Kujala, Rainer; Weckström, Johan; Mladenovic, Milos; Saramäki, Jari

**Travel times and transfers in public transport: Comprehensive accessibility analysis based on Pareto-optimal journeys**

Published in:
COMPUTERS ENVIRONMENT AND URBAN SYSTEMS

**DOI:**
10.1016/j.compenvurbsys.2017.08.012

Published: 01/01/2018

*Please cite the original version:*
Travel times and transfers in public transport: Comprehensive accessibility analysis based on Pareto-optimal journeys

Rainer Kujala\textsuperscript{a,}\textsuperscript{*}, Christoffer Weckström\textsuperscript{b}, Miloš N. Mladenović\textsuperscript{b}, Jari Saramäki\textsuperscript{a}

\textsuperscript{a}Department of Computer Science, Aalto University, Finland
\textsuperscript{b}Department of Built Environment, Aalto University, Finland

\textbf{ARTICLE INFO}

Article history:
Received 3 March 2017
Received in revised form 28 August 2017
Accepted 28 August 2017
Available online xxxx

Keywords:
Transit
Routing
Temporal distance profile
Temporal network
GTFS
OpenStreetMap

\textbf{ABSTRACT}

Efficient public transport (PT) networks are vital for well-functioning and sustainable cities. Compared to other modes of transport, PT networks feature inherent systemic complexity due to their schedule-dependence and network organization. Because of this, efficient PT network planning and management calls for advanced modeling and analysis tools. These tools have to take into account how people use PT networks, including factors such as demand, accessibility, trip planning and navigability. From the PT user perspective, the common criteria for planning trips include waiting times to departure, journey durations, and the number of required transfers. However, waiting times and transfers have typically been neglected in PT accessibility studies and related decision-support tools. Here, we tackle this issue by introducing a decision-support framework for PT planners and managers, based on temporal networks methodology. This framework allows for computing pre-journey waiting times, journey durations, and number of required transfers for all Pareto-optimal journeys between any origin–destination pair, at all points in time. We visualize this information as a \textit{temporal distance profile}, covering any given time interval. Based on such profiles, we define the best-case, mean, and worst-case measures for PT travel time and number of required PT vehicle boardings, and demonstrate their practical utility to PT planning through a series of accessibility case studies. By visualizing the computed measures on a map and studying their relationships by performing an all-to-all analysis between 7463 PT stops in the Helsinki metropolitan region, we show that each of the measures provides a different perspective on accessibility. To pave the way towards more comprehensive understanding of PT accessibility, we provide our methods and full analysis pipeline as free and open source software.

© 2017 The Authors. Published by Elsevier Ltd. This is an open access article under the CC BY license (http://creativecommons.org/licenses/by/4.0/).

\textbf{1. Introduction}

Efficient, easy-to-use public transport (PT) networks are a vital element of functional, sustainable cities (Banister, 2008; Newman & Kenworthy, 1989). If planned carefully, PT is a space efficient transport mode with low emission levels, offering mobility for users spanning all ages and income levels (Church, Frost, & Sullivan, 2000). One prerequisite for good PT network planning is a set of tools and measures for evaluating PT network designs. In particular, tools for measuring PT travel impedance are required, as they help practitioners identify potential problems, such as poor connectivity, and assess the impacts of public transport investments and network redesigns.

Among urban transport modes, PT has three distinguishing features that make the assessment of travel impedance difficult. First, PT journeys are usually multi-modal, as a completed journey requires access and egress legs with another mode, typically walking. Second, unlike other modes, PT is a scheduled service that offers connections between stops only at specific points in time. Third, PT provides services through a network that should operate efficiently while maintaining significant spatial coverage. These PT features are also transferred to the passenger perspective. Common factors affecting PT user experience include waiting times to departure, access and egress walking distances, journey durations, and the number of required transfers.

The challenges in assessing PT travel impedance have resulted in a variety of analysis frameworks. While some studies have used static representations of PT networks for computing travel times (Curtis & Scheurer, 2010; Delmelle & Casas, 2012; Mavoa, Witten, McCreanor, & O’Sullivan, 2012; O’Sullivan, Morrison, & Shearer, 2000; Tribby & Zandbergen, 2012) and the number of required vehicle boardings (Hadas & Ranjitkar, 2012; Wang & Yang, 2011), the recent trend has been towards more accurate modeling of travel

\textsuperscript{*} Corresponding author.
E-mail address: rainer.kujala@aalto.fi (R. Kujala).

http://dx.doi.org/10.1016/j.compenvurbsys.2017.08.012
0198-9715 © 2017 The Authors. Published by Elsevier Ltd. This is an open access article under the CC BY license (http://creativecommons.org/licenses/by/4.0/).
times using actual PT schedules with information on departure and arrival times (Benenson, Ben-Elia, Rofé, & Geyzersky, 2017; Benenson, Martens, Rofé, & Kwartler, 2011; Farber & Fu, 2017; Farber, Morang, & Widener, 2014; Lei & Church, 2010; Salonen & Toivonen, 2013). To address the dynamic nature of PT travel time, travel times have been computed at different times of day with time resolutions as high as 1 min (Farber & Fu, 2017; Farber et al., 2014; Owen & Levinson, 2015). This methodology has enabled meaningful computation of the minimum and maximum travel times together with the estimation of typical service headways using Fourier analysis of the travel time profile (Farber & Fu, 2017). Moreover, the spatial resolution of travel time analyses has been increasing, and recently door-to-door travel times have been computed even at the level of individual buildings (Benenson et al., 2017).

Despite the ongoing progress, previous research leaves room for methodological improvements in assessing PT travel impedance. Especially, we have identified two areas of improvement related to quantifying pre-journey waiting times, journey durations, and transfers, which are known to cause discomfort to PT users (Iseki & Taylor, 2009; Litman, 2008; Wardman, 2004). First, how PT travel time is measured varies across studies and it is typically considered single-faceted. While some studies only aim to capture the journey duration (Benenson et al., 2017; Salonen & Toivonen, 2013; Tenkanen, Heikinheimo, Järv, Salonen, & Toivonen, 2016), others include the pre-journey waiting time as part of PT travel time (Farber & Fu, 2017; Farber et al., 2014; Lei & Church, 2010; Owen & Levinson, 2015). The former approach effectively assumes that the PT user plans her travel according to schedules, while the latter assumes that travel takes place spontaneously. Despite this, there has been little discussion on the differences of these two alternative definitions of PT travel time.

The second area of improvement relates to quantifying the required number of transfers between an origin-destination pair. Even though transfers are an integral part of PT travel impedance, there are no studies quantifying the number of PT vehicle boardings between origin-destination pairs that would fully take the time-dependence of PT operations into account.

One potential reason why the above aspects of PT travel impedance have not been considered before might be rooted in the methodology used by most PT accessibility studies. In particular, many studies rely on Dijkstra’s algorithm for computing travel times in the PT network (Dijkstra, 1959). However, Dijkstra’s algorithm can only optimize PT travel time while it ignores the number of required transfers. Using Dijkstra’s algorithm also necessitates that PT travel times are sampled, i.e., travel times computed only at certain departure times. Even though sampling yields an approximate picture of the dynamic travel time profile, disentangling pre-journey waiting times from journey durations remains difficult.

These challenges can be overcome by realizing that PT travel times and numbers of required boardings are determined by the journey alternatives enabled by the PT network, assessed through the concept of Pareto-optimality. Pareto-optimality can be explained with a simple example. Let us assume that a PT user is traveling from an origin O to destination D at time t, and compares PT journey alternatives. Further, let us assume that her decision-making criteria only include the time to reach the destination (\(t_{\text{arr}}-t\)) and the number of PT vehicles (b) she needs to board. Then, each PT journey alternative can be summarized as a tuple \((t_{\text{arr}}-t, b)\). If the user prefers to reach her destination fast and dislikes transfers, i.e., prefers small values of \(t_{\text{arr}}-t\) and b, her rational choice alternatives correspond to the Pareto-frontier of all journey alternatives, as illustrated in Fig. 1.

The above setting corresponds to spontaneous travel, where the departure time of the travel is pre-determined. However, in reality a user can plan and adjust her departure time based on PT schedules. Therefore, the departure times \(t_{\text{dep}}\) of the journey alternatives should be taken into account too. Then, PT journeys are summarized as triplets \((t_{\text{dep}}, t_{\text{arr}}, b)\). To minimize the journey duration \((t_{\text{arr}} - t_{\text{dep}})\), it is now natural to prefer large values of \(t_{\text{dep}}\). Given all journey alternatives, the Pareto-frontier contains the fastest journey alternatives for reaching the destination with different numbers of boardings, at all departure times. Such sets of Pareto-optimal journey alternatives fully describe the dynamic accessibility between origin-destination pairs in terms of journey durations, pre-journey waiting times, and transfers.

The routing algorithms used in typical PT accessibility studies cannot compute Pareto-optimal journey alternatives over a given time interval. However, many algorithms specifically tailored for PT have been developed recently (Bast et al., 2015; Delling, Pajor, & Werneck, 2012; Dibbelt, Pajor, Strasser, & Wagner, 2013). While the main motivation in their development has been to decrease the response times of on-line journey planners, they can also compute all Pareto-optimal journey alternatives between an origin-destination pair that depart within a given time interval.

However, to the best of the authors’ knowledge, such Pareto-optimal journey alternatives have not been used as the basis of PT accessibility studies, and there is no methodological framework for their analysis. Thus, we develop such a framework based on temporal networks methodology (Gallotti & Barthelemy, 2015; Holme & Saramäki, 2012; Holme & Saramäki, 2013). Especially, we show how sets of Pareto-optimal journey alternatives can be used to construct temporal distance profiles that provide full temporal information on the time to reach a destination over a specified time interval (Pan & Saramäki, 2011). These profiles can be augmented with information on the required numbers of vehicle boardings. Using the temporal distance profiles, we define the best-case, mean, and worst-case measures for PT travel time and the number of required vehicle boardings. Additionally, we study the trade-offs between travel time and the required number of vehicle boardings.

Regarding our analysis pipeline, we adopt an open science approach in terms of data and software. For PT timetables, we use data provided in the General Transit Feed Specification (GTFS) format, and for computing the walking network between PT stops, we rely on open data provided by the OpenStreetMap project (OpenStreetMap contributors, 2017). Moreover, we provide our full analysis pipeline as free and open source software.

To demonstrate the utility of our methodology for PT planning, we discuss a series of accessibility case studies in the Helsinki metropolitan area. Through temporal distance profiles and map visualizations, we show how each of the suggested measures can be useful depending on the focus of the analysis – each measure provides a different perspective on accessibility. Finally, we perform an all-to-all analysis between the 7463 PT stops in the Helsinki metropolitan area. Through temporal distance profiles and map visualizations, we show how each of the suggested measures can be useful depending on the focus of the analysis – each measure provides a different perspective on accessibility. Finally, we perform an all-to-all analysis between the 7463 PT stops in the Helsinki metropolitan area.
metropolitan area, revealing general relationships between the different definitions on PT travel time and the number of required PT vehicle boardings.

2. Methods

In this section, we first introduce our main methodological contributions to the analysis of Pareto-optimal journey alternatives. As the first step, we show how to construct a fastest-path temporal distance profile from a set of schedule-based Pareto-optimal journey alternatives when only the departure and arrival times of the journeys are considered. Building on these profiles, we provide definitions for the minimum, mean and maximum time to reach a destination. Then, we augment the journey alternatives with information on the number of required vehicle boardings, leading to boarding-count-augmented temporal distance profiles. Based on these, we define measures for quantifying the number of vehicle boardings required for reaching the destination, and measures for the trade-offs between the number of vehicle boardings and the previously defined temporal distance statistics. To provide schematic examples of the temporal distance profiles and fastest-path distributions, we discuss fictitious PT services between an origin-destination pair shown in Fig. 2, for which we have listed all available journey alternatives in Table 1. Last, we describe how we compute the sets of Pareto-optimal journey alternatives based on GTFS and OpenStreetMap data, and describe our analysis pipeline in more detail.

2.1. Fastest-path temporal distance profiles

The question of how much time it takes to reach a destination from an origin in a PT network turns out to be more intricate than what it seems at first sight. The origin of this intricacy lies in the schedule-dependence of PT operations. In addition to the actual time spent traveling, a PT user may need to wait at the origin before departing for the journey. How the user experiences this waiting time depends on several factors: e.g. whether the user simply goes to the nearest stop and waits for the vehicle, or plans the journey ahead based on known schedules. If the user travels spontaneously, to the nearest stop and waits for the vehicle, or plans the journey ahead based on known schedules. If the user travels spontaneously, to the nearest stop and waits for the vehicle, or plans the journey ahead based on known schedules. If the user travels spontaneously, to the nearest stop and waits for the vehicle, or plans the journey ahead based on known schedules.

To avoid ambiguity between different interpretations of “PT travel time”, we adopt the following terminology. We use the term journey duration to describe the actual origin–destination journey time including the access and egress walking legs, and the term temporal distance to describe the sum of the journey duration and the pre-journey waiting time. Further, a journey is a fastest-path journey if at some point in time it is the fastest way for reaching the destination.

To quantify journey durations and temporal distances between origin–destination pairs, information on the fastest-path journey alternatives is required at all points in time. Now, at a given travel departure time $t$, one has to identify the journey alternative that reaches the destination fastest, i.e. the journey that has the smallest arrival time $t_{arr}$. If several journeys arrive at the same time, we define the optimal journey as the one that has the latest departure time because it minimizes the time spent traveling. Thus, one needs to simultaneously optimize for late departure time and early arrival time.

When considering all fastest-path journey alternatives at all points in time, each journey is Pareto-optimal in terms of departure time $t_{dep}$ (larger better) and arrival time $t_{arr}$ (smaller better). Now, a journey alternative $a$ is Pareto-optimal, if there is no other journey alternative $\beta$ that is better than $a$ both in terms of $t_{dep}$ and $t_{arr}$.

To illustrate how the concept of Pareto-optimality works, let us consider journeys $B(t_{dep} = 08:08, t_{arr} = 08:28)$ and $F(t_{dep} = 08:19, t_{arr} = 08:39)$ listed in Table 1. Now, even though journey $B$ arrives at the destination earlier than $F$, journey $B$ does not dominate journey $F$ as it departs earlier than $F$. In fact, neither of the journeys are dominated by any other journey listed in Table 1 in terms of $t_{dep}$ and $t_{arr}$, and thus they both belong to the set of fastest-path journey alternatives.

To define the temporal distance from origin $o$ to destination $d$ at time $t$, we first identify the next journey $j^*$ from the set of fastest-path journey alternatives $S$:

$$j^* = j^*(t) = \arg\min_{j \in S} \left( t^*_{dep} - t_{dep} \geq t \right).$$

Now the temporal distance from $o$ to $d$ using PT can be written as

$$\tau_{PT}(t) = \tau_{wait}(t) + \tau_{journey}^t = \left( t^*_{dep} - t \right) + \left( t_{arr}^* - t_{dep} \right) = t_{arr}^* - t,$$

where $\tau_{journey}^t$ stands for the duration of journey $j^*$.

Fig. 2. PT services between an origin–destination pair. Each circle labeled with a letter corresponds to a PT stop. The numbers below travel mode icons indicate the travel duration on the trip segment, and the departure times of PT vehicles are indicated on the right-hand side of the icon. The icons for different travel modes are adapted from Google’s Material Design icon collection ([https://material.io/icons/](https://material.io/icons/)), licensed under Apache License version 2.0.
Above, we have assumed that each PT journey includes at least one PT vehicle boarding. However, walking to destination \( d \) can be a faster alternative than using any of the Pareto-optimal PT journey alternatives. To take this possibility into account, we cap the temporal distance function to the walk duration \( t_{\text{walk}} \) between the origin \( o \) and destination \( d \), and obtain our definition for the fastest-path temporal distance:

\[
\tau(t) = \min \left( \tau_{\text{PT}}(t), \tau_{\text{walk}} \right).
\] (3)

The evolution of temporal distance over a time window ranging from \( t_{\text{start}} \) to \( t_{\text{end}} \) can be visualized as a fastest-path temporal distance profile. Fig. 3a shows a fastest-path temporal distance profile that is constructed from the journeys B, F, H, and I of Table 1. Here, the 60-minute walk to the destination is never the fastest option, and thus the cutoff implied by Eq. (3) is never applied.

### 2.2. Fastest-path temporal distance statistics

A fastest-path temporal distance profile in the interval from \( t_{\text{start}} \) to \( t_{\text{end}} \) can be summarized with a set of temporal distance statistics.

The **minimum temporal distance**

\[
\tau_{\text{min}} = \min_{t \in [t_{\text{start}}, t_{\text{end}}]} \tau(t)
\] (4)
describes the minimum time to travel from origin to destination. Typically, \( \tau_{\text{min}} \) corresponds to the minimum journey duration within the analysis time window, and is thus a good indicator of PT travel time when a PT user plans her departure time well.

The **mean temporal distance**, \( \tau_{\text{mean}} = \frac{1}{t_{\text{end}} - t_{\text{start}}} \int_{t_{\text{start}}}^{t_{\text{end}}} \tau(t) \, dt \) (5) describes the average time to reach the destination when travel takes place spontaneously without planning.

The **maximum temporal distance**, \( \tau_{\text{max}} = \max_{t \in [t_{\text{start}}, t_{\text{end}}]} \tau(t) \), (6) describes the worst-case, or guaranteed, travel time from an origin to a destination. Large values of \( \tau_{\text{max}} \) are thus indicative of service gaps between the origin and the destination.

In addition to the minimum, mean, and maximum temporal distances, their differences enable characterization of service level variations. In this study we focus on the difference \( \tau_{\text{mean}} - \tau_{\text{min}} \) and its scaled variant \((\tau_{\text{mean}} - \tau_{\text{min}})/\tau_{\text{min}}\). The difference \( \tau_{\text{mean}} - \tau_{\text{min}} \) gives information on the general variation in the temporal distance. For completely regular service between \( o \) and \( d \), \( \tau_{\text{mean}} - \tau_{\text{min}} \) corresponds

![Fig. 3. A fastest-path temporal distance profile (a) and its fastest-path temporal distance distribution (b). The temporal distance profile consists of the fastest-path journey alternatives B, F, H and I of Table 1. Note that, as the analysis time window covers only the departure times from \( t_{\text{start}} = 08:00 \) to \( t_{\text{end}} = 08:30 \), journey I departure time is outside of the analysis time window, but nonetheless affects the profile after the departure of journey H. The values for the minimum, mean, and maximum temporal distance are plotted as horizontal lines.](image-url)
to the average pre-journey waiting time or, equivalently, half of the headway between $a$ and $d$. Thus, $\tau_{\text{mean}} - \tau_{\text{min}}$ can be interpreted as the effective waiting time. When this difference is divided by the minimum temporal distance, we obtain an indicator of the effective waiting time’s role with respect to the minimum temporal distance: $\frac{\tau_{\text{mean}} - \tau_{\text{min}}}{\tau_{\text{min}}}$.

To summarize the variation of temporal distance over time, we compute fastest-path temporal distance distributions $P(\tau)$. In Fig. 3b we show the fastest-path temporal distance distribution corresponding to the fastest-path temporal distance profile shown in Fig. 3a.

Fastest-path temporal distance statistics and distributions can be computed exactly without any need for time discretization. In practice, we first split the fastest-path profile into non-overlapping trapezoidal blocks $[B_1, B_2, \ldots, B_M]$ that are determined based on the departure times of the fastest-path journey alternatives and cover the area under the temporal distance profile. Each block $B$ consists of its start and end times, as well as the temporal distance values at those points in time: $B = (\tau_{\text{start}}^B, \tau_{\text{end}}^B, t_{\text{start}}^B, t_{\text{end}}^B)$. Now, e.g. the integral $\int_{\tau_{\text{start}}^B}^{\tau_{\text{end}}^B} \tau(t)dt$ required for the mean temporal distance can be computed by summing up the areas of the trapezoidal blocks:

$$\int_{\tau_{\text{start}}^B}^{\tau_{\text{end}}^B} \tau(t)dt = \sum_{i=1}^{M} \frac{(\tau_{\text{end}}^i - \tau_{\text{start}}^i) (\tau_{\text{start}}^i + \tau_{\text{end}}^i)}{2}.$$ (7)

The fastest-path temporal distance distributions can be computed by first discovering all distinct temporal distance values that appear as the blocks’ start ($\tau_{\text{start}}^i$) or end ($\tau_{\text{end}}^i$) temporal distance values. When ordered, these values define the bins of the temporal distance distribution. Finally, the temporal distance distribution is obtained by computing how many times each of these bins is covered by the temporal distance intervals of the trapezoidal blocks, and normalizing the distribution suitably.

2.3. Boarding-count-augmented temporal distance profiles

In addition to temporal distance, the number of vehicle boardings required to reach a destination is an important factor of PT travel impedance. The number of required boardings to reach destination $d$ from origin $o$ could in principle be computed using a static graph presentation of PT lines. However, the time domain should be taken into account for at least two reasons. First, some PT lines may run during some times of the day only. Second, the path with minimum number of vehicle boardings may not be the most lucrative one, as shorter journey durations may be preferred at the cost of additional vehicle boardings.

To this end, we extend our approach for analyzing Pareto-optimal journey alternatives to also take into account the number of required vehicle boardings $b$ on each journey. Now each journey alternative $j$ is characterized by a triplet $(t_{\text{dep}}^j, t_{\text{arr}}^j, b^j)$, and preference for low values of $b$ is assumed.

To illustrate the effect of considering boarding counts on the Pareto-optimal set of journey alternatives, let us discuss journeys $E = (t_{\text{dep}}^E = 08:13, t_{\text{arr}}^E = 08:50, b^E = 1)$ and $F = (t_{\text{dep}}^F = 08:19, t_{\text{arr}}^F = 08:39, b^F = 2)$ of Table 1. Even though journey $E$ departs earlier and arrives later than journey $F$, it is still included in the set of Pareto-optimal journey alternatives, as $E$ requires only one vehicle boarding while $F$ requires two.

A set of journeys with boarding counts can be visualized as a boarding-count-augmented temporal distance profile. Fig. 4a provides an example profile constructed from the Pareto-optimal journeys of Table 1. The profile allows inspecting all Pareto-optimal journey alternatives for reaching the destination, and investigating their trade-offs. For instance, if one were to depart at 08:00, there would be three Pareto-optimal journey alternatives (journeys $B$, $C$ and $E$ of Table 1) to choose from in addition to walking. These four Pareto-optimal options correspond to the Pareto-frontier of Fig. 1, when $t$ equals 08:00 and pre-journey waiting times are also accounted for. Again, we summarize the fastest-path temporal distance profile as a distribution, also including information on the number of required vehicle boardings (Fig. 4b).

2.4. Boarding-count statistics and time-transfer trade-offs

Based on a boarding-count-augmented temporal distance profile and the associated set of Pareto-optimal journey alternatives, we define three statistics to describe the number of required boardings between an origin and a destination within a time interval $[t_{\text{start}}, t_{\text{end}}]$.

![Fig. 4. A boarding-count-augmented temporal distance profile (a) and its corresponding fastest-path temporal distribution (b). The profile has been created based on the journeys of Table 1, assuming a walk duration of 60 min. Individual journeys are plotted as circles, and the fastest-path temporal distance profile is highlighted using a dashed line. The flat profile with the least number of boardings corresponds to walking ($b_{\text{min}} = 0$).](image-url)
First, the minimum number of vehicle boardings, $b_{\min}$, describes how many vehicle boardings are at least required to reach the destination. More precisely, $b_{\min}$ is defined as the minimum number of vehicle boardings of any discovered journey alternative departing after $t_{\text{start}}$.

Second, the mean number of vehicle boardings on fastest paths, $b_{\text{mean f.p.}}$, describes the expected number of boardings on the next-departing fastest-path journey at time $t$, i.e., $b_{\text{mean f.p.}}$ can be expressed and computed as

$$b_{\text{mean f.p.}} = \frac{1}{t_{\text{end}} - t_{\text{start}}} \int_{t_{\text{start}}}^{t_{\text{end}}} b^{(i)}(t) \, dt$$

where $j_i$ denotes the fastest-path journeys departing between $t_{\text{start}}$ and $t_{\text{end}}$, and $j_n$ is the next fastest-path trip departing after $t_{\text{end}}$.

Finally, the maximum number of vehicle boardings on fastest path, $b_{\text{max f.p.}}$, describes the number of boardings needed in the worst case, i.e., whenever a transfer is possible between the stops by walking. To quantify the trade-offs between the fastest-path temporal distance profile and the profile requiring least boardings, we define the following two measures. First, the difference between $b_{\text{max f.p.}} - b_{\min}$ describes the additional number of vehicle boardings required to reach the destination in the fastest possible time compared to the profile requiring the least number of boardings. Second, the difference $\tau_{\text{mean}, b_{\min}} - \tau_{\text{mean}}$ captures the time saved by choosing the fastest-path journeys instead of journeys requiring the least number of boardings.

2.5. Computation of Pareto-optimal journeys

The presented methodology relies on the provision of Pareto-optimal journey alternatives. There are several algorithms for computing these alternatives (Bast et al., 2015; Delling et al., 2012; Dibbelt et al., 2013). For the purposes of this study, we have implemented the recently-introduced multi-criteria profile connection scan algorithm (mcpCSA) (Dibbelt et al., 2013). Below, we briefly describe how mcpCSA operates and how we have slightly modified it.

mcpCSA models PT timetables as a collection of elementary PT connections each containing information on the departure stop $s_{\text{dep}}$, arrival stop $s_{\text{arr}}$, departure time $t_{\text{dep}}$, arrival time $t_{\text{arr}}$, and the trip id $t$ identifying the PT line and vehicle used. Each connection $c$ can thus be presented as a tuple of five elements $(s_{\text{dep}}, s_{\text{arr}}, t_{\text{dep}}, t_{\text{arr}}, t)$.

In essence, mcpCSA models PT operations as a temporal network consisting of many elementary “events” occurring between nodes (stops) (Holme & Saramäki, 2013). Such temporal networks can be visualized using a node–time diagram, as shown in Fig. 5.

To take into account transfers between stops, transfer connections, or “pseudo-connections” as in Dibbelt et al. (2013), are created whenever a transfer is possible between the stops by walking. In practice, this is done during a fast pre-computation step, taking into account the walking distance and speed and an optional safety margin for transferring between vehicles. In the original description of the mcpCSA algorithm, footpaths are assumed to be transitively closed, meaning that transfers could in practice take place only at certain transfer stations to which multiple PT stops can be associated.

In our implementation of the mcpCSA algorithm, we have adapted the algorithm such that walking transfers are allowed between all stops when the walking distance is below a pre-defined threshold. To disallow long transfers on foot, we enforce that no journey can consist of multiple sequential transfer connections. This required adding several new logical checks in different stages of the original mcpCSA algorithm. The length of the maximum walking distance also strongly affects the number of transfer connections, and thus also the running time of our algorithm. In Fig. 5b, we show the created transfer connections in addition to the original PT connections.

In addition to the combined list of PT and transfer connections, it is necessary to specify the destination node to which access times are computed as well as the start and end times of the routing. Then, the basic idea behind mcpCSA is to scan over the list of connections...
in decreasing order of connection departure time, effectively moving backwards in time. At all times, every stop in the network keeps track of the Pareto-optimal set of journey alternatives for reaching the destination stop. When scanning a connection, the journey alternatives of the connection’s arrival stop are progressed to the connection’s departure stop, which then updates its set of Pareto-optimal journeys. To discover journeys with an access walk leg, a post-processing step is required after scanning of the connections. In this step, each origin node combines its set of journey alternatives with the access-leg-augmented journey alternatives of other nodes that are reachable within the specified walking distance. The Pareto-frontier of this combined pool of journeys yields then the final set of Pareto-optimal journeys for that origin.

Note that instead of specifying destination nodes, the mcpCSA algorithm can also be adjusted to run starting from one or multiple source nodes. In this case, the algorithm would run similarly, but progress forward in time (Dibbelt et al., 2013). For further information on the algorithm, we refer the reader to Dibbelt et al. (2013), and to the source code of our adapted version of the mcpCSA (Section 2.7), where further implementation details can be investigated.

2.6. Analysis pipeline

To construct the list of PT and transfer connections, data on PT timetables, street network, and the locations of origins and destinations are required. Our analysis pipeline relies on timetable data complying with the General Transit Feed Specification (GTFS) standard (Google Inc., 2017), for which data is openly available for many cities. For the computation of the walking distances between stops, we use open data from the Open Street Map project (OpenStreetMap contributors, 2017). As we solely use freely available data sources based on open standards, all our analyses can be easily carried out for any city where GTFS timetable data is available.

In more detail, our analysis pipeline consists of the following steps:

1. Import GTFS data into an SQLite database.
2. For computing door-to-door travel times, add any other origins and destinations into the SQLite database as stops.
3. Compute walking distances between stops and other locations, and add them to the SQLite database.
4. Extract PT connections and footpaths from the database and create transfer connections.
5. Run mcpCSA to obtain Pareto-optimal sets of journeys.
6. Based on the sets of Pareto-optimal sets of journeys, construct temporal distance profiles and statistics and produce visualizations.

A schematic illustration of our analysis pipeline is provided in Fig. 6.

2.7. Software

We provide our complete analysis pipeline as open source software. The main component is the free Python package gtfspy, which can be accessed at http://github.com/CxAalto/gtfspy. With the package, all presented analysis steps can be carried out in Python, except for step 2 for which we have used a router written in Java.

There are other open source alternatives for performing accessibility analyses on PT networks, such as OpenTripPlanner (http://www.opentripplanner.org/) and R5 (http://github.com/conveyal/r5). While these may be more complete in terms of features and faster than our gtfspy package, they have been written using Java. Our gtfspy package benefits from the extensive Python data science ecosystem which enables seamless integration of the computationally intensive mcpCSA runs and the subsequent analyses. Furthermore, Python is fairly easy to learn as compared to Java. Thus we hope that gtfspy will turn out accessible to transport planners and analysts without extensive programming background.

2.8. Setup for this study

The scripts for producing our results are provided freely at http://github.com/rkujala/ptn_temporal_distances. The GTFS data have been obtained through the Reittiopas API provided by the Helsinki Region Transport (2016). While the data downloaded through this API cover multiple weeks, here we limit our analyses to PT operations taking place on a typical Monday (October 3rd, 2016). For the pedestrian routing, an Open Street Map extract covering whole Finland was downloaded (Geofabrik GmbH, 2017). In all our analyses, we use walking speed of 70 m per minute and a 3 min safety margin for transferring between vehicles. These values were chosen to match with the default values in the popular journey planner, Reittiopas, used within the Helsinki metropolitan area (Helsinki Region Transport, 2017). The maximum walking distance for direct walk from an origin to a destination as well as the access, egress, and transfer legs was set to 1000 m. The selection of this value is supported by results on walking behavior in the Helsinki metropolitan region, as most (~85%) realized walking trips are shorter than 1000 m (Weckström, 2016) and as the surveyed maximum tolerance for walking to a metro stop has a median value of 1000 m (Suomalainen, 2014). While PT users typically prefer shorter distances, we have opted for this conservative value of 1000 m because it also allows PT users to walk longer distances when it is beneficial for them in terms of travel time or transfers.

3. Results

We now apply our methodology to the public transport network of the Helsinki metropolitan area, shown in Fig. 7. In the following, we discuss a set of case studies relevant to measuring travel impedance for the purposes of PT planning. In Section 3.1.1,
we discuss two examples of fastest-path temporal distance profiles that allow PT planners to investigate service frequencies and journey durations between a chosen origin-destination pair. To provide a spatial overview of PT travel time towards a selected destination, we summarize these individual profiles with temporal distance measures, and provide map visualizations for more holistic analysis (Section 3.1.2). In Section 3.1.3, we incorporate the number of boardings to the fastest-path temporal distance profiles, which allows investigating the potential trade-offs between PT travel time and number of required boardings. Again, summarizing these statistics and displaying them on a map provides a more holistic overview (Section 3.1.4). To investigate service variations, we show in Section 3.2 a temporal distance profile that tells how temporal distances and boarding counts vary during one day. Sometimes, a PT planner needs to measure the ease of access of multiple, interchangeable destinations such as grocery stores or change-over points to other modes of transport. To this end, we discuss the ease of access to long-distance trains heading north of Helsinki in Section 3.3. Last, in Section 3.4, we show the results of an all-to-all analysis between the PT stops in the Helsinki metropolitan area, allowing us to understand the general relationships between the introduced statistics, and to assess the overall status of PT travel impedance in the region.

3.1. Access to Aalto University campus

We showcase our approach with a case-study of traveling to the Aalto University main campus using PT from all other PT stops and two additional locations in the Helsinki metropolitan area. The precise destination is the main building of the Aalto University campus. For now, we focus on the journey alternatives departing during the morning rush hour (08:00–09:00). The routing interval for the mcpcSA algorithm is set to 08:00–11:00, limiting the duration of any discovered journey to 3 h.

3.1.1. Fastest-path temporal distance profile examples

As all of our statistics are based on the understanding of the fastest-path temporal distance profiles and distributions, we start by analyzing the access times to the Aalto University campus from two locations: the Itäkeskus and Munkkiniemi shopping centers. The locations of these three places are indicated in Fig. 7. The computed temporal distance profiles and distributions are shown in Fig. 8.

For the first origin, the Itäkeskus commercial center (Fig. 8a), we observe that there are many journey alternatives with similar durations. While no clear service gaps are present, the departure times of the journeys are irregular, which can indicate that multiple different PT lines are used by the fastest-path journeys. The fastest-path temporal distance distribution shown in Fig. 8a summarizes the profile: the temporal distances lie within a narrow range between \( \tau_{\text{min}} = 46.3 \) min and \( \tau_{\text{max}} = 52.3 \) min. The effective pre-journey waiting time is small both in absolute and relative terms: \( \tau_{\text{mean}} - \tau_{\text{min}} = 49.1 - 46.3 = 2.8 \) min; \( (\tau_{\text{mean}} - \tau_{\text{min}}) / \tau_{\text{min}} \approx 0.06 \).

In Fig. 8b we show the fastest-path profile from the second origin, Munkkivuori, to the Aalto University campus. Compared to the previous profile, there are now fewer journey alternatives to choose from. On closer investigation, the profile seems to be a combination of two recurring journey alternatives having durations of approximately 20 and 26 min. Typically the fastest option towards the destination is to wait for one of the 20-minute journeys, but when the waiting time for such a 20-minute journey is long, it can be faster to travel to the destination using a 26-minute journey. Overall, the effect of the 26-minute journeys to the mean temporal distance is nonetheless small, as the total area under the temporal distance profile would not increase much if the 26-minute journeys were not available.

The temporal distance statistics for this profile are as follows: \( \tau_{\text{min}} = 18.9 \) min, \( \tau_{\text{mean}} = 25.7 \) min, \( \tau_{\text{max}} = 31.9 \) min. Unlike in the previous profile, now the pre-journey waiting time is a large component of the mean temporal distance \( (\tau_{\text{mean}} - \tau_{\text{min}} = 6.8 \) min), especially in relative terms: \( (\tau_{\text{mean}} - \tau_{\text{min}}) / \tau_{\text{min}} \approx 0.36 \).

3.1.2. Visualizing temporal distance statistics on a map

Although the previous temporal distance profiles provide considerable insight, it is not feasible nor desirable for a PT planner to go through hundreds of individual profiles. Thus, for a more holistic understanding of the accessibility of Aalto University campus, we visualize the minimum, mean, and maximum temporal distance statistics from all PT stops in the Helsinki metropolitan area in Fig. 9a–c. In these visualizations, the differences between the three

![Organization of public transport in Helsinki. In the figure we visualize the public transport lines in the Helsinki region. Additionally, we have pinpointed the locations of Aalto University campus (A), Itäkeskus commercial center (I), and Munkkivuori (M), which are used as the origins and destinations in our example case-studies.](image-url)
statistics become evident: while the campus area is quickly accessible from many PT stops when measured with the minimum temporal distance, the visualizations for the mean and maximum temporal distance show that the difference to the best case situation can be significant.

To better understand the differences between the mean $T_{\text{mean}}$ and minimum temporal distance $T_{\text{min}}$, we visualize them in Fig. 9d. In general, we notice that the differences $(T_{\text{mean}} - T_{\text{min}})$ increase with the distance from the destination, which could be explained through increased effective headways.

However, when we visualize the relative difference $(T_{\text{mean}} - T_{\text{min}})/T_{\text{min}}$ in Fig. 9e, the situation is the opposite: the shorter the distance, the larger the relative role of the pre-journey waiting time. As minor exceptions, there are a few more distant areas that become highlighted. From these areas, there probably are rare fast PT connections to the campus that take place infrequently within the time interval.

### 3.1.3. Boarding-count-augmented temporal distance profile examples

In addition to travel time, the numbers of vehicle boardings between an origin–destination pair need to be considered and understood by a PT planner. Fig. 10a shows the temporal distance profile between Itäkeskus and Aalto University, augmented with information on the number of boardings for each journey alternative. It can be seen that the journeys giving rise to the fastest-path temporal distance profile require two PT vehicle boardings ($b_{\text{mean f.p.}} = b_{\text{max f.p.}} = 2$). Additionally, there is a direct service between the stops.
Fig. 10. Two real-world boarding-count-augmented temporal distance profiles. On the left (a), we show the profile and the temporal distance distribution from Itäkeskus to Aalto University campus, where the fastest-path temporal distance profile differs from the profile requiring least number of boardings. On the right (b), the profile and distribution from Munkkivuori to Aalto University campus is shown. Now the fastest-path temporal distance profile mostly coincides the profile requiring least number of boardings.

For the first origin (Itäkeskus), the boarding-count-augmented fastest-path temporal distance distribution shown in Fig. 10a provides little new information. However, the distribution for the latter profile in Fig. 10b shows the dependencies between fastest-path boarding counts and temporal distances. For instance, the distribution shows that the journey alternatives with two vehicle boardings are not the overall fastest options for reaching the destination.

3.1.4. Visualizing boarding-count statistics on a map

To obtain a spatial overview on the number of vehicle boardings required to reach Aalto University campus, we visualize the different boarding-count statistics on a map. In Fig. 11a–c we show the minimum number of boardings $b_{\text{min}}$, along with the mean $b_{\text{mean f.p.}}$.

For the first origin (Itäkeskus), the boarding-count-augmented fastest-path temporal distance distribution shown in Fig. 10a provides little new information. However, the distribution for the latter profile in Fig. 10b shows the dependencies between fastest-path boarding counts and temporal distances. For instance, the distribution shows that the journey alternatives with two vehicle boardings are not the overall fastest options for reaching the destination.

3.1.4. Visualizing boarding-count statistics on a map

To obtain a spatial overview on the number of vehicle boardings required to reach Aalto University campus, we visualize the different boarding-count statistics on a map. In Fig. 11a–c we show the minimum number of boardings $b_{\text{min}}$, along with the mean $b_{\text{mean f.p.}}$.

For the first origin (Itäkeskus), the boarding-count-augmented fastest-path temporal distance distribution shown in Fig. 10a provides little new information. However, the distribution for the latter profile in Fig. 10b shows the dependencies between fastest-path boarding counts and temporal distances. For instance, the distribution shows that the journey alternatives with two vehicle boardings are not the overall fastest options for reaching the destination.

3.1.4. Visualizing boarding-count statistics on a map

To obtain a spatial overview on the number of vehicle boardings required to reach Aalto University campus, we visualize the different boarding-count statistics on a map. In Fig. 11a–c we show the minimum number of boardings $b_{\text{min}}$, along with the mean $b_{\text{mean f.p.}}$.

For the first origin (Itäkeskus), the boarding-count-augmented fastest-path temporal distance distribution shown in Fig. 10a provides little new information. However, the distribution for the latter profile in Fig. 10b shows the dependencies between fastest-path boarding counts and temporal distances. For instance, the distribution shows that the journey alternatives with two vehicle boardings are not the overall fastest options for reaching the destination.
and maximum fastest-path boarding-count statistics $b_{\text{max f.p.}}$. These three figures provide us with three different perspectives: While the map for $b_{\text{max f.p.}}$ shows that the campus can be reached using one or two vehicles from most PT stops, the maps for $b_{\text{mean f.p.}}$ and $b_{\text{max f.p.}}$ show that more vehicle boardings are required when opting for fastest-path journeys.

Now we can also investigate the trade-offs between the fastest-path journeys and the journeys requiring the least number of vehicle boardings ($b_{\text{min}}$). In Fig. 11d we show the differences between $b_{\text{mean f.p.}}$ and $b_{\text{min}}$, which are large especially in the eastern part of the Helsinki metropolitan area. In these areas journey alternatives requiring few boardings exist, but allowing for additional transfers enables the traveler to reach the campus area faster. In addition to boarding counts, in Fig. 11e we show the differences between the mean temporal distance computed for the profile with least boardings $\tau_{\text{mean,b min}}$ and the mean fastest-path temporal distance $\tau_{\text{mean}}$. Now, in addition to the eastern part of the Helsinki metropolitan area, for instance the Lauttasaari neighborhood (south-east from Aalto University) becomes highlighted due to few direct connections to the Aalto University campus during the morning.

3.2. Service level variations through a day

So far, the time interval in our analyses has spanned only 1 h. To analyze service level variations through one day, we computed a temporal distance profile ranging from 6:00 to 21:00 between Itäkeskus and Aalto University, shown in Fig. 12a. While there are frequent direct journey alternatives between the origin and destination ($b_{\text{min}} = 1$), alternatives with two or more vehicle boardings are always faster. The computed statistics indicate considerable variance in the fastest-path temporal distance ($\tau_{\text{max}} \approx 62.2$ min, $\tau_{\text{mean}} \approx 52.2$ min, $\tau_{\text{min}} \approx 44.3$ min), and trade-offs between temporal distance and numbers of required vehicle boardings: $\tau_{\text{mean,b min}} \approx 66.6$, $b_{\text{mean f.p.}} \approx 2.5$, $b_{\text{min}} = 1$.

The temporal distance profile also enables detailed analysis of daily patterns in PT service levels. First, the more frequent service patterns during rush hours are visible in the profile corresponding to the direct trunk bus route 550 operating between Itäkeskus and Aalto University. During the morning rush hour peak (centered approximately at 08:00), the scheduled journey durations for the direct route tend to be longer, which is most likely caused by increased passenger load and congestion. Interestingly, the afternoon rush hour is not visible in the fastest-path profile. Based on our investigations of the actual time-tables, this is most likely due to bus line 102 that provides fast and frequent service to the campus from the city center in the morning, but less so during the afternoon.

The fastest-path temporal distance distribution shown in Fig. 12b now efficiently summarizes the fastest-path profile. The distribution shows that the fastest path to the destination usually requires two or three vehicle boardings and that the overall fastest options for reaching the destination require two boardings.

3.3. Access to multiple long-distance train stations

For PT planning, it is sometimes necessary to compute access times towards multiple alternative destinations, such as shopping malls or transfer stations. This can be easily done with the mcpCSA algorithm, as the only necessary modification to the algorithm is to initialize it with multiple destinations instead of a single destination.

To demonstrate this possibility, we compute and analyze the access times and boarding counts to the three train stations (Helsinki central, Pasila, Tikkurila) on the long-distance railroad track heading north from Helsinki, during the weekday morning rush hour (08:00–09:00). Fig. 13a illustrates that the best access in terms of mean temporal distance is along the main trunk lines of the metropolitan area: all railroads provide good connections to the stations in addition to the bus routes operating to the west from the Helsinki city center. When traveling longer distances, passengers typically have more luggage. As additional luggage makes transferring between vehicles often more difficult, the role of transfers is now especially important (Fig. 13b). The fastest paths that take place by train require only one boarding, whereas getting to one of the three stations from the west of the city center requires an additional transfer.

3.4. All-to-all rush hour analysis

So far we have analyzed accessibility only to a single destination or multiple destinations. These case studies have provided us with some hints on the relationship between the temporal distance measures and boarding-count statistics. To validate these relationships, we computed these statistics between all pairs of PT stops during the morning rush hour (08:00–09:00; mcpCSA routing time 08:00–11:00). We present the most interesting findings in Fig. 14.

First, we show the distributions of the three fastest-path temporal distance measures $\tau_{\text{min}}, \tau_{\text{mean}}$, and $\tau_{\text{max}}$ in Fig. 14a. The distributions clearly differ from each other and peak at different values indicating that the selection of the PT travel time measure affects the outcome of analysis.

Next, we investigated the differences between $\tau_{\text{mean}}$ and $\tau_{\text{min}}$ in more detail. As shown in Fig. 14b, the difference between $\tau_{\text{mean}}$ and $\tau_{\text{min}}$ increases on average as a function of $\tau_{\text{min}}$. In other words, the pre-journey waiting time increases with travel duration. However,
the relative difference \( \frac{\tau_{\text{mean}} - \tau_{\text{min}}}{\tau_{\text{min}}} \) decreases as a function of \( \tau_{\text{min}} \) (Fig. 14c), and therefore the relative role of the waiting time decreases with travel duration. In combination these two results also indicate that the dependency between \( \tau_{\text{mean}} \) and \( \tau_{\text{min}} \) is non-trivial and cannot be explained by a constant offset or a multiplication factor. Thus, \( \tau_{\text{mean}} \) and \( \tau_{\text{min}} \) describe truly different aspects of PT accessibility.

In Fig. 14d, we show how the distributions for the three different measures for the number of required boardings \( (b_{\text{min}}, b_{\text{mean f.p.}}, b_{\text{max f.p.}}) \) differ from each other. When boardings are measured with \( b_{\text{min}} \), three boardings are almost always sufficient. However, the distribution for \( b_{\text{mean f.p.}} \) shows that additional vehicle boardings are often necessary on fastest-path journeys. Thus, the proper incorporation of time-domain into transfer analysis clearly affects analysis outcomes. Furthermore, the distribution for \( b_{\text{max f.p.}} \) shows that for a significant fraction of origin–destination pairs the fastest paths require even up to five vehicle boardings.

In Fig. 14e, we show how the mean number of required boardings on fastest paths increases as a function of the minimum temporal distance. The result is as expected: the longer the minimum temporal distance, the more vehicle boardings are required.

We also discovered that the number of Pareto-optimal journey alternatives (taking into account also the number of boardings) decreases as a function of the minimum temporal distance. This is shown in Fig. 14f. Although one might think that on longer distances there would be more options to choose from, in terms of Pareto-optimal alternatives the situation is the opposite.

3.5. Notes on running times

For our approach to be usable in practice, running times of the analysis should be feasible. In practice, the computation times are dominated by the computation of the Pareto-optimal journeys using mcPCA. For the rush-hour case studies that required mcPCA to scan over 3 h worth of PT and transfer connections, the routing took approximately 1 min per each destination on modern hardware. When we computed the Pareto-optimal journeys and statistics for the all-to-all analysis, we parallelized the computations over 64
CPUs, which enabled us to finish the computations in less than 4 h on average (per CPU). Note that significant speedups to the run times can be obtained by keeping track only of departure and arrival times, by decreasing the maximum allowed walking distance, or by implementing the mcppCSA algorithm using a lower-level programming language such as C++ that allows for low-level code optimization.

4. Discussion

For the purposes of PT planning, this paper introduced an approach for computing PT travel times and required numbers of transfers based on the analysis of Pareto-optimal journey alternatives. In particular, we visualized these journeys as temporal distance profiles depicting the temporal variation of the time to reach destination and the number of boardings required. Based on these profiles, we defined multiple measures characterizing PT travel time and the number of required vehicle boardings, while taking into account the schedule-dependence of PT operations. Furthermore, we showed that each of the suggested measures captures a different perspective on PT accessibility, demonstrated through a series of examples and statistics of travel times and transfers computed for all PT stop pairs in the Helsinki metropolitan region.

When analyzing PT travel times, we first defined three measures for PT travel time. The minimum temporal distance provides a proxy for the PT travel time when the user is willing to schedule her travel, while the mean and maximum temporal distances capture the expected and worst-case PT travel time when travel takes place spontaneously. Furthermore, we showed that the dependencies between these three measures are not trivial, i.e., they cannot be explained by a constant offset or a multiplication factor. In addition to overall variation, we found that the differences between the measures systematically depend on travel duration: the longer the duration, the larger are the differences. Because of this, we argue that travel time in PT is multi-faceted, and the different aspects of PT travel time should be considered separately. How much each of these aspects should be emphasized depends on the preferences of PT users.

We also introduced measures for the number of PT vehicle boardings, i.e., transfers, between origin–destination pairs. Especially, we argue that for computing the typical number of required vehicle boardings between an origin–destination pair, it is necessary to take the time-dependence of PT into account, in contrast to previous approaches based on a static network presentation. To this end, we proposed to compute the typical number of required boardings as the pre-journey waiting-time weighted average of the vehicle boardings of the fastest-path journeys. The measure thus describes the expected number of transfers assuming that the departure time is random and that the user always chooses the fastest path towards the destination. One should note that the fastest path is not always optimal with respect to the number of boardings, as there may be paths with fewer boardings but longer travel times. In addition, as the number of required boardings vary in time and between origin–destination pairs, we argue that the number of required vehicle boardings should be considered as a key component of all PT accessibility studies.

Overall, we believe that the suggested definitions for the fastest-path temporal distance and the boarding-count statistics provide a good starting point for more comprehensive PT accessibility studies. Here, we purposefully defined the measures to be as simple as possible, as this is often preferred by practitioners. However, more refined measures could be defined based on the sets of Pareto-optimal journeys. For instance, one could give a certain weight to the pre-journey waiting time based on user preferences, or limit the analysis only to journeys having at most a certain number of PT vehicle boardings.

Our approach presents a conceptually different approach for computing PT travel times and transfers, as we formulate these quantities using sets of Pareto-optimal journey alternatives containing information on the departure and arrival times of journeys. Given that modern algorithms can efficiently compute all such Pareto-optimal journey alternatives within a given time-frame, sampling of departure times can now be avoided. Also, the results of an accessibility analysis can now be stored compactly, as it is enough to only store the Pareto-optimal journey alternatives instead of recording travel times at the sampled departure times. In addition to departure and arrival times, we also included information on the number of required vehicle boardings for each journey. However, one could also consider other components of PT journeys such as walking time or transfer waiting time, which would enable more refined quantification of PT travel time components and their trade-offs.

The computation of the Pareto-optimal journey alternatives nonetheless requires the specification of certain parameters, and the outcomes are affected by their choice. First, to compute the Pareto-optimal journey alternatives in a reasonable time and to ignore journeys with excessively long walking distances, the maximum walking distance was set to 1 km on the access, transfer, and egress legs of the journey. While this hard limit can cause some artifacts, we expect them to be small as the speed difference between walking and traveling on a PT vehicle is large. Additionally, the values used for the safety margin for transferring between vehicles and walking speed can affect the results. Especially, the comfortable walking speed is known vary across individuals, and to depend on age and sex (Bohannon, 1997). Because of this, users of our tools should define the parameter values to suit their analyses and perform sensitivity analyses when working on critical real-world applications. However, such sensitivity analyses have not been typically done in PT accessibility studies, and thus further research on the impacts of parameter choices on PT travel impedance measures is required.

A qualitatively different limitation is that we rely purely on schedule data and assume that PT vehicles operate with perfect precision. However, delays and vehicle breakdowns are common in most PT networks, which can have a large impact on the accessibility and reliability experienced by PT users. Thus, further research should aim to compare the differences in accessibility when computed using schedule data and data on realized PT operations. As data on real-time locations of public transport vehicles is becoming increasingly available, such comparisons should soon become feasible.

In this paper, we modeled PT operations as a temporal network consisting of elementary connections between stops. Furthermore, the main idea behind the temporal distance profile was adopted from the network science literature (Pan & Saramäki, 2011). In addition to measuring temporal distances between network nodes, the field of network science provides tools for measuring e.g. network resilience (Albert, Jeong, & Barabasi, 2000; Williams & Musolesi, 2016) and the ease of navigation (Lee & Holme, 2012). These ideas could be also used to supplement current PT network analysis methodologies and tools.

To facilitate the adoption of our method, we have provided our full analysis pipeline as free open source software. As our pipeline relies solely on open data, similar studies can be carried out for any city where GTFS data is available.

Funding

All authors thank the support from the Academy of Finland through DecoNet-project (No. 295499). In addition, this research has received partial support from BEMINE-project (No. 303538).

Acknowledgments

The computational resources provided by the Aalto Science IT project are acknowledged. We thank the following public transport
planners and analysts at Helsinki Region Transport for discussions on the project: Jonne Virtanen, Matti-Pekka Laaksonen, Teemu Känsäkangas, and Niko-Matti Ronkonmäki. Moreover, we thank Mikko Kivelä for valuable comments on the manuscript, and Nils Haglund for proofreading the manuscript.

References