# Integrating network science and public transport accessibility analysis for comparative assessment

Luo, Ding; Cats, Oded; van Lint, Hans; Currie, Graham

**DOI**
10.1016/j.jtrangeo.2019.102505

**Publication date**
2019

**Document Version**
Final published version

**Published in**
Journal of Transport Geography

Journal of Transport Geography 80 (2019) 102505

Contents lists available at ScienceDirect

Journal of Transport Geography

journal homepage: www.elsevier.com/locate/jtrangeo

# Integrating network science and public transport accessibility analysis for comparative assessment

Ding Luo[a,*], Oded Cats[a], Hans van Lint[a], Graham Currie[b]

[a] Department of Transport and Planning, Faculty of Civil Engineering and Geosciences, Delft University of Technology, Stevinweg 1, P.O. Box 5048, 2600, GA, Delft, the Netherlands

[b] Institute of Transport Studies, Department of Civil Engineering, Monash University, Clayton, VIC 3800, Australia

ARTICLE INFO

*Keywords:*
Public transport
Accessibility
Network science
General transit feed specification (GTFS)
Comparative assessment
Tram networks

ABSTRACT

Network science offers powerful concepts and methods for studying complex systems, such as public transport networks. However, many existing studies on complex network analysis of public transport networks were primarily motivated to test network science concepts using real-life networks. Consequently, important properties related to public transport services in these studies have been overlooked. This has led to claims made by transport researchers that these works do not necessarily improve the understanding of system properties and performance. To bridge this gap, this study integrates network science and public transport accessibility analysis for comparative assessment across multiple public transport networks. The primary contribution of this study pertains to developing a method based on network science for computing public transport accessibility measured as the average travel impedance. The travel impedance metric is defined based on the generalized travel cost between stop pairs, containing initial and transfer waiting times, in-vehicle travel times and time-equivalent transfer penalty costs. To perform efficient computation, we propose a new type of weighted graph representation of public transport networks, which explicitly incorporates travel costs that are derived from general transit feed specification (GTFS) data. Besides the methodological contribution, the secondary one of this study consists in performing a comparative assessment of worldwide tram networks' accessibility. The analysis shows insights into how different travel components (e.g., in-vehicle travel times and waiting and transfer times) specifically account for the variance in accessibility across different networks. Additional references for improving the designing and planning of public transport networks can be obtained from such comparative assessments.

## 1. Introduction

*Network science*, a research field built upon graph theory, is dedicated to studying the connection and interaction between components in complex systems (Newman, 2010). It provides researchers with powerful toolkit to quantitatively investigate the collective dynamics resulting from the interactions among system elements. Given this advantage, an increasing amount of research has been conducted to apply theories and methods from network science to study transport systems over the past decades (Lin and Ban, 2013). In particular, there has been a focus on studying the topological characteristics of public transport networks (PTNs) (e.g., Sienkiewicz and Holyst, 2005; von Ferber et al., 2007, 2009; Berche et al., 2009; Louf et al., 2014; de Regt et al., 2018).

Applications of network science to transport systems have often been studied by network scientists seeking real-life examples of networks rather than by transport engineers and planners (Derrible and Kennedy, 2011). Consequently, a majority of studies ended up solely showing topological analyses without embedding information fundamental to transport systems, leading to claims that these works do not necessarily contribute to the transport community. For instance, Dupuy (2013) pointed out that these had provided limited recommendations to network planners, and thus impeded potential applications and impacts due to the absence of features related to transport and urban planning. Presumably, One of the main causes is that most topological analyses were performed with unweighted networks, thus leading to findings and conclusions that can provide limited knowledge and insights for improving the planning and operations of public transport (PT) (Cats, 2017). Although the significance of analyzing weighted networks for more meaningful findings was already demonstrated almost two decades ago (Barrat et al., 2003), unweighted networks have still been

* Corresponding author.
*E-mail addresses:* D.Luo@tudelft.nl (D. Luo), O.Cats@tudelft.nl (O. Cats), J.W.C.vanLint@tudelft.nl (H. van Lint), graham.currie@monash.edu (G. Currie).

https://doi.org/10.1016/j.jtrangeo.2019.102505
Received 28 April 2019; Received in revised form 22 July 2019; Accepted 26 August 2019
0966-6923/ 

commonly used in this specific research domain. Moreover, the perspective has often neglected the features associated with service attributes which are fundamental to PTNs. This is largely attributed to the difficulty in obtaining consistent PT attributes for creating meaningful weighted networks. This research gap needs to be bridged for network science to become more applicable in the transport research community.

To this end, this study is dedicated to performing an exemplary integration of network science and PT analysis that can enhance the understanding of system properties and performance. Our primary contribution pertains to proposing a new method based on network science for computing public transport accessibility measured as the average travel impedance. Note that the concept “accessibility” in this study is defined solely based on the travel impedance associated with reaching any potential destination across the network. This definition thus does not account for the intensity and diversity of destinations, which stay beyond the scope of this study and are hence neglected in the remaining of this paper. The travel impedance metric is defined based on the generalized travel cost (GTC) between stop pairs, containing initial and transfer waiting times, in-vehicle travel times and time-equivalent transfer penalty costs. To perform efficient computation, a new type of weighted graph representation of public transport networks is proposed, which explicitly incorporates the aforementioned components that are derived from general transit feed specification (GTFS) data. The secondary contribution of this study consists in performing a comparative assessment of worldwide tram networks' accessibility. The analysis shows insights into how different travel components (e.g., in-vehicle travel times and waiting and transfer times) specifically contribute to the variance in accessibility across different networks. Such latitudinal comparative assessments can provide additional knowledge for the PTN design, benchmark and planning, but are still scarce in the current literature due to the requirements imposed by existing methods that heavily rely on geographical information systems (GIS).

The remainder of this paper is organized as follows. Section 2 presents related research on PTN analysis and PT accessibility. The proposed method is then detailed in section 3. In section 4, the selected eight tram networks for the case study are described, followed by the presentation of the results and analysis in section 5. Section 6 concludes the study with the insights derived from the comparative study across different tram networks and suggestions for future research.

## 2. Related research

The research on PTN analysis has for a long time been a significant topic among transport scholars given its fundamental role in influencing network design and planning (van Nes, 2002). According to a review by Derrible and Kennedy (2011), several chronological development stages of this research topic can be identified. Pioneering contributions were made early by Vuchic and Musso who introduced a graph theory approach for analyzing PTNs (Musso and Vuchic, 1988; Vuchic and Musso, 1991). A variety of metrics were established to quantitatively evaluate PTNs (Vuchic, 2005). Derrible and Kennedy (2009, 2010a, 2010b) later on adopted graph theory principles for studying PTNs in a series of contributions. For example, they adapted a variety of concepts of graph theory to describe the features of metro networks. *State*, *form* and *structure* were proposed to measure the complexity of a network, the link between systems and the built environment, and the connectivity and directness of networks, respectively. The latest development has been focused on applying concepts, theories, and methods from network science to study the topological characteristics of PTNs. This has been largely facilitated more recently by the increasing availability of data sets (e.g., Gallotti and Barthelemy, 2015; Kujala et al., 2018a) and software functionalities (e.g., Hagberg et al., 2008).

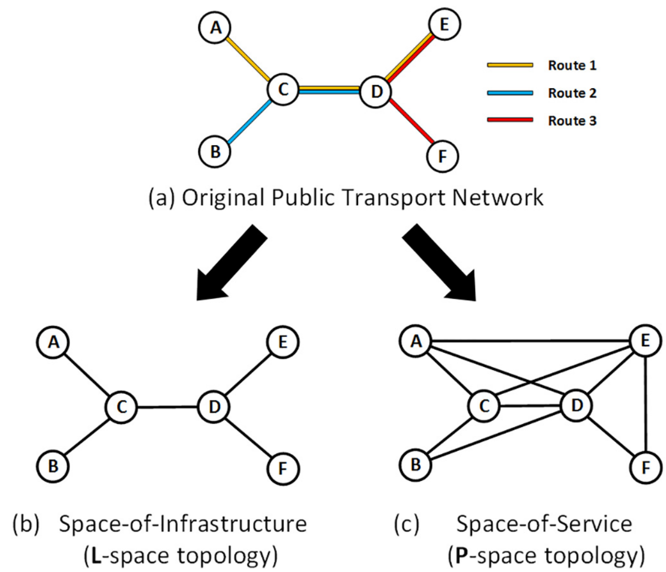



**Fig. 1.** Illustration of two commonly adopted topological (space) representations of PTNs (adapted from von Ferber et al. (2009)). The terms *space-of-infrastructure* (**L**-space) and *space-of-service* (**P**-space) are used in the following to better reflect the context of PT.

### 2.1. Network science analysis of PTNs

At the early stage, PTNs drew network scientists' attention when they searched for real-life examples of networks to test their theories and models (e.g., Latora and Marchiori, 2001, 2002; Sen et al., 2003). For example, many studies explored whether PTNs possess *scale-free* (Barabási and Albert, 1999) and *small-world* (Watts and Strogatz, 1998) features, which are considered among the most significant topological properties for characterizing networks. For more detailed summary and discussion of these studies, readers are referred to the reviews by Derrible and Kennedy (2011) and Zanin et al. (2018).

Unraveling PTNs' topological characteristics requires considering the key feature of PTNs. In particular, PTNs consist of both infrastructure and service dimensions with the latter being superimposed on the former. Consequently, establishing meaningful graph representations of PTNs lays the foundation for any further network analyses. Two types of graph representations that have been mostly employed in the literature are the so-called **L**-space and **P**-space topology, which are illustrated in Fig. 1 and also interpreted below.

- **L**-space: A straightforward representation of PTNs from the perspective of infrastructure. Each node represents a stop, and a link between two stops is formed if two stops are adjacent on at least one infrastructure segment (i.e. road or rail). This is one of the most extensively utilized representations by researchers (e.g., Latora and Marchiori, 2002; Sienkiewicz and Holyst, 2005; von Ferber et al., 2007, 2009). In many of the studies adopting this representation, only unweighted graphs are used in the analysis, hence containing no information of service properties.
- **P**-space: A representation solely based on PT service routes. The nodes represent stops and are linked if they are served by at least one common route. As a result, the neighbors of a node in this space are all stops that can be reached without performing a transfer. **P**-space has been widely used to investigate transfer possibilities (e.g., Sienkiewicz and Holyst, 2005; Xu et al., 2007; von Ferber et al., 2007, 2009). This space does not contain any information about the infrastructure layer (i.e., the physical sequence of links traversed between stops).

In the following, the terms *space-of-infrastructure* (**L**-space) and *space-of-service* (**P**-space) are used to better reflect the context of PT (Luo et al., 2019).

Based on the pioneering studies mentioned above, scholars have recently striven to better capitalize on network science in the context of PT research. This has led to efforts to incorporate PT specific features into complex network analysis of PTNs, such as travel demand and service attributes (e.g., passenger flows, transfers, service frequency, travel times, etc.). More weighted complex network analyses have emerged to account for demand and supply patterns in PTNs (e.g., Soh et al., 2010; Haznagy et al., 2015; Feng et al., 2017). Furthermore, investigations into the vulnerability, robustness and (node and link) criticality of PTNs have explicitly considered passenger demand and flow assignment (e.g., Cats and Jenelius, 2014; Cats, 2016; Cats et al., 2016, 2017).

### 2.2. Public transport accessibility

Accessibility has been widely studied in the field of urban planning (e.g., Batty, 2009) and transport planning (e.g., Geurs and van Wee, 2004; van Wee, 2016). According to the summary by Nassir et al. (2016), a consensus has been reached that accessibility can be measured based on two main components, which are: (i) locations and attractiveness of urban opportunities (benefit side); and, (ii) impedance of traveling to these locations from residential areas in the network (cost side). In a nutshell, more accessible areas are defined as those that can be reached with lower travel impedance. PT accessibility can thus be defined in a similar fashion, with the travel mode restricted to PT, and the travel impedance calculated based on PTN attributes.

PT accessibility research has so far been mostly performed based on GIS techniques with limited data availability (e.g., Lei and Church, 2010; Currie, 2010; Saghapour et al., 2016). The recent wide spread of GTFS data, along with the improvement of software's capability in processing them (e.g., ArcGIS), has further facilitated these GIS-based approaches (e.g., Farber et al., 2014, 2016; Farber and Fu, 2017). Moreoever, new analytical methods have also been gradually proposed in parallel, such as the travel impedance metrics based on utility (Nassir et al., 2016) and Pareto-optimal journey (Kujala et al., 2018b), and a faster computational algorithm by Fayyaz et al. (2017).

According to the literature review presented above, it can be summarized that previous studies have mostly performed mechanical analyses without incorporating service information fundamental to PT systems, thus providing the transport community with limited insights. Although some studies have attempted to address this challenge, more efforts are still needed. To further bridge the gap, the research presented in what follows demonstrates how network science can be effectively integrated with PT accessibility analysis.

## 3. Method

This section presents the proposed method. An overview is first presented in Fig. 2 along with an introduction of all the individual steps. The details of each step is then depicted in the following subsections.

### 3.1. Building graph representation of PTNs from GTFS data

We first define that PTNs are comprised of two layers: the infrastructure layer (i.e., roads and rails) and the service layer superimposed on the physical one (i.e., routes). A PTN is then represented as a directed graph which can be denoted by a triple $G = (V, E, R)$, where $V$, $E$, $R$ represent the set of nodes, links and routes, respectively. Each node $v$ represents a stop, while each link $e_{ij}$ is defined by an ordered pair of nodes $(v_i, v_j)$, where $v_i$ and $v_j$, respectively, denote the source and target nodes. Each route $r$ is directional and characterized by an ordered sequence of links $r = (e_{r,\ 1}, e_{r,\ 2}, \ldots, e_{r,|r|})$. Note that a link can constitute parts of different routes when they share the same corridors. In addition, the stop here relates to a service location (as commonly shown in PT maps) which can contain more than one individual boarding and alighting spot in the operational network.

Based on the above definition, PTNs are constructed from GTFS data which are now commonly available as a standard data format for PT operators to share their schedules and network information with the public. Structured as a relational database, one GTFS feed is comprised of multiple files that are interconnected via common keys (Google, 2019). To obtain required graph representation, a dedicated program was developed in MATLAB. In principle, a full-scan of all the trips associated with a route during a period needs to be performed in order to obtain the complete stop sequence for this route.

### 3.2. Constructing the unweighted space-of-service network

Based on the basic graph representation, we further establish the unweighted space-of-service network $G^s = (V^s, E^s)$ for PTNs. As introduced in section 2.1, this topological representation characterizes PTNs merely from a service perspective. A node $v_i^s$ in this case represent a PT stop which is the same as it in the basic graph G (i.e., $V^s = V$). A link $e_{ij}^s \in E^s$ is created if $v_j$ can be reached from $v_i$ without performing transfers.

### 3.3. Adding travel times as weights to the space-of-service network

Using the scheduling information from GTFS data, we are able to combine in-vehicle travel times and waiting times into link weights in the space-of-service network. Note that both components are time-dependent based on schedules and are averaged values for a given time period. For simplicity the temporal element is not incorporated in the following formulation. For link $e_{ij}^s$ in $G^s$, its weight $w_{ij}^s$ is defined as follows:

$$w_{ij}^s = t_{ij}^v + t_{ij}^w \tag{1}$$

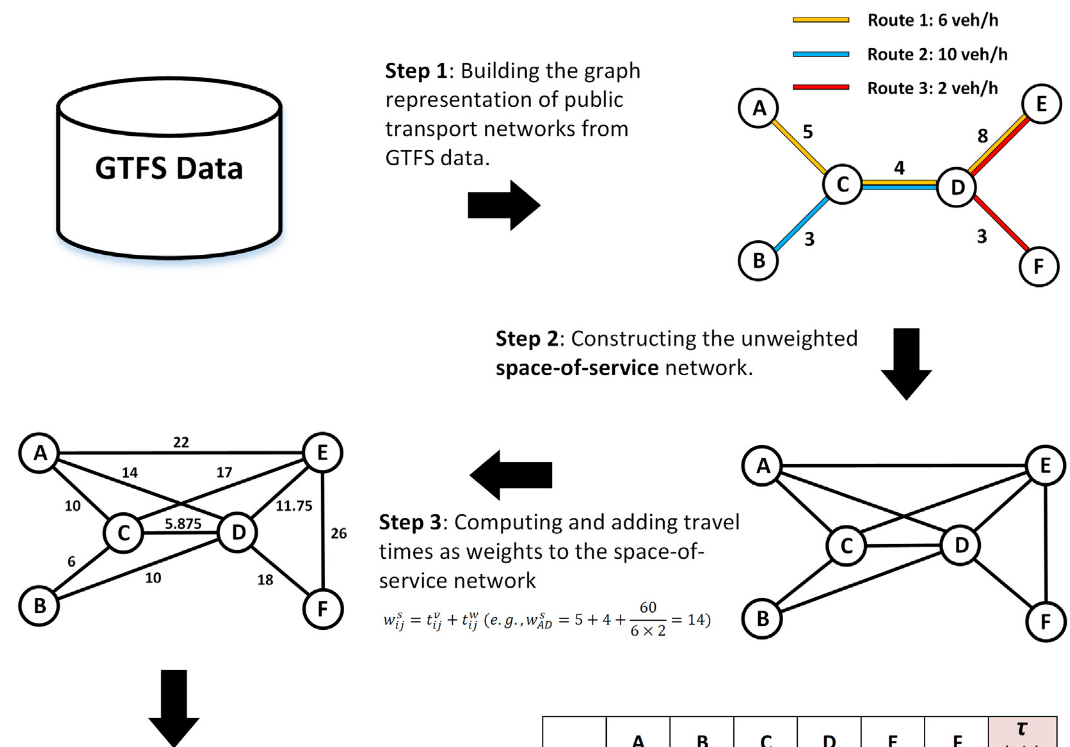
where $t_{ij}^v$ denotes the in-vehicle travel time of the direct service from stop $i$ to stop $j$ according to the schedule from GTFS. $t_{ij}^w$ represents the expected waiting time of travelers for the direct service from stop $i$ to stop $j$ during the same time period. This is estimated as half of the headway based on the joint service frequency of all direct services from stop $i$ to stop $j$. It needs to be pointed out that if there is any synchronization among different PT routes, it is neglected in this definition. An example is presented at step 3 in Fig. 2 to illustrate the proposed weight.

### 3.4. Measuring the average travel impedance per stop

The average travel impedance associated with individual stop $\tau_i$ in this study is defined as follows:

$$\tau_i = \frac{\sum_{j \neq i} u_{ij}}{|V| - 1} \tag{2}$$

where $|V|$ denotes the number of stops in the PTN. $u_{ij}$ represents the lowest GTC from stop $i$ to $j$. Specifically, $u_{ij}$ is the sum of total in-vehicle travel times, total waiting times for each journey leg (i.e., the initial waiting time and subsequent transfer waiting times), and time-equivalent penalty per transfer. The last component expresses the additional penalty associated with the inconvenience of performing a transfer beyond the additional travel time. The lowest GTC in this case can be efficiently computed using the proposed weighted space-of-service network $G^s(V^s, E^s, W^s)$. For instance, for the GTC from stop $i$ to $j$, $u_{ij}$, the shortest path $P_{ij}$ between them in $G^s$ is first searched for, which is denoted as a sequence of links $P_{ij} = (e_1^s, e_2^s, \ldots, e_K^s)$. Then the additional penalty cost is added based on the number of transfers along this shortest path. The definition is specified as follows:




| | A | B | C | D | E | F | $\tau$ (min) |
|---|---|---|---|---|---|---|---|
| A | 0 | 21 | 10 | 14 | 22 | 37 | 20.8 |
| B | 21 | 0 | 6 | 10 | 26.75 | 33 | 19.35 |
| C | 10 | 6 | 0 | 5.875 | 17 | 28.875 | 13.55 |
| D | 14 | 10 | 5.875 | 0 | 11.75 | 18 | 11.925 |
| E | 22 | 26.75 | 17 | 11.75 | 0 | 26 | 20.7 |
| F | 37 | 33 | 28.875 | 18 | 26 | 0 | 28.575 |

**Fig. 2.** Illustration of the proposed method.

$$u_{ij} = \sum_{k=1}^{K} w_k^s + \lambda(K-1) \tag{3}$$

where $w_k{}^s$ denotes the weight (i.e., travel time, including in-vehicle time and waiting time) of link $e_k{}^s$. $K$ represents the number of links, namely the minimal number of journey legs from stop $i$ to $j$. $\lambda$ represents the time-equivalent penalty cost per transfer which is a constant in this case. The applied metric is conceptually similar to the so-called *closeness* centrality (Bavelas, 1950), but it is much more adapted in the context of PT. It allows to quantify the travel impedance of each stop to all other stops in terms of the total journey time in the PTN. The simple average function can be replaced by a weighted average to enrich the metric by accounting for the importance of different origin-destination relations based on land-use or demand data if those are available.

## 4. Case study: assessing the accessibility of tram networks worldwide

To demonstrate the proposed method, eight tram networks worldwide, including Melbourne (Australia), Vienna (Austria), Milan (Italy), Toronto (Canada), Budapest (Hungary), Zurich (Switzerland), Amsterdam and The Hague (The Netherlands), are employed for the case study. Tram networks are used because there is, to the best of our knowledge, a lack of dedicated research on tram networks with sufficient samples in the literature compared to other PT modes, particularly metros (e.g., Derrible, 2012; Roth et al., 2012). This was presumably largely due to the unavailability of standardized data on tram networks. Therefore, by investigating eight tram networks worldwide in this case study, we aim to further highlight one of the main merits of the proposed method, which is its high transferability.

The network selection criteria include the following aspects: (i) the tram network plays a dominant role in the local urban transport system; and (ii) the GTFS data are available. All the networks were then generated using the GTFS data from the second half of 2018, which are subject to temporal variations possibly caused by construction and maintenance. In addition, for this case study the weekday morning peak schedules (8 am – 9 am) are used for computing the weight of space-of-service networks. For the time-equivalent transfer penalty cost $\lambda$ in Eq. 3, we apply 5 min per transfer in this study based on the result from Yap et al. (2018).

To obtain a better understanding of the basic properties of these tram networks, a graph displaying the numbers of stops, (physical) links and routes (i.e., $G(V,E,R)$) is presented in Fig. 3. Noticeably, the stop here relates to a service location (as commonly shown in PT maps) which can contain more than one individual boarding and alighting spot in the operational network. From the graph, it can be seen that these eight tram networks exhibit considerable differences in terms of their physical scale. With more than 800 stops and 1500 links, the Melbourne tram network is significantly larger than the others. Notwithstanding, it does not contain the largest number of routes (24

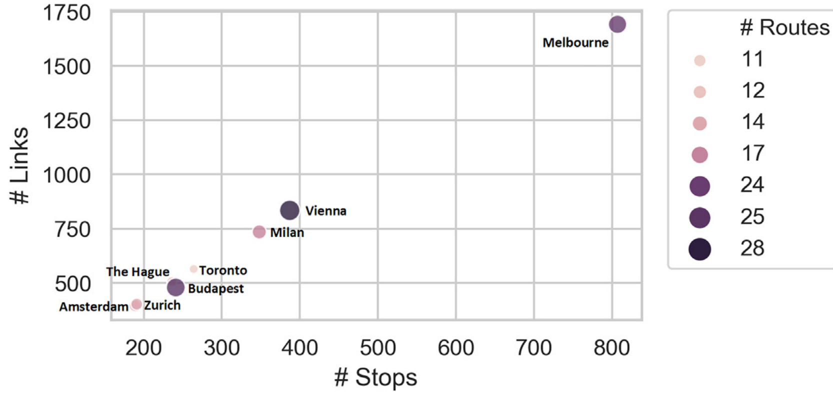



**Fig. 3.** Illustration of the basic properties of the studied tram networks. Note that the stop here relates to a service location (as commonly shown in PT maps) which can contain more than one individual boarding and alighting spot in the operational network.

routes only). Following the world's largest tram network in terms of the track-km length, the second tier cluster in our analysis encompasses Vienna and Milan, both of which possess some 300 stops and 700 links. However, Vienna boasts the largest number of regular routes out of all the eight networks, while Milan's number of routes seems to be proportionate to its scale. On the lower left corner in this diagram lie the remaining five networks, with Amsterdam and Zurich being the smallest. The Budapest tram network also stands out since its number of routes is remarkably disproportionate to its scale. Overall, it can be concluded that the numbers of stops and links are linearly correlated, while the number of routes does not necessarily coincide with the size of the networks.

## 5. Results and analysis

This section presents the results of applying the proposed method to the case study tram networks. Section 5.1 first introduces a benchmark metric used in the analysis. Then the results are presented in section 5.2 with visualizations, followed by a discussion in section 5.3. More insights are further presented in section 5.4 through an investigation into the variance in the proposed GTC-based travel impedance metric across different networks.

### *5.1. Additional benchmark travel impedance metric*

To better demonstrate the merit of the proposed GTC-based travel impedance metric, especially in the context of PT, we include an additional one in this analysis as a benchmark. It is computed using the space-of-infrastructure (**L**-space) representation described in section 2.1. This metric is derived in the same way as implied by Eq. 2, except that the GTC is replaced with the topological shortest path length in the unweighted space-of-infrastructure network. In other words, the travel impedance is solely determined based on the minimal number of links that are traversed between stop pairs in the space-of-infrastructure network. Neither travel times nor transfer attributes are incorporated in this benchmark metric. The travel impedance is thus completely viewed from a topological perspective. The reason for adopting this specific metric as the benchmark against the proposed one is the following: existing topological analyses of PTNs have used the space-of-infrastructure representation to study PTNs' properties related to what this study is examining. For example, several studies have studied the efficiency of PTNs based on the shortest path length in the unweighted space-of-infrastructure network without taking into account of PT transfers (e.g., Latora and Marchiori, 2001; de Regt et al., 2018). By comparing the proposed metric based on the GTC to this benchmark, the impact of including information on service properties can be assessed.

### *5.2. Results*

In this subsection, the computational results are presented in the form of travel impedance maps with explanations. As Fig. 4 shows, each column displays three graphs for a given tram network, corresponding to the result of, from top to bottom, (i) the benchmark metric; (ii) the proposed GTC-based metric, and; (iii) the difference between (i) and (ii). For the first two graphs, the same color scheme is used to illustrate different magnitudes of the travel impedance, therefore allowing for analyzing the spatial variation in impedance. Additionally, the distribution of metrics is presented in the top right corner for a more quantitative description. Note that there are black “**x**” markers in the middle and bottom rows, which represent the stops disconnected from the rest of the network. This is due to the fact that in this study the weighted space-of-service network pertains to service provision rather than infrastructure availability and is thus time-dependent. Besides, there is a tram route that is completely disconnected from the rest of the network in Toronto. This is because this tram route is integrated into the network through two other metro routes which are not included in this analysis.

The graphs on the bottom row in Fig. 4 visualizes the gap between the two different travel impedance metrics. The gap is quantified by the residual of a linear regression model in which the dependent and independent variables are the GTC-based and benchmark metrics, respectively. The rationale is that the bigger the residual (absolute value) for a stop is, the larger the gap between the two metrics for this stop is. For the network-wide visualization, we apply two colors, i.e., red and blue, to indicate the sign of residuals, while the magnitude is reflected through the depth of the color: the deeper the color is, the larger the gap between the two metrics is. Specifically, the blue spots illustrate where the travel impedance by the benchmark metric is higher than that by the proposed GTC-based metric, while the opposite holds for the red spots. The scatter plot in the upper right corner further shows more information about the linear regression between the two metrics. With the color for describing the gap retained, this plot allows for a more intuitive understanding of the consistency between the two metrics, particularly across different networks with the indication of the Pearson correlation coefficient *r*.

### *5.3. Discussion*

Some important patterns from the visualization are summarized in this subsection. First, according to Fig. 4, it can be seen that low travel impedance (i.e., high accessibility) is pronounced for those stops in the central area of the network in both benchmark and GTC-based cases. The travel impedance gradually increases as one moves away from the center toward the periphery for all the networks. Nevertheless, certain differences are also noticeable pertaining to the general pattern and the

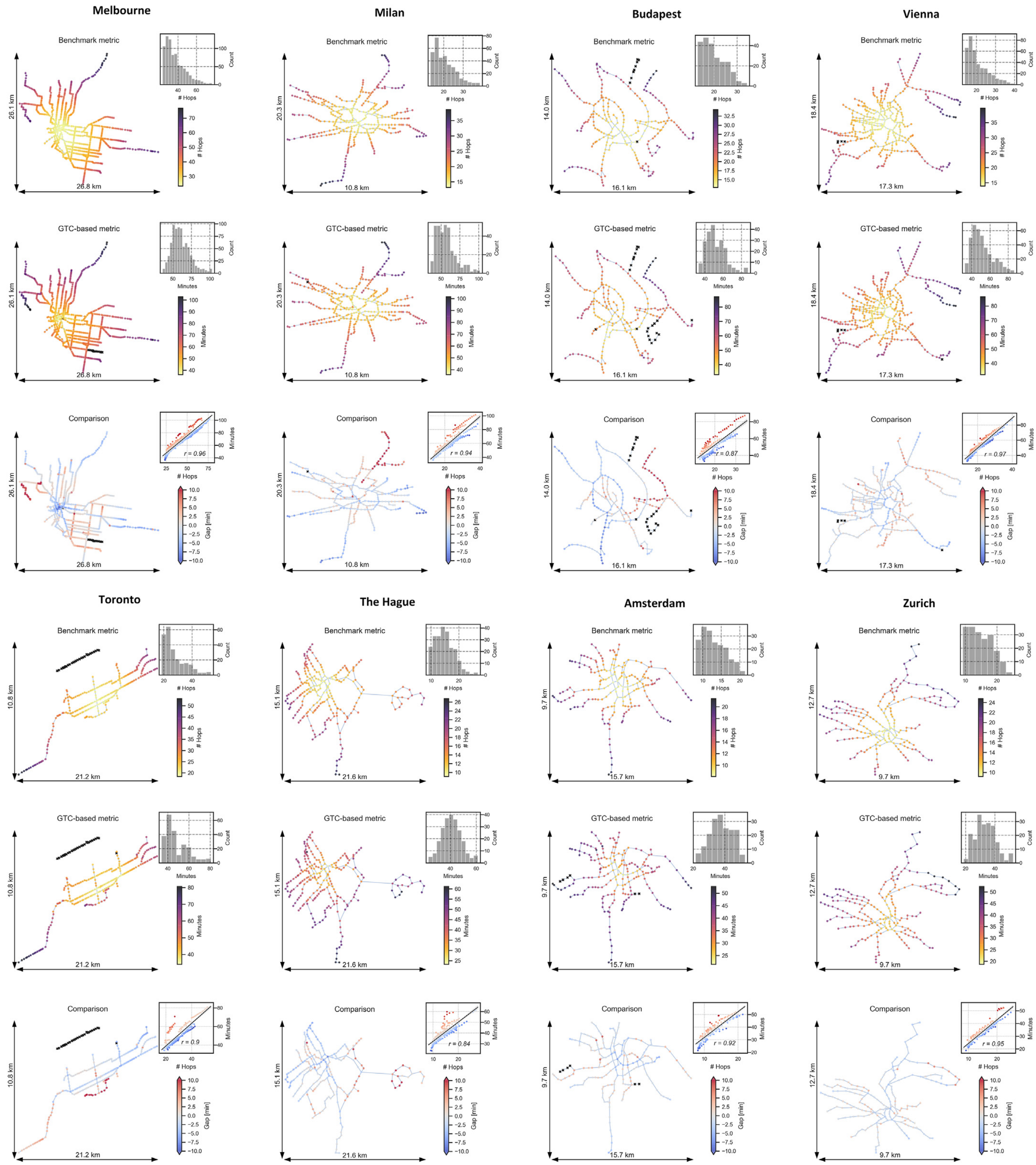



**Fig. 4.** Visualization of the travel impedance maps for case study tram networks. The benchmark metric, newly proposed GTC-based metric, and the comparison between them are respectively displayed from top to bottom for each city. The physical scale of all the networks are also provided on axes.

distribution. For example, the spatial disparity in the travel impedance is visually more remarkable when it is measured in terms of the GTC (the second row). In other words, in the GTC-based case, the stops with low travel impedance are more concentrated in the central part of the network. This is particularly remarkable in the case of Amsterdam, The Hague, Vienna and Melbourne. Second, in several cities, e.g., The Hague, Toronto, Milan, Budapest and Melbourne, some stops on the periphery of the network display largely underestimated impedance by the benchmark metric (reflected by the red spots). These patterns pertaining to the inconsistency between the two metrics can stem from three factors: (i) Less waiting times are needed in the central area given higher service frequency, particularly in the studied morning peak period; (ii) It is more likely that reaching a stop distant from the center requires more transfers, thus incurring more transfer penalty costs. (iii) It is common that the physical inter-stop space increases as the location moves from the center to the periphery of a network, therefore requiring more in-vehicle travel times.

Some intriguing findings can be also observed from the Pearson correlation coefficient. It can be seen that the values of Vienna (0.97), Melbourne (0.96) and Zurich (0.95) are higher than those of the rest of the tram networks, implying a good consistency between the two metrics. The Hague (0.84) turns out to be significantly lower than any other network, followed by Budapest (0.87). In the case of the former, this is due to a cluster of suburban stops on the east since the long corridor connecting the suburb to the city largely increases the cost in the GTC-based situation, resulting in a bias in the benchmark metric which underestimates the impedance.

We further analyze the difference between the two metrics from a distributional perspective. From Fig. 4, it can be already observed that the distributions of the two impedance metrics are significantly different from each other. The distribution of the benchmark metric is positively skewed, meaning that a large proportion of the distribution mass lies on the left side and there is a long right-side tail. While for the GTC-based metric, the majority in the distribution significantly moves to the right for most networks. Overall, the GTC-based metric quantifies the travel impedance in a less extreme way than the benchmark one does, suggesting that the spatial disparity in accessibility might be lower than assumed when neglecting service properties.

A more global comparison between them is further presented in Fig. 5 using the complementary cumulative distribution plot. Fig. 5a shows that several cities clearly overlap when the impedance is measured by the benchmark method (e.g., the group of Budapest, Vienna and Milan; the group of Amsterdam, The Hague and Zurich), while Fig. 5b demonstrates that the difference in the impedance among these networks becomes pronounced when the GTC is considered. In other words, once accounting for service properties that reflect network design and resource allocation choices, there are more pronounced differences between different networks than if those are neglected. Another finding is that the relative positions of the curves representing Budapest, Vienna and Milan shift from Toronto's left to its right when the GTC-based metric is applied. This could be attributed to the fact that the number of routes in Toronto is significantly lower than the other three, albeit the similar network size in terms of the number of nodes. As a result, much fewer transfers need to take place in Toronto than in the others, hence incurring lower transfer penalty costs.

## 5.4. Variance analysis

This subsection is devoted to investigating the variance in the resulting GTC-based travel impedance metrics across different networks. In addition to looking at the variance in the total cost, we also examine how individual components (i.e., (i) in-vehicle travel times and (ii) waiting and transfer times (with penalty)) varies in different networks. The so-called violin plot is used for visualization as shown in Fig. 6. As a type of enhanced visualization technique for the conventional box plot (The median of the data is marked by a white dot inside the plot with the interquartile ranges lying on its both sides), the violin plot for each network additionally illustrates the probability density of the data smoothed by a kernel density estimator. In this case, the individual violin plots are arranged in a descending order from left to right in terms of the median travel impedance value of each city.

Fig. 6a well corresponds to the intuition that larger networks in terms of the spatial scale indeed exhibit higher travel impedance. It can be observed that the median value of the top one (Melbourne, 60 min) is almost twice as that of the bottom (Zurich, 30 min). In addition, the diagram reveals that the variance in travel impedance is proportional to the spatial scale as well. As the median value of the impedance metric declines, the variance also drops in a way that the range between the maximum and minimum shrinks. For large networks, such as Melbourne and Milan, they appear to have long tails on the top (hence, right-hand tails) as a result of dramatic network sprawl from the center to suburbs. The opposite holds for some much smaller networks, including The Hague, Amsterdam and Zurich. Their shapes look much more compact in comparison to the others. Besides, the unique shape of Toronto reveals that its tram network is differently organized from many others.

Fig. 6b and Fig. 6c further unravel how (i) in-vehicle travel times and (ii) waiting and transfer times (with penalty)) respectively contribute to the variance in the GTC. It can be seen in Fig. 6b that the sequence derived based on the total GTC can still be roughly applied to the case of in-vehicle travel times and the shapes are also similar. This makes sense as the overall GTC is dominated by the in-vehicle travel time in all the networks. The only remarkable inconsistency occurs to Toronto where its median value is larger than that of Budapest and Vienna.

Fig. 6c then displays a very different pattern from the previous two diagrams. By excluding the component of in-vehicle travel times, which is largely dependent on the spatial scale of networks, we can actually find out how planned services on top of the physical network (i.e., route

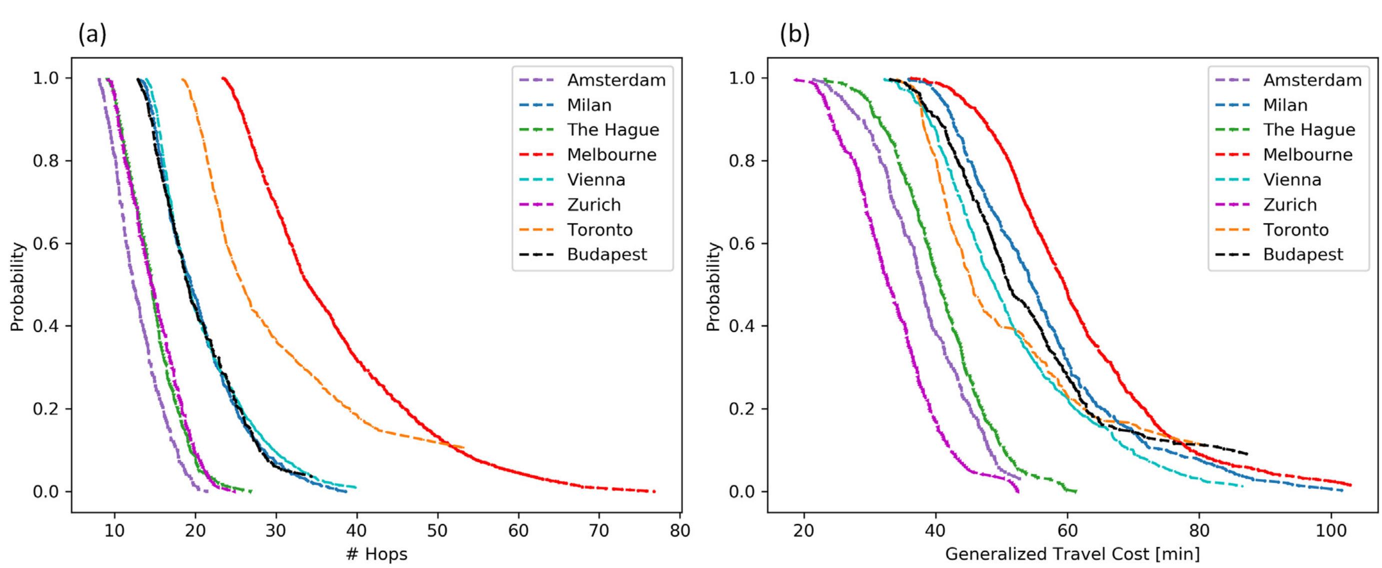



**Fig. 5.** Complementary cumulative distributions of the travel impedance for the studied tram networks. (a) The benchmark metric; (b) The GTC-based metric.

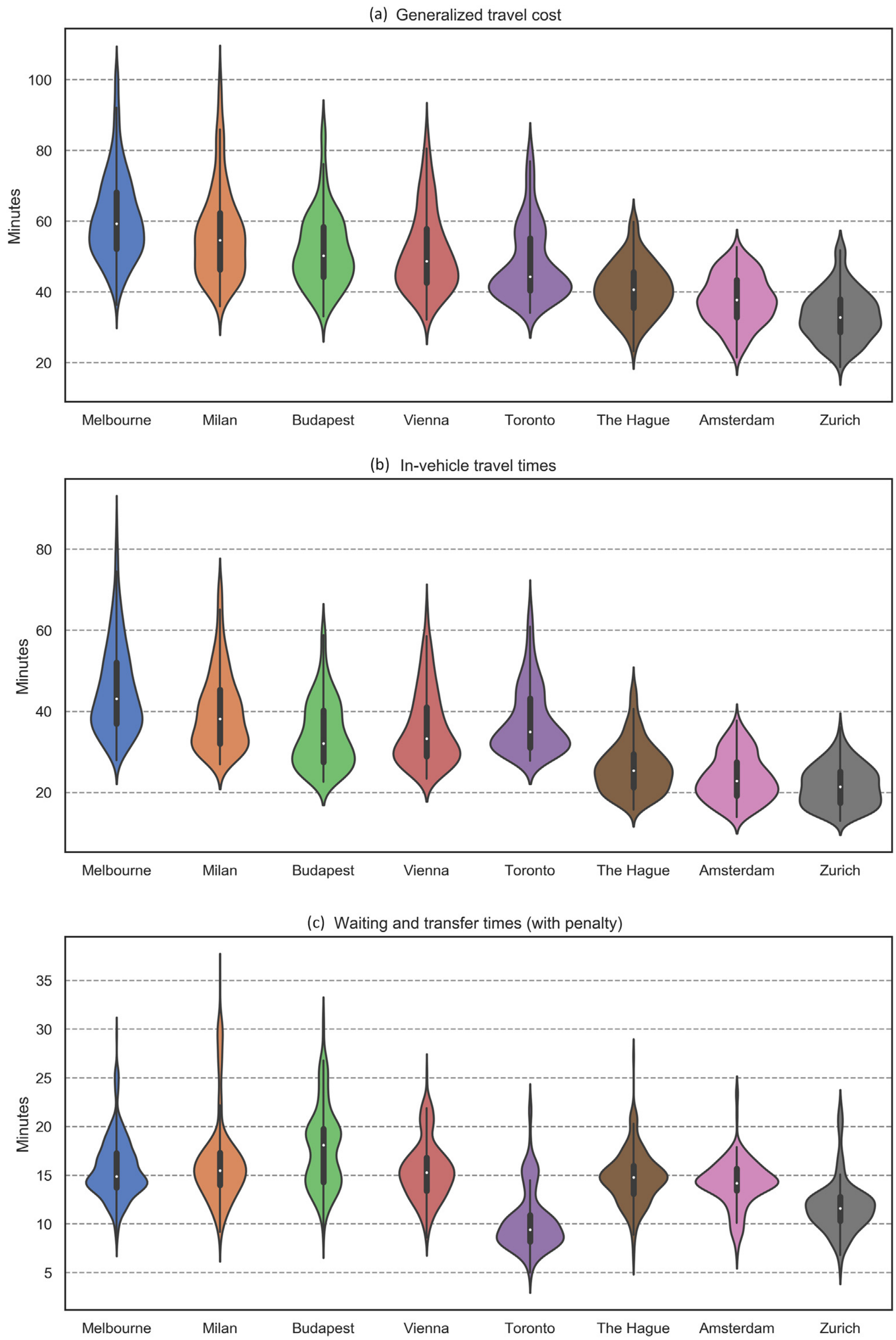



**Fig. 6.** Visualization of the variance in travel impedance for eight tram networks. The violin plot displays the median, quartiles and probability density of the data smoothed by a kernel density estimator.

and frequency design) influence the travel impedance. This can be explained from two aspects: (i) The number of transfers needed for connecting every stop pairs in the entire network determines waiting times and transfer penalty costs; (ii) Service frequency determines waiting times for each ride. It can be seen that the majority of the networks exhibit a median value of around 15 min, with the largest and smallest being close to 20 min (Budapest) and lower than 10 min (Toronto), respectively. Surprisingly, most significant variance can be found in Milan which shows a long tail on the top. This explains why Melbourne and Milan are similar in terms of the shape of the GTC (Fig. 6a) despite the fact that the in-vehicle travel times in Milan have less variance.

## 6. Conclusion

Existing complex network analyses of PTNs have mostly been focused on performing topological analyses without incorporating information fundamental to public transport services for the sake of PT planning. Although these studies have generated many insights into PT systems previously unknown to PT scholars, more efforts are still needed to enable network science to further address challenges encountered in PT planning and operations. To this end, this study proposes an exemplary integration of network science and public transport accessibility analysis. The primary contribution lies in developing a method based on network science for computing public transport accessibility measured as the average travel impedance. With the proposal of an innovative weighted graph representation of PTNs that explicitly incorporates travel costs according to the planned services contained in GTFS data, we are able to efficiently compute the minimal generalized travel cost between stop pairs. Such cost is comprised of initial and transfer waiting times, in-vehicle travel times and time-equivalent transfer penalty costs. The secondary contribution pertains to performing a comparative assessment of worldwide tram networks' accessibility based on the proposed method. Such latitudinal comparative assessments can provide additional insights into the PTN design, benchmark and planning, but are still scarce in the current literature due to the requirements imposed by existing methods that are heavily reliant on GIS.

New insights derived from the comparative case study are summarized as follows. According to the comparison between the benchmark and proposed GTC-based metrics, the main conclusion is that the spatial disparity in PT accessibility can be higher when planned service properties (i.e., travel times, initial and transfer waiting times, and transfer penalties) are considered than when they are not. In addition, the subsequent investigation shows that larger networks in terms of the spatial scale indeed exhibit higher median total travel impedance. However, this is primarily attributed to the component of in-vehicle travel times which is intrinsically positively correlated to the spatial scale of networks. By excluding in-vehicle travel times, we can see that the majority of the networks exhibit a median cost of around 15 min regarding waiting and transfers, with the largest and smallest being close to 20 min (Budapest) and lower than 10 min (Toronto), respectively. Among all the studied networks, Milan shows the largest variance in this specific cost component.

This study can be further continued in following directions. First, as mentioned earlier, information of the origin-destination demand can be incorporated when computing the average travel impedance. This would result in a more comprehensive way of measuring PT accessibility. Second, the proposed method can be extended to be more suitable for multimodal PT systems, possibly by adopting the multilayer network theory (Kivelä et al., 2014).

## Acknowledgements

This study was supported by the My-TRAC project funded by the European Union‘s Horizon 2020 research and innovation program under grant agreement No. 777640. The first author also acknowledges the financial support of the Strategic Alliance Student Exchange Program offered by Department of Civil Engineering at Monash University.

## References


Barabási, A.L., Albert, R., 1999. Emergence of scaling in random networks. Science 286 (5439), 509–512.
Barrat, A., Barthelemy, M., Pastor-Satorras, R., Vespignani, A., 2003. The architecture of complex weighted networks. Proc. Natl. Acad. Sci. U. S. A. 101 (11), 3747–3752.
Batty, M., 2009. Accessibility: in search of a unified theory. Environ and Plan. B: Plan and Design 36 (2), 191–194.
Bavelas, A., 1950. Communication patterns in task-oriented groups. J. Acoust. Soc. Am. 22 (6), 725–730.
Berche, B., Von Ferber, C., Holovatch, T., Holovatch, Y., 2009. Resilience of public transport networks against attacks. Eur. Phys. J. B 71 (1), 125–137.
Cats, O., 2016. The robustness value of public transport development plans. J. Transp. Geogr. 51, 236–246.
Cats, O., 2017. Topological evolution of a metropolitan rail transport network: the case of Stockholm. J. Transp. Geogr. 62, 172–183.
Cats, O., Jenelius, E., 2014. Dynamic vulnerability analysis of public transport networks: mitigation effects of real-time information. Netw. Spat. Econ. 14 (3–4), 435–463.
Cats, O., Yap, M., van Oort, N., 2016. Exposing the role of exposure: public transport network risk analysis. Transp. Res. A Policy Pract. 88, 1–14.
Cats, O., Koppenol, G.J., Warnier, M., 2017. Robustness assessment of link capacity reduction for complex networks: application for public transport systems. Reliab. Eng. Syst. Saf. 167, 544–553.
Currie, G., jan 2010. Quantifying spatial gaps in public transport supply based on social needs. J. Transp. Geogr. 18 (1), 31–41.
de Regt, R., von Ferber, C., Holovatch, Y., Lebovka, M., 2018. Public transportation in great Britain viewed as a complex network. Transportmetrica A: Transport Sci. https://doi.org/10.1080/23249935.2018.1529837.
Derrible, S., 2012. Network centrality of metro systems. PLoS One 7 (7), e40575.
Derrible, S., Kennedy, C., 2009. Network analysis of world subway systems using updated graph theory. Trans Res Rec: J Trans Res Board 2112, 17–25.
Derrible, S., Kennedy, C., 2010a. Characterizing metro networks: state, form, and structure. Transportation 37 (2), 275–297.
Derrible, S., Kennedy, C., 2010b. The complexity and robustness of metro networks. Phys A: Stat Mech and its Appl 389 (17), 3678–3691.
Derrible, S., Kennedy, C., 2011. Applications of graph theory and network science to transit network design. Transp. Rev. 31 (4), 495–519.
Dupuy, G., 2013. Network geometry and the urban railway system: the potential benefits to geographers of harnessing inputs from “naive” outsiders. J. Transp. Geogr. 33, 85–94.
Farber, S., Fu, L., 2017. Dynamic public transit accessibility using travel time cubes: comparing the effects of infrastructure (dis)investments over time. Comput. Environ. Urban. Syst. 62, 30–40 mar.
Farber, S., Morang, M.Z., Widener, M.J., 2014. Temporal variability in transit-based accessibility to supermarkets. Appl. Geogr. 53, 149–159.
Farber, S., Ritter, B., Fu, L., may 2016. Space-time mismatch between transit service and observed travel patterns in the Wasatch front, Utah: a social equity perspective. Travel Behav. Soc. 4, 40–48.
Fayyaz, S., K, S., Liu, X.C., Zhang, G., 2017. An efficient general transit feed specification (GTFS) enabled algorithm for dynamic transit accessibility analysis. PLoS One 12 (10), e0185333.
Feng, J., Li, X., Mao, B., Xu, Q., Bai, Y., 2017. Weighted complex network analysis of the Beijing subway system: train and passenger flows. Phys A: Stat Mech and its Appl 474, 213–223.
Gallotti, R., Barthelemy, M., 2015. The multilayer temporal network of public transport in Great Britain. Scientific Data 2, 140056 jan.
Geurs, K.T., van Wee, B., jun 2004. Accessibility evaluation of land-use and transport strategies: review and research directions. J. Transp. Geogr. 12 (2), 127–140.
Google, 2019. GTFS Static Overview. URL https://developers.google.com/transit/gtfs/.
Hagberg, A., Chult, D.S., Swart, P., 2008. Exploring network structure, dynamics, and function using NetworkX. In: Varoquaux, G., Vaught, T., Millman, J. (Eds.), Proceedings of the 7th Python in Science Conference, pp. 11–15 (SciPy 2008). Pasadena, CA, USA.
Haznagy, A., Fi, I., London, A., Nemeth, T., 2015. Complex network analysis of public transportation networks: a comprehensive study. In: 2015 International Conference on Models and Technologies for Intelligent Transportation Systems (MT-ITS). IEEE, Budapest, Hungary, pp. 371–378.
Kivelä, M., Arenas, A., Barthelemy, M., Gleeson, J.P., Moreno, Y., Porter, M.A., 2014. Multilayer networks. J Complex Netw 2 (3), 203–271.
Kujala, R., Weckstrom, C., Darst, R.K., Mladenovic, M.N., Saramaki, J., 2018a. Data descriptor: a collection of public transport network data sets for 25 cities. Scientific Data 5, 180089 may.
Kujala, R., Weckström, C., Mladenovic', M.N., Saramäki, J., 2018b. Travel times and transfers in public transport: comprehensive accessibility analysis based on Pareto-optimal journeys. Comput. Environ. Urban. Syst. 67, 41–54.
Latora, V., Marchiori, M., 2001. Efficient behavior of small-world networks. Phys. Rev. Lett. 87, 198701.
Latora, V., Marchiori, M., 2002. Is the Boston subway a small-world network? Phys A: Stat Mech and its Appl 314 (1–4), 109–113.
Lei, T.L., Church, R.L., feb 2010. Mapping transit-based access: integrating GIS, routes and schedules. Int. J. Geogr. Inf. Sci. 24 (2), 283–304.
Lin, J., Ban, Y., 2013. Complex network topology of transportation systems. Transp. Rev. 33 (6), 658–685.
Louf, R., Roth, C., Barthelemy, M., 2014. Scaling in transportation networks. PLoS One 9 (7), 102007.
Luo, D., Cats, O., van Lint, H., 2019. Can passenger flow distribution be estimated solely based on network properties in public transport systems? Transportation. https://doi.org/10.1007/s11116-019-09990-w.
Musso, A., Vuchic, V.R., 1988. Characteristics of metro networks and methodology for their evaluation. Transp. Res. Rec. 1162, 22–33.
Nassir, N., Hickman, M., Malekzadeh, A., Irannezhad, E., jun 2016. A utility-based travel impedance measure for public transit network accessibility. Transp. Res. A Policy Pract. 88, 26–39.
Newman, M., 2010. Networks: An Introduction. Oxford University Press.
Roth, C., Kang, S.M., Batty, M., Barthelemy, M., 2012. A long-time limit for world subway networks. J. R. Soc. Interface 9 (75), 2540–2550.
Saghapour, T., Moridpour, S., Thompson, R.G., 2016. Public transport accessibility in metropolitan areas: a new approach incorporating population density. J. Transp. Geogr. 54, 273–285.
Sen, P., Dasgupta, S., Chatterjee, A., Sreeram, P.A., Mukherjee, G., Manna, S.S., 2003. Small-world properties of the Indian railway network. Phys. Rev. E Stat. Phys. Plasmas Fluids Relat. Interdiscip. Topics 67 (3), 5.
Sienkiewicz, J., Holyst, J.A., 2005. Statistical analysis of 22 public transport networks in

Poland. Phys. Rev. E 72 (4), 46127.
Soh, H., Lim, S., Zhang, T., Fu, X., Lee, G.K.K., Hung, T.G.G., Di, P., Prakasam, S., Wong, L., 2010. Weighted complex network analysis of travel routes on the Singapore public transportation system. Phys A: Stat Mech and its Appl 389 (24), 5852–5863.
van Nes, R., 2002. Design of Multimodal Transport Networks: A Hierarchical Approach. Ph.D. thesis. Delft University of Technology.
van Wee, B., 2016. Accessible accessibility research challenges. J. Transp. Geogr. 51, 9–16 Feb.
von Ferber, C., Holovatch, T., Holovatch, Y., Palchykov, V., 2007. Network harness: Metropolis public transport. Phys A: Stat Mech and its Appl 380 (1–2), 585–591.
von Ferber, C., Holovatch, T., Holovatch, Y., Palchykov, V., 2009. Public transport networks: empirical analysis and modeling. The European Phys J B 68, 261–275.
Vuchic, V.R., 2005. Transit lines and networks. In: Urban Transit: Operations, Planning, and Economics. John Wiley & Sons (Ch. 4).
Vuchic, V., Musso, A., 1991. Theory and practice of metro network design. Public Transport Int 40 (3), 298.
Watts, D.J., Strogatz, S.H., 1998. Collective dynamics of 'small-world' networks. Nature 393 (6684), 440–442.
Xu, X., Hu, J., Liu, F., Liu, L., 2007. Scaling and correlations in three bus-transport networks of China. Phys A: Stat Mech and its Appl 374 (1), 441–448.
Yap, M., Cats, O., van Arem, B., 2018. Crowding valuation in urban tram and bus transportation based on smart card data. Transportmetrica A: Transport Sci. https://doi.org/10.1080/23249935.2018.1537319.
Zanin, M., Sun, X., Wandelt, S., 2018. Studying the topology of transportation systems through complex networks : handle with care. J. Adv. Transp. 2018 (Article ID 3156137).