A New Sequential Approach to Periodic Vehicle Scheduling and Timetabling

Paul Bouman
Erasmus School of Economics, Erasmus University Rotterdam, The Netherlands
bouman@ese.eur.nl

Alexander Schiewe
TU Kaiserslautern, Germany
a.schiewe@mathematik.uni-kl.de

Philine Schiewe
TU Kaiserslautern, Germany
p.schiewe@mathematik.uni-kl.de

Abstract

When evaluating the operational costs of a public transport system, the most important factor is the number of vehicles needed for operation. In contrast to the canonical sequential approach of first fixing a timetable and then adding a vehicle schedule, we consider a sequential approach where a vehicle schedule is determined for a given line plan and only afterwards a timetable is fixed. We compare this new sequential approach to a model that integrates both steps. To represent various operational requirements, we consider multiple possibilities to restrict the vehicle circulations to be short, as this can provide operational benefits. The sequential approach can efficiently determine public transport plans with a low number of vehicles. This is evaluated theoretically and empirically demonstrated for two close-to real-world instances.

2012 ACM Subject Classification Mathematics of computing → Discrete mathematics; Applied computing → Transportation

Keywords and phrases Vehicle Scheduling, Timetabling, Integrated Planning

Digital Object Identifier 10.4230/OASIcs.ATMOS.2020.6

Funding This work was partially funded by DFG under SCHO 1140/8-2.

1 Introduction

In public transport planning, the problem of designing a public transport plan is traditionally separated into multiple sequential problems, [7, 10]. Commonly, one of the first steps is to obtain a line plan, for an overview see [25]. The lines in such a plan are a sequence of stops at which a vehicle picks up and drops off passengers. The lines are operated at a certain frequency. Designing a good line plan is a challenging problem where a trade-off between service quality and operational costs has to be made. The service quality is mostly determined by the number of transfers that passengers need to reach their destination, and the time their journey takes. There are numerous works focusing either on optimizing service quality or operational costs, e.g. [1, 4, 6, 27].

The travel times of the passengers and the vehicles are dependent on the timetable, which dictates at which time each line is operated. Typically, this occurs in a periodic fashion where for example the same line departs at the same time every hour. An overview on timetabling can be found in [14], some models and solution approaches on periodic timetabling include [8, 13, 18, 20].

The number of vehicles needed to operate a line plan under a certain timetable is an important factor in the operational costs. Determining which vehicles operate which lines in which order is called vehicle scheduling, for an overview, see [5]. Especially aperiodic

© Paul Bouman, Alexander Schiewe, and Philine Schiewe;
licensed under Creative Commons License CC-BY
20th Symposium on Algorithmic Approaches for Transportation Modelling, Optimization, and Systems (ATMOS 2020).
Editors: Dennis Huisman and Christos D. Zaroliagis; Article No. 6; pp. 6:1–6:16
OpenAccess Series in Informatics
OASICS Schloss Dagstuhl – Leibniz-Zentrum für Informatik, Dagstuhl Publishing, Germany

vehicle scheduling is well researched, see e.g. [15, 21]. Recently, periodic vehicle scheduling received more attention. In the case considered in this paper, [3] showed that periodic vehicle scheduling is a viable alternative.

When we have a periodic timetable, a periodic vehicle schedule consists of vehicle circulations that are operated by a number of vehicles. While it is possible to work with long vehicle circulations, this results in strong sequential dependencies of the activities that must be performed on different lines. Constraints on the length of the circulations, on the other hand, make the vehicle scheduling problem more challenging and restrict the solution space. In this paper, we investigate the impact of such constraints on which circulations are allowed when a vehicle schedule is determined.

While traditionally the steps of line planning, timetabling and vehicle scheduling are performed sequentially in that canonical order, integrating multiple planning stages has proven to be promising, see [2, 11, 24]. Due to the increased intricacy of the integrated problems, there exist various heuristic approaches that incorporate some form of integration, e.g. [12, 16, 19]. A general scheme for deriving heuristic solution approaches is the so-called eigenmodel, see [26], where the single stages line planning, timetabling and vehicle scheduling are re-ordered. First approaches on the reordering proposed in this paper were done in [19], where a simple form of aperiodic vehicle scheduling is considered. In this paper, we assume a line plan has been constructed and consider both an integrated method that jointly optimizes a periodic timetable and vehicle schedule, as well as a sequential method that first computes a vehicle schedule and determines a timetable based on that. The practical application we focus on is long-term strategic planning which typically occurs when future demand is highly uncertain. As such our main objective is vehicle scheduling and the minimization of the number of vehicles needed rather than the optimization of passenger convenience which is a common concern in the tactical and operational planning phases of public transport planning. For an overview on the different planning phases, see [10].

In [31], the problem of finding a vehicle schedule based on a line plan is analyzed. The concept of strict circulations is introduced, where a line is always covered by a single circulation. In this paper, we mainly consider strict circulations and investigate additional circulation restrictions as well as the effect of adding a timetable. An integrated model for periodic vehicle scheduling and timetabling is presented by [30] but without additional restrictions on the circulations. We use the model proposed in [30] as a basis for the integrated formulation in Section 5 and show how the sequential process developed in Sections 3 and 4 can already find optimal solutions to the integrated problem while reducing the problem size. All models are implemented and computationally evaluated using the open source framework LinTim, see [22, 23], in Section 6.

2 Problem Definition

In this section we formally introduce the problems considered in this paper. All these problems take a line plan as input.

▶ **Definition 1.** *A* line plan $\mathcal{L}$ *contains a set of lines* $l \in \mathcal{L}$*, which are paths in the infrastructure network PTN=(V,E) with stations* V *and direct connections* E *between them. For each line, there is a* forward trip l^+ *and a* backward trip l^-*. The* trip time t_{l^+}, t_{l^-} *is the minimal time needed to make a trip in one direction of the line.* Frequency f_l *indicates how often line* l *should be serviced per period, whose length is denoted by* T*.*

With these lines, we define a *trip graph* where stations V form the nodes and lines $\mathcal{L}$ the edges. We consider both the directed and the undirected case. In the undirected trip graph $L = (V, E(\mathcal{L}))$ each edge $e(l) = \{u, v\}$, $l \in \mathcal{L}$, is the pair of terminal stations for line l. The directions of a line form a directed trip graph $\overleftrightarrow{L} = (V, A(\mathcal{L}))$, with arcs $a(l^+) = (u, v)$ and $a(l^-) = (v, u)$, $l \in \mathcal{L}$. We now use the trip graph to construct a periodic vehicle schedule and a periodic timetable, to determine the number of vehicles needed to operate the lines.

In the vehicle scheduling problem, we consider circulations which are cycles in the directed trip graph $\overleftrightarrow{L} = (V, A(\mathcal{L}))$. The time it takes to operate all lines in the circulation c is defined as $t_c = \sum_{a(l^\cdot)\in c} t_l$. For shorter notation, we here without loss of generality assume that $t_{l^+} = t_{l^-} = t_l$. The minimal number of vehicles needed to operate a circulation in every period is given by $k_c = \left\lceil \frac{t_c}{T} \right\rceil$, which can alternatively be interpreted as the least number of periods a single vehicle spends on a circulation.

▶ **Definition 2.** *Let a line line $\mathcal{L}$ and a set of possible circulations $\mathcal{C}$ be given. A* feasible periodic vehicle schedule *is a subset $C' \subseteq \mathcal{C}$ such that every line $l \in \mathcal{L}$ is covered f_l times in both directions or equivalently where every arc in the directed trip graph $\overleftrightarrow{L}$ is covered exactly f_l times.*

In reality, interdependence of lines are imposed, e.g. by security constraints in the form of headways. These cannot be respected without knowing the actual departure and arrival times of the lines. Hence we need to add a periodic timetable to obtain the correct number of vehicles needed to operate the vehicle schedule.

▶ **Definition 3.** *For a given line plan $\mathcal{L}$ and a set of circulations $\mathcal{C}$, an event-activity-network (EAN) is a directed graph containing the departure and arrival of all lines $l \in \mathcal{L}$ at their respective stops as vertices (events) and arcs (activities) stating the interdependencies between these events. These can contain drive activities, wait activities, circulation activities and headway activities. A feasible periodic timetable assigns a periodic time $\pi_i \in \{0, \dots, T-1\}$ for every event i, such that for all activities the duration is in given time bounds.*

While drive and wait activities are directly related to the given line plan $\mathcal{L}$, headway activities represent restrictions of the infrastructure network such as safety restrictions on tracks. Circulation activities model the turnaround time of the vehicles between trips and are therefore given by the chosen circulations.

Note that opposed to most literature on periodic timetabling, the EAN described here does not contain transfer activities. Due to the periodicity of the timetable, transferring between lines can be assumed to always be feasible and we do not consider passenger convenience here.

The problem we want to solve overall is the following.

▶ **Definition 4.** *Let a line plan $\mathcal{L}$ with frequencies $f_l, l \in \mathcal{L}$, and a set of possible circulations $\mathcal{C}$ be given.* (LinToTimVeh) *is the problem of finding a feasible periodic vehicle schedule and a corresponding feasible periodic timetable such that the number of vehicles needed to operate the line plan is minimal.*

Different solution methods for (LinToTimVeh) are presented in Figure 1. While it is possible to solve the problem integratedly, we also consider a sequential solution approach. In contrast to the standard sequential planning process presented in [7, 10], we change the order of the optimization problems as suggested in [26]. For a given line plan, we therefore first fix a vehicle schedule by determining periodic circulations in (LinToVeh) that minimize the lower bound of vehicles needed to operate the chosen circulations while covering every trip. For these circulations, a periodic timetable is determined in (LinVehToTim). As we want to minimize the number of vehicles needed to operate the circulations, we cannot use a standard PESP model from literature, see [24].

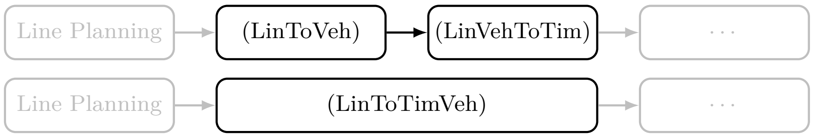


Figure 1 Overview of the presented problems. The naming scheme of the problems is given by the following notation: Different sequential planning stages are divided in their in- and output by “To”, “Lin” refers to line planning, “Tim” refers to timetabling, “Veh” refers to vehicle scheduling and “TimVeh” refers to the integrated timetabling and vehicle scheduling problem.

For both the sequential and the integrated approach we consider different sets of possible circulations as discussed in Section 2.1. We especially differentiate between *general* and *linked* circulations where linked circulations contain both directions of each covered line.

2.1 Circulations

Since the set of possible circulations available for the vehicle scheduling problem described in Definition 2 is crucial for the obtained number of vehicles needed, we first describe our assumptions for those sets. We assume to have a symmetric directed trip graph $\overleftrightarrow{L}$, so an Eulerian cycle exists for each connected component. This provides a solution that minimizes the number of vehicles, since the gap of the $\lceil\cdot\rceil$-operator in k_c is minimized.

However, there are practical reasons to look for solutions that involve circulations with fewer trips. It is unlikely that a good timetable can be constructed, as the Eulerian cycle imposes strong interdependence on the arrival and departure times of all the lines. Furthermore, delays and disruptions can propagate through the vehicle schedule. The Eulerian cycle based solution would make all trips dependent on all other trips, which is bad from a robustness perspective. Thus, if the same number of vehicles can be achieved with a solution that has multiple shorter circulations, this is preferable.

In order to find a set of shorter circulations, we can impose restrictions on the type of circulations that are allowed in our solution. We refer to a circulation c as an (α, β) circulation if the number of trips in c is α and the number of unique lines covered by c is β. If additionally a circulation c contains both directions of each line, i.e., if $\forall l \in \mathcal{L}$ it holds that $l^+ \in c$ iff $l^- \in c$, we call c a *linked* circulation. We will refer to a linked circulation c as a β circulation if exactly β lines are covered by it, and thus it must contain 2β trips. Therefore, a β circulation is also a $(2\beta, \beta)$ circulation.

In order to express a limit on the number of trips and lines in a circulation, we refer to $\leq \beta$, $(\alpha, \leq \beta)$ and $(\leq \alpha, \beta)$ circulations as a circulation that have no more than β lines, or α trips. If we only want to impose a limit on the number of trips or lines used, we use the notation $(\leq \alpha, \bullet)$ or $(\bullet, \leq \beta)$, respectively.

In Figure 2 we present an example where we get a better solution when $(\leq 6, \leq 4)$ circulations are allowed compared to the situation where ≤ 4 linked circulations are allowed. The main insight is that in the non-linked case, we can sometimes avoid downtime by assigning the forward direction of a line to one circulation, while the other direction is assigned to a different circulation.

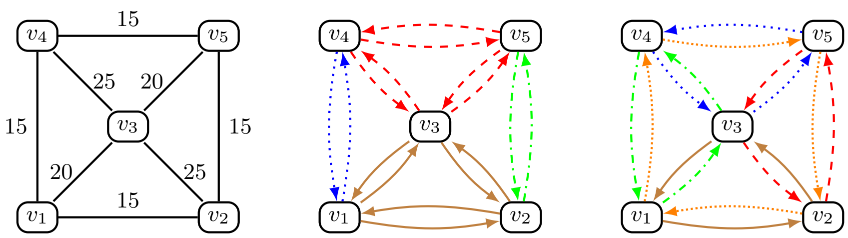


Figure 2 Example for a disadvantage when using linked circulations. The period is 60 and at most 6 trips and 4 lines can be used in a single circulation. The different circulations are marked by color and line style. The trip length of the lines is such that we can do better when we use general circulations (five vehicles, on the right hand side) than when we use linked circulations (six vehicles, the middle).

3 Vehicle Scheduling Based on a Line Plan

We now introduce a model for (LinToVeh), as defined in Definition 2. Using the notation from Section 2, we can model the problem using binary variables z_c, indicating whether a circulation c is chosen.

$$\text{(LinToVeh)} \quad \min \sum_{c\in\mathcal{C}} k_c z_c$$
$$\sum_{c\in\mathcal{C}:l\in c} z_c = f_l \qquad l \in \mathcal{L}$$
$$z_c \in \{0,1\} \qquad c \in \mathcal{C}$$

As mentioned already in the introduction, allowing larger circulations can result in a lower minimal number of vehicles needed. However, Example 8 in Section 4 shows that adding a timetable for these larger circulations might lead to actually needing more vehicles.

▶ **Lemma 5.** *For increasing k, the minimal number of vehicles computed by (LinToVeh) decreases monotonically for linked $\leq k$ circulations as well as general $(\leq k, \bullet)$ and $(\bullet, \leq k)$ circulations.*

Proof. The statement follows directly from the fact that the solution space for k is contained in the solution space for $k+1$. ◀

3.1 Comparing Linked to General Circulations

When comparing the linked and the general case, we get that we may need more vehicles in the linked case when both solutions may contain at most 2β trips.

▶ **Theorem 6.** *Let $\beta \in \mathbb{N}$, $\beta \geq 2$ be given. Denote $I = (L, t, f, T)$ an instance of (LinToVeh) with $\mathcal{C}_l$ an optimal solution in the linked case, i.e., the circulations $c \in \mathcal{C}_l$ are $\leq \beta$ circulations and $\mathcal{C}_u$ an optimal solution in the general case, i.e., the circulations $c \in \mathcal{C}_u$ are $(\leq 2\beta, \bullet)$ circulations. Then we get*

$$\max_I \frac{\sum_{c\in\mathcal{C}_l} k_c}{\sum_{c\in\mathcal{C}_u} k_c} \geq \frac{3}{2}.$$

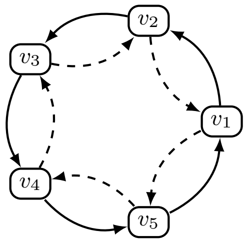


Figure 3 Directed trip graph for $K = 5$. The solid arcs represent the lines in forward direction while the dashed arcs represent the lines in backward direction.

Proof. Consider the following instance $I = (L, t, f, T)$ of the (LinToVeh) problem where $K = \beta + 1$ if β is even and $K = \beta + 2$ if β is odd. Let $V = \{v_1, \dots, v_K\}$ and $\mathcal{L} = \{l_1, \dots, l_K\}$ with directed trip graph $\overleftrightarrow{L} = (V, A(\mathcal{L}))$ and

$$\begin{aligned} a(l_i^+) &= (v_i, v_{i+1}), \qquad i \in \{1, \dots, K-1\}, \qquad a(l_K^+) = (v_K, v_1), \\ a(l_i^-) &= (v_{i+1}, v_i), \qquad i \in \{1, \dots, K-1\}, \qquad a(l_K^-) = (v_1, v_K). \end{aligned}$$

The directed trip graph is depicted in Figure 3. Furthermore, set $f_l = 1$ for all lines $l \in \mathcal{L}$ and $t_l = \frac{T}{K}$ for all trips for lines $l \in \mathcal{L}$.

For the general case, an optimal solution consists of two (K, K) circulations, $c_1 = (l_1^+, \dots, l_K^+)$ and $c_2 = (l_1^-, \dots, l_K^-)$ with $t_{c_1} = t_{c_2} = T$ and thus $k_{c_1} = k_{c_2} = 1$ such that two vehicles are needed.

However, with K odd, we get for any $(2k, \bullet)$ circulation c with $k \le \beta < K$ and thus especially for linked k circulations,

$$t_c = \sum_{l \in c} t_l = \frac{|c|}{K} \cdot T = \frac{2k}{K} \cdot T \neq n \cdot T, \text{ for any } n \in \mathbb{N}$$

and therefore $k_c = \lceil \frac{t_c}{T} \rceil > \frac{t_c}{T}$. For a set $\mathcal{C}$ of $(2k, \bullet)$ circulations with $k \le \beta$ covering all lines in $\mathcal{L}$ we therefore get

$$\sum_{c \in \mathcal{C}} k_c = \sum_{c \in \mathcal{C}} \left\lceil \frac{t_c}{T} \right\rceil > \sum_{c \in \mathcal{C}} \frac{t_c}{T} = 2.$$

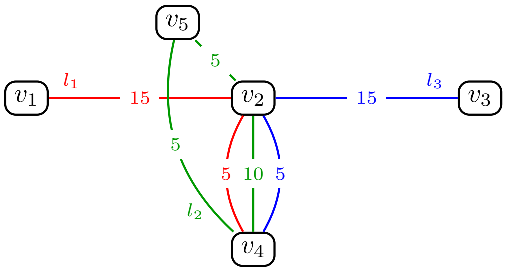
With $k_c \in \mathbb{N}$, we get $\sum_{c \in \mathcal{C}} k_c \ge 3$ and thus

$$\max_I \frac{\sum_{c \in \mathcal{C}_l} k_c}{\sum_{c \in \mathcal{C}_u} k_c} \ge \frac{3}{2}. \qquad \blacktriangleleft$$

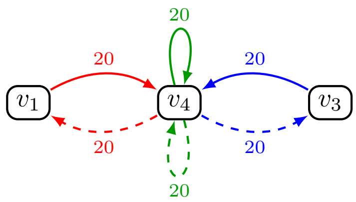
There are also special cases where solutions for linked and general circulations coincide.

▶ **Lemma 7.** *Let $I = (L, t, f, T)$ be an instance of (LinToVeh). If there is no cycle in L with length smaller or equal to β, then any general $(\le 2\beta, \bullet)$ circulation is linked. This is especially true when L is a tree.*

Proof. As there is no cycle of length smaller or equal to β in L, each $(\le 2\beta, \bullet)$ circulation c in $\overleftrightarrow{L}$ containing a trip of line l^+ also contains a trip of its backwards line l^- and vice versa. Therefore, only linked $\le \beta$ circulations can be found in $\overleftrightarrow{L}$ such that the linked and the general case coincide. ◀



(a) Infrastructure network with line plan. Each line is operated with frequency one. The minimal duration of the edges is given on the edges.



(b) Directed trip graph where forward trips are depicted solid and backward trips dashed. The minimal duration of the trips is given on the arcs.

Figure 4 Example for impairment of (LinVehToTim).

4 Adding a Periodic Timetable

Computing the number of vehicles needed to operate the line plan in Section 3 is only an approximation for the actual number of vehicles needed in the complete public transport system, since the timetable has an important effect on this property as well.

For modeling (LinVehToTim), we consider an event-activity network containing circulation activities as described in Definition 3 resulting in periodic event scheduling constraints as introduced in [28]. However, the number of vehicles needed to operate a given circulation cannot be expressed using the standard objective of PESP, see, e.g. [24]. We therefore use the integrated model presented in Section 5 with fixed circulation variables to find a timetable respecting the given circulations and minimizing the number of vehicles needed.

The number of vehicles needed when adding a feasible timetable is always at least as high as the one computed by (LinToVeh), as the line trip times t_l are lower bounds on the actual trip times respecting headways. Note that this number can even increase when the vehicle number determined by (LinToVeh) decreases, as shown in the following example.

▶ **Example 8.** Consider the instance of (LinVehToTim) given in Figure 4 with three lines l_1, l_2, l_3 and period length $T = 60$.

As the minimal duration for each line is 20, the optimal solution $\mathcal{C} = \{c_1, c_2, c_3\}$ for $(\leq 2, \bullet)$ circulations consists of 3 circulations. Each circulation c_i contains both directions of line l_i such that three vehicles are needed.

For $(\leq 3, \bullet)$ circulations, an optimal solution $\mathcal{C}' = c'_1, c'_2$ is given by $c'_1 = (l_1^+, l_2^+, l_1^-)$, $c'_2 = (l_3^+, l_2^-, l_3^-)$ such that only two vehicles are required.

For (LinVehToTim), we consider the case without wait times such that for each station in each line it suffices to determine a departure time. We impose headway constraints at station v_2 such that departures at this station have to be scheduled at least ten time units apart. As station v_2 is part of all six trips, there is a departure at station v_2 every ten time units.

For circulation set $\mathcal{C}$ there is a timetable resulting in 3 vehicles needed by extending the duration of each drive activity to 15 time units and starting the circulations scheduled 10 time units apart. The corresponding departure times can be found in Table 1.

However, for circulation set $\mathcal{C}'$, a feasible timetable results in needing at least four vehicles as operating a circulation c'_i by one vehicle leads to infeasibility: If circulations c'_1 is to be operated by one vehicle, each edge has to be operated with the minimum duration. This leads to three departures of station v_2 scheduled at $(\tau, \tau + 15, \tau + 30) \mod 60$. Due to the headway constraints, this leaves a time window of ten time units in which for circulation c'_2 three departures at station v_2 have to be scheduled which is infeasible. For c'_2, we can use an analogue argument.

The timetable constructed for circulation set $\mathcal{C}$ can also be operated for circulation set $\mathcal{C}'$ but here four vehicles are needed.

Table 1 Periodic departure times for Example 8. Departure times at stations v_2 are marked bold. Note that the timetable for c_1' cannot be extended for c_2' such that the headway constraints are satisfied.

line	l_1^+	l_1^+	l_1^-	l_1^-	l_2^+	l_2^+	l_2^-	l_2^-	l_3^+	l_3^+	l_3^-	l_3^-
station	v_1	v_2	v_4	v_2	v_4	v_2	v_4	v_2	v_3	v_2	v_4	v_2
c_1	0	**15**	30	**45**								
c_2					10	**25**	40	**55**				
c_3									20	**35**	50	**5**
c_1'	0	**15**	40	**45**	20	**30**						

For a fixed set of circulation computed by (LinToVeh), we investigate a worst case bound on the approximation error.

▶ **Theorem 9.** *When considering infrastructure headways and strictly positive, integer minimal activity durations, the optimal objective value of (LinVehToTim) is at most $\frac{T}{2}$ times the number of vehicles computed by (LinToVeh), if there are feasible solutions for both problems.*

Proof. Since the duration of the trips in (LinToVeh) are based on t_l, i.e., the minimal amount of time needed to operate a line, we need to consider the maximal increase in duration of a line in an optimal periodic timetable. The maximal amount of headway possible between two activities is $\frac{T}{2} - 1$, since otherwise there is no possibility of both activities covering the same infrastructure edge in the same period, i.e., there is no feasible periodic timetable.

Therefore, the worst case for any activity in a line is an increase in duration by factor $\frac{T}{2}$, increasing the number of vehicles needed of every circulation by at most $\frac{T}{2}$. ◀

Additionally, there exist instances where this worst case bound is obtained.

▶ **Example 10.** Consider a star shaped undirected trip graph L with 30 lines, a time period of 60, a trip time of 1 per line and a headway between leaving and entering a infrastructure edge of a vehicle of 29. Additionally, all circulations are allowed. Then (LinToVeh) will choose a single $(60, 30)$ circulation, covering all lines with a single vehicle. When respecting the headway constraints in (LinVehToTim), this circulation now needs 30 periods, i.e., 30 vehicles in total.

5 Integrated Planning

As a comparison to the sequential planning process presented in Sections 3 and 4, we additionally investigate the integrated problem (LinToTimVeh) of finding a periodic timetable and a vehicle schedule for a given line plan and set of possible circulations $\mathcal{C}$. For this, we use the model described in [30] while adding the possibility to restrict feasible vehicle schedules to a given set of circulations, i.e., we add the constraints

$$\sum_{\substack{c\in\mathcal{C}:\\ l\in c}} z_c = 1 \qquad l \in \mathcal{L} \tag{1}$$

$$y_a \geq z_c \qquad c \in \mathcal{C}, a \in A_{\text{turn}} : a \in c \tag{2}$$

Table 2 Sizes of the different datasets. Average trip time t_l is given in minutes.

	Stops in PTN	Edges in PTN	Lines	Nodes in $\overleftrightarrow{L}$	Arcs in $\overleftrightarrow{L}$	Avg t_l
`Toy`	8	8	13	8	26	8
`Sprinter`	416	448	32	38	64	39
`Intercity`	416	448	23	25	46	103

where constraints (1) are the cover constraints for the possible circulations $c \in \mathcal{C}$ and (2) couple the circulation constraints to the rest of the problem, where y_a determines whether a circulation activity $a \in A_{\text{turn}}$ is used. For the resulting complete model, see Appendix A.

We now investigate the connection between the integrated model and the sequential models described in Section 3 and 4.

▶ **Lemma 11.** *The optimal objective value of (LinToVeh) is a lower bound on (LinToTimVeh).*

Proof. We show that (LinToVeh) is a relaxation of (LinToTimVeh). The constraints of (LinToVeh), i.e., to cover each trip by exactly one vehicle circulation, are also constraints for (LinToTimVeh), such that the feasible set of (LinToTimVeh) is contained in the feasible set of (LinToVeh). For (LinToVeh), a circulation c contributes $k_c = \lceil \frac{t_c}{T} \rceil$ to the objective function. With $t_c = \sum_{a(l\cdot) \in c} t_l$ and t_l being the minimal time needed to operated a trip on line l, k_c is a lower bound on the actual number of vehicles needed to operate circulation c. ◀

In addition to this lower bound on the integrated problem, we also get an upper bound from (LinVehToTim).

▶ **Lemma 12.** *For a feasible periodic vehicle schedule, a corresponding feasible solution to (LinVehToTim) gives an upper bound on (LinToTimVeh).*

Proof. As (LinVehToTim) corresponds to solving (LinToTimVeh) for fixed circulation variables, any optimal solution of (LinVehToTim) remains feasible for the integrated problem thus giving an upper bound on the optimal objective value. ◀

We therefore have a validation criterion for the optimality of the sequential process:

▶ **Corollary 13.** *If the optimal objective values of (LinToVeh) and the corresponding problem (LinVehToTim) coincide, the corresponding solution is also optimal for (LinToTimVeh).*

If this is the case for possible circulations $\mathcal{C}'$ which consist of Eulerian tours for each connected component of $\overleftrightarrow{L}$, the number of vehicles is a lower bound on the objective of (LinToTimVeh) for any set $\mathcal{C}$ of possible circulations.

This result is especially helpful, since the runtime of the sequential models is much faster than of the integrated problem, as observed in Section 6.

6 Computational Results

For evaluating the developed models, we use the open source software library LinTim ([22, 23]). LinTim offers a variety of algorithms for various stages of public transport planning, such as line planning, timetabling, vehicle scheduling, delay management etc. As additionally all linking stages (e.g. constructing an even-activity network for a given line plan) as well as evaluation routines are implemented and test datasets are provided, new algorithms can easily be evaluated.

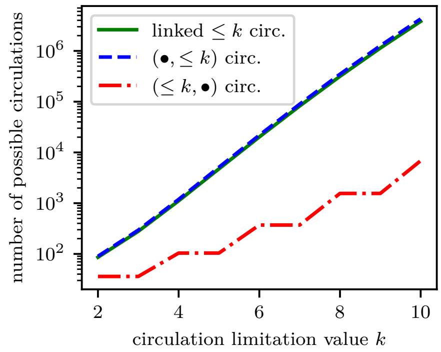


Figure 5 Number of possible circulations for `Sprinter`.

Table 3 Average runtime in seconds for the different models, average objective value and gap in parentheses, using all circulation limitation values β discussed in this section. (LinToTimVeh)* is provided a starting solution.

Dataset	(LinToVeh)	(LinVehToTim)	(LinToTimVeh)	(LinToTimVeh)*
`Toy`	1 (6.5, 0%)	222 (7.3, 0%)	3166 (9.7, 39%)	3366 (7.3, 22%)
`Sprinter`	1 (59.7, 0%)	31 (59.9, 0%)	3132 (64.1, 11%)	2687 (59.9, 3%)
`Intercity`	1 (102.5, 0%)	1626 (102.9, 0%)	3614 (110.9, 10%)	3283 (102.9, 2%)

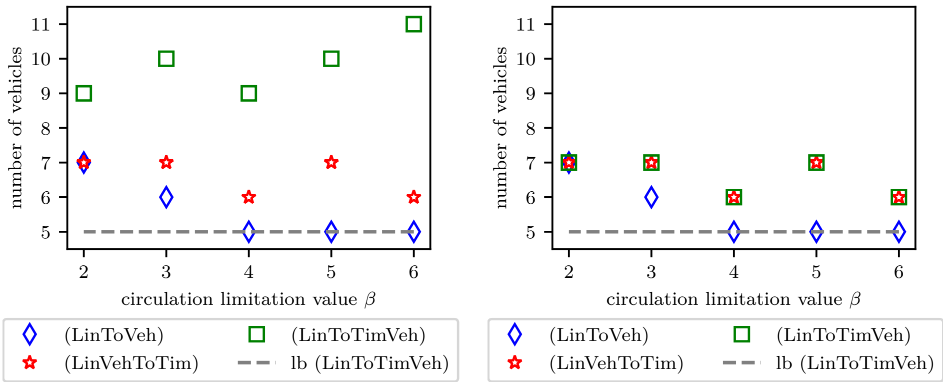
We use three different datasets, a small test dataset `Toy` and two close-to real world datasets `Sprinter` and `Intercity`, which are based on the railway network in the Netherlands. For an overview of the dataset sizes, see Table 2, and for the corresponding infrastructure networks as well as trip graphs Appendix B. To generate the possible circulations for different limitations, we use the open source library jGraphT ([17]), and especially the algorithm of Szwarcfiter and Lauer ([29]), to enumerate all possible cycles in the directed trip graphs while filtering the admissable ones. All models are solved using Gurobi 8.1.1 ([9]) on a compute server with a Intel Xeon E5-2670 and 96.6GB of RAM.

6.1 Investigating the Circulations

In Figure 5, we compare the number of circulations of a given form for dataset `Sprinter`. Here, we are comparing linked $\leq k$ circulations, $(\bullet, \leq k)$ circulations and $(\leq k, \bullet)$ circulations. Note that the first two circulation sets limit the number of lines, while the third one limits the number of trips, i.e, there are always fewer $(\leq k, \bullet)$ circulations. All circulation set sizes grow approximately exponential in size, resulting in problems with computing and storing the full sets for large k. The sizes for the sets of linked $\leq k$ circulations and $(\bullet, \leq k)$ circulations are nearly identical, with the set of linked $\leq k$ circulations being on average 7.6% smaller.

6.2 Comparing Sequential and Integrated Process

When comparing the integrated and the sequential planning process, there is a large disparity in the runtime of the different algorithms, see Table 3. On the one hand, (LinToVeh) can be solved to optimality within seconds for all datasets. (LinVehToTim) finds an optimal solution for the smaller datasets `Toy` and `Sprinter` within minutes and even for the largest instance `Intercity`, the average runtime is significantly lower than the time limit of one hour. On



(a) Without a starting solution. **(b)** With a starting solution.

Figure 6 Dataset `Toy`, comparing with or without starting solution for (LinToTimVeh) for $(\bullet, \leq \beta)$ circulations. The lower bound of the model is depicted as lb (LinToTimVeh).

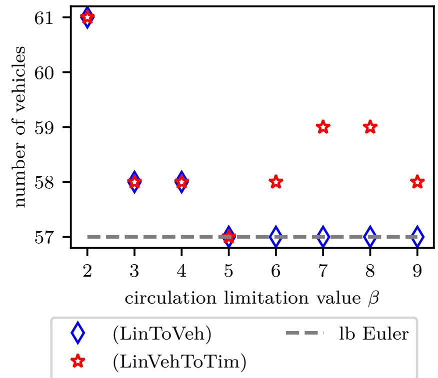


Figure 7 (LinToVeh) vs (LinVehToTim) for linked $\leq \beta$ circulations on `Sprinter`, including a lower bound provided by Eulerian circulations.

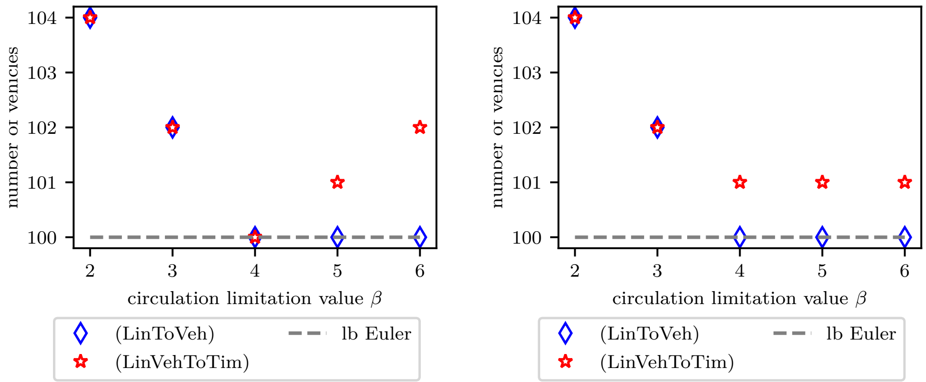
the other hand, (LinToTimVeh) seldom finds an optimal solution within the time limit. We therefore used the solution for the sequential process as a warm start for the integrated model. Figure 6 shows the effect of using a warm start for data set `Toy`. While the solution quality for the integrated model improves significantly, there is no instance where (LinToTimVeh) finds a better solution than sequentially solving (LinToVeh) and (LinVehToTim) within the time limit of one hour. Note especially that for $\beta = 5$, (LinToTimVeh) does not find the solution found for $\beta = 4$ which is still feasible with lower objective value.

6.3 Comparing (LinToVeh) to (LinVehToTim)

As shown in Lemma 5, the minimal number of vehicles needed to operate a set of circulations computed by (LinToVeh) decreases monotonically with increasing β. However, this does not hold for the number of vehicles needed for the corresponding timetable as illustrated in Figure 7 for dataset `Sprinter` and linked $\leq \beta$ circulations (this non-monotonic behavior can also be observed in Figures 6 and 8). Futhermore, we see in Figure 7 that for small limitation values $\beta \leq 5$, (LinToVeh) and (LinVehToTim) yield the same objective value which is therefore optimal for the integrated problem (LinToTimVeh), see Corollary 13. For larger $\beta > 5$, the number of vehicles needed for (LinVehToTim) surpasses the number computed by

Table 4 Number of different circulations sizes used in the optimal solutions of (LinToVeh) for $(\bullet, \leq \beta)$ circulations on `Intercity`. For comparison, the optimal objective values for (LinToVeh) and (LinVehToTim) are given as well.

limitation value β	$(\bullet,1)$	$(\bullet,2)$	$(\bullet,3)$	$(\bullet,4)$	$(\bullet,5)$	$(\bullet,6)$	obj (LinToVeh)	obj (LinVehToTim)
2	15	4	0	0	0	0	104	104
3	9	1	4	0	0	0	102	102
4	5	0	2	3	0	0	100	100
5	4	1	1	1	2	0	100	101
6	4	1	0	0	1	2	100	102



(a) $(\bullet, \leq \beta)$ circulations. **(b)** linked $\leq \beta$ circulations.

Figure 8 Effect of restricting to linked circulations on `Intercity`, including a lower bound provided by Eulerian circulations.

(LinToVeh). This is due to the fact that the solution for (LinToVeh) contains circulations with more trips as shown in Table 4. However, the number of vehicles computed by (LinToVeh) does not change such that the solution for $\beta = 5$ would still be optimal yielding a lower number of actually needed vehicles. This emphasizes that in practice shorter circulations are preferable but also shows that it might be beneficial to test a variety of circulation limitations. Additionally, Figure 7 contains the lower bound of (LinToVeh) provided by a Eulerian circulation, i.e., the minimal amount of vehicles possibly needed by any set of circulations, which is 57 vehicles in this instance. This means that the solution found by our sequential solution approach for $\beta = 5$ is an optimal solution when not considering circulation limitations.

6.4 Comparing Linked to General Circulations

When comparing linked and general circulations, Figure 8 shows that the solutions quality of (LinVehToTim) varies although the solution space of (LinToVeh) is smaller for linked circulations. This emphasizes again that the objective value of (LinToVeh) alone does not suffice to judge the solution quality of the sequential process and it is beneficial to test various sets of possible circulations. Note again that, as for `Sprinter` in Figure 7, the sequential approach for $(\bullet, \leq 4)$ circulations is able to find the minimal number of vehicles possible for any circulation set, provided by the Eulerian circulations.

7 Conclusion

In this paper we investigate the minimal number of vehicles needed to operate a given line plan. Instead of the traditional sequential approach of fixing a timetable first and a vehicle schedule second, we start by computing a periodic vehicle schedule. In order to limit a reduction of solution quality when a timetable is added, we restrict the set of circulations from which the vehicle schedule can be chosen. The resulting sequential approach is able to outperform an integrated formulation in terms of runtime and matches the solution quality on close-to real world datasets. For several instances, we can prove the optimality of the sequential approach for a given circulation limitation.

As the limitation of the circulation has a crucial influence on the solution quality, we suggest to further investigating this limitation. While we propose three ideas for limiting the circulations in this paper, further preprocessing of the admissable circulation set for finding "good" circulations beforehand may improve runtime and quality of the algorithms. Additionally, there may be possibilities to limit the maximal length of circulations needed for an instance beforehand, without losing solution quality.

In addition to the operational costs of a public transport system, which is the focus of this paper, passenger convenience is an important factor for gauging its quality. Thus adding passenger convenience into the models, e.g. by computing lexicographically optimal solutions concerning operational costs and passenger convenience, would extend the utility of the proposed model further.

References

1 R. Borndörfer, M. Grötschel, and M. Pfetsch. A column-generation approach to line planning in public transport. *Transportation Science*, 41(1):123–132, 2007.

2 R. Borndörfer, H. Hoppmann, and M. Karbstein. Passenger routing for periodic timetable optimization. *Public Transport*, 9(1-2):115–135, 2017.

3 R. Borndörfer, M. Karbstein, C. Liebchen, and N. Lindner. A Simple Way to Compute the Number of Vehicles That Are Required to Operate a Periodic Timetable. In Ralf Borndörfer and Sabine Storandt, editors, *18th Workshop on Algorithmic Approaches for Transportation Modelling, Optimization, and Systems (ATMOS 2018)*, volume 65 of *OpenAccess Series in Informatics (OASIcs)*, pages 16:1–16:15, Dagstuhl, Germany, 2018. Schloss Dagstuhl–Leibniz-Zentrum fuer Informatik. doi:10.4230/OASIcs.ATMOS.2018.16.

4 S. Bull, J. Larsen, R. Lusby, and N. Rezanova. Optimising the travel time of a line plan. *4OR*, October 2018. doi:10.1007/s10288-018-0391-5.

5 S. Bunte and N. Kliewer. An overview on vehicle scheduling models. *Public Transport*, 1(4):299–317, 2009.

6 M. Claessens, N. van Dijk, and P. Zwaneveld. Cost optimal allocation of rail passenger lines. *European Journal of Operational Research*, 110(3):474–489, 1998.

7 G. Desaulniers and M. Hickman. Public transit. *Handbooks in operations research and management science*, 14:69–127, 2007.

8 M. Goerigk and A. Schöbel. Improving the modulo simplex algorithm for large-scale periodic timetabling. *Computers & Operations Research*, 40(5):1363–1370, 2013.

9 Gurobi Optimizer. http://www.gurobi.com/, 2018. Gurobi Optimizer Version 8.1.1, Houston, Texas: Gurobi Optimization, Inc.

10 D. Huisman, L. Kroon, R. Lentink, and M. Vromans. Operations research in passenger railway transportation. *Statistica Neerlandica*, 59(4):467–497, 2005.

11 K. Li, H. Huang, and P. Schonfeld. Metro Timetabling for Time-Varying Passenger Demand and Congestion at Stations. *Journal of Advanced Transportation*, 2018, 2018.

12 C. Liebchen. Linien-, Fahrplan-, Umlauf-und Dienstplanoptimierung: Wie weit können diese bereits integriert werden? In *Heureka'08*, 2008.

13 C. Liebchen and R. Möhring. The modeling power of the periodic event scheduling problem: railway timetables—and beyond. In *Algorithmic methods for railway optimization*, pages 3–40. Springer, 2007.

14 R. Lusby, J. Larsen, M. Ehrgott, and D. Ryan. Railway track allocation: models and methods. *OR spectrum*, 33(4):843–883, 2011.

15 G. Maróti. *Operations research models for railway rolling stock planning.* PhD thesis, Eindhoven University of Technology, 2006.

16 M. Michaelis and A. Schöbel. Integrating Line Planning, Timetabling, and Vehicle Scheduling: A customer-oriented approach. *Public Transport*, 1(3):211–232, 2009.

17 D. Michail, J. Kinable, B. Naveh, and J. V. Sichi. Jgrapht—a java library for graph data structures and algorithms. *ACM Trans. Math. Softw.*, 46(2), May 2020.

18 K. Nachtigall. *Periodic Network Optimization and Fixed Interval Timetables.* PhD thesis, University of Hildesheim, 1998.

19 J. Pätzold, A. Schiewe, P. Schiewe, and A. Schöbel. Look-Ahead Approaches for Integrated Planning in Public Transportation. In Gianlorenzo D'Angelo and Twan Dollevoet, editors, *17th Workshop on Algorithmic Approaches for Transportation Modelling, Optimization, and Systems (ATMOS 2017)*, volume 59 of *OpenAccess Series in Informatics (OASIcs)*, pages 17:1–17:16, Dagstuhl, Germany, 2017. Schloss Dagstuhl–Leibniz-Zentrum für Informatik. doi:10.4230/OASIcs.ATMOS.2017.17.

20 J. Pätzold and A. Schöbel. A Matching Approach for Periodic Timetabling. In Marc Goerigk and Renato Werneck, editors, *16th Workshop on Algorithmic Approaches for Transportation Modelling, Optimization, and Systems (ATMOS 2016)*, volume 54 of *OpenAccess Series in Informatics (OASIcs)*, pages 1:1–1:15, Dagstuhl, Germany, 2016. Schloss Dagstuhl–Leibniz-Zentrum für Informatik. doi:10.4230/OASIcs.ATMOS.2016.1.

21 M. Reuther and T. Schlechte. Optimization of Rolling Stock Rotations. In *Handbook of Optimization in the Railway Industry*, pages 213–241. Springer, 2018.

22 A. Schiewe, S. Albert, J. Pätzold, P. Schiewe, A. Schöbel, and J. Schulz. LinTim: An integrated environment for mathematical public transport optimization. Documentation. Technical Report 2018-08, Preprint-Reihe, Institut für Numerische und Angewandte Mathematik, Georg-August-Universität Göttingen, 2018. URL: http://num.math.uni-goettingen.de/preprints/files/2018-8.pdf.

23 A. Schiewe, S. Albert, J. Pätzold, P. Schiewe, and A. Schöbel. LinTim - Integrated Optimization in Public Transportation. Homepage. http://lintim.math.uni-goettingen.de/, 2018. URL: http://lintim.math.uni-goettingen.de/.

24 P. Schiewe. *Integrated Optimization in Public Transport Planning*, volume 160 of *Optimization and Its Applications*. Springer, 2020. doi:10.1007/978-3-030-46270-3.

25 A. Schöbel. Line planning in public transportation: models and methods. *OR spectrum*, 34(3):491–510, 2012.

26 A. Schöbel. An eigenmodel for iterative line planning, timetabling and vehicle scheduling in public transportation. *Transportation Research Part C: Emerging Technologies*, 74:348–365, 2017.

27 A. Schöbel and S. Scholl. Line planning with minimal transfers. In *5th Workshop on Algorithmic Methods and Models for Optimization of Railways*, number 06901 in Dagstuhl Seminar Proceedings, 2006.

28 P. Serafini and W. Ukovich. A mathematical model for periodic scheduling problems. *SIAM Journal on Discrete Mathematics*, 2(4):550–581, 1989.

29 J. Szwarcfiter and P. Lauer. *Finding the elementary cycles of a directed graph in O (n+ m) per cycle.* University of Newcastle upon Tyne, 1974.

30 R. N. van Lieshout. Integrated periodic timetabling and vehicle circulation scheduling. Preprint, August 2019. URL: http://hdl.handle.net/1765/118655.

31 R. N. van Lieshout and P.C. Bouman. Vehicle Scheduling Based on a Line Plan. In Ralf Borndörfer and Sabine Storandt, editors, *18th Workshop on Algorithmic Approaches for Transportation Modelling, Optimization, and Systems (ATMOS 2018)*, volume 65 of *OpenAccess Series in Informatics (OASIcs)*, pages 15:1–15:14, Dagstuhl, Germany, 2018. Schloss Dagstuhl–Leibniz-Zentrum fuer Informatik. doi:10.4230/OASIcs.ATMOS.2018.15.

A IP Model for (LinToTimVeh), based on [30]

Notation

A:	Activities
$\lambda_a,\ a \in A$:	Passenger weight of activity a
$l_a, u_a,\ a \in A$:	Bounds of activity a
$\mathcal{B}$:	Integral cycle basis of EAN
$a_C, b_C,\ C \in \mathcal{B}$:	Bounds on the cycles in $\mathcal{B}$
$M_1 = T$	

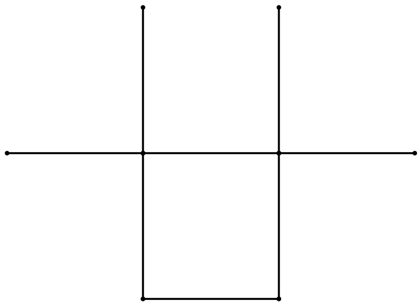
Variables

$x_a,\ a \in A$:	Tension of activity a
n:	Number of vehicles needed
$q_C,\ C \in \mathcal{B}$:	Periodicity variable of C
$y_a,\ a \in A_{\text{turn}}$:	Cover-variable for circulation activity a
$w_e,\ e \in E_{\text{end}}$:	Duration of activity after end event e
$z_c,\ c \in \mathcal{C}$:	Cover-variable for circulation c

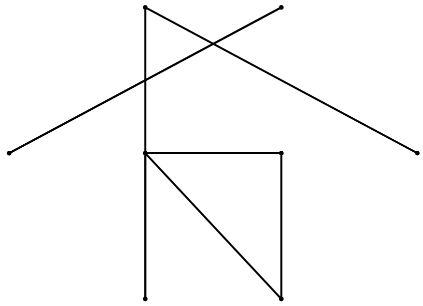
IP Model

$$
\begin{array}{lll}
\min & n & \\
\text{s.t.} & l_a \le x_a \le u_a & a \in A \\
& \displaystyle\sum_{a \in C^+} x_a - \sum_{a \in C^-} x_a = q_C \cdot T & C \in \mathcal{B} \\
& A_C \le q_C \le b_C & C \in \mathcal{B} \\
& n \ge \dfrac{1}{T}\left(\displaystyle\sum_{a \in A_{\text{veh}}} x_a + \sum_{e \in E_{\text{end}}} w_e\right) & \\
& w_e \ge x_a - M_1(1 - y_a) & e \in E_{\text{end}} \\
& w_e \le x_a + M_1(1 - y_a) & e \in E_{\text{end}} \\
& w_e \ge 0 & e \in E_{\text{end}} \\
& \displaystyle\sum_{\substack{c \in \mathcal{C}: \\ l \in c}} z_c = 1 & l \in \mathcal{L} \\
& y_a \ge z_c & c \in \mathcal{C}, a \in A_{\text{turn}} : a \in c \\
& n \in \mathbb{N} & \\
& x_a, y_a \in \mathbb{N} & a \in \mathcal{A} \\
& q_C \in \mathbb{N} & C \in \mathcal{B} \\
& w_e \in \mathbb{N} & e \in E_{\text{end}} \\
& z_c \in \{0, 1\} & c \in \mathcal{C}
\end{array}
$$

B Datasets



(a) PTN of dataset `Toy`.



(b) Trip graph L of dataset `Toy`.

Figure 9 PTN and trip graph of dataset `Toy`.

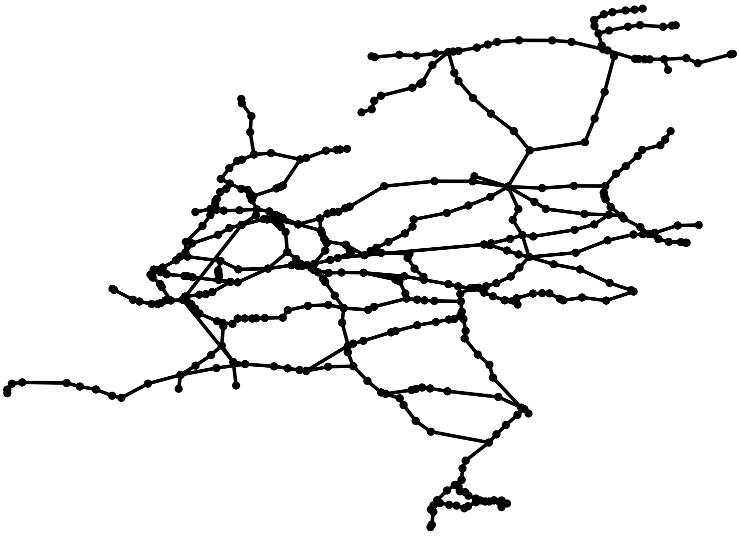
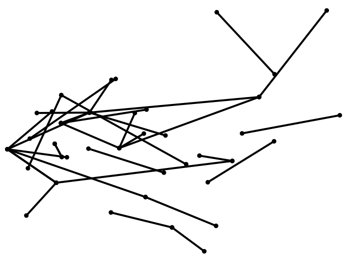
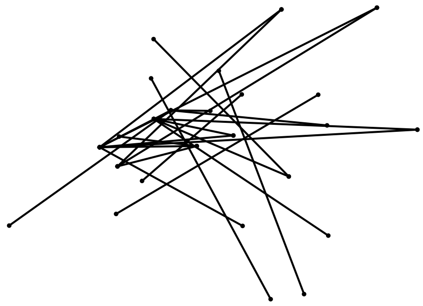


Figure 10 PTN of dataset `Sprinter` and `Intercity`.



(a) Trip graph L of dataset `Sprinter`.



(b) Trip graph L of dataset `Intercity`.

Figure 11 Trip graphs of dataset `Sprinter` and `Intercity`.