scientific reports

Check for updates

OPEN

Urban segregation on multilayered transport networks: a random walk approach

Mateo Neira[1,2✉], Carlos Molinero[2], Stephen Marshall[3] & Elsa Arcaute[1,2]

We present a novel method for analysing socio-spatial segregation in cities by considering constraints imposed by transportation networks. Using a multilayered network approach, we model the interaction probabilities of socio-economic groups with random walks and Lévy flights. This method allows for evaluation of new transport infrastructure's impact on segregation while quantifying each network's contribution to interaction opportunities. The proposed random walk segregation index measures the probability of individuals encountering diverse social groups based on their available means of transit via random walks. The index incorporates temporal constraints in urban mobility with a parameter, $\alpha \in [0, 1)$, of the probability of the random walk continuing at each time step. By applying this to a toy model and conducting a sensitivity analysis, we explore how the index changes dependent on this temporal constraint. When the parameter equals zero, the measure simplifies to an isolation index. When the parameter approaches one it represents the city's overall socio-economic distribution by mirroring the steady-state of the random walk process. Using Cuenca, Ecuador as a case study, we illustrate the method's applicability in transportation planning as a valuable tool for policymakers, addressing spatial distribution of socio-economic groups and the connectivity of existing transport networks, thus promoting equitable interactions throughout the city.

Cities exist to connect people, influencing human interactions through their structural networks[1–3]. A successful city maximises face-to-face interactions[1], providing equal opportunities for all inhabitants[4]. However, understanding the dynamics in cities remains a challenge due to the complexity of urbanisation, social effects, and policy concerns[5]. One critical factor influencing these dynamics is urban segregation, which affects socio-spatial interactions within cities.

Urban segregation refers to the spatial separation of different social, ethnic, racial, or economic groups within a city[6]. It is characterised by the unequal distribution of these groups across neighbourhoods or areas within a city - presenting distinct socio-spatial patterns[7]. Urban segregation can occur due to various factors, such as historical policies, social preferences, economic disparities, or discrimination[8]. The unequal distribution of population groups in cities can have significant consequences, affecting access to resources, opportunities (both social and economic) while reinforcing social isolation and perpetuating inequalities[9]. Understanding urban segregation is essential for fostering social integration within cities, reducing disparities and promoting more equitable and inclusive places[10].

Although urban segregation has been extensively studied, there is limited research investigating how coupled transport systems affect urban segregation. To address this gap, the current study integrates concepts from complexity science and complex network analysis, proposing a novel approach to examine urban segregation in relation to transport networks to quantify the likelihood of interactions between different population groups. This study aims to shed light on the intricate connections between urban mobility and segregation, thereby providing insights that can inform more inclusive and equitable urban planning and policies.

Residential segregation can be defined as any pattern in the spatial distribution of population groups that deviates significantly from a random distribution[11]. This distribution can be a product of social and spatial differentiation, as people have different preferences and resources. People's individual choices can lead to an aggregate outcome that is completely different than what one would expect[12]. These ideas were formalised by physicists who coined the term *sociodynamics*[13] and *sociophysics*[14]. Segregation, diffusion, and collective behaviour were explored through this lens with the seminal work of Thomas Schelling[15] and Mark Granovetter[16]. Most statistical

[1]Alan Turing Institute, British Library, London NW1 2DB, UK. [2]Centre for Advanced Spatial Analysis, University College London, London W1T 4TJ, UK. [3]Bartlett School of Planning, University College London, London WC1H 0QB, UK. ✉email: mateo.neira.16@ucl.ac.uk

methods used to measure segregation found in the literature aim to assess how evenly different population groups are distributed in different areas of the city.

Although many methods to measure urban segregation had been developed, it wasn't until the work of Massey and Denton[7] that they were systematically reviewed and organised. Massey and Denton defined segregation as the degree to which two or more groups live separately from one another. The concept of urban segregation now encompasses various types, including residential, workplace, and experienced segregation. Residential segregation refers to the spatial separation of social groups within neighbourhoods, while workplace segregation pertains to the unequal distribution of social groups across occupations and workplaces. Experienced segregation, on the other hand, concerns segregation as experienced through the daily activities people undertake in urban areas.

Residential and workplace segregation have long been the focus of urban segregation research, with a growing emphasis on experienced segregation in more recently years with the increase amount of data about people's mobility patterns as captured through location enabled devices[17,18]. This shift towards examining daily mobility patterns unveils a more nuanced understanding of segregation. Similarly recent work has also employed an intersectional approach using mobility surveys to reveal variation in mobility patterns based on gender, age, and social classes in cities[19].

Segregation has also been a problem explored within network science, specifically in the study in social networks; although the framework is different from residential segregation, it has been extended in interesting ways to spatially embedded networks. In social networks, individuals have a tendency to relate with others who are similar to them across different characteristics; a property known as *homophily*[20], and known within network science as *assortative mixing*[21]. This can have important implications in the information people in the network receive, attitudes they form, and interactions they experience[22,23]. Measures of assortative mixing can be related to the measures of exposure from residential segregation we discussed previously. Many measures have been proposed by different authors, and in general, they can be divided into two approaches; *descriptive graph statistics* and *spectral graph theory*[22].

More recent studies have increasingly focused on understanding how segregation manifests itself at different scales[24–26]. However, the scale at which to study segregation still remains an open question, with most research studying segregation using different methods but mostly working with census track data and their adjacencies to derive segregation measures. Underlying these studies there is an assumption that all persons sharing a tract, whether they are located in the centre of the tract or towards the periphery, have equal proximity to residents outside the tract as well as being equally proximate to everyone within its boundaries. This assumption stems from treating tracts as as spatially homogeneous irrespective of their relative distances and the connectivity patterns provided by the street network and additional transport networks that might be present. In our work we address this problem by providing a method that explicitly takes into account a city's connectivity structure by modelling all modes of transport as a multilayered network. This methodological innovation allows us to analyse the impacts of different transport modes on socio-spatial segregation, offering a more detailed perspective on the potential of transport infrastructure to both exacerbate and mitigate segregation.

The proposed method for quantifying segregation using multilayer networks and random walks takes into account the heterogeneity of the connectivity different groups of people have dependent of where they are located in a city. Segregation is understood as unequal opportunities for encounters, measured as the lack of *exposure* between different population groups. To measure the lack of exposure between groups their spatial relationships must be taken into account, as well as the constraints that the available transport networks impose on the opportunities for encounters. The paper is structured as follows: we first provide a brief overview of similar studies that have used random walks to measure segregation. Then we introduce the methodology and apply the method to a toy model in order to explore the sensitivity of the parameters. Finally, we apply the framework to study segregation in the city of Cuenca, Ecuador, and we show how the measure can be used to assess the impact of transport infrastructure on segregation.

Related works

Socio-spatial segregation, the uneven distribution of population groups within a city, has been a focal point of research due to its implications for societal inequalities[7,11]. Traditional studies have predominantly used aggregate-level indices to measure the extent of segregation within predefined spatial units, focusing mainly on the dimensions of evenness and exposure[22]. However, these methods often overlook the complex social and spatial interactions within urban environments.

Recent advancements in network science, including measures like the assortativity coefficient[21] and the spectral segregation index[27], have begun to address these complexities by considering social and spatial network structures. Despite their advancements, these measures face limitations, such as reliance on residential proximity and sensitivity to group composition.

To address these issues, Ballester and Vorsatz[28] introduced a random-walk based approach, utilising an eigenvector-based centrality measure to estimate the likelihood of inter-group encounters. Further building on random-walk methods, Sousa and Nicosia[29] proposed non-parametric measures that use random walk trajectories to quantify segregation, accounting for the diversity of urban systems through a null model. Although these graph-theoretic approaches provide a natural way to capture the dimensions of residential segregation, their application has been limited to arbitrary definitions of areal units in which different groups reside. However, they provide an intuitive way to include information about the links between places in a city and capture segregation at a more dis-aggregate spatial scale as a result of limitations on probabilities of interactions between different groups.

More recently a number of studies have quantified how segregated urban areas are by analysing individual mobility patterns and the places people visit[18,30–32]. These studies emphasises the use of large-scale, high-resolution

data (e.g., GPS data from smartphones, credit card transactions, mobile phone data) to analyse urban segregation and human mobility. Within this context—we complement these studies by providing a method that can help to interrogate how the urban structure—in terms of its transport network, once established, can exacerbate or mitigate segregation patterns. The choice of a multilayered network and a random walk approach[33] is driven by the need to model these complex interactions comprehensively.

Building on this foundation, we propose a novel approach using multilayered networks to model the city's transport systems, integrating socio-economic data to assess segregation more dynamically. This method employs a modified random-walk-based segregation index[28] to quantify interaction probabilities among population groups, factoring in the spatio-temporal constraints of the transport network. This approach not only overcomes the spatial limitations of previous methods but also elucidates the role of transport infrastructure in shaping urban segregation patterns.

Multilayered networks, with their multiple types of connections among nodes, offer a comprehensive framework for analyzing complex phenomena like information spread and disease transmission. In our work, each layer represents a different transport system, providing a nuanced view of how urban planning and policy interventions can leverage these networks to create more integrated environments. By challenging spatial homogeneity assumptions and highlighting targeted infrastructure improvements, our approach advocates for proactive urban planning to mitigate segregation and promote equity.

Methods

Random walk segregation on multilayered networks

We can study the structure and interactions in urban systems through their networks. In particular, multilayered networks allow us to capture the coupling of multiple transport systems, and by doing so it is able to better capture the spatio-temporal constraints they impose on different places and can be used to measure urban segregation at a fine spatial resolution.

Here we set a formal description of the independent transport network graphs and their inter-modal coupling to create the multilayered network. We describe the assignment of socio-economic variables to the multilayered network, as well as the measures that will be used to assess change in urban structure and segregation.

Each transport network is modelled as a directed graph in their *primal* representation, where each intersection or station is modelled as a node, and their connections—such as streets, routes, or transport lines—are modelled as temporally weighted links. All the transport networks are represented by an ordered list of networks, $\vec{G}$ given by:

$$\vec{G} = (G_1, G_2, ..., G_i, ...G_M) \tag{1}$$

where M indicates the total number of transport modes available in the city, and $G_i = (N_i, L_i, w_i)$ is the transport network i. N_i is a set of nodes, L_i is a set of links, and w_i is a function that takes a link and returns a weight ($w_i : L_i \to \mathbb{R}$) equal to the travel time between the links of the node in minutes. In the case of street networks, this is calculated using an average walking speed of 5 km/h.

Additionally, we define the coupling of the different transport networks through a $M \times M$ list of bipartite networks $\mathcal{G}_{i,j} = (N_i, N_j, L_{i,j}, w_{i,j})$ for each $i < j$ and $i, j \in \{1, 2...M\}$. $\mathcal{G}$ indicates the bipartite network with node sets N_i and N_j and the link set $L_{i,j}$. The links of the network $\mathcal{G}_{i,j}$ are called interlinks and connect the nodes of layer i to nodes of a different layer j and $w_{i,j}$ is a function that takes a link and returns a weight ($w_{i,j} : E_{i,j} \to \mathbb{R}$) equal to the transfer time in minutes between two transport layers.

Finally, the multilayer network $\mathcal{M}$ is given by the triple:

$$\mathcal{M} = (Y, \vec{G}, \mathcal{G}) \tag{2}$$

where Y indicates the set of layers $Y = \{i | i \in \{1, 2, ..., M\}\}$ for each transport network. This multilayered network can be described by a supra-adjacency matrix $A_{\mathcal{M}}$[34] and the corresponding time-weighted supra-adjacency matrix $T_{\mathcal{M}}$.

The minimum number of transport networks a city can have is one, corresponding to the street network. Socio-economic data needs to be incorporated into the nodes of the street network to measure urban segregation using the multilayered model. Formally, this is captured by classifying each individual from a total set $N = \{1, 2, ..., n\}$ of n individuals in a city, into different socio-economic groups B. Let $n_{b,i}$ be the number of individuals of group $b \in B$ that will start a journey through the city in the multilayered nodes $i \in N_\mu$, where μ denotes the layer, and in particular $\mu = 1$ corresponds to the street network. The number of individuals who belong to a group b is $n_b = \sum_i n_{b,i}$, and the number of individuals who will start their journey from node i is $n_i = \sum_{b \in B} n_{b,i}$. The column vectors $c_b = (\frac{n_{b,i}}{n_i})_i$ and $d_b = (\frac{n_{b,i}}{n_b})_i$ are referred to as the vectors of group concentrations and group densities, respectively.

This representation allows us to create time-weighted paths, as illustrated in Fig. 1b, and calculate the probabilities of different nodes in the system being occupied by different population groups to measure segregation.

Dynamic isolation index

To calculate segregation in a city we define a transition matrix P that contains entries $\pi_{i \to j}$, which indicate the probability that a random walker transitions from node i to node j at each time step. Additionally, we define a parameter $\alpha \in [0, 1)$ that encodes the probability that at each time step the random walk continues, or stops (with probability $1 - \alpha$). This parameter is important in calibrating the expected duration of the random walk to align with empirical observations of urban mobility. Given this transition matrix and the parameter α, the probability of a walk starting in i and ending in j is defined as q_{ij} in a matrix Q such that $Q = (1 - \alpha)(I - \alpha P)^{-1} P$.
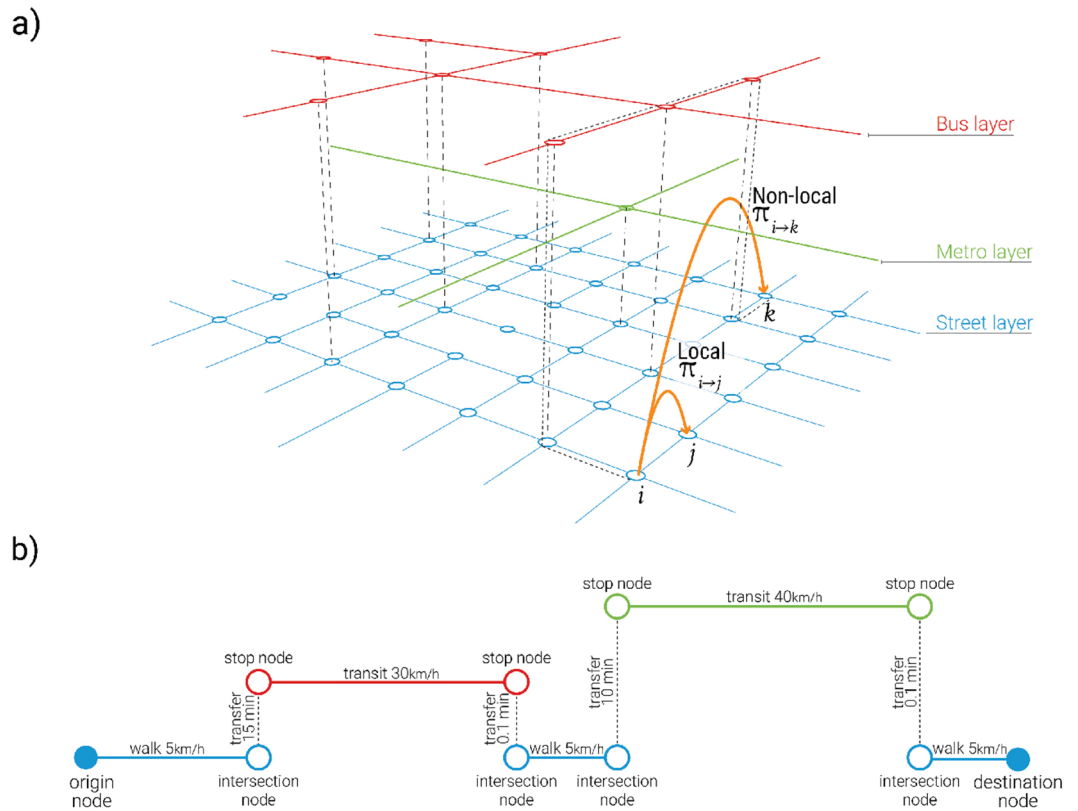
Figure 1. (**a**) Example of a multilayer network modelling different transport systems and their interconnections along with possible types of transitions that can be modelled using local and non-local random walks. (**b**) An illustration of a time-weighted path along the multilayered network.

Given the initial concentrations c_b and densities d_b of populations groups across the city, the normalised segregation ($\bar{\sigma}_{b,i}$) of each group b in the node i is defined as follows:

$$\bar{\sigma}_{b,i} = \left(\frac{n_b}{n}\right)^{-1} d_{b,i} \sum_j q_{ij} c_{b,j} \tag{3}$$

where $\bar{\sigma}(b,i)$ is equal to the probability, up to the multiplicative scalar n, that a randomly chosen individual from group b encounters another randomly chosen individual from the same group. Note that if $\bar{\sigma}(b,i) > 1$ we can say that group b is isolated. The segregation index of group b in the city is the average of the segregation indices of all the nodes for that group: $\bar{\sigma}_b = \sum_i \bar{\sigma}_{b,i}$, and the segregation of the city is defined as the weighted average over the segregation indices of the groups: $\bar{\sigma} = \sum_b \frac{n_b}{b} \bar{\sigma}_b$.

Given this definition, the segregation index depends on two values: α and the probability of transition $\pi_{i \to j}$. α encodes the temporal constraints on mobility and can be related directly to the amount of time people are willing to spend on travel τ, where the expected τ of a random walk given α is:

$$\mathbb{E}(\tau) = \overline{T_{\mathcal{M}}} \frac{1}{1-\alpha}. \tag{4}$$

An important aspect of our model is the interpretation of the alpha parameter. When α approaches 1, the model assumes that the random walk almost always continues, leading to a scenario where the transition matrix P mirrors the steady-state distribution of the random walk. This state reflects a scenario where the movement patterns within the city reach a dynamic equilibrium, allowing us to examine the long-term behaviour and connectivity within the urban network.

Random walk strategies

Diverse types of random walk strategies can be explored in terms of the weight matrices[33]. In this work we look at the diffusion process of a random walk to measure segregation by using two different strategies: local and non-local random walks. For local strategies, a random walker is restricted to adjacent nodes on the network, where as for non-local strategies a random walker can hop beyond nearest neighbours with a probability given by a generalised cost of moving to a particular location as shown in Fig. 1a.

Local random walks
Normal random walk: in this case, the weights coincide with the elements of the adjacency matrix[35], from which we can calculate the transition matrix given by :

$$\pi_{i\rightarrow j} = \frac{A_{ij}}{k_i}.$$

(5)

where A_{ij} is the adjacency matrix element indicating the presence (1) or absence (0) of a link between nodes i and j, and k_i is the degree of node i. This approach ensures that the normal random walker transitions with equal probability to any of its immediate neighbours, embodying a uniform distribution of transition probabilities that reflect an unbiased exploration of the network's local structure.

Preferential navigation: in the preferential navigation case, a random walker transitions to a neighbour with a probability that depends on a quantity $q_i > 0$ assigned to each node i of the network. The value q_i can represent a topological feature of the respective node or a value independent of the network structure, that quantifies an existing resource at each node,

$$\pi_{i\rightarrow j} = \frac{A_{ij}q_j^{\beta}}{\sum_{l=1}^{N} A_{il}q_l^{\beta}},$$

(6)

where β controls the influence of the incorporated features into the random walk. Such features can encompass information about the global structure of the network. This weighting introduces a bias in the transition probabilities, making transitions to more central nodes more likely. This approach can indeed lead to a transition matrix that approximates a power-law distribution, especially in networks where node centrality is unevenly distributed. The initial transition matrix for preferential navigation is generated by first calculating the centrality measure for each node and then normalising these values to sum to 1 for each row in the matrix, ensuring that they represent valid probabilities.

Non-local random walks
Non-local random walks on the network are motivated by the possibility of transitioning from one node to another irrespective if it is a direct neighbour. One particular type of non-local random walks are Lévy flights. These are random walks with displacements of length l, and a probability distribution given by an inverse power-law relation. There have been empirical studies showing that human mobility patterns display this type of behaviour[36–38].

Lévy flights are characterized by step lengths that follow an inverse power-law distribution, allowing for both short and long-range moves within the network. The transition probabilities for a Lévy flight are determined by the inverse of the distance d_{ij} between nodes, raised to the power of β. Lévy flights on networks can be described in terms of weights $d_{ij}^{-\beta}$, where $\beta \in \mathbb{R}_{>0}$, $\mathbb{R}_{>0} = \{x \in \mathbb{R} \mid x > 0\}$. In the case of Lévy flights the probability transition matrix is equal to $d_{ij}^{-\beta}$—where the probability of transitioning from node i to node j in the network is a function of the time it takes to travel within the network from i to j. We can use this probability transition matrix to calculate Q such that $Q = (1-\alpha)(I - \alpha P)^{-1}P$.

Incorporating both local and non-local random walk strategies enables our model to simulate a wide range of mobility behaviours, from local movements to occasional long-distance travel, providing a comprehensive framework for analysing urban segregation dynamics.

Results
Measuring segregation on a synthetic city
We test the method by first running the different types of random walks and measuring the resultant segregation on a toy model represented by a mono-layered network comprised of 50 nodes and a population of 500 divided into two groups, as shown in Fig. 2. For this toy model, we test the three types of random walks to visualise how the network structure affects segregation, and run a sensitivity analysis for the parameters on both local and non-local random walks.

Figure 3 shows the influence that different dynamics have on the measure of segregation on the toy model. The measure is able to capture the influence of the network structure on segregation, where not only are the nodes that act like bridges between the two communities less segregated, but also the nodes that are within a short network distance. When the dynamics change from normal to preferential random walk, segregation decreases, as most people will tend towards a smaller subset of final positions regardless of their initial position. In the case of a Lévy flight, the segregation is the lowest, this is not directly comparable to the previous two types of random walks since, in this case, the α value no longer represents the same temporal constraint. This is because it's the relationship between α and expected travel time is influence by the step length distribution rather than the temporal sequence of steps. This changes the characteristics of the walk, allowing more long-distance jumps, which impacts segregation differently compared to the other two dynamics.

We perform a sensitivity analysis of both preferential and Lévy flight models of segregation, in order to show the impact of the parameter values on the segregation index, see Fig. 4. It is important to note that a segregation index closer to zero indicates no segregation, while increasing values signify higher segregation levels.

In the preferential random walk model, the segregation index is primarily influenced by parameter α; as α and β increase, segregation decreases due to greater spatial mobility and individuals converging to the same subset of nodes. For the Lévy flight model, higher values of β lead to reduced spatial mobility, and as β approaches two, segregation increases dramatically.
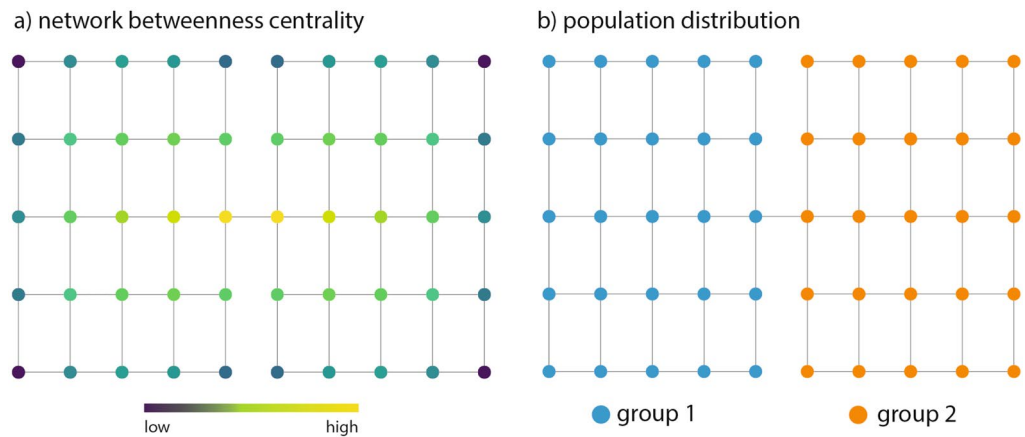
Figure 2. Features of toy model. (**a**) Street network structure and betweenness centrality values, (**b**) population distribution by groups, each node contains 10 people.
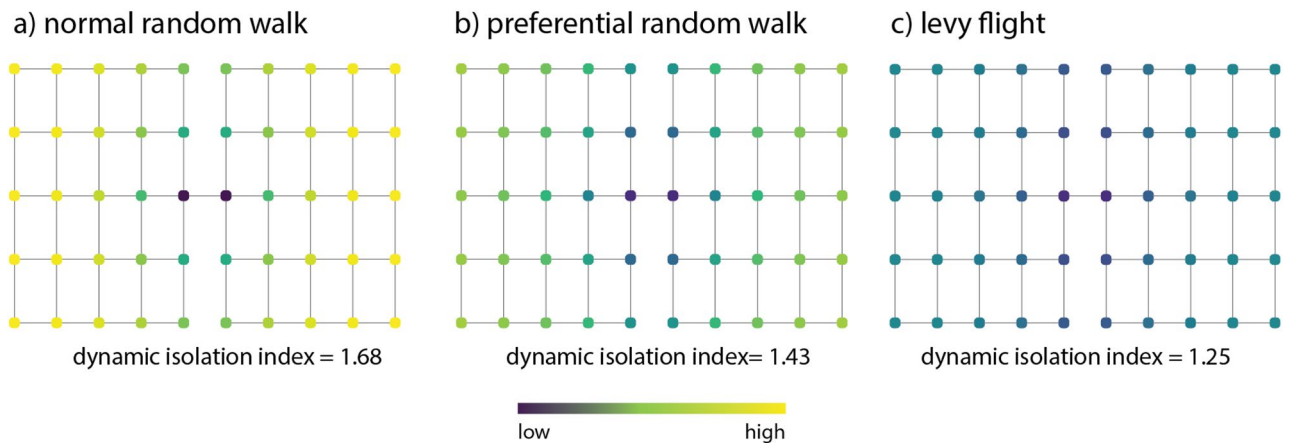
Segregation on network

Figure 3. Segregation measures on the toy model city for: (**a**) normal random walk with $\alpha = 0.85$, (**b**) preferential random walk with $\alpha = 0.85$ and $\beta = 1$, and c) Lévy flight with $\alpha = 0.85$ and $\beta = 2$.

Evaluating a transport network intervention strategy

We test the proposed model on the city of Cuenca, Ecuador to investigate how the introduction of a new mode of transport affects segregation of different socio-economic groups. Cuenca, with a population of 505,585 according to the 2010 census, is the third largest agglomeration in Ecuador after Guayaquil and Quito. It serves as a service and market centre for the southeastern region of the country[39], and the urban area extends over 72.48 Km^2, housing 331,885 people.

Previous research on urban development in Cuenca which studied the evolution of city's size and population density has emphasised the importance of planned densification policies to counteract the negative impacts of urban sprawl and create liveable urban areas[40]. Research on socio-spatial segregation in the city, which used measures of evenness, exposure, and clustering, has shown that there are two parallel processes of spatial segregation occurring within the urban area: segregation of low socio-economic status households towards the north and west periphery of the city, and self-segregation of high socio-economic status households in areas along the Tomebamba River[41].

In our study, we analyse the impact of both the bus network, and the tram which began its operation on the 25 May 2020, on Cuenca. We use the various measures described in the previous section to assess segregation in the city at a finer spatial scale by analysing population data at census block level and incorporating the spatio-temporal constraints that the various transport networks impose.

To achieve this, we first classify the population into four groups based on socio-economic data from the census. We then merge this data to the street network, and build three distinct multilayered networks which contained only the street network ($\mathcal{M}s$), the street network and bus network ($\mathcal{M}sb$), and finally the street network, bus network and tram network ($\mathcal{M}sbt$) respectively. Finally we apply the different types of random walk
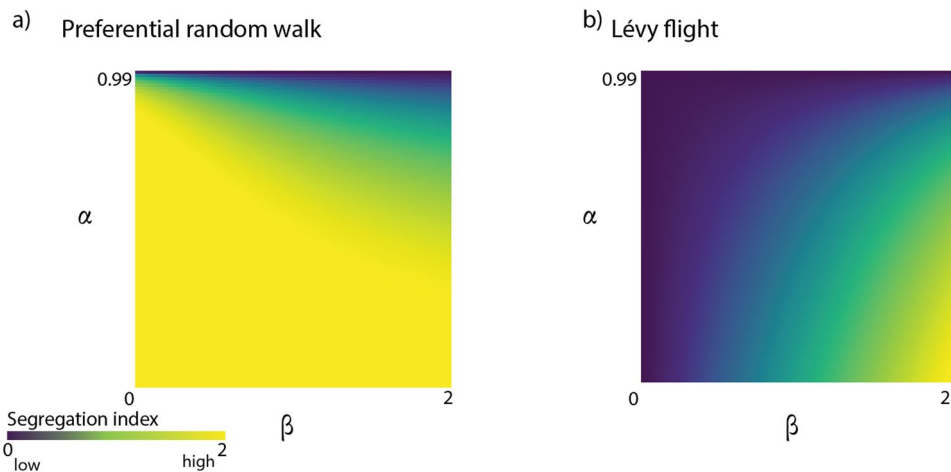
Figure 4. Sensitivity analysis for (**a**) local (preferential random walk) model and (**b**) Non-local (Lévy flight) model. In the case of local random walks the model is mainly driven by α—as it controls the spatial mobility of the population, with values closer to zero resulting in interactions that are constrained to direct neighbours. In the case of non-local random walks the model is mainly driven by β—with segregation being the lowest for values close 0 where there is equal probability of transitions between nodes regardless of distance.

segregation measures introduced previously to each network, in order to understand the effects that each has on the segregation of the different groups within the city.

To classify the population in Cuenca into four distinct socio-economic groups we first calculate an index of life conditions[41] at the household level. The index incorporates various factors such as the physical characteristics of dwellings, basic services of the household, education levels of residents, and access to health care. The index of life conditions (ICV) ranges from 0 to 2 where households with less than one express deprivation and above one present life conditions above standard. The ICV values are calculated at the dwelling level, and then assigned to all individuals who reside within the dwelling. Using this index, we create four groups by dividing the resultant data into quartiles and we classify each person within the city into one of these four groups. We then aggregate the population for each group at the urban block level, resulting in a distribution of the population at a fine-spatial scale. The spatial distribution of the mean values of the index of life conditions at the block level can be found in the Supplementary Material, and the distribution of each group within the city in Fig. 5.

Most individuals in the city have ICV values less than one, showing that they do not meet the minimum threshold for well-being across one or more variables. The ICV values across the population follow a normal distribution with a mean of 0.9. Given this distribution the population is divided into quartiles, $B = (Q1, Q2, Q3, Q4)$. Figure 5 shows the spatial distribution of concentration c_b of each group at the census block level. Quartiles Q1 and Q4 show clear spatial patterns, individuals with high ICV values locate along the Tomebamba River and towards the north-east while individuals with low ICV values locate towards the periphery, with some concentration within the city's historic center.

Socio-economic information is obtained from census data captured at the block level. The block geometries define the urban area that will be studied. Since these geometries are not connected, their nodes are extracted as points and an alpha shape[42] is used to determine the bounding polygon to define our study area. Different alpha values were tested to arrive at an optimal urban boundary for the case study. The resulting area is used to obtain the street network data, as well as set spatial limits on the other transport networks.

Transport network data is obtained from Open Street Map (OSM) and shapefiles provided by llactaLab—Sustainable Cities Research Group at the University of Cuenca. To construct the street network, osmnx[43] is used to download and construct the network. All street segments are included, except those that relate to private streets, emergency access, steps, cycleways, and paths. Since the street network has to be modelled as a walkable layer, street directionality is disregarded by adding additional reciprocal links to all oneway streets. In addition, the distance weight attribute of the links is turned to a time-weighted attribute by multiplying the distance of each link by an average walking speed equal to 5 Km/h.

To construct the bus and tram networks similar approaches are taken. For both networks the available data consists of two shapefiles, one containing the line geometries of the transport routes and another containing point geometries of the stops or stations within the city. Custom functions are developed to transform these shapefiles into directed multigraphs, which preserve information about their geometric properties.

The first step involved processing the line geometries so that all routes are represented as a single polyline. Since the point geometries, representing the stops or stations do not always match the polyline geometries and contain no information about which line a specific stop belongs to additional processing is needed to match stops/stations to their routes.

To address this, a 50 m buffer is set for each route. All points which fall within this buffer are aligned or 'snapped' to the corresponding route. Once the geometries of the routes and stops/stations are matched a directed multigraph for each route is created by cutting the line geometries by the points and creating the
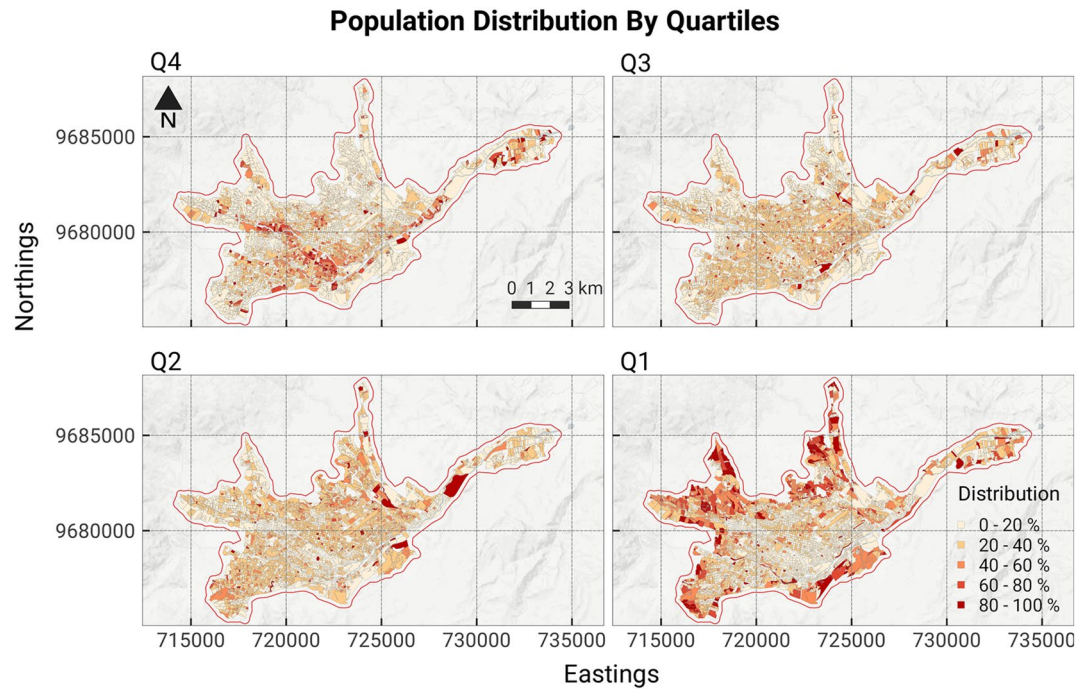
7

Population Distribution By Quartiles

Figure 5. Population distribution by ICV index quartiles, with Q4 representing the highest life conditions and Q1 the lowest. Q1 clusters are primarily in the city's north and west periphery, while Q4 clusters align with the Tomebamba river on a west-east axis. Q2 and Q3 groups show a more even distribution throughout the urban area.

corresponding nodes and links. Geometric properties are conserved for visualisation purposes, and a temporal weight is added to each link by calculating travel time using an average travel speed of 30 Km/h for the bus network and 40 Km/h for the tram network.

After creating a directed multigraph for each route for the bus and tram network, transfer links are created for routes within each that shared the same stop/station. These transfer links are weighted by an average waiting time of 10 min for the bus network and 5 min for the tram network. The resulting graphs are: (1) a strongly connected multidigraph for the street network, defined as $G_s = (N_s, L_s, w_s)$; (2) a strongly connected multidigraph for the bus network, defined as $G_b = (N_b, L_b, w_b)$; and (3) a strongly connected multidigraph for the tram network, defined as $G_t = (N_t, L_t, w_t)$. Each graph can be described by their adjacency and time-weighted adjacency matrix. The topological and geometric structure of these graphs are shown in the Supplementary Material.

Once each individual transport network is modelled as a graph, we follow the procedure described in the methodology section to create the different multilayered networks $\mathcal{M}s$, $\mathcal{M}sb$, $\mathcal{M}sbt$. For each multilayered network we assign the calculated socio-economic groups described previously to the nodes of the street network layer. To achieve this, we first create a Voronoi tessellation using the street intersection geometries. We then use a weighted area overlay interpolation to calculate population values for the Voronoi polygons. Finally, we assign the Voronoi polygon values to their corresponding node in the street network. This approach assumes that each individual living in a particular block will always start their journey from the same intersection, for the sake of simplicity. A representation of the multilayer network which captures all three transport modes is shown in Fig. 6. The properties of each individual transport network, as well as each resultant multilayered network can be seen in Table 1.

We analyse segregation given the population distribution through the $\mathcal{M}_s$, $\mathcal{M}_{sb}$ and $\mathcal{M}_{sbt}$, to capture the effects of the spatio-temporal constraints these networks impose, and how the bus network and the introduction of the tram affect segregation in the city. For all three networks a value of $\alpha = 0.85$ is used, which is equal to an expected weighted walk length of 20 min on the street network and 25 min on the multilayered networks. These values were chosen to match the mean travel time of the population as captured through travel surveys. To plot out the spatial distribution of the calculated segregation values, we aggregate the values calculated at each node of the street network to a hexagonal grid using H3 geospatial indexing system at level 9, which corresponds to hexagons that have side lengths of roughly 200 m and with areas of roughly 0.1 km^2, and then we take the mean. Additionally, we run the three types of random walks described previously and highlight how each type of random walk can reveal different aspects of how the transport system affects segregation.

Given that the groups Q1 and Q4 present the highest levels of spatial clustering, and have been found to have the highest levels of segregation in the city in previous studies[41], we focus on these two groups to explore the results of this study in more detail, while reporting overall values of segregation for all groups at the city level. As it will be shown, the addition of the tram does not significantly reduce segregation in the city—and change is mostly driven by considering the bus network. Because of this, we only plot the spatial distribution of change
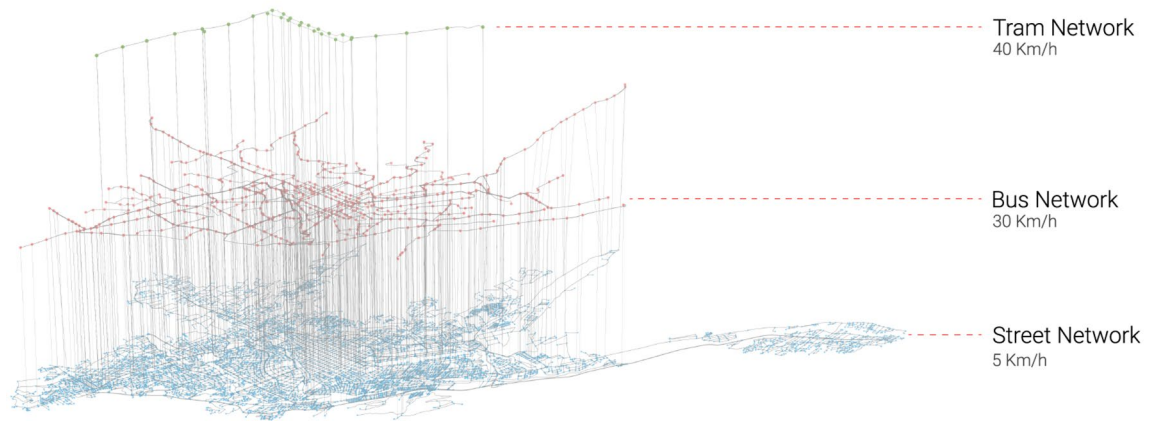
Multiplex Network

Figure 6. Multilayer network $\mathcal{M}^{sbt}$ of Cuenca showing the streets, bus, and tram networks as time-weighted graphs, along with their interlinks.

Network	N	L	$\bar{\ell}_{ij}$ Km	$\bar{\tau}_{ij}$ min	$\langle \ell \rangle$ Km	$\langle \tau \rangle$ min
G_s	8836	24,554	6.21	74.55	23.045	276.54
G_b	1090	15,848	4.79	31.97	18.20	71.74
G_t	41	73	4.68	18.28	11.2	42
$\mathcal{M}^{sb}$	11,016	44,762	5.61	38.31	22.23	170.72
$\mathcal{M}^{sbt}$	11,059	44,921	5.60	37.75	22.23	170.72

Table 1. Transport network measures of Cuenca outlining the number of nodes N and links L for each graph, as well as the average distance weighted shortest path $\bar{\ell}_{ij}$, average travel-time weighted shortest path $\bar{\tau}_{ij}$, distance weighted diameter $\langle \ell \rangle$ and travel-time weighted diameter $\langle \tau \rangle$.

in segregation caused by considering all transport networks, and report only the city wide results for all three multilayered network models.

Normal random walk
The normal random walk segregation measure is only affected by the topological structure and connectivity of the transport networks available. Figure 7 shows the resultant spatial distribution of the segregation values for groups Q1 and Q4 when considering all transport networks. Overall, the group with the higher values of
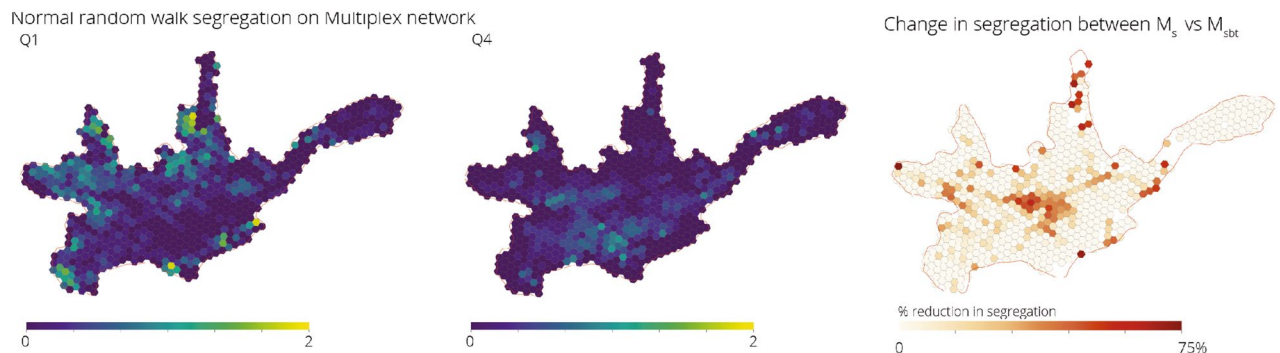

Figure 7. Spatial distribution of normalised segregation index for groups Q1 and Q4 in Cuenca measured using normal random walks on $\mathcal{M}^{sbt}$. Although the distribution on the Q4 population is concentrated along the east–west axis following the Tomebamba river, only those in the south–east present high levels of segregation due to a lack of transport connections to other parts of the city. Relative change in the segregation index caused by the introduction of the bus and tram network, measured using normal random walks, shows that the bus and tram reduce segregation mainly in the city centre and in specific areas in the periphery.

segregation, Q1, corresponds to the one with the lowest index of life conditions, and this is localised mostly in the western and northern periphery of the city. Q4 also presents some spatial clustering in the southern area of the city along one of the rivers. It is interesting to note that the segregation values that result from the normal random walk method in the street network yield similar results to the segregation values obtained through the relative size of the population groups in each census to the total size of the population group in the city. This might be due to the fact that temporal mobility of 20 min introduced through α is similar to the group interactions that one would expect at the census track level.

In Fig. 7 we can see the relative change in the values of the segregation index when taking into account all transport networks available, as opposed to only considering the street network. The central area of the city presents the highest decrease in segregation due to the increased connectivity to most other areas of the city through both the bus and tram network. Additionally, other areas such as specific regions in the north and west of the city also see significant reduction, and correspond to the areas in the periphery that have good coverage by the bus network.

Preferential random walk
We can also take into account the fact that people will tend to visit certain areas with more or less frequency depending on what those areas have to offer. In this case, we can use a preferential random walk that assigns probabilities of transition not only based on the connectivity and topological structure of the network, but also on the additional information about how attractive different places are, which can be quantified as an additional parameter q.

For this case study, we define q as the betweenness centrality of the nodes. The resultant segregation values for Q1 and Q4 can be seen in Fig. 8, they present similar spatial patterns of segregation to the normal random walk, with mostly a decrease of magnitude for all the areas. In this case however the reduction in segregation when considering all transport networks as opposed to only the street network, is much more pronounced, and it affects a much wider area in the city, as seen in Fig. 8. This is mainly driven by the fact that people will tend to visit the same areas regardless of their residence[44], increasing the probabilities that different groups end up in the same places. The areas which exhibit the highest change are areas that are well served by public transport, such as the city centre and along the linear corridors towards both the west and north of the city.

Lévy flights
Finally we use Lévy flights to better model how people move in urban areas based on how far or close different places are. In this case, the distance of the shortest paths between all nodes in the system are calculated and used to estimate a probability transition matrix P, this is then employed to calculate the probabilities of different individuals being present in the same area. Figure 9 shows the segregation values for Q1 and Q4 which exhibit very different spatial-patterns to both the normal and preferential random walks. Firstly, segregation values tend to be less extreme in all cases, with higher segregation values clustering near the centre and certain places in the west of the city for Q1, and towards the south east for Q4. The areas with the highest relative reduction in segregation when considering all transport modes as compared to the street network, are mostly concentrated in the west of the city as shown in Fig. 9.

Table 2 shows the normalised segregation index for each quartile group for the three networks as measured using normal random walks, preferential random walks, and Lévy flights. As mentioned in the previous section, values greater than 1 indicate that a group is over represented in a particular area. This means that the group is less exposed to other groups when considering the probabilities of two randomly chosen individuals encountering
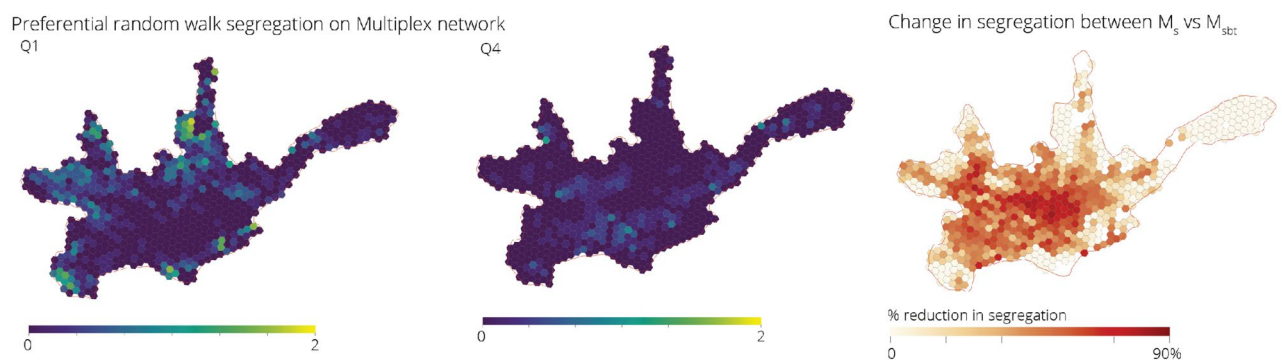

Figure 8. Spatial distribution of the normalised segregation index for groups Q1 and Q4 in Cuenca measured using preferential random walks on $\mathcal{M}^{sbt}$. Although there is an overall lower segregation due to transition probabilities being concentrated in few places, there are still high segregation values for Q1 in the north and south-west of the city due to poor connectivity of the transport network in these areas. Relative change in the segregation index caused by the introduction of the bus and tram network, measured using preferential random walks, shows that a higher overall reduction in segregation is evident throughout most of the city, with the exception of the eastern part of the city where there is a lack of street network connectivity and no additional transport connections.
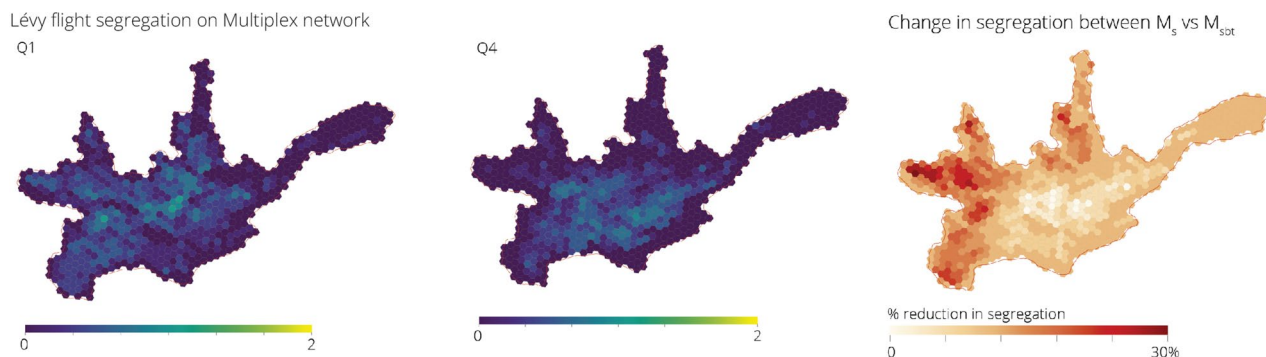
Figure 9. Spatial distribution of the normalised segregation index for groups Q1 and Q4 in Cuenca measured using Lévy flights on $\mathcal{M}^{sbt}$. Segregation values tend to be less extreme overall, with higher values generally coinciding with the spatial clustering of different groups. Relative change in the segregation index, caused by the introduction of the bus and tram network and measured using Lévy flights, shows that the tram and bus networks mostly affect the western part of the city by greatly reducing travel times towards the centre.

Network	Normal				Preferential				Lévy flights			
	Q1	Q2	Q3	Q4	Q1	Q2	Q3	Q4	Q1	Q2	Q3	Q4
$\mathcal{M}_s$	1.37	1.07	1.09	1.42	1.29	1.04	1.03	1.31	1.17	0.91	0.89	0.99
$\mathcal{M}_{sb}$	1.11	0.85	0.87	1.18	0.88	0.65	0.64	0.84	1.17	0.89	0.86	0.94
$\mathcal{M}_{sbt}$	1.10	0.84	0.87	1.17	0.87	0.64	0.63	0.83	1.15	0.86	0.81	0.88

Table 2. Segregation measures for Cuenca, Ecuador by ICV quartile for different network models and random walk types.

each other as they move through the city. The value is then zero when the probabilities match the make up of the population of the city.

If we only consider the road network using normal random walks we arrive at similar results as those of Orellana and Osorio[41], where groups Q1 and Q4 present normalised segregation values $\bar{\sigma} > 1$, showing that the process of socio-spatial segregation is more pronounced in the case of individuals with the lowest and highest ICV values. In all cases we see an overall decrease in segregation when adding the bus network, and a smaller decrease when adding the tram network, showing how increased connectivity can help integrate otherwise segregated parts of the city.

The normalised segregation in the multilayered network, $\mathcal{M}_{sb}$, shows no group over-representation. Considering the city's public transportation, $\bar{\sigma}$ across all quartiles displays no group isolation, meaning buses effectively improve mobility and access across diverse city areas despite physical constraints. However, trams don't enhance this effect possibly due to similar routes with buses, serving already integrated groups.

The results from the Lévy flight method indicate that segregation is overall lower for Q4 compared to both the normal walk and preferential walk case—with only Q1 showing higher segregation values. We hypothesise that this is because the areas where Q4 is located in the city are much more integrated and present higher connectivity through the transport network than those areas where group Q1 tends to cluster.

Through this analysis we showed that the introduction of the tram network did not have an important observed effect on increasing interactions between different groups in the city. However, the introduction of the tram presents an opportunity to restructure bus lines, reduce redundancies between these two systems, and possibly increase efficiency as well as decrease socio-economic segregation. The framework developed here, could hence provide useful information for such a restructuring, by providing a means of comparing the impact of different proposed interventions in the city.

Discussion

In this work, we presented a framework to measure segregation beyond the residential level, by estimating the interaction probabilities for different socio-economic groups considering the available transport networks. The different transport modes and mobility constraints were modelled as random walks in a multilayered network. By modelling these networks as multilayered systems through which random walks of various types can occur, we provide a nuanced understanding of how transport infrastructure influences segregation. Our approach highlights that network configuration can either facilitate or hinder socio-economic interactions, aiming to quantify the extent to which different transport modes support or impede the integration of diverse socio-economic groups.

The method introduced assesses the impact of new infrastructure and quantifies each transport network's contribution to the interaction opportunities. Our measure includes a parameter $\alpha \in [0, 1]$, that represents the temporal constraints in urban movement. For example, applying this method to a toy model, and conducting a

sensitivity analysis, we find that $\alpha = 0$ results in isolation index, whereas increasing α captures the steady-state random walk process and the city's socio-economic distribution.

In our empirical analysis, we studied the city of Cuenca, Ecuador, using block-level socio-economic census data. We showed that the measure is able to capture the influence of the network structure on segregation, where not only are the areas of the city that act like bridges between the two communities less segregated, but also the areas that are within a short network distance, revealing patterns of segregation otherwise hidden to traditional residential segregation measures. Finally, we showed how the measure can be used to assess the introduction of new transport infrastructure, by applying the method to evaluate the introduction of a new tram to the city. Our findings show that unless the current bus network is reorganised, the effects of the new tram on integrating different socio-economic groups by increasing probabilities of interactions across different parts of the city are negligible. While our multilayer model reveals the complex interplay between different transport modes and urban segregation, we also conducted a comparative analysis using only the street network layer. This analysis showed that relying solely on the street network oversimplifies the urban mobility landscape and fails to capture the significant contributions of other transport modes to segregation dynamics. The results indicate that the multilayer model provides a more accurate and holistic understanding of urban segregation, revealing patterns and relationships that a single-layer model cannot.

Our study underscores the significance of transportation networks in mitigating urban segregation, particularly emphasising the potential of targeted infrastructure interventions. The case of Cuenca, Ecuador, serves as an example, where our analysis offers specific insights for policymakers aiming to leverage transport infrastructure for socio-economic integration. Our findings suggest that the introduction of the tram system holds potential for enhancing connectivity across socio-economic divides. However, this potential can only be realised if accompanied by a strategic reorganisation of existing transport services, like the bus network, restructuring the bus routes that currently overlap with the tram network, prioritising the development of transport infrastructure in underserved areas to improve access to key resources and opportunities, thereby reducing spatial inequalities.

Beyond the specific context of Cuenca, our study offers generic recommendations for urban policymakers globally to use transport networks as tools for mitigating segregation: (1) conducting comprehensive analyses of cities' transport networks as multilayered systems to identify critical gaps and opportunities for enhancing socio-economic integration. (2) Leveraging detailed socio-economic and mobility data to inform the development and adjustment of transport services, ensuring they effectively address the mobility needs of diverse urban populations.

There are several limitations to our study that should be considered. First, our analysis is based on a single city, so it is not clear how well the results would generalise to other cities. Second, our measure of segregation is based on random and Lévy walks, which may not capture all the dynamics of segregation in real cities—specifically they don't capture interactions that might happen throughout a journey. Additionally transfer penalties, which refer to the time, cost, and inconvenience associated with changing modes or lines within a transport system are not fully considered, and can significantly affect the accessibility and desirability of public transport options, especially for marginalised or socio-economically disadvantaged groups. Finally, our analysis does not take into account other factors that may influence segregation, such as cultural preferences and the distribution of amenities and jobs that act as attractors to different places.

Despite these limitations, the segregation index proposed is a valuable tool for studying the effects of the transportation network on segregation. Overall, our study provides a new perspective on the dynamics of segregation in urban environments and offers a promising approach for measuring and analysing this complex phenomenon.

Data availibility

References

1. Comunian, R. Rethinking the creative city: The role of complexity, networks and interactions in the urban creative economy. *Urban Stud.* **48**, 1157–1179 (2011).
2. Barthélemy, M. *The Structure and Dynamics of Cities* (Cambridge University Press, 2016).
3. Batty, M. *The new science of cities* (MIT Press, 2013).
4. Pereira, R. H. M., Schwanen, T. & Banister, D. Distributive justice and equity in transportation. *Trans. Rev.* **37**, 170–191 (2017).
5. Scott, A. J. & Storper, M. The nature of cities: The scope and limits of urban theory. *Int. J. Urban Reg. Res.* **39**, 1–15 (2015).
6. Reardon, S. F. & Firebaugh, G. Measures of multigroup segregation. *Sociol. Methodol.* **32**, 33–67 (2002).
7. Massey, D. S. & Denton, N. A. The dimensions of residential segregation. *Soc. Forces* **67**, 281–315 (1988).
8. Lee, B. A. *et al.* Beyond the census tract: Patterns and determinants of racial segregation at multiple geographic scales. *Am. Sociol. Rev.* **73**, 766–791 (2008).
9. Louf, R. & Barthélemy, M. Patterns of residential segregation. *PloS One* **11**, e0157476 (2016).
10. Louail, T., Lenormand, M., Arias, J. M. & Ramasco, J. J. Crowdsourcing the Robin Hood effect in cities. *Appl. Netw. Sci.* **2**, 11 (2017).
11. Winship, C. A revaluation of indexes of residential segregation. *Soc. Forces* **55**, 1058–1066 (1977).
12. Bouchaud, J.-P. Crises and collective socio-economic phenomena: Simple models and challenges. *J. Stat. Phys.* **151**, 567–606. https://doi.org/10.1007/s10955-012-0687-3 (2013).
13. Weidlich, W. The statistical description of polarization phenomena in society. *Br. J. Math. Stat. Psychol.* **24**, 251–266 (1971).
14. Galam, S., Gefen, Y. & Shapir, Y. Sociophysics: A new approach of sociological collective behaviour. I. Mean-behaviour description of a strike. *J. Math. Sociol.* **9**, 1–13 (1982).
15. Schelling, T. C. Dynamic models of segregation. *J. Math. Soc.* **1**, 143–186 (1971).

16. Granovetter, M. & Soong, R. Threshold models of diffusion and collective behavior. *J. Math. Soc.* **9**, 165–179 (1983).
17. Morales, A. J., Dong, X., Bar-Yam, Y. & 'Sandy'Pentland, A. Segregation and polarization in urban areas. *R. Soc. Open Sci.* **6**, 190573 (2019).
18. Moro, E., Calacci, D., Dong, X. & Pentland, A. Mobility patterns are associated with experienced income segregation in large us cities. *Nat. Commun.* **12**, 4633 (2021).
19. Vallée, J. & Lenormand, M. Intersectional approach of everyday geography. *Environ. Plan. B Urban Anal. City Sci.* https://doi.org/10.1177/23998083231174025 *(2023)*.
20. Lazarsfeld, P. F. *et al.* Friendship as a social process: A substantive and methodological analysis. *Freedom Control Mod. Soc.* **18**, 18–66 (1954).
21. Newman, M. E. J. Mixing patterns in networks. *Phys. Rev. E* **67**, 26126 (2003).
22. Rodriguez-Moral, A. & Vorsatz, M. An overview of the measurement of segregation: Classical approaches and social network analysis. In *Complex Networks and Dynamics* (eds Commendatore, P. *et al.*) 93–119 (Springer, 2016).
23. Shalizi, C. R. & Thomas, A. C. Homophily and contagion are generically confounded in observational social network studies. *Sociol. Methods Res.* **40**, 211–239 (2011).
24. Jones, K., Johnston, R., Manley, D., Owen, D. & Charlton, C. Ethnic residential segregation: A multilevel, multigroup, multiscale approach exemplified by London in 2011. *Demography* **52**, 1995–2019 (2015).
25. Randon-Furling, J., Olteanu, M. & Lucquiaud, A. From urban segregation to spatial structure detection. *Environ. Plan. B Urban Anal. City Sci.* **47**, 645–661 (2020).
26. Olteanu, M., Randon-Furling, J. & Clark, W. A. Segregation through the multiscalar lens. *Proc. Natl. Acad. Sci.* **116**, 12250–12254 (2019).
27. Echenique, F. & Fryer, R. G. A measure of segregation based on social interactions. *Q. J. Econ.* **122**, 441–485 (2007).
28. Ballester, C. & Vorsatz, M. Random walk-based segregation measures. *Rev. Econ. Stat.* **96**, 383–401 (2014).
29. Sousa, S. & Nicosia, V. Quantifying ethnic segregation in cities through random walks. *Nat. Commun.* **13**, 1–12 (2022).
30. Dong, X. *et al.* Segregated interactions in urban and online space. *EPJ Data Sci.* **9**, 20 (2020).
31. Athey, S., Ferguson, B., Gentzkow, M. & Schmidt, T. Estimating experienced racial segregation in us cities using large-scale GPS data. *Proc. Natl. Acad. Sci.* **118**, e2026160118 (2021).
32. Nilforoshan, H. *et al.* Human mobility networks reveal increased segregation in large cities. *Nature* **624**, 1–7 (2023).
33. Riascos, A. P. & Mateos, J. L. Random walks on weighted networks: A survey of local and non-local dynamics. *J. Complex Netw.* **9**, cnab032 (2021).
34. Kivelä, M. *et al.* Multilayer networks. *J. Complex Netw.* **2**, 203–271 (2014).
35. Noh, J. D. & Rieger, H. Random walks on complex networks. *Phys. Rev. Lett.* **92**, 118701. https://doi.org/10.1103/PhysRevLett.92.118701 (2004).
36. Gallotti, R., Bazzani, A., Rambaldi, S. & Barthélemy, M. A stochastic model of randomly accelerated walkers for human mobility. *Nat. Commun.* **7**, 12600 (2016).
37. Alessandretti, L., Sapiezynski, P., Sekara, V., Lehmann, S. & Baronchelli, A. Evidence for a conserved quantity in human mobility. *Nat. Hum. Behav.* **2**, 485 (2018).
38. Gonzalez, M. C., Hidalgo, C. A. & Barabasi, A.-L. Understanding individual human mobility patterns. *Nature* **453**, 779–782 (2008).
39. Bolay, J.-C. & Rabinovich, A. Intermediate cities in Latin America risk and opportunities of coherent urban development. *Cities* **21**, 407–421 (2004).
40. Hermida, M., Hermida, C., Cabrera, N. & Calle, C. . La. densidad urbana como variable de an{á}lisis de la ciudad: El caso de Cuenca, Ecuador. *EURE (Santiago)* **41**, 25–44 (2015).
41. Orellana, D. *et al.* Segregación socio-espacial urbana en Cuenca, Ecuador. *Analitika J. Stat. Anal.* **8**, 27–38 (2014).
42. Edelsbrunner, H., Kirkpatrick, D. & Seidel, R. On the shape of a set of points in the plane. *IEEE Trans. Inf. Theory* **29**, 551–559 (1983).
43. Boeing, G. OSMnx: New methods for acquiring, constructing, analyzing, and visualizing complex street networks. *Comput. Environ. Urban Syst.* **65**, 126–139 (2017).
44. Alessandretti, L., Aslak, U. & Lehmann, S. The scales of human mobility. *Nature* **587**, 402–407 (2020).

Acknowledgements

Author contributions

M.N. developed the theoretical framework, designed the experiments and conducted the analysis. C.M., S.M., and E.A. contributed to the theoretical framework. M.N. wrote the paper, and all authors contributed to the structure and fine tuning of the manuscript.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary Information The online version contains supplementary material available at https://doi.org/10.1038/s41598-024-58932-9.

Correspondence and requests for materials should be addressed to M.N.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.