



Forward and Line-Based Cycle Bases for Periodic Timetabling

Berenike Masing¹ · Niels Lindner² · Patricia Ebert¹

Received: 27 January 2023 / Accepted: 19 May 2023 / Published online: 23 June 2023
© The Author(s) 2023

Abstract

The optimization of periodic timetables is an indispensable planning task in public transport. Although the periodic event scheduling problem (PESP) provides an elegant mathematical formulation of the periodic timetabling problem that led to many insights for primal heuristics, it is notoriously hard to solve to optimality. One reason is that for the standard mixed-integer linear programming formulations, linear programming relaxations are weak, and the integer variables are of pure technical nature and in general do not correlate with the objective value. While the first problem has been addressed by developing several families of cutting planes, we focus on the second aspect. We discuss integral forward cycle bases as a concept to compute improved dual bounds for PESP instances. To this end, we develop the theory of forward cycle bases on general digraphs. Specifically for the application of timetabling, we devise a generic procedure to construct line-based event-activity networks and give a simple recipe for an integral forward cycle basis on such networks. Finally, we analyze the 16 railway instances of the benchmark library PESPLib, match them to the line-based structure, and use forward cycle bases to compute better dual bounds for 14 out of the 16 instances.

Keywords Periodic event scheduling problem · Periodic timetabling · Cycle spaces · Cycle bases

1 Introduction

The optimization of timetables is a method to improve public transport in an inexpensive way: By improving transfer times, coordinating frequently used lines, selecting a timetable which is particularly stable with respect to delays, etc.,

✉ Berenike Masing
masing@zib.de

Niels Lindner
lindner@zib.de

Patricia Ebert
ebert@zib.de

¹ Zuse Institute Berlin, Berlin, Germany

² Freie Universität Berlin, Berlin, Germany

customer satisfaction can be increased significantly by simply re-planning trips without the need to invest in new infrastructure. In practice, particularly in the context of local transport systems, vehicle trips are oftentimes scheduled periodically, meaning that vehicles arrive and depart at regular intervals, e.g., every 20 or 60 min. A mathematical foundation for such a timetabling problem was formally introduced by Serafini and Ukovic as the *periodic event scheduling problem* (PESP) [1]. Several NP-hardness results have been derived in the context of PESP [2-4]. The problem is notoriously hard in practice as well, e.g., none of the 22 instances of the benchmark library `PESPLib` [5] could be solved to optimality so far. Although PESP can be formulated efficiently as a mixed integer linear programming problem, a great drawback is that the natural LP relaxation provides no information, so that standard methods such as branch-and-cut do not work particularly well.

In the past, a myriad of customized methods have been developed for the periodic event scheduling problem, many of which are based on cycles in the underlying so-called event-activity network: Nachtigall presents a cycle-based mixed-integer programming formulation for PESP using a cycle matrix of a fundamental cycle basis, i.e., a matrix whose rows are incidence vectors of the fundamental circuits of a spanning tree [6]. Liebchen and Peeters extend this concept to integral cycle bases [7]. The *cycle* [4], *change-cycle* [6], *multi-circuit* [8], *flip* [9] cutting planes are all supported on cycles. Lindner and Liebchen suggest a *timetable merging* strategy which among other features analyzes cycle offset variables [10]. *Tropical neighbourhood search* moves along columns of a cycle matrix [11]. In short, cycles play a prominent role in the context of periodic scheduling.

However, the number of potential cycle matrices that can be used for the cycle-based mixed-integer programming formulation is typically huge, and the choice has a significant impact on the progression of the dual bound. In our context, cycles can have both forward and backward arcs. When a cycle is a *forward cycle*, i.e., it uses exclusively forward arcs, it turns out that increasing the value of the integer variable associated to this cycle leads to an increase of the objective value, which is not necessarily the case if a cycle contains arcs in both directions. In particular, branching along such variables should be advantageous, and it is therefore expected that integral cycle bases consisting of forward cycles only will make standard mixed-integer programming techniques more effective. This has been experimentally confirmed for a single `PESPLib` instance [12].

Our goal now is to shed some light on the theoretical aspects of forward cycle bases and examine whether their good performance can be replicated on more instances. Our approach is thus threefold: Firstly, we formally introduce the concept of forward cycle bases and their properties and position it within the hierarchy of general cycle bases [13]. Secondly, we examine forward cycle bases specifically in the context of periodic timetabling. To that end, we propose a generic construction of an event-activity network based on line plans as are typically used in public transport planning. For such a *line-based event activity network*, we construct a structured integral forward cycle basis as required for the mixed-integer programming formulation for PESP. Lastly, we want to examine whether forward cycles indeed perform better with respect to lower bounds in practice. We analyze

the structure of the 16 railway instances in PESPLib, allowing us to apply our findings about line-based event-activity networks in practice. We conclude our study by running computational experiments for all those railway instances and compare the performance of our structured forward cycle basis with other promising candidates. Forward cycle bases turn out to indeed be the preferred choice: In all of our tests, a forward cycle formulation resulted in the best lower bounds. Even more, this approach contributed to an improvement of the lower bounds on 14 of the 16 considered instances in PESPLib.

This article is structured as follows: We shortly introduce PESP in Section 2. In Section 3, we cover the theory of cycle bases: Standard notions are recalled in Section 3.1, and the relevance of cycles in the context of PESP is motivated in Section 3.2. We discuss the concept of forward cycle bases and existence properties in Section 3.3, followed by the introduction of cycle bases of subspaces and their inherited properties in Section 3.4. In Section 4, we propose a generalized modeling approach for timetabling: Section 4.1 describes the generic construction of line-based event-activity networks, followed by the construction of a structured forward integral cycle basis in Section 4.2. Finally, we put the theory into practice in Section 5, by reverse-engineering a line plan from the PESPLib instances (Section 5.1), discussing how to algorithmically compute extremal forward cycle bases in Section 5.2, as well as performing computational experiments in Section 5.3. We conclude the paper in Section 6.

2 Periodic Event Scheduling

The periodic timetabling problem in public transport is commonly modeled in the language of the *periodic event scheduling problem* (PESP) introduced by Serafini and Ukovich [1]. A PESP instance consists of a 5-tuple (D, T, ℓ, u, w) , where

- D is a directed graph, called *event-activity network*, with vertex set $\mathcal{V}(D)$ and arc set $\mathcal{A}(D)$,
- $T \in \mathbb{N}_{\geq 3}$ is a *period time*,
- $\ell \in \mathbb{R}_{\geq 0}^{\mathcal{A}(D)}$ is a vector of *lower bounds* with $0 \leq \ell < T$,
- $u \in \mathbb{R}_{\geq 0}^{\mathcal{A}(D)}$ is a vector of *upper bounds* such that $0 \leq u - \ell < T$,
- $w \in \mathbb{R}_{\geq 0}^{\mathcal{A}(D)}$ is a vector of *weights*.

Applying preprocessing techniques (see, e.g., [14]) if necessary, we will assume that D has no loops and is weakly connected. A *periodic timetable* is a vector $\pi \in [0, T)^{\mathcal{V}(D)}$ such that there is a *periodic tension* $x \in \mathbb{R}^{\mathcal{A}(D)}$ satisfying

$$\forall a = (i, j) \in \mathcal{A}(D) : \ell_a \leq x_a \leq u_a \quad \text{and} \quad x_a \equiv \pi_j - \pi_i \pmod{T}. \quad (1)$$

PESP may be defined in an abstract way with a general digraph D . In the setting of periodic timetabling in public transport however, the vertices in $\mathcal{V}(D)$ model departure or arrival events of vehicles of a line at a station, and the arcs in $\mathcal{A}(D)$

represent activities such as, e.g., driving between two adjacent stations, dwelling or turning at a station, or passenger transfers. The graph D is sometimes called an *event-activity network*, as it encodes all relations between events and activities, e.g., if i describes the arrival of some train at some station, and j its subsequent departure, then an arc $a = (i, j)$ can be seen as a waiting activity. The bounds ℓ_a and u_a would then be interpreted as the minimum and maximum dwell time at that station. We refer to Liebchen and Möhring's in-depth discussion of the capabilities of the PESP model [15]. A periodic timetable π hence assigns timings in $(0, T)$ to each event, and a periodic tension x sets activity durations that are compatible with the event timings modulo the period time T . The weights w can be seen as a score of importance on the activities. Typically, the weights w are penalty factors for turning or waiting activities or estimate the number of passengers using an activity. It is therefore desirable to minimize $w^\top x$, e.g., the total idle time of vehicles or the total travel time of all passengers, among all periodic tensions x . Equivalently, one may seek to minimize the weighted periodic slack $w^\top x - w^\top \ell$:

Definition 2.1 (Periodic Event Scheduling Problem [1]) Given (D, T, ℓ, u, w) , the *periodic event scheduling problem (PESP)* is to find a periodic timetable π with a periodic tension x such that the weighted periodic slack $\sum_{a \in \mathcal{A}(D)} w_a(x_a - \ell_a)$ is minimum, or to decide that no periodic timetable exists.

The constraints (1) allow to formulate PESP as a mixed-integer linear program in a straightforward fashion. We are however interested in a different formulation based on cycles. To set up the stage for this formulation, we need to discuss some theory about cycle bases at first.

3 Theory of Cycle Bases

We briefly review the theory of the cycle space and cycle bases of directed graphs in Section 3.1 following the treatment as described by Kavitha et al. [13]. How cycle bases are tied back to PESP is covered in Section 3.2. We proceed with discussing forward cycle bases in Section 3.3 and introducing cycle bases of subspaces in Section 3.4.

3.1 Cycle Space and Cycle Bases

Let D be a weakly connected digraph with vertex set $\mathcal{V}(D)$ and arc set $\mathcal{A}(D)$.

Definition 3.1 (Cycle Space) The *cycle space* of D is

$$\mathcal{C}(D) := \left\{ \gamma \in \mathbb{Z}^{\mathcal{A}(D)} \mid \sum_{a \in \delta^+(v)} \gamma_a = \sum_{a \in \delta^-(v)} \gamma_a \right\}.$$

An element $\gamma \in \mathcal{C}(D)$ is called a *cycle*. For $a \in \mathcal{A}(D)$, we write $a \in \gamma$ if $\gamma_a \neq 0$. An arc $a \in \gamma$ is called a *forward arc* of γ if $\gamma_a > 0$ and *backward* if $\gamma_a < 0$. A cycle γ is a *circuit* if $\gamma \in \mathcal{C}(D) \cap \{-1, 0, 1\}^{\mathcal{A}(D)}$. The subgraph of D induced by $\{a \in \mathcal{A}(D) \mid a \in \gamma\}$ need not be weakly connected. A circuit γ is *simple* if no vertex of D is incident with more than two arcs $a \in \gamma$.

Observation 3.2 The cycle space $\mathcal{C}(D)$ is a free abelian group. Its rank is given by the *cyclomatic number*

$$\mu(D) = |\mathcal{A}(D)| - |\mathcal{V}(D)| + 1.$$

One may also extend the cycle space to coefficients in a field κ , i.e., the κ -vector space

$$\mathcal{C}(D) \otimes_{\mathbb{Z}} \kappa = \left\{ \gamma \in \kappa^{\mathcal{A}(D)} \mid \sum_{a \in \delta^+(v)} \gamma_a = \sum_{a \in \delta^-(v)} \gamma_a \right\},$$

which is of dimension $\mu(D)$ (see, e.g., [13, Theorem 2.3])

Definition 3.3 (Cycle Basis) A *cycle basis* \mathcal{B} of D is a set of circuits that constitutes a basis of the \mathbb{Q} -vector space $\mathcal{C}(D) \otimes_{\mathbb{Z}} \mathbb{Q}$.

A cycle basis \mathcal{B} hence consists of $\mu := \mu(D)$ circuits $\gamma_1, \dots, \gamma_{\mu} \in \mathcal{C}(D) \cap \{-1, 0, 1\}^{\mathcal{A}(D)}$ such that every cycle $\gamma \in \mathcal{C}(D) \otimes_{\mathbb{Z}} \mathbb{Q}$ with rational coefficients can be written as a unique rational linear combination of $\gamma_1, \dots, \gamma_{\mu}$.

Let $G(D)$ denote the underlying undirected graph of D . For each $a \in \mathcal{A}(D)$, we write $e(a) \in \mathcal{E}(G)$ for the edge of $G(D)$ corresponding to the arc a . By a *spanning tree* of D , we mean a subgraph $\mathcal{T} \subseteq D$ such that the corresponding subgraph $G(\mathcal{T}) \subseteq G(D)$ is a spanning tree. To each *co-tree arc* $a \in \mathcal{A}(D) \setminus \mathcal{A}(\mathcal{T})$, we can associate the *fundamental circuit* of a w.r.t. \mathcal{T} , which is the unique circuit γ in D that satisfies $\gamma_a = 1$ and projects to the unique cycle in the undirected graph that arises by adding $e(a)$ to $G(\mathcal{T})$.

Definition 3.4 (Classes of Cycle Bases [13, Definition 3.1]) Let $\mathcal{B} = \{\gamma_1, \dots, \gamma_{\mu}\}$ be a cycle basis of D . We call \mathcal{B}

- i) *fundamental* if there exists a spanning tree \mathcal{T} of D such that \mathcal{B} is the set of fundamental circuits of the μ co-tree arcs of \mathcal{T} ;
- ii) *weakly fundamental* if there exists a permutation σ such that

$$\forall i \in \{2, \dots, \mu\} : \gamma_{\sigma(i)} \setminus (\gamma_{\sigma(1)} \cup \dots \cup \gamma_{\sigma(i-1)}) \neq \emptyset; \tag{2}$$

- iii) *integral* if \mathcal{B} is a basis of the free abelian group $\mathcal{C}(D)$, i.e., every cycle $\gamma \in \mathcal{C}(D)$ can be written as a unique *integral* linear combination of the elements of \mathcal{B} ;
- iv) *undirected* if \mathcal{B} is a basis of the \mathbb{F}_2 -vector space $\mathcal{C}(D) \otimes_{\mathbb{Z}} \mathbb{F}_2$, where \mathbb{F}_2 denotes the finite field with two elements.

Since any directed graph has a spanning tree, there are always fundamental cycle bases. While Definition 3.3 is about rational coefficients of cycles, iii) deals with integral coefficients, and iv) with coefficients in \mathbb{F}_2 . Any fundamental cycle basis is weakly fundamental, and by algebraic considerations, any integral cycle basis is undirected. It is also true that weakly fundamental cycle bases are integral [16, Corollary 30]. This implies the following hierarchy of cycle bases:

fundamental \implies weakly fundamental \implies integral \implies undirected.

Definition 3.5 (Cycle Matrix) Let \mathcal{B} be a cycle basis of D . The matrix $\Gamma \in \{-1, 0, 1\}^{\mathcal{B} \times A(D)}$, whose rows are the elements of \mathcal{B} , is called the cycle matrix of \mathcal{B} .

3.2 Cycles in the Context of PESP

The significance of cycle bases for periodic timetabling becomes evident from the following theorem:

Theorem 3.6 (Cycle Periodicity Property [14, 17]) For a PESP instance (D, T, ℓ, u, w) and an integral cycle basis \mathcal{B} of D , the vector $x \in \mathbb{Q}^{A(D)}$ with $\ell \leq x \leq u$ is a periodic tension if and only if there exists a vector $z \in \mathbb{Z}^{\mathcal{B}}$ such that $\gamma^\top x = Tz_\gamma$ for each $\gamma \in \mathcal{B}$.

This allows us for the following MIP formulation of the periodic event scheduling problem using a cycle matrix:

Corollary 3.7 (Cycle-Based MIP Formulation of PESP) For a PESP instance (D, T, ℓ, u, w) , an integral cycle basis \mathcal{B} of D with corresponding cycle matrix Γ , the vector $x \in \mathbb{Q}^{A(D)}$ is an (optimal) periodic tension of PESP if and only if x is an (optimal) solution to

$$\begin{aligned} \min \quad & w^\top(x - \ell) \\ \text{s.t.} \quad & \ell \leq x \leq u \\ & \Gamma x = Tz \\ & z \in \mathbb{Z}^{\mathcal{B}} \\ & x \in \mathbb{Q}^{A(D)} \end{aligned} \tag{3}$$

It is easy to see that an optimal solution to the LP relaxation of (3) is trivially obtained, namely, with $x = \ell$ and $z = \Gamma\ell/T$, which results in a weighted slack of 0. Consequently, the LP relaxation does not provide additional information. To improve this lower bound, one could add valid inequalities to the LP relaxation. A useful inequality for PESP was introduced by Odijk [4]:

Lemma 3.8 Let (D, T, ℓ, u, w) be a PESP instance and x a periodic tension. Then for all cycles $\gamma \in \mathcal{C}(D)$, the cycle inequality

$$\left| \frac{\gamma_+^\top \ell - \gamma_-^\top u}{T} \right| \leq \frac{\gamma^\top x}{T} \leq \left| \frac{\gamma_+^\top u - \gamma_-^\top \ell}{T} \right| \tag{4}$$

holds, where $\gamma_+ := \max(\gamma, 0)$ and $\gamma_- := \max(-\gamma, 0) \in \mathbb{Z}_{\geq 0}^{A(D)}$ are the positive and negative parts of γ , respectively, so that $\gamma = \gamma_+ - \gamma_-$.

The cycle inequalities belong to the larger family of *flip inequalities* [9]. In the case that $\gamma_- = 0$, the cycle inequality for γ is equivalent to the *change-cycle inequality* for γ [6]. Odijk’s cycle inequalities can be interpreted as an information of how much aggregated tension $\gamma^\top x$ needs to be distributed along a cycle. Adding these inequalities to the LP relaxation essentially corresponds to increasing the tension among arcs of certain cycles away from the trivial solution.

An intuitive approach to quickly increase the dual bound is to pick cycles whose inequality contributes to a particularly high lower bound, thus leading to a significant increase of aggregated tensions. In the dual sense, the objective is to maximize; one wants to select cycles that lead to high costs. Suppose the addition of a cycle inequality leads to an increase of tension along said cycle. This can be achieved by an increase of the tension on an arc $a \in \gamma$ if a is forward ($\gamma_a = 1$) or by a decrease if a is backward ($\gamma_a = -1$). Observe that the tension of each arc contributes to the objective function, independent of its orientation within cycles. Thus, an increase in the objective can be obtained if the additional tension is distributed along some forward arcs. If the cycle contains also backward arcs, the increased tension requirement may result in only marginal changes on the dual bounds, due to the opposite signs. Consequently, cycles with forward arcs only seem more promising for quickly raising the dual bounds.

Among the cycles using arcs in one direction only, which ones should be selected for a basis? Since we have $\gamma^\top x = Tz_\gamma$ for a cycle γ of the cycle basis, it is preferable to use basis elements, where the upper and the lower bound of Odijk’s cycle inequalities are particularly close together, since this restricts the number of possible integral values for z_γ , which is relevant for a branch-and-bound approach. A reasonable choice thus tends to be cycles with small aggregated span $\gamma^\top (u - \ell)$ [7]. An alternative choice could be high weight cycles, as they contribute to the objective significantly, making them promising candidates for a quick increase of the dual bound.

Example 3.9 For a better understanding, consider the small PESP instance in Fig. 1. This graph contains (up to sign) 7 circuits, namely, $\gamma_1, \gamma_2, \gamma_3, \gamma_1 + \gamma_2, \gamma_1 + \gamma_3, \gamma_2 + \gamma_3, \gamma_1 + \gamma_2 + \gamma_3$. The circuits $\gamma_1, \gamma_2, \gamma_3$ are the ones with smallest span and would be the traditional first

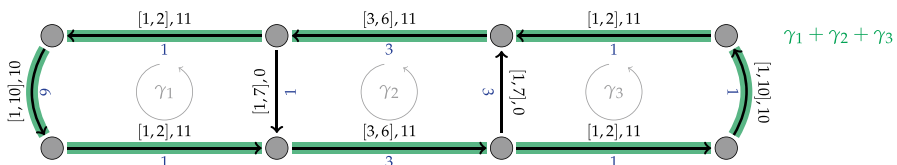


Fig. 1 Example instance by Lindner et al. [12] with period time $T = 10$ and labels $\frac{[\ell_a, u_a], w_a}{x_a}$

choice to increase the dual bound. Yet, their addition will not lift the dual bound, as the corresponding cycle inequalities do not cut off the trivial solution $x = \ell$ to the LP relaxation. In contrast, the addition of the cycle inequality of γ_2 and $\gamma_1 + \gamma_2 + \gamma_3$ (marked in green) to the LP relaxation is enough to detect the optimal solution of this instance, as is indicated by the blue tension values. Observe that both are forward cycles, γ_2 is one of the cycles with the tightest bounds w.r.t. (4) and $\gamma_1 + \gamma_2 + \gamma_3$ is the cycle whose minimum arc weight is maximum. In fact, the latter is solely responsible for increasing the dual bound. For a more detailed discussion of this example, we refer to Lindner, Liebchen, and Masing [12].

We want to evaluate whether our intuition pays off in practical situations: Do cycle basis formulations perform significantly better if the cycles are all forward? Does the cycle selection play a role, i.e., are small span, or heavy weight cycles better than generic forward cycles? To that end, we will first discuss the theory of forward cycles in Section 3.3, then show how to strategically construct forward cycle bases on a line-based event-activity network in Section 4, and finally reach the computational experiments in Section 5 where we compare the performance of different types of cycle bases.

3.3 Forward Cycle Bases

Definition 3.10 (Forward Cycle Basis) A cycle $\gamma \in \mathcal{C}(D)$ is called a *forward cycle* if it uses only forward edges, i.e., if $\gamma_a > 0$ for all $a \in \gamma$. A cycle basis consisting exclusively of forward circuits is called a *forward cycle basis*.

Let D be digraph. By a *block* of D , we mean a subgraph of D that projects to a biconnected component of the underlying undirected graph $G(D)$. Analogously, a *weakly 2-edge-connected component* is a subgraph of D that projects to a 2-edge-connected component of $G(D)$.

Theorem 3.11 (Existence Criteria of Forward Cycle Bases) *The following are equivalent for a directed graph D :*

- i) D has a forward cycle basis.
- ii) Each block of D is either strongly connected or a single arc.
- iii) Each weakly 2-edge-connected component of D is strongly connected.

Proof The equivalence of i) and ii) was shown by Gleiss, Leydold, and Stadler [18, Theorem 7]. Since for all $\gamma \in \mathcal{C}(D)$ holds $\gamma_a = 0$ when $e(a)$ is a bridge of $G(D)$, the cycle space of D is the direct sum of the cycle spaces of its weakly 2-edge-connected components. Strongly connected digraphs admit a forward cycle basis [19, Proposition 2.1], thus iii) implies i).

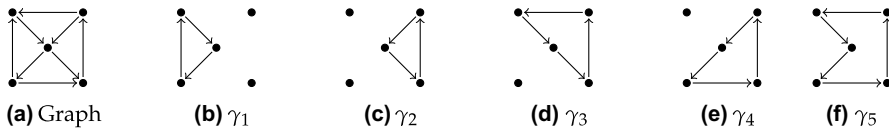


Fig. 2 Strongly connected graph without a fundamental forward cycle basis (cf. Lindner et al. [12])

To see that i) implies iii), let $a \in \mathcal{A}(D)$ be an arc within one weakly 2-edge-connected component C , and let \mathcal{B} constitute a forward cycle basis of D . Since $e(a)$ is not a bridge of $G(D)$, a is contained in some cycle $\gamma \in \mathcal{C}(D)$, but as γ is a linear combination of the circuits in \mathcal{B} , a is contained in at least one of the circuits in \mathcal{B} . In particular, if $a = (i, j)$, then there is a j - i -path in C . This implies that C is strongly connected.

Example 3.12 Not every digraph with a forward cycle basis has a fundamental forward cycle basis. Consider for example the graph in Fig. 2a. Its cyclomatic number is 4, and it has exactly five forward cycles, which are displayed in Fig. 2b–f. Observe that $\gamma_2 + \gamma_5 = \gamma_3 + \gamma_4$: Any combination of γ_1 with three of the cycles $\gamma_2, \dots, \gamma_5$ forms a cycle basis of the graph. A spanning tree of the graph corresponds to four co-tree arcs. However, any combination of the forward cycles forming a basis has at most three arcs covered by one unique cycle only. Consequently, this graph cannot have a fundamental forward cycle basis. It has a *weakly* fundamental forward cycle basis however, e.g., the ordering $\gamma_4, \gamma_2, \gamma_3, \gamma_1$ fulfils the property (2).

Lemma 3.13 Every undirected graph G has an orientation such that the induced directed graph D with $G(D) = G$ has a fundamental forward cycle basis.

Proof Without loss of generality, we can assume that G is connected. Consider a depth-first search tree \mathcal{T} of G from some root $r \in \mathcal{V}(G)$. Since G is connected, \mathcal{T} is spanning. Any co-tree edge $e \in \mathcal{E}(G) \setminus \mathcal{E}(\mathcal{T})$ is a back edge and not a cross edge [20, Theorem 22.10]. By inducing an orientation along each edge $e \in \mathcal{E}(\mathcal{T})$ from r towards the leaves of \mathcal{T} , and backwards along the co-tree arcs, i.e., towards the ancestor, all fundamental circuits in the resulting digraph become forward cycles.

3.4 Cycle Bases of Subspaces

Let D be a weakly connected digraph. For later use, we will generalize the concept of cycle bases to subspaces of the cycle space.

Definition 3.14 (Cycle Bases of Subspaces) Let $\mathcal{B}' = \{\gamma_1, \dots, \gamma_{\mu'}\}$ be a set of circuits in $\mathcal{C}(D)$ and consider $\text{span}_{\mathbb{Q}}(\mathcal{B}')$, the \mathbb{Q} -linear subspace of $\mathcal{C}(D) \otimes_{\mathbb{Z}} \mathbb{Q}$ generated by \mathcal{B}' . If \mathcal{B}' is linearly independent over \mathbb{Q} , we call \mathcal{B}' a *cycle basis of the subspace* $\text{span}_{\mathbb{Q}}(\mathcal{B}')$. Furthermore, in this case, the basis \mathcal{B}' is called

- i) *fundamental* if for each $\gamma \in \mathcal{B}'$ there exists an arc $a \in \gamma$ which is uniquely in γ , i.e., $a \notin \gamma'$ for all $\gamma' \in \mathcal{B}' \setminus \{\gamma\}$;
- ii) *weakly fundamental* if there is a permutation σ of the cycles in \mathcal{B}' such that

$$\forall i \in \{2, \dots, \mu'\} : \gamma_{\sigma(i)} \setminus (\gamma_{\sigma(1)} \cup \dots \cup \gamma_{\sigma(i-1)}) \neq \emptyset; \tag{5}$$

- iii) *integral* if $\text{span}_{\mathbb{Q}}(\mathcal{B}') \cap \mathbb{Z}^{A(D)} = \text{span}_{\mathbb{Z}}(\mathcal{B}')$, i.e., every integer cycle in $\text{span}_{\mathbb{Q}}(\mathcal{B}')$ can be expressed as an integer linear combination of cycles in \mathcal{B}' ;
- iv) *undirected* if \mathcal{B}' is linearly independent over \mathbb{F}_2 .

Remark 3.15 Note that while the cycle space $\mathcal{C}(D')$ of subgraph $D' \subseteq D$ gives rise to a subspace of $\mathcal{C}(D) \otimes_{\mathbb{Z}} \mathbb{Q}$, not every subspace $\text{span}_{\mathbb{Q}}(\mathcal{B}')$ of $\mathcal{C}(D) \otimes_{\mathbb{Z}} \mathbb{Q}$ is also a cycle space of a subgraph. This becomes evident when considering the cycles in Example 3.12 again: The three cycles $\gamma_3, \gamma_4, \gamma_5$ span the cycle space of the subgraph D' with the leftmost arc removed and $\mu(D') = 3$. In contrast, the 3-dimensional space $\text{span}_{\mathbb{Q}}(\{\gamma_1, \gamma_2, \gamma_5\})$ covers every arc of D , but $\mu(D) = 4$.

The following proposition assures the compatibility with the previous classification of cycle bases.

Proposition 3.16 *Let \mathcal{B} be a (fundamental/weakly fundamental/integral/undirected) cycle basis of D . Then, any subset $\mathcal{B}' \subseteq \mathcal{B}$ is a (fundamental/weakly fundamental/integral/undirected) cycle basis of the subspace $\text{span}_{\mathbb{Q}}(\mathcal{B}')$.*

Proof Let $\mathcal{B} = \{\gamma_i, i \in [\mu]\}$ be a cycle basis of D , and let $\mathcal{I} \subseteq [\mu]$ be the index set such that $\mathcal{B}' = \{\gamma_i \in \mathcal{B} \mid i \in \mathcal{I}\}$. It is clear that \mathcal{B}' is a cycle basis of the subspace $\text{span}_{\mathbb{Q}}(\mathcal{B}')$. We will now go through the four classes of cycle bases of Definitions 3.4 and 3.14:

- i) Suppose that \mathcal{B} is fundamental w.r.t. some spanning tree \mathcal{T} . Then each cycle of \mathcal{B} contains a unique co-tree arc of \mathcal{T} .
- ii) Suppose that \mathcal{B} is weakly fundamental. Reordering the cycles, we can assume that \mathcal{B} satisfies (2) with $\sigma = \text{id}$. But then \mathcal{B}' satisfies (5) with $\sigma = \text{id}$.
- iii) Suppose that \mathcal{B} is integral. If $\gamma \in \text{span}_{\mathbb{Q}}(\mathcal{B}') \cap \mathbb{Z}^{A(D)}$, then $\gamma \in \mathcal{C}(D)$, thus γ can be written as a unique integer linear combination of the cycles in \mathcal{B} . Since $\gamma \in \text{span}_{\mathbb{Q}}(\mathcal{B}')$, the coefficients w.r.t. cycles not in \mathcal{B}' must vanish.
- iv) Suppose that \mathcal{B} is undirected. Since \mathcal{B} is linearly independent over \mathbb{F}_2 , so is the subset \mathcal{B}' .

The hierarchy of cycle bases extends as well to cycle bases of subspaces:

Proposition 3.17 *Let $\mathcal{B}' = \{\gamma_1, \dots, \gamma_{\mu'}\}$ be a cycle basis of the subspace $\text{span}_{\mathbb{Q}}(\mathcal{B}')$. Then, the following implications hold:*

$$\text{fundamental} \xRightarrow{\text{i)}} \text{weakly fundamental} \xRightarrow{\text{ii)}} \text{integral} \xRightarrow{\text{iii)}} \text{undirected}$$

Proof

- i) If \mathcal{B}' is fundamental, then \mathcal{B}' satisfies (5) with $\sigma = \text{id}$.
- ii) Suppose that \mathcal{B}' is weakly fundamental. By (5), we can assume that there is a sequence $a_1, \dots, a_{\mu'}$ of arcs such that for all $i \in \{2, \dots, \mu'\}$ holds $a_i \in \gamma_i$, but $a_i \notin \gamma_j$ for $j < i$. Let $\gamma \in \text{span}_{\mathbb{Q}}(\mathcal{B}') \cap \mathbb{Z}^{A(D)}$. Since \mathcal{B}' is a \mathbb{Q} -basis of $\text{span}_{\mathbb{Q}}(\mathcal{B}')$, we can express γ as a linear combination of $\gamma_1, \dots, \gamma_{\mu'}$ by means of a coefficient vector λ . In particular, λ satisfies the system of linear equations

$$\sum_{j=1}^{\mu'} \gamma_{j,a_i} \lambda_j = \gamma_{a_i}, \quad i \in \{1, \dots, \mu'\}. \tag{6}$$

The coefficient matrix of (6) is upper triangular, and since all cycles in \mathcal{B}' are circuits, all diagonal entries are ± 1 . We conclude that the coefficient matrix has an integer inverse, so that by Cramer’s rule, λ is unique and integral.

- iii) Suppose that \mathcal{B}' is integral. Let $\lambda_1, \dots, \lambda_{\mu'} \in \mathbb{F}_2$ such that $\sum_{j=1}^{\mu'} \lambda_j \gamma_j = 0$ over \mathbb{F}_2 . Choosing $v_j \in \mathbb{Z}$ with $v_j \equiv \lambda_j \pmod{2}$ for $j \in \{1, \dots, \mu'\}$, this means that $\sum_{j=1}^{\mu'} v_j \gamma_j = 2\gamma$ for some $\gamma \in \text{span}_{\mathbb{Q}}(\mathcal{B}') \cap \mathbb{Z}^{A(D)}$. This implies that $\gamma = \sum_{j=1}^{\mu'} \frac{v_j}{2} \gamma_j$ over \mathbb{Q} , and since \mathcal{B}' is an integral cycle basis of $\text{span}_{\mathbb{Q}}(\mathcal{B}')$, the coefficients of γ w.r.t. \mathcal{B}' must be integral, so that all v_j are divisible by 2. We conclude that $\lambda_1 = \dots = \lambda_{\mu'} = 0 \in \mathbb{F}_2$.

4 Line-Based Cycle Bases

While PESP can be applied to general periodic scheduling problems, e.g., for automated production systems [21] or coordinated traffic signals [22], the main application case is timetabling in the context of public transport. We will present a way to construct a generic event-activity network based on a given line plan and an explicit line-based integral forward cycle basis.

To this end, we represent the infrastructure of a given city by a simple undirected graph \mathcal{I} . Its vertices $S := \mathcal{V}(\mathcal{I})$ correspond to *stations* and its edges $\mathcal{E}(\mathcal{I})$ to possible connections. We consider a *line plan* \mathcal{L} as a set of bidirectional lines represented by simple undirected paths on \mathcal{I} . The *line network* is then an undirected graph \mathcal{N} whose vertices are the stations and whose edges are induced by the lines. If two lines share the same edge $e \in \mathcal{E}(\mathcal{I})$, we create two distinct parallel edges in \mathcal{N} . More formally, we define the line network as follows:

Definition 4.1 (Line Network \mathcal{N}) The *line network* \mathcal{N} associated to $(\mathcal{I}, \mathcal{L})$ is the undirected graph with $\mathcal{V}(\mathcal{N}) := S$ and $\mathcal{E}(\mathcal{N}) := \{\{u, v\}^{(\ell)} \mid \{u, v\} \in \mathcal{E}(\ell) \text{ for } \ell \in \mathcal{L}\}$. For $s \in S$, we will denote the degree of the node s in \mathcal{N} by d_s .

We will tacitly assume that \mathcal{N} is connected.

4.1 Line-Based Event-Activity Networks

Based on the line network, we now want to determine a timetable, i.e., departure and arrival times for each stop on each line in \mathcal{L} . For transfers, we make the following assumptions:

- If two distinct lines meet in a station, transfers are allowed to and from both lines.
- Transfers within the same line are prohibited (both in the same and return direction).

We propose the following generic construction for an event-activity network based on a line network, which is also illustrated in Fig. 3.

Definition 4.2 (Line-Based Event-Activity Network EAN) Let \mathcal{N} be a line network associated to $(\mathcal{I}, \mathcal{L})$, and let d_s denote the degree of station $s \in S$ in \mathcal{N} . Fix a bijective labeling $v_s : \delta(s) \rightarrow [d_s]$ of the edges of \mathcal{N} incident with s . We construct the event-activity network EAN based on \mathcal{N} and v as the following directed graph:

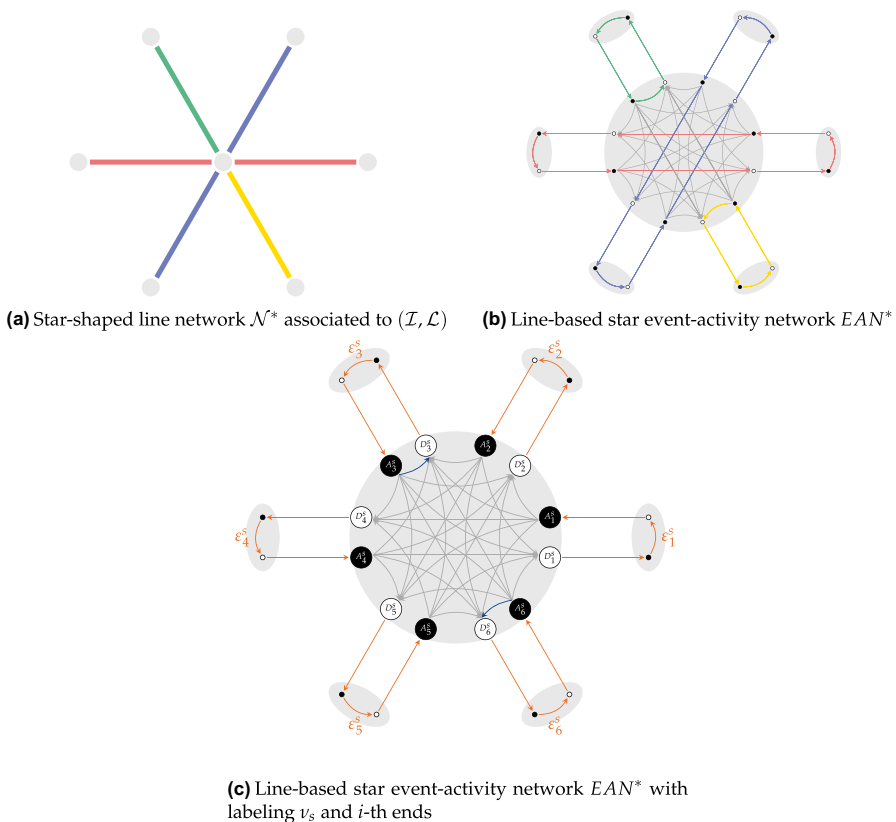


Fig. 3 Example of Definitions 4.2 and 4.4

- i) Start with an empty digraph.
- ii) For each station $s \in S$ and each $i \in [d_s]$, add a *departure node* D_i^s and an *arrival node* A_i^s .
- iii) For each line $\ell \in \mathcal{L}$, add
 - two *driving arcs* $(D_{v_s(e)}^s, A_{v_t(e)}^t)$ and $(D_{v_t(e)}^t, A_{v_s(e)}^s)$ for each $e = \{s, t\}^{(\ell)} \in \mathcal{E}(\ell)$,
 - two *dwell arcs* $(A_{v_s(e)}^s, D_{v_s(e')}^s)$ and $(A_{v_s(e')}^s, D_{v_s(e)}^s)$ for each non-terminal station (i.e., internal node) $s \in \mathcal{V}(\ell)$ with the two incident edges e and e' ,
 - two *turn arcs* $(A_{v_s(e)}^s, D_{v_s(e)}^s)$ and $(A_{v_t(e')}^t, D_{v_t(e')}^t)$, where s and t are the terminal stations (i.e., path endpoints) of ℓ , and e and e' are the first and last edge of ℓ , respectively.
- iv) For each station $s \in S$ and $i, j \in [d_s]$ with $i \neq j$, add a *transfer arc* (A_i^s, D_j^s) unless this arc has been added previously as a dwell arc.

The driving arcs (D_i^s, A_j^t) and (D_j^t, A_i^s) are called *antiparallel partner arcs*. For a departure node D_i^s resp. arrival node A_i^s , we define the *line association* $\ell(D_i^s)$ resp. $\ell(A_i^s)$ as the unique line in \mathcal{L} to which the edge $v_s^{-1}(i)$ belongs. Finally, we define the *i -th end at s* as the unique directed simple path ε_i^s from D_i^s to A_i^s using only nodes with the same line association, i.e., using no transfer arcs.

Observation 4.3 For an event-activity network *EAN*, we can identify arc types by their labels:

- i) If $(D_i^s, A_j^t) \in \mathcal{A}(EAN)$, then it is a driving arc of line $\ell(D_i^s) = \ell(A_j^t)$.
- ii) If $(A_i^s, D_j^s) \in \mathcal{A}(EAN)$, then it is a turn arc of line $\ell(D_i^s)$.
- iii) Any $(A_i^s, D_j^s) \in \mathcal{A}(EAN)$ with $i \neq j$ is either a transfer or a dwell arc.
- iv) If $(A_i^s, D_j^s) \in \mathcal{A}(EAN)$ is a turn arc, then any (A_i^s, D_j^s) with $i \neq j$ is a transfer arc.
- v) If $(A_i^s, D_j^s) \notin \mathcal{A}(EAN)$, then there exists a unique $j \in [d_s], i \neq j$ such that (A_i^s, D_j^s) is a dwell arc of line $\ell(D_i^s)$, and all other arcs (A_i^s, D_k^s) with $k \in [d_s] \setminus \{i, j\}$ are transfer arcs to other lines.
- vi) The two arcs (A_i^s, D_j^s) and (A_j^s, D_i^s) are of the same type.

Note that every i -th end ε_i^s contains exactly one turn arc outside of station s . Such an ε_i^s can be interpreted as a sub-sequence of activities of an associated line: It consists of the entire sub-path of a line starting at station s all the way to its terminus station (in direction as indicated by e with $v_s(e) = i$), turning in t and returning back to s .

Please note that we do not add any headway arcs in the line-based event-activity network, which might seem like a unrealistic oversight. In practice, headways are an important tool to evade scheduling conflicts when vehicles use the same infrastructure. However, they are highly dependent on the specific infrastructure, e.g., a headway arc might be needed within a station when two lines are routed over the same platform, but not if they use different platforms. Observe that this level of detail is not reflected in a given line network \mathcal{N} —there is no

information about station size, available switches, capacity restrictions, etc. As our construction is line-based, in the sense that it relies on the information which can be obtained from a given line network \mathcal{N} only, we do not want to make any assumptions about headway requirements. The line-based construction should thus be seen as an idealized case, which has to be adjusted slightly to fit practical purposes. In fact, we will see in Section 5.1 that practical instances have a similar structure, but *do* include headway arcs and not as many transfer arcs.

We will first consider a special case of a line-based event-activity network, namely, one which arises from a star-shaped line network. The *star event-activity network* EAN^* is easy to analyze and—more importantly—any generic line-based event-activity network in the sense of Definition 4.2 may locally be considered as a special case of EAN^* . This will allow us to extend local properties obtained from the star to the generic event-activity network.

Definition 4.4 (Star Event-Activity Network EAN^*) A *star event-activity network of degree d* is a line-based event-activity network EAN^* of a line network associated to $(\mathcal{I}, \mathcal{L})$, where \mathcal{I} is a star graph with central node s and d leaves v_1, \dots, v_d , and \mathcal{L} consists of lines that connect s with either one or two leaves in a way that all leaves are covered by some line and two distinct lines intersect only in s .

Example 4.5 Figure 3 shows a line network \mathcal{N}^* based on a star graph \mathcal{I} with $d = 6$ leaves and a line plan \mathcal{L} of four lines (Fig. 3a), the corresponding line-based event-activity network EAN^* (Fig. 3b), and the labeling v_s and i -th ends ϵ_i^s (Fig. 3c).

Lemma 4.6 (Cyclomatic Number of EAN) The cyclomatic number of a line-based event-activity network EAN based on a line network \mathcal{N} associated to $(\mathcal{I}, \mathcal{L})$ is given by

$$\mu(EAN) = \sum_{s \in S} (d_s - 1)^2 - |S| + 2|\mathcal{L}| + 1,$$

where S denotes the set of stations (i.e., nodes) in \mathcal{N} .

Proof By Definition 4.2, $|\mathcal{V}(EAN)| = \sum_{s \in S} 2d_s$. The combined number of turn and transfer arcs within a station s is $d_s(d_s - 1)$, as we add an arc (A_i^s, D_j^s) for every $i, j \in [d_s]$ with $i \neq j$. For each edge $\{s, t\}^{(\ell)} \in \mathcal{E}(\mathcal{N})$, exactly two driving arcs are added. Lastly, for each line $\ell \in \mathcal{L}$, there are exactly two turn arcs at the terminal stations of ℓ . In total, we have

$$|\mathcal{A}(EAN)| = \sum_{s \in S} d_s(d_s - 1) + 2|\mathcal{E}(\mathcal{N})| + 2|\mathcal{L}|$$

arcs. Using that $\sum_{s \in \mathcal{V}(\mathcal{N})} d_s = 2|\mathcal{E}(\mathcal{N})|$ by the handshaking lemma, the formula for the cyclomatic number (Observation 3.2), and observing that EAN is weakly connected since \mathcal{N} is

$$\begin{aligned}
 \mu(EAN) &= |\mathcal{A}(EAN)| - |\mathcal{V}(EAN)| + 1 \\
 &= \sum_{s \in S} d_s(d_s - 1) + 2|\mathcal{E}(\mathcal{N})| + 2|\mathcal{L}| - 4|\mathcal{E}(\mathcal{N})| + 1 \\
 &= \sum_{s \in S} (d_s - 1)^2 + \sum_{s \in S} (d_s - 1) + 2|\mathcal{L}| - 2|\mathcal{E}(\mathcal{N})| + 1 \\
 &= \sum_{s \in S} (d_s - 1)^2 - |S| + 2|\mathcal{L}| + 1.
 \end{aligned}$$

Let \mathcal{A}_{turn}^s denote the set of turn arcs (A_i^s, D_i^s) at a station $s \in S$.

Corollary 4.7 (Cyclomatic Number of EAN^*) *The cyclomatic number of a star line-based event-activity network EAN^* of degree d with central station s is*

$$\mu(EAN^*) = (d - 1)^2 + |\mathcal{A}_{turn}^s|.$$

Proof Since the underlying line network \mathcal{N}^* has d stations of degree 1 and one (central) station s of degree d , by Lemma 4.6, it suffices to show that $|\mathcal{A}_{turn}^s| = 2|\mathcal{L}| - d$. The number of lines in \mathcal{L} is at most d , and it equals d if and only if all d lines turn at s , so that the formula holds in this case. Whenever two lines with one leaf are joined to a larger line connecting two leaves, $|\mathcal{A}_{turn}^s|$ decreases by two, and $|\mathcal{L}|$ decreases by one, so that the formula remains valid.

4.2 Construction of a Line-Based Forward Cycle Basis

In the following, we will exploit the special structure of line-based event-activity networks, as they allow us to construct fairly intuitive forward cycles guaranteeing an integral forward cycle basis.

Definition 4.8 (q -End Cycles) For $s \in S$ and $q \in [d_s]$, a q -end cycle at s is a forward circuit in EAN^* which uses exactly q distinct ends at s . In particular, using $+$ to denote the concatenation of paths, we define the

- 1-end cycles, denoted by α_i^s for $i \in [d_s]$ such that $(A_i^s, D_i^s) \in \mathcal{A}_{turn}^s$,

$$\alpha_i^s := \epsilon_i^s + (A_i^s, D_i^s),$$

- 2-end cycles, denoted by β_{ij}^s for $i, j \in [d_s], i \neq j$,

$$\beta_{ij}^s := \epsilon_i^s + (A_i^s, D_j^s) + \epsilon_j^s + (A_j, D_i),$$

- 3-end cycles, denoted by γ_{ijk}^s for pairwise distinct $i, j, k \in [d_s]$,

$$\gamma_{ijk}^s := \epsilon_i^s + (A_i^s, D_j^s) + \epsilon_j^s + (A_j^s, D_k^s) + \epsilon_k^s + (A_k, D_i).$$

We denote by C_{123}^s the subspace of $\mathcal{C}(EAN) \otimes_{\mathbb{Z}} \mathbb{Q}$ spanned by the 1-, 2-, and 3-end cycles at station s .

Example 4.9 For an intuitive understanding of the 1, 2, 3-end cycles, consider the star line-based event-activity EAN^* of degree $d = d_s = 3$ in Fig. 4a. Only one line ends at the central node s , as the only turn arc is (A_1^s, D_1^s) . In this case, there is only one 1-end cycle, namely, α_1^s . Furthermore, in this EAN , there are three 2-end cycles— $\beta_{12}^s, \beta_{13}^s$, and β_{23}^s . Lastly, in this small example, γ_{123}^s and γ_{132}^s are the only 3-end cycles in EAN . All 1, 2, 3-end cycles in this graph can be found in Fig. 4b–g. In fact, in this example, the six 1, 2, 3-end cycles are the only simple forward circuits that exist. The only simple forward circuit containing (A_1^s, D_1^s) is α_1^s . As $\mu(EAN^*) = 5$, any forward cycle basis must contain α_1^s and a linearly independent subset of four of the five cycles Fig. 4c–g.

Example 4.10 The instance in Example 4.9 also serves as an example that a star network does not necessarily have a fundamental forward cycle basis: Suppose there is a fundamental forward cycle basis on this network induced by some spanning tree \mathcal{T} . In this case, we have $|\mathcal{A}(\mathcal{T})| = 11$. Any subset of four cycles of $\{\beta_{12}^s, \beta_{13}^s, \beta_{23}^s, \gamma_{123}^s, \gamma_{132}^s\}$ covers the end sections $\epsilon_1^s, \epsilon_2^s$ and ϵ_3^s at least three times. Then, all nine arcs of the three end sections must be part of the spanning tree. Consequently, two more arcs (A_i^s, D_j^s) and (A_k^s, D_l^s) are needed as tree arcs. In particular, we need $i \neq j$ and $k \neq l$ as well as $\{1, 2, 3\} = \{i, j, k, l\}$ for the tree to be connected. Up to symmetry, there are only three cases, all of which are displayed with one representative in Fig. 5: If $i = l$ and $j \neq k$, then the tree is spanning, but the co-tree arc

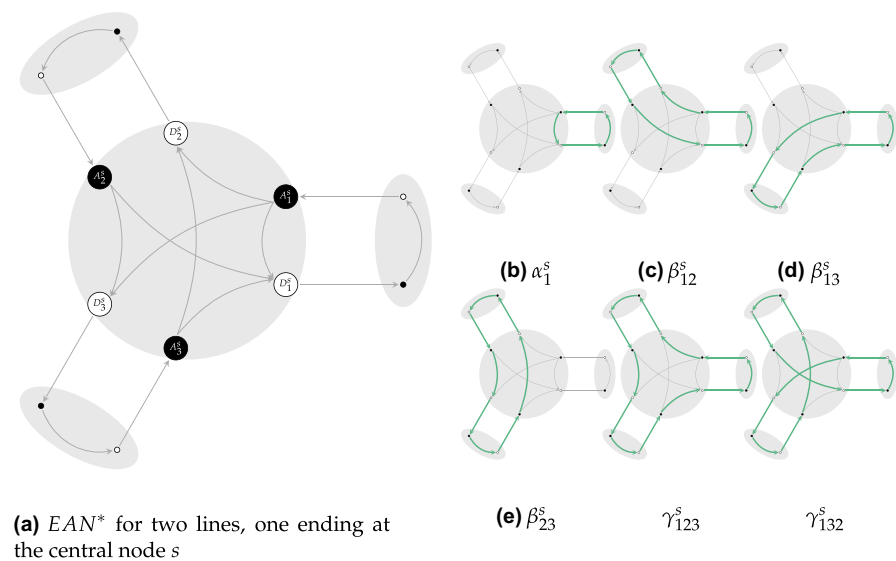


Fig. 4 Example network EAN^* of degree $d = 3$ with all 1, 2, 3-end cycles

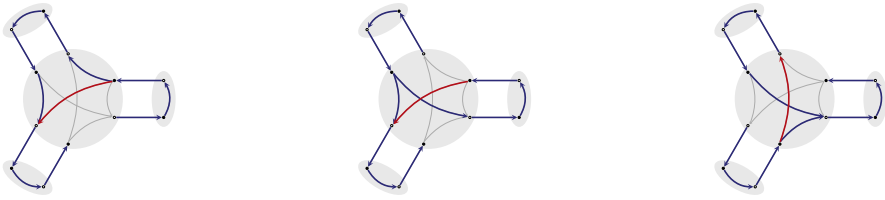


Fig. 5 Explanation of the non-existence of a fundamental forward cycle basis. Blue arcs are necessary tree arcs, and red is the co-tree arc inducing a non-forward cycle

(A_k^s, D_j^s) induces a non-forward cycle. If $i = k$ and $j \neq l$, then the co-tree arc (A_j^s, D_l^s) induces a non-forward cycle. Lastly, if $i \neq k$ and $j = l$, then the co-tree arc (A_i^s, D_k^s) also induces a non-forward cycle. Consequently, there cannot exist a fundamental forward cycle basis.

Lemma 4.11 *For any star event-activity network EAN^* , there exists a weakly fundamental forward cycle basis of $\mathcal{C}(EAN^*)$ using only 1-, 2-, 3-end cycles.*

Proof We give an explicit choice and order of cycles such that each new cycle contains at least one arc not present in the previous cycles, cf. (2). We proceed in three blocks (see Fig. 6 for an example):

- Block 1: $\{\alpha_i^s \mid i \in [d] \text{ and } (A_i^s, D_i^s) \in \mathcal{A}_{turn}^s\}$,
- Block 2: $\{\beta_{1i}^s \mid i \in [d] \setminus \{1\}\}$,
- Block 3: $\{\gamma_{1ij}^s \mid i, j \in [d] \setminus \{1\}, i \neq j\}$.

As the α_i^s cycles are arc-disjoint, they fulfil the property (2) within Block 1. Furthermore, by definition of the 2-end cycles, each β_{1i}^s contains the antiparallel partner arcs (A_1^s, D_i^s) and (A_i^s, D_1^s) . Clearly, these arcs are uniquely in β_{1i}^s in Block 2, so that Block

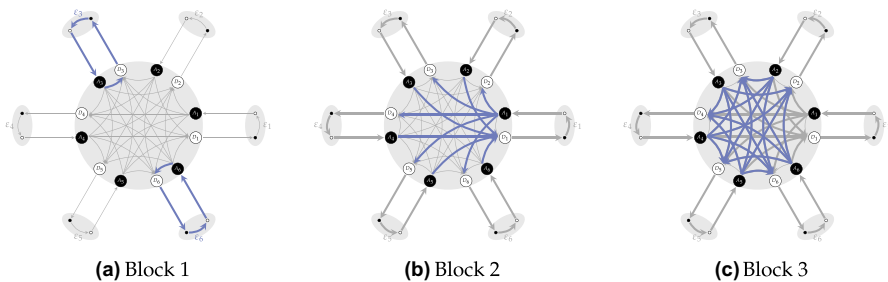


Fig. 6 Construction of the three blocks ($d = 6$). Arcs in block cycles are bold, and new arcs certifying the weak fundamentality (2) are blue

2 satisfies (2) within itself. Since none of the cycles in Block 1 contains any transfer or dwell arcs, neither (A_i^s, D_i^s) nor (A_i^s, D_j^s) are in Block 1 for any $i \in [d]$. Lastly, the arc (A_i^s, D_j^s) is uniquely contained in γ_{1ij}^s and in none of the other cycles of Blocks 1, 2, or 3. As a consequence, Blocks 1, 2, and 3 in this ordering satisfy (2). In total, we have

$$\begin{aligned} |\text{Block 1}| + |\text{Block 2}| + |\text{Block 3}| &= |\mathcal{A}_{turn}^s| + (d - 1) \\ &+ (d - 2)(d - 1) = (d - 1)^2 + |\mathcal{A}_{turn}^s| = \mu(EAN^*) \end{aligned}$$

cycles (Lemma 4.6), all of which are forward, and the given ordering certifies weak fundamentality.

While the star event-activity network EAN^* is a very specific and easy case, it serves as an illustrative basis case to generalize 1, 2, 3-end cycles to arbitrary line-based cycle bases.

Definition 4.12 We define the *space of 1, 2, 3-end cycles of EAN* as the \mathbb{Q} -linear space spanned by all 1, 2, 3-end cycles at each station $s \in S$ and denote it by $\mathcal{C}_{123}(EAN)$.

Note that for the star event-activity network with central station s , we have $\mathcal{C}_{123}(EAN^*) = \mathcal{C}_{123}^s = \mathcal{C}(EAN^*)$. For an outer station t , \mathcal{C}_{123}^t is generated by α_1^t . Such an α_1^t corresponds to a 1-end cycle at the central node s if the line terminates at s , or to a 2-end cycle at s if it passes through the central station. In any case, $\mathcal{C}_{123}^t \subseteq \mathcal{C}_{123}^s$.

Let us emphasize that for any station s , the end section ε_i^s is the entire directed path from the departure event D_i^s to arrival A_i^s along the line $\ell(D_i^s)$. In particular, it is well possible that ε_i^s passes through multiple other stations before returning to s , but it never contains transfer arcs. Consequently, 1-end cycles exist only at terminal stations of lines and correspond exactly to the cycle composed of driving, dwell and turn activities of line $\ell := \ell(D_i^s)$. In particular, we have $\alpha_i^s = \alpha_j^s$ for some $i \in [d_s]$ and $j \in [d_i]$ for the two terminal stations $s \neq t$ of ℓ . For ease of notation, we will denote this cycle by $\alpha(\ell)$.

Lemma 4.13 Consider the subspace $\mathcal{C}_{123}(EAN)$ of the cycle space $\mathcal{C}(EAN) \otimes_{\mathbb{Z}} \mathbb{Q}$ spanned by the 1, 2, 3-end cycles of an arbitrary line-based event-activity network induced by line network \mathcal{N} . Then, the dimension of $\mathcal{C}_{123}(EAN)$ is $\sum_{s \in S} (d_s - 1)^2 - |\mathcal{E}(\mathcal{N})| + 2|\mathcal{L}|$, and there is a weakly fundamental forward cycle basis of $\mathcal{C}_{123}(EAN)$ consisting of 1, 2, 3-end cycles.

Proof Similarly to the proof of Lemma 4.11, we give a specific ordering of a subset of 1, 2, 3-end cycles:

- Block 1: $\{\alpha(\ell) \mid \ell \in \mathcal{L}\}$,
- Block 2: $\{\gamma_{1kl}^s \mid s \in S, 1 < k < l \leq d_s \text{ such that } (A_k^s, D_l^s) \text{ is a dwell arc}\}$,

Block 3: $\{\beta_{1i}^s \mid s \in S, 1 < i \leq d_s \text{ such that } (A_1^s, D_i^s) \text{ is a transfer arc}\}$,

Block 4: $\{\gamma_{1kl}^s \mid s \in S, 1 < k \leq d_s, 1 < l \leq d_s, k \neq l \text{ such that } (A_k^s, D_l^s) \text{ is a transfer arc}\}$.

We now show that this ordering satisfies the weak fundamentality property (5).

Block 1 is clear, as the cycles $\alpha(\ell)$ are pairwise disjoint.

Block 2 contains a 3-end cycle γ_{1kl}^s if (A_k^s, D_l^s) is a dwell arc. Due to Observation 4.3v), the two arcs $(A_1^s, D_k^s) \in \gamma_{1kl}^s$ and $(A_l^s, D_1^s) \in \gamma_{1kl}^s$ must be transfer arcs. For $k \neq k'$ and $l \neq l'$, neither $(A_k^s, D_{l'}^s)$ nor $(A_{k'}^s, D_l^s)$ can be a dwell arc, so that neither $\gamma_{1kl'}$ nor $\gamma_{1k'l}$ is contained in Block 2. Consequently, the only cycle in Block 2 covering the two transfer arcs (A_1^s, D_k^s) and (A_l^s, D_1^s) is γ_{1kl} . One can observe by the same argumentation that no cycle in Block 2 contains one of their antiparallel partners (A_k^s, D_1^s) and (A_1^s, D_l^s) , since $k < l$, which is a property which will be needed for Block 3. As the cycles in Block 1 do not contain any transfer arcs, Blocks 1 and 2 are (even strictly) fundamental in the sense of Definition 3.14.

Block 3 is clearly fundamental with respect to itself. Each cycle β_{1i}^s in Block 3 contains exactly two transfer arcs, namely, (A_1^s, D_i^s) and (A_i^s, D_1^s) . By the previous argument, at least one of these antiparallel partner arcs is not covered by any cycle in Block 2, and certainly not by Block 1. Blocks 1, 2, and 3 are hence weakly fundamental.

So far, no cycle contains a transfer arc (A_k^s, D_l^s) for $k \neq l$ with $k, l > 1$. We conclude that Blocks 1, 2, 3, and 4 provide a weak fundamental ordering in the sense of (5). Note that arcs of the form (A_1^s, D_i^s) and (A_i^s, D_1^s) are part of Block 3 and Block 4 cycles, so that the ordering cannot be assumed to be strictly fundamental.

Having established weak fundamentality and hence linear independence, we observe that for a station s and pairwise distinct $i, j, k \in [d_s] \setminus \{1\}$ holds

$$\beta_{ij}^s = \gamma_{1ij}^s + \gamma_{1ji}^s - \beta_{1i}^s - \beta_{1j}^s, \tag{7}$$

$$\beta_{ji}^s = \beta_{ij}^s, \tag{8}$$

$$\gamma_{ijk}^s = \gamma_{1ij}^s + \gamma_{1jk}^s + \gamma_{1ki}^s - \beta_{1i}^s - \beta_{1j}^s - \beta_{1k}^s, \tag{9}$$

$$\gamma_{ijk}^s = \gamma_{jki}^s = \gamma_{kij}^s. \tag{10}$$

Note that all cycles on the right-hand side are in the span of the cycles of Blocks 1–4. If β_{1i}^s was omitted in Block 3, then (A_1^s, D_i^s) is a dwell arc, so that β_{1i}^s is in fact $\alpha(\ell(D_i^s))$ in Block 1. Similarly, it is possible that γ_{1kl}^s with $k > l$ is neither in Block 2 nor Block 4. Then, (A_k^s, D_l^s) is a dwell arc of line $\ell(D_l^s)$, and $\beta_{kl} = \alpha(\ell(D_l^s))$, so that γ_{1kl}^s is in the span of Blocks 1–4 by relation (7).

In particular, the cycles of Blocks 1–4 generate $C_{123}(EAN)$.

It remains to compute the number of cycles in Blocks 1–4: Block 1 contains $|\mathcal{L}|$ cycles. For Blocks 2–4, we can simply count the number of all 2-end β_{1i}^s and all 3-end cycles γ_{1kl}^s at each station and then subtract the number of omitted cycles. At each of the non-terminal stations s , the cycle $\alpha(\ell)$ passes through dwelling arcs. If ℓ connects the 1st end with the i -th end, then β_{1i}^s in Block 3 is already given by $\alpha(\ell)$. If

it connects ends k and l , say $k < l$, then we need to count γ_{1kl} , but omit γ_{1lk}^s . We thus need to drop exactly one cycle for each of the $|\mathcal{V}(\ell)| - 2 = |\mathcal{E}(\ell)| - 1$ non-terminal stations of each line. The total number of cycles in Blocks 1–4 is therefore

$$\begin{aligned} & \underbrace{|\mathcal{L}|}_{\text{Block 1}} + \underbrace{\sum_{s \in S} (d_s - 1)}_{\text{2-end cycles } \beta_{1l}, 1 < i} + \underbrace{\sum_{s \in S} (d_s - 1)(d_s - 2)}_{\text{3-end cycles } \gamma_{1kl}, 1 < k < l} - \underbrace{\sum_{\ell \in \mathcal{L}} (|\mathcal{E}(\ell)| - 1)}_{\text{omitted cycles}} \\ &= \sum_{s \in S} (d_s - 1)^2 - \sum_{\ell \in \mathcal{L}} |\mathcal{E}(\ell)| + 2|\mathcal{L}| \\ &= \sum_{s \in S} (d_s - 1)^2 - |\mathcal{E}(\mathcal{N})| + 2|\mathcal{L}|. \end{aligned}$$

Consider a line-based event activity network EAN of \mathcal{N} , and let D be an arbitrary orientation of \mathcal{N} so that $G(D) = \mathcal{N}$. We define a \mathbb{Z} -linear map $\rho : \mathcal{C}(EAN) \rightarrow \mathcal{C}(D)$ via

$$\rho(\gamma) = (\rho(\gamma)_a)_{a=(s,t) \in \mathcal{A}(D)}, \quad \rho(\gamma)_{(s,t) \in \mathcal{E}} := \gamma_{(D_i^s, A_j^t)} - \gamma_{(D_j^s, A_i^t)},$$

where $a = (s, t) \in \mathcal{A}(D)$ corresponds with $e(a) = \{s, t\} \in \mathcal{E}(\mathcal{N})$, and $((D_i^s, A_j^t), (D_j^s, A_i^t))$ is the antiparallel pair of driving arcs of line $\ell = \ell(D_i^s)$.

The map ρ is well-defined: Let $\gamma \in \mathcal{C}(EAN)$ and $s \in S = \mathcal{V}(D)$. For $i \in [d_s]$, let a_i^+ denote the unique driving arc from D_i^s , and let a_i^- denote the unique driving arc to A_i^s . Then $W := \{A_i^s \mid i \in [d_s]\} \cup \{D_i^s \mid i \in [d_s]\}$, the set of all events at s , defines a cut in EAN such that $\delta^+(W) = \{a_1^+, \dots, a_{d_s}^+\}$ and $\delta^-(W) = \{a_1^-, \dots, a_{d_s}^-\}$. By definition of ρ and since γ is a cycle,

$$\sum_{a \in \delta^+(s)} \rho(\gamma)_a - \sum_{a \in \delta^-(s)} \rho(\gamma)_a = \sum_{i=1}^{d_s} \gamma_{a_i^+} - \sum_{i=1}^{d_s} \gamma_{a_i^-} = \sum_{a \in \delta^+(W)} \gamma_a - \sum_{a \in \delta^-(W)} \gamma_a = 0.$$

Consequently, $\rho(\gamma)$ is indeed an element of $\mathcal{C}(D)$.

Now, let $c \in \mathcal{C}(D)$ be a forward circuit with arc sequence $((s_1, s_2)^{\ell_1}, \dots, (s_k, s_1)^{\ell_k})$. We can lift c to the forward circuit $\lambda(c)$ with node sequence

$$(D_{i_1}^{s_1}, A_{j_1}^{s_2}, D_{i_2}^{s_2}, \dots, A_{j_{k-1}}^{s_k}, D_{i_k}^{s_k}, A_{j_k}^{s_1}, D_{i_1}^{s_1}),$$

where for $l \in [k]$, i_l , and j_l are chosen such that $(D_{i_l}^{s_l}, A_{j_l}^{s_{l+1}})$ is the unique driving arc of line ℓ_l from station s_l to s_{l+1} , with the convention that $s_{k+1} := s_1$. In other words, $\lambda(c)$ is simply the forward cycle in EAN obtained by taking all driving arcs in the direction as indicated by c , and then connecting two consecutive driving arcs by the suitable transfer or dwell arc.

Lemma 4.14 *Let EAN be a line-based event-activity network based on a line network \mathcal{N} , and let \mathcal{B}_{123} be an integral forward cycle basis of the subspace $\mathcal{C}_{123}(EAN)$. Let D be an orientation of \mathcal{N} such that there is an integral forward cycle basis \mathcal{B}_D of D . Then, \mathcal{B}_{123} and the lifts via λ of the cycles in \mathcal{B}_D constitute an integral forward cycle basis of EAN .*

Proof Every cycle $c \in \mathcal{B}_D$ is a forward circuit and can hence be lifted to a forward circuit $\lambda(c) \in \mathcal{C}(EAN)$. Since \mathcal{B}_D is an integral basis of $\mathcal{C}(D)$, λ extends to a \mathbb{Z} -linear map $\lambda : \mathcal{C}(D) \rightarrow \mathcal{C}(EAN)$. Observe that for any forward circuit, $c \in \mathcal{C}(D)$ holds $\rho(\lambda(c)) = c$. We deduce that $\rho \circ \lambda$ is the identity map, so that $\mathcal{C}(EAN)$ decomposes as the direct sum of two free abelian groups: The image of λ and the kernel of ρ . To construct an integral forward cycle basis of $\mathcal{C}(EAN)$, it hence suffices to construct integral forward cycle bases of each summand. Since λ is injective, $\{\lambda(c) \mid c \in \mathcal{B}_D\}$ is a basis for the image of λ as an abelian group. We will now show that $\ker(\rho) \otimes_{\mathbb{Z}} \mathbb{Q} = \mathcal{C}_{123}(EAN)$. Since \mathcal{B}_{123} is integral, this implies that \mathcal{B}_{123} is a basis of $\ker(\rho)$ as an abelian group.

By construction of the 1, 2, 3-end cycles, if any $\gamma \in \mathcal{B}_{123}$ uses a driving arc (D_i^s, A_j^t) , also its antiparallel counterpart (D_j^t, A_i^s) must be used. In particular, $\mathcal{C}_{123}(EAN) \subseteq \ker(\rho) \otimes_{\mathbb{Z}} \mathbb{Q}$. Using the dimension formulas of Lemmas 4.6 and 4.13,

$$\begin{aligned} \dim(\ker(\rho) \otimes_{\mathbb{Z}} \mathbb{Q}) &= \mu(EAN) - \mu(D) \\ &= \sum_{s \in S} (d_s - 1)^2 - |\mathcal{V}(\mathcal{N})| + 2|\mathcal{L}| + 1 - |\mathcal{E}(\mathcal{N})| + |\mathcal{V}(\mathcal{N})| - 1 \\ &= \sum_{s \in S} (d_s - 1)^2 - |\mathcal{E}(\mathcal{N})| + 2|\mathcal{L}| \\ &= \dim(\mathcal{C}_{123}(EAN)), \end{aligned}$$

and we conclude $\mathcal{C}_{123}(EAN) = \ker(\rho) \otimes_{\mathbb{Z}} \mathbb{Q}$.

The previous considerations give a constructive proof of the following result:

Theorem 4.15 *Every line-based event-activity network has an integral forward cycle basis.*

Proof The proof of Lemma 4.13 describes an explicit construction of a weakly fundamental and hence integral forward cycle basis of $\mathcal{C}_{123}(EAN)$. By Lemma 3.13, \mathcal{N} can always be oriented in a way that a fundamental and hence integral forward cycle basis exists. It remains to apply Lemma 4.14.

5 Application to Periodic Timetabling

Lindner et al. performed preliminary tests which indicate that forward bases provide good dual bounds—on a single base instance [12]. We now want to examine whether this is the case for more instances. We therefore choose a two-stage approach: First of all, in the spirit of Lindner et al. [12], we extend our network analysis to all 16 railway instances of PESPLIB and detect structural patterns, which allow us to identify lines. Adding turn arcs at the terminal stations of the

lines, the networks become strongly connected, so that forward cycle bases exist by Theorem 3.11. Secondly, we want to investigate whether forward cycle bases have an advantage in comparison to arbitrary integral cycle bases: For each of those 16 instances, we compare the performance of the dual bounds of three different forward bases, namely, minimum forward span, forward bottleneck, and a modification of the 1, 2, 3-end cycles, to a non-forward cycle basis. Since Borndörfer et al. “empirically observed stronger dual bounds” [23, p. 13] for the minimum span cycle basis, this is also our choice for the non-forward case.

5.1 Structural Analysis of PESPLib Instances

The railway instances of PESPLib [5] have been analyzed in the past. Goerigk and Liebchen do so for the purpose of preprocessing and reducing the model size [24]. As a side product, they also interpret some of the arcs, e.g., they assume that *free arcs* are “waiting times (e.g., of passengers at transfers, of trains during turnarounds, of both during stops)” [24, p. 5]. Our aim now is to better understand these instances and to interpret them from a practical perspective. Lindner et al. performed such a structural analysis for the smallest instance already [12], and we want to extend this approach to all railway instances of PESPLib, with the goal of identifying activity types and line-structures.

Each of the instances is a standard PESP instance $I = (EAN, T, \ell, u, w)$ with period time $T = 60$. It turns out that the 16 instances have a similar structure. As Lindner et al. do for the smallest instance [12, Section 4], we can partition the set $\mathcal{A}(EAN)$ of arcs of any of the instances into four categories:

- $\mathcal{A}_{head} := \{a \in \mathcal{A}(EAN) \mid \ell_a = u_a = 0\}$
- $\mathcal{A}_{trans} := \{a \in \mathcal{A}(EAN) \mid \ell_a = 3, u_a = 62\}$
- $\mathcal{A}_{dwell} := \{a \in \mathcal{A}(EAN) \mid \ell_a = 1, u_a = 5\}$
- $\mathcal{A}_{drive} := \{a \in \mathcal{A}(EAN)\} \setminus (\mathcal{A}_{head} \cup \mathcal{A}_{trans} \cup \mathcal{A}_{dwell})$

When removing all $a \in \mathcal{A}_{head}$ and all $a \in \mathcal{A}_{trans}$ from $\mathcal{A}(I)$, the remaining graph still contains all vertices and decomposes into $2k$ components for some $k \in \mathbb{N}$, each of which is a simple directed path. Moreover, each of the paths p consists of a sequence of nodes (v_0, \dots, v_m) for some odd m such that $(v_{i-1}, v_i) \in \mathcal{A}_{dwell}$ for even i and $(v_{i-1}, v_i) \in \mathcal{A}_{drive}$ for odd i and $i \in [m]$. Lastly, it is possible to match each path with another path of the same length, which has the same arc bounds in reverse order, i.e., for p , there exists another path p' with $p' = (v'_0, \dots, v'_m)$ such that $\ell_{v_{i-1}v_i} = \ell_{v'_{m-i}v'_{m-i+1}}$ and $u_{v_{i-1}v_i} = u_{v'_{m-i}v'_{m-i+1}}$. Let \overline{PP} denote the set of matched pairs of paths.

This allows us to reinterpret each instance. The set \mathcal{A}_{drive} correspond to driving arcs and \mathcal{A}_{dwell} to dwell arcs. Each pair of matched simple paths $(p, p') \in \overline{PP}$ can then be interpreted as the alternating sequence of drive and dwell arcs per direction of a bidirectional line. Furthermore, for $p = (v_0, \dots, v_m)$ as above, we call the nodes with even index the departure nodes and those with odd index arrival nodes. From a modeling perspective, this identification is reasonable: Assuming that the period time $T = 60$ is measured in minutes, vehicle dwell times in stations of 1 to 5 min

seem realistic. Furthermore, the headway arcs \mathcal{A}_{head} are only between arrival nodes, and since their lower and upper bounds are both set to zero, they ensure a simultaneous arrival of vehicles. Please note PESPLIB should be regarded as a collection practical instances. Contrary to our somewhat idealized construction of line-based event-activity networks, they do include (what we identify as) headway arcs.

In this view, we consider the set \mathcal{A}_{trans} as the transfer arcs from one line to another, as each of the arcs link an arrival node of one line with a departure node of a different line with a minimum transfer time of 3 minutes. Observe that $u_a = 62 = \ell_a + T - 1$ for $a \in \mathcal{A}_{trans}$, which means that any integer periodic tension on the subset $\mathcal{A}_{head} \cup \mathcal{A}_{dwell} \cup \mathcal{A}_{drive}$ can be extended to \mathcal{A}_{trans} as well.

In the spirit of the line-based event activity network (Section 4.1) and since vehicle routes should be forward cycles in a periodic setting, we propose the following extension of the instances: For each pair of matched simple paths $(p, p') = ((v_0, \dots, v_m), (v'_0, \dots, v'_m))$ associated to the forward and backward journey of a line, we add two turn arcs $(v_m, v'_0), (v'_m, v_0)$ to model direction change of vehicles. The turn arcs in combination with the arcs of p and p' form a simple circuit, which we interpret as an $\alpha(l)$ -cycle as in Section 4.2, i.e., the activity sequence of a single line l .

The arc categorization and the set of $\alpha(l)$ -cycles lets us deduce the underlying line network \mathcal{N} as well, such that we now go beyond the identification similar to Lindner et al. [12]. We assume that transferring and waiting is possible within a station only and that each line travels bidirectionally, i.e., it visits the same stations in both directions. Correspondingly, we assign each event u to a station $s(u) \in S = \mathcal{V}(\mathcal{N})$ and associate multiple events with the same station, $s(u) = s(v)$ if $(u, v) \in \mathcal{A}_{trans}$ or if $(u, v) \in \mathcal{A}_{dwell}$ or if there exists a $(p, p') \in PP$ with $v_i \in p$ and $v'_{m-i} \in p'$ for $i \in \{0, \dots, m\}$. It turns out that with this station assignment, all driving activities $(u, v) \in \mathcal{A}_{drive}$ are between distinct stations, i.e., $s(u) \neq s(v)$. This allows for an explicit line identification as the undirected path along visited stations in the spirit of Section 4.1:

$$\mathcal{L} := \{ (s(v_0), s(v_2), \dots, s(v_{m-1}), s(v_m)) \mid (p, p') = ((v_0, \dots, v_m), (v'_0, \dots, v'_m)) \in \overline{PP} \},$$

i.e., the sequence of stations of all departure events along p in addition to the station of the last arrival event v_m for each matched pair of paths.

As mentioned, we extend the PESPLIB instances by some artificial edges \mathcal{A}_{art} , to better fit with our concept of the line-based event-activity network. The overall goal remains to solve the original instances, so any additional free arc should not inhibit a feasible solution on the original. For all $a \in \mathcal{A}_{art}$ we therefore set

$$\ell_a := 0, \quad u_a := 59, \quad w_a := 0.$$

The additional artificial edges consist mostly of the turning arcs \mathcal{A}_{turn} as induced by the matched simple path pairs $(p, p') \in PP$. Two instances, namely R2L4 and R3L4 had to be extended by two and a single artificial transfer arc, respectively, in order to make the instances strongly connected.

Note that we do not extend the original instances to include all possible transfer arcs. The resulting event-activity networks (excluding the headways) are thus only sub-graphs of a full line-based event activity network as defined in Definition 4.2. An extension to the

full line-based event activity network seems unreasonable in this case, as this would mean adding more artificial transfer arcs than the total number of arcs in the original instance and would increase the cyclomatic number by a factor of 4.4–11.4. As a consequence, not all 1,2 and 3-end cycles exist, so that a one-to-one application from the construction of strategic forward cycles is not entirely possible. We will address this issue in Section 5.2.

As a side note we should mention that this analysis and identification does not work in the same manner for the other four PESPLIB instances, which are considered bus timetabling instances [5]: The partition into categories is not as clear-cut as for the railway instances, nor can one easily identify forward and backward directions of the vehicles' activity sequences. As busses are not restricted to tracks and empty rides to starting or from end points are less of an issue from a practical point of view, we suspect that such activities are not included in the data set, which makes reverse-engineering to obtain the underlying network significantly harder.

For the railway instances, Table 1 gives an overview of the key values of our identified line plan \mathcal{L} , the corresponding line network \mathcal{N} and the event-activity network EAN as obtained by extending the PESPLIB instances as described.

5.2 Cycle Basis Selection

As discussed in Example 3.9, forward cycles seem particularly promising for an increased lower bound, and the structure of Odijk's cycle inequalities (4) suggests that cycles with low span or heavy minimum weight lead to better dual bounds. To evaluate if this holds true in practice, let us first formally define the extremal cycle basis problem.

Table 1 Overview of the (extended) PESPLIB instances and the underlying network as a result of the structural analysis

Instance	$ \mathcal{L} $	$ \mathcal{V}(\mathcal{N}) $	$ \mathcal{E}(\mathcal{N}) $	$\mu(\mathcal{N})$	$ \mathcal{V}(EAN) $	$ \mathcal{A}(EAN) $	$\mu(EAN)$	$ \mathcal{A}_{arr} $
R1L1	55	522	916	397	3664	6495	2832	110
R1L2	54	520	917	398	3668	6651	2984	108
R1L3	65	605	1046	444	4184	7161	2978	130
R1L4	71	646	1190	545	4760	8670	3911	142
R2L1	66	596	1039	446	4156	7493	3338	132
R2L2	66	592	1051	461	4204	7695	3492	132
R2L3	80	750	1262	513	5048	8446	3399	160
R2L4	116	1112	1915	806	7660	13407	5748	234
R3L1	73	614	1129	518	4516	9291	4776	146
R3L2	70	598	1113	517	4452	9391	4940	140
R3L3	95	755	1431	679	5724	11359	5636	190
R3L4	120	1122	2045	925	8180	15898	7719	241
R4L1	86	602	1233	632	4932	10434	5503	172
R4L2	89	609	1262	655	5048	10913	5866	178
R4L3	115	734	1592	860	6368	13468	7101	230
R4L4	133	1019	2096	1078	8384	18020	9637	266

Definition 5.1 (Extremal Cycle Basis Problem) Let \mathbf{B} be a set of cycle bases of a digraph D with a weight function $c : \mathcal{C}(D) \rightarrow \mathbb{R}_{\geq 0}$. The minimum weight cycle basis problem is to find a cycle basis $\mathcal{B} \in \mathbf{B}$ such that

$$\sum_{\gamma \in \mathcal{B}} c(\gamma) = \min_{\mathcal{B}' \in \mathbf{B}} \sum_{\gamma \in \mathcal{B}'} c(\gamma).$$

Analogously, the maximum weight cycle basis problem is to find a basis with maximum weight.

In our computations, we thus compare three different extremal cycle bases to one based on the 1, 2, 3-end cycles to assess the impact of the concrete basis choice. More precisely, we want to compute a

- B1) minimum span, where \mathbf{B} is the set of all (not necessarily forward) integral cycle bases on EAN , and the weight is given by $c(\gamma) := \sum_{a \in \gamma} (u_a - \ell_a)$ for $\gamma \in \mathcal{C}(EAN)$,
- B2) minimum forward span, where \mathbf{B} is the set of all integral forward cycle bases on EAN , and $c(\gamma) := \sum_{a \in \gamma} (u_a - \ell_a)$,
- B3) maximum forward bottleneck, where \mathbf{B} is the set of all integral forward cycle bases on EAN , and $c(\gamma) := \min_{a \in \gamma} w_a$

cycle basis.

As no practical algorithm to compute extremal integral cycle bases is available, we relax \mathbf{B} to the potentially larger the set of (forward) undirected cycle bases. Then, (E, I) forms a vector matroid, with E as the ground set of circuits in D and $I \subseteq \mathcal{P}(E)$ as the set of (\mathbb{F}_2) -linearly independent sets of circuits in D , as was shown by Horton [25]. Also when restricting both the ground set and the independent set to forward cycles, (E, I) remains a vector matroid, as the independence properties are inherited from the underlying vector space.

A greedy approach will thus result in an optimal solution to the extremal cycle basis problem when applied to all circuits. Clearly, an enumeration of all circuits for graphs of the size of the PESPLIB instances is not feasible in practice. Instead, we use (variations) of Horton’s algorithm [25] to obtain high quality cycle bases.

Definition 5.2 ((Forward) Horton Set) Let D be a digraph, and let c be a non-negative weight function on paths in D . For each node $z \in \mathcal{V}(D)$, fix a shortest path tree T_z^+ with source z w.r.t. c and a shortest path tree T_z^- with target z w.r.t. c . A *Horton circuit* is a circuit of the form $p_{zu} + (u, v) + q_{vz}$ such that

- $(u, v) \in \mathcal{A}(D)$,
- p_{zu} is the unique z - u path in T_z^+ ,
- q_{vz} is the unique v - z path in T_z^- ,
- p_{zu} and q_{vz} are arc-disjoint.

The *Horton Set* \mathcal{H} consists of all Horton cycles.

In our setting,

- B1) $T_z^+ = T_z^-$ is a subgraph of D projecting to a shortest path tree rooted at z w.r.t. edge weights $u_a - \ell_a$ on the undirected graph $G(D)$,
- B2) T_z^+ is a shortest path tree with z as source w.r.t arc weights $u_a - \ell_a$, and similarly, T_z^- is a shortest path tree with z as target,
- B3) T_z^+ is a widest path tree with z as the source, i.e., any z - u path p_{zu} on T_z^+ is forward and maximum with respect to the bottleneck costs $c(p) = \min_{a \in A(p)} w_a$ among all z - u -paths p . Analogously, T_z^- is a widest path tree with z as target.

For B1, a Horton cycle can then have both forward and backward arcs, whereas the Horton cycles for B2 and B3 are forward.

The Horton algorithm can be summarized as computing the Horton set \mathcal{H} and subsequently extracting a set of linearly independent cycles from it by a greedy procedure. For a general weight function and optimization sense, this approach is clearly just a heuristic. In the case of B1 and B2, however, the Horton algorithm guarantees an optimal solution: The B1 case follows from the classical result as described by Horton [25]. The case of B2 is described with different terminology by Gleiss et al. and follows from Corollary 16, Lemma 18, and Theorem 19 [18].

There are minimum weight cycle basis algorithms that asymptotically perform better than that of Horton, such as de Pina's [26], which was improved upon, e.g., by Hariharan et al. [27] or Amaldi et al. [28]. However, it is not as straightforward to adapt the algorithms to work for forward cycle bases only, and the performance of Horton's algorithm turned out to be sufficient in our case.

Finally, we want to compare the performance of the 1, 2, 3-end forward cycles to the introduced extremal cycle bases. As mentioned in the previous section, the instances are not complete, in the sense that not all possible transfer arcs per station exist. We therefore propose the following approach: At each station, we compute all possible 1, 2, 3-end cycles. Due to the sparse amount of transfer arcs, we generate few more q -end cycles to cover all forward cycles between two lines meeting at a station. We then construct our basis by iterating over the cycles and add a cycle whenever linear independence persists. Our ordering of the cycles is dependent on the number of used transfer arcs, such that cycles with fewer transfers are prioritized. However, there is still a large degree of freedom—it could be worth exploring whether short spans or heavy weight q -ended cycles should be preferred. If this procedure stops before the full cycle space is generated, we complete the basis with cycles from a minimum forward span basis (B2). In the following, this extended 1, 2, 3-end cycle basis will be abbreviated by B4.

We will briefly discuss our decision to restrict our experiments to these four bases. Our primary goal is to evaluate whether forward bases are advantageous in comparison to non-forward ones, but also secondarily, whether a formulation tailored specifically for Odijk's cycle inequalities (4) has a significant impact. Consequently, we make the choice of comparing one structural (B4) and two optimized (B2 and B3) forward cycle bases to a reasonable choice of non-forward cycle basis (B1). Liebchen et al. observed a better performance for strictly fundamental cycle

bases in comparison to minimum span [29]. Our decision to refrain from using a fundamental cycle basis as reference is due to three factors: The key point is that it allows a cleaner comparison: A minimum span cycle basis as obtained by a Horton set should have a similar structure as its forward counterpart, such that their key difference is w.r.t. orientation. If we compared a minimum span forward cycle basis to a fundamental non-forward one, it would be hard to extrapolate whether the forward or the structural properties were the influencing factors. Secondly, the existence of fundamental forward cycle basis is not guaranteed even if the event-activity network is strongly connected (see Example 3.12). Thirdly, tests by Borndörfer et al. suggest that a minimum span performs better in practice than a fundamental cycle basis [23].

5.3 Computational Results

In our computational experiments, we computed each of the four bases (B1 – B4) for each of the 16 instances described in Section 5.1.¹ Note that for timetabling, integral cycle basis is required, which is not guaranteed by our approach for computing B1 – B3. We can, however, check whether a cycle basis is integral simply by computing the determinant of the cycle basis [13]. It turns out that the computed matrices have determinant ± 1 , which ensures that the choices of bases are indeed integral (cf. Kavitha et al. [13]).

In the framework of the PESP-specialized solver `ConcurrentPESP` [23] invoking CPLEX 12.10 [30], we optimize the dual bounds for each instance-basis pair on an Intel i7-9700K CPU machine with 64 GB RAM with a wall time limit of 6 h each. `ConcurrentPESP` makes use of Odijk's cycle inequalities and more sophisticated cutting planes, e.g., flip inequalities [9].

Note that similar tests for the R1L1 instance already suggested that forward bases provide good dual bounds [12]. Our current results underline this claim:

First, let us consider the final results at the end of the runs. In Fig. 7, one can observe the relative improvement at the end of the 6 h run for each of the bases in comparison to the previously best dual bound. What becomes obvious is that generally, the forward cycles (B2–B4) are significantly better than the non-forward ones (B1). With the exception of R3L4, where B3 provides the worst dual bound, B1 always comes in last. We observe a trend in the rankings: The minimum forward span basis (B2) is the clear best choice, as it provides the best dual bound in 11 out of the 16 instances. In contrast, the 1, 2, 3-end cycle basis (B4) is only 3 times in the first place, closely followed by the maximum bottleneck (B3) with 2 best dual bounds. When comparing the quality of the improvement, this ranking is still evident, but not as prominent. When looking at Fig. 7, the the red line of (B4) is often quite close to B2's red line. On average,² within the 6 h runs, B2 could improve the dual bound by 4.56% and B4 by 4.04%. In contrast, the average improvement of B3 is only by 2.3%, the non-forward basis B1 reached

¹ The only exception here is the instance R4L4—here, we were unable to compute a large bottleneck cycle basis, as its Horton set alone surpassed a memory usage of 120 GB. The relatively poor performance of B3 on the other instances did not justify further time and resource consuming computations for this single instance.

² We exclude the results of R4L4 here, as we want a comparison of all four basis choices.

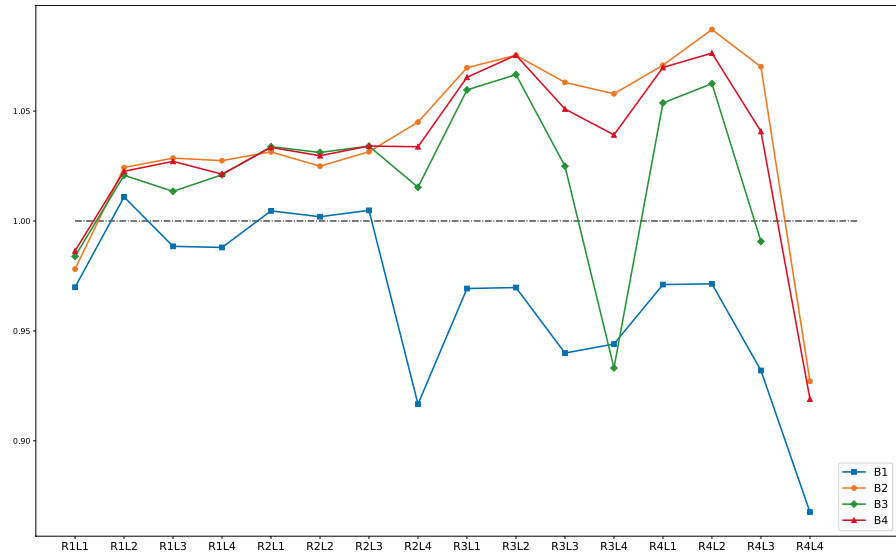


Fig. 7 Relative improvement of dual bounds in comparison to the previous best bound for each of the 16 railway PES_{lib} instances

solutions which are 2.78% worse than the previous best solution. Note that these results need to be taken with a grain of salt, as we compare the new dual bounds to previously computed ones, meaning that it highly depends on how much the instances have been treated in the past. This is evident when looking at the instances R1L1 and R4L4: These are the only two instances without improvement. We cannot hope to contend with the previous dual bounds of R1L1, which have been treated with forward cycle bases for 24 instead of 6 h of wall time [12]. An overview over the new bounds obtained by our

Table 2 New dual bounds

Instance	Dual bound	Provided by basis
R1L2	19886799	B2
R1L3	19323821	B2
R1L4	17283850	B2
R2L1	25929643	B3
R2L2	25642692	B3
R2L3	23941492	B4
R2L4	19793447	B2
R3L1	26825864	B2
R3L2	27178406	B4
R3L3	23007043	B2
R3L4	17432725	B2
R4L1	29174444	B2
R4L2	28664399	B2
R4L3	24293621	B2

computations can be found in Table 2. On average, we were able to reduce the relative gap from 42.35 to 39.84%.

We also compare the progression of the bounds over time (see Fig. 8). The minimum forward span basis B2 tends to have the higher quality bound earlier than the other bases, followed by the B4 basis. What is surprising is that the B3 basis, which we expected to perform particularly well, does not provide the best bounds as fast. Particularly in light of our motivating example, the superiority of the minimum span (B2) over maximum bottleneck (B3) is noteworthy: Even though the basis B2 was chosen regardless of its arc weights, it seems to outperform B3. One explanation could be that Odjik’s cycle inequalities should be “stricter” and thus provide stronger cuts for low-span in comparison to heavy weight cycles. It is possible that the cuts from the heavy weight cycles increase the tension along a cycle only marginally, which then—despite the large weight—does not contribute significantly to the objective. Here, the quality of the cuts seems to outweigh the heavy weight of

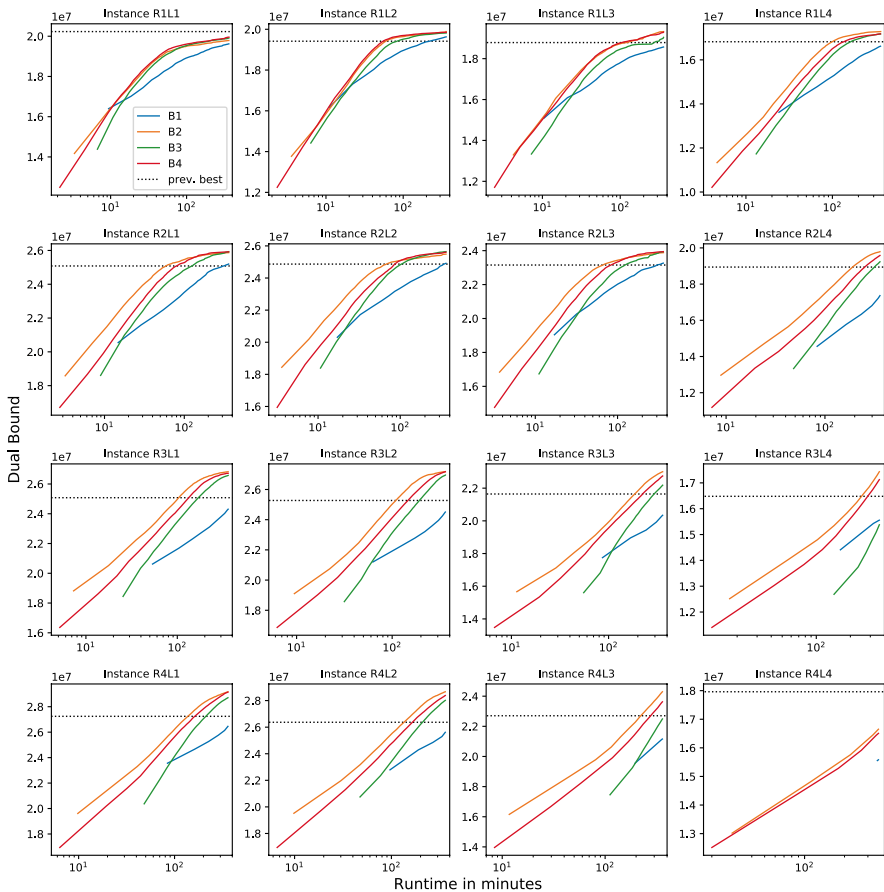


Fig. 8 Progression of dual bounds over time

the cycles in B3. In any case, the monotonicity property of adding cuts arising from forward cycles seems to have a significant impact on the quality of the dual bounds.

The slope of the progression of the forward bases is significantly steeper than that of B1. In contrast, the slope of the forward bases seems to be comparable to each other. Furthermore, we can observe that after a certain time, when the dual bounds begin to stagnate, there is not a clear winner anymore. This can be observed in the smaller instances, where the lines of the forward bases begin to be entangled. This is particularly surprising, since one could expect B4 to perform poorly in comparison to any other optimized forward cycle basis. Recall that 1, 2, 3-end cycles are constructed only by arc type without considering arc weights or spans. In contrast, both B2- and B3-bases are optimized from the beginning. However, in the bound behavior, this does not seem to play such a prominent role.

We conclude that in practice, forward cycle bases are indeed a better choice than non-forward bases when optimizing dual bounds. Which explicit forward basis should be chosen is not as clear, as each of them provided an improvement of some dual bounds. For short runs, the minimum forward span basis B2 seems like the best choice, as it quickly finds high quality bounds. In the long run, however, the other two basis choices can compete with B2. Since the computation of B2 and B3 can be quite resource consuming, it might be sufficient to use a strategic construction of a forward cycle basis if structural properties of the graph are available.

Nevertheless, while the improvements due to forward cycle bases are significant, it cannot be concealed that the dual gap still remains very large. It is clear that further research is needed for the dual, but also the primal side.

6 Conclusion and Outlook

Our main contributions are on one side on the extension of the theory of cycle bases: We apply standard notions, such as (weak) fundamentality, and integrality, to forward cycle bases and discuss existence properties. Furthermore, we consider bases of cycle subspaces and extend the hierarchy of classes of cycle basis to this setting

On the other hand, we highlight the importance of forward cycle bases in the context of periodic timetabling for transportation networks: Firstly, our computational experiments provide better lower bounds for almost all railway instances of the benchmark library `PESPLib`. Secondly, the results suggest that forward cycle bases are clearly preferable to arbitrary cycle bases. However, in general networks, an integral forward cycle basis, as is needed in the context of PESP, might not always be available. We therefore introduced a construction of a line-based event-activity network, which may be used for practical modeling applications. For such a generic network, not only can we guarantee the existence of an integral forward cycle basis, but also describe an explicit construction procedure.

This directly opens the door for future work: Is there some characteristic property of a graph which ensures the existence of a forward cycle basis which is also (weakly) fundamental, integral, and undirected? What is the complexity of deciding whether a certain type of forward cycle basis exists? How hard is it to find an extremal cycle basis? Another question is whether the concept of cycle bases of subspaces can be utilized for a primal heuristic for PESP, e.g., with a divide and conquer approach of solving PESP on subspaces.

As we highly recommend to use forward cycle basis formulations for PESP, one could ask the question of how to deal with networks that do *not* contain a forward cycle basis. For such a case, one could either add artificial free arcs of zero weight (cf. Section 5.1) to make the graph strongly connected and thus increase the problem size. An alternative could be to use as many forward cycles in the basis as possible. This could lead to a new, interesting extremal cycle basis problem: Find a cycle basis such that the number of forward cycles is maximal. This would again relate back to the concept of cycle bases of subspaces.

Author Contribution Berenike Masing: conceptualization, methodology, writing—original draft, writing—review and editing, validation, formal analysis, visualization; Niels Lindner: conceptualization, methodology, writing—original draft, writing—review and editing, supervision, software; Patricia Ebert: formal analysis, investigation, visualization.

Funding Open Access funding enabled and organized by Projekt DEAL. Berenike Masing is funded by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) under Germany's Excellence Strategy - The Berlin Mathematics Research Center MATH+ (EXC-2046/1, project ID: 390685689).

Availability of Data and Materials Please contact the authors.

Declarations

Ethical Approval Not applicable.

Competing Interests The authors declare no competing interests.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

References

1. Serafini P, Ukovich W (1989) A mathematical model for periodic scheduling problems. *SIAM J Discret Math* 2(4):550–581
2. Borndörfer R, Hoppmann H, Karbstein M, Lindner N (2020a) Separation of cycle inequalities in periodic timetabling. *Discret Optim* 35:100552
3. Lindner N, Reisch J (2020) Parameterized complexity of periodic timetabling. Technical Report 20-15, Zuse Institute Berlin
4. Odijk MA (1994) Construction of periodic timetables, part 1: a cutting plane algorithm. Technical Report 94-61, TU Delft
5. Goerigk M (2012) PESPLib – a benchmark library for periodic event scheduling. <http://num.math.uni-goettingen.de/~m.goerigk/pesplib/>

6. Nachtigall K (1996) Cutting planes for a polyhedron associated with a periodic network. Technical Report 112-96/17, Deutsches Zentrum für Luft- und Raumfahrt
7. Liebchen C, Peeters L (2009) Integral cycle bases for cyclic timetabling. *Discret Optim* 6(1):98–109
8. Liebchen C, Swarat E (2008) The second chvatal closure can yield better railway timetables. In: Fischetti M, Widmayer P (eds) 8th Workshop on Algorithmic Approaches for Transportation Modeling, Optimization, and Systems (ATMOS'08), volume 9 of OpenAccess Series in Informatics (OASiCs), Dagstuhl, Germany. Schloss Dagstuhl-Leibniz-Zentrum fuer Informatik, ISSN, pp 2190–6807
9. Lindner N, Liebchen C (2020) Determining all integer vertices of the PESP polytope by flipping arcs. In: Huisman D, Zaroliagis CD (eds) 20th Symposium on Algorithmic Approaches for Transportation Modelling, Optimization, and Systems (ATMOS 2020), volume 85 of OpenAccess Series in Informatics (OASiCs), pages 5:1–5:18, Dagstuhl, Germany. Schloss Dagstuhl-Leibniz-Zentrum für Informatik, ISSN, pp 2190–6807
10. Lindner N, Liebchen C (2021) Timetable merging for the periodic event scheduling problem. Technical Report 21-06. Zuse Institute Berlin
11. Bortoletto E, Lindner N, Masing B (2022) Tropical neighbourhood search: a new heuristic for periodic timetabling. In: D'Emidio M, Lindner N (eds) 22nd Symposium on Algorithmic Approaches for Transportation Modelling, Optimization, and Systems (ATMOS 2022), volume 106 of Open Access Series in Informatics (OASiCs), pages 3:1–3:19, Dagstuhl. Schloss Dagstuhl - Leibniz-Zentrum für Informatik, Germany
12. Lindner N, Liebchen C, Masing B (2021) Forward cycle bases and periodic timetabling. In: Müller-Hannemann M, Perea F (eds) 21st Symposium on Algorithmic Approaches for Transportation Modelling, Optimization, and Systems (ATMOS 2021), volume 96 of Open Access Series in Informatics (OASiCs), pages 2:1–2:14, Dagstuhl. Schloss Dagstuhl - Leibniz-Zentrum für Informatik, Germany
13. Kavitha T, Liebchen C, Mehlhorn K, Michail D, Rizzi R, Ueckerdt T, Zweig KA (2009) Cycle bases in graphs characterization, algorithms, complexity, and applications. *Comput Sci Rev* 3(4):199–243
14. Liebchen C (2006) Periodic timetable optimization in public transport. PhD thesis, Technische Universität Berlin
15. Liebchen C, Möhring RH (2007) The modeling power of the periodic event scheduling problem: railway timetables - and beyond. In: Garaets, F., Kroon, L., Schöbel, A., Wagner, D., and Zaroliagis, C.D., editors, *Algorithmic Methods for Railway Optimization*, Lecture Notes in Computer Science, pages 3–40, Berlin, Heidelberg. Springer
16. Liebchen C, Rizzi R (2007) Classes of cycle bases. *Discret Appl Math* 155(3):337–355
17. Liebchen C, Peeters L (2002) On cyclic timetabling and cycles in graphs. Technical report, Technische Universität Berlin
18. Gleiss PM, Leydold J, Stadler PF (2003) Circuit bases of strongly connected digraphs. *Discussiones Mathematicae Graph Theory* 23(2):241
19. Seymour P, Thomassen C (1987) Characterization of even directed graphs. *Journal of Combinatorial Theory, Series B* 42(1):36–45
20. Cormen TH, Leiserson CE, Rivest RL, Stein C (2009) *Introduction to algorithms*, Third Edition. The MIT Press, 3rd edition
21. Helmberg C, Hofmann T, Wenzel D (2022) Periodic event scheduling for automated production systems. *INFORMS J Comput* 34(2):1291–1304
22. Wunsch G (2008) Coordination of traffic signals in networks. PhD thesis, Technische Universität Berlin
23. Borndörfer R, Lindner N, Roth S (2020b) A concurrent approach to the periodic event scheduling problem. *J Rail Trans Plann Manag* 15:100175. Best Papers of RailNorrköping 2019
24. Goerigk M, Liebchen C (2017) An improved algorithm for the periodic timetabling problem. In ATMOS, volume 59 of OASiCs 12:1–12:14. Schloss Dagstuhl - Leibniz-Zentrum für Informatik
25. Horton JD (1987) A polynomial-time algorithm to find the shortest cycle basis of a graph. *SIAM J Comput* 16(2):358–366
26. de Pina J (1995) Applications of shortest path methods. PhD thesis, University of Amsterdam
27. Hariharan R, Kavitha T, Mehlhorn K (2006) A faster deterministic algorithm for minimum cycle bases in directed graphs. In: Bugliesi, M., Preneel, B., Sassone, V., and Wegener, I., editors, *Automata, Languages and Programming*, pages 250–261, Berlin, Heidelberg. Springer Berlin Heidelberg
28. Amaldi E, Iuliano C, Jurkiewicz T, Mehlhorn K, Rizzi R (2009) Breaking the $o(m^2n)$ barrier for minimum cycle bases. p 301–312
29. Liebchen C, Proksch M, Wagner FH (2008) Performance of algorithms for periodic timetable optimization. volume 600, pages 151–180, Berlin, Heidelberg. Springer Berlin Heidelberg. Book Title:

Computer-aided Systems in Public Transport Series Title: Lecture Notes in Economics and Mathematical Systems

30. IBM (2019) IBM CPLEX optimizer 12.10 user's manual. <https://www.ibm.com/products/ilog-cplex-optimization-studio/cplex-optimizer>

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.