Assessing the evolution of educational accessibility with self-avoiding random walk insights from Helsinki

Aydin, Nazli Yonca; Yigitbasi, Emre; Casali, Ylenia; van Wee, Bert

DOI
10.1007/s41109-023-00581-4

Publication date
2023

Document Version
Final published version

Published in
Applied Network Science

Applied Network Science

RESEARCH

Open Access

Assessing the evolution of educational accessibility with self-avoiding random walk: insights from Helsinki

Nazli Yonca Aydin[1*], Emre Yigitbasi[2], Ylenia Casali[3] and Bert van Wee[3]

*Correspondence:
N.Y.Aydin@TuDelft.nl

[1] Multi Actor Systems, Faculty of Technology, Policy and Management, Delft University of Technology, Jaffalaan 5, 2628 BX Delft, The Netherlands
[2] Faculty of Technology, Policy and Management, Delft University of Technology, Delft, The Netherlands
[3] Transport and Logistics Group, Faculty of Technology, Policy and Management, Delft University of Technology, Delft, The Netherlands

Abstract

Rapid urbanization has posed challenges to accessibility to critical services that require in-depth analysis. Complex networks theory has been used to evaluate the evolution of network topologies or the overall accessibility of transportation systems. However, topological metrics to explain the temporal changes in accessibility levels do not fully capture the dynamics and implications of accessibility to specific critical services. In this study, we address this gap and investigate the opportunities of using a self-avoiding random walk (SARW) algorithm to evaluate and explain the evolution of spatial accessibility to education facilities. We used hotspot analysis to understand the temporal changes and investigated changes in hot and cold spots over time. Furthermore, we explored the relationship between the network indicators and the SARW-based accessibility metric. We illustrated this method in a case study from Helsinki, where large-scale open data spanning from 1991 to 2016 is available. Our findings indicate that the SARW-based metric delivers more detailed node-level results than the traditional isochrone-based metric. The latter generates accessibility zones where accessibility is assumed to be uniform, while the SARW metric captures the dynamic nature of educational facility accessibility more accurately. The developed methodology helps to identify the impacts on the historical development of accessibility and can be applied to investigate accessibility to other critical services.

Keywords: Accessibility, Transportation, Complex networks, Geographic information systems, Temporal analysis

Introduction

Ensuring access to critical services such as healthcare, employment, or education during the rapid urbanization era has been a growing concern in cities (Vecchio et al. 2020). Accessibility has frequently been used to measure equality, equity, and fairness in transport research. It is defined as the degree to which transport and land use system enables individuals to reach an activity or destination by providing at least one transport mode (Geurs and van Wee 2004, p. 128). "Good accessibility" enables participation in services and social interaction (Li et al. 2021).

Accessibility to activities and services is an important indicator of development policies regarding transport and land use systems (Geurs and van Eck 2003). Among these activities and services are economic opportunities (e.g., job locations), and critical services (e.g., education, healthcare, and other fundamental facilities) (Curl 2018). The lack of access to these opportunities can remarkably affect the quality of life. For example, accessibility to educational opportunities is an important decision factor for continuing higher education (Dickerson and McIntosh 2013). Thus, it impacts economic participation, as an individual's education level influences their economic stature (Paolo et al. 2017). Despite its significance, there are fewer studies that evaluate the accessibility of education in comparison to the accessibility of employment or other essential services (Sharma and Patil 2021), a recent example being (An evaluation of primary schools and its accessibility using GIS techniques 2023). Acknowledging the need for accurate measurement of accessibility, extensive literature on measurement approaches exists (Geurs and Wee 2004). However, each type of measure brings a different perspective on accessibility, making them useful in certain use cases.

Complex network theory focuses on understanding how complex systems function by using mathematical and computational methods to investigate their organization and behavior (see "Complex networks approaches" section). This approach allows for the mapping and spatial analysis of urban systems, providing various possibilities for understanding their functioning. Some researchers focus on topological growth to understand how transportation networks evolve through observing the changes in connectivity indices, such as alpha, beta, gamma, and centrality metrics (Casali and Heinimann 2019; Cats 2017; Strano et al. 2012), while others focus on the efficiency of transport networks (Brussel et al. 2019). Specifically, transport systems benefit from the graph theory-driven approach for quantitatively analyzing road network characteristics (Barthélemy 2011), robustness (Casali and Heinimann 2020; He et al. 2021), and identifying communities and groups (Viljoen and Joubert 2019).

Complex network theory is thoroughly examined for its application in urban land-use and transportation research and practice by Ding (2019). Despite its potential, limited studies apply it to identify and measure the evolution of accessibility in transport and land use systems (Ding 2019). Barthélemy and Flammini (2009) investigated the relationship between population density and how transportation networks develop over time. However, their research didn't examine the spatial patterns of evolution such as the accessibility to certain land uses. We argue that studying the evolution of accessibility to a particular critical service using complex network theory can provide insight into future changes and contribute greatly to the evidence-based planning of cities.

To address these shortcomings, we examine the evolution of accessibility to education facilities and explore the use of complex networks theory to understand: "*to what extent can complex network analysis be used to evaluate and explain the evolution of spatial accessibility to education?*" Our contributions are twofold: (1) we use the Self-Avoiding Random Walk (SARW) as an accessibility metric to evaluate the accessibility to education facilities. The results are compared to the most commonly used accessibility measure of isochrone-based accessibility. (2) We evaluate the evolution of accessibility education facilities and investigate the spatiotemporal changes using hotspot analysis with Getis-Ord Gi* statistic. Furthermore, we discuss the relationships between the

self-avoiding random walk-based school accessibility metric and network topology indicators. Our approach is illustrated in a case study from Helsinki.

Background

School accessibility factors

Accessibility measurements are used as quantitative indicators of spatial availability of socioeconomic opportunities, which are categorized by Geurs and van Wee (2004) under four main perspectives: infrastructure, location, utilities, and people. Isochrone-based accessibility is a location-based measure that assesses accessibility to services by measuring the travel time or distance to a location, but it doesn't take into account the combined effect of land use and transport systems, which can limit its ability to fully capture accessibility.

Studies focusing on spatial accessibility often evaluate the proximity to education opportunities. Dickerson and McIntosh (2013) showed that distance plays a role in determining whether young people who are on the fence about participating in post-compulsory education will continue their education or not. A similar study by Sá et al. (2006) conducted on Dutch high school students showed that proximity to professional education increases the probability of continuing their education. Zooming into the impact of proximity, Andersson et al. (2012) conducted a comparative analysis of the effect of distance on schools from 2000 to 2006. Their findings indicate that low-income groups are often constrained to schools closer to home due to not owning personal vehicles. Therefore, affordability indirectly impacts the access range of individuals, and an overall increase in the distances to schools impairs the accessibility for disadvantaged groups. Mei et al. (2019) studied the accessibility to schools in the Shenyang area of China, and found that schools were clustered in the city center, resulting in longer travel distances and times for those in peripheral areas.

Xu et al. (2018) conduct a historical analysis of the socio-spatial accessibility to urban education in a case study in Nanjing. Their method involves three distinct accessibility indices: geographic accessibility, opportunity availability, and economic affordability. These indices reflect three main factors they identify concerning education accessibility: the proximity to schools, the supply of schools compared to housing, and the affordability of access to school districts. Similarly, Bertolini (2012) explains how transportation and land use interact in a feedback cycle and how exogenous factors such as innovations, policy, and land availability can affect accessibility to job and education opportunities. The study identifies proximity, availability of opportunities, and affordability as the key factors in accessibility. Also, it suggests that urban form plays a role in the availability and spatial distribution of activities.

Accessibility to both education and employment is crucial for sustainable development, but education accessibility has been given less attention (Sharma and Patil 2021). Although the two are interrelated and have similar enabling factors, education has a direct impact on the likelihood of employment (Paolo et al. 2017). There is a wide range of literature that defines accessibility factors (Geurs and Wee 2004; Li et al. 2021; Wee and Mouter 2021), and these factors are essential for evaluating accessibility in the context of education. Some critical services (e.g., job centers (Hu and Downs 2019), healthcare (Aydin et al. 2019; Cheng et al. 2020)) and their

accessibility had been studied more than others. Especially, novel methodologies developed to measure the accessibility to educational services have fewer examples than the other services. Here, we aim to fill this gap by developing a novel modeling and analysis of the accessibility to educational opportunities.

Complex networks approaches

Network theory is an analytical approach to get a deeper understanding of complex systems that are difficult to envision by solely evaluating the behaviors of its individual components. It relies on a network system such as transportation networks where individual elements are represented as nodes (e.g., road intersections, junctions), and their connections or interactions are represented as edges (e.g., road segment of a transportation network) in a comprehensive graph (Mata 2020). In our case, a graph is a road network and the connectivity of it is defined based on the existence of edges between every pair of nodes. On the node level, centrality metrics (e.g., betweenness centrality, closeness centrality) are used to identify the node's role in connecting other node pairs (Barthélemy 2011). Community indicators are used to identify clusters within the network, and topological indicators investigate the structure of the network based on the size and density of network components (Casali and Heinimann 2019; Cats 2017). Lastly, accessibility indicators measure how network topology affects human movement and the reachability of nodes (Lee and Kim 2021).

Transport networks are analyzed through complex network metrics to understand road network characteristics, dynamic processes, communities within systems, and resilience properties (Ding 2019). For example, Aydin et al. (2018) used topological metrics to measure the resilience of transportation networks after the major Gorkha earthquake in 2015 in Kathmandu, Nepal. Aydin et al. (2019) used a modified version of betweenness centrality to identify the critical locations when traveling to a healthcare service using a modified betweenness centrality. Wang et al. (2020) examined the relationship between road network structure and ride-sharing accessibility by analyzing degree, closeness, and betweenness centrality. Their results showed that high degree and low closeness centrality in the road network is correlated with improved ride-sharing accessibility, while betweenness centrality has no significant impact. Wen et al. (2021) used network size and commonly used complex network metrics of average path length, and average clustering coefficient to identify the relationship between land value and topological properties of urban rail transit. The authors reported that land price is positively correlated with the number of nodes and the average clustering coefficient and negatively correlated with the average path length. Topological indicators are useful to analyze existing network structures or the impacts of interventions to provide evidence for future planning. However, most studies focus on the network structure and ignore the accessibility to critical functions, such as education, when using network theory approaches.

The random walk, a topological movement mechanism through networks, has been used in complex network analysis and is gaining attention in road transport network studies (Lee and Kim 2021). The dynamics of the random walk have been used in physics and the research of linear dynamics of diffusion (Travençolo and Costa 2008). The random walk can help to investigate the accessibility in transport networks, specifically

when large-scale mobility (i.e., human movements) data is unavailable. Lee and Kim (2021) use the random walk to model access diversity in a road network and propose an accessibility metric based on the geometric distance calculated by the summation of edge weights (lengths) in the network. Hanna (Hanna 2020) suggests that human movements can be modeled using the random walk to predict the movement of agents and network centrality measures. Their model assumes random movement without memory or direction, based on the angle of intersections.

The random walk method is particularly useful for including the network topology effects on accessibility. However, studies using random walks often lack the land use component and focus on the network structure to measure the diversity of reachable locations. Lee and Kim (2021) identify the shortcoming of their approach and propose the inclusion of origin–destination pairs based on real activity data. Although useful to assess accessibility based on network topology, this method could benefit from the inclusion of land use and activity components and has the potential to identify accessibility to educational opportunities.

Overall, network theory offers significant opportunities for measuring accessibility. Yet, it is noted that most studies using complex network theory and applications neglect the relationships between a land use function and transport systems but solely focus on the network characteristics. In this study, we explore the possibility of using complex network theory, specifically the random walk method, to evaluate the accessibility to educational facilities.

Methodology

Case study and data

The City of Helsinki is located in the Greater Helsinki metropolitan area and is Finland's largest city and capital. Since the 1970s, the city has rapidly transformed by contracting suburban areas (Nevanlinna 2016). As this rapid urbanization and transformation of the built environment have implications for the accessibility to critical services in Helsinki, we use Helsinki as a case study to investigate the evolution of accessibility to educational facilities. The study area in Helsinki includes 142 subdistricts within a 10 km radius of the city center. Helsinki Region Infoshare (HRI) platform provides large-scale and open socio-spatial data, including the road network and built environment (City of Helsinki 2021). From this database, we collected spatial data (i.e., road infrastructure) in *shapefile* format for 2016 and used ArcGIS to digitize guide maps for 1991, 1999, and 2007 to study changes in accessibility over time. To guarantee that the road data were consistent with the historical geometry of roads, we followed the data preprocessing steps described in Sect. 2.3 by. To process road data as networks, we first exported the attribute table of the road shapefiles as a txt file and we imported it on Python (see Casali and Heinimann 2019). All the network modeling and analysis were performed by using Igraph package in Python.

Helsinki school register, *Koulurekisteri* was used to collect historical school location data (Koulurekisteri 2020). This database contains information about all levels of schools and buildings since 1550. The data was obtained through a REST API in *json* format. Based on the start and end year, the data was sorted and divided between the years 1991, 1999, 2007, and 2016 (i.e., total of 4 timesteps) using Python *pandas* library.

For example, if a school has a start year of 1985 and end date of 2005, it can be found on both 1991 and 1999 datasets. Using the geopy library, we determined the geographical coordinates of schools based on a given address (Geopy 2022) and stored them as point data in a *shapefile* format in ArcGIS (2022).

The school locations represented in *shapefile* format were not consistently aligned with the road network data. In other words, schools were not represented as intersections/junctions in the model but as standalone points. For these cases, we used "Near tool" in ArcGIS to locate the closest edge (road segment) to each school node. The two nodes (i.e., starting node and the end node) were identified from the selected edges, and then the school location is assigned to these two nodes using the NetworkX library (NetworkX 2022). Therefore, each school node was represented by two nodes of the road network. This enables better integration of the school locations into the network nodes for simulating random walks.

School accessibility

Self-avoiding random walk (SARW)

This algorithm simulates the movement of a group of individuals starting from a common origin within a city and tracking their random movements within a specified threshold (i.e., distance range). The visited locations are recorded and the individuals are constrained to not revisit the same locations. The diversity of reached locations within a given threshold is analyzed using the collected data at the end of the simulation. The algorithm of this complex network method is presented in pseudo-code form in Algorithm 1.

Algorithm 1 Self-avoiding random walk.

Algorithm 1: Self-avoiding Random Walk

Input: weighted network $G(N, E, w)$, distance threshold D

Output: list of visited nodes V

```
for s ∈ N do
    initialize V;
    initialize total distance of walk W
    while W ≤ D do
        Add s to V;
        Get list of neighbor nodes (L);
        Remove visited nodes V from L;
        Choose random neighbor node r from L;
        Calculate distance w between r and s add to W
            r is the new s;
```

Throughout a single walk, the path selection is completely randomized. This suggests the probability assigned to any road segment is equal to all other available road segments when the random walker makes a decision. Self-avoidance property is imposed on the random walker, meaning that in a single walk, an already visited node cannot be

revisited. This property eliminates the possibility of the random walker being stuck in a certain area or even going back and forth between two same nodes. At the end of every walk, the list of visited nodes is reset which enables visits to the same nodes in different walks. Three conditions could potentially stop a SARW. (1) The random walk finalizes when a predefined distance threshold (D) is reached. The total distance traveled is calculated by summing the weights of each edge traveled during the walk. (2) The walk ends when the walker reaches a node with a degree of 1 (i.e., the dead-end node). (3) The walk ends when all the neighbors of the last visited node have been previously visited, and no available options exist.

This model provides a very simplistic movement simulation within the network. No prior information is provided regarding the network. The model demonstrates locations that can be reached with this movement pattern. For a real-life case, it is unrealistic that an individual would take random trips through the network. However, this model depends on the topology around a specific location for measuring the effect of road connectivity in determining paths.

SARW-based accessibility metric

For the case of accessibility to schools, we propose a metric using the SARW algorithm that relies on the node-specific analysis of visited nodes and counts the number of times a school node is visited. The number of school visits is summed across walks and divided by the number of walks to obtain the metric *vpw* (Eq. 1). We refer to this metric as "*accessibility metric*" which describes the number of schools within reach of a starting node.

$$vpw = \frac{\sum_{w=1}^{N_w} |\{S \cap V_m\}|}{N_w} \tag{1}$$

where S is the set of school nodes, V_m is the set of visited nodes in walk w, N_w is the number of walks in the simulation. Parameter selection for the SARW-based accessibility metric is explained in "Parameter selection" section.

Comparison with isochrone-based accessibility

We examined the usefulness and limitations of using a SARW-based accessibility metric by analyzing accessibility to educational facilities using the isochrone-based accessibility using ArcGIS ("ARCGIS". Accessed 2022). This technique identifies a catchment area of a critical function based on a predefined threshold and Dijkstra shortest path algorithm (Ertugay et al. 2016). Dijkstra's algorithm, a well-known approach, solves the problem of finding the shortest path between two locations in a weighted graph. To do this, the algorithm maintains a collection of junctions, denoted as S, for which the final shortest path from the starting location, s, has been determined. In each iteration, the algorithm selects the junction with the smallest estimated shortest path from the set of junctions and adds it to S. It then updates the shortest-path estimates for the neighboring junctions that have not yet been included in S. This process continues until the destination junction is included in S, signifying that the shortest path from s to the destination location, d, has been found (ArcGIS Desktop Help 2023).

Isochrone method is one of the most commonly utilized and widely recognized approaches for measuring accessibility and offers a straightforward approach for understanding variations in accessibility to facilities (Cascetta et al. 2013). In this study, we examine accessibility to educational facilities by "walking" as a travel mode for both accessibility metrics. Furthermore, it is assumed that pedestrian pathways exist alongside the road network due to the lack of pedestrian network data.

Spatial and temporal evolution of accessibility: hotspot analysis

Our study utilized a SARW-based accessibility metric to evaluate accessibility levels in a single year. To understand the changes in accessibility over time, we employed the hotspot analysis method. This method is commonly used to identify clusters of high and low accessibility values, and its outputs allow for spatial comparisons across historical time steps. It calculates p-values and z-scores using the Getis-Ord Gi* statistic, which compares the attribute values of a location with its neighbors. A location with a high attribute value and surrounded by other high-value locations is considered a hotspot. This is determined by comparing the local sum of attribute values with the overall sum of values in the area of interest, and determining if the difference is statistically significant using a z-score-based p-value (Kalinic and Krisp 2018).

Here, we used the Hotspot Analysis tool developed by ESRI (2022). The *Optimized Hotspot Analysis* tool selects the most suitable analysis parameters based on a set of conditions. The parameter decisions made are related to how spatial relationships are defined, which often involves a fixed distance threshold. This value establishes the search radius around the location of interest and must contain at least one neighbor. To select the parameters, the Incremental Spatial Autocorrelation strategy is used to identify spatial clustering and underlying processes. This approach calculates the Global Moran's I statistic for increasing distances, evaluating clustering intensity using the z-score. Peaks in the z-score signify distances where clustering is most prominent. The Optimized Hot Spot Analysis tool utilizes Incremental Spatial Autocorrelation to determine the scale of analysis based on peak distances. If no peak distance is found, the tool computes the average distance yielding K neighbors for each feature, with K determined as 0.05 * N which is the number of features. K is then adjusted to be minimum 3 and maximum 30. In case there are features greater than 500, the incremental analysis is skipped and the average distance is set to 30 neighbors. The resulting scale of analysis is used for the subsequent Hot Spot Analysis (Getis-Ord Gi*) tool (ESRI 2023). Note that the methodological improvements to this optimization method is beyond the scope of this study.

Network topology indicators

Two major factors are relevant to investigate school accessibility using hotspot analysis which are network topology indicators and the school locations. By comparing these indicators, the impact of the transport network and land use can also be distinguished. We use a connectivity indicator called "gamma index," which is used to assess the connectivity level of networks and evaluated by $\left(\gamma = \frac{\#of\ edges}{3(Number\ of\ nodes-2)}\right)$ (Kansky 1963). The gamma index represents the ratio between the observed number of edges and the maximum number of edges. For a large number of nodes, it is proportional to the

average node degree (Casali and Heinimann 2019). We evaluate the gamma indices for each district and applied a hotspot analysis to node degrees for each year to identify if there is a relationship between the SARW-based accessibility measure and the node degree.

Results and discussion

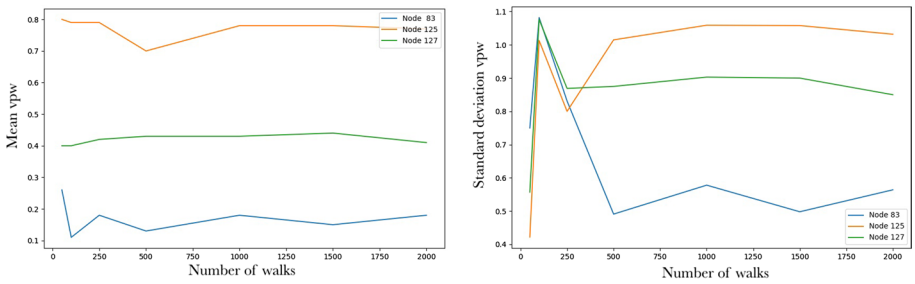
Parameter selection

There are two significant parameters to determine before moving forward with accessibility analysis. The first parameter is the threshold for the SARW and isochrone-based accessibility metrics. To determine the travel threshold, we considered the average pedestrian walking speed of 70 m per minute for Helsinki. This value includes impedances of traffic lights and crosswalks, and assumed a constant average speed (Tenkanen and Toivonen 2019). As for the travel times, we used Helsinki Travel Survey as a guide. Based on the data, 20 min per trip is the average amount of time an individual takes to travel to a school in Helsinki (Kaupunki 2016). This value and the walking speed are used to determine the distance threshold by simply multiplying speed and time. Thus, the distance threshold for assessing accessibility metric via walking is determined to be 1400 m. This threshold is used for measuring SARW-based accessibility and isochrone-based accessibility metrics.

The second parameter is the total number of walks, which will be used in the case of a SARW. Due to the random nature of the algorithm, having a low number of walks would result in unrealistic outputs. Considering the number of nodes and the average node degrees in the network, every starting node offers a variety of paths that can be taken across iterations. However, increasing the number of walks could mean observing more variety in distinct paths. It should be noted that a large number of walks would produce results with less precision and increase the computational time. Therefore, selecting the number of walks for an outcome that is more representative for the area is an important part of this study. Lee & Kim (Lee and Kim 2021) used 1000 walk iterations to model access diversity in a road network using random walk. Their approach relies on the sequential draws of road segments (edges) on each intersection (node) such that each edge is given the same probability of selection. At each step, the probability of the walk passing through an edge becomes inversely proportional to the node degree.

In this research, we applied sensitivity analysis to identify the optimal number of walks in a selected area from the sub-districts of the Helsinki urban area. The chosen sub-district, Etela-Haaga, has an area of 2.3 square kilometers and 187 nodes. The district included five school nodes in 2016. The sample run was formulated as follows:

(1) Starting nodes must be in the selected region
(2) Walks are not restricted to the selected region, free to continue on the edges outside
(3) The walk threshold is 1400 m based on pedestrian walk case
(4) The analysis is conducted using seven separate numbers of walks: 50, 100, 250, 500, 1000,1500, and 2000



(a) Mean *vpw* (b) Standard deviation *vpw*

Fig. 1 Sensitivity analysis for selecting the number of walks. Mean and standard deviation in *vpw* are displayed on three randomly selected nodes

Table 1 Number of nodes and edges in Helsinki over time

Year	# of nodes	# of edges/ road segments
1991	9574	12,962
1999	10,856	14,279
2007	11,222	14,722
2016	11,330	14,831

Finally, by comparing the mean and standard deviation of average school visits across walks (i.e., *vpw*), the sensitivity of the metric to the number of walks were analyzed. The results are given in Fig. 1.

The mean value of the *vpw* showed a relatively steady trend irrespective of the number of walks, especially for the total number of random walks, which is over 250. However, the standard deviation displayed a volatile nature and started to level out only after 500 walks. Therefore, 500 walks or more would be needed for the accessibility metric to converge on its variance. Considering that the computational load is directly proportional to the number of walks parameter, going beyond 500 walks was not preferred.

Network characteristics of Helsinki

Helsinki underwent several urban planning and transport system transformations in the twentieth century. The city expanded beyond its historical center in the 1950s (Nevanlinna 2016). The south harbor was closed in 1970, leading to a shift in the freight port to the city's east. These changes resulted in a polycentric urban form (Söderström et al. 2015) and increased car dependency and travel distances in the 1960s (METREX 2020). The construction of suburbs and road network investments, including highways, connected peripheral towns.

We observed this growth by inspecting changes in the network nodes and edges/ road segments over time (see Table 1). Car ownership also has steadily increased since the 1980s in Finland. However, the Helsinki region is consistently below the national average (Liljamo et al. 2021). City planning has played a role in limiting the number

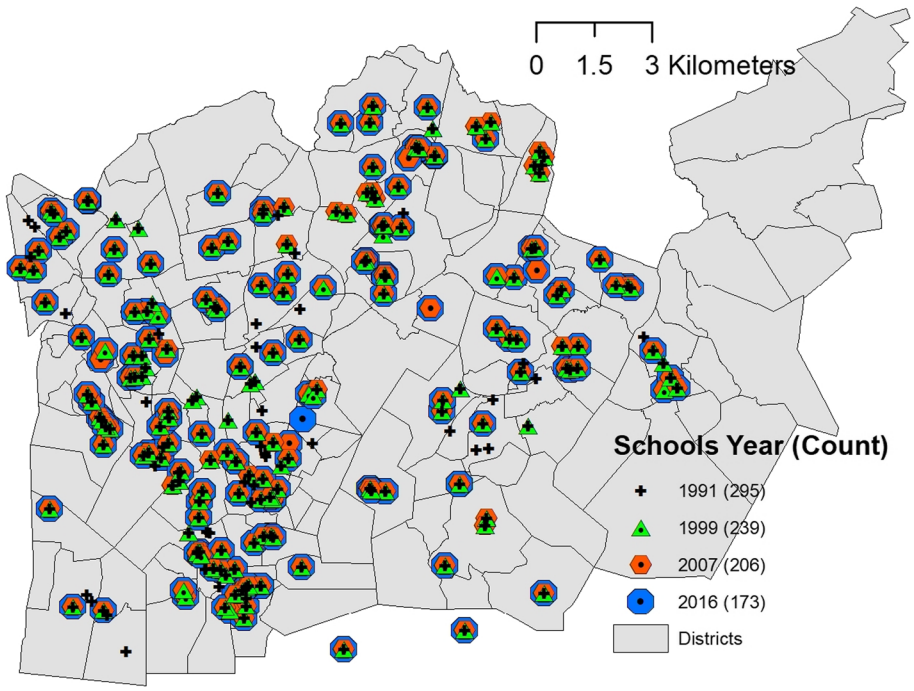


Fig. 2 Spatial distribution of schools over time

Table 2 Descriptive statistics of accessibility based on school visits per walk in the whole network

School visits per walk statistics/year	1991	1999	2007	2016
Non-school visiting node percentage	12.7%	16.5%	18.6%	24.3%
Mean	0.45	0.32	0.29	0.24
Standard deviation	0.55	0.46	0.43	0.42
Median	0.23	0.13	0.1	0.06

of cars by creating pedestrian zones in and around the city center starting in 1989 (City of Helsinki 2020). Therefore, Helsinki provides a good case study to evaluate the accessibility to schools as the empirical evidence shows that the road infrastructure changed over time. Figure 2 shows the school locations for 2016, 2007, 1999, and 1991. Interestingly, the number of schools has decreased over the years from 295 in 1991 to 173 in 2016.

School accessibility

Self-avoiding random walk

SARW-based accessibility analysis for schools is carried out for the selected years of 1991, 1999, 2007, and 2016. Table 2 shows that the percentage of nodes that have not reached a school after 500 random walks increased over time. Clearly, as the number of schools decreases, it is more likely that random walks reaching a school will decrease, which shows the impact of the reduced number of schools over the years. It should be noted that our analysis is limited to accessibility by walking. It is possible that the decrease in the number of schools observed over time in the city is accompanied by an

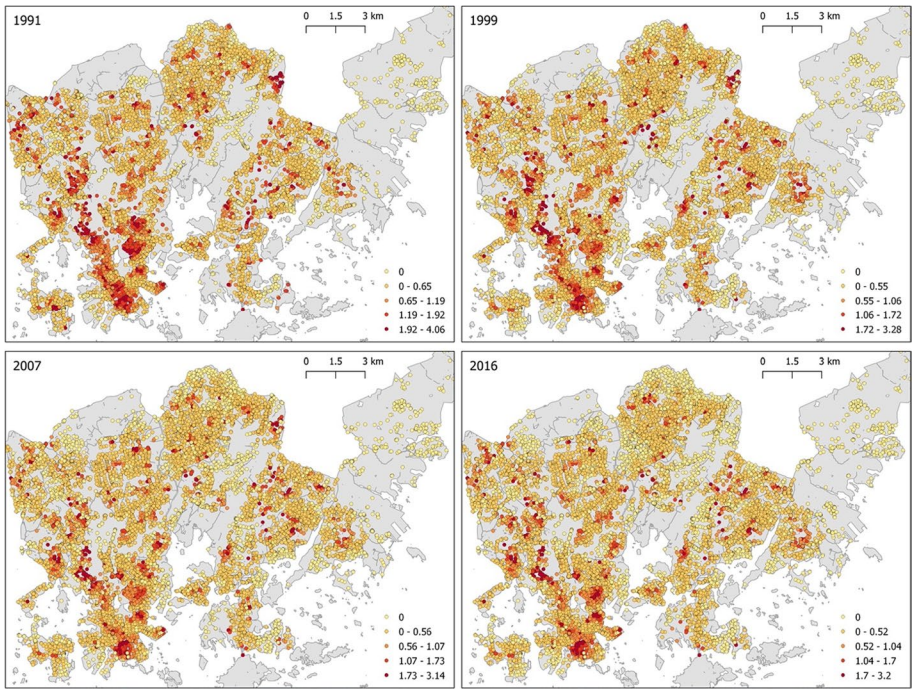


Fig. 3 Spatial distribution of the accessibility metric (*vpw*) for each node based on SARW simulation in Helsinki

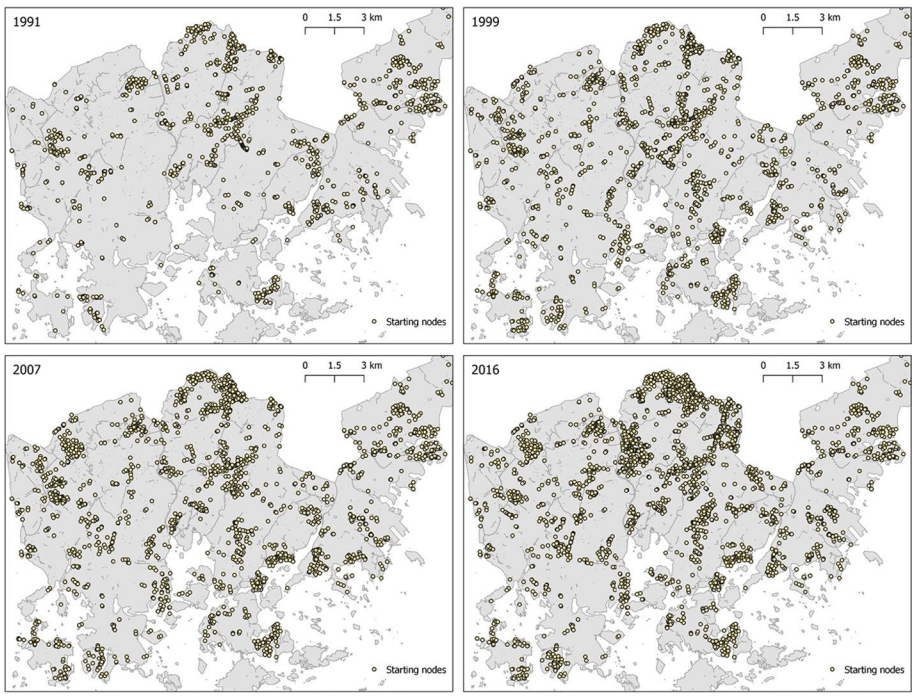


Fig. 4 Spatial distribution of nodes with zero school visits across 500 random walks

increase in the capacity of remaining schools. This could result in students needing to travel greater distances to reach their schools, potentially making alternate modes of transportation such as motorized vehicles or biking more favorable.

Figure 3 displays the spatial distributions of the accessibility metric (*vpw*) over time. Figure 4 shows the spatial distribution of low-access nodes for each year, specifically the nodes that never visit a school in any of the 500 walks. The nodes with low school visits accumulated in the northern region in 1991, but they get more dispersed over time (see Fig. 4). Throughout the years, the southwestern peninsula, corresponding to the center of Helsinki, has displayed higher accessibility to schools than the peripheries (see red points in Fig. 3, also see Fig. 4, where the zero school visits are less than the other districts). However, the center had lower school accessibility in 2016 compared to 1991, as evidenced by the increase in zero school visits over time, as shown in Fig. 4.

In this study, schools' capacities (i.e., the number of students or classrooms) are not included due to the lack of data. However, based on random walk approach, we can infer which schools are more easily accessible to pedestrians, and therefore may be more likely to attract higher demand (see Fig. 5). Figure 5 illustrates that most visits are accumulated in the historical center of the city in the southwestern peninsula of Helsinki. A higher frequency of school visits indicates greater attractiveness and the relative appeal of schools in this area. There is also a relatively large number of schools located.

We compared the results with the distribution of the population of school-aged children from ages 7–19 (see Fig. 6), representing the school-aged children from ages 7–19. The results show that districts with the most students are consistently located outside the historical center over the years. Yet, these areas have relatively lower accessibility value than the center of the city. This might be because the peripheral regions of the city

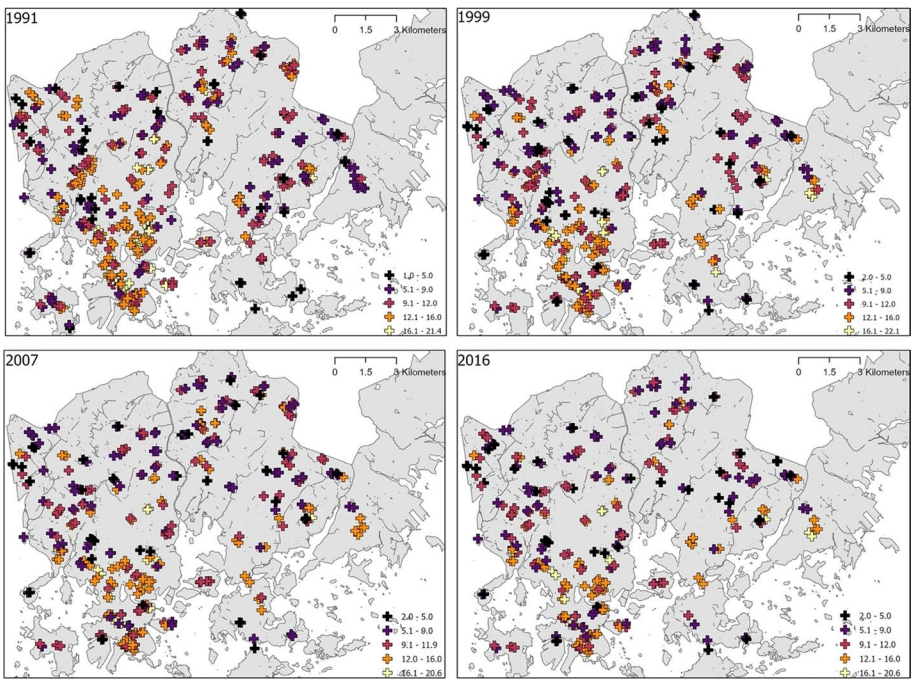


Fig. 5 Average number of visits per school per walk

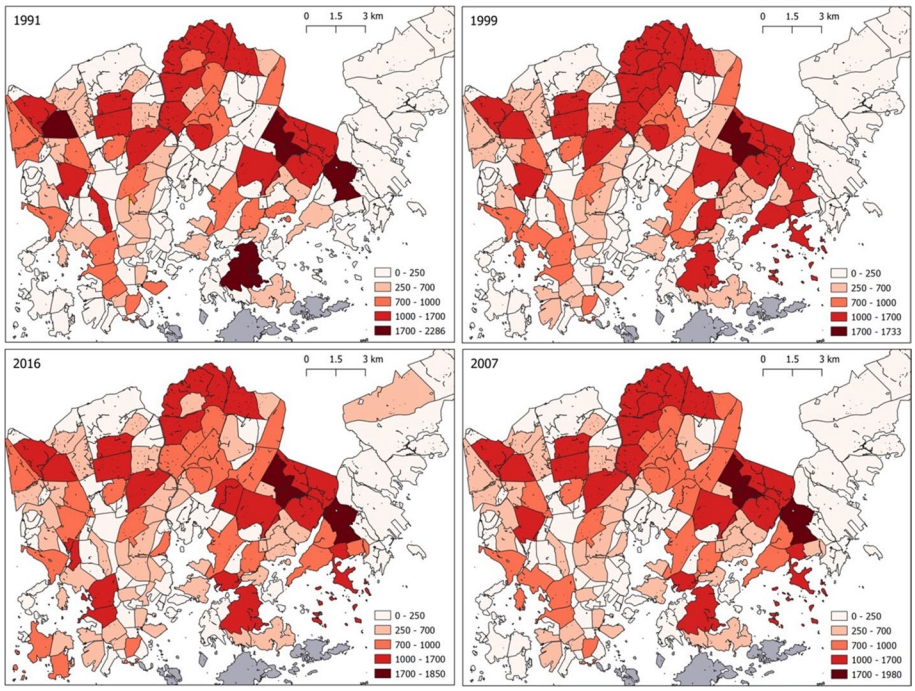


Fig. 6 Population of school-aged children from ages 7–19 by subdistrict

Fig. 7 Isochrone-based accessibility results

Table 3 Optimized hotspot parameters for every timestep

Year	1991	1999	2007	2016
K-nearest neighbors	30	30	30	30
Distance bands (m)	374	366	520	361

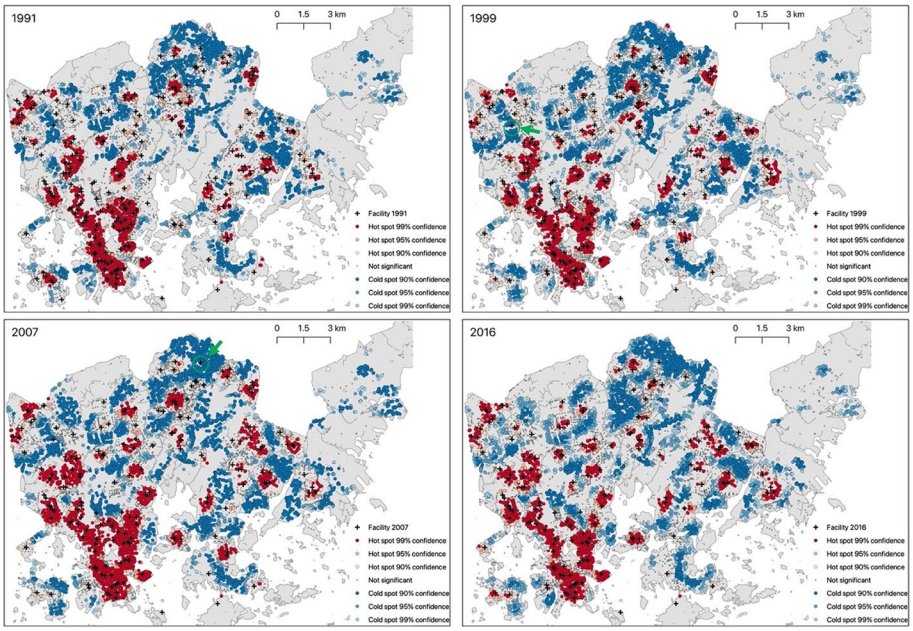


Fig. 8 Hotspot analysis of school accessibility for each year

rely more on alternative modes of transportation, while our analysis only considered a threshold of 1400 m for walking distance.

Comparison with isochrone-based accessibility

We compared the random walk results with the isochrone-based accessibility metric. The school locations are assigned as facilities in ArcGIS, and the service area is calculated as 1400 m away from the facilities.

Figure 7 shows that the accessibility at the center of Helsinki city has increased while it has reduced at the north side of the study area. This might be because, over the years, new schools have been placed strategically to cover a wider geographical area at the center of Helsinki. The disadvantage of isochrone-based measure is that it assumes that the level of accessibility is identical within a zone. This means that every node within the zone is considered to have the same degree of accessibility, which may not reflect the real-world situation accurately. On the other hand, the advantage of the random walk-based accessibility metric is its granularity. It allows evaluating accessibility and observing the changes in accessibility at the node level (i.e., road intersection). Nodes might have different levels of accessibility even if they are located in proximity to each other.

Spatial and temporal evolution of accessibility: hotspot analysis

Hotspot analysis was conducted to identify statistically significant hot- and cold- spots (i.e., clusters) on the accessibility metric (vpw) for each timestep. Based on the optimized parameters of this tool, the most appropriate distance band was selected for the K-nearest neighbor's approach. The parameters for each timestep are given in Table 3. The results of the hotspot analysis are mapped in Fig. 8.

We found that in 1991 the prominent hotspots were located in the southwestern peninsula (see Fig. 8). By 2007, hotspots become larger specifically those located in the north. As new schools emerged in 2007 and the node degrees increased in the northeast part of the city, hotpots became more prominent in this area. From 2007 to 2016, the size of accessibility hotspots decreased in the northern region of the city. This can be attributed to the decreased availability of schools in this area (see Fig. 2). We also found that some schools are surrounded by cold spots indicating the existence of significantly low accessibility values. Our analysis shows that when network topology has low connectivity, the SARW-based accessibility metric may still indicate reduced accessibility to schools, even if a nearby school exists in these areas. We marked some of these schools with green circles in Fig. 8. This phenomenon shows that sparse networks with low connectivity around school locations result in low accessibility scores. "Spatial and temporal evolution of accessibility: hotspot analysis" section will discuss the relationship between network topological metrics and SARW-based accessibility.

We examined the relationship between the identified clusters of low accessibility and the distribution of the population of school-aged children from ages 7–19 (see Fig. 6), to determine if the level of accessibility aligns with the distribution of students. Our analysis showed that the identified clusters of low accessibility correspond to the population of school-aged children from ages 7–19 (as seen in Fig. 6) over all time steps. Despite

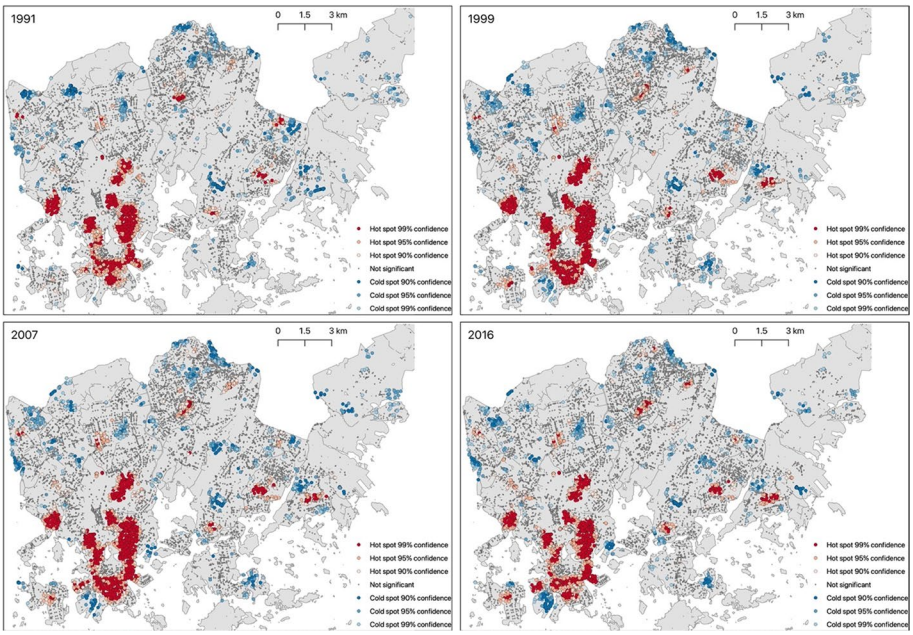


Fig. 9 Hotspot analysis of the average node degree for each year

Fig. 10 Spatial distribution of gamma index values in districts over time

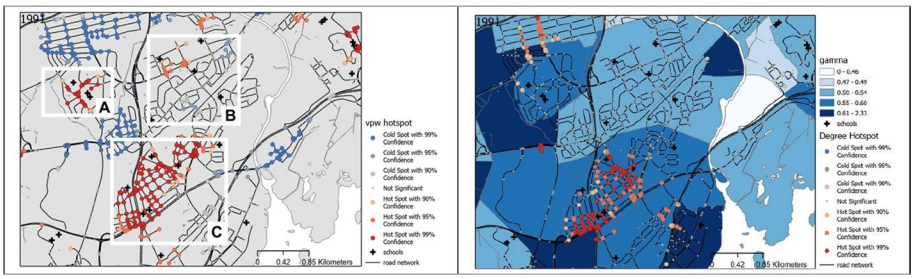


Fig. 11 A sample study area in the City of Helsinki. Accessibility hotspots are displayed on the left side, node degree hotspots and gamma values are displayed on the right side

high student populations being located outside the historical center, these areas displayed relatively low accessibility values.

Network topology indicators

We explore the relationship between network topology indicators and SARW-based accessibility metrics. Figures 9 and 10 illustrate the results of the hotspot analysis for node degree and the gamma index values for each district, respectively.

The results indicate that northern, eastern, and northeastern areas have a low level of connectivity. Between 1991 and 2016 the connectivity decreased in the northern districts, while the southwestern peninsula maintained a high level of connectivity in 2016 (see Fig. 10). A comparison between Figs. 8 and 9 shows that schools surrounded by cold spots (see Fig. 8) are mainly located in areas with low node degrees. Furthermore, these areas correspond to the districts with low gamma index values in Fig. 10.

In contrast, we observed that the historical center of Helsinki has high node degrees in Fig. 9, where the average number of visits per school per walk is also high (see Fig. 5). These results indicate that the number of times a school node is visited during the random walk depends on the network's average node degree and connectivity.

Figure 11 shows a focus area for which we analyzed school accessibility using network topology indicators for 1991. Here, Region A illustrates a high accessibility cluster due to the spatial density of schools in the region. Region B displays an example of lower accessibility due to the low node degree clusters. Region C shows a small high-access cluster with high numbers of schools and the network node degree. This finding validates that the measured level of accessibility is influenced by the transport network and the school locations (Bertolini 2012). Overall, we found that the network topology highly influences the accessibility metrics results indicating that dense networks with a high number of connections result in increased school visits.

Conclusions and implications

This study explores the application of complex network analysis to analyze the evolution of accessibility to education. We proposed an accessibility metric that combines the SARW with the hotspot analysis using the Getis-Ord Gi* statistic. The results indicate that overall accessibility to schools by walking has decreased over time in the center of Helsinki, as measured by the SARW-based accessibility metric. This decline is likely associated with the reduction in the number of schools in the city, from 295 in 1991 to 173 in 2016. However, the remaining schools may have increased their capacities, but we cannot confirm this due to the lack of data. Moreover, the results of the isochrone-based accessibility analysis have not changed significantly over time. We observed some decrease in accessibility in the north and an increase in accessibility at the center of Helsinki. This could be due to the strategic placement of remaining schools to serve a wider geographic area. Overall, our analysis demonstrates that using the SARW accessibility metric and hotspot analysis provides a comprehensive understanding of accessibility evolution, which can aid decision-making and evidence-based spatial planning for improved accessibility in the future.

The advantage of this SARW method is using a road network to determine the range of movements in complex network analysis by introducing the concept of randomness for reaching various nodes. It includes activity locations and movements, which makes it more consistent with the transport and land use cycle proposed by Bertolini (2012). Furthermore, individual movements can be modelled without considering prior information, such as travelling behavior or school capacities. Compared to the commonly used methods, such as isochrone-based accessibility metrics, which assume the same level of accessibility within a zone, the key benefit of this metric is its granularity (i.e., node-level) providing a more detailed spatial accessibility analysis.

Furthermore, the stochastic nature of the SARW method allows for modelling accessibility that captures the inherent randomness of human activity patterns. In contrast, isochrone-based methods, which employ deterministic shortest path algorithms, primarily emphasize network distances when determining accessibility outcomes. Nevertheless, each method could be useful in different context and brings a different perspective on accessibility. Recognizing these distinctions is crucial for leveraging the

strengths of each method and tailoring their application to different analytical scenarios. The network's connectivity heavily influences the accessibility results obtained by the SARW method, which returns low accessibility values in sparse networks (see green circles in Fig. 8). Therefore, it provides a better reflection of the accessibility and connectivity level of networks.

The random walk model's computational performance depends on the parameter selection (see "Comparison with Isochrone-based accessibility" section). Increasing the number of walks and the distance threshold proportionately increases the model's computation time. We restricted the number of walks to 500 as no significant change was observed in the consistency of the output with a higher number of walks. In addition, the distance threshold was selected considering the real-world situation where schools are often accessed by walking. These parameters led to a short computational time. However, for implementations of a larger scope, high-performance computing approaches could be needed. The reason stems from the way the random walk approach is designed, that is, compensating the randomness factor by a high number of repetitions to obtain a well-rounded probability distribution.

In terms of analysis of the evolution of accessibility, we found that the hotspot analysis when applied through historical timesteps in a repeatable fashion allows identifying significant clusters (i.e., hot- and cold- spots) is useful to generalize the SARW-based accessibility metric results. Note that the selection of input parameters in the optimized hot spot analysis tool has the potential to affect the resulting cluster structure. Nevertheless, spatial comparison of these clusters over time enables urban planners and decision-makers to identify contextual problems with infrastructure and land use decisions.

This study has some limitations. First, the capacity of schools which influences the number of school visits is not considered. When such data is available, capacities can be included by limiting the random walk visits to schools based on the capacity value. However, this would require a simultaneous simulation of random walks from all nodes and increase the computational load on the model. We took a sequential approach, simulating random walks one starting node at a time. Second, this study included primary, secondary, and high school data. However, a more detailed analysis should be done for each type of school and compared with the population characteristics of target age groups. The third limitation is related to the theoretical basis of the random walk method. The SARW algorithm does not represent real-life mobility patterns but rather random behavior. Therefore, it cannot be used to simulate human behavior, but only as a proxy for accessibility. Another limitation of this study pertains to the transportation network data. Although we focused solely on walking accessibility, we utilized the transportation network data assuming pedestrians can freely traverse the road network. However, factors such as fences and other barriers may impact the actual reachability to educational facilities.

In addition, the parameters selected were based on the walking speed and distance range. There are three main points of concern regarding the SARW model. First, due to the self-avoidance property, increasing the threshold beyond a certain point does not proportionately increase the reach. This is due to the increased likelihood of encountering the stopping conditions of the random walk. Further analysis is

necessary to determine the sensitivity of the distance threshold and its impact on the output results. Secondly, as long travel distances are considered, the effect of competition in the transport modes becomes more pronounced. Therefore, the model choice of not including public transport or cycling becomes more questionable when considering larger thresholds. For the inclusion of public transport, a multi-layer approach can be used, which has applications in the literature (Ding et al. 2021), but is not included in this study. Lastly, in our study, we specifically chose to compare the SARW-based accessibility metric with the isochrone-based accessibility approach. However, it is worth noting that further exploration of SARW's capabilities would involve comparing it with other accessibility methods. By doing so, we can gain a comprehensive understanding of SARW's effectiveness in assessing accessibility. Further research is necessary to fully explore the advantages of SARW-based accessibility, specifically over isochrone-based accessibility, using different sets of parameters and various network topologies.

Acknowledgements
Not applicable.

Author contributions
NYA contributed by conceptualization, investigation, methodology, resources, formal analysis, supervision, validation, visualization, writing—original draft. EY contributed by conceptualization, investigation, methodology, data curation, formal analysis, visualization, writing—review & editing. YC contributed by conceptualization, investigation, methodology, supervision, data curation, writing—review & editing. BVW contributed by conceptualization, supervision, writing—review & editing.

Funding
Not applicable. The authors received no financial support for the research, authorship, and/or publication of this article.

Availability of data and materials
The datasets generated and/or analysed during the current study are available in the "Helsinki Region Infoshare." repository online available: https://hri.fi/en_gb/. The digitized datasets used and/or analysed during the current study are available from the corresponding author on reasonable request.

Declarations

Ethics approval and consent to participate
Not applicable. This study does not use human and/or animal subjects.

Consent for publication
Not applicable.

Competing interests
The authors declare that they have no competing interests.

Received: 21 February 2023 Accepted: 7 August 2023
Published online: 30 August 2023

References

An evaluation of primary schools and its accessibility using GIS techniques: a case study of Prayagraj district, India. https://doi.org/10.1007/s10708-022-10715-3. Accessed on 13 Jul 2023

Andersson E, Malmberg B, Östh J (2012) Travel-to-school distances in Sweden 2000–2006: changing school geography with equality implications. J Transp Geogr 23:35–43. https://doi.org/10.1016/j.jtrangeo.2012.03.022

ArcGIS Desktop Help 9.3—algorithms used by network analyst. https://webhelp.esri.com/arcgisdesktop/9.3/index.cfm?TopicName=Algorithms_used_by_Network_Analyst. Accessed on 05 Jul 2023

ARCGIS (2022) Accessed on 30 Nov 2022. [Online]. ESRI.com

Aydin NY, Duzgun HS, Wenzel F, Heinimann HR (2018) Integration of stress testing with graph theory to assess the resilience of urban road networks under seismic hazards. Nat Hazards 91(1):37–68. https://doi.org/10.1007/s11069-017-3112-z

Aydin NY, Casali Y, Duzgun HS, Heinimann HR (2019) Identifying changes in critical locations for transportation networks using centrality. In: Geertman S, Zhan Q, Allan A, Pettit C (eds) Computational urban planning and management for smart cities. Springer International Publishing, Cham, pp 405–423

Barthélemy M (2011) Spatial networks. Phys Rep 499(1–3):1–101. https://doi.org/10.1016/j.physrep.2010.11.002

Barthélemy M, Flammini A (2009) Co-evolution of density and topology in a simple model of city formation. Netw Spat Econ 9(3):401–425. https://doi.org/10.1007/s11067-008-9068-5
Bertolini L (2012) disP—the planning review integrating mobility and urban development agendas: a manifesto. Plann Rev 48(October):16–26
Brussel M, Zuidgeest M, van den Bosch F, Munshi T, van Maarseveen M (2019) Can the urban poor reach their jobs? Evaluating equity effects of relocation and public transport projects in Ahmedabad, India. In: Measuring transport equity. Elsevier Inc., Amsterdam. https://doi.org/10.1016/B978-0-12-814818-1.00006-8
Casali Y, Heinimann HR (2019) A topological analysis of growth in the Zurich road network. Comput Environ Urban Syst 75(March):244–253. https://doi.org/10.1016/j.compenvurbsys.2019.01.010
Casali Y, Heinimann HR (2020) Robustness response of the Zurich road network under different disruption processes. Comput Environ Urban Syst 81:101460. https://doi.org/10.1016/j.compenvurbsys.2020.101460
Cascetta E, Cartenì A, Montanino M (2013) A New measure of accessibility based on perceived opportunities. Procedia Soc Behav Sci 87:117–132. https://doi.org/10.1016/j.sbspro.2013.10.598
Cats O (2017) Topological evolution of a metropolitan rail transport network: the case of Stockholm. J Transp Geogr 62:172–183. https://doi.org/10.1016/j.jtrangeo.2017.06.002
Cheng L, Yang M, De Vos J, Witlox F (2020) Examining geographical accessibility to multi-tier hospital care services for the elderly: a focus on spatial equity. J Transp Health 19:100926
City of Helsinki (2020) Walking|City of Helsinki. 2020. [Online]. Available on https://www.hel.fi/helsinki/en/maps-and-transport/cycling/walking/
City of Helsinki (2021) Helsinki Region Infoshare 2021. [Online]. https://hri.fi/en_gb/
Curl A (2018) The importance of understanding perceptions of accessibility when addressing transport equity: a case study in Greater Nottingham, UK. J Transp Land Use 11(1):1147–1162. https://doi.org/10.5198/jtlu.2018.1003
Di Paolo A, Matas A, Raymond JL (2017) Job accessibility and job-education mismatch in the metropolitan area of Barcelona. Pap Reg Sci 96:S91–S112. https://doi.org/10.1111/pirs.12179
Dickerson A, McIntosh S (2013) The impact of distance to nearest education institution on the post-compulsory education participation decision. Urban Stud 50(4):742–758. https://doi.org/10.1177/0042098012455717
Ding R (2019) The complex network theory-based urban land-use and transport interaction studies. Complexity. https://doi.org/10.1155/2019/4180890
Ding R, Zhang Y, Zhang T, Ma C (2021) Development of a complex network-based integrated multilayer urban growth and optimisation model for an efficient urban traffic network. Complexity. https://doi.org/10.1155/2021/3467485
Ertugay K, Argyroudis S, Düzgün HŞ (2016) Accessibility modeling in earthquake case considering road closure probabilities: a case study of health and shelter service accessibility in Thessaloniki, Greece. Int J Dis Risk Reduct 17:49–66. https://doi.org/10.1016/j.ijdrr.2016.03.005
ESRI (2023) How optimized hot spot analysis works. https://pro.arcgis.com/en/pro-app/latest/tool-reference/spatial-statistics/how-optimized-hot-spot-analysis-works.htm. Accessed on 06 Jul 2023
Geopy (2022) Geocoding library for Python. [Online]. Available on https://github.com/geopy. Accessed 07 Jul 2022
Geurs KT, Ritsema van Eck JR (2003) Evaluation of accessibility impacts of land-use scenarios: the implications of job competition, land-use, and infrastructure developments for the Netherlands. Environ Plann B Plann Des 30(1):69–87. https://doi.org/10.1068/b12940
Geurs KT, van Wee B (2004) Accessibility evaluation of land-use and transport strategies: review and research directions. J Transp Geogr 12(2):127–140. https://doi.org/10.1016/j.jtrangeo.2003.10.005
Hanna S (2020) Random walks in urban graphs: a minimal model of movement. EPB: Urban Anal City Sci 48(6):1711–1967. https://doi.org/10.1177/2399808320946766
He Z, Navneet K, van Dam W, Mieghem PV (2021) Robustness assessment of multimodal freight transport networks. Reliab Eng Syst Saf 207:107315. https://doi.org/10.1016/j.ress.2020.107315
Hu Y, Downs J (2019) Measuring and visualizing place-based space-time job accessibility. J Transp Geogr 74:278–288. https://doi.org/10.1016/j.jtrangeo.2018.12.002
Kalinic M, Krisp JM (2018) Kernel density sstimation (KDE) vs. hot-spot analysis—detecting criminal hot spots in the city of San Francisco. In: 21st International Conference on Geographic Information Science (AGILE), June, pp 1–5
Kansky KJ (1963) Structure of transportation networks: relationships between network geometry and regional characteristics. The University of Chicago, Chicago
Kaupunki H (2016) HLT 16—kaupunkijulkaisu. Helsinki
Koulurekisteri (2020) Koulurekisteri 2020. [Online]. https://dev.hel.ninja/projects/koulurekisteri/
Lee M, Kim JB (2021) Spatial distribution of access diversity on urban road networks. J Korean Phys Soc 79:504–511. https://doi.org/10.1007/s40042-021-00257-1
Li Y, Lin Y, Geertman S, Hooimeijer P, Xu W (2021) Accessibility-based equity of public facilities: a case study in Xiamen, China. Appl Spat Anal Policy 14(4):947–968. https://doi.org/10.1007/s12061-021-09387-2
Liljamo T, Liimatainen H, Pöllänen M, Viri R (2021) The effects of mobility as a service and autonomous vehicles on people's willingness to own a car in the future. Sustainability 13(4):1–28. https://doi.org/10.3390/su13041962
Mata AS (2020) Complex networks: a mini-review. Braz J Phys 50:658–672
Mei D, Xiu C, Feng X, Wei Y (2019) Study of the school-residence spatial relationship and the characteristics of travel-to-school distance in Shenyang. Sustainability 11(16):1–15. https://doi.org/10.3390/su11164432
METREX (2020) Helsinki City Boulevards Strategy and Projects
NetworkX (2022) Software for complex networks. [Online]. https://networkx.org/. Accessed 07 Jul 2022
Nevanlinna AK (2016) Ports, industries and economics in the urban history of Helsinki. 2016. [Online]. https://www.kvartti.fi/en/articles/ports-industries-and-economics-urban-history-helsinki
Sá C, Florax RJGM, Rietveld P (2006) L'accessibilité à l'enseignement supérieur importe-t-elle? Comportement du choix des diplômés de l'enseignement secondaire au pays-bas. Spat Econ Anal 1(2):155–174. https://doi.org/10.1080/17421770601009791
Sharma G, Patil GR (2022) Spatial and social inequities for educational services accessibility—a case study for schools in Greater Mumbai. Cities 122:103543. https://doi.org/10.1016/j.cities.2021.103543

Söderström P, Schulman H, Ristimäki M (2015) Urban Form in the Helsinki and Stockholm City Regions
Strano E, Nicosia V, Latora V, Porta S, Barthélemy M (2012) Elementary processes governing the evolution of road networks. Sci Rep 2:1–8. https://doi.org/10.1038/srep00296
Tenkanen H, Toivonen T (2019) Helsinki region travel time matrix. [Online]. Available on https://zenodo.org/record/3247564
Travençolo BA, Costa LD (2008) Accessibility in complex networks. Phys Lett Sect a: General Atomic Solid State Phys 373(1):89–95. https://doi.org/10.1016/j.physleta.2008.10.069
Vecchio G, Tiznado-Aitken I, Hurtubia R (2020) Transport and equity in Latin America: a critical review of socially oriented accessibility assessments*. Transp Rev 40(3):354–381. https://doi.org/10.1080/01441647.2020.1711828
Viljoen NM, Joubert JW (2019) Supply chain micro-communities in urban areas. J Transp Geogr 74:211–222. https://doi.org/10.1016/j.jtrangeo.2018.11.011
Wang M, Chen Z, Mu L, Zhang X (2020) Road network structure and ride-sharing accessibility: a network science perspective. Comput Environ Urban Syst 80:101430. https://doi.org/10.1016/j.compenvurbsys.2019.101430
van Wee B, Mouter N (2021) Evaluating transport equity (1st edn, vol 7). In: Advances in transport policy and planning. Elsevier Inc., Amsterdam. https://doi.org/10.1016/bs.atpp.2020.08.002
Wen S, Shi J, Zhang W (2021) Impact of urban rail transit network on residential and commercial land values in China: a complex network perspective. Complexity 2021:8849066. https://doi.org/10.1155/2021/8849066
Xu Y, Song W, Liu C (2018) Social-spatial accessibility to urban educational resources under the school district system: a case study of public primary schools in Nanjing, China. Sustainability. https://doi.org/10.3390/su10072305

Publisher's Note