Citation (APA)
Šfiligoj, T., Peperko, A., & Cats, O. (2026). Access graph: A novel graph representation of public transport networks for accessibility analysis. *Physica A: Statistical Mechanics and its Applications*, *699*, Article 131901.
https://doi.org/10.1016/j.physa.2026.131901

Physica A 699 (2026) 131901

Contents lists available at ScienceDirect

Physica A

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

Access graph: A novel graph representation of public transport networks for accessibility analysis

Tina Šfiligoj [a],*, Aljoša Peperko [b,c], Oded Cats [d]

[a] Faculty of Maritime Studies and Transport, University of Ljubljana, Pot pomorščakov 4, 6320 Portorož, Slovenia
[b] Faculty of Mechanical Engineering, University of Ljubljana, Aškerčeva 6, 1000 Ljubljana, Slovenia
[c] Institute of Mathematics, Physics and Mechanics, Jadranska ulica 19, 1000 Ljubljana, Slovenia
[d] Department of Transport and Planning, Delft University of Technology, P.O. Box 5048, 2600 GA Delft, The Netherlands

ARTICLE INFO

Keywords:
Public transport
Accessibility
Network science
Complex systems
Metro network
Equity

ABSTRACT

Accessibility, defined as travel impedance between spatially dispersed opportunities for activity, is one of the main determinants of public transport use. In-depth understanding of its properties is crucial for optimal public transport systems planning and design. Although the concept has been around for decades and there is a large body of literature on accessibility operationalisation and measurement, a standardised approach is lacking. To this end, we introduce a dedicated graph representation of public transport networks, termed the Access Graph, or A-space, based on the generalised travel times between nodes. We introduce an edge between two nodes in the access graph if the travel time between them is below a certain threshold time budget. In this representation, node degree directly measures the number of nodes reachable within a predetermined time, reproducing the cumulative opportunities measure of access at each specific value of the time budget. We study the threshold-dependent evolution of the degree distribution of the access graph and define a set of accessibility indicators. The indicators are observed at network-specific and passenger-based characteristic times. We apply the methodology to a dataset of 51 metro networks worldwide. For the Washington DC metro, we validate the approach by comparing supply to demand at the origin–destination level. The new representation addresses accessibility at the network structure level, offering a network science-grounded and flexible conceptual framework for accessibility studies.

1. Introduction

Accessibility is key in providing individuals with the opportunity to take part in employment, education, and recreational activities. The key objective of transport systems, and public transport services in particular, is therefore defined in terms of providing access. The Hansen definition of accessibility as travel impedance between spatially dispersed opportunities for activity is at the core of defining accessibility indicators [1]. It has been consistently in use since its conception, as evident from a large body of literature, and has facilitated accessibility-based planning [2]. In Hansen-like measures of access, also termed primal access, travel time is the independent variable and the number of reachable opportunities is the dependent variable. Complementary to this, dual access measures invert the independent and dependent variables and measure the time needed to reach a predetermined number of nodes [3].

* Corresponding author.
E-mail address: tina.sfiligoj@fpp.uni-lj.si (T. Šfiligoj).

https://doi.org/10.1016/j.physa.2026.131901
Received 10 March 2026; Received in revised form 24 July 2026
Available online 1 August 2026

Despite its wide theoretical and empirical developments, the exact measurement of accessibility remains an open problem. Although the general idea of accessibility is straightforward, the concrete operationalisation and related indicators vary significantly between implementations. The exact definitions and functions of the travel impedance differ, while determining reliable values for the number of opportunities is even more difficult to perform in a unified manner [4].

This has led to recent efforts to establish unified approaches to accessibility analysis. [5] proposed a general model across different modalities and categories of activity. This was followed by proposing a method, which allows for the comparison of different modes of urban accessibility across different cities [6]. In [7], the authors return to the spatial principles of accessibility and review metrics within a physical framework, highlighting the need for scientifically and mathematically rigorous approaches to access measurement. In [8,9], defining unified accessibility indicators has been identified as a major research direction in the coming decades. The lack of unified concepts and measures is seen as a key obstacle to accessibility-based planning in practice (e.g. [10,11]). In the afore-referenced works, the overarching thread is the recognition of the importance of the conceptual framework underlying accessibility studies — first its understanding, then its measurement.

Accessibility is fundamentally about connectedness between spatially dispersed locations, which is naturally modelled with graphs. We argue that a network science-based framework will offer new possibilities to conceptualise access by providing a framework that is both deeply rooted in physics and able to model and predict social and technical phenomena.

The preceding discussion on challenges in accessibility research and planning is especially relevant in the context of public transport network (PTN) planning. Public transport offers a sustainable and space-efficient means of travel, and it is paramount to design accessible PTNs for sustainable and livable cities. Graph-theoretical and network science approaches have become increasingly important in the field of public transport systems research. Although network science methods have been present in PTN research for several decades, and gained traction since the publication of a seminal paper by [12], there has been a widely recognised concern that this research has been driven primarily by network scientists using PTNs as real-world examples for theoretical developments (e.g. [13]). As a result, the practical implications for transportation researchers and planners have often been overlooked.

Since the 2010s, this gap has begun closing down with transport researchers increasingly using network science approaches in modelling and studying PTN properties [14,15]. This is especially evident in network robustness and vulnerability studies [16–19]. In contrast, network-based accessibility research has received comparatively less attention. Past studies focused either on analysing the relative importance of nodes using centrality indicators [20,21] or compared different networks using an aggregate metric such as the average shortest path from each node to all other nodes, known as network (in)efficiency [22,23]. In a pioneering study, [24] suggested connecting network science and accessibility by calculating average shortest paths for each node to all other nodes, providing a Hansen-like indicator similar to closeness centrality. Their application of the method to eight tram networks worldwide demonstrated the potential of network-based methods for comparative assessment. An alternative approach to accessibility that takes into account different components of travel time was presented in [25] where the authors investigated accessibility in terms of Pareto-optimal journeys on temporal networks [25]. As evident from recent literature, the increasing availability of standardised timetable data in the General Transit Feed Specification (GTFS) format has made accessibility studies increasingly accessible [26].

In this study, we advance the argument for using graph-theoretical approaches for accessibility research. While the development and subsequent analysis of the standard PTN graph representations offer original insights into PTN properties, we argue that the existing graph representations of PTNs may not offer the level of detail needed for accessibility analysis. At stop level, the two standard PTN representations are the so-called L- and P-space representations, where nodes represent service stops. In L-space, or space-of-infrastructure, there is an edge between two nodes if they represent consecutive stops on a line, whereas in P-space, or space-of-service, there is an edge between each pair of nodes lying along the same line (i.e., each line is a clique), reflecting the ability of passengers to travel between nodes without a transfer. Edge weights may represent in-vehicle and waiting time for L-space and P-space, respectively [27]. However, from the accessibility perspective, where travel impedance should include all components of travel time, it is not straightforward to reconcile those in a single representation [25].

Related to the previous observation, the network properties used to quantify accessibility do not provide sufficiently complete and insightful information when comparing accessibility across networks. At the stop level, centrality measures, even those in which the global properties of the network are implicit in their definition, offer local and isolated node-level metrics. In addition, centrality measures originate from social network analysis, and while they have been used successfully in PT studies, their interpretation in the accessibility context may be obscured to some extent as they cannot be unequivocally translated into meaningful access properties. This may consequently limit the generalisation of these measures to potential new models, such as including land use, and thereby diminish their interpretability. At the network level, average shortest path length and network efficiency (defined as the inverse of the sum of all shortest path lengths) are often understood as global measures of accessibility. However, the results will differ based on whether they are calculated in L- or P-space, and in neither representation can the total travel time of passengers be fully captured.

PTNs are spatial networks [28]; therefore their infrastructure obeys constraints of planarity, with a possible exception of an occasional tunnel or viaduct. In planar networks, two edges intersect if and only if the intersection constitutes a node. This is encoded in the L-space representation. However, the operational system overlaid on the infrastructure can have a highly non-planar structure, as is the case in the P-space representation. In the latter, the edges are spatially abstracted, and the P-space representation has consistently been shown to exhibit great predictive power on passenger flows, offering great relevance to transport researchers and planners (e.g. [21,27]). In P-space, the node degree directly represents the number of stops reachable without a transfer and the path length effectively counts the number of transfers. These properties, crucial for understanding passenger behaviour, cannot be modelled in the infrastructural representation.

In line with the above reasoning, and to address the discussed shortcomings, we propose a novel graph representation of PT networks dedicated for accessibility analysis, called the *Access Graph*, or *A-space*, where two nodes share an edge if they can be reached from one another within a given travel time budget t_b. The construction of the access graph is based on the matrix of generalised travel times obtained from L- and P-space representations. The generalised travel time incorporates three components: in-vehicle time, waiting time, and transfer penalties, the latter two weighted with empirically obtained perceived relative costs. We note here that the travel time in this work includes only the travel within the studied metro system, i.e. excluding access and egress times by other modes and implicitly assuming uniform demand.

The resulting representation is time-budget dependent, and the edge set of the access graph will differ for each value of t_b. When the time budget equals zero, the access graph will be empty, and when it reaches the maximum generalised travel time, the access graph will form a clique. The edges of the access graph can be seen as representing functional connections between nodes. Importantly, in the resulting graph representation, while retaining the same set of nodes, the edges directly indicate accessibility between two nodes. A crucial distinctive feature of the A-space representation is that it allows for conducting an analysis directly at the link level. This is of particular importance also to PT planners, as the origin–destination formulation of supply exhibited by the A-space exists at the same level as demand and the realisation of which in terms of passenger flows. Therefore, this approach offers a supply-based representation that describes network function at the origin–destination level, making it consistent with demand representation by construction. In this formulation, structural features can be studied at the link level, which stands in contrast to the traditional geographical and network-based approaches that consider access at the node level.

In addition, at the node level, the node degree in the access graph directly provides the number of reachable nodes within a specific time budget limit, corresponding to the traditional cumulative opportunities accessibility measure [5]. This offers a flexible framework to study the cumulative opportunities measures of access at a higher level: instead of observing the measure at one or a few fixed values of the cut-off time, we can study the structural evolution of access with systematically varying the cut-off times. Based on this representation, we study the temporal dependence of node degree distributions.

To connect supply with demand, an analogous methodology to that of the access graph is applied to the OD passenger flow network for the Washington, D.C. metro. Here, the origin–destination matrix is reconstructed from smart card data [29,30]. Then, a thresholded approach is used to identify OD pairs that carry at least a threshold flow, and its value is systematically decreased. We then observe the evolution of both the access and flow graphs and identify the threshold values of travel time and flow at which the two graphs are structurally closest to each other. In addition, this approach allows us to identify regions of under- and over-supply.

Furthermore, the distributional aspect of accessibility is one of the most pressing issues in accessibility-based design, where large (spatial) disparities in access to PT impact available opportunities for different social groups [31–35]. To this end, the Gini coefficient of the degree distribution is used as an indicator of access inequality [36].

The contribution of this study is fourfold:

1. **Methodological:** We introduce a dedicated PTN graph representation, the *access graph*, or the *A-space*, which connects pairs of nodes based on their reachability on generalised time-weighted shortest paths. This is the principal contribution of this work.
2. **Applied:** Based on the topology of the access graph, we define a set of accessibility indicators, including access inequality indicators.
3. **Empirical:** We apply the methodology to a dataset of 51 metro networks worldwide to study the properties of the access graph and compare accessibility measures across networks.
4. **Validation:** for one of the networks (Washington, DC) we compare the structure of the access graph to the structure of the flow graph and identify regimes of over- and under-supply.

The remainder of this article is structured as follows. In Section 2, the methodology for constructing the access graph and the definitions of the proposed indicators are detailed. In Section 3, the results of the empirical analysis on the metro dataset and the validation against demand are presented. Section 4 concludes this study and offers directions for further research.

2. Methodology

2.1. Preliminaries

Many real-world systems, among them (public) transport systems, can be modelled as networks, comprising sets of elements and their interconnections. A network is modelled as a graph that possesses certain properties of the real-world objects represented by nodes and edges. A graph is a structure $G(V,E)$ with a set of vertices, or nodes, $V = \{v_1, \dots, v_N\}$ and a set of edges $E = \{e_{ij}\}$, where an edge $e_{ij} = (v_i, v_j)$ connects nodes v_i and v_j. $N = |V|$ and $M = |E|$ are the dimension and the size of the graph, respectively. A graph with no edges is an empty graph, and a graph where all pairs of nodes share an edge is a complete graph, or a clique. A graph is most commonly represented in matrix form with the adjacency matrix A of dimension $N \times N$ and with entries $a_{ij} = 1$ if there is an edge between nodes i and j, and $a_{ij} = 0$ otherwise. A weighted representation provides a generalisation where edges are assigned weights, corresponding to their properties (e.g., cost or strength of connection). Node degree $k(i)$ of node i counts the number of the node's direct neighbours and equals the sum of the corresponding row in the adjacency matrix, $k(i) = \sum_{j=1}^{N} a_{ij}$. The $N \times N$ distance matrix of the graph$D$ contains as elements d_{ij} the lengths of the shortest paths between all pairs of nodes. If the graph is disconnected, the respective elements are set to infinity by convention.

Table 1
A table of PTN graph representations.

Representation	Name	Nodes	Edges
L-space	Space-of-infrastructure	Stops	Infrastructural connections
P-space	Space-of-service	Stops	Operational connections
C-space	Space-of-inter-lining	Routes	Shared stops
B-space	Space-of-line-interchange	Stops and Routes	Stop lies on route
A-space	**Access graph**	Stops	Stops reachable within threshold time

2.2. PTN graph representations

There exist four standard graph representations of PTNs: at the node-level, the L- and P-space representations; the C-space representation at the route level; and the bipartite B-space representation at the combined stop-route level. The L- and P-space representations were introduced in [37], and their properties were analysed for 22 Polish cities. In the seminal work of [12], the authors added the C- and B-space representations and consolidated the use of the four representations in the future PTN network studies that followed. In the following, we briefly describe the four representations and their properties.

- **L-space**: this is the most basic representation in which the nodes represent stops and the links connect stops that are adjacent on a route. It is also termed space-of-infrastructure [24], reflecting the infrastructural nature of connections.
- **P-space**: similar to L-space, the nodes represent stops, and there is an edge between each pair of nodes that lie on the same route, i.e., each route forms a clique. Also called space-of-service [24], it reflects the ability of passengers to travel without transfer, and the path length in P-space effectively counts the number of transfers.
- **C-space**: nodes represent the routes, and there is an edge between two nodes if they share at least one stop. C-space focuses on route-level behaviour and is suitable for studying synchronisation, etc.
- **B-space**: a bipartite representation where the two sets of nodes represent stops and routes, respectively, and there is an edge between two nodes if the corresponding stop lies on the corresponding route. Explicitly connects the two levels.

The representations can be seen as increasing in levels of spatial abstraction of edges and nodes. In the P-space, the edges reflect an important feature of connectivity and passenger journeys in PT systems, i.e., the transfer behaviour. In C-space, the routes are spatially abstracted into nodes. Each of the representations can offer a different perspective on PTN behaviour. In this vein, we see the access graph, or the A-space representation, as an addition to this family of representations. The node set represents stops, and the edges represent pairs of nodes that can be reached from one another within a given cut-off value of generalised travel time, comprised of in-vehicle and waiting time, plus a time-equivalent transfer penalty. In this sense, the edges in the access graph amount to a form of functional relationships between the nodes, by design dedicated to accessibility analysis. In line with the discussion above, the edges are spatially abstracted to directly reflect reachability between nodes. In contrast to the existing representations, the access graph is time budget-dependent, resulting in a layered representation where consecutive layers represent the functional connections at consecutive values of cut-off times. The representations and their properties are summarised in Table 1 and visualised in Fig. 1.

In comparison to all other representations, the A-space representation exists at the origin–destination level, which is the level relevant for studying accessibility and passenger flows. All the existing graph representations are supply-oriented, whereas A-space is explicitly devised to reflect the level-of-service delivered in terms of accessibility, thereby reflecting its most important feature in catering to travel demand. Consequently, in this representation, the simplest node-level property – node degree – is a measure of access. This circumvents the problems of interpretability and generalisability of centrality measures in L- and P-space, as well as the integration of information from both representations (e.g., [21,24,27]).

At the same time, in this representation, the original information on infrastructure and service is lost; we thus emphasise that the A-space is an addition to the family of graph representations, designed specifically to study accessibility, and does not replace other representations. The L-space is a natural representation for resilience analysis, while the P-space is useful for studying service connectivity. The C-space representation can be used to study line-level properties and address problems such as synchronisation at the line-level. The B-space representation is the only one that explicitly connects stops and lines, thus enabling the representation of travel paths; however, this is the least explored representation in the literature. For a comprehensive understanding of PTN function, insights gained from all representations should be integrated to enable informed planning.

In the following section, we describe the construction of the access graph in detail.

2.3. Access graph

In PT access studies, generalised travel time is typically comprised of in-vehicle and waiting times, as well as the time-equivalent representation of transfer penalties [38,39]. The latter include the value of disutility of transfer above and beyond the (weighted) waiting time at the interchange. In the access graph $\mathcal{G}_A$, the nodes represent the stops of a PTN, and the edge set is obtained depending on the generalised travel times as follows. First, the generalised travel time matrix D is constructed:

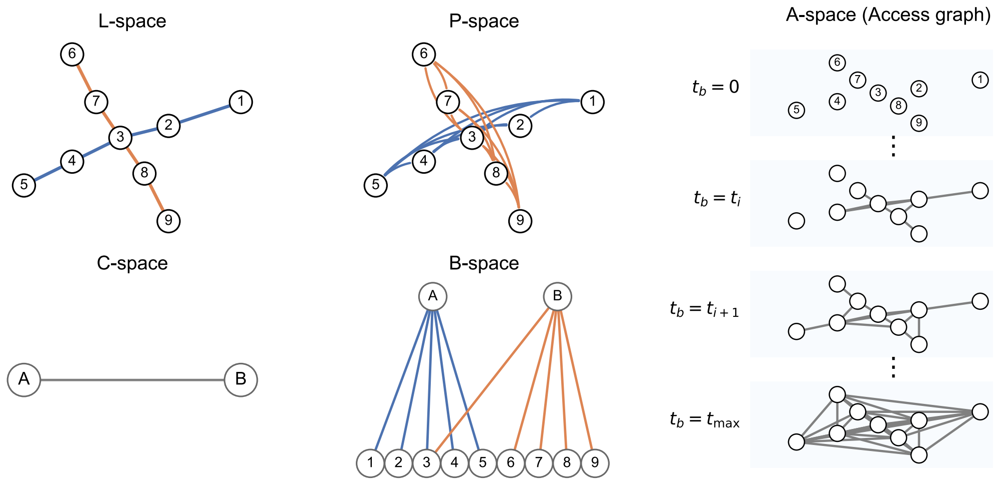

Fig. 1. Overview of existing PTN graph representations and placement of our new A-space representation in the group. On the right-hand side, the temporal progression of the A-space representation is shown. The topmost layer is shown for the threshold time t_b=0 where $\mathcal{G}_A$ is an empty graph. We then show schematically in the next two layers the structure of $\mathcal{G}_A$ at two consecutive intermediate time steps t_i and t_{i+1}. In the bottom layer, $t_b = t_{max}$, and $\mathcal{G}_A$ forms a clique.

1. Construct the standard graph representations: frequency-weighted P-space representation with edge weights representing service frequencies f_{ij} (given as the number of rides per hour) and in-vehicle time-weighted L-space representation with edge weights t^L_{ij}.
2. Transform the edge weights of the P-space representation into average waiting time t^{wait} in minutes, defined as half of the headway: $t^{wait}_{ij} = 30/f_{ij}$.
3. Assign in-vehicle time weights t^{in-veh}_{ij} to P-space edges by calculating the weight of the corresponding path in L-space (summing in-vehicle times on all consecutive links).
4. Calculate the distance matrix $\mathcal{D}^{P^{wait}}$ of the waiting time-weighted P-space representation.
5. Determine the distance matrix of in-vehicle times $\mathcal{D}^{P^{in-veh}}$ by calculating the in-vehicle time weights of the paths obtained in the previous step.
6. Similarly, determine the distance matrix $\mathcal{D}^{P^u}$ of corresponding path lengths in unweighted P-space. Note that $([\mathcal{D}^{P^u}]_{ij} - 1)$ counts the number of transfers along the path between i and j.
7. Construct the generalised travel time matrix $\mathcal{D}$:

$$\mathcal{D} = \mathcal{D}_{L^{in-veh}} + w^{\text{wait}}\mathcal{D}_{P^{wait}} + w^{\text{transfer}}(\mathcal{D}_{P^u} - J_N), \tag{1}$$

where J_N is an $N \times N$ matrix of ones, $w^{\text{wait}} = 2$ is the waiting time weight and $w^{\text{transfer}} = 5$ min is the transfer penalty. The values are determined based on the literature on empirical behavioural analysis and reflect average passengers' valuation of time [38,39].

Note that the entry $d_{ij} = [\mathcal{D}]_{ij}$ of the generalised travel time matrix represents the generalised travel time between stops i and j. In the next step, the time budget t_b is introduced, i.e., the maximum total travel time allowed. In the access graph of t_b, there exists an edge between nodes i and j if and only if $d_{ij} \leq t_b$. The adjacency matrix A with entries a_{ij} of the access graph is thus obtained from the generalised travel time matrix:

$$a_{ij} = \begin{cases} 1 & d_{ij} \leq t_b \\ 0 & \text{otherwise.} \end{cases} \tag{2}$$

The edge set of the access graph is thus time-budget dependent. Note that this approach and the resulting $\mathcal{G}_A(t_b)$ adjacency matrix is similar to the thresholded reachability matrix of a graph, where any pair of nodes that are reachable by a path are connected, and the threshold cuts off the connections above a certain length (see e.g. [40]). While similar, a key distinction between our approach and the original reachability matrix approach is that while the thresholding is based on several representations (and further allows for other modifications, e.g. to add other components of perceived travel cost), our A-graph constitutes a single representation. It thereby addresses exactly the long-standing issue of integrating infrastructure and service (and possibly other) information in a single PTN graph representation. Encoding this matrix as an adjacency matrix of a graph thus constitutes a genuinely new representation that is not trivially reducible to either L- or P-space, while incorporating the information from both. This representation enables an advanced network-based analysis of access: first and foremost, it enables the structural analysis at the level of origin–destination pairs, compared to node-level analysis typical in previous approaches, both in network science and traditional geographical analysis of access. In addition, it allows for the evolution approach to access, compared to single time points used in existing analyses. Moreover, in contrast to the thresholded reachability matrix, it allows for studying the changes in reachability structure, inference of critical time points, and comparison to demand (Sections 2.4 and 2.5).

The $\mathcal{G}_A(t_b)$ obtained in this process is an undirected, unweighted graph. Edges in the access graph connect nodes reachable within the time budget by following the shortest path. Therefore, the node degree D of $\mathcal{G}_A$ reproduces the cumulative opportunities, or, more specifically isochrone-based measure of accessibility [41,42]. The access graph as a whole can be thought of as a temporal network with consecutive layers representing the status of the access graph at consecutive values of t_b and thus levels of connectedness at each time step [43].

2.4. Access indicators

The evolution of the access graph is observed by increasing the time budget t_b from 0 to $t_{max} = \max_{i,j} d_{ij}$. At $t_b = 0$, $\mathcal{G}_A$ is an empty graph, and at $t_b = t_{max}$, $\mathcal{G}_A$ is a complete graph. The average degree $\langle k \rangle$ in these boundary cases is 0 and $N-1$, respectively. To examine global accessibility indicators, we study the change in degree distribution with increasing t_b in the interval $[0, t_{max}]$ for two complementary notions: the growth of average degree and the overall structural change of the network. Based on the shapes of the examined relationships, we identify characteristic values of t_b that mark notable transitions in access behaviour.

The average degree $\langle k \rangle(t_b)$ in the access graph is an increasing function of t_b. By the design of real-world transport systems – where the spatial density of stops and routes is highest in a central area, and the distance between adjacent stops as well as parallel routes then increases towards the periphery – the edges will typically first emerge in the core, and then the periphery gradually becomes better connected. This will result in a sigmoid (or S-shape) dependence of $\langle k \rangle$ w.r.t. t_b, meaning that at some point there is a point of inflection where the curve changes from convex to concave, and this is where typically the second derivative $\langle k \rangle''$ equals zero. This corresponds to the maximum of the first derivative $\langle k \rangle' = \frac{d\langle k \rangle}{dt_b}$, and we will take the value of t_b where this maximum occurs as an access indicator. Note that here we are working with discrete time steps and the derivatives are approximated with difference quotients $\frac{\Delta\langle k \rangle}{\Delta t_b}$. From the access perspective, the point where the maximum of the first derivative is achieved is interesting. This point can be understood as the time where the network has reached an "accessible state" through a second-order phase transition, after which the connections between the remaining pairs of nodes emerge at a (distinctively) lower rate. This perspective gives insights into accessibility behaviour from the system performance point of view.

For a complementary perspective, we adopt the passengers' point of view by considering characteristic values of t_b that correspond to typical commute times. Specifically, we observe the properties of the access graph at $t_b = 30$ min. The reasoning behind this choice is that the individual daily time budget does not increase, or increases only marginally with city size, and it was found to be around one hour [44,45]. This value appears to have remained approximately constant throughout the several decades since the first dedicated studies of travel time budget began in the 1970s. However, recent studies suggest that this value might be increasing over time [46]. We settle for the 60-minute average and note that our methodology enables straightforward modifications of this value. Thresholds near the average commuting times were set as cut-off values for the cumulative opportunities measures of access in past studies, for example, in [42].

Based on this reasoning, we propose a set of global accessibility indicators. First, the value of t_b at the occurrence of the point of maximum average degree growth, t_M, is proposed as an absolute-value dual access indicator. The reasoning behind taking an absolute value of time for performance measurement is again based on the 30 min goal, regardless of city (or network) size. For a complementary measure, an additional indicator δt_M is proposed that measures the point of maximum growth relative to the maximum generalised travel time observed in the network, i.e. $\delta t_M^{\Delta} = t_M^{\Delta}/t_{max}$. This gives a measure of performance that is normalised in relation to the farthest away origin–destination pair (which is equivalent to the notion of graph diameter), i.e. how early in the interval the maximal access growth is reached. Note that lower values of t_M and δt_M generally imply higher accessibility. In addition to the values of the (relative) time budgets at maximal growth, the values of the average degree at t_M and $t_b = 30$ min, $\langle k \rangle_M$ and $\langle k \rangle_{30}$, respectively, constitute complementary primal indicators. We emphasise that although the indicators are based on the average degree, the choice of the values at which we measure the average degree is based on the overall evolution of the access graph. In this sense, the indicators capture the global behaviour of accessibility within this averaged measure.

Alongside the average degree, we examine the t_b-dependent degree distributions of the access graph for the purpose of assessing the structural access equity in terms of spatial disparities in accessibility outcomes. A standard measure of inequality is the Gini coefficient, defined as:

$$G = \frac{\sum_{i=1}^{n} \sum_{j=1}^{n} |x_i - x_j|}{2n^2 \overline{x}}, \tag{3}$$

where $\overline{x} = \frac{1}{n} \sum_{i=1}^{n} x_i$ is the distribution mean [36]. G can assume values in the interval $[0, 1]$, where $G = 0$ represents perfect equity and $G = 1$ means maximum disparity. In this study, the values of the Gini coefficients of the node degree distributions of the access graph are observed. Following a similar reasoning as above, G_M and G_{30}, representing the values of the Gini coefficients of the degree distribution of the access graph at $t_b = t_M$ and $t_b = 30$ min, respectively, are proposed as indicators of access inequality. The Gini coefficient is chosen here rather than skewness, because the degree distributions in our empirical analysis were found to often exhibit a bimodal shape. We emphasise that this measure of inequality is strictly structural and does not involve the socio-economic aspect that are often considered in access equity literature (e.g., [9,47,48]).

Table 2
Accessibility indicators used in the analysis.

Notation	Name	Definition
t_{max}	Network diameter	maximum generalised travel time in the network
t_M	Access inflection point	time budget at maximum growth of average degree
δt_M	Normalised access inflection point	t_M/t_{max}; t_M relative to maximum generalised travel time
t_{JS}	Access structural change point	characteristic time budget at maximum structural change in of the access graph
δt_{JS}	Normalised access structural point	t_{JS}/t_{max}; t_{JS} relative to maximum generalised travel time
$\langle k\rangle_M$	Access at inflection point	average degree of the access graph at t_M
$\langle k\rangle_{JS}$	Access at structural change point	average degree of the access graph at t_{JS}
$\langle k\rangle_{30}$	30-min access	average degree of the access graph at $t_b = 30$ min
G_M	Inflection access inequality	the Gini coefficient of the degree distribution at t_M
G_{30}	30-min access inequality	the Gini coefficient of the degree distribution at $t_b = 30$ min

In addition, we investigate how degree distributions evolve for consecutive time steps. We adopt the Jensen–Shannon divergence (JSD) as a metric of statistical distance between the distributions, which is commonly used as a measure to quantify structural differences across layers in multiplex networks (see e.g. [49]). It is defined as:

$$\mathrm{JSD}(P \parallel Q) = \frac{1}{2}\ \mathrm{KL}(P \parallel M) + \frac{1}{2}\ \mathrm{KL}(Q \parallel M), \qquad M = \frac{1}{2}(P+Q), \tag{4}$$

where KL is the Kullback–Leibler divergence, which measures relative entropy of two probability distributions P and Q. For a discrete distribution, it is defined as:

$$\mathrm{KL}(P \parallel Q) = \sum_i P(i)\ \log_2 \frac{P(i)}{Q(i)}. \tag{5}$$

The JSD is then the symmetrised version of KL. JSD measures the "surprise value" of distribution P given the distribution Q based on the relative entropy and is thus sensitive to the overall degree distribution shape. For the access graph, $Q(i)$ and $P(i)$ will be the degree distributions at consecutive values of t_b: if $Q(i)$ is the degree distribution at $t_b = t_j$ (i.e. jth step) then $P(i)$ is the degree distribution at $t_b = t_{j+1}$. The values of JSD range from 0 when the two distributions are identical, to 1 (if the logarithm base is 2, which is typically the case in information entropy measures). The value 1 will only be reached for two distributions at disjoint domains, which does not happen for the degree distributions on the same set of nodes as is the case for the access graph (i.e., the domain is $[0, N-1]$ for all t_b values).

In contrast to the average degree, JSD is not a monotonous function of t_b and is sensitive to the choice of the time step. For this reason, a single value of t_b where JSD reaches a maximum is not a robust measure. To overcome this, we smooth the curve with a sliding window and observe the cumulative JSD in the intervals with the width equal to 10% of the total t_b interval, i.e., $w = 0.1t_{max}$. We then calculate the cumulative values with the sliding window and identify the interval of width w at which the total change in JSD is the largest. We then take the midpoint t_b value of this interval as a dual access indicator t_{JS}. Then the time $t_{JS} \pm w/2$ is the time of the fastest change in access graph structure. Similar as in the case of t_M, we propose the time relative to the maximum time budget, $\delta t_{JS} = t_{JS}/t_{max}$, and the average degree at t_{JS}, $\langle k\rangle_{JS}$, as additional indicators.

The accessibility indicators used in the analysis are summarised in Table 2.

2.5. Validation against demand

We test the ability of the new A-space representation to predict demand by comparing the structural similarity of $\mathcal{G}_A$ to a similar concept of a thresholded origin–destination, or flow, graph $\mathcal{G}_F$. The flow graph is constructed as follows:

1. Construct the origin–destination matrix (ODM) from smart card data.
2. Apply a threshold flow f_{th} on ODM, but in a reverse manner compared to $\mathcal{G}_A$; there is an edge between two nodes in $\mathcal{G}_F$ if the flow between them is at least f_{th}:

$$a_{ij} = \begin{cases} 1 & f_{ij} \geq f_{th} \\ 0 & \text{otherwise.} \end{cases} \tag{6}$$

   Note the reversed inequality sign compared to Eq. (2). This is due to the flow being a strength-type weight; the higher the flow, the better connected the nodes are. This is in inverse relationship to GTT, which is of cost-type, where large costs indicate poorer connection (larger effort to travel between the nodes). Thus, we expect the connections with lower GTT (cost) to experience larger flows (strength), which is incorporated by this reverse thresholding in $\mathcal{G}_F$ compared to $\mathcal{G}_A$.
3. Progressively decrease the flow threshold f_{th} from $f_{max} + 1$ (maximum flow across all pairs) down to $f_{min} > 0$. The $+1$ in $f_{max} + 1$ is added for $\mathcal{G}_F$ to be an empty graph in the first step, on par with the access graph at $t_b = 0$. However, note that in contrast to the access graph, $\mathcal{G}_F$ does not necessarily become a clique at the last step if there are some pairs of nodes between which no passengers travel.

We then systematically increase t_b and decrease f_{th}, and compare the structure of $\mathcal{G}_A$ and $\mathcal{G}_F$ at each (t_b, f_{th}). For this purpose, we use the graph edit distance (GED) as a measure of similarity. $GED(G_1, G_2)$ between two graphs G_1 and G_2 measures the number of elementary operations (vertex or edge deletions or insertions) needed for G_1 and G_2 to become isomorphic. It is in general an NP-hard problem and is intractable for large graphs. In our case, where the node set is constant and we are working with unweighted graphs, this simplifies to calculating the Hamming distance (or string distance) between adjacency matrices. E.g., if we have the access graph at $t_b(i)$ and its corresponding adjacency matrix of zeroes and ones, if at the time $t_b(i+1)$ one edge between nodes k and l is added, this means that the adjacency matrix in this step will be that from the previous step, except for the elements $a_{kl}(t_b = i+1) = a_{lk}(t_b = i+1) = 1$ while $a_{kl}(t_b = i) = a_{lk}(t_b = i) = 0$, to account for the symmetry of the undirected adjacency matrix. Thus, the Hamming distance between the two matrices will be $d(G(i), G(i+1)) = 2$, corresponding to the graph edit distance between two layers of the access graph $GED(G(i), G(i+1)) = d(G(i), G(i+1))/2 = 1$ (i.e., one edge is added).

If we calculate the Hamming distance between the access graph and the flow graph, it measures the total number of edges that are present in G_A and not in G_F, as well as the number of edges present in G_F and not in G_A.

In addition, we calculate the asymmetric distances d^{A-F} (number of edges that exist in $\mathcal{G}_A$ and not in $\mathcal{G}_F$) and d^{F-A} (vice versa). Then, we can identify the parts of the network that are underserved (edge exists in G_F but not in G_A) or underutilised (edge exists in G_A but not in G_F) at each point (t_b, f_{th}); we will be specifically interested in the points in which the (symmetric) Hamming distance is the smallest.

We then observe the values of (t_b, f_{th}) for which the graphs exhibit the highest similarity. This allows us to infer how the generalised travel time relates to demand.

3. Results and analysis

In this section, we apply the methodology to real-world metro networks. First, in Section 3.1, we illustrate our new A-space and the proposed global indicators using a detailed explanation of the results and a comparison between the metro networks of Washington, D.C. and Vienna. In Section 3.2, we perform a sensitivity analysis for different sets of GTT parameters, and then proceed in Section 3.3, to apply the Access Graph methodology on a dataset of metro networks of 51 cities worldwide. Finally, we explore the relationship between the access graph and the flow graph in Section 3.4for the Washington metro. For the construction of access graphs, we use publicly available in-vehicle time-weighted L-space and frequency-weighted P-space representations, which were constructed using GTFS data [50,51]. For each metro network, access graphs are built for varying time budgets. t_b is varied in the interval $[0, t_{max}]$, where t_{max} is the maximum generalised travel time in the network, with progressively increasing steps of 2 min. For the construction of the flow graph, we use Washington DC metro smart card data operated by the Washington Metropolitan Area Transit Authority (WMATA). The Python NetworkX library was used for graph construction and analysis [52].

3.1. Access graph illustration: case studies of Washington, D.C. and Vienna metro networks

We illustrate our approach for two metro networks: Washington, D.C. and Vienna. The networks were chosen based on their similar dimension of approximately 100 nodes, which is large enough for the network to exhibit smooth behaviour yet small enough to allow for meaningful visualisations. The Washington network has $N = 89$ nodes and a maximum generalised travel time of $t_{max} = 104$ minutes, and for Vienna, the values are $N = 98$ and $t_{max} = 74$ minutes.

The growth of the access graph in progressive steps of 2 min of t_b is shown in Figs. 2 for Washington and 3 for Vienna. Each subplot shows the access graph $\mathcal{G}_A(t_b)$. The edges that first appear at the respective t_b are shown in red, and the edges that were added earlier are shown in grey. In each subplot, there is an inset plot of the degree distribution at the respective value of t_b.

This behaviour is more compactly presented as heatmaps in Fig. 4. In the heatmap, the x-axis represents the time budget t_b, and y-axis represents the degree k in the respective access graph. The width of t_b and k bins are 2 min and 5, respectively. The colours of the heatmap cells represent the percentage of nodes in each degree bin. The average degree at each t_b is shown with a red marker.

At lower values of t_b, the degree distribution is skewed heavily to the right, which is also observed in Figs. 2 and 3. As the travel time budget increases, the values of the average degree, along with the minimum and maximum degree, increase and the degree distribution is progressively skewed to the left. The distributions for the intermediate values of t_b are often bimodal or even multimodal with several local maxima, which is observed for Washington for example in the histogram at $t_b = 38$ min, where the distribution is bimodal (one larger peak at $k \approx 20$, and a smaller peak at $k \approx 40$). Conversely, the degree distributions for Vienna exhibit roughly unimodal behaviour throughout the interval, where the peak of the distribution moves progressively to higher values. This can also be observed in the heatmap plots, where for Washington, two distinct S-shapes to the right of the averaged curve can be observed, marking observably faster growth of degree of a group of nodes compared to the bulk of the distribution. For the Vienna metro, the bulk stays around the average degree and gradually decreases away from it on both sides. We also observe steeper growth of average degree for the Vienna network early in the interval. From the above, we qualitatively expect that the Vienna metro is more accessible and has higher access equity than the Washington metro.

We quantitatively compare the pace at which access progresses in both network by examining how node degree distributions change across progressive Access graphs and based on the time at maximum growth of average degree, i.e., the point where its first derivative is maximal. To examine the behaviour of the derivative in more detail, we observe the shape of the average degree derivative in the whole t_b interval. For a complementary view, sensitive to the overall structure, we examine the Jensen–Shannon divergence alongside it. The results are shown in Fig. 5. We see that the shape of $\langle k \rangle'(t_b)$ is wider for Washington compared to Vienna, and the peak for Vienna is more prominent. This is in line with the disparities in level of accessibility for the Washington

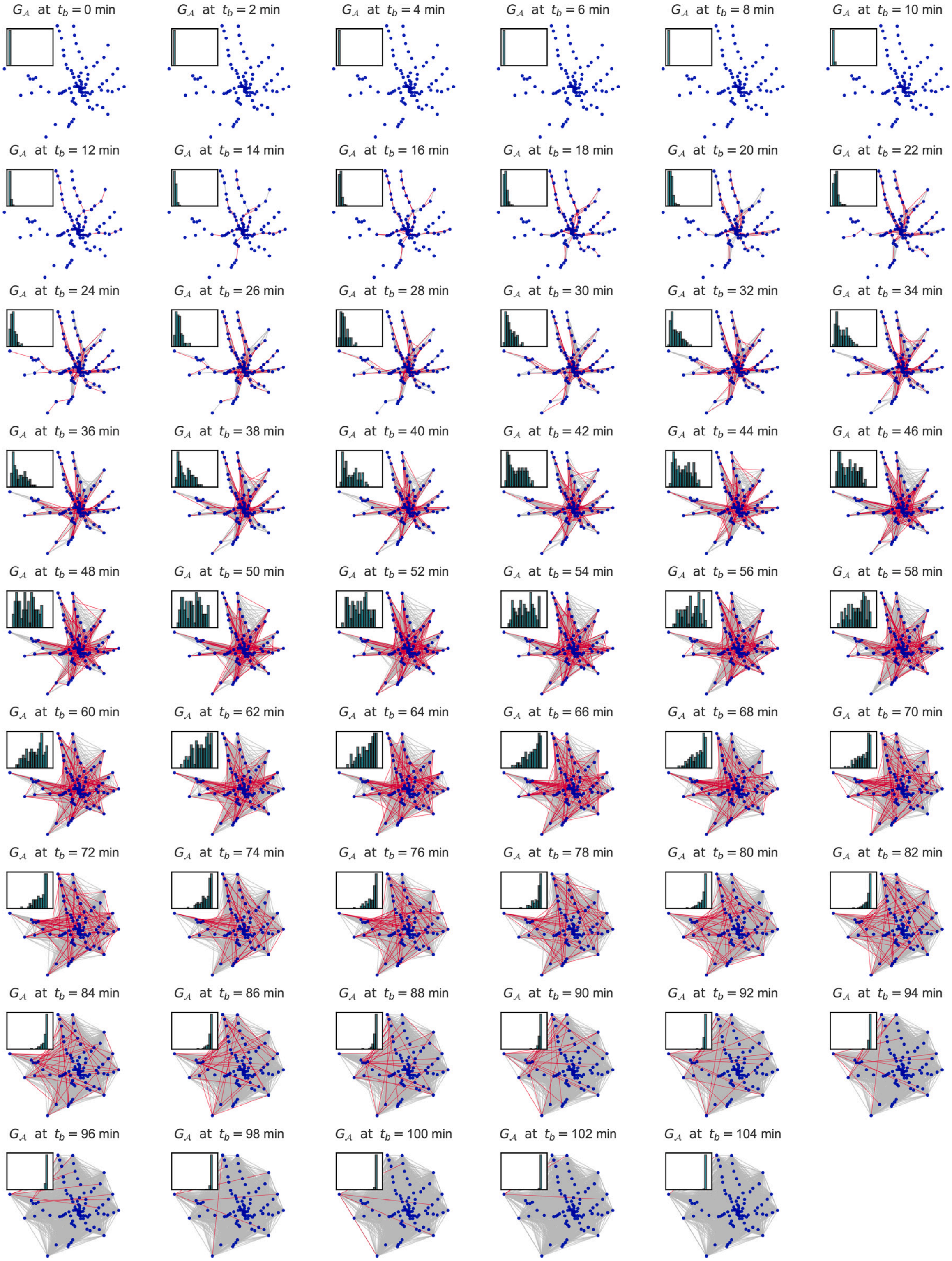

Fig. 2. A visualisation of the time budget-dependent access graph for the Washington, D.C. metro. In successive subplots, access graphs at respective values of t_b, as indicated in each subplot title, are shown. The edges that first appear at a given t_b are shown in red. Other edges (i.e., those present in $\mathcal{G}_A$ already earlier than at t_b) are shown in grey. The Washington metro has $N = 89$ nodes and a maximum generalised travel time of $t_{max} = 104$ min.

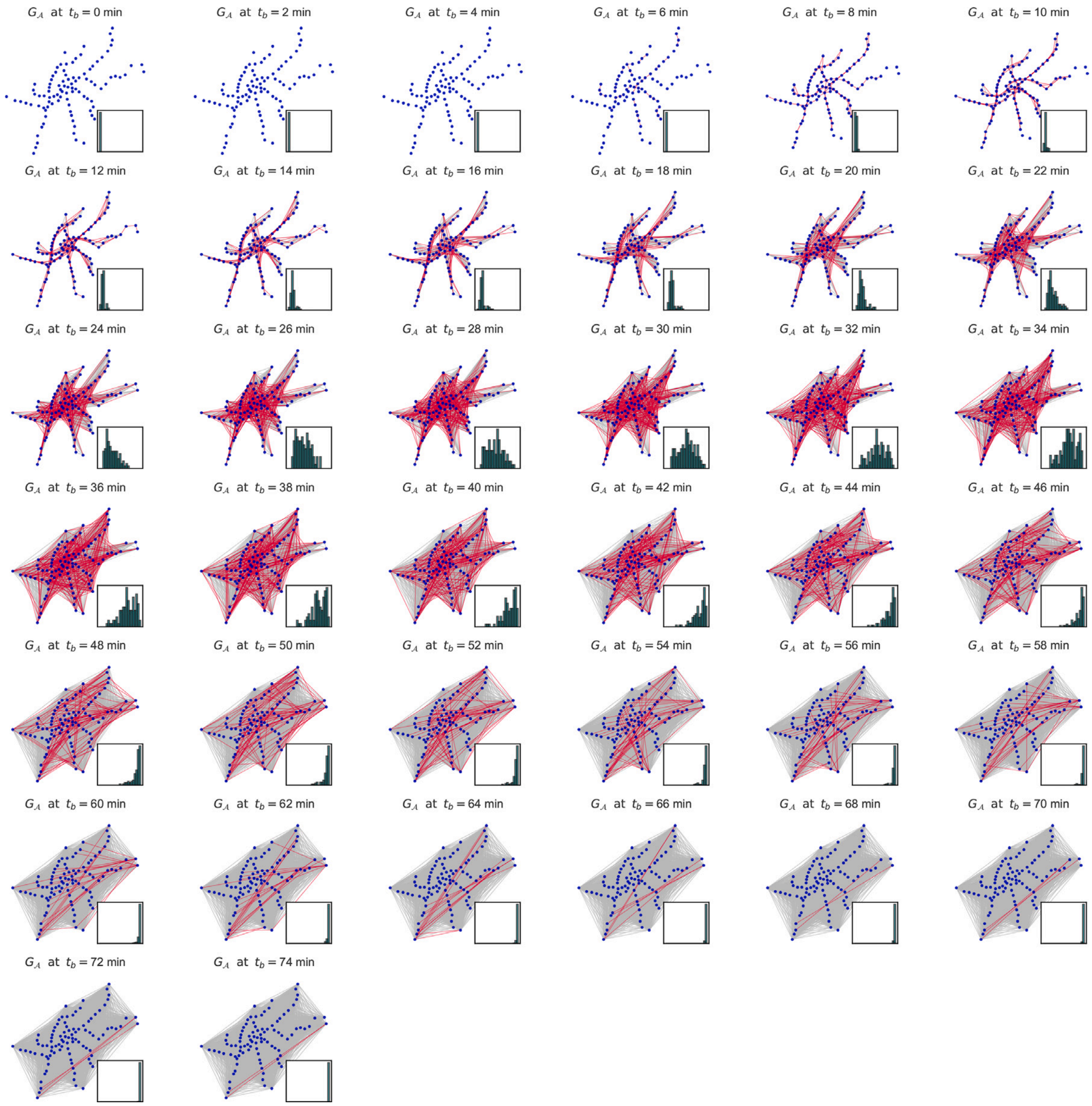

Fig. 3. A visualisation of the time budget-dependent access graph for the Vienna metro. In successive subplots, access graphs at respective values of t_b, as indicated in each subplot title, are shown. The edges that first appear at any given t_b are shown in red. Other edges (i.e., those present in $\mathcal{G}_A$ already earlier than at t_b) are shown in grey. In each subplot, the degree distribution of $\mathcal{G}_A$ is shown in the inset plot. The Vienna metro has $N = 98$ nodes and a maximum generalised travel time of $t_{max} = 74$ min.

network where the group of better connected nodes drives the change faster, but is spread out by the slower growth in degree for the majority of nodes. For the Vienna metro, the behaviour is much more regular, in line with observed higher equity of degree growth across the whole network.

An interesting observation is that JSD offers a qualitatively different perspective. For Washington, it has two peaks, first at about $t_b = 20$ min and the second near $t_b = 90$ min. This means that there are two points at which the overall structure of the degree distribution changes significantly. Returning to Fig. 2, we see that at $t_b = 22$ min, the degree distribution first exhibits a peak (compared to strictly decaying before that time point), and by $t_b = 86$ min, almost all of the nodes have the maximum degree and the distribution is heavily left-skewed, with the remaining nodes catching up in connectedness. For Vienna, the only prominent JSD peak occurs at about $t_b = 15$ min. This coincides with the value at which the access graph becomes fully connected. Note that in the access graph, by definition, degree is a monotonically increasing function, so positive JSD here always means an increase in accessibility, and a peak in JSD corresponds to a larger change.

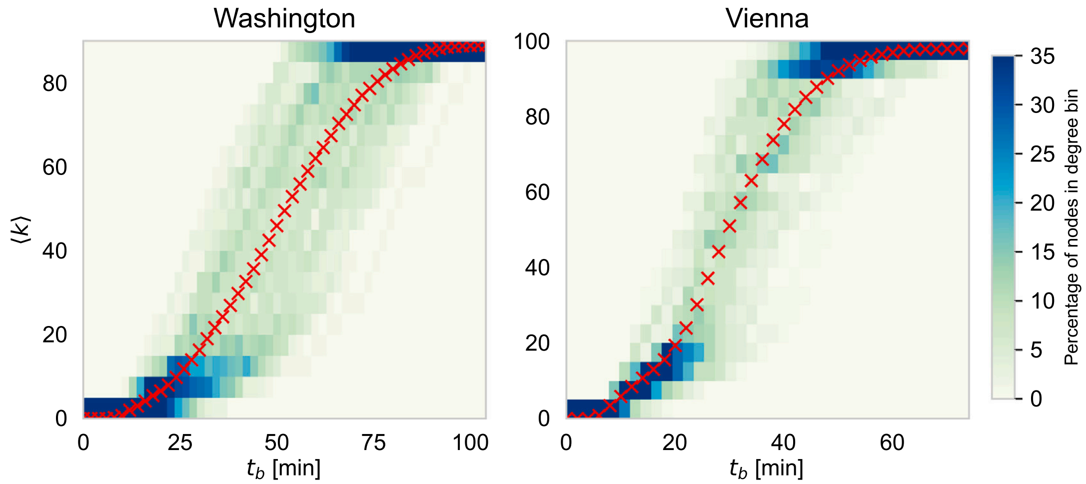

Fig. 4. Heatmaps of degree distributions with varying time budgets for the Washington (left) and Vienna (right) metros. t_b is increased in steps of 2 min. The x-axis represents the time budget, and the y-axis represents the node degree in the respective access graph. Heatmap cell colour represents the percentage of nodes in each degree bin. Thus, each vertical column represents the colour-coded histogram of the node degree distribution. The average degree at each t_b bin is plotted with red markers.

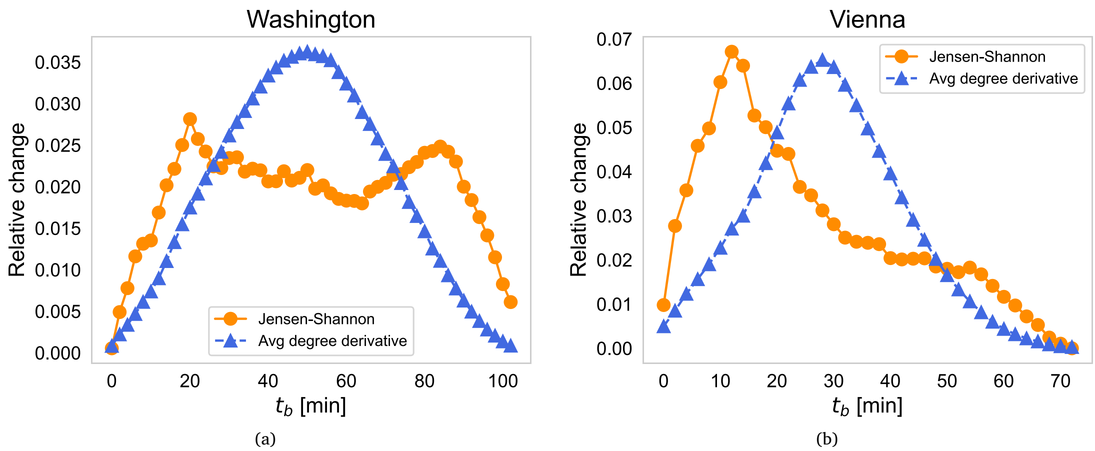

Fig. 5. Washington (left) and Vienna (right).

3.2. Sensitivity analysis

In this section, we perform a sensitivity analysis of the introduced access indicators to the values of the weight of waiting time and transfer penalty in the definition of GTT (Eq. (1)) for the Washington metro. We vary w^{wait} from 0 to 10, and $w^{transfer}$ from 0 to 14. The results are shown as heatmaps in Fig. 6.

The results show that absolute value indicators, which are fundamentally constrained by time (t_{max}, $\langle k\rangle_{30}$, G_{30}), worsen monotonically with increasing weight parameter values, which is expected. t_M, which represents a network characteristic point, shows a generally monotonic behaviour, albeit with some noise compared to the absolute values, which is even more prominent for δt_M, $\langle k\rangle_M$ and G_M which derived from t_M. Rather than a structural observation, this is likely an indicator artefact, as t_M is obtained by the numerical calculation of the first derivative, and might be sensitive to discrete steps in w^{wait} and $w^{transfer}$. At the same time, specific combinations of $(w^{wait}, w^{transfer})$ may interact to cause a local change in the structure. Notable is also that for all indicators, the indicator behaviour is relatively stable with $w^{transfer}$ and almost all the change is due to w^{wait}. However, the Washington metro has a diameter of 2 in P-space, which means that it is possible to travel between all pairs of points with at most one transfer. A different result might present in larger networks with larger diameters, such as bus networks. An interesting observation is that δt_M remains remarkably stable across almost the full range of parameters, strengthening the point that it is a structural performance parameter influenced primarily by the network topology and thus useful for characterising a network.

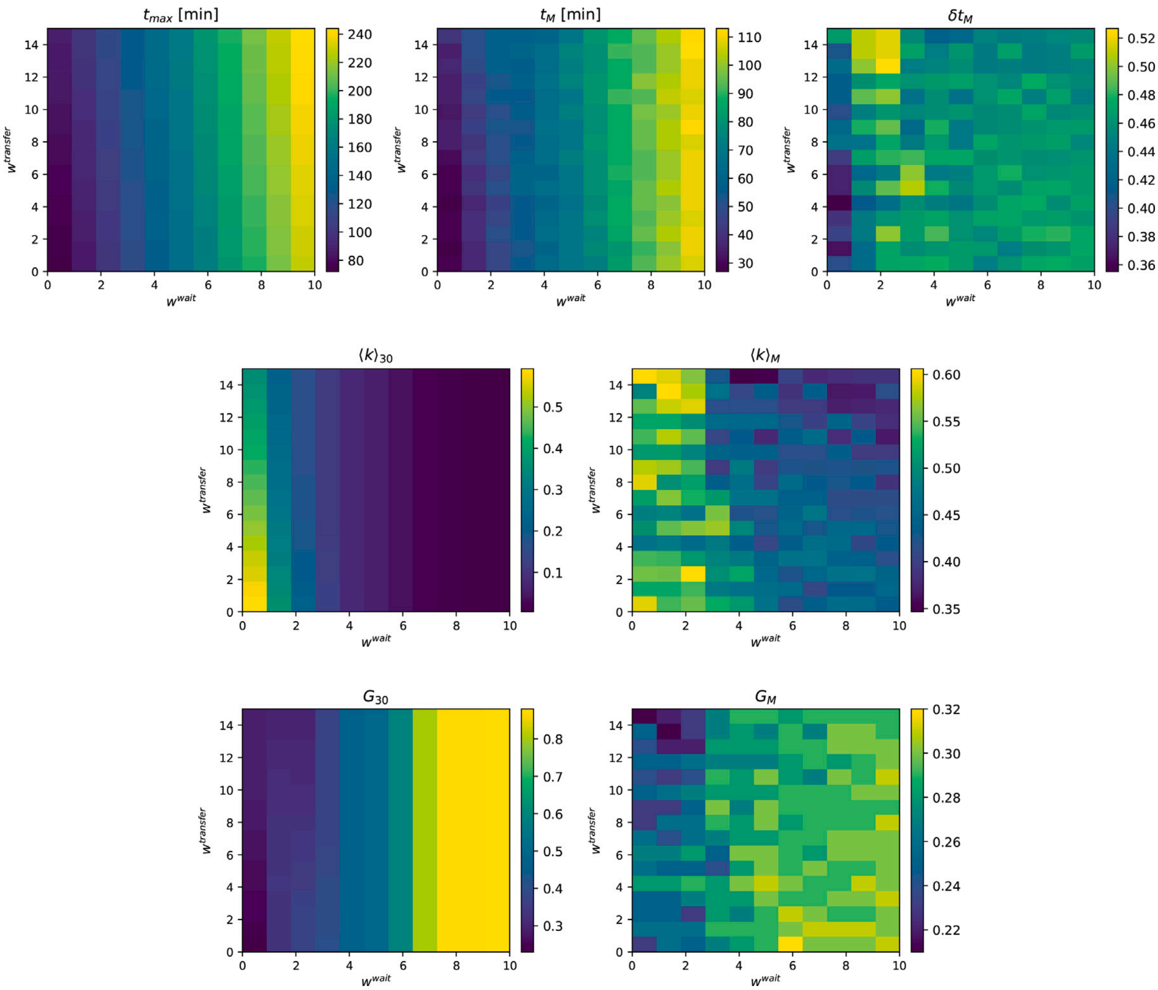

Fig. 6. Sensitivity analysis of the introduced accessibility indicators with respect to varying the w^{wait} (x-axis) and $w^{transfer}$ (y-axis) parameters in the definition of GTT (Eq. (1)).

3.3. Results for 51 metro networks worldwide

In this section, we turn to numerical results for the access indicators introduced in 2.4 for all 51 metro networks included in our analysis. The discussion in the previous Section 3.1 serves as a basis for interpreting the reported values. The heatmap plots of degree distribution and the $\langle k\rangle'(t_b)$ and $JSD(t_b)$ plots that were shown for Oslo and Vienna in Figs. 4 and 5 are given for all cities in Figs. 7 and 8. In the latter figures, each of the subplots represents a separate network (city) and shows the time evolution of the observed quantities.

Average degree-based access indicators

In almost all cases, a logistic-like S-shaped convergence of average degree $\langle k\rangle$ to the maximum degree is observed. The most regular S-shaped behaviour is observed for the largest networks (New York, London and Paris, among others). In these cases, the value of t_M corresponds to the inflection point (i.e. the "turning point" of the S-curve). Several networks exhibit a more complex behaviour with multiple points of inflection in $\langle k\rangle(t_b)$ (e.g. Athens, Los Angeles and Naples), implying several local extrema of $\langle k\rangle(t_b)$. Upon closer inspection, we observe line-level effects (e.g. line length and the importance of transfers in connecting node pairs) and the general topology of the L-space graph as likely predictors of this behaviour.

The numerical values of all indicators are summarised in Table 3. The networks are of remarkably diverse dimensions, ranging from $N = 8$ (Genoa) to $N = 421$ (New York). Maximum generalised travel time ranges from $t_{max} = 20$ (Rennes) to $t_{max} = 152$ minutes (New York, San Francisco). Large metropolitan areas are expected to have large values of t_{max}, e.g. 120 and 138 min for London ($N = 261$) and Madrid ($N = 240$), respectively. An exception is Paris with $N = 303$ and $t_{max} = 66$ min suggesting a high level of accessibility. On the other hand, moderately sized networks with high values of t_{max}, e.g. Valencia ($N = 95, t_{max} = 140$), Chicago

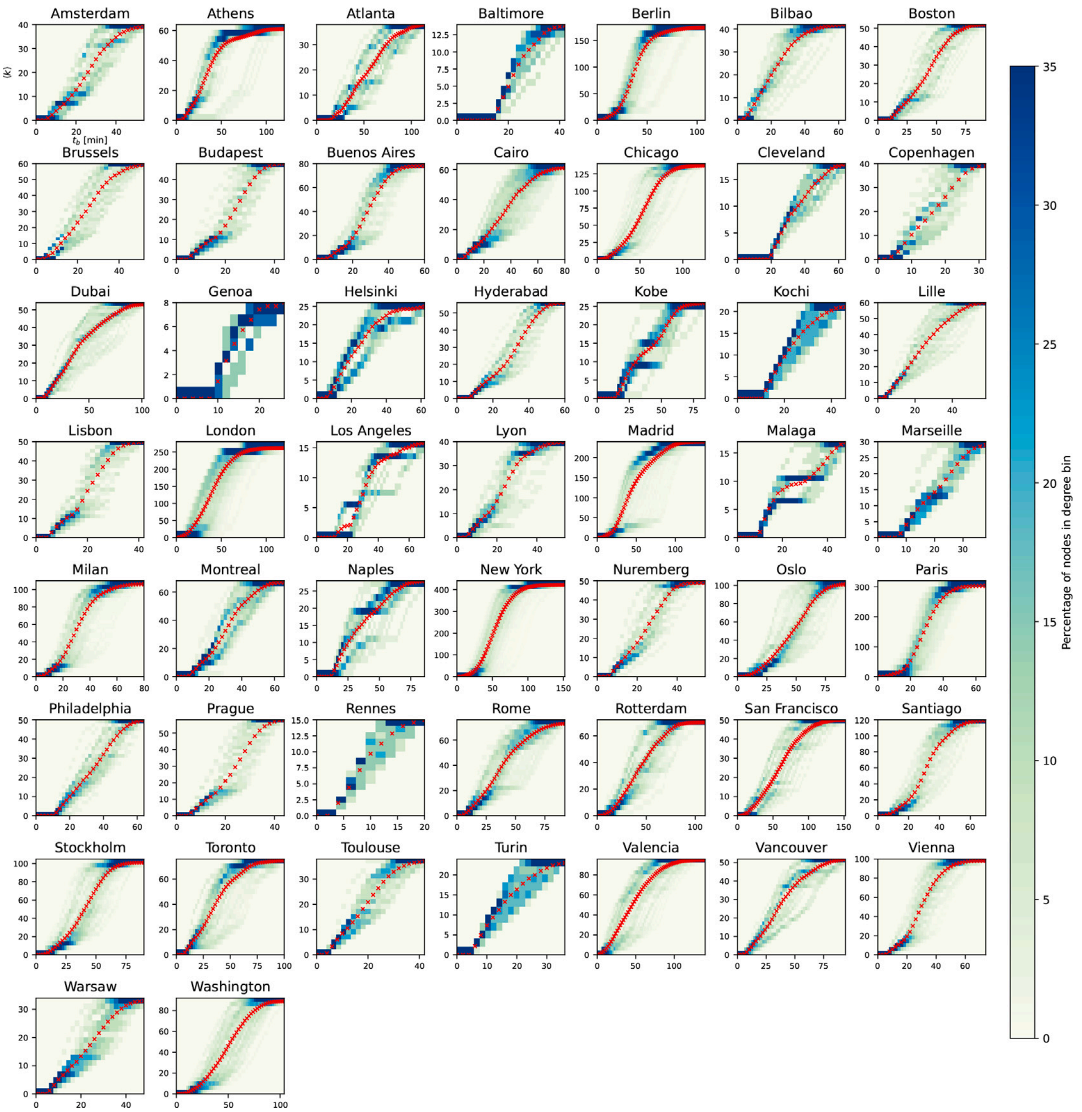

Fig. 7. Degree distributions of access graphs with varying t_b. Each subplot shows the time evolution of the access graph with the heatmaps representing node degree distributions. t_b is increased in steps of 2 min. In each subplot, the x-axis represents the time budget, and y-axis represents the degree in the corresponding access graph. Heatmap cell colours represent the percentage of nodes in each degree bin. Average degree of the access graph at t_b is shown with red markers.

($N = 137, t_{max} = 126$) or Athens ($N = 61, t_{max} = 120$) are less accessible, with the extreme case of San Francisco with $N = 50$ and $t_{max} = 152$.

Nine of the networks comprise of a single line: these are the metros in Baltimore, Cleveland, Genoa, Helsinki, Kochi, Los Angeles, Malaga, Rennes, and Turin. Their dimension ranges from 8 (Genoa) to 25 (Helsinki). For one-line networks, the growth of the access graph depends only on in-vehicle times as there are no transfers, and waiting time is uniform across the network.

The access inflection point values range from $t_M = 7$ (Rennes) to $t_M = 61$ minutes (San Francisco). Note that almost half of the networks (23 networks) have t_M in the interval between 25 and 35 min, meaning that the operational characteristic time is close to the passengers'-oriented value of 30 min. The values of δt_M range from 0.15 (Valencia) 0.66 (Copenhagen). In general, low values of the normalised access inflection point indicate higher accessibility, but this is not necessarily the case. For example, Valencia has the lowest value of δt_M, however the value of access at inflection point $\langle k \rangle_M = 0.19$ shows that degree growth is far from saturation. Conversely, the Copenhagen metro reaches the highest value among networks with $\langle k \rangle_M = 0.75$ and is one of the best connected networks with $t_{max} = 32$ and $t_M = 21$ minutes at $N = 39$. The lowest value of access at inflection point is $\langle k \rangle_M = 0.14$ for Marseille, and the largest $\langle k \rangle_M = 0.79$ for Copenhagen ($\langle k \rangle_M = 0.75$ for Oslo). However, for Oslo we have seen in Section 3.1 that the peak of average degree derivative is reached late in the interval, confirmed by the value of $\delta t_M = 0.64$. Therefore, the value of $\langle k \rangle_M$ alone is

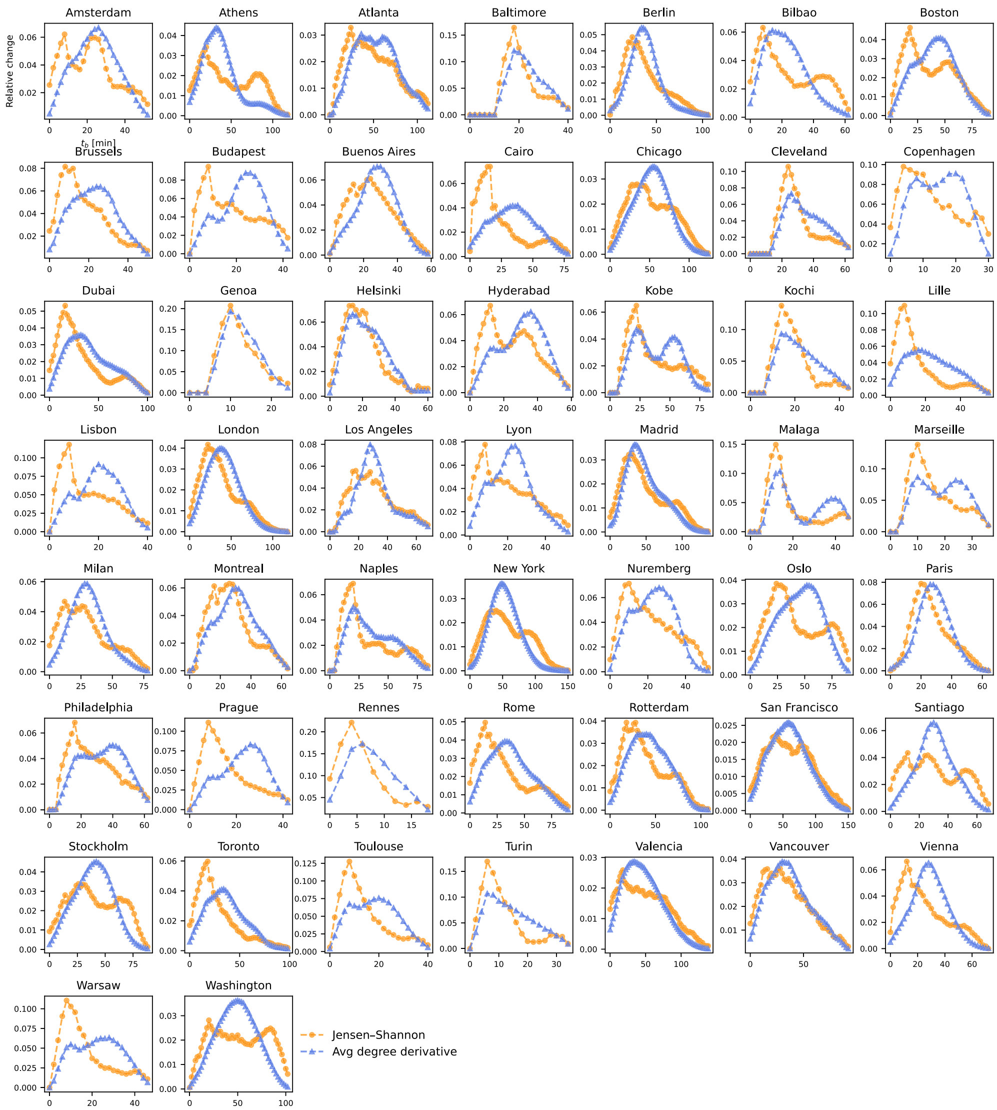

Fig. 8. Average degree derivative and Jensen–Shannon divergence vs. t_b plots for all cities.

not a sufficient indicator of access. These results indicate that for a more complete view of accessibility, a combination of primal and dual access indicators is necessary for assessing the overall network performance with respect to access. For the 30-minute access indicators, the average degree at $t_b = 30$ min ranges from $\langle k \rangle_{30} = 0.11$ for New York to $\langle k \rangle_{30} = 1.0$ for Genoa and Rennes. Since the maximum travel time varies largely across networks, the value of $\langle k \rangle_{30} = 0.11$ is expected to be lower for cities with large values of t_{max} (152 min for New York), and in the case of Genoa and Rennes the maximum travel time is below 30 min, with $t_{max} = 26$ min and $t_{max} = 20$ min, respectively.

JSD-based access indicators

The biggest structural transition of the access graph, as reflected in the Jensen–Shannon divergence, consistently occurs earlier in the interval than the access inflection point. The values of the access structural change point t_{JS} range from 3 min (Rennes) to 48 min (San Francisco). The normalised values range from $\delta t_{JS} = 0.08$ for Bilbao and Valencia to $\delta t_{JS} = 0.38$ for Baltimore. The values of access at structural change point range from $\langle k \rangle_{JS} = 0.02$ for Bilbao to $\langle k \rangle_{JS} = 0.35$ for San Francisco. An exception is the Genoa metro where $\langle k \rangle_{JS} = 0.0$, due to the short time interval and simple degree change behaviour for a network with one line and 8 nodes. These results indicate that the largest structural reorganisation of the access graph, and thus the change in accessibility levels, will occur relatively early in the interval and consequently at low average degrees. In comparison, the point at the maximum growth of average degree often occurs when the structural changes are relatively small (e.g. Boston, Cairo, Oslo, or Washington, Fig. 8). We also stress that whereas the average degree derivative vs. t_b exhibits one distinctive peak for most cities (the exceptions

Table 3
Values of the studied variables for each network (city). N: number of nodes in the network; t_{max}: maximum generalised travel time (in minutes); t_M: access inflection point; $\delta t_M = \frac{t_M}{t_{max}}$ normalised access inflection point; t_{JS}: access structural change point; $\delta t_M = \frac{t_{JS}}{t_{max}}$ normalised access structural point; $\langle k \rangle_M$: access at inflection point; $\langle k \rangle_{JS}$: access at structural change point; $\langle k \rangle_{30}$: 30-min access; G_M: inflection access inequality; G_{30}: 30-min access inequality.

City	N	t_{max}	t_M	δt_M	t_{JS}	δt_{JS}	$\langle k \rangle_M$	$\langle k \rangle_{JS}$	$\langle k \rangle_{30}$	G_M	G_{30}
Amsterdam	39	54	29	0.54	7	0.13	0.67	0.06	0.67	0.18	0.18
Athens	61	120	31	0.26	14	0.12	0.47	0.11	0.42	0.27	0.28
Atlanta	38	114	39	0.34	20	0.18	0.33	0.06	0.17	0.27	0.38
Baltimore	14	42	21	0.50	16	0.38	0.51	0.11	0.85	0.16	0.11
Berlin	174	108	37	0.34	19	0.18	0.55	0.10	0.31	0.22	0.29
Bilbao	42	64	17	0.27	5	0.08	0.42	0.02	0.76	0.20	0.13
Boston	52	92	45	0.49	15	0.16	0.59	0.08	0.28	0.21	0.25
Brussels	59	52	29	0.56	11	0.21	0.72	0.17	0.72	0.16	0.16
Budapest	48	44	25	0.57	6	0.14	0.62	0.04	0.80	0.20	0.13
Buenos Aires	78	60	29	0.48	21	0.35	0.59	0.23	0.59	0.21	0.21
Cairo	61	80	35	0.44	12	0.15	0.55	0.12	0.42	0.22	0.24
Chicago	137	126	49	0.39	38	0.30	0.45	0.26	0.16	0.29	0.37
Cleveland	18	64	25	0.39	23	0.36	0.29	0.17	0.44	0.11	0.10
Copenhagen	39	32	21	0.66	4	0.12	0.79	0.03	1.00	0.14	0.00
Dubai	53	102	33	0.32	11	0.11	0.46	0.07	0.37	0.17	0.19
Genoa	8	26	11	0.42	9	0.35	0.45	0.00	1.00	0.28	0.00
Helsinki	25	62	11	0.18	17	0.27	0.21	0.32	0.76	0.28	0.14
Hyderabad	56	60	35	0.58	9	0.15	0.64	0.03	0.44	0.17	0.19
Kobe	26	84	21	0.25	18	0.21	0.22	0.09	0.40	0.26	0.19
Kochi	21	46	13	0.28	12	0.26	0.20	0.09	0.82	0.12	0.11
Lille	60	58	15	0.26	7	0.12	0.33	0.11	0.68	0.17	0.15
Lisbon	50	42	19	0.45	6	0.14	0.50	0.05	0.90	0.22	0.08
London	261	120	41	0.34	16	0.13	0.56	0.09	0.31	0.24	0.37
Los Angeles	16	70	31	0.44	26	0.37	0.59	0.28	0.48	0.16	0.20
Lyon	40	54	25	0.46	7	0.13	0.63	0.09	0.78	0.20	0.14
Madrid	240	138	33	0.24	31	0.22	0.31	0.27	0.23	0.37	0.40
Malaga	17	48	11	0.23	12	0.25	0.20	0.18	0.62	0.13	0.14
Marseille	29	38	9	0.24	10	0.26	0.15	0.14	0.90	0.17	0.08
Milan	106	80	29	0.36	8	0.10	0.52	0.04	0.52	0.26	0.26
Montreal	67	66	33	0.50	13	0.20	0.62	0.08	0.49	0.17	0.21
Naples	28	86	17	0.20	16	0.19	0.19	0.11	0.47	0.34	0.23
New York	421	152	51	0.34	44	0.29	0.49	0.33	0.11	0.25	0.42
Nuremberg	49	54	25	0.46	9	0.17	0.58	0.06	0.71	0.20	0.17
Oslo	101	92	59	0.64	29	0.32	0.75	0.19	0.22	0.15	0.30
Paris	303	66	27	0.41	19	0.29	0.49	0.18	0.57	0.25	0.22
Philadelphia	50	64	39	0.61	13	0.20	0.65	0.03	0.42	0.14	0.18
Prague	58	44	27	0.61	6	0.14	0.69	0.03	0.77	0.18	0.14
Rennes	15	20	7	0.35	3	0.15	0.51	0.12	1.00	0.12	0.00
Rome	73	92	35	0.38	11	0.12	0.50	0.08	0.37	0.21	0.24
Rotterdam	70	110	39	0.35	24	0.22	0.47	0.18	0.28	0.24	0.28
San Francisco	50	152	61	0.40	48	0.32	0.55	0.35	0.17	0.23	0.35
Santiago	119	70	31	0.44	10	0.14	0.53	0.04	0.45	0.22	0.25
Stockholm	101	90	45	0.50	30	0.33	0.63	0.27	0.28	0.21	0.33
Toronto	75	100	29	0.29	13	0.13	0.41	0.09	0.41	0.28	0.28
Toulouse	37	42	21	0.50	6	0.14	0.66	0.09	0.91	0.16	0.07
Turin	23	36	7	0.19	6	0.17	0.22	0.09	0.99	0.09	0.04
Valencia	95	140	21	0.15	11	0.08	0.20	0.07	0.32	0.40	0.33
Vancouver	52	94	31	0.33	25	0.27	0.42	0.26	0.38	0.20	0.20
Vienna	98	74	27	0.36	10	0.14	0.46	0.06	0.53	0.25	0.23
Warsaw	33	48	27	0.56	8	0.17	0.67	0.09	0.74	0.17	0.15
Washington	89	104	51	0.49	27	0.26	0.56	0.16	0.19	0.25	0.35

are mostly the one line networks), the JSD shape is not as universal across networks and often exhibits another peak, or several peaks, later in the interval, which is not captured in the devised indicators (e.g. Oslo, Santiago, Stockholm).

Access equity indicators

Next, we turn to investigating access equity. Values of inflection access inequality G_M and 30-min access inequality G_{30} are in the interval between ≈ 0.05 and ≈ 0.4, with the exception of networks with $t_{max} \lesssim 30$ min where $G_{30} = 0$. This happens for two of the one-line networks, Rennes and Genoa, as well as Copenhagen. The value of $G_M = 0.14$ is a further indicator of the quality of the Copenhagen metro. The highest disparities are observed for Valencia ($G_M = 0.40, G_{30} = 0.33$) and Madrid ($G_M = 0.37, G_{30} = 0.40$). The largest value of $G_{30} = 0.42$ is observed for New York however, with its large size and maximum travel time, this is expected and the value $G_M = 0.25$ indicates moderate inequality.

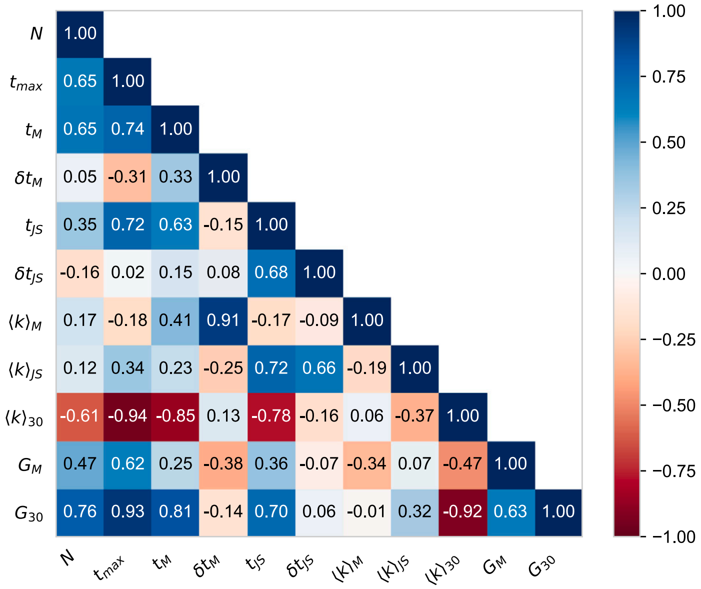

Fig. 9. Spearman correlation matrix for accessibility variables. N: number of nodes in the network; t_{max}: maximum generalised travel time (in minutes); t_M: access inflection point; $\delta t_M = \frac{t_M}{t_{max}}$ normalised access inflection point; t_{JS}: access structural change point; $\delta t_M = \frac{t_{JS}}{t_{max}}$ normalised access structural point; $\langle k \rangle_M$: access at inflection point; $\langle k \rangle_{JS}$: access at structural change point; $\langle k \rangle_{30}$: 30-min access; G_M: inflection access inequality; G_{30}: 30-min access inequality.

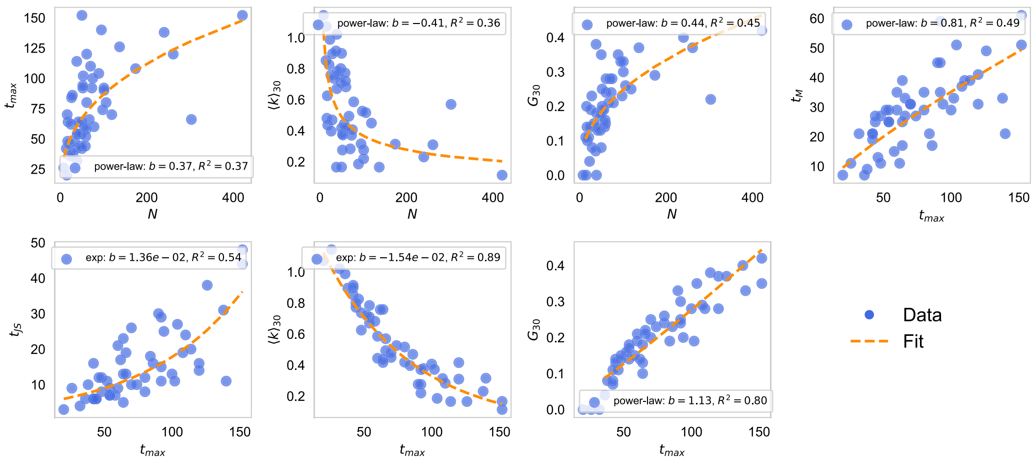

Fig. 10. Scatter plots for selected pairs of variables with the highest observed values of Spearman correlation coefficients. In each subplot, one data point corresponds to a metro network. Included are the best power-law or exponential fits with parameter values and goodness of fit. The fits are performed in log–log (log) scale for power-law (exponential) fits, and R^2 is determined in linear scale.

For the illustrative example of the Oslo and Vienna metros in the previous Section 3.1, the values of the accessibility indicators are (for each indicator, the values are given first for Oslo, second for Vienna): $t_M = 27$ and $t_M = 59$; $\delta t_M = 0.36$ and $\delta t_M = 0.64$; $\langle k \rangle_M = 0.46$ and $\langle k \rangle_M = 0.75$; $\langle k \rangle_{30} = 0.53$ and $\langle k \rangle_{30} = 0.22$; $G_M = 0.25$ and $G_M = 0.15$; $G_{30} = 0.23$ and $G_{30} = 0.30$, confirming the qualitative observation of Vienna having higher overall accessibility as well as access equity than Oslo.

3.3.1. Statistical analysis

To better understand the relationships between different indicators, the Spearman correlation matrix for all variables is shown in Fig. 9. Spearman correlation is chosen over Pearson due to non-linear relations between several pairs of variables. To examine in more detail the relations of the most significantly correlated pairs of variables with non-linear relationships, their scatter plots alongside the best power-law or exponential fits are shown in Fig. 10.

The maximum travel time t_{max} varies per city and increases sub-linearly with the size of the network N (Fig. 10). This indicates a relatively better performance for larger networks. This is expected due to the approximately universal values of individual travel

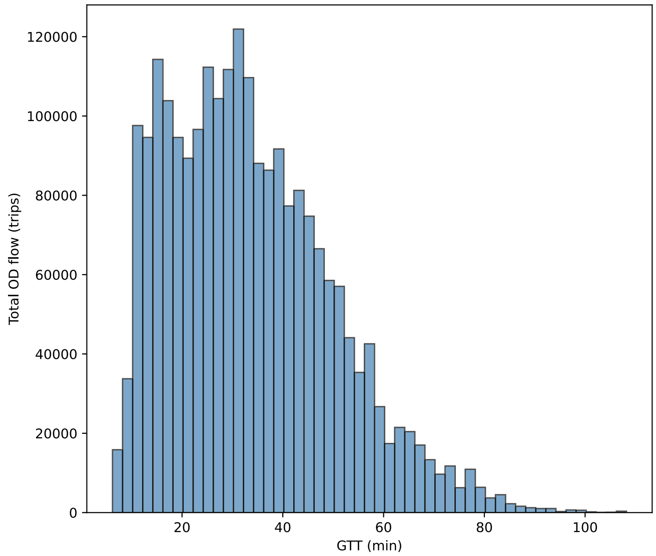

Fig. 11. A histogram of passenger flows, as measured in total number of trips, with respect to GTT for the Washington metro.

time budgets, and in PTN planning t_{max} must have a reasonable upper bound regardless of the size of the network. Note here that N is correlated with city geographical area and population size.

Interestingly, the correlation of the access inflection point t_M with its normalised counterpart $\delta t = t_M/t_{max}$ is relatively low ($r_S \approx 0.4$), suggesting the potential added value in using the combination of these indicators for measuring and comparing accessibility. Notable is the lack of correlation between δt_M and N, indicating that this indicator offers a size-independent measure of access.

The correlation of t_{JS} with N is $r = 0.35$, significantly lower than for t_M with N. This suggests that the position of the time interval where the structural reorganisation of the access graph is the largest is largely independent of network size. In contrast, the correlation with t_{max}, $r = 0.72$ is significant and comparable to $r = 0.74$ for t_M and N. The correlations of the normalised values δt_{JS} with N ($r = -0.16$) and t_{max} ($r = 0.02$) are largely insignificant and indicate a lack of uniform behaviour of the JSD-based measures.

The values of the 30-min type indicators, the 30-min access $\langle k \rangle_{30}$ and the 30-min access inequality G_{30} correlate significantly with the size of the network N and maximum generalised travel time t_{max}. The relations are shown in Fig. 10. Strong negative correlations and negative power-law relations are observed for $\langle k \rangle_{30}$ vs. N, while the shape of $\langle k \rangle_{30}$ vs. t_{max} is best described by exponential decay. In contrast, the values of G_{30} increase sub-linearly, following a power-law. For the largest networks, with size $N \gtrapprox 150$, the value of the Gini coefficient is approximately constant, indicating that beyond a certain size, larger networks tend to offer equally unequal distributions of access, in contrast to the relation observed for small and moderately sized networks (i.e. the growth of G_M stalls after this value of N).

The strong negative correlation of $\langle k \rangle_{30}$ and G_{30} complements the above observations, and the high positive correlation is expected due to similar patterns of behaviour for all networks. For the primal indicators, the strong positive correlation of δt with access at inflection point $\langle k \rangle_M$ points to a universal behaviour of average degree growth (observed to be logistic-like in most cases) across network sizes. The moderate positive correlations of inflection access inequality G_M and N, and G_M and t_{max} indicate that network size does impact access inequality, with larger networks tending to have greater disparities in access. The significant negative correlation between the Gini index and δt indicates that inequity decreases with δt, complementing the previous observation on the relation to N.

3.4. Structural comparison of supply and demand

In this section, we validate the access graph methodology against demand by comparing the structure of the access graph $\mathcal{G}_A$ and the flow graph $\mathcal{G}_F$, where we systematically vary the threshold parameters t_b and f_t, respectively. To construct the flow graph, we use Washington DC metro smart card data, operated by WMATA, for a period of 12 days in September 2019 (September 1st to 12th). Note that this was before the COVID pandemic, which brought severe disruptions to PT use worldwide.

During this time, a total of over 4.5 million trips were recorded. The Washington metro is a tap-in, tap-out system, so the origin–destination matrix is readily obtainable from data. Fig. 11 shows the histogram of passenger flows, measured in the total number of trips. We see that the majority of the trips are made for values of GTT between approximately 20 and 40 min, with a mean value of 33.5 min, which is in good agreement with our theoretically determined fixed value of 30 min.

In the following, we systematically compare the structural differences of $\mathcal{G}_A$ and $\mathcal{G}_F$ at the link level. For each pair of threshold values (t_b, f_t), we observe the symmetric and asymmetric Hamming distances between them: by d, we denote the symmetric distance,

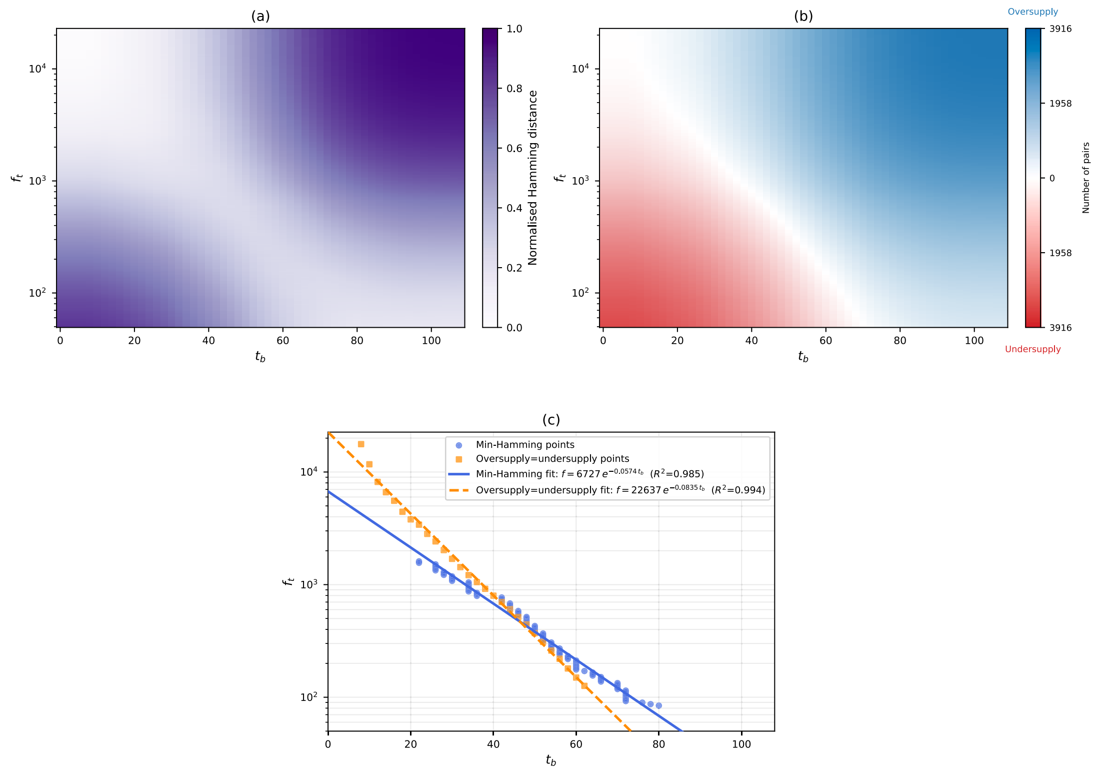

Fig. 12. Structural differences between $\mathcal{G}_A$ and $\mathcal{G}_F$. Figure (a) shows the heatmap of the Hamming distance for each pair (t_b, f_t); darker shades mean larger distance and thus larger dissimilarity. Figure (b) shows the asymmetric Hamming distance where oversupply ($d^{A-F} > d^{F-A}$) is shown in blue and undersupply ($d^{A-F} < d^{F-A}$) is shown in red, the colour strength indicating the distance. In the white belt, the number of oversupply edges equals the undersupply edges; note that this does not indicate equality of graphs, as the Hamming distance is still about 20% as seen in Figure (a). In figure (c), the (t_b, f_t) points for minimum Hamming distance and over- and undersupply equality points are extracted and the exponential fit is performed on both. Note that the y-axis is shown in logarithmic scale in all plots.

while d^{A-F} (d^{F-A}) measures the number of edges present in $\mathcal{G}_A$ ($\mathcal{G}_F$) and not in $\mathcal{G}_F$ ($\mathcal{G}_A$). We are specifically interested in the values of parameters at which the structural differences are the smallest, i.e., when the two graphs most resemble each other. The results are shown in Fig. 12.

Fig. 12(a) shows the heatmap of the Hamming distance for each pair (t_b, f_t); darker shades mean larger distance and thus larger dissimilarity. Fig. 12(b) shows the asymmetric Hamming distance where oversupply ($d^{A-F} > d^{F-A}$) is shown in blue and undersupply ($d^{A-F} < d^{F-A}$) is shown in red, the colour strength indicating the distance. In the white belt, the number of oversupply edges equals the number of undersupply edges; note that this does not indicate equality of graphs, as the Hamming distance is still about 20% as seen in Figure (a). In Fig. 12(c), the (t_b, f_t) points for minimum Hamming distance and over- and undersupply equality points are extracted and the exponential fit is performed on both. Note that the y-axis is shown in logarithmic scale in all plots.

Before the crossing point at $t_b = 46$ min, the difference between the access and the flow graph is due primarily to undersupply, and after that point it changes to the oversupply regime. This is in line with the known characteristics of PTN design, where less travelled paths are oversupplied to provide minimal access, while a percentage of the busiest OD pairs will remain undersupplied. This analysis enables to find the travel time, at which the regime changes. We note also that the regime change value of $t_b = 46$ min is close to the t_M indicator value of 51 min, although we emphasise that this does not constitute a statistical finding.

The asymmetric difference between $\mathcal{G}_A$ and $\mathcal{G}_F$ is shown in Fig. 13 for one of the structural transition points, at the value of $t_b = 25$ minutes, and the corresponding value $f_t = 506$ trips. Red and blue edges indicate undersupply and oversupply, respectively. Oversupply edges are mostly those between stations that are close on the same line or geographically close, while undersupply at $t_b = 25$ min is mostly present along the more distant stations on the same lines. This is in line with the design of the Washington DC metro, where the lines are constructed to cater to high demand between terminal stations and the city centre. Here, the oversupply edges are primarily those along the way between these high-demand nodes. At the same time, at $t_b = 25$ minutes, the physical distance and the corresponding in-vehicle times are the main drivers of undersupply. Note that the same line can exhibit both over- and undersupply, depending on the OD pairs considered. We discuss some possible modifications to account for inter-station passenger loads in Section 4.2.

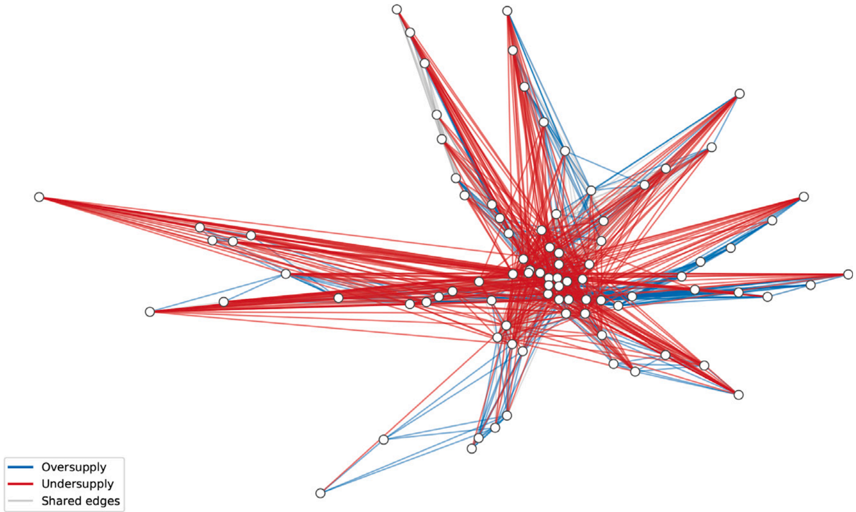

Fig. 13. Difference between $\mathcal{G}_A$ and $\mathcal{G}_F$ at $t_b = 25$ minutes. Red = undersupply; blue = oversupply.

4. Discussion and conclusion

We proposed a novel graph representation of public transport networks, the access graph $\mathcal{G}_A$, also denoted here as A-space. The edges in $\mathcal{G}_A$ are based on shortest paths weighted with generalised travel times, as determined from the weighted L- and P-space representations, and directly connect nodes reachable from one another within a given time budget.

We studied the time budget-dependent topology of the access graph, specifically the node degree evolution. We compared the structure of the access graph to the analogously constructed flow graph and found the values of travel times and flows for which the two network are structurally most similar. We found distinct patterns of GTT threshold and passenger flows, which follow an exponential decay. Moreover, using this approach, we were able to identify different regimes of network performance as observed by over- and undersupply with respect to demand. Average degree and degree distributions of the access graphs for 51 metro networks were examined, and a set of global accessibility indicators was proposed. First, the time budget at the fastest growth of average degree t_M, and average degree at this time, $\langle k \rangle_M$, along with the relative value of the maximal growth time to maximum travel time, δt_M, examine the access graph behaviour at a critical point. Although the observed values are averaged quantities, the determination of the point at which they are observed follows from the global evolution of the whole access graph. The Gini coefficient G_M of the degree distribution at the same value of time budget was used as an access equity indicator. These measures reflect network properties and can also be understood as network quality indicators. On the other hand, the value of the average degree $\langle k \rangle_{30}$ and the Gini coefficient G_{30} at a fixed time budget of 30 min provide a complementary view from the passengers' perspective, offering a yardstick for comparing networks. In addition to studying the averaged values, we examine the overall shape of the degree distribution and its change between consecutive time steps by observing the Jensen–Shannon divergence of the degree distribution. An important insight is that it offers an alternative perspective to the average degree change, which amounts to a complementary metric.

The introduced approach addresses an important gap in accessibility research. Our operationalisation of accessibility, based on graph theory and network science, contributes a methodology to unify accessibility studies. Instead of devising intricate metrics to include different properties of the transport system in the existing representations, we propose an algorithmic approach where each step is well-defined, and the end result is a new graph representation where the simplest metric, the node degree, offers reliable, interpretable, and transferable indicators. The widespread availability of standardised GTFS data is leveraged to build the standard PT graph representations on which our methodology is based. This makes the proposed approach readily generalisable to other networks of all PT modalities at different scales. Importantly, the methodology can be applied with modifications to each step, for example, using different edge weights or different functions of distance, and arrive at similar definitions of indicators without compromising interpretability. Moreover, the observed similarity in the behaviour of the proposed measures across all 51 metro networks included in our analysis supports the relevance of our approach for accessibility analysis.

The presented framework also carries concrete implications for planners. First, the use of the A-space representations allows urban and transport planners to assess what parts of the network meet the X-minutes objectives and, by way of simulation, how potential interventions (e.g. changes in line planning, service frequencies and speed improvements) will impact the accessibility of different parts of a city in terms of PTN reachability. Second, where demand data is available, the access graph approach enables planners to identify over- and under-supplied parts of the public transport network and assess how changes in service plans may reduce the discrepancy between supply and demand in the network.

In the following, we discuss some possibilities for modifications to the access graph methodology and propose some ways to adapt the construction to different transport modes.

4.1. Future modifications and adaptations of the access graph

Definition of GTT. The definition of GTT used in this work incorporates three components of travel time together with their relative weights, as reported in the literature on the perception of travel time. This formulation enables easy modifications of the components, as well as including other components of perceived travel time and modelling the time-equivalent disutility penalties for several known determinants of PT use and satisfaction. We list some options below:

- In our formulation, we use the constant waiting time as half of the headway. Underlying this choice is the assumption of random passenger arrivals for high-frequency services (which holds for most metro lines) but does not hold for low-frequency lines. To account for the non-randomness of passenger arrivals in the latter case, the waiting time component can be modelled by truncating at the maximum allowed waiting time value: $t^{wait} = \max\{30/f_{ij}, t^{wait}_{max}\}$, where t^{wait}_{max} is a parameter (for example, 10 min for metro networks) [53,54].
- In this work, we use a constant transfer penalty; this can be modelled more realistically by weighting each subsequent transfer more heavily.
- Other penalties can be added in a straightforward way: for example, a time-equivalent penalty for crowding, which can be obtained from smart card data; or reliability of service as extracted from historical operation data.
- The logsum of utility of several route alternatives for the same OD pair can be incorporated into the model (e.g., [55].

Other modes. In this work, we only consider metro networks. The construction of any other PT modality will proceed in the same manner. In addition, access graphs for other modes can be constructed in a similar manner; for example, road networks with junctions as nodes and street segments between them as edges; then by calculating the expected travel times based on length and average speed, the access graph construction proceeds in the same way as described here. This approach allows us to incorporate mode-specific influences on travel times, such as expected congestion, time needed to find a parking space, etc., in the case of road networks. This is modelled in the very first step; after that, the analysis proceeds in the same way. This also holds for active modes, such as walking or cycling; shared mobility, etc.

Multimodal networks. Related to the previous point, an extension to multi-modal systems is a natural way forward. In the present analysis of access, the metro networks are treated as isolated systems. However, travel is often multi-modal, where a passenger uses more than one mode in a single journey. The analysis of the overall PT accessibility of an urban area will be an interesting extension of the access graph methodology. In this case, the multi-modal system may be represented as a multilayer network [56], with each mode representing a separate aspect. Relative contributions of each mode to overall accessibility, as measured, for example, by node participation, will offer a framework for integrated assessment of access.

Generalised cost. In the present work, we only considered generalised travel time as an example of the more general term generalised travel cost. The access graph methodology can be applied to any other cost definition, including monetary cost. While this is not the major factor of PT use in urban areas, it becomes an important factor in long distance travel. For example, for the European air or rail networks, we can ask: between which pairs of cities can one travel for at most 100 euros? 200? 500? etc. Then progressively increase the monetary threshold and build the access graph based on financial accessibility. Even in urban or regional PT systems, such an approach might guide planners in devising pricing strategies, determining tariff zones based on partitions of the access graph.

Temporal variability. Service levels often change significantly over time, either in the short term, such as for different times-of-the-day (rush hours versus night-time schedules), or for different days-of-the-week (weekday vs. weekend service), or long-term, such as changes to line frequencies or line plans. In this work, we used averaged values of frequencies for the entire operating time window. Constructing the access graph for different values of service frequency will enable a comparison of access levels in different time periods. Specifically, simulation of adding new lines or changes of frequency will enable planners to systematically assess changes to network accessibility.

Passenger variability in perceived travel costs. In this work, we used a fixed set of waiting time and transfer penalty parameters. Sensitivity analysis shows the variation of indicators with varying parameter values; however, the resulting access graph is ultimately constructed for a fixed pair of $(w^{wait}, w^{transfer})$. In reality, perceived costs vary across individual passengers, for example as derived from mixed logit discrete choice models [57]. This might be included in the framework by adopting a probabilistic approach, e.g., by modelling a joint distribution of parameters $f(w^{wait}, w^{transfer})$. We see (at least) two options to approach this: (i) sampling from the distribution and calculating the expected values of GTT at each link, then proceeding as before, and (ii) calculating the intervals of GTT and use interval algebra methods to assess variability in the structure of the access graph stemming from the probabilistic modelling.

In conclusion, the access graph methodology provides a flexible framework that can incorporate empirical behavioural findings in the definition of perceived (generalised) travel time, offering a powerful tool for both researchers and planners to systematically evaluate the impacts of different sources of perceived travel costs on experienced access levels.

4.2. Limitations and future work

Alongside the aforementioned contributions of this work, our study is also subject to certain limitations, which we discuss in the following, and outline several related venues for future research to address those. The most obvious omission of the model is the exclusion of data on the number of opportunities for different activities available at accessible destinations. The access graph in this original form addresses only the transport component of accessibility, implicitly assuming the number of reachable stops as an approximation for opportunities for activity. Including land use data in the model and developing suitable indicators for this extended representation constitutes an important further research direction. In addition, our analysis of over- and undersupply (Section 3.4) at the OD-level should be enriched by an investigation at the service link-level. This can be performed by examining inter-station passenger volume over capacity ratios which will offer a more operational comparison of supply and demand. Furthermore, this will allow establishing capacity utilisation profiles at the line-level. The treatment of lines as primary objects of interest within the access graph approach appears as a promising future direction, and will open new approaches to the line planning and frequency setting problems.

Next, modelling the full journeys, including access from origin to the station and from the station to the destination, together with the distribution of origin and destination points, will provide measures of access in a more comprehensive sense. We see modelling the PT network as a network of networks, with each PT station positioned in a network of walkable paths, as an interesting direction of future research. Related to the previous point, the access equity analysis considers only topological distribution of reachable opportunities which is a limited definition of equity. Including socio-economic perspectives will offer a more complete picture of access inequalities. Overall, a statistical analysis of the introduced indicators as predictors of demand, either in terms of ridership or modal share, would provide an external validation of the proposed measures of access.

In the present work, we focus on global indicators of access. An interesting extension will be a node-level analysis, where beside node degree, as a measure of isochrone-based access, the structure of its neighbourhood, such as observing the clustering coefficient, will offer insights into access beyond traditional cumulative opportunity measures at the local level.

Furthermore, the geographical scale of the networks can jeopardise the comparison of networks of such varying sizes as in our dataset, especially for absolute-value or fixed-time indicators, such as t_{max}, t_M, or $\langle k \rangle_{30}$. Notwithstanding our effort to address this issue to a certain extent by introducing the normalised versions of average degree and the times relative to the maximum time, this leaves open the normative question of the maximum generalised travel time as an accessibility indicator in itself (i.e., does the observed t_{max} make sense for the geographical scope of this network?). This has been examined in detail in a follow-up study [58], where we compare the access behaviour of each metro network to its idealised counterpart based on maximally connected subgraphs of the access graph.

Metro networks are relatively small (the number of nodes < 500), whereas bus networks in these cities consist of thousands of nodes and longer total travel times, which may become prohibitive from both the time and space computational complexity. Moreover, with the construction of the functional access connections, we lose the information on infrastructure and service. Consequently, any changes in either of those requires the full construction of the access graph from scratch. We thus emphasise that it is an addition to the family of PTN graph representations, and is not intended as a replacement. Each serves its purpose; e.g. L-space for resilience analysis, P-space for transfer patterns, A-space for access.

While we presented the sensitivity analysis for one of the networks with respect to travel time parameters, we note that in each instant, the travel time is considered constant. In real-world PTNs, actual travel times are subject to inherent variability due to changing conditions that may cause disruptions to the scheduled service. A more realistic view of access levels will be gained by incorporating these variabilities, for example, by using a probabilistic approach to model the randomness of travel times, the accumulated impact thereof and observing the changes to the access graph. Specifically, a systematic simulation of delays at different nodes and links will help planners identify those that have the largest impact on the overall network accessibility, thereby helping decision makers prioritise parts of the network for employment of control strategies.

Finally, the access graph is modelled as an unweighted undirected graph. This is the result of taking the simplest decay function of distance, i.e. the cut-off function. In reality, individual time budgets and weights for different components of travel time may vary across passengers. Taking a probabilistic approach to edge inclusion or using fuzzy cut-off values will be an interesting next step.

We see several promising generalisations in the direction of multilayer networks [56]. The structure of the access graph itself hints at multilayer networks; we believe there is significant potential in bridging the theory of multilayer networks and accessibility to address the problem in its full complexity. First, the formalisation of the access graph as a temporal network will enable the ready application of existing methods and deeper insights, such as temporal motif detection [59]. Second, the access graph can be used to model different types of interactions, such as financial costs, perceived accessibility, etc. Including layers with different thresholds for different variables or aspects will be an interesting generalisation, such as modelling the X-minute, Y-cost city [60]. Heterogeneous networks where different entities (stops, routes, passengers, etc.) can be modelled to interact with each other can guide the development of accessibility studies.

In a wider sense, the access graph can be seen as introducing a method to formulate graph representations with functional connections, which was recently reviewed for robustness-related studies in [61]. Modelling the PT system properties at the level of network structure, not at the level of its properties, itself emerges as a promising direction for PTN studies.

We note that the relationships between the introduced access indicators and the size of the network, as reflected in its topological dimension N and network diameter t_{max}, as observed in Fig. 10, exhibit power-law behaviour in most cases. Although the fits do not exhibit strong statistical significance, they point to (weak) evidence of allometric scaling [62]. The analysis of access behaviour with

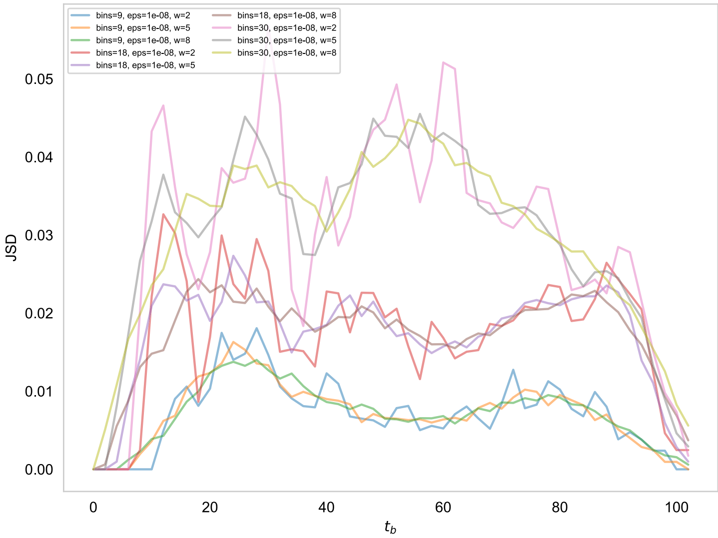

Fig. A.14. JSD sensitivity analysis for the Washington network; sliding time window and binning are varied.

a focus on maximal achievable levels of access within the framework of allometric scaling offers an interesting avenue for future research. Beyond (public) transport networks, we note the potential of the Access graph methodology to inspire notions similar to accessibility and network science approaches in other application domains, complementing related approaches such as that in [63].

The new representation can also guide planners and policymakers in designing accessible public transport networks. First, by transferring the access formalism into the network structure, the indicators it offers are highly interpretable and reproducible across all modes of PT systems. The L- and P-space representations, on which the access graph construction is based, use the GTFS data, making the construction of the access graph readily available. Detailed analysis and visualisation can help identify regions with lower access levels, and simulations can provide strategies for equitable planning in the strategic or tactical stage by studying the impact of changes to the network.

CRediT authorship contribution statement

Tina Šfiligoj: Writing – original draft, Visualization, Software, Methodology, Investigation, Formal analysis, Conceptualization. **Aljoša Peperko:** Writing – review & editing, Supervision, Methodology, Conceptualization. **Oded Cats:** Writing – review & editing, Supervision, Methodology, Conceptualization.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

The first and second authors were partially supported by ARIS-Slovenian Agency for Research and Innovation (grants P2-0394 (A) and P1-0222, respectively). They also acknowledge the financial support from the Slovenian Research and Innovation Agency (ARIS) through the UL VIP project (KTTK21) under contract no. SN-ZRD/22-27/510. The authors gratefully acknowledge WMATA for the data provision.

Appendix. Sensitivity analysis for JSD

Here, we provide sensitivity analysis for the JSD analysis for the Washington and Vienna networks. For our results, we used a fixed binning with $n_{bins} = \text{round}(N/5)$ to construct the degree distribution and a sliding window of width 5 to smooth the curve $JSD(t_b)$. Here, we systematically vary the binning and the width of the sliding window. The results suggest robustness with respect to both parameters for both cities, as the shapes and the location(s) of the peaks stay similar across all combinations of parameters (Fig. A.14 for Washington and Fig. A.15 for Vienna). When the binning is too coarse, the shapes show less variance, but a binning of only 10 bins distorts the distributions. As the number of bins rises and the point-wise variations become larger, the shapes remain similar in all cases.

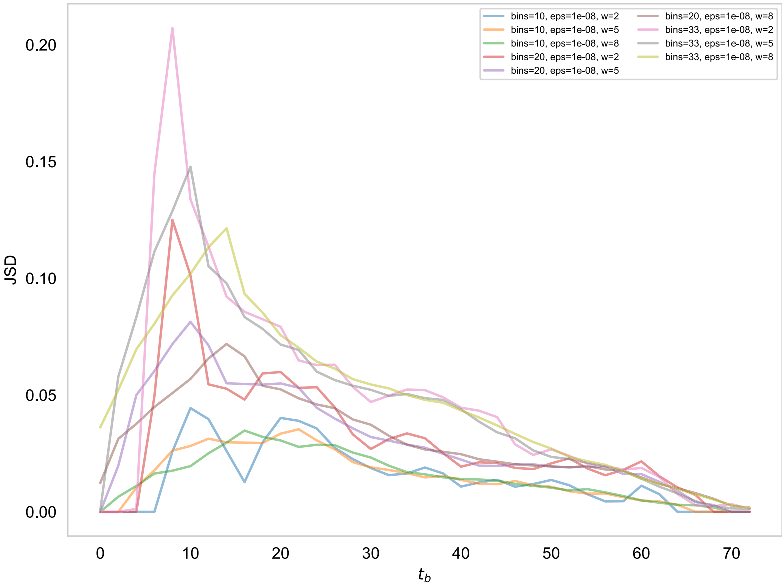

Fig. A.15. JSD sensitivity analysis for the Vienna network; sliding time window and binning are varied.

Data availability

The supply data used is publicly available and referenced in the manuscript. The authors do not have permission to share the demand data.

References

[1] W.G. Hansen, How accessibility shapes land use, J. Am. Inst. Planners 25 (2) (1959) 73–76.
[2] K.T. Geurs, B. Van Wee, Accessibility evaluation of land-use and transport strategies: Review and research directions, J. Transp. Geogr. 12 (2) (2004) 127–140.
[3] M. Cui, D. Levinson, Primal and dual access, Geogr. Anal. 52 (3) (2020) 452–474.
[4] B. Lahoorpoor, H. Rayaprolu, H. Wu, D.M. Levinson, Access-oriented design? disentangling the effect of land use and transport network on accessibility, Transp. Res. Interdiscip. Perspect. 13 (2022) 100536.
[5] D. Levinson, H. Wu, Towards a general theory of access, J. Transp. Land Use 13 (1) (2020) 129–158.
[6] H. Wu, P. Avner, G. Boisjoly, C.K. Braga, A. El-Geneidy, J. Huang, T. Kerzhner, B. Murphy, M.A. Niedzielski, R.H. Pereira, et al., Urban access across the globe: An international comparison of different transport modes, Npj Urban Sustain. 1 (1) (2021) 16.
[7] A. Soukhov, R.H. Pereira, C.D. Higgins, A. Páez, A family of accessibility measures derived from spatial interaction principles, PLoS One 20 (11) (2025) e0335951.
[8] B. Van Wee, Accessible accessibility research challenges, J. Transp. Geogr. 51 (2016) 9–16.
[9] B. Van Wee, Accessibility and equity: A conceptual framework and research agenda, J. Transp. Geogr. 104 (2022) 103421.
[10] G. Boisjoly, A.M. El-Geneidy, The insider: A planners' perspective on accessibility, J. Transp. Geogr. 64 (2017) 33–43.
[11] S. Handy, Is accessibility an idea whose time has finally come? Transp. Res. Part D: Transp. Environ. 83 (2020) 102319.
[12] C. Von Ferber, T. Holovatch, Y. Holovatch, V. Palchykov, Public transport networks: Empirical analysis and modeling, Eur. Phys. J. B 68 (2009) 261–275.
[13] G. Dupuy, Network geometry and the urban railway system: The potential benefits to geographers of harnessing inputs from "naive" outsiders, J. Transp. Geogr. 33 (2013) 85–94.
[14] S. Derrible, C. Kennedy, Applications of graph theory and network science to transit network design, Transp. Rev. 31 (4) (2011) 495–519.
[15] R. Ding, N. Ujang, H.B. Hamid, M.S.A. Manan, R. Li, S.S.M. Albadareen, A. Nochian, J. Wu, Application of complex networks theory in urban traffic network researches, Netw. Spat. Econ. 19 (2019) 1281–1317.
[16] K. Adjetey-Bahun, B. Birregah, E. Châtelet, J.-L. Planchet, A model to quantify the resilience of mass railway transportation systems, Reliab. Eng. Syst. Saf. 153 (2016) 1–14.
[17] O. Cats, The robustness value of public transport development plans, J. Transp. Geogr. 51 (2016) 236–246.
[18] O. Cats, G.-J. Koppenol, M. Warnier, Robustness assessment of link capacity reduction for complex networks: Application for public transport systems, Reliab. Eng. Syst. Saf. 167 (2017) 544–553.
[19] O. Cats, P. Krishnakumari, Metropolitan rail network robustness, Phys. A 549 (2020) 124317.
[20] O. Cats, Topological evolution of a metropolitan rail transport network: The case of stockholm, J. Transp. Geogr. 62 (2017) 172–183.
[21] T. Šfiligoj, A. Peperko, P. Bajec, O. Cats, Node importance corresponds to passenger demand in public transport networks, Phys. A (2025) 130354.
[22] S.D. Dimitrov, A.A. Ceder, A method of examining the structure and topological properties of public-transport networks, Phys. A 451 (2016) 373–387.
[23] R. de Regt, C. von Ferber, Y. Holovatch, M. Lebovka, Public transportation in great britain viewed as a complex network, Transp. A: Transp. Sci. 15 (2) (2019) 722–748.
[24] D. Luo, O. Cats, H. van Lint, G. Currie, Integrating network science and public transport accessibility analysis for comparative assessment, J. Transp. Geogr. 80 (2019) 102505.

[25] R. Kujala, C. Weckström, M.N. Mladenović, J. Saramäki, Travel times and transfers in public transport: Comprehensive accessibility analysis based on pareto-optimal journeys, Comput. Environ. Urban Syst. 67 (2018) 41–54.
[26] GoogleDevelopers, General transit feed specification (GTFS), 2025, (Accessed 01 February 2025).
[27] D. Luo, O. Cats, H. van Lint, Can passenger flow distribution be estimated solely based on network properties in public transport systems? Transportation 47 (6) (2020) 2757–2776.
[28] M. Barthélemy, Spatial networks, Phys. Rep. 499 (1–3) (2011) 1–101.
[29] E. Hussain, A. Bhaskar, E. Chung, Transit od matrix estimation using smartcard data: Recent developments and future research challenges, Transp. Res. Part C: Emerg. Technol. 125 (2021) 103044.
[30] O. Cats, Identifying human mobility patterns using smart card data, Transp. Rev. 44 (1) (2024) 213–243.
[31] T. Neutens, Accessibility, equity and health care: Review and research directions for transport geographers, J. Transp. Geogr. 43 (2015) 14–27.
[32] A. El-Geneidy, D. Levinson, E. Diab, G. Boisjoly, D. Verbich, C. Loong, The cost of equity: Assessing transit accessibility and social disparity using total travel cost, Transp. Res. Part A: Policy Pr. 91 (2016) 302–316.
[33] A. Karner, Assessing public transit service equity using route-level accessibility measures and public data, J. Transp. Geogr. 67 (2018) 24–32.
[34] T.A. Bittencourt, M. Giannotti, Evaluating the accessibility and availability of public services to reduce inequalities in everyday mobility, Transp. Res. Part A: Policy Pr. 177 (2023) 103833.
[35] H. Tenkanen M. Pönkänen, M. Mladenović, Spatial accessibility and transport inequity in finland: Open source models and perspectives from planning practice, Comput. Environ. Urban Syst. 116 (2025) 102218.
[36] S. Gori, L. Mannini, M. Petrelli, Equity measures for the identification of public transport needs, Case Stud. Transp. Policy 8 (3) (2020) 745–757.
[37] J. Sienkiewicz, J.A. Hołyst, Statistical analysis of 22 public transport networks in poland, Phys. Rev. E—Statistical Nonlinear Soft Matter Phys. 72 (4) (2005) 046127.
[38] M. Yap, O. Cats, B. Van Arem, Crowding valuation in urban tram and bus transportation based on smart card data, Transp. A: Transp. Sci. 16 (1) (2020) 23–42.
[39] M. Yap, H. Wong, O. Cats, Passenger valuation of interchanges in urban public transport, J. Public Transp. 26 (2024) 100089.
[40] M.J. Williams, M. Musolesi, Spatio-temporal networks: Reachability, centrality and robustness, R. Soc. Open Sci. 3 (6) (2016) 160196.
[41] A.M El-Geneidy, D.M. Levinson, Access to destinations: Development of accessibility measures, 2006.
[42] B. Kapatsila, M.S. Palacios, E. Grisé, A. El-Geneidy, Resolving the accessibility dilemma: Comparing cumulative and gravity-based measures of accessibility in eight canadian cities, J. Transp. Geogr. 107 (2023) 103530.
[43] P. Holme, J. Saramäki, Temporal networks, Phys. Rep. 519 (3) (2012) 97–125.
[44] A. Ahmed, P. Stopher, Seventy minutes plus or minus 10—A review of travel time budget studies, Transp. Rev. 34 (5) (2014) 607–625.
[45] P.R. Stopher, A. Ahmed, W. Liu, Travel time budgets: New evidence from multi-year, multi-day data, Transportation 44 (5) (2017) 1069–1082.
[46] W. Wang, T. Osaragi, Lognormal distribution of daily travel time and a utility model for its emergence, Transp. Res. Part A: Policy Pr. 183 (2024) 104058.
[47] L.A. Guzman, D. Oviedo, C. Rivera, Assessing equity in transport accessibility to work and study: The bogotá region, J. Transp. Geogr. 58 (2017) 236–246.
[48] F. Di Ciommo, Y. Shiftan, Transport equity analysis, 2017.
[49] M. De Domenico, V. Nicosia, A. Arenas, V. Latora, Structural reducibility of multilayer networks, Nat. Commun. 6 (1) (2015) 6864.
[50] S. Vijlbrief, O. Cats, P. Krishnakumari, S. van Cranenburgh, R. Massobrio, A curated data set of l-space representations for 51 metro networks worldwide, 2022.
[51] S. Vijlbrief, O. Cats, P. Krishnakumari, S. van Cranenburgh, R. Massobrio, A curated data set of p-space representations for 51 metro networks worldwide, 2022.
[52] A. Hagberg, P.J. Swart, D.A. Schult, Exploring Network Structure, Dynamics, and Function Using Networkx, Tech. rep., Los Alamos National Laboratory (LANL), Los Alamos, NM (United States), 2008.
[53] L.A. Bowman, M.A. Turnquist, Service frequency, schedule reliability and passenger wait times at transit stops, Transp. Res. Part A: Gen. 15 (6) (1981) 465–471.
[54] J.B. Ingvardson, O.A. Nielsen, S. Raveau, B.F. Nielsen, Passenger arrival and waiting time distributions dependent on train service frequency and station characteristics: A smart card data analysis, Transp. Res. Part C: Emerg. Technol. 90 (2018) 292–306.
[55] N. Nassir, M. Hickman, A. Malekzadeh, E. Irannezhad, A utility-based travel impedance measure for public transit network accessibility, Transp. Res. Part A: Policy Pr. 88 (2016) 26–39.
[56] M. Kivelä, A. Arenas, M. Barthelemy, J.P. Gleeson, Y. Moreno, M.A. Porter, Multilayer networks, J. Complex Netw. 2 (3) (2014) 203–271.
[57] M. Yap, O. Cats, Taking the path less travelled: Valuation of denied boarding in crowded public transport systems, Transp. Res. Part A: Policy Pr. 147 (2021) 1–13.
[58] T. Šfiligoj, A. Peperko, O. Cats, Network accessibility as the emergence of cliques, Sci. Rep. (2026).
[59] A. Paranjape, A.R. Benson, J. Leskovec, Motifs in temporal networks, in: Proceedings of the Tenth ACM International Conference on Web Search and Data Mining, 2017, pp. 601–610.
[60] Y. Chen, S. Jia, Q. Xu, Evaluating the accessibility and equity of key amenities in the x-minute and y-monetary-cost city: A double-threshold cumulative opportunity measure, J. Public Transp. 27 (2025) 100128.
[61] C.M. Büth, M. Zanin, Functional connectivity networks for transportation delay analysis: From theory to software, 2025, arXiv preprint arXiv:2510.05143.
[62] H. Samaniego, M.E. Moses, Cities as organisms: Allometric scaling of urban road networks, J. Transp. Land Use 1 (1) (2008) 21–39.
[63] H.H. Lentz, T. Selhorst, I.M. Sokolov, Unfolding accessibility provides a macroscopic approach to temporal networks, Phys. Rev. Lett. 110 (11) (2013) 118701.