



# On Finding $k$ Earliest Arrival Time Journeys in Public Transit Networks

Ali Al-Zoobi, David Coudert, Arthur Finkelstein, Jean-Charles Régim

## ► To cite this version:

Ali Al-Zoobi, David Coudert, Arthur Finkelstein, Jean-Charles Régim. On Finding  $k$  Earliest Arrival Time Journeys in Public Transit Networks. ICORES 2022 - 11th International Conference on Operations Research and Enterprise Systems, Feb 2022, Virtual event, France. pp.314-325, 10.5220/0010977200003117 . hal-03559992

**HAL Id: hal-03559992**

**<https://inria.hal.science/hal-03559992>**

Submitted on 7 Feb 2022

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

# On Finding $k$ Earliest Arrival Time Journeys in Public Transit Networks<sup>a</sup>

Ali Al-Zoobi<sup>1</sup>, David Coudert<sup>1</sup>, Arthur Finkelstein<sup>2</sup> and Jean-Charles Régin<sup>3</sup>

<sup>1</sup>Université Côte d’Azur, Inria, CNRS, I3S, Sophia Antipolis, France

<sup>2</sup>Instant System, Sophia Antipolis, France

<sup>3</sup>Université Côte d’Azur, CNRS, I3S, Sophia Antipolis, France

**Keywords:** Public Transit Routing, shortest path, dissimilar paths.

**Abstract:** Journey planning in (schedule-based) public transit networks has attracted interest from researchers in the last decade. In particular, many algorithms aiming at efficiently answering queries of journey planning have been proposed. However, most of the proposed methods give the user a single or a limited number of journeys in practice, which is undesirable in a transportation context. In this paper, we consider the problem of finding  $k$  earliest arrival time journeys in public transit networks from a given origin to a given destination, i.e., an earliest arrival journey from the origin to the destination, a second earliest arrival journey, *etc.* until the  $k^{\text{th}}$  earliest arrival journey. For this purpose, we propose an algorithm, denoted by Yen - Public Transit (Y-PT), which extends to public transit networks the algorithm proposed by Yen to find the top- $k$  shortest simple paths in a graph. Moreover, we propose a more refined algorithm, called Postponed Yen - Public Transit (PY-PT), enabling a considerable speed up in practice. Our experiments on several public transit networks show that, in practice, PY-PT is faster than Y-PT by an order of magnitude.

## 1 INTRODUCTION

In the context of multimodal transportation, journey planning in (schedule-based) public transit networks and accelerating queries for efficient journey planning is a long-standing problem (Bast et al., 2016). In the last decade, many algorithms have been developed not only to efficiently answer basic queries like the quickest or the earliest arrival journey, but also to optimize additional criteria like the number of transfers, the cost of the trip, *etc.* or even to offer Pareto optimal solutions combining several criteria (Bast et al., 2016; Dellling et al., 2015; Dibbelt et al., 2018).

A *transit network* is a set of stops (such as bus stops or train stations), a set of routes (such as bus, tramways, ferries, metro or train lines), and a set of trips. Trips correspond to individual vehicles that visit the stops along a certain route at a specific time of the day. Trips can be further subdivided into sequences of elementary connections, each given as a pair of (origin/destination) stops and (departure/arrival) times between which the vehicle travels without stopping.

In addition, footpaths model walking transfers between nearby stops. A *journey* is a sequence of trips one can take within a transit network (also referred to as a transportation network or a *timetable*).

**The  $k$  Shortest Simple Paths Problem** A directed graph (*digraph* for short) is a set of vertices connected by arcs. A *path* from a source to a destination in a digraph is a sequence of vertices starting from the source and ending at the destination, such that consecutive vertices are connected by an arc. A path is *simple* if it has no repeated vertices. The *length* (or weight or cost) of a path is the sum of the lengths (or weights or costs) of its arcs. In this context, the  $k$  shortest simple paths ( $k$ SSP) problem asks to find a set  $S$  of  $k$  distinct simple paths from a source to a destination such that no path outside  $S$  has a length strictly lower than any path in  $S$ . This problem can be solved in time  $O(kn(m + n \log n))$  using the algorithm proposed by Yen (Yen, 1971), where  $n$  is the number of vertices and  $m$  is the number of arcs. Since this is the best known time complexity for this problem, a significant research effort has been put on the design of algorithms for efficiently solving the  $k$ SSP problem in practice (Kurz and Mutzel, 2016; Al Zoobi et al., 2020; Al Zoobi et al., 2021a). Note that, if the paths of  $S$  are not required to be simple, the problem can be solved by Eppstein’s algorithm in time  $O(k + m + n \log n)$  (Eppstein, 1998).

<sup>a</sup>This work has been supported by the French government, through the UCA<sup>JEDI</sup> Investments in the Future project managed by the National Research Agency (ANR) with the reference number ANR-15-IDEX-01 and the ANR project MULTIMOD with the reference number ANR-17-CE22-0016.

In fact, a road network can be modelled using a weighted directed graph where crossroads are represented by vertices and routes by arcs with length corresponding to the distances or the travel time between crossroads. So, finding  $k$  “best” (shortest, fastest or cheapest) paths from a given origin to a given destination in a road network is straightforward using any  $k$ SSP algorithm. Unfortunately, this problem becomes harder in public transit networks. First, because public transit networks are time dependent, i.e., certain segments of the network can only be traversed at specific times. Second, several additional optimization criteria are considered in public transit network such as the arrival time, the departure time, the number of transfers, *etc.*

### **Journey Planning Queries in Public Transit Networks**

A plethora of algorithms were proposed to efficiently answer queries of optimal journeys from a given origin  $o$  to a given destination  $d$  after a departure time  $t_0$  in a public transit network. For instance, the *Connection Scan Algorithm (CSA)* (Dibbelt et al., 2018) is the fastest algorithm, without any preprocessing routine, enabling to find an earliest arrival journey from  $o$  to  $d$  departing after  $t_0$ . With the help of a heavy preprocessing routine, the *Transfer Patterns algorithm* (Bast et al., 2010) can achieve a tremendous speed up with respect to the CSA. Besides, *Round Based Public Transit Routing (RAPTOR)* (Delling et al., 2015) is the fastest algorithm (also without any preprocessing routine) enabling to compute a Pareto optimal set of journeys optimizing the arrival time and the number of transfers of a journey. Recently, Bast et al. (Bast et al., 2016) presented an extensive survey on the topic of journey planning in road and public transit networks.

**Related Work** Vo et al. (Vo et al., 2015) proposed a time dependent graph modeling a bus network. Then, they adapt Yen’s algorithm to find alternative journeys in this network model. Precisely, they select a set of alternative journeys (journeys sharing only a limited part of their common edges) among those given by Yen’s adaptation.

As shown below, Yen’s algorithm uses Dijkstra’s algorithm as a basic brick to compute shortest detours of a given path. Analogously, Vo et al. (Vo et al., 2015) used the time-dependent shortest path (TDSP) algorithm of (Schulz et al., 2000) to compute earliest detours of a journey in a bus network. They evaluated their method on a single network of around 4000 stops and 8000 connections, resulting in an average running time of around 1 second to find 5 journeys.

On the other hand, Scano et al. (Scano et al., 2015) modeled a transportation network as a labeled directed graph where a label is an object composed

of the transportation mode (foot, car, bus, *etc.*) and a travel time. This model merges road and public transport networks together. Then, it is shown how the  $k$  shortest path algorithms can be adapted for this model. Precisely, they adapted Yen’s and Eppstein’s algorithm to work on their model. In both algorithms, a Dijkstra-like algorithm called *Dijkstra Regular Language Constraint (DRegLC)* (Barrett et al., 2008) is used to answer earliest arrival journeys queries. Moreover, an *Iterative Enumeration Algorithm (IEA)* is proposed to extract only simple journeys using Eppstein’s algorithm. i.e., using Eppstein’s  $k$  shortest paths algorithm as an iterator and then selecting the simple corresponding journeys (a journey is *simple* if it does not visit a stop more than once).

Experimentally, Scano et al. showed that their IEA is faster than Yen’s straightforward adaptation on the transportation network of Toulouse (75000 nodes, 500000 road edges and 43000 public transport edges). On this network, the average running time of Yen’s adaptation to find 100 journeys is 250 seconds while it is 0.6 seconds using their refined IEA. However, IEA is not a polynomial-time algorithm, and its memory consumption is too high (Scano et al., 2015). In addition, using the labelled directed graph model described in (Scano et al., 2015) may cause a duplication of the public transit part in practice, i.e., a large number of journeys given by the algorithms proposed in (Scano et al., 2015) may only differ on the footpath part while sharing the exact same public transit part. This is undesirable in applications requesting diverse public transit journeys.

**Our contributions** In this paper, we aim at answering  $k$  earliest arrival journeys queries from a given origin to a given destination in a public transit network. To this end, we use the timetable model of public transit networks as in (Bast et al., 2016; Dibbelt et al., 2018; Delling et al., 2015). First, we propose a performant adaptation of Yen’s  $k$  shortest simple paths algorithm to public transit networks (Yen - Public Transit, Y-PT algorithm). In contrast with (Scano et al., 2015; Vo et al., 2015), we use the *Connection Scan Algorithm (CSA)* to answer earliest arrival journey queries in our algorithm.

Our main contribution is a novel algorithm, called *Postponed Yen’s algorithm for Public Transit networks (PY-PT)*. With the help of a lower bound on the arrival time of a detour journey (a journey that may be one of the  $k$  earliest arrival journeys), PY-PT postpones the effective computation of such detour (and so the corresponding earliest arrival journey queries using CSA) with the aim of skipping it.

Our experimental results on several train and public transit networks show that the running time of our

adaptation of Yen’s algorithm is acceptable in practice. Moreover, on the same dataset, the PY-PT algorithm performs 10 to 30 times faster than the Y-PT algorithm on average.

Finally, we evaluate the mutual similarity of the journeys given by our algorithms. And we show experimentally that our algorithms can be used to extract earliest arrival journeys that are mutually dissimilar.

## 2 PRELIMINARIES

In this section we formalize the inputs and algorithms used in this work. We use almost the same formalization used in (Dibbelt et al., 2018) for the CSA and as in (Al Zoobi et al., 2020) for Yen’s algorithm.

### 2.1 Graph - Definitions and Notations

Let  $D = (V, A)$  be a digraph with  $n = |V|$  vertices and  $m = |A|$  arcs, let  $N^+(v) = \{w \in V \mid vw \in A\}$  be the set of out-neighbors of a vertex  $v \in V$ , and let  $\ell_D : A \rightarrow \mathcal{R}^+$  be a length function over the arcs.

For every  $s, t \in V$ , a *path* from  $s$  to  $t$  in  $D$  is a sequence  $P = (s = v_0, v_1, \dots, v_l = t)$  of vertices with  $v_i v_{i+1} \in A$  for all  $0 \leq i < l$ . An arc  $uv$  belongs to a path  $P$  ( $uv \in P$ ) if and only if  $u$  and  $v$  are two consecutive vertices of  $P$ , i.e., there is  $0 \leq i < l$  such that  $u_i = u$  and  $u_{i+1} = v$ . A path is *simple* if all of its vertices are distinct, i.e.,  $v_i \neq v_j$  for all  $0 \leq i < j \leq l$ . The *length* of the path  $P$  is the sum of the lengths of its arcs,  $\ell_D(P) = \sum_{0 \leq i < l} \ell_D(v_i, v_{i+1})$ . The *distance*  $d_D(s, t)$  between two vertices  $s, t \in V$  is the length of a shortest  $s$ - $t$  path, i.e., a path with the smallest length among all the  $s$ - $t$  paths. Given two paths  $P = (v_0, \dots, v_r)$  and  $Q = (w_0, \dots, w_p)$ , and an arc  $v_r w_0 \in A$ , we denote by  $P.Q$  the  $v_0$ - $w_p$  path resulting from the concatenation of  $P$  and  $Q$ . That is,  $P.Q = (v_0, \dots, v_r, w_0, \dots, w_p) = (v_0, \dots, v_r, Q) = (P, w_0, \dots, w_p)$ .

Given  $s, t \in V$ , a *set of top- $k$  shortest simple  $s$ - $t$  paths* is any set  $S$  of  $s$ - $t$  simple paths such that  $|S| = k$  and  $\ell(P) \leq \ell(P')$  for every  $s$ - $t$  path  $P \in S$  and  $s$ - $t$  path  $P' \notin S$ . The  $k$  shortest simple paths problem takes as input a digraph  $D = (V, A)$ , a length function over the arcs  $\ell_D : A \rightarrow \mathcal{R}^+$  and a pair of vertices  $(s, t) \in V^2$  and asks to find a set of top- $k$  shortest simple  $s$ - $t$  paths.

Dijkstra’s algorithm finds an  $s$ - $t$  shortest path in  $D$  with worst-case time complexity in  $O(m + n \log n)$ .

Let  $P = (v_0, v_1, \dots, v_l)$  be any path in  $D$ . Let  $0 \leq i < l$ , any path  $P' = (v_0, \dots, v_i, v', v'_1, \dots, v'_r = v_l)$  s.t.  $v' \neq v_{i+1}$  is called a *detour* of  $P$  at  $v_i$ . Note that neither  $P$  nor  $P'$  are required to be simple. However, if  $P'$  is simple, it will be called a *simple detour* of  $P$  at  $v_i$ . In addition,  $P'$  is called a shortest (simple) detour at

$v_i$  if and only if  $P'$  is a detour with minimum length among all (simple) detours of  $P$  at  $v_i$ . Finally, the subpath  $\pi_i = (v_0, \dots, v_{i-1})$  of  $P$  starting from  $s$  and ending at  $v_{i-1}$  for  $0 \leq i \leq l$  is called  *$i$ -prefix path* of  $P$  (the 0-prefix of any path is an empty path)

### 2.2 Yen’s Algorithm

We now describe Yen’s algorithm for finding a set of top- $k$  shortest simple  $s$ - $t$  paths in  $D$ . For the sake of simplicity, we assume that  $D$  has at least  $k$   $s$ - $t$  simple paths.

Yen’s algorithm starts by computing a shortest  $s$ - $t$  path  $P_0 = (s = v_0, v_1, \dots, v_l = t)$  by applying Dijkstra’s algorithm. Note that  $P_0$  is simple since the weights of  $D$  are non-negative. Clearly, a second shortest simple  $s$ - $t$  path is a shortest simple detour of  $P_0$  at one of its vertices. Yen’s algorithm computes, for every vertex  $v_i$  in  $P_0$ , a shortest simple detour of  $P_0$  at  $v_i$ . For this purpose, for  $0 \leq i < r$ , Yen’s algorithm removes the vertices of the  $i$ -prefix path  $\pi_i = (v_0, \dots, v_{i-1})$  of  $P_0$  and the arc  $v_i v_{i+1}$ , then it computes, using Dijkstra’s algorithm, a shortest path  $Q_i$  from  $v_i$  to  $t$ . Let  $C_i = \pi_i.Q_i$  be the concatenation of  $\pi_i$  and  $Q_i$ . First,  $C_i$  is simple as  $Q_i$  is computed after removing  $\pi_i$ . Second,  $v_i v_{i+1} \notin C_i$  as the arc  $v_i v_{i+1}$  of  $P_0$  is removed before computing  $Q_i$  and constructing  $C_i$ . Therefore,  $C_i$  is a shortest simple detour of  $P_0$  at  $i$ . Note that the index  $i$  (called below *deviation-index*) where the path  $(v_0, \dots, v_{i-1}, Q_i)$  deviates from  $P_0$  is kept explicit, i.e., the path is stored with its deviation index. Finally,  $C_i$  is added to a set *Candidates* (initially empty) for every  $0 \leq i < l$ . Once  $C_i$  has been added to *Candidates* for all  $0 \leq i < l$ , by remark above, a path with minimum length in *Candidates* is a second shortest simple  $s$ - $t$  path.

Now, let us assume that a set  $S$  of top- $k'$  (with  $0 < k' < k$ ) shortest simple  $s$ - $t$  paths has been computed and the set *Candidates* contains a set of simple  $s$ - $t$  paths such that there exists a shortest path  $Q \in \text{Candidates}$  with  $S \cup \{Q\}$  a top- $(k' + 1)$  set of shortest  $s$ - $t$  simple paths.

Let  $R = (v_0 = s, \dots, v_j, \dots, v_r = t)$  be a path in *Candidates* with minimum length and let  $j$  be its deviation index. Similarly to the procedure of finding a second shortest path, Yen’s algorithm iterates over the vertices  $v_i$  ( $j \leq i < r$ ) of  $R$ . At each vertex  $v_i$ , a shortest simple detour of  $R$  at  $v_i$  is added to *Candidates* (since one of these detours may be a  $k' + 1^{\text{th}}$  shortest simple  $s$ - $t$  path). Let, again,  $\pi_i = (v_0, \dots, v_{i-1})$  be the  $i$ -prefix of  $R$ . Yen’s algorithm removes  $\pi_i$  from  $D$ . Then, it removes each arc  $v_i w$  such that  $S$  contains a path with  $(v_0, \dots, v_i, w)$  as a  $i + 1$ -prefix. Finally, a shortest  $v_i$ - $t$  path  $Q_i$  is computed, using Dijkstra’s

algorithm, and the path  $\pi_i.Q_i$  is added to *Candidates* with  $i$  as deviation index. This process is repeated until  $k$  paths have been found, i.e., when  $k' = k$ .

Therefore, for each path  $R$  that is extracted from *Candidates*,  $O(|V(R)|)$  calls of Dijkstra's algorithm are done. This results in a worst-case time-complexity in  $O(kn(m + n \log n))$ .

## 2.3 Timetable - Definitions and Notations

In this section, we describe the data structures used by the Connection Scan Algorithm (CSA) with the same formalism as in (Dibbelt et al., 2018). Then we will describe briefly the CSA and one of its variant called the profile CSA (PCSA).

**Timetable** A timetable represents for one specific day the vehicles that exist (train, bus, tram, ferry, ...), when they travel, where they travel and how a passenger can go from one vehicle to another. Formally, a timetable is a quadruple  $\mathcal{T} = (S, T, C, F)$  of stops  $S$ , trips  $T$ , connections  $C$  and footpaths  $F$ :

**Stop:** a position outside a vehicle where a passenger can wait. At a stop (and only at a stop) a vehicle can halt and passengers can leave or get on.

**Trip:** defined by a vehicle going through stops at fixed times. Precisely, a trip is a scheduled vehicle, i.e., a journey done by a unique vehicle from a starting stop to a last stop at a fixed time.

**Connection:** a vehicle going from one stop to another with no intermediate stops. Formally, a connection  $c$  is a quintuple  $(c_{dep\_stop}, c_{arr\_stop}, c_{dep\_time}, c_{arr\_time}, c_{trip})$  whose attributes are the departure stop, the arrival stop, the departure time, the arrival time and the trip of  $c$ , respectively. A connection must respect two conditions: (1) it cannot be a self loop, i.e.,  $c_{dep\_stop} \neq c_{arr\_stop}$  and (2) it has a non-zero travel time, i.e.,  $c_{dep\_time} < c_{arr\_time}$ .

**Footpath:** used to model a transfer from a vehicle to another. Formally, a footpath  $f$  is a triple  $(f_{dep\_stop}, f_{arr\_stop}, f_{dur})$  whose attributes are the departure stop, the arrival stop and the duration of the footpath, respectively. Note that, footpaths are neither trips, nor connections.

Note that, all the connections of a trip form a sequence  $c^1, c^2 \dots c^\phi$ , such that  $c_{arr\_stop}^i = c_{dep\_stop}^{i+1}$  and  $c_{arr\_time}^i < c_{dep\_time}^{i+1}$  for all  $0 \leq i \leq \phi$ .

Going from a connection  $c$  to a connection  $c'$  with  $c_{trip} \neq c'_{trip}$  is possible if and only if there is a footpath  $f^t$  from  $c_{arr\_stop}$  to  $c'_{dep\_stop}$  such that  $c'$  is reachable

via  $f^t$ , i.e.,  $f_{dur}^t \leq c'_{dep\_time} - c_{arr\_time}$ . A loop is introduced on each stop to allow a passenger to get off at a stop and take another trip going through this stop.

**Journeys** A journey describes how a passenger can travel through a public transit network. It is made of *legs* that are sequences of connections of the same trip. Formally, a journey is a sequence of alternating footpaths and legs  $J = (f^0, l^0, f^1, l^1 \dots f^{r-1}, l^r, f^r)$ , where  $l^i = (c_0^i, \dots, c_{\delta_i}^i)$ . That is, a passenger takes the footpath  $f^0$  from  $f_{dep\_stop}^0$  to  $f_{arr\_stop}^0$ , then takes the connection  $c_0^1, c_1^1, \dots, c_{\delta_1}^1$ , proceeds to take the footpath  $f^1$  from  $f_{dep\_stop}^1$  to  $f_{arr\_stop}^1$  etc. until reaching  $f_{arr\_stop}^r$ . A journey must start and end with a footpath, which can be a self loop. In this paper, we sometimes denote a journey as a sequence of footpaths and connection, i.e.,  $J = (f^0, c^0, c^1, \dots, c^\alpha, f^1, c^{\alpha+1}, \dots, f^{r-1}, c^{\gamma+1}, \dots, c^\phi, f^r)$  where  $c^0 = c_0^0, c^1 = c_1^0, \dots, c^\phi = c_{\delta_r}^r$ .

Given two stops  $o$  and  $d$  in  $S$ , an  $o$ - $d$  journey  $J$  is a journey  $(f^0, c^0, \dots, c^\phi, f^r)$  such that  $f^0$  starts from  $o$  and  $f^r$  ends at  $d$ . We define the departure time of a journey  $dep_t(J)$  as the departure time of its first footpath, formally,  $dep_t(J) = c_{dep\_time}^0 - f_{dur}^0$ . Similarly, the arrival time of a journey  $arr_t(J)$  is the arrival time of its last footpath, i.e.,  $arr_t(J) = c_{arr\_time}^\phi + f_{dur}^r$ .

A journey is called simple if it does not visit twice the same stop (except for self loop footpaths). Formally, let  $J = (f^0, l^0 = (c_0^0, \dots, c_{\delta_0}^0), \dots, f^j, l^j = (c_0^j, \dots, c_{\delta_j}^j), \dots, f^j, l^j = (c_0^j, \dots, c_{\delta_j}^j), \dots, l^r = (c_0^r, \dots, c_{\delta_r}^r), f^r)$  be a journey. For all  $0 \leq i < j \leq r$ , let  $c_{dep\_stop}^\alpha$  be the departure stop of  $c_\alpha^i$  for  $0 \leq \alpha \leq \delta_i$ . Similarly, for  $0 \leq \beta \leq \delta_j$ , let  $c'_{dep\_stop}^\beta$  be the departure stop of  $c_\beta^j$  and  $c'_{arr\_stop}^\beta$  be the arrival stop of  $c_\beta^j$ . We have  $c_{dep\_stop}^\alpha \neq c'_{dep\_stop}^\beta$  and  $c_{dep\_stop}^\alpha \neq c'_{arr\_stop}^\beta$ .<sup>1</sup>

The concatenation of two journeys  $J = (f^0, l^0, \dots, l^r, f^r)$  and  $J' = (f'^0 = f^r, l'^0, \dots, l'^\ell, f'^\ell)$  such that  $f^r = f'^0$  and  $arr_t(J) \leq dep_t(J')$  is the journey that starts by  $f^0$ , follows  $J$  until  $f^r$ , and then follows  $J'$  until  $f'^\ell$ . We denote  $J'' = J.J' = (f^0, l^0, \dots, f^r = f'^0, \dots, l'^\ell, f'^\ell)$ .

Given a journey  $J = (f^0, c^0, \dots, c^i, \dots, c^\phi, f^r)$ , a journey  $Q = (f'^0, c'^0, \dots, c'^i, \dots, c'^w, f'^\ell)$  is called a *detour* of  $J$  at  $i$  if  $f'^0 = f^0, c'^0 = c^0, \dots, c'^{i-1} = c^{i-1}$  but  $c'^i \neq c^i$  and  $f'_{arr\_stop}^\ell = f_{arr\_stop}^r$ . If  $Q$  is simple, it is called a *simple detour* of  $J$  at  $i$ , and  $Q$  is called an earliest arrival (simple) detour of  $J$  at  $i$ , if  $arr_t(Q) \leq arr_t(Q')$  for each (simple) detour  $Q'$  of  $J$  at  $i$ .

Two journeys are equal if and only if all of their attributes are the same.

<sup>1</sup>We suppose that a leg cannot have a loop, as a user may get off and wait outside the corresponding vehicle.

We denote by  $\mathcal{J}_{o,d}^{t_0,t_{\max}}$  the set of  $o$ - $d$  simple journeys starting from  $o$  after  $t_0$  and reaching  $d$  before  $t_{\max}$ , i.e.,  $\mathcal{J}_{o,d}^{t_0,t_{\max}} = \{J \text{ s.t. } J \text{ is a simple } o\text{-}d \text{ journey with } \text{dep}_t(J) \geq t_0 \text{ and } \text{arr}_t(J) \leq t_{\max}\}$ .

## 2.4 Connection Scan Algorithm

The CSA answers earliest arrival time journey queries from a given origin  $o$  to a given destination  $d$ . That is, departing after a given time  $t_0$ , how to get from  $o$  to  $d$  as soon as possible.

Similarly to Dijkstra’s algorithm, the CSA will store an earliest arrival time for each stop in an array. A connection is considered *reachable* if a passenger can sit in the public transit vehicle of the connection. However, the main difference between Dijkstra’s algorithm and the CSA is the fact that the CSA does not use a priority queue. Instead, the CSA iterates over all the connections sorted by their departure time (the same ordering is used for all queries). The CSA checks whether a connection is reachable or not. If so, it improves the arrival time at the arrival stop of the connection. Once all the connections have been scanned, the earliest arrival time to a stop is the current arrival time stored for the stop. The main advantage of avoiding the use of a priority queue is that, while more connections are scanned, the amount of work per connections is significantly reduced. Therefore, the CSA is significantly faster than Dijkstra’s algorithm (Dibbelt et al., 2018).

## 2.5 Profile Connection Scan Algorithm

The result of the Profile Connection Scan Algorithm (PCSA) is a mapping between a departure time from a departure stop onto the earliest arrival time at the arrival stop. In other words, the profile problem solves simultaneously the earliest arrival problem for all departure times.

Compared with the CSA, the PCSA iterates on the connections sorted decreasingly by departure time, which leads to the fact that it solves the all-to-one problem. The PCSA constructs journeys from late to early and exploits the fact that an early journey can only have later journeys as subjourneys. It has been reported in (Dibbelt et al., 2018) that the PCSA is one order of magnitude slower than the CSA, which is acceptable considering the fact that it solves the all-to-one problem.

Note that, the PCSA offers, from each stop  $s$  to the arrival stop  $d$ , a single earliest arrival  $s$ - $d$  journey departing after  $t_0$  and reaching  $d$  before  $t_{\max}$ .

Let  $M$  be the output of the PCSA, we denote by  $M_{o,d}^{t_0,t_{\max}}$  an earliest arrival journey starting from  $o$  and

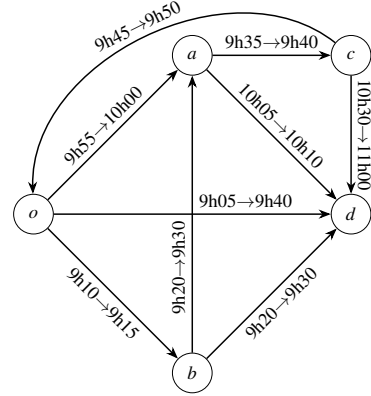


Figure 1: Toy network for  $k$  earliest arrival time journeys

reaching  $d$ , departing after  $t_0$  and arriving before  $t_{\max}$ .

## 3 PROBLEM DEFINITION

In this section, we formalize the  $k$  Earliest Arrival Time problem.

**$k$  Earliest Arrival Time ( $k$ EAT) Problem** In this paper, we aim at finding  $k$  earliest arrival time ( $k$ EAT) simple journeys from a given origin to a given destination. Formally, the problem takes as input a timetable  $\mathcal{T} = (S, T, C, F)$ , origin and destination stops  $o, d$  in  $S$ , a departure time  $t_0$ , a maximum arrival time  $t_{\max}$  (often  $t_{\max} = t_0 + 24h$  or  $t_{\max} = t_0 + 48h$ ) and an integer  $k$ . It asks to find a set  $\mathcal{J}^* = \{J_1, J_2, \dots, J_k\}$  of top- $k$  earliest arrival  $o$ - $d$  simple journeys i.e.,  $J_i \neq J_j$  for  $0 \leq i < j \leq k$ , and for every  $J$  in  $\mathcal{J}^*$ ,  $J' \in \mathcal{J}_{o,d}^{t_0,t_{\max}}$ ,  $\text{arr}_t(J) \leq \text{arr}_t(J')$ .

**Example** In the example of Figure 1, we look for the four earliest arrival time journeys from  $o$  to  $d$  departing after 9h00. The earliest arrival journey  $J_0 = (o, d, b)$  arrives at  $d$  at 9h30, starts with  $o$  and reaches  $d$  via  $b$ , the passenger arrives at  $b$  at 9h15 and waits 5 minutes before boarding the connection going from  $b$  to  $d$ . The second journey  $J_1 = (o, d)$  arrives at 9h40 and goes directly from  $o$  to  $d$ . The third journey  $J_2 = (o, b, a, d)$  arrives at 10h10 and goes from  $o$  to  $b$  then  $a$  then  $d$ , the passenger arrives at  $b$  at 9h15, waits 10 minutes then boards the connection going from  $b$  to  $a$ , arrives at 9h30 and waits 35 minutes before boarding the connection going from  $a$  to  $d$ . The fourth journey  $J_3 = (o, a, d)$  arrives at 10h10 and goes from  $o$  to  $a$  then  $d$ .

Note that the journey  $J_{ns} = (o, b, a, c, o, a, d)$  arriving at 10h10 is not a part of the solution as it is not simple (it visits the station  $o$  twice). Note also that there are other  $o$ - $d$  journeys in this example, each arriving after 10h10. Therefore,  $\{J_0, J_1, J_2, J_3\}$  are the

four earliest arrival simple  $o$ - $d$  journeys.

Each edge in the graph belongs to a specific trip, i.e, there is a self loop path between each step in the examples.

## 4 PUBLIC TRANSIT YEN'S ALGORITHM (Y-PT)

In this section, we describe our adaption of Yen's algorithm on public transit networks, called Y-PT algorithm. As described before, Y-PT algorithm solves the  $k$ EA problem. So, it takes as input a timetable  $\mathcal{T} = (S, T, C, F)$ , origin and destination stops  $o, d$  in  $S$ , a departure time  $t_0$ , a maximum arrival time  $t_{max}$  ( $= t_0 + 48h$ ) and an integer  $k$ , and returns a set  $Output = \{J_1, J_2, \dots, J_k\}$  of top- $k$  earliest arrival  $o$ - $d$  simple journeys in  $\mathcal{T}$ .

Roughly, Y-PT algorithm starts by computing a first earliest arrival journey, iterates over its connections in order to compute its earliest arrival simple detours and adds their minimum (the detour with minimum arrival time) to the output. Then, Y-PT algorithm repeats this process until  $k$  journeys are added to the output.

Now, let us give a precise and formal description of Y-PT algorithm whose pseudocode is presented in Algorithm 1. Analogously to Yen's algorithm, Y-PT starts by computing an earliest arrival journey  $J_0$  and adding it (with 0 as deviation index) to a set of candidate journeys called *Candidates*. The journeys of the set *Candidates* are non-decreasingly sorted by their arrival time. Also, the algorithm initializes the output set *Output* as an empty set. After this initialization phase, the algorithm extracts a minimum element from the set *Candidates*, i.e, a journey  $J = (f^0, c^0, \dots, c^\phi, f^r)$  with minimum arrival time among those in *Candidates* and adds it to *Output*. Let  $C_J = (c^0, c^1, \dots, c^\phi)$  be the sequence of connections of  $J$ . The algorithm iterates over the connections in  $C_J$  starting from the deviation index of  $J$ . Precisely, let  $j$  be the deviation index of  $J$ , for each connection  $c^i = (c_{dep\_stop}^i, c_{arr\_stop}^i, c_{dep\_time}^i, c_{arr\_time}^i, c_{trip}^i)$  for  $j \leq i \leq \phi$ , the algorithm removes the prefix stations, i.e, each station visited by one of the connections  $c^0, \dots, c^{i-1}$ , (equivalent to the prefix path of Yen's) from  $\mathcal{T}$ . This is done to ensure that the candidate journey is simple.

Moreover, in order to avoid duplications of journeys, for each journey  $J$  in *Output* starting with the connections  $c^0, c^1, \dots, c^{i-1}, c'$ , the connection  $c'$  is removed from  $\mathcal{T}$ . Then, using the CSA, the Y-PT algorithm computes an earliest arrival journey  $Q = (f_Q^0, c_Q^0, \dots, c_Q^\omega, f_Q^\ell)$  from  $c_{arr\_stop}^{i-1}$  to  $d$  with  $c_{arr\_time}^{i-1}$

---

### Algorithm 1 Public Transit - Yen's algorithm (PT-Y)

---

```

1: Input A timetable  $\mathcal{T} = (S, T, C, F)$ , an origin and
   a destination stops ( $o$  and  $d$ ), departure and maximum
   arrival time  $t_{dep}, t_{max}$  and an integer  $k$ 
2: Output a set of top- $k$  earliest arrival journeys
   from  $o$  to  $d$  departing after  $t_{dep}$ 
3:  $J_0 \leftarrow CSA(\mathcal{T}, o, d, t_{dep}, t_{max})$ 
4:  $Candidates \leftarrow \{(J_0, 0)\}$ 
5:  $Output \leftarrow \emptyset$ 
6: while  $|Output| < k$  and  $Candidate \neq \emptyset$  do
7:    $\epsilon = (J, j) \leftarrow extractmin(Candidates)$ 
8:   Let  $C_J = (c^0, \dots, c^\phi)$  be the sequence of connections
     of  $J$ 
9:   add  $J$  to  $Output$ 
10:  for each connection  $c^i$  with  $j \leq i \leq \phi$  in  $C_J$ 
    do
11:     $c_{arr\_stop} \leftarrow$  the arrival stop of  $c^{i-1}$ 
12:     $c_{arr\_time} \leftarrow$  the arrival time of  $c^{i-1}$ 
13:     $\pi = (f^0, c^0, \dots, c^{i-1})$ 
14:     $S_\pi \leftarrow$  the set of stations visited by one of
     the connections  $(c^0, \dots, c^{i-1})$ 
15:     $C_{dev} \leftarrow \{c' \text{ s.t. there is a journey } J' \text{ in }
     Output \text{ starting with } (c^0, \dots, c^{i-1}, c')\}$ 
16:     $\mathcal{T}' = (S \setminus S_\pi, T, C \setminus C_{dev}, F)$ 
17:     $Q \leftarrow CSA(\mathcal{T}', c_{arr\_stop}, d, c_{arr\_time}, t_{max})$ 
18:     $J_{new} \leftarrow \pi.Q$ 
19:    add  $(J_{new}, i)$  to  $Candidates$ 
20:  end for
21: end while
22: Return  $Output$ 

```

---

as departure time <sup>2</sup>. Let  $J_{new}$  be the concatenation of the prefix of  $J$  and  $Q$ , i.e,  $J_{new} = (f^0, c^0, \dots, c^{i-1}, f_Q^0, c_Q^0, \dots, c_Q^\omega, f_Q^\ell)$ . The journey  $J_{new}$  is added to *Candidates* with  $i$  as deviation index.

Y-PT algorithm repeats this process until  $k$  journeys are added to *Output*.

## 5 PUBLIC TRANSIT POSTPONED YEN'S ALGORITHM (PY-PT)

We now present the Postponed Yen algorithm for public transit (PY-PT algorithm) whose pseudocode is presented in Algorithm 2. It is inspired from the Postponed Node Classification algorithm (PNC) for the  $k$ SSP described in (Al Zoobi et al., 2021a).

---

<sup>2</sup>If the element right before  $c^i$  is a footpath, i.e,  $J = (f^0, \dots, f^\lambda, c^i, \dots, f^r)$ , it is possible to have journeys with two consecutive footpaths. In order to avoid such scenario, the CSA call is forced to compute a journey starting with a self loop footpath.

PY-PT algorithm has the same input as Y-PT algorithm, and it also returns a set of top- $k$  earliest arrival simple journeys from the origin to the destination in a timetable. However, the journeys given by Y-PT are not necessarily the same as those given by PY-PT, i.e., the order of extraction of journeys is not necessarily the same. This may occur in scenarios where several journeys from the origin to the destination have the same arrival time.

The main drawback of Y-PT algorithm is its excessive number of calls of the CSA. Here, with the help of lower bounds on the arrival time of simple detours, we propose to postpone these calls in order to avoid some of them. We show that this can be done while preserving the correctness of the algorithm. In contrast with Y-PT algorithm where all journeys in the set *Candidates* are simple, the PY-PT algorithm may add non-simple journeys to the set *Candidates*. As shown below, this corresponds to detours whose effective computation (and so their corresponding CSA calls) are postponed.

Let us now describe PY-PT algorithm in details.

For a query from the origin  $o$  to the destination  $d$  starting at time  $t_0$ , the PY-PT algorithm first uses the Profile CSA (PCSA). Let  $M$  be the mapping output by PCSA. The mapping  $M$  associates to each station  $s \in S$  and each departure time  $t \geq t_0$  the earliest arrival  $s$ - $d$  journey, providing it is possible to reach  $d$  from  $s$  before  $t_{max}$  when starting at  $t$  (we let  $t_{max} = t_0 + 48h$  in our experiments).

Similarly to Y-PT algorithm, PY-PT algorithm starts by adding an earliest arrival time journey  $J_0$  to a set of candidate journeys called *Candidates*. An element  $\varepsilon$  in *Candidates* has three attributes, the journey  $J$ , its deviation index  $i$  and a boolean flag  $\zeta$  indicating whether  $J$  is simple or not. So, the element  $\varepsilon_0 = (J_0, 0, 1)$  is added to *Candidates*. In contrast with Y-PT algorithm where a CSA call is consumed to compute  $J_0$ , PY-PT algorithm extract  $J_0$  from the already computed mapping  $M$ . Precisely,  $J_0 = M_{o,d}^{t_0, t_{max}}$ . Then, also like Y-PT algorithm, the *Output* set is initialized with an empty set. After these initializations steps, the algorithms starts by extracting an earliest arrival journey  $(J, j, \zeta)$  among those in *Candidates*. Suppose  $J = (f^0, c^0, \dots, c^\alpha, f^1, c^{\alpha+1}, \dots, c^\beta, f^2, \dots, c^{\gamma+1}, \dots, c^\phi, f^r)$ . Two cases are distinguished:

**if  $\zeta = 1$  ( $J$  is simple):**  $J$  is added to the *Output*, then all the earliest arrival detours of  $J$  are added to *Candidates*. This is done as follows, let  $C_J = (c^0, c^1, \dots, c^\phi)$  be the sequence of connections of  $J$ , at each connection  $c^i$  (for  $j \leq i < \phi$ ) in  $C_J$ , an earliest arrival detour  $J_{new}$  of  $J$  at  $i$  is extracted. This operation uses  $M$  as described below.

---

**Algorithm 2** Public Transit - Postponed Yen's algorithm (PY-PT)

---

```

1: Input A timetable  $\mathcal{T}$ , an origin and a destination
   stops ( $o$  and  $d$ ), departure and maximum arrival
   time ( $t_{dep}$  and  $t_{max}$ ), and an integer  $k$ 
2: Output a set of top- $k$  earliest arrival simple jour-
   neys from  $o$  to  $d$  departing after  $t_{dep}$ 
3:  $M \leftarrow PCSA(\mathcal{T}, o, d, t_{dep}, t_{max})$ 
4:  $J_0 \leftarrow M_{o,d}^{t_{dep}, t_{max}}$ 
5: Candidates  $\leftarrow \{(J_0, 0, \zeta = 1)\}$ 
6: Output  $\leftarrow \emptyset$ 
7: while Candidates  $\neq \emptyset$  and  $|Output| < k$  do
8:    $\varepsilon = (J, j, \zeta) \leftarrow extractmin(Candidates)$ 
9:   Let  $C_J = (c^0, \dots, c^\phi)$  be the sequence of con-
   nections of  $J$ 
10:  if  $\zeta = 1$  ( $J$  is simple) then
11:    add  $J$  to Output
12:  for each connection  $c^i$  in  $C_J$  ( $c_j, \dots, c^\phi$ )
  do
13:     $J_{new} \leftarrow EarliestArrivalDetour(J, i, M)$ 
14:     $\zeta' \leftarrow 0$ 
15:    if  $J_{new}$  is simple then
16:       $\zeta' \leftarrow 1$ 
17:    end if
18:    add  $(J_{new}, i, \zeta')$  to Candidates
19:  end for
20:  else
21:     $S_\pi \leftarrow$  the set of stations visited by one of
    the connections  $(c^0, \dots, c^{j-1})$ 
22:     $C_{dev} \leftarrow \{c \text{ s.t. there is a journey } J' \text{ in}$ 
    Output starting with  $(c^0, \dots, c^{i-1}, c)\}$ 
23:     $\mathcal{T}' = (S \setminus S_\pi, \mathcal{T}, C \setminus C_{dev}, F)$ 
24:     $Q \leftarrow CSA(\mathcal{T}', c_{arr\_stop}, d, c_{arr\_time}, t_{max})$ 
25:    if  $Q$  exists then
26:       $J_{new} \leftarrow (f^0, c^0, \dots, c^j, Q)$ 
27:      add  $(J_{new}, j, \zeta = 1)$  to Candidates
28:    end if
29:  end if
30: end while
31: return Output

```

---

The journey  $J_{new}$  may not be simple (also described below). However,  $J_{new}$  will be added to the set *Candidate* with  $i$  as deviation index and  $\zeta = 1$  if  $Q$  is simple (and  $\zeta = 0$  otherwise).

**if  $\zeta = 0$  ( $J$  is not simple):**  $J$  is “repaired”, i.e., it is replaced (if possible) by its corresponding earliest arrival simple journey. For this purpose, the algorithm applies almost the same routine as Y-PT algorithm. Precisely, let  $c^j = (c_{dep\_stop}^j, c_{arr\_stop}^j, c_{dep\_time}^j, c_{arr\_time}^j, c_{trip}^j)$  be the connection at the deviation index, the algorithm removes the prefix stations, i.e., each station vis-



---

**Algorithm 3** EarliestArrivalDetour( $J, i, M$ )

---

- 1:  $c^i \leftarrow$  the  $i^{th}$  connection of  $J$
  - 2:  $c_{arr\_stop} \leftarrow$  the arrival stop of  $c^{i-1}$
  - 3:  $c_{arr\_time} \leftarrow$  the arrival time of  $c^{i-1}$
  - 4:  $C_{dev} \leftarrow \{c' \text{ s.t. there is a journey } J' \text{ in } Output \text{ starting with } (c^0, \dots, c^{i-1}, c')\}$
  - 5:  $C^N = \{c' \in C \text{ s.t. } c'_{dep\_stop} = c_{arr\_stop}, c'_{dep\_time} \geq c_{arr\_time} \text{ and } c' \notin C_{dev}\}$
  - 6:  $c^{LB} \leftarrow$  a connection in  $C^N$  leading to a minimum arrival time from  $c_{arr\_stop}$  to  $d$  after  $c_{arr\_time}$  following  $M$
  - 7:  $F_{dev} \leftarrow \{f \text{ s.t. there is a journey } J' \text{ in } Output \text{ starting with } (c^0, \dots, c^{i-1}, f)\}$
  - 8:  $F^N = \{f \in F \text{ s.t. } f_{dep\_stop} = c_{arr\_stop} \text{ and } f \notin F_{dev}\}$
  - 9:  $f^{LB} \leftarrow$  a footpath in  $F^N$  leading to a minimum arrival time from  $c_{arr\_stop}$  to  $d$  following  $M$
  - 10:  $J_{c^{LB}} \leftarrow c^{LB} \cdot M_{c_{arr\_stop}, d}^{c_{arr\_time}^{LB}, t_{max}^{LB}}$
  - 11:  $J_{f^{LB}} \leftarrow f^{LB} \cdot M_{f_{arr\_stop}, d}^{c_{arr\_time}^{f^{LB}} + f_{dur}^{f^{LB}}, t_{max}^{f^{LB}}}$
  - 12:  $J_{min} \leftarrow$  the earliest arrival journey among  $J_{c^{LB}}$  and  $J_{f^{LB}}$
  - 13:  $\pi = (f^0, c^0, \dots, c^{i-1})$
  - 14:  $J_{new} \leftarrow \pi \cdot J_{min}$
  - 15: **return**  $J_{new}$
- 

ited by one of the connections  $c^0, \dots, c^{j-1}$ , from  $\mathcal{T}$ . Also, for each journey  $J'$  in *Output* starting with the connections  $c^0, c^1, \dots, c^{j-1}, c'$ , the connection  $c'$  is removed from  $\mathcal{T}$ . Then, using the CSA, PY-PT algorithm computes an earliest arrival journey  $Q = (f_Q^0, c_Q^0, \dots, f_Q^{\ell-1}, l_Q^\emptyset, f_Q^\ell)$  from  $c_{arr\_stop}^{j-1}$  to  $d$  with  $c_{arr\_time}^{j-1}$  as departure time. Let  $J_{new}$  be the concatenation of the prefix of  $J$  and  $Q$ , i.e.,  $J_{new} = (f^0, c^0, \dots, c^{j-1}, f_Q^0, c_Q^0, \dots, f_Q^\ell)$ . The journey  $J_{new}$  is added to the *Candidates* with  $j$  as deviation index and with  $\zeta = 1$  (as  $J_{new}$  is simple).

The PY-PT algorithm repeats this process until  $k$  journeys are added to *Output*.

Now, let us explain how the journey  $J_{new}$  is computed (in the case where  $\zeta = 1$ ). The pseudocode of this procedure is described in Algorithm 3. Let  $c^i = (c_{dep\_stop}^i, c_{arr\_stop}^i, c_{dep\_time}^i, c_{arr\_time}^i, c_{trip}^i)$  be the  $i^{th}$  connection of  $C_J$  (for  $j \leq i < \phi$ ), the following procedure is applied:

- First, the algorithm scans the connections starting with  $c_{arr\_stop}^i$  after  $c_{arr\_time}^i$  leading to new journeys, i.e., different from those in *Output*. Precisely, let  $C_{dev} = \{c_{old} \in C \text{ s.t. there is a journey in } Output \text{ starting with the connections } c^0, \dots, c^{i-1}, c_{old}\}$ , let  $C^N = \{c \in C \text{ s.t. } c_{dep\_stop} =$

$c_{dep\_stop}^i, c_{dep\_time} \geq c_{dep\_time}^i \text{ and } c \notin C_{dev}\}$  be the set of new deviating connections. The algorithm scans the connections of  $C^N$ . Let  $c^{LB}$  be a connection of  $C^N$  leading to an earliest arrival journey from  $c_{dep\_stop}^i$  to  $d$  using  $M$ . Formally, for each  $c$  in  $C^N$ , let  $J_c$  be the journey via  $c$  following  $M$ , i.e., let  $J_c = c \cdot M_{c_{arr\_stop}, d}^{c_{arr\_time}^c, t_{max}^c}$ , then  $c^{LB}$  is a connection in  $C^N$  s.t.  $arr_t(J_{c^{LB}}) \leq arr_t(J_c)$  for each  $c$  in  $C^N$ .

- Second, the algorithm scans the footpaths starting with  $c_{dep\_stop}^i$  leading to new journeys, i.e., different from those in *Output*. Again, let  $F_{dev} = \{f_{old} \text{ s.t. there is a journey in } Output \text{ starting with the connections } c^0, \dots, c^{i-1} \text{ followed by } f_{old}\}$ , let  $F^N = \{f \in F \text{ s.t. } f_{dep\_stop} = c_{dep\_stop}^i \text{ and } f \notin F_{dev}\}$  be the set of the new deviating footpaths and let  $f^{LB}$  be a footpath of  $F^N$  leading to an earliest arrival journey from  $c_{dep\_stop}^i$  to  $d$  using  $M$ . Precisely, for each  $f$  in  $F^N$ , let  $J_f$  be the journey via  $f$  following  $M$ , i.e.,  $J_f = f \cdot M_{f_{arr\_stop}, d}^{c_{dep\_time}^f + f_{dur}^{f^{LB}}, t_{max}^{f^{LB}}}$ , then  $f^{LB}$  is a footpath in  $F^N$  s.t.  $arr_t(J_{f^{LB}}) \leq arr_t(J_f)$  for each  $f$  in  $F^N$ .

Now let  $Q_{min}$  be the journey with minimum arrival time among  $J_{c^{LB}}$  and  $J_{f^{LB}}$  and let  $J_{min}$  be the journey formed by the concatenation of the prefix journey of  $J$  and  $Q_{min}$ , i.e.,  $J_{min} = (f^0, c^0, \dots, c^{i-1}, Q_{min})$ . Note that,  $J_{min}$  may not be simple as the sub-journey extracted from  $M$  may revisit a station of one of the prefix connections. For instance, a station that is visited by  $c^0$  or  $c^1, \dots$ , or  $c^{i-1}$  may be visited again by  $J_{c^{LB}}$  (or by  $J_{f^{LB}}$ )<sup>4</sup>.

To conclude, in contrast with Y-PT algorithm where an earliest arrival simple detour is computed at each index of an extracted journey using the CSA, PY-PT algorithm consider an earliest arrival detour (not necessarily simple) given by the already computed PCSA at each index, and two cases are distinguished: If the earliest arrival detour is simple, then a CSA call is saved and a shortest simple detour is added to *Candidates*. If not, i.e., the earliest arrival de-

<sup>3</sup>If the element right before  $c^i$  is a footpath, i.e.,  $J = (f^0, \dots, f^\lambda, c^i, \dots, f^r)$ . It is possible to have journeys with two consecutive footpaths. In order to avoid such scenario, the footpaths starting with  $c_{arr\_stop}^i$  will not be scanned.

<sup>4</sup>When scanning the connections starting with  $c_{arr\_stop}^i$  after  $c_{arr\_time}^i$ , the journey  $M_{c_{arr\_stop}, d}^{c_{arr\_time}^i, t_{max}^i}$  can start either with a self loop footpath or a footpath. On the other hand, when scanning footpaths starting with  $c_{dep\_stop}^i$  the journey

$M_{f_{arr\_stop}, d}^{c_{dep\_time}^i + f_{dur}^{f^{LB}}, t_{max}^{f^{LB}}}$  cannot start with anything other than a self loop footpath, to do so the PCSA stores journeys in two separate data structures, one for journeys starting with a self loop footpath and one for the other journeys.

tour is not simple, PY-PT algorithm inserts this non-simple detour to the set of *Candidates* with a flag indicating that it is not simple. Recall that journeys in *Candidates* are non-decreasingly stored by their arrival time. So, only when this non-simple detour is extracted from *Candidates*, its simple version will be computed using the CSA. In other words, the actual computation of such simple detour is “postponed”. Such postponement may end up saving some CSA calls, typically when  $k$  earliest arrival journey have been added to *Output* while some non-simple journeys remain in *Candidates*, i.e, their whole “repair” procedure is skipped.

Note that, despite these postponements, the order of extraction of simple journeys from *Candidates* remains valid. This is because a journey  $J$  in *Candidate* is either inserted with its real arrival time (the case where  $J$  is simple) or with a lower bound on its arrival time (the case where  $J$  is non-simple, by Claim 1).

**Claim 1.** Let  $J = (f^0, c^0, f^1 \dots, c^\phi, f^r)$  be an  $o$ - $d$  journey with  $J_{ns}$  an earliest arrival detour of  $J$  at  $i$  and  $J_s$  with an earliest simple arrival detour of  $J$  at  $i$  (where  $0 \leq i \leq \phi$ ). Then,  $arr_t(J_{ns}) \leq arr_t(J_s)$

*Proof.* The proof follows from the fact that an earliest arrival detour of  $J$  at  $i$  arrives earlier than any detour of  $J$  at  $i$ . In particular, it arrives earlier than any earliest arrival simple detour of  $J$  at  $i$ .  $\square$

## 6 EXPERIMENTAL EVALUATION

In this section, we start describing our implementation and settings (Section 6.1). Then we discuss our experimental results on train and public transit networks (Section 6.2).

### 6.1 Experimental Settings

Here we describe the details of the implementation and the setting used in our experiments.

We have implemented Y-PT and PY-PT algorithms in Java and our code is publicly available at (Al Zoobi and Finkelstein, 2021).

Note that in our implementations the parameter  $k$  is not part of the input, this enables the use of these methods as iterators, able to return a next earliest arrival itinerary as long as one exists. Despite the fact that some additional optimizations could be added to the implementation if  $k$  is a part of the input.

**Networks Setting** We have evaluated the performances of our algorithms on two train networks (Germany and Switzerland) and three public transit net-

Table 1: Characteristics of the PT networks: number of stops, connections, lines, trips and footpaths.

Network	Stops	Connections	Lines	Trips	Footpaths
Germany	74 398	3 601 420	3 599	168 024	599 284
Switzerland	29 844	2 599 675	5 645	248 826	27 202
Paris	44 534	3 209 401	1 864	150 963	502 291
Berlin	28 651	1 379 755	1 296	63 569	62 456
Stockholm	14 258	703 326	664	34 799	22 138

Table 2: Running time (s) of the algorithms on PT networks, ( $k = 100$ )

		Germany	Switzerland	Paris	Berlin	Stockholm
Y-PT	avg	94.6	42.0	66	22.7	7.2
	med	47.3	30.6	25.1	14	3.5
PY-PT	avg	3.6	1.9	5.4	0.8	0.2
	med	1.7	1.4	3.8	0.5	0.1

works (Paris, Berlin and Stockholm). The characteristics of these networks are presented in Table 1. This dataset is publicly available via a GTFS feed (<https://transitfeeds.com/>), we downloaded this dataset in October 2019.

The public transit networks are denser than the train networks, i.e. the connections to stops ratio is smaller on train networks than public transit networks. This can be easily explained because the train networks can only use trains whereas the public transit networks can use buses, trains, ferries and many other means of transportation. Therefore, we will show the performances of our algorithms on those two types of networks.

In our experiments, we have randomly chosen 1000 source-destination pairs (of stops) for each public transit network, and we have run each algorithm for each pair for  $k$  going from 2 to 100.

We have measured the execution time and the number of CSA calls. Note that the number of CSA calls is an indication of the running time which is independent of the implementation and the architecture of the machine.

All reported computations have been performed on computers equipped with an Intel(R) Core(TM) i7-1185G7 at 3.00GHz and 32 GB of RAM.

Table 3: Number of CSA calls using each of the algorithms on PT networks, ( $k = 100$ )

		Germany	Switzerland	Paris	Berlin	Stockholm
Y-PT	avg	2132	2158	1355	1788	2072
	med	1729	1749	1262	1604	1510
PY-PT	avg	32	77	39	7.6	8.3
	med	12	56	26	7	2

## 6.2 Experimental Results

In this section, we describe and analyse our experimental results on public transit networks.

We have measured the average and the median of the algorithms' running time in the considered networks. The data (the running time and the number of CSA calls) in Tables 2 and 3 and fig. 3 correspond to the biggest experienced value of  $k$  ( $k = 100$ ). While the data in Figure 2 corresponds to their evolution with respect to the values of  $k$ .

The average and median running times reported in Table 2 show that the PY-PT algorithm is significantly faster than the Y-PT algorithm for every considered network (the average speed up of the running time is bigger than a factor of 10 for  $k = 100$ ). Moreover, a refined comparison on the Germany and Paris networks (Figure 3) shows that PY-PT is faster than Y-PT for almost all queries. In addition, Figures 2a and 2b shows that this speed up remains considerable even for small values of  $k$  (even for  $k = 2$ ) for Stockholm and Switzerland networks. This means that the time consumed for the PCSA computation routine is compensated by the extraction of simple detours, even for  $k = 2$ . In addition, very similar results were obtained on the remaining networks. Based on these remarks, we conclude that, in practice, PY-PT is faster than Y-PT for almost every scenario (the value of  $k$ , the query specifications and the network structure).

Furthermore, on the Stockholm and Switzerland networks, Table 3 and figs. 2c and 2d show that the number of CSA calls is significantly reduced using PY-PT. This ensures that a similar speed up is guaranteed for any experimental settings (Johnson, 2002).

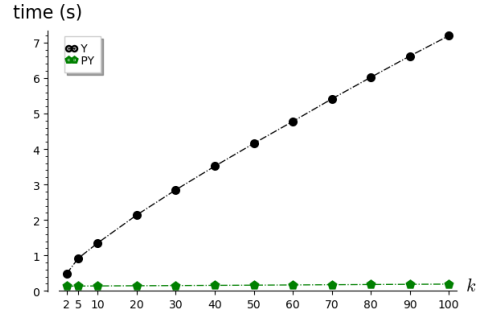
As the obtained results are similar, we only displayed data obtained from experiments on selected networks (Stockholm and Switzerland for Figure 2, Paris and Germany for Figure 3). However, the results/plots corresponding to the remaining networks are very similar.

To conclude, on average, PY-PT algorithm is more than 10 times faster than Y-PT algorithm, it is also faster than Y-PT for almost every scenario.

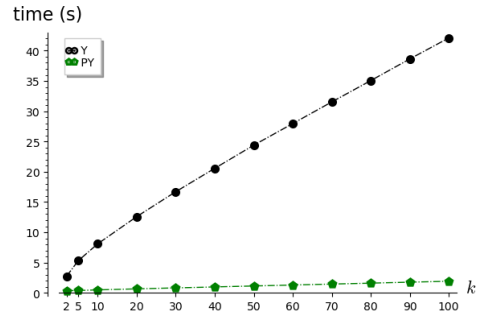
## 7 (DIS)SIMILARITY OF THE OUTPUT JOURNEYS

In this section, we experimentally evaluate the dissimilarity of the journeys extracted by our algorithms on the described dataset in order to check whether our algorithms can be used to extract dissimilar journeys.

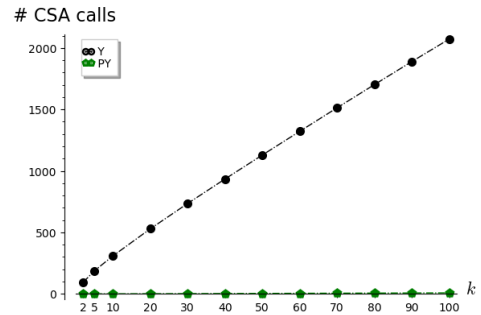
Roughly, two journeys are similar if they share a major part of their connections and footpaths. Here,



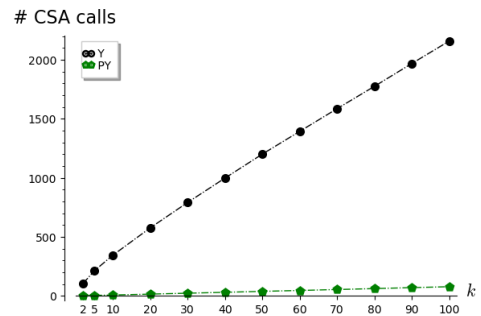
(a) Average running time on Stockholm



(b) Average running time on Switzerland

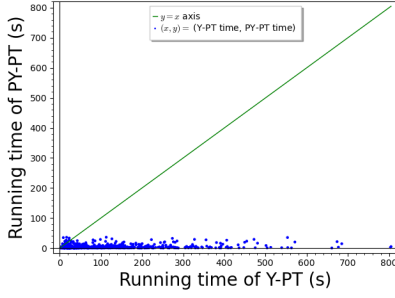


(c) Average number of CSA calls on Stockholm

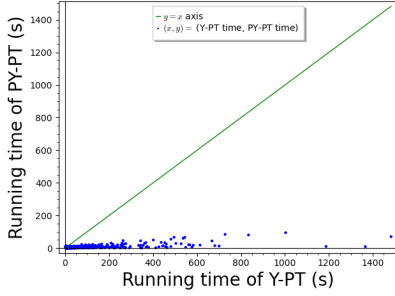


(d) Average number of CSA calls on Switzerland

Figure 2: The running time of the  $k$ EAT algorithms on Switzerland train network and Stockholm public transit network with respect to the values of  $k$



(a) Running time of Y-PT and PY-PT on Germany



(b) Running time of Y-PT and PY-PT on Paris

Figure 3: Comparison of the running time of Y-PT and PY-PT on a train network and a public transit network

analogously with the Jaccard similarity measure of paths in graphs, that measures the similarity of two paths by the ratio of the length of the arcs they have in common over the length of their union (eq. 1) (Chondrogiannis et al., 2017; Al Zoobi et al., 2021b).

$$S(P, Q) = \frac{\ell(P \cap Q)}{\ell(P \cup Q)} \quad (1)$$

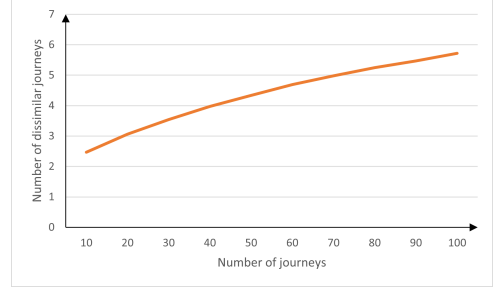
We define the similarity of two journeys as the ratio between the travel time of their connections and footpaths in common over the travel time of the union of their connections and footpaths. Precisely, we define the similarity of two journeys as follows.

Let  $c = (c_{dep\_stop}, c_{arr\_stop}, c_{dep\_time}, c_{arr\_time}, c_{trip})$  be a connection, we define the travel time of  $c$  as the time consumed by  $c$ , i.e.,  $dur(c) = c_{arr\_time} - c_{dep\_time}$ . Now, let  $J_1, J_2$  be two journeys, the similarity between  $J_1$  and  $J_2$  can be measured as follows.

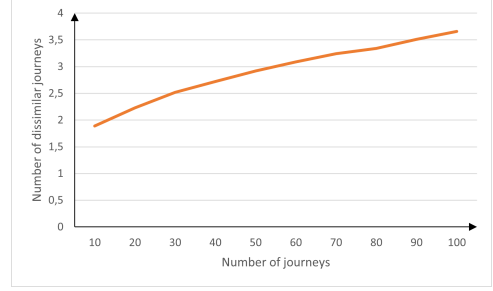
$$S_{PT}(J_1, J_2) = \frac{\sum_{c \in J_1 \cap J_2} dur(c) + \sum_{f \in J_1 \cap J_2} f_{dur}}{\sum_{c \in J_1 \cup J_2} dur(c) + \sum_{f \in J_1 \cup J_2} f_{dur}} \quad (2)$$

So, given a threshold value  $\theta \in [0, 1]$ , two journeys  $J_1$  and  $J_2$  are called  $\theta$  dissimilar if their similarity does not exceed  $\theta$ , i.e., if  $S_{PT}(J_1, J_2) \leq \theta$ .

In order to evaluate the number of dissimilar journeys given by our algorithms, we measured the number of journeys that are  $\theta$  dissimilar among the first 100 journeys given by our algorithm. As shown in



(a) Paris



(b) Germany

Figure 4: Number of paths that are 0.5 dissimilar among those given by our algorithms

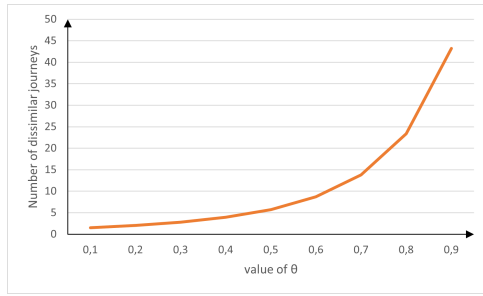
Figure 4, our algorithms can be used to extract, on average, 3 to 5 journeys that are 0.5 dissimilar for the Paris network. Moreover, Figure 5 shows the number of dissimilar journeys among the 100 journeys given by our algorithms with respect to the value of the similarity measure  $\theta$ . We notice that, starting from  $\theta = 0.8$ , the number of dissimilar journeys starts to be considerable (more than 15 journeys).

The plots presented in this section corresponds to the results obtained on the Paris and Germany networks. However, very similar results were obtained on the remaining networks.

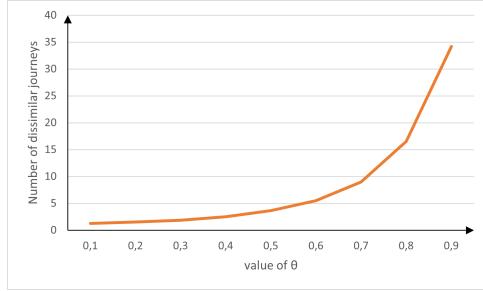
Summarizing, after measuring the similarity of the journeys given by our algorithms in practice, we claim that algorithms Y-PT and PY-PT can be used as a tool to extract journeys that are reasonably dissimilar.

## 8 CONCLUSION

In this paper, we have explored alternative journey planning in public transit networks, offering a vast set of interesting solutions. This is done by adapting the  $k$  shortest simple paths problem to the public transit network context. We proposed a straightforward adaptation of Yen's algorithm and a more refined version answering the proposed problem in a reasonable running time. Finally, we evaluated the similarity of the output journeys and showed that our algorithms can be used to extract dissimilar journeys.



(a) Paris



(b) Germany

Figure 5: Number of dissimilar paths among the first 100 paths given by our algorithms with respect to the variation of the similarity threshold  $\theta$

Interesting questions are asked about designing algorithms answering  $k$  earliest arrival journeys query faster. Whether by improving / proposing faster methods than PY-PT algorithm, or even with the help of a preprocessing routine. For instance, a more specific question is whether one can use journey planning algorithms like Transfer Patterns algorithm (Bast et al., 2010) to answer  $k$  earliest arrival journeys queries.

Another interesting question concerns the design of algorithms dedicated to extracting dissimilar journeys in public transit networks. This could be done by adapting some of the algorithms proposed to find shortest dissimilar paths in a graph (Chondrogiannis et al., 2017).

## REFERENCES

Al Zoobi, A., Coudert, D., and Nisse, N. (2020). Space and time trade-off for the  $k$  shortest simple paths problem. In *18th International Symposium on Experimental Algorithms (SEA)*, volume 160, page 13. Schloss Dagstuhl–Leibniz-Zentrum für Informatik.

Al Zoobi, A., Coudert, D., and Nisse, N. (2021a). Finding the  $k$  Shortest Simple Paths: Time and Space trade-offs. Research Report hal-03196830, Inria ; I3S, Université Côte d’Azur.

Al Zoobi, A., Coudert, D., and Nisse, N. (2021b). On the complexity of finding  $k$  shortest dissimilar paths in a graph. Research Report hal-03187276, Inria ; CNRS

; I3S ; UCA.

Al Zoobi, A. and Finkelstein, A. (2021). PT-KSSP github repository. Available at <https://github.com/fink-arthur/PT-KSSP>.

Barrett, C., Bisset, K., Holzer, M., Konjevod, G., Marathe, M., and Wagner, D. (2008). Engineering label-constrained shortest-path algorithms. In *International conference on algorithmic applications in management*, pages 27–37. Springer.

Bast, H., Carlsson, E., Eigenwillig, A., Geisberger, R., Harrelson, C., Raychev, V., and Viger, F. (2010). Fast routing in very large public transportation networks using transfer patterns. In *18th Annual European Symposium on Algorithms (ESA)*, pages 290–301.

Bast, H., Delling, D., Goldberg, A., Müller-Hannemann, M., Pajor, T., Sanders, P., Wagner, D., and Werneck, R. F. (2016). Route planning in transportation networks. In *Algorithm engineering*, pages 19–80. Springer.

Chondrogiannis, T., Bouros, P., Gamper, J., and Leser, U. (2017). Exact and approximate algorithms for finding  $k$ -shortest paths with limited overlap. In *20th International Conference on Extending Database Technology (EDBT)*, pages 414–425.

Delling, D., Pajor, T., and Werneck, R. F. (2015). Round-based public transit routing. *Transportation Science*, 49(3):591–604.

Dibbelt, J., Pajor, T., Strasser, B., and Wagner, D. (2018). Connection scan algorithm. *ACM Journal of Experimental Algorithmics (JEA)*, 23:1–56.

Eppstein, D. (1998). Finding the  $k$  shortest paths. *SIAM Journal on Computing*, 28(2):652–673.

Johnson, D. S. (2002). A theoretician’s guide to the experimental analysis of algorithms. *Data structures, near neighbor searches, and methodology: fifth and sixth DIMACS implementation challenges*, 59:215–250.

Kurz, D. and Mutzel, P. (2016). A sidetrack-based algorithm for finding the  $k$  shortest simple paths in a directed graph. In *27th International Symposium on Algorithms and Computation (ISAAC)*, volume 64 of *LIPICs*, pages 49:1–49:13. Schloss Dagstuhl - Leibniz-Zentrum für Informatik.

Scano, G., Huguet, M.-J., and Ngueveu, S. U. (2015). Adaptations of  $k$ -shortest path algorithms for transportation networks. In *International Conference on Industrial Engineering and Systems Management (IESM)*, pages 663–669. IEEE.

Schulz, F., Wagner, D., and Weihe, K. (2000). Dijkstra’s algorithm on-line: An empirical case study from public railroad transport. *ACM Journal of Experimental Algorithmics (JEA)*, 5:12–es.

Vo, K. D., Pham, T. V., Nguyen, H. T., Nguyen, N., and Van Hoai, T. (2015). Finding alternative paths in city bus networks. In *2015 International Conference on Computer, Control, Informatics and its Applications (IC3INA)*, pages 34–39. IEEE.

Yen, J. Y. (1971). Finding the  $k$  shortest loopless paths in a network. *Management Science*, 17(11):712–716.