enough data to assess *Pedibus* effects on the medium term time horizon. Actually, the most challenging issue is not how to start a *Pedibus* service but how to keep it running steadily. According to Kingham and Ussher (2005) the main reason for service dismissal is the lack of new children once the former ones have become independent commuters. This fact has pros and cons: it confirms the educational function of *Pedibus*: living a *Pedibus* experience at youngest ages makes children confident with sustainable travel modes so that they reach school autonomously at an earlier age than peers. It also shows the drawbacks of initiatives relying exclusively on the goodwill of few volunteers: once children have grown up younger kids do not take the baton, their parents do not replace those in charge of the service, and both functions, i.e., escorting and lines management, remain pending until the service implodes. Another frequent reason for service dismissal when service management is not centralized but it is run by parents is the lack of supervising adults (so called *Pedibus line drivers, Plds*). All these considerations justify the need for a methodological study of the *Pedibus* lines planning problem, in order to provide school managers a decision support tool able to run the service in an integrated manner and serve the whole community exploiting economies of scale.

**Methodologies behind *Pedibus* lines planning** The methodological studies devoted to the development of an optimization based approach to *Pedibus* line design are much less frequent and can be differentiated according to the operational mode the specific service at hand conforms to. In all cases, the problem is mapped on an oriented graph whose nodes map the school (node 0), the children's homes and the (potential) lines stops (these two sets coincide in some cases), and whose arcs denote the shortest safe walking connections between two nodes. The solution to the line

design problem has a spanning tree (or Steiner tree) like structure since all routes converge to school and all children must be served. Several variants can be distinguished according to the following features, which obviously impact on the structure and complexity of the corresponding problem.

*Shepherding*: escorting is compulsory (yields a constraint on the number of tree leaves if *Plds* are limited) or not.

*Capacity*: when shepherding is needed, a maximum ratio of kids per adult along the line must be ensured vs the *uncapacitated* case where the number of adults per line is fixed.

*Stops-at-home*: line stops at children homes (lines yield a spanning tree) vs the variant in which stops must be located at some nodes of the graph while children walk alone from home to the closest stop (a location problem adds to the routing problem and the lines yield a Steiner tree shaped network).

*Duration*: maximum route duration with respect to the shortest path vs unlimited duration (affects lines number and number of stops per line).

*Limited drivers*: in case of required adult supervision, the primary objective is to minimize the number of *Plds*. Two cases are possible: in the *uncapacitated* case it consists of minimizing the number of tree leaves while in the *capacitated* case *Plds*' number is optimized if children are properly partitioned into lines.

*Routes risk and length*: the objective is to minimize risk and (or) duration.

Not any combination of the above features makes sense while some others lead to well known problems: for example, when shepherding from home is required but there is no capacity nor maximum route duration and the aim is to minimize the number of lines as primary objective and line duration as the second, then the problem boils down to the shortest Hamiltonian path ending at node 0. In this study we consider the most common version of the problem, according to which i) each line is managed by two adults, one leading the line and the second one at its rear, ii) children are picked up at home (full shepherding), iii) maximum walking distance of each child is constrained, iv) the primary objective is to minimize the number of lines (and so the number of supervisors), v) total risk of the arcs in the network is minimized as a secondary objective. Due to i), *Plds* are as many as twice the lines, whatever the number of children per line is. The secondary objective aims to design a compact network by favoring lines merging before destination since larger pedestrian groups are safer in traffic. This feature may also take into account potential upgrading of the pedestrian infrastructure along the lines. In the more general variant with capacity, the lines merging feature is even more important since it allows to make a better use of limited supervisors when demand at the nodes is not unitary. This is peculiar to *MLPLP* and it forbids the straightforward application of solution approaches developed in the field of Demand Responsive Transit Systems, where lines are intended for motorized vehicles and do not physically merge. We call this problem the Minimum Leaves Pedibus Line Design Problem, *MLPLP*. As line stops coincide with children homes, the network is made of a set of directed paths (one path coincides with one line) with constrained duration, rooted at node 0 and potentially sharing the last part of the journey. Since each node must be covered and has to be within a given distance from the root, in the field of combinatorial optimization *MLPLP* can be described as a constrained spanning tree rooted at the school node: the primary objective iv) minimizes the number of tree leaves while the secondary objective tends to aggregate the terminal part of the routes. Since this particular problem has never been addressed, the present modeling effort is justified. A literature search returned few papers addressing similar problems, as listed hereafter.

In network design problems stemming from telecommunication applications, Fernandes and Gouveia (1998) tackles the minimum spanning tree with a given number of leaves and presents two MILP models. In terms of Pedibus, it corresponds to the escorted *uncapacitated* case, say one adult for each line, with unlimited route duration and a given number of supervisors (lines/leaves). The objective function is a linear function of the arcs in the tree so it may represent the risk as in our secondary objective (see 1). Another related problem in telecommunication concerns the design of delay-constrained minimum spanning trees (Salama et al. (1997)). In Pedibus terms it corresponds to the minimum risk network (our secondary objective function) with equal maximum walking distance for each child.

The *capacitated* Pedibus variant with compulsory escorting and line stops location has been studied in Porro (2015); a heuristic algorithm already developed for the SBR problem has been adapted to the Pedibus case, not allowing the lines merging option which on the contrary can be quite helpful in case of supervisors shortages. The problem is decomposed into three steps and each one is heuristically solved: first, children are partitioned and clustered to yield a line for each cluster, then for each cluster the line stops are first located and then sequenced. Large instances can be solved but performance depends on the careful calibration of few parameters that are instance dependent. Concerning exact approaches to SBR, mixed integer linear programming (MILP) models are discussed in Riera-Ledesma and

Salazar-González (2013) among which a multi-commodity flow arc-based formulation. Nevertheless they are of little use for solving our problem as they do not contemplate lines merging: we rather exploit the special features of *MLPLP* to propose a new two-index arc flow based formulation. The third related study addresses the case of unaccompanied children. In Japan it is quite common for children living in urban environments to walk to school unescorted. There is evidence, however, that safety increases the more the journey is shared with peers. On this basis Tanaka et al. (2016) considers the distance of the subpath that each child walks alone as a measure of risk. Since individual shortest paths to school are likely to differ from child to child, sharing part of the itinerary leads to routes with increased distance. The objective function is thus a bi-criteria one, minimizing total risk and total walking distance. An integer programming formulation is proposed and solved via a state of the art solver. The Pareto optimal solutions show that safety can be largely increased for a slight raise in distance with respect to the shortest path solution (the default solution). None of these approaches can solve *MLPLP* since children are supervised (cfr. Tanaka et al. (2016)), line merging must be allowed (cfr. Porro (2015)), routes are constrained in distance (cfr. Fernandes and Gouveia (1998)) and we minimize the number of leaves (cfr. Salama et al. (1997)). In Section 3 we present an arc based MILP model which provides an optimal solution to *MLPLP* if solved to optimality or a heuristic one and a lower bound if the solver is stopped before convergence at time out.

### 3. An arc-based flow model for *MLPLP*

Let us introduce the mathematical notation required to formalize *MLPLP* as a MILP problem. The problem is described on an abstract network given by an oriented graph  $G = (N, A)$  where only children homes and the school are present (lines stops are set at children homes). In the flow formulation, each child is represented as a unit of flow coming out of a source node associated to the child home. Sets and parameters are summarized in Table 1. Four kinds of variables are needed (see Table 2): integer flow variables  $x_{ij}$  and binary design variables  $y_{ij}$  for each arc  $(i, j)$  in  $A$ ; for each node  $i \in N$ , the binary variable  $z_i$  represents the leaf status and the real variable  $\pi_i$  the duration of the itinerary starting from  $i$ . The MILP model is given in (1-10). The objective function (1) is made of two parts: minimization of the number of leaves and minimization of total risk. The coefficient  $\epsilon$  tunes the impact of the secondary objective such that  $(\epsilon \sum_{(i,j) \in \mathcal{A}} d_{ij} y_{ij}) < 1$  and the number of lines is the floor of the solution value.

Table 1: Sets and Input Parameters

|  |   |
|--|---|
| $N$  | set of nodes: 0 denotes the school while the other nodes represent children homes.  |
| $\mathcal{A}$  | set of arcs representing connections (shortest paths) between line stops.   |
| $c_{ij} > 0 \quad \forall (i, j) \in \mathcal{A}$    | denotes the length of the shortest safe walking connection from $i$ to $j$ , so that $c_{i0}$ is the length of the shortest path from $i$ to school.                                  |
| $d_{ij} \geq 0 \quad \forall (i, j) \in \mathcal{A}$ | is a risk measure, denoting how dangerous the path from $i$ to $j$ is   |
| $\delta > 1$   | is the maximum allowed ratio between the length of the itinerary from home to school along the walking bus line and the length of the shortest path from the same bus stop to school. |
| $\epsilon > 0$                                       | denotes the weight of the risk component in the objective function.   |
| $Q$  | a large enough value.   |

Table 2: Model Variables

|                           |                                  |  |
|---------------------------|----------------------------------|--|
| $x_{ij} \in \mathbb{Z}^+$ | $\forall (i, j) \in \mathcal{A}$ | integer variable representing the flow going from node $i$ to node $j$                       |
| $y_{ij} \in \{0, 1\}$     | $\forall (i, j) \in \mathcal{A}$ | binary variable equal to 1 if arc $(i, j)$ is used, 0 otherwise                              |
| $z_i \in \{0, 1\}$        | $\forall i \in N$                | binary variable equal to 1 if node $i$ is a leaf, 0 otherwise                                |
| $\pi_i > 0$               | $\forall i \in N$                | real variable associated to the length of the itinerary along the pedibus line from $i$ to 0 |

$$\min \sum_{i \in N} z_i + \epsilon \sum_{(i,j) \in \mathcal{A}} d_{ij} y_{ij} \tag{1}$$

$$\sum_{(j,i) \in \mathcal{A}} x_{ji} - \sum_{(i,j) \in \mathcal{A}} x_{ij} = -1 \quad \forall i \in N \setminus \{0\} \tag{2}$$

$$\sum_{(i,0) \in \mathcal{A}} x_{i0} = |N| - 1 \tag{3}$$

$$\sum_{(i,j) \in \mathcal{A}} y_{ij} = 1 \quad \forall i \in \mathcal{N} \setminus \{0\} \quad (4)$$

$$y_{ij} - x_{ij} \leq 0 \quad \forall (i, j) \in \mathcal{A} \quad (5)$$

$$x_{ij} - |\mathcal{N}|y_{ij} \leq 0 \quad \forall (i, j) \in \mathcal{A} \quad (6)$$

$$-z_i - \sum_{(j,i) \in \mathcal{A}} y_{ji} \leq -1 \quad \forall i \in \mathcal{N} \quad (7)$$

$$\sum_{(j,i) \in \mathcal{A}} y_{ji} + |\mathcal{N}|z_i \leq |\mathcal{N}| \quad \forall i \in \mathcal{N} \quad (8)$$

$$\pi_j + c_{ij}y_{ij} - M(1 - y_{ij}) - \pi_i \leq 0 \quad \forall (i, j) \in \mathcal{A} : i \neq 0 \quad (9)$$

$$\pi_i - \delta c_{i0} \leq 0 \quad \forall i \in \mathcal{N} \setminus \{0\} \quad (10)$$

Constraints (2) and (3) define the flow on  $G$ : since each children home is source of one flow unit while the school (node 0) is the only sink, for each node in  $\mathcal{N} \setminus \{0\}$  the flow going out is equal to the incoming flow plus one. Flow variables  $x_{ij}$  denote the number of nodes present in the subtree rooted in  $j$ , or, likewise, how many stops are present along the line(s) up to  $j$ . Constraint (4) imposes that for each node only one outgoing arc must be selected. indeed, if two walking bus lines arrive at the same node they are joined together and cannot be split again, thus sharing the same itinerary until school is reached. Inequalities (5) and (6) represent linking and activation constraints for variables  $x_{ij}$  and  $y_{ij}$ . Constraints (7) and (8) assign the leaf status ( $z_i = 1$ ) to each node  $i \in \mathcal{N}$  with no incoming flow. Constraints (9) define  $\pi_i$  as the distance from node  $i$  to school along the line node  $i$  belongs to. Finally, constraints (10) set a limit on the maximum traveled distance from home to school with respect to the shortest path length  $c_{i0}$ .

This formulation is quite compact and not hampered by symmetry. Moreover, the model can be extended to handle capacity. However, especially when the allowed walking distance is not binding, i.e., much higher than the shortest path, the linear relaxation may be quite weak which slows down convergence, as discussed in Sec.4.

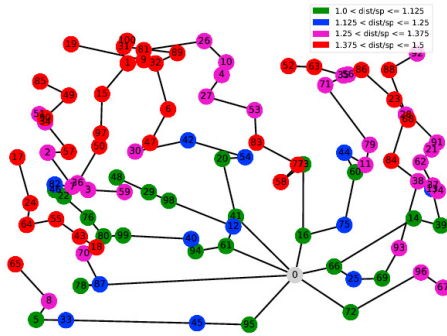
#### 4. Computational experiments

**Random instances** An extensive computational campaign has been carried out to test the model effectiveness in terms of its ability to solve real and random generated instances within a reasonable time limit (7200 seconds). The random data set contains 50 instances. Nodes range from 10 to 300 plus one extra node for the school. They are randomly located on the plane, uniformly distributed inside a squared area whose edge length is  $|\mathcal{N}|$ . Travel times are defined as euclidean distances between points. Risk coefficients are randomly generated using a uniform distribution function in the range  $[0.01, 0.99]$ . Ten different graphs have been generated, then the model has been solved for five different values of  $\delta$ , namely 1.1, 1.2, 1.5, 1.8, 20. The largest  $\delta$  values are just meant to stress the model (the problem becomes more difficult as  $\delta$  increases) as in practice a walking distance twice the minimum is not a viable option. Cplex (version 12.7.0) a state of the art commercial MILP solver is used with default settings on a cluster of virtual machines equipped with Intel Xeon E31220 3.4GHz (4 core/ 8 threads), 8 GB of ram, running Ubuntu 16.04 64 Bit. Results are shown in Table 4, whose columns report, from left to right: concerning the instance, number of nodes (school not included)  $|\mathcal{N}|$  and maximum route duration allowed with respect to the shortest path  $\delta$ ; from column 3 onwards, we report linear relaxation at the root node (LP), best integer solution at time out when solving the model from scratch (plain model, column M) and when starting from the incumbent (column H+M), percentage residual gap for the plain model (G% M) and for the empowered model (G% H+M); the gap is computed as the percentage ratio of the solution minus the best lower bound over the bound. Then, we compute the improvement of the model over the heuristic (column H vs H+M). In the last 4 columns we report results regarding the number of stops per line in the solution of the empowered model, i.e., the average (L avg), the maximum (L Max), and the minimum (L Min) number of stops per line and the standard deviation (L std). Linear relaxation running time is negligible (below 1 second) up to 50 nodes, below 3.2 seconds up to 100 nodes, and below 90 seconds for the biggest instances. This is related to the fact that it provides a weak lower bound. Indeed, what prevents the solution from being a single Hamiltonian path are distance constraints (9-10) which become loose when flow is fractional. Indeed, the linear solution value is always below 2 since flow tends to circulate as much as 9 allows. This confirms previous experience on the weakness of the linear relaxation of arc based MILP models for a distance constrained Vehicle Routing Problem (Aringhieri et al. (2017)) with inequalities such as 9. As a consequence, optimality can not be reached (or proved) for the 30 node

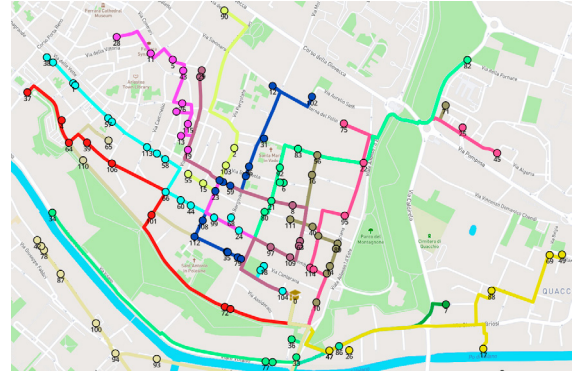
Table 3: Results on random instances

| $ N $ | $\delta$ | LP    | M      | H+M    | G% M   | G% H+M | H vs H+M | L avg | L Max | L Min | L std |
|-------|----------|-------|--------|--------|--------|--------|----------|-------|-------|-------|-------|
| 10    | 1.1      | 1.267 | 5.571  | 5.571  | 0.000  | 0.000  | 2.486    | 2.2   | 3     | 2     | 0.4   |
| 10    | 1.2      | 1.265 | 4.693  | 4.693  | 0.000  | 0.000  | 0.000    | 2.5   | 4     | 2     | 0.866 |
| 10    | 1.5      | 1.262 | 3.482  | 3.482  | 0.000  | 0.000  | 2.629    | 3.7   | 4     | 3     | 0.471 |
| 10    | 1.8      | 1.259 | 3.418  | 3.418  | 0.000  | 0.000  | 4.043    | 3.7   | 5     | 3     | 0.943 |
| 10    | 2.0      | 1.257 | 2.550  | 2.550  | 0.000  | 0.000  | 28.969   | 5     | 5     | 5     | 0     |
| 20    | 1.1      | 1.024 | 7.114  | 7.114  | 0.000  | 0.000  | 0.138    | 3     | 4     | 2     | 0.756 |
| 20    | 1.2      | 1.024 | 6.115  | 6.115  | 0.000  | 0.000  | 14.115   | 3.5   | 6     | 2     | 1.384 |
| 20    | 1.5      | 1.023 | 4.084  | 4.084  | 0.000  | 0.000  | 19.885   | 5     | 6     | 4     | 0.707 |
| 20    | 1.8      | 1.023 | 3.090  | 3.090  | 0.000  | 0.000  | 0.226    | 6.7   | 7     | 6     | 0.471 |
| 20    | 2.0      | 1.023 | 3.071  | 3.071  | 0.000  | 0.000  | 1.012    | 7     | 8     | 6     | 0.816 |
| 30    | 1.1      | 1.030 | 10.143 | 10.143 | 0.000  | 0.000  | 0.473    | 3.3   | 6     | 1     | 1.552 |
| 30    | 1.2      | 1.030 | 8.132  | 8.132  | 0.000  | 0.000  | 11.386   | 4.1   | 7     | 1     | 1.965 |
| 30    | 1.5      | 1.029 | 5.109  | 5.109  | 0.000  | 0.000  | 0.890    | 6.2   | 8     | 4     | 1.6   |
| 30    | 1.8      | 1.029 | 4.097  | 3.146  | 24.863 | 0.000  | 24.204   | 10    | 11    | 8     | 1.414 |
| 30    | 2.0      | 1.029 | 3.111  | 3.111  | 33.151 | 33.035 | 0.826    | 10    | 11    | 8     | 1.414 |
| 50    | 1.1      | 1.031 | 12.281 | 12.281 | 0.000  | 0.000  | 7.758    | 4.3   | 7     | 1     | 2.055 |
| 50    | 1.2      | 1.031 | 10.255 | 10.255 | 0.000  | 0.000  | 9.151    | 5.2   | 9     | 1     | 2.64  |
| 50    | 1.5      | 1.030 | 8.193  | 8.194  | 12.524 | 12.442 | 1.066    | 6.4   | 10    | 2     | 2.45  |
| 50    | 1.8      | 1.030 | 7.146  | 7.145  | 49.275 | 43.889 | 1.412    | 7.6   | 12    | 5     | 2.06  |
| 50    | 2.0      | 1.029 | 6.148  | 6.150  | 55.256 | 54.490 | 1.494    | 8.9   | 11    | 5     | 1.95  |
| 80    | 1.1      | 1.034 | 10.342 | 11.287 | 10.086 | 17.419 | 9.117    | 8     | 12    | 5     | 2     |
| 80    | 1.2      | 1.033 | 9.263  | 9.264  | 40.720 | 37.152 | 1.585    | 9.7   | 12    | 7     | 1.76  |
| 80    | 1.5      | 1.032 | 7.183  | 6.269  | 55.898 | 49.363 | 2.189    | 14.2  | 17    | 11    | 2.27  |
| 80    | 1.8      | 1.032 | 8.131  | 5.236  | 73.971 | 59.498 | 2.590    | 16.8  | 21    | 8     | 4.87  |
| 80    | 2.0      | 1.032 | 8.123  | 5.202  | 74.212 | 59.648 | 3.309    | 16.8  | 18    | 16    | 0.74  |
| 100   | 1.1      | 1.036 | 21.488 | 21.486 | 4.971  | 0.000  | 12.623   | 5.2   | 9     | 1     | 2.06  |
| 100   | 1.2      | 1.036 | 17.367 | 17.384 | 25.752 | 22.984 | 6.340    | 6.4   | 10    | 1     | 2.38  |
| 100   | 1.5      | 1.036 | 15.226 | 10.387 | 59.185 | 40.046 | 9.853    | 10.3  | 17    | 3     | 3.67  |
| 100   | 1.8      | 1.036 | 14.217 | 7.446  | 63.878 | 30.865 | 1.500    | 14.4  | 18    | 8     | 3.42  |
| 100   | 2.0      | 1.036 | 13.245 | 7.417  | 72.520 | 53.213 | 1.326    | 14.4  | 21    | 7     | 4.37  |
| 150   | 1.1      | 1.004 | 21.051 | 20.056 | 33.298 | 24.899 | 4.865    | 8.4   | 13    | 2     | 3.01  |
| 150   | 1.2      | 1.004 | 17.048 | 15.056 | 44.999 | 35.869 | 0.162    | 10.9  | 16    | 4     | 3.42  |
| 150   | 1.5      | 1.004 | 18.035 | 10.067 | 74.037 | 50.050 | 0.110    | 15.5  | 28    | 10    | 4.92  |
| 150   | 1.8      | 1.004 | 21.027 | 8.070  | 85.683 | 62.571 | 0.071    | 19.2  | 30    | 9     | 6.24  |
| 150   | 2.0      | 1.004 | 28.029 | 7.076  | 92.833 | 84.699 | 0.011    | 22.4  | 37    | 13    | 7.74  |
| 200   | 1.1      | 1.004 | 33.080 | 33.081 | 12.109 | 12.077 | 13.216   | 6.8   | 12    | 3     | 2.08  |
| 200   | 1.2      | 1.004 | 28.064 | 27.067 | 34.348 | 31.756 | 7.034    | 8.2   | 11    | 4     | 1.83  |
| 200   | 1.5      | 1.004 | 28.049 | 19.070 | 71.356 | 57.825 | 0.189    | 11.5  | 17    | 6     | 2.87  |
| 200   | 1.8      | 1.004 | 37.038 | 14.100 | 91.835 | 78.503 | 0.033    | 14.5  | 30    | 7     | 6.62  |
| 200   | 2.0      | 1.004 | 37.040 | 13.095 | 91.828 | 76.923 | 0.044    | 16    | 32    | 6     | 6.43  |
| 250   | 1.1      | 1.005 | 39.088 | 38.093 | 25.584 | 23.550 | 11.699   | 7.58  | 15    | 3     | 2.77  |
| 250   | 1.2      | 1.005 | 35.068 | 31.080 | 46.425 | 38.607 | 3.270    | 9.29  | 15    | 5     | 3.02  |
| 250   | 1.5      | 1.005 | 31.065 | 20.114 | 79.716 | 71.579 | 0.080    | 13.1  | 30    | 6     | 6.09  |
| 250   | 1.8      | 1.005 | 48.062 | 16.119 | 93.784 | 81.193 | 0.031    | 16.1  | 38    | 4     | 8.18  |
| 250   | 2.0      | 1.005 | 48.061 | 13.120 | 93.694 | 76.960 | 0.040    | 19.7  | 30    | 9     | 6.13  |
| 300   | 1.1      | 1.005 | 43.096 | 41.096 | 27.862 | 24.310 | 6.948    | 8.3   | 12    | 2     | 2.66  |
| 300   | 1.2      | 1.005 | 41.075 | 34.107 | 58.416 | 49.882 | 2.984    | 10.1  | 17    | 4     | 3.4   |
| 300   | 1.5      | 1.005 | 53.075 | 21.148 | 87.355 | 70.690 | 0.011    | 14.5  | 29    | 4     | 6.13  |
| 300   | 1.8      | 1.005 | 61.064 | 14.149 | 94.841 | 77.773 | 0.008    | 21.8  | 36    | 8     | 7.11  |
| 300   | 2.0      | 1.005 | 82.068 | 13.145 | 97.581 | 84.602 | 0.020    | 23.2  | 35    | 11    | 7.49  |

instances with  $\delta \geq 1.8$ , for the 50 node ones with  $\delta \geq 1.5$ , and from 80 nodes onwards when the model is solved from scratch (when residual gap is positive timeout is reached). The number of lines (integer part of the solution value) decreases as  $\delta$  increases as more nodes can fit into the same line, and when this monotonic trend is not followed it is just due to the poor solution quality at time out. To increase convergence rate despite the model weakness, a heuristic solution is provided as the starting incumbent, allowing the solver to prune a substantial part of the search tree. We refer to this option as to the *empowered model*. The heuristic is based on a greedy procedure that builds routes starting from node 0 according to the nearest neighbor principle; at each iteration the node nearest to the current one



(a) The solution for the 100-1.5 case, depicted on the plane

(b) The solution for the real case of Ferrara with  $\delta = 1.2$ 

is appended at the end of the route, if feasible, and it becomes the new current node; when no nodes can be added at the end of the current route, the route is closed and a new route is started from 0. Then, local improvements are sought by moving terminal nodes between routes. The process is iterated perturbing the first node assignment. Total running time is negligible - much quicker than the linear relaxation time. The empowered model (column H+M) solves to proven optimality only one additional instance but in the most challenging cases ( $\delta \geq 1.8$  and  $|N| \geq 150$ ) solution quality substantially improved with respect to the plain model (column M) which, starting from scratch, sometimes it is not even able to reach the solution quality it achieves on the same graph for lower  $\delta$ . The heuristic does not dominate the plain model; obviously the empowered model always improves over the heuristic and it often saves lines. The percentage ratio of such improvement over the heuristic value is reported in 8<sup>th</sup> column (H vs H+M).

The impact of  $\delta$  on problem hardness is evident: the larger  $\delta$  is, the higher the number of stops in a single line is, the larger the feasible region is, the more fractional the linear solution is, and the looser the lower bound becomes so that the solver spends most of the time carrying out a complete search without much pruning until time out. This is confirmed by looking at the number of line stops of the empowered MILP solution: on average they increase three times as  $\delta$  grows from 1.1 to 2, and for the same  $\delta$  they increase with  $|N|$  almost linearly: note that when  $Lavg \cdot [LP] > |N|$  at least two lines merged before reaching destination (nodes visited after merging are counted as belonging to both lines). While we lack a reliable measure of solution quality, the graphical representation allows evaluations at a glance. Figure 1a depicts the solution for the 110 nodes case with  $\delta = 1.5$ . The colors help appreciate the deterioration in route length affecting individual nodes. The lines are rather similar regarding the number of line stops, which is a good feature. Note that some lines merge and share the last part of the route.

**Real data** We tested the empowered model on a real instance for the city of Ferrara, Italy. Children data concerning the students of the primary school Biagio Rossetti living within 1.5 km (walking distance) from school yield 116 nodes (17 of the 133 children share their home address with at least another child). Distances are real walking distances. The extra walking distance ratio  $\sigma$  is at most 1.3, which may lead a child to walk almost 2 km. Risk has been assumed to be proportional to distance, but it may be set based on any indicator affecting the quality of the pedestrian environment, such as: average traffic, speed limits, sidewalk width, traffic signs, roundabout and pedestrian crossing features. Table 4 reports, from left to right,  $\delta$ , the linear relaxation value (column LP), the heuristic solution value (column H), the enhanced model solution (column H+M, \* if optimality is reached), the residual percentage gap at timeout (G%H+M), the heuristic running time and the solver running time (timeout reached for positive gaps), the percentage gap improvement the model achieved on the incumbent heuristic solution (column H vs H+M) which is computed as the percentage ratio of the heuristic solution minus the solution of the empowered model over the heuristic solution value, the number of lines (tree leaves); the last four columns concern the number of stops per line and report the average, maximum, minimum value and the standard deviation, in this order, as in Table 3. As observed on random data, the number of lines decreases as  $\delta$  increases and residual gap can hardly be taken as a reliable indicator of solution quality due to the loose linear relaxation. Fig. 1b depicts the solution for the real case with  $\delta = 1.2$  on the road map (on shared road segments only one color is visible).

Table 4: Results on real instance

| $\delta$ | LP    | H      | H+M     | G% H+M | T(H)  | T(H+M)   | H vs H+M | # L | L avg | L Max | L Min | L std |
|----------|-------|--------|---------|--------|-------|----------|----------|-----|-------|-------|-------|-------|
| 1.1      | 1.056 | 27.062 | 25.060* | 0.000  | 1.298 | 14.250   | 7.39783  | 25  | 5.28  | 13    | 1     | 2.807 |
| 1.2      | 1.056 | 21.062 | 19.061* | 0.000  | 0.763 | 870.337  | 9.50052  | 19  | 6.421 | 15    | 2     | 3.529 |
| 1.3      | 1.056 | 17.062 | 16.059  | 24.906 | 0.817 | 7204.283 | 5.87856  | 16  | 8.375 | 17    | 2     | 4.106 |

## 5. Conclusions and work in progress

Abundant literature on Pedibus applications testifies its importance, however the optimization problem underlying the lines design process has been scarcely addressed in the scientific literature. The lack of decision support tools leaves the burden of service organization on volunteers who rely on common sense based practices and hardly deal with scarce *Plds* and (or) tend to increase walking time more than necessary. Indeed, a challenging combinatorial optimization problem lies behind Pedibus lines planning even for the simplest variant tackled in this paper, where children are escorted along the whole itinerary from home to school and route duration is bounded. The objective is to minimize the number of lines (and thus the number of *Plds*) and as a secondary objective the total risk of the selected arcs. Since lines merging are allowed, this problem differs from the School Bus Routing problem; it can be rather seen as a new variant of the Minimum Spanning Tree with Constrained Number of Leaves. We present a new arc based MILP model and test it on random and realistic data. While it is effective for small instances, when the size approaches the scale of a realistic elementary school population the exact solution is out of reach within a 2 hours time out, which is a fair time limit for a design problem. A feasible solution is always returned, however its quality is hard to assess due to the lack of tight lower bounds. This lack affects solution quality the most when the constraint on route duration is loose, i.e. up to twice the shortest path duration, which is quite unlikely in practice. In such cases, substantial improvement is achieved by providing the solver with a heuristic starting solution - whose running time is negligible - that allows the solver to prune much earlier suboptimal solutions, speeds up the search process and achieves better solutions in the allotted time. On realistic data, a tighter threshold on walking time is used and good solutions are obtained (optimality is reached on two instances over three). Further studies are ongoing to improve on two aspects: path based formulations are being developed to reach and certify global optimality on larger instances exploiting stronger lower bounds; besides, the capacitated variant is tackled by generalizing the current uncapacitated model.

## References

- Chilln, P., Evenson, K.R., Vaughn, A., Ward, D.S., 2011. A systematic review of interventions for promoting active transportation to school. *International Journal of Behavioral Nutrition and Physical Activity*, 8:10.
- D. Engwicht, 1992. Is the Walking Bus stalled?
- Fernandes, L.M., Gouveia, L., 1998. Minimal spanning trees with a constraint on the number of leaves. *Eur. J. Oper. Res.* 104, 250–261.
- Aringhieri, R., Bruglieri, M., Malucelli, F., Nonato, M. 2017. A Special Vehicle Routing Problem Arising in the Optimization of Waste Disposal: A Real Case. *Transport. Sci.* <https://doi.org/10.1287/trsc.2016.0731>.
- Kingham, S., Ussher, S., 2005. Ticket to a sustainable future: An evaluation of the long-term durability of the Walking School Bus program in Christchurch, New Zealand. *Transport Policy* 12(4), 314–323.
- Riera-Ledesma, J., Salazar-González, J.J., 2013. A column generation approach for a school bus routing problem with resource constraints. *Computers & Operations Research* 40, 566–83.
- Mammen, G., Faulkner, G., Buliuna, R., Lay, J., 2012. Understanding the drive to escort: a cross-sectional analysis examining parental attitudes towards childrens school travel and independent mobility. *BMC Public Health* 12, 826.
- N.C. McDonald and A. E. Aalborg, 2009. Why Parents Drive Children to School. *Journal of the American Planning Association* 75(3), 331–342.
- McDonald, N.C., 2007. Travel and the social environment: Evidence from Alameda County, California. *Transp Res D* 12(1), 53–63. <http://www.pedibus-italia.it/>. Last accessed on March 2017.
- Porro, M., 2015. Sviluppo di un algoritmo di ottimizzazione dei percorsi in un servizio Pedibus. <http://hdl.handle.net/10589/117289>
- Salama, H.F., Reeves, D.S., Viniotis, Y., 1997. The delay-constrained minimum spanning tree problem. *ISCC* 97, 699–703. [http://guide.saferoutesinfo.org/walking\\_school\\_bus](http://guide.saferoutesinfo.org/walking_school_bus). Accessed on line March 2017.
- Tanaka, K., Miyashiro, R., Miyamoto, Y., 2016. A Bi-objective Optimization Model for Designing Safe Walking Routes for School Children. *Geographical Analysis*, doi: 10.1111/gean.12095