# How transport systems create opportunities for social interaction

**Ed Manley**
E.J.Manley@leeds.ac.uk
University of Leeds https://orcid.org/0000-0002-8904-0513
**Yitao Yang**
University of Leeds https://orcid.org/0000-0002-9650-9981
**Erjian Liu**
Beijing Jiaotong University
**Bin Jia**
Beijing Jiaotong University https://orcid.org/0000-0001-5071-2573

**Article**

**Keywords:** social interaction, mobility, transport infrastructure, mobility data

**Posted Date:** October 28th, 2025

**DOI:** https://doi.org/10.21203/rs.3.rs-6716648/v1



**Additional Declarations:** There is **NO** Competing Interest.

# How transport systems create opportunities for social interaction

Yitao Yang[1†], Erjian Liu[2†], Bin Jia[2], Ed Manley[1*]

[1*]School of Geography, University of Leeds, Leeds, UK.
[2]School of Systems Science, Beijing Jiaotong University, Beijing, 100044, China.

*Corresponding author(s). E-mail(s): e.j.manley@leeds.ac.uk;
Contributing authors: y.yang@leeds.ac.uk; 17120752@bjtu.edu.cn; bjia@bjtu.edu.cn;
†These authors contributed equally to this work.

**Abstract**

Social mixing in cities emerges from encounters between individuals of different backgrounds in shared spaces [1–3]. These opportunities for inter-group contact are not randomly distributed but shaped by urban transport systems that channel millions of daily trajectories. As cities face rising segregation, understanding and quantifying these opportunity structures has become critical for designing evidence-based policies that foster social inclusion [4–6]. Yet existing approaches face significant limitations in capturing the probabilistic nature of these opportunities within complex, multimodal cities [7]. This study addresses this gap by developing a computational framework that treats encounters as likelihoods shaped by behavioral uncertainty. Encounter probabilities among individuals are calculated within mode-specific encounter spaces—from individual roads, city blocks to rail and bus service segments—by aggregating potential trajectories across socioeconomic groups. Using city-scale mobility data, the outcomes reveal how infrastructure, daily rhythms, and travel choices interweave to create spatially and temporally varying opportunity structures across multimodal transport systems. We then extend these findings through agent-based simulations, demonstrating how transport policies designed to promote sustainable mobility may produce unintended social consequences. The study underscores that effective policymaking for social inclusion must account for how transport interventions reshape encounter opportunities in citizens' daily mobility.

**Keywords:** social interaction, mobility, transport infrastructure, mobility data

## Introduction

Socially cohesive cities thrive on interactions between diverse individuals, which foster mutual understanding and shared identity [8, 9]. The foundation for these interactions, however, is the simple, yet crucial, opportunity for people to be in the same place at the same time [6, 10, 11]. While not every instance of co-presence—whether a shared bus ride, or a crosswalk encounter—leads to meaningful interaction, it constitutes a necessary structural condition, what social and geographical scientists refer to as interaction potential[12] rooted in the principles of time geography [13] that govern interaction opportunities between pairs of people. In an era of increasing urban inequality, understanding how to create more opportunities for cross-group contact has become a critical priority for policymakers seeking to build inclusive cities [14, 15].

In urban environments, these opportunities for interaction are not randomly distributed; they are shaped by transportation infrastructure that channels millions of daily trajectories into predictable patterns of social mixing [16, 17]. Every journey, whether by foot, car, bus, or train, traces a trajectory through physical and social space, creating chances for contact with others whose paths intersect [18]. Transport systems do more than move people efficiently; they structure who is likely to share space with whom during daily mobility.

Despite growing recognition of mobility's role in shaping social life, current methods for mapping interaction opportunities across urban transport networks remain limited. Many computational studies using mobile phone data [4, 5, 19, 20], transit records [21–23], and social networks [24–26] typically infer encounters from spatiotemporal overlaps, assuming co-presence when two individuals' recorded positions coincide within distance and time thresholds. Such deterministic methods not only overestimate the interaction certainty when mobility records happen to overlap, but more fundamentally, underestimate the interaction potential in the unobserved intervals between recorded locations due to the sampling sparsity of real-world mobility data [27]. While other approaches—such as space-time prisms [28–30], Bayesian location models [31, 32], and entropy-based tie inference [33]—have been developed to address uncertainty in data and activity, they remain largely theoretical or confined to small-scale demonstrations with restrictive assumptions. Crucially, they fail to capture the reality of large, multimodal cities, where vertically layered transport infrastructures create mode-specific interaction potentials that cannot be measured by simple proximity [34, 35].

Here we introduce a novel probabilistic framework that quantifies cross-group encounter opportunities across multimodal transport systems at city scale. We conceptualize co-presence not as a binary event but as a probabilistic opportunity—a likelihood shaped by individuals' route choices, mode preferences, and behavioral uncertainty inherent in human mobility. We develop two core indices: the Probabilistic Mixing Index (*PMI*), which assesses the diversity of potential encounters using each transport mode (active, private, bus, railway), and the Multimodal Uniformity Index (*MUI*), which measures the consistency or divergence of these mixing opportunities across different mobility layers within the same geographic area. Applied to Beijing's metropolitan area using mobile phone data from 11 million users, our framework uncovers the hidden social opportunity structures for cross-group contact embedded within transport systems that vary across modes, times, and urban contexts. To translate the findings into actionable guidance, we further develop an agent-based model to explore how targeted policies can reshape these opportunity structures to foster more inclusive urban encounters.

# Results

## A probabilistic framework for measuring encounter opportunities

To quantify how urban transport systems create opportunities for cross-group encounters, we develop a probabilistic framework (Fig. 1a) using high-resolution mobile phone data from 11 million anonymous users in Beijing, China. We infer individuals' socioeconomic status by matching home locations to community property values from LianJia (China's largest real estate platform), categorizing users into four income quartiles (see Methods). Our approach conceptualizes an "encounter opportunity" not as a confirmed interaction, but as the probabilistic co-presence of individuals within specific urban environments, allowing us to account for the inherent uncertainty in mobility data and focus on the structural potential for interaction.

We define potential encounter spaces differently across transport modes: 1 km grids for diffuse modes like active travel and driving, and station-to-station segments for contained modes like bus and rail, where individuals share a vehicle or platform (Fig. 1b). Co-presence likelihood is assessed within 1-hour windows, balancing temporal resolution with computational feasibility (see Supplementary Note 2.3 for sensitivity analysis of spatiotemporal scales). By aggregating encounter probabilities across income groups, we compute the Probabilistic Mixing Index ($PMI_{s,m}$), a mode-specific measure that quantifies the diversity of potential encounters within a given spatial unit $s$ for mode $m$ (a higher *PMI*, approaching 1, signifies greater potential for cross-group encounters). To assess how these mixing opportunities vary across different transport layers in the same area, we introduce the Multimodal Uniformity Index ($MUI_A$), which measures the consistency of *PMI* values across all transport modes within the same geographic context (e.g., a specific region $A$) (see Methods).

## Divergent opportunity structures across mobility layers

Urban transport systems generate distinct patterns of mixing opportunities across Beijing's metropolitan area (Fig. 1c). Active mobility shows the most localized mixing patterns ($PMI_{\text{active}} = 0.4881$), with concentrated opportunities in mixed-use districts (e.g., business centers where diverse workers converge) but limited cross-group contact in homogeneous residential neighborhoods. Private vehicle travel shows intermediate mixing potential ($PMI_{\text{private}} = 0.7102$) along arterial roads where diverse traffic streams converge. Public transport reveals striking contrasts: railways create the most extensive opportunities ($PMI_{\text{railway}} = 0.9050$) by connecting disparate neighborhoods, except in suburbs, while buses sustain limited mixing ($PMI_{\text{bus}} = 0.7257$) due to localized routes serving similar demographics.

Encounter opportunities generally diminish peripherally (declining *PMI*; Fig. 1e), yet railway-scarce suburbs show unexpected mixing where shared bus dependency unites residents lacking alternatives. The *MUI* distributions (Fig. 1d,f) further reveal that central areas show high uniformity (high *MUI*), where different modes offer similarly high mixing opportunities, whereas peripheral zones exhibit highly divergent, mode-dependent

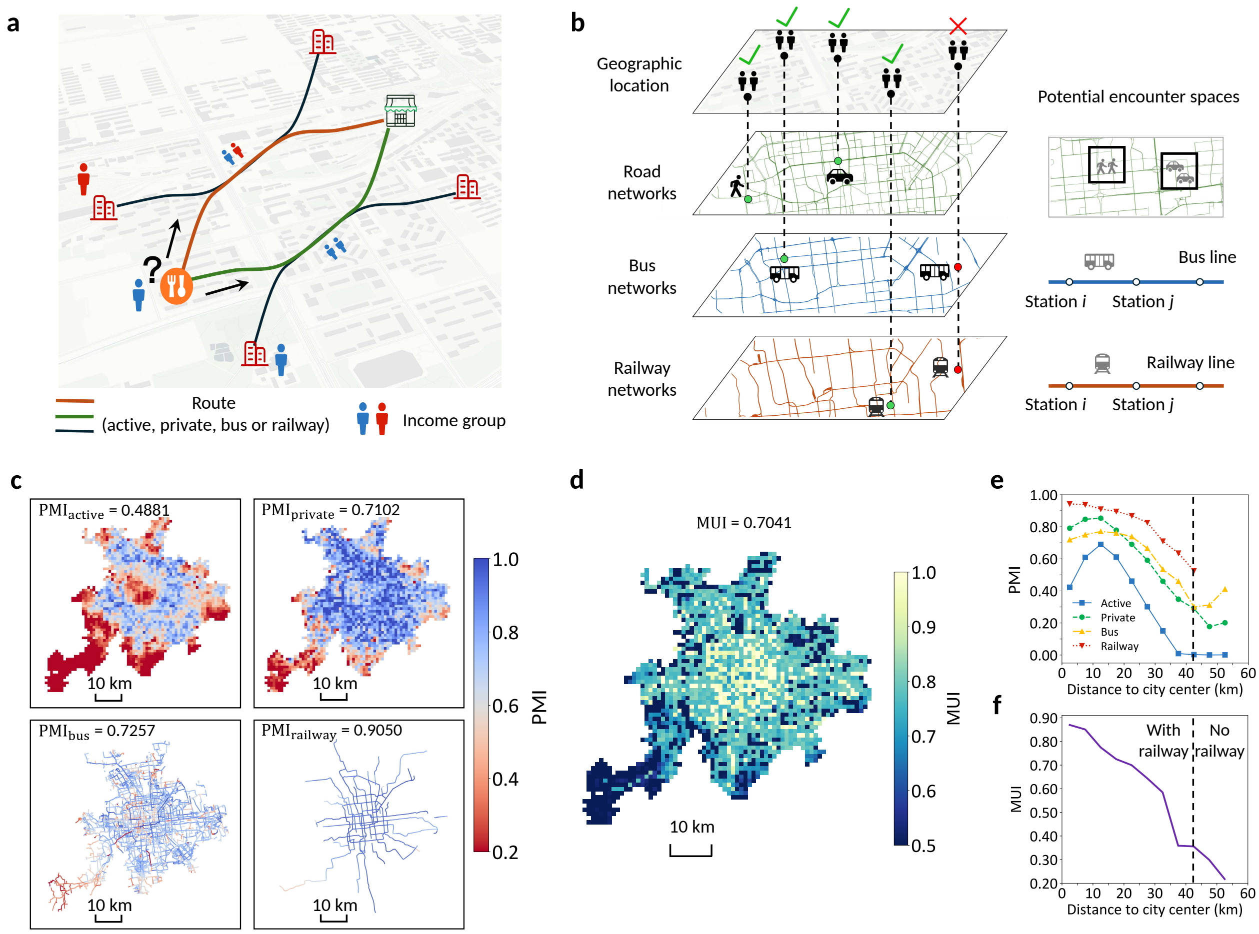


**Fig. 1 | Probabilistic framework quantifies encounter opportunities in multimodal transport systems.** **a**, Conceptual overview of the methodology. Individual trajectories from mobile phone data estimate probabilities of co-presence between pairs of individuals traveling via different transport modes (active, private, bus, railway). **b**, Mode-specific spatial units for calculating co-presence probabilities. Road-based encounter opportunities (active/private modes) are assessed in 1 km $\times$ 1 km grids representing potential roadside proximity. Public transport (bus/railway modes) opportunities are assessed along station-to-station segments, representing shared vehicle or platform environments. **c**, Spatial maps of the Probabilistic Mixing Index ($PMI$) for each transport mode across Beijing metropolitan area. Distinct patterns emerge depending on the mode, with mean citywide $PMI$ values noted above each map. **d**, Spatial distribution of Multimodal Uniformity Index ($MUI$), measuring consistency of $PMI$ values across the four transport modes in the same region (1 km $\times$ 1 km grids shown). **e**, Distance-dependent mixing patterns. Points show $PMI$ averages in 5 km concentric zones from the city center. Dashed line marks railway-deprived suburbs with elevated mixing opportunities. **f**, Average $MUI$ versus distance from the center, showing declining cross-modal uniformity toward the periphery.

opportunity structures. This highlights that transport infrastructure creates geographically varied landscapes of potential social interaction [36].

## Encounter opportunities unfolds with residents' daily rhythms

Temporal patterns across Beijing's downtown, peri-urban, and outskirts zones (Fig. 2a) show how routines influence mixing opportunities. A midday window (e.g., 11:00–13:00) boosts diversity across modes, with synchronized $PMI$ and $MUI$ peaks (Fig. 2b–e), driven by balanced, multidirectional flows (mean angular dispersion 90°, Fig. 2g-i; see Methods). This reflects diverse activities enhancing cross-group co-presence potential. In contrast, morning (06:00–08:00) and evening (post-18:00) commutes limit opportunities, as stratified flows converge downtown and disperse peripherally. Buses notably enhance mixing in peripheral areas, with higher relative $PMI$ from downtown to outskirts (Fig. 2b-d). These findings challenge conventional assumption that extensive public transit ensures equity [37], showing that encounter potential is dynamically shaped by the interplay of daily schedules, transport infrastructure, and urban spatial organization.

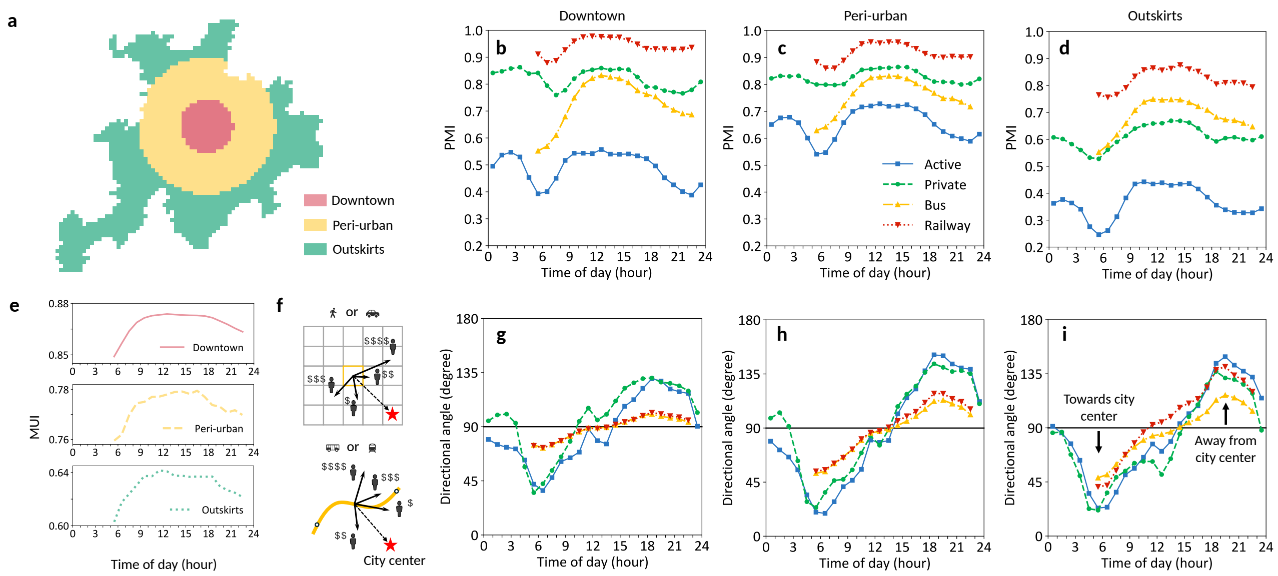


**Fig. 2 | Daily rhythms drive multimodal mixing opportunities across urban contexts. a**, Beijing metropolitan area partitioned into three zones—downtown, peri-urban, and outskirts—based on equal population distribution. **b-d**, Hourly variation in *PMI* shown for each transport mode within the three spatial contexts. Points indicate the average *PMI* for a given mode and context during each one-hour interval. **e**, Temporal fluctuations in the *MUI*. **f**, Schematic defining the mobility flow directional angle $\theta_s$ for spatial units (grids or transit segments), calculated relative to the city center (0° towards, 180° away). **g-i**, Hourly rhythms of the context-specific average directional angle. Midday periods exhibit more balanced, multidirectional travel (mean angular dispersion approaching 90°), correlating with enhanced mixing opportunities. In contrast, commuting hours show strongly directional flows (towards center in morning, away in evening) that limit cross-group co-presence.

## Transport infrastructure shapes opportunities for social interaction

To unravel how transport infrastructure unevenly shapes opportunities for social interaction, we use OLS regression models linking grid-level transport infrastructure features to mode-specific contact opportunities ($PMI$) and cross-modal uniformity ($MUI$) (see Methods; Supplementary Note 3). Dense infrastructure generally enhances interaction potential within respective modes (Fig. 3a; Supplementary Table 2): high motorways accessibility increase opportunities for private vehicle users ($\beta^{***}_{\mathrm{PMI,private}} > 0$), tertiary roads accessibility benefit active travelers ($\beta^{***}_{\mathrm{PMI,active}} > 0$), and subway stations boost railway encounters ($\beta^{***}_{\mathrm{PMI,railway}} > 0$). However, mode-exclusive infrastructure creates divergent opportunity structures: motorways decrease cross-modal uniformity ($\beta^{**}_{\mathrm{MUI}} < 0$) by enhancing private-mode encounters while excluding other users, whereas diverse road networks promote equitable opportunities across modes ($\beta^{***}_{\mathrm{MUI}} > 0$). Temporal analysis shows infrastructure's impact varies with daily rhythms (Fig. 3b). For example, tertiary roads maximize encounter potential during commute hours for road users but shift peak contribution to public transport midday, while weekend patterns support more equitable cross-modal opportunities. These findings demonstrate infrastructure's social impact depends not just on physical presence but on when and how diverse groups utilize it.

## Model simulation for policy interventions

To explore how policy can actively shape these opportunity structures, we developed an agent-based model that simulates morning commutes based on income-differentiated travel preferences (see Methods; Supplementary Note 4.1). The model, calibrated to match empirically observed encounter diversity $PMI$ (Fig. 4b; Supplementary Note 4.2), allows us to test the distributional effects of common transport policies (private car control [38, 39], public transport subsidies [40], active travel promotion [41, 42]).

Our simulations reveal that interventions often produce complex and counter-intuitive social outcomes. For instance, a uniform citywide increase in driving costs $\Delta\beta_{\mathrm{private}}$ (akin to fuel taxes [43] or parking fees [44]) disproportionately shifts lower-income users to other modes, paradoxically making the pool of remaining drivers less socioeconomically diverse ($PMI$ decreases; Fig. 4c) even as the policy has a progressive financial impact (Fig. 4d).

The spatial application of policy is also critical. A downtown congestion charge (mimicking congestion pricing [39]; Supplementary Note 4.3), for example, creates a different trade-off, enhancing encounter diversity among

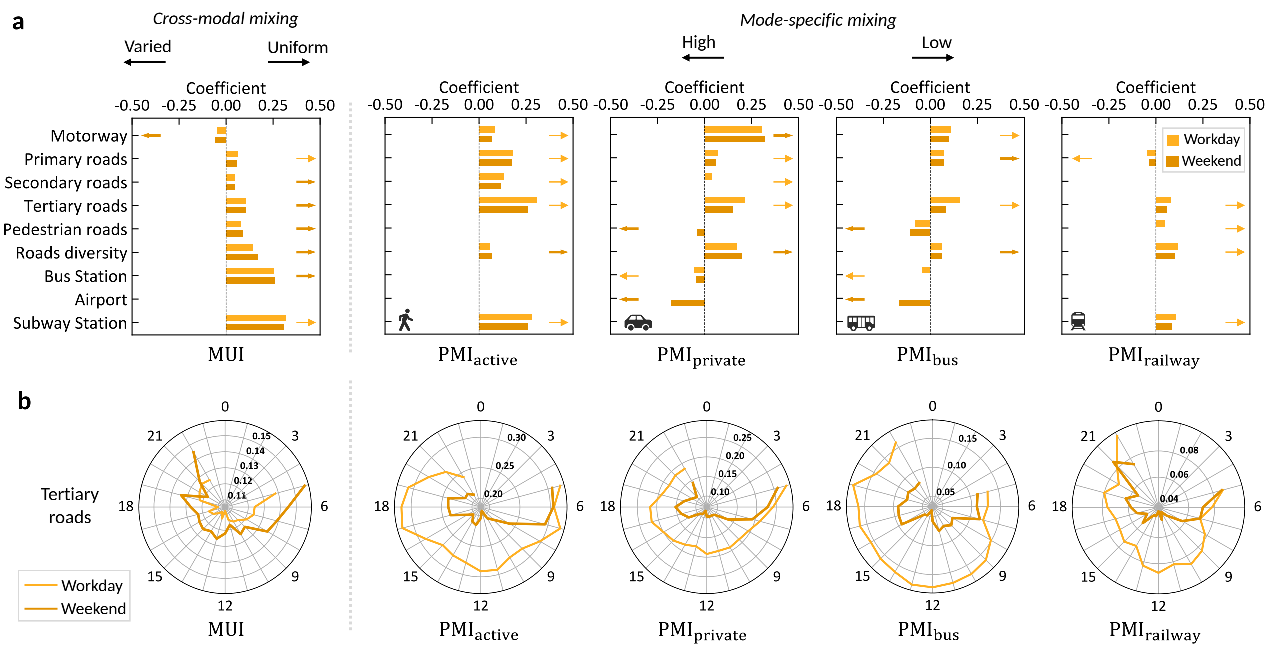


**Fig. 3 | OLS regression models explain encounter opportunities. a**, Contribution of transport infrastructure features on encounter diversity estimated using the daily granularity model. The leftmost column shows feature names. Each feature is associated with one or two bars, representing significant coefficients ($\beta > 0$) on workdays and weekends respectively. A positive coefficient ($\beta > 0$, rightward arrow) indicates the feature enhances co-presence potential (for *PMI*) or cross-modal uniformity (for *MUI*), while a negative coefficient ($\beta > 0$, leftward arrow) indicates reduced opportunities or more varied patterns among modes. Bold arrow denotes more pronounced weekend effects, non-bold marks stronger workday effects. **b**, Variation in the influence of selected feature (tertiary roads) throughout the day estimated using the hourly granularity model. In the compass plot, each point's angular position represents a specific hour (0:00 to 23:00), with numeric labels indicating the coefficient's value for that hour and day type.

drivers citywide but reducing it on other modes by altering specific commuter flows. Similarly, while public transport subsidies provide clear benefits to lower-income groups, they can inadvertently concentrate them onto rail, reducing interaction potential within that mode and showing diminishing returns (Supplementary Note 4.4). Promoting active travel (through infrastructure or safety improvements [42]) also yields spatially polarized results, fostering diverse encounters downtown but reinforcing socially sorted patterns in suburbs (Supplementary Note 4.5). These findings underscore a crucial insight for urban governance: transport policies designed for efficiency or environmental goals have profound, spatially-dependent impacts on the social fabric, presenting critical trade-offs between policy goals and social equity that demand integrated, context-aware planning.

# Discussion

Cities shape social life through the encounter opportunities they create or constrain[15, 36]. Our probabilistic framework reveals how transport systems structure these opportunities, moving beyond deterministic measures that assume precise spatiotemporal overlaps. By treating co-presence as likelihood rather than certainty, our metrics (*PMI* and *MUI*) capture the inherent uncertainty in human behavior and mobility data. Our framework reconceptualizes social dynamics not as fixed patterns but as probabilistic opportunities emerging from the interplay of infrastructure, mobility choices, and daily rhythms.

Our findings challenge assumptions that physical proximity automatically fosters social cohesion [45–47]. Railway systems create extensive encounter opportunities yet remain temporally constrained; buses in peripheries paradoxically enhance cross-group contact through shared dependency; motorways generate mode-based stratification despite improving connectivity. These patterns demonstrate that infrastructure's social impact depends critically on who uses it, when, and under what constraints—aligning with time-geography principles [48] while extending them to multimodal contexts. These findings compel infrastructure-centric urban design [49, 50] to move beyond simple expansion and instead embrace strategies that coordinate infrastructure, policy, and operations with urban rhythms, a perspective central to the concept of chrono-urbanism [51].

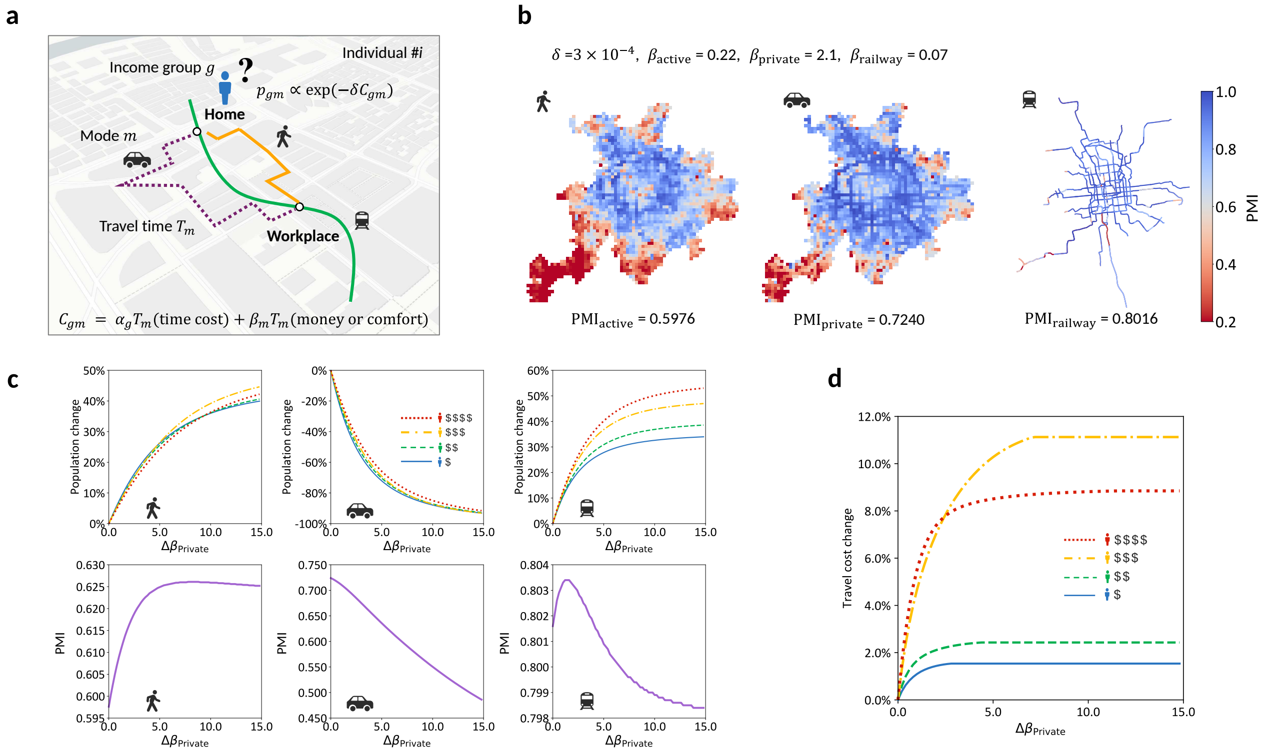


**Fig. 4 | Agent-based model simulates policy impacts on encounter opportunities.** **a**, Schematic of the agent-based mobility model. Individuals choose between active, private car, or railway modes probabilistically to minimize perceived travel cost, which incorporates travel time ($T_m$), income-specific value of time ($\alpha_g$), and mode-specific costs ($\beta_m$). **b**, Spatial distribution of model-predicted *PMI* for active, private, and railway travel during the morning commuting hour. The optimal model parameters calibrated using empirical data are displayed above the maps. **c**, Simulated impact of a uniform citywide increase in private car cost ($\Delta\beta_{\text{private}}$). Top row: Proportional change in mode usage relative to baseline by income groups. Bottom row: Change in overall citywide encounter potential (*PMI*) for each mode. **d**, Impact of the uniform private car policy on average travel costs relative to baseline by income group, showing a generally progressive burden.

Policy simulations reveal tensions between transport objectives and social goals. Interventions optimizing efficiency or sustainability may inadvertently reduce encounter opportunities, highlighting that transport policies reshape cities' social opportunity landscapes in complex, often unexpected ways. These outcomes underscore that transport policies cannot be evaluated solely on efficiency or environmental grounds—their impacts on the social opportunity structure demand equal consideration.

Finally, while our Beijing case provides detailed insights, the framework's probabilistic approach and focus on structural opportunities offers a transferable methodology for examining how different urban contexts create or constrain possibilities for cross-group contact. Future comparative research across cities with different transport legacies, urban forms, and social compositions will be crucial to test the universality of these patterns. Such work can help build a more comprehensive theory of how urban mobility systems shape the social fabric of our cities, paving the way for the design of environments that are not only connected, but also cohesive.

# Methods

## Datasets

**Mobile phone data.** The anonymous mobile phone dataset come from a telecommunications service provider in China. The dataset was collected over one-month period (June) in 2023, and consists of anonymized records of GPS locations ("pings") from users that have signed up to provide data through explicit consent agreements outlined in the telecoms company's privacy notice. The mobile phone users were informed about how their data will be used and stored under the regulation of China's Personal Information Protection Law (PIPL). In alignment with China's data protection and cybersecurity regulations, all data are thoroughly anonymized to remove personally identifiable information and processed to ensure privacy. This work was exempt from the ethical review by the Business, Environment, Social Sciences Faculty Research Ethics Committee in the

University of Leeds (Reference: BESS+ FREC - 2024 1663-2108). The raw data consist of 11,018,253 users (covering 50% population) and 8,386,564,428 pings in Beijing, China. Each ping consists of a de-identified user ID, latitude, longitude and timestamp. The mean number of raw pings associated with a user in a month is 1008 and the median number of pings is 872. To ensure data reliability, we removed duplicate pings to ensure accuracy. We filtered out users with fewer than 300 pings to eliminate noise. The filtered dataset consists of 7,562,482 users and 4,815,969,539 pings. The population representativeness of the dataset is validated in Supplementary Note 1.2.

**LianJia housing.** The LianJia housing dataset [52], sourced from China's largest real-time property transaction platform, offers a detailed repository of residential property information in Beijing. As of June 2023, this dataset encompasses 9,501 communities, covering 97.3% of Beijing's residential market. This dataset provides granular metrics including average transaction prices (RMB/$m^2$), architectural typologies, household counts, and precise geographic coordinates. As China's residential neighborhoods are predominantly planned, gated communities with homogeneous pricing and shared amenities, the dataset captures near-complete market dynamics, enabling robust socioeconomic inference.

**Transport facilities.** We collected a dataset of 64,805 transport facilities in Beijing, China, using Application Programming Interface (API) of Amap [53], a leading mapping and navigation service provider in China. Each record includes name, address, latitude, longitude and category. The categories of transport facilities encompassed in this dataset include: Airport, Train Station, Port, Intercity Bus Station, Subway Station, Bus Station, Parking Lot, Toll Station, and Highway Service.

**Transport networks.** We retrieved the urban road networks from OpenStreetMap [54], including both driving roads and pedestrian pathways. Urban road networks in Beijing consist of 307,978 junctions and 437,337 segments. We used Amap API [53] to obtain public transport networks, encompassing bus and railway routes. The API generates 4,238 directional bus routes (15,876 stations and 44,370 segments) and 118 directional railway routes (432 stations and 1,028 segments) in Beijing. Each record includes the route names, station names, station coordinates (latitude and longitude), average travel duration between stations, and route operational hours.

**Geolife.** The Geolife GPS Trajectory dataset [55], a widely-used public benchmark for travel mode inference, released by Microsoft Research Asia comprises 17,621 trajectories collected from 182 volunteer users between April 2007 and August 2012. It includes raw GPS points (latitude, longitude, timestamp) and labels for transport modes. We used a processed subset of this dataset, totaling 4,425 trips with unambiguous labels (1,819 active, 881 private, 1,725 public) for training and validating our travel mode inference model (Supplementary Note 1.4) and validating route generation accuracy (Supplementary Note 1.5).

**MemDA.** The MemDA (Memory-Augmented Deep Architecture) dataset [56] is an open-sourced dataset providing traffic speeds on major road segments in Beijing over a 75-day period (May 12 - July 25, 2022). This dataset serves as an independent ground truth to validate the reliability of our inferred mobility patterns, particularly the travel mode inference and route generation processes (Supplementary Note 1.6).

## Inferring home locations and workplaces

We first detected significant stays during individuals' trips through a spatiotemporal clustering approach [57]. These stays were defined by high-density clusters of trajectory points within a 50-meter radius and a minimum of 10 points, subsequently linked to nearby POIs within 100 meters for contextual validation. Stays shorter than 15 minutes or longer than 24 hours were excluded to ensure focus on meaningful activities. Home locations were determined by analyzing stays during nighttime hours (21:00–6:00) across multiple days, selecting the location with the highest cumulative duration and at least 25 visits over the 30-day period, situated in residential areas as confirmed by Amap POIs. Workplaces were identified from stays during typical working hours (9:00–17:00) on workdays, requiring a minimum of 4 visits per 5 workdays, cross-referenced with commercial POIs. This method reliably assigned home/work locations to 604,7368 individuals. The inference reliability was validated by comparing home-based trip frequencies with China's National Population Census in 2020 [58], adjusted for population growth (see Supplementary Note 1.1 and 1.2).

## Inferring income quartiles

To estimate socioeconomic status, we leveraged China's distinct residential structure—characterized by uniformly priced, gated communities with shared amenities—which enables housing expenditure to serve as robust proxies for income stratification. Individuals' geolocated homes were systematically matched to property transaction data from LianJia, China's largest real estate platform ($n$ = 9,501 Beijing communities). We constructed Voronoi polygons around each community to delineate localized socioeconomic zones, ensuring spatial contiguity while minimizing edge effects through nearest-neighbor allocation. Individuals within a polygon were assigned corresponding community's average transaction price as a proxy for socioeconomic status. This method capitalizes on the inherent homogeneity of Chinese residential communities where housing prices strongly correlate with residents' economic capacity. Validation confirms 80% of inferred home locations resided within 250 meters (90% within 500 meters) of matched communities, ensuring spatial reliability. Individuals were categorized into four income quartiles based on 25th, 50th and 75th percentiles of community price distributions. Spatial analysis reproduces the distinct residential segregation patterns in Beijing, reinforcing the validity of our approach (see Supplementary Note 1.3).

## Quantifying encounter opportunities

To quantify the opportunities for cross-group encounters shaped by multimodal mobility, we developed two entropy-based measures grounded in probabilistic co-location patterns. The foundation is the calculation of the likelihood that individuals are simultaneously present at specific locations along their journeys. This involves first estimating the probability of individuals choosing specific transport modes—active (walking/cycling), private (car/taxi), bus, or railway—for each trip using a machine learning model pre-trained on the Geolife GPS trajectory dataset (see Supplementary Note 1.4). The most probable route for each potential mode is then generated using the Amap navigation API, incorporating real-time traffic, transit schedules, and original GPS waypoints to enhance realism (see Supplementary Note 1.5). The likelihood of any two individuals encountering each other via the same mode $m$ within a specific spatial unit $s$ is calculated as the product of the probabilities that their routes intersect within that unit and their estimated travel times overlap. The reliability of these co-location likelihood estimates was confirmed via cross-validation using MemDA traffic speed data (see Supplementary Note 1.6). Aggregating these encounter probabilities across all individuals belonging to the four inferred income quartiles ($q$), we estimated the expected population mix in each spatial unit $pop_{qsm} = \sum_{i \in q} p_{ism}$, where $p_{ism}$ is the probability that individual $i$ from group $q$ occupies spatial unit $s$ via mode $m$ during a specific time window. Our first measure, the mode-specific Probabilistic Mixing Index ($PMI$), quantifies the income group diversity within this expected population mix using a normalized entropy metric [59]:

$$PMI_{s,m} = -\frac{1}{\log(4)} \sum_{q} \tau_{qsm} \log\left(\tau_{qsm}\right), \tag{1}$$

where $\tau_{qsm} = pop_{qsm} / \sum_q pop_{qsm}$. $PMI_{s,m}$ spans from 0 (indicating highly sorted co-presence, dominated by a single group) to 1 (representing maximum mixing potential with equal group representation).

Our second measure, the Multimodal Uniformity Index ($MUI$), evaluates how consistently these encounter opportunities are distributed across the different transport modes within a specific geographic area $A$. It is calculated by computing the normalized entropy [59] of the mode-averaged $PMI$ values within that region:

$$MUI_A = -\frac{1}{\log(4)} \sum_{m} r_{A,m} \log\left(r_{A,m}\right), \tag{2}$$

where $r_{A,m} = PMI_{A,m} / \sum_m PMI_{A,m}$. Here, $PMI_{A,m}$ represents the average mixing level across all the mode-specific spatial units located in region $A$. A higher $MUI_A$ value (close to 1) indicates that encounter opportunities are equitable across modes within region $A$. Conversely, a lower value (closer to 0) signals that the potential for social encounters is highly stratified by mode choice. Together, these measures provide a framework for evaluating how urban transport systems function as social opportunity structures, shaping the likelihood of socioeconomic encounters in cities.

## Mobility flow directionality

We adopt a vector-based approach [60] to capture the spatial orientation of population movements across urban contexts. The city center is first defined as the point of highest population density within the Beijing metropolitan area. For each spatial unit (1 km × 1 km grid for active and private modes; transit segment for bus and railway modes), mobility flow vector for each groups is calculated based on trip origin-destination pairs recorded over hourly intervals. The vector magnitude represents the total number of trips weighted by the probabilities of individuals from the specific group being present in that spatial unit during the given hour. Next, the group-specific vectors within each spatial unit $s$ are summed into a composite mobility flow vector $\vec{V}_s$. The directional angle of this composite vector is then computed as the angle between the vector and the reference line connecting the unit's centroid to the city center: $\theta_s = \arccos\left(\frac{\vec{V}_s \cdot \vec{e}_c}{\|\vec{V}_s\|}\right)$, where $\vec{e}_c$ is the unit vector pointing from $s$ to city center. Angles measured in radians are converted to degrees on a 0°–180° scale (0°: strictly toward center, 180°: strictly outward). For each urban context $\mathcal{C}$ (downtown, peri-urban, or outskirts), the mean directional angle $\bar{\theta}_{\mathcal{C}}$ in a given hour is computed as the arithmetic average of directional angles from all spatial units within the context: $\bar{\theta}_{\mathcal{C}} = \frac{1}{N}\sum_{s\in\mathcal{C}} \theta_s$, where $N$ denotes the number of spatial units.

## OLS regression models

We employ ordinary least squares (OLS) regression models [61] to quantify how transport infrastructure shapes the spatiotemporal distribution of encounter opportunities. These models operate at the 1 × 1 km grid level, linking grid-specific indices of social mixing to transport-related explanatory variables $\{T_i\}$. Specifically, either the Probabilistic Mixing Index for a given mode $m$ ($PMI_{A,t,m}$) or the Multimodal Uniformity Index ($MUI_{A,t}$) within grid $A$ at time $t$ (collectively denoted $M_t$) is modeled as a function of these transport infrastructure features. The features $\{T_i\}$ encompass characteristics such as the lengths of different road types, road type diversity, and counts of transport facilities. A detailed list and description of these grid-level variables are provided in Supplementary Table 1. To capture fine-grained temporal dynamics, we fitted separate models for each hour of the day, distinguishing between workdays and weekends (see Supplementary Note 3 for further details on the model specification).

## Individual mobility model

We develop an agent-based model to simulate how policy interventions can reshape travel choices and, consequently, the city's structure of social encounter opportunities. For a given home-to-work trip, an individual $i$ from income group $g \in \mathcal{G}$ ($|\mathcal{G}| = 4$) probabilistically chooses a mode $m \in \mathcal{M} = \{\text{active, private, railway}\}$ to minimize a perceived travel cost $\mathcal{C}_{gm}$ using a multinomial logit formulation [62]:

$$p_{gm} = \frac{\exp(-\delta C_{gm})}{\sum_{m' \in \mathcal{M}} \exp(-\delta C_{gm'})}, \tag{3}$$

where $\delta$ is a sensitivity parameter. The perceived cost $\mathcal{C}_{gm}$ combines the income-specific value of time ($\alpha_g$) with other mode-specific costs ($\beta_m$), such as monetary expenses or inconvenience, applied to the travel time $T_m$ (computed as the duration of the shortest path $\mathbf{R}_m$ for mode $m$):

$$\mathcal{C}_{gm} = \alpha_g T_m + \beta_m T_m. \tag{4}$$

We use the model to simulate the travel preferences of individuals during the morning peak hour (9:00–10:00 AM). The model's outputs, including individual choice probabilities $p_{gm}$ and shortest paths $\mathbf{R}_m$, are used as inputs for calculating our mode-specific measure of encounter diversity ($PMI$). Model parameters are calibrated using empirical data: relative time values ($\alpha_g$) are fixed based on group incomes, while the sensitivity ($\delta$) and mode-specific costs ($\beta_m$) are estimated by minimizing the difference between the model's predicted $PMI$ and the empirically observed values during the peak hour. Detailed calibration procedures and performance validation are provided in Supplementary Note 4.2.

We then utilize the calibrated model to evaluate the distributional effects of three common transport policy archetypes. These policies are simulated by systematically modifying the relevant mode-specific cost parameters ($\beta_m$). For each policy scenario, we recalculate individual mode choices and analyze the resulting changes in encounter diversity ($PMI$), mode shares, and average travel costs across income groups to assess the trade-offs between policy goals and social equity. Detailed simulation setups and results are provided in Supplementary Notes 4.3–4.5.

# Data availability

The source data supporting the findings of this study are available online (https://github.com/UrbanMobility-y/multimodal-segregation/tree/main/Source%20data). The raw mobile phone mobility data are not publicly available to preserve individual privacy and user confidentiality. The datasets of transport facilities and public transport networks are commercially available and may be requested for research use (https://lbs.amap.com/). Road networks were from OpenStreetMap (https://www.openstreetmap.org/). LianJia housing transaction data (https://bj.lianjia.com/xiaoqu/), China National Population Census 2020 data (https://www.stats.gov.cn/english/PressRelease/202105/t20210510_1817185.html), Geolife trajectory data (https://www.microsoft.com/en-us/download/details.aspx?id=52367) and MemDA traffic speed data (https://github.com/deepkashiwa20/Urban_Concept_Drift) are available online.

# Code availability

All analysis was conducted using Python. Code is publicly available at GitHub (https://github.com/UrbanMobility-y/multimodal-segregation).

# Acknowledgments

We thank our funders who supported this research - the Leverhulme Trust via Philip Leverhulme Prize funding (Ref: PLP-2022-262) and the National Natural Science Foundation of China (Refs: 72288101, 72242102).

# Author contributions

Y. Y. and E. M. conceived the study, overall research design, and developed the probabilistic framework. Y. Y. and E. L. performed data processing, implemented the software, ran the formal analyses and simulations, prepared the figures, and drafted the manuscript. B. J. curated the mobility data and provided transport-systems expertise in result interpretation. E. M. supervised the project, provided critical feedback, and reviewed and edited the manuscript. All authors discussed the results and approved the final version of the manuscript.

# Competing interests

The authors declare no competing interests.

# References

[1] de la Prada, a. G. & Small, M. L. How people are exposed to neighborhoods racially different from their own. *Proceedings of the National Academy of Sciences of the United States of America* **121** (2024).

[2] Cagney, K., York Cornwell, E., Goldman, A. & Cai, L. Urban mobility and activity space. *Annual Review of Sociology* **46**, 623–648 (2020).

[3] Phillips, N., Levy, B., Sampson, R., Small, M. & Wang, R. The social integration of American cities: Network measures of connectedness based on everyday mobility across neighborhoods. *Sociological Methods and Research* **50**, 1110–1149 (2021).

[4] Nilforoshan, H. *et al.* Human mobility networks reveal increased segregation in large cities. *Nature* **624**, 586 (2023).

[5] Moro, E., Calacci, D., Dong, X. & Pentland, A. Mobility patterns are associated with experienced income segregation in large us cities. *Nature Communications* **12** (2021).

[6] Farber, S., Neutens, T., Miller, H. J. & Li, X. The social interaction potential of metropolitan regions: A time-geographic measurement approach using joint accessibility. *Annals of the Association of American Geographers* **103**, 483–504 (2013).

[7] Liao, Y., Gil, J., Yeh, S., Pereira, R. H. & Alessandretti, L. Socio-spatial segregation and human mobility: A review of empirical evidence. *Computers, Environment and Urban Systems* **117**, 102250 (2025).

[8] Jacobs, J. *The death and life of great american cities* (Random House, New York City, 1961).

[9] Whyte, W. H. *The social life of small urban spaces* (Project for Public Spaces, New York City, United States, 1980).

[10] Hägerstrand, T. What about people in regional science? *Papers of the Regional Science Association* **24**, 6–21 (1970).

[11] Farber, S., Neutens, T., Carrasco, J.-A. & Rojas, C. Social interaction potential and the spatial distribution of face-to-face social interactions. *Environment and Planning B: Planning and Design* **41**, 960–976 (2014).

[12] Miller, H. J. A measurement theory for time geography. *Geographical Analysis* **37**, 17–45 (2005).

[13] Hägerstraand, T. What about people in regional science? *Papers in Regional Science* **24**, 7–21 (1970).

[14] Farber, S. & Li, X. Urban sprawl and social interaction potential: an empirical analysis of large metropolitan regions in the united states. *Journal of Transport Geography* **31**, 267–277 (2013).

[15] Farber, S., Neutens, T., Miller, H. J. & Li, X. The social interaction potential of metropolitan regions: A time-geographic measurement approach using joint accessibility. *Annals of the Association of American Geographers* **103**, 483–504 (2013).

[16] Xian, S., Qi, Z. & ming Yip, N. Beyond home neighborhood: Mobility, activity and temporal variation of socio-spatial segregation. *Journal of Transport Geography* **99**, 103304 (2022).

[17] Aiello, L. M., Vybornova, A., Juhász, S., Szell, M. & Bokányi, E. Urban highways are barriers to social ties. *Proceedings of the National Academy of Sciences* **122**, e2408937122 (2025).

[18] Loo, B. P. Y., Fan, Z. & Moro, E. Residential and experienced social segregation: the roles of different transport modes, metro extensions, and longitudinal changes in hong kong. *Humanities & Social Sciences Communications* **11** (2024).

[19] Athey, S., Ferguson, B., Gentzkow, M. & Schmidt, T. Estimating experienced racial segregation in us cities using large-scale gps data. *Proceedings of the National Academy of Sciences* **118**, e2026160118 (2021).

[20] Fan, Z. *et al.* Diversity beyond density: Experienced social mixing of urban streets. *PNAS Nexus* **2**, pgad077 (2023).

[21] Abbasi, S., Ko, J. & Min, J. Measuring destination-based segregation through mobility patterns: Application of transport card data. *Journal of Transport Geography* **92**, 103025 (2021).

[22] Gao, Q.-L. *et al.* Revealing transport inequality from an activity space perspective: A study based on human mobility data. *Cities* **131**, 104036 (2022).

[23] Kolkowski, L. *et al.* Measuring activity-based social segregation using public transport smart card data. *Journal of Transport Geography* **110**, 103642 (2023).

[24] Wang, Q., Phillips, N. E., Small, M. L. & Sampson, R. J. Urban mobility and neighborhood isolation in america's 50 largest cities. *Proceedings of the National Academy of Sciences* **115**, 7735–7740 (2018).

[25] Tóth, G. *et al.* Inequality is rising where social network segregation interacts with urban topology. *Nature Communications* **12**, 1143 (2021).

[26] Candipan, J., Phillips, N. E., Sampson, R. J. & Small, M. From residence to movement: The nature of racial segregation in everyday urban mobility. *Urban Studies* **58**, 3095–3117 (2021).

[27] Barreras, F. & Watts, D. J. The exciting potential and daunting challenge of using GPS human-mobility data for epidemic modeling. *Nature Computational Science* **4**, 398–411 (2024).

[28] Song, Y. & Miller, H. J. Simulating visit probability distributions within planar space-time prisms. *International Journal of Geographical Information Science* **28**, 104–125 (2014).

[29] Elias, D. & Kuijpers, B. Visit probability in space–time prisms based on binomial random walk. *ISPRS International Journal of Geo-Information* **9** (2020).

[30] Winter, S. & Yin, Z.-C. Directed movements in probabilistic time geography. *International Journal of Geographical Information Science* **24**, 1349–1365 (2010).

[31] Huang, L. *et al.* Reconstructing human activities via coupling mobile phone data with location-based social networks. *Travel Behaviour and Society* **33**, 100606 (2023).

[32] Crandall, D. J. *et al.* Inferring social ties from geographic coincidences. *Proceedings of the National Academy of Sciences* **107**, 22436–22441 (2010).

[33] Pham, H., Shahabi, C. & Liu, Y. Inferring social strength from spatiotemporal data. *ACM Trans. Database Syst.* **41** (2016).

[34] Alessandretti, L., Orozco, L. G. N., Saberi, M., Szell, M. & Battiston, F. Multimodal urban mobility and multilayer transport networks. *Environment and Planning B: Urban Analytics and City Science* **50**, 2038–2070 (2023).

[35] Liao, Y., Gil, J., Yeh, S., Pereira, R. H. & Alessandretti, L. Socio-spatial segregation and human mobility: A review of empirical evidence. *Computers, Environment and Urban Systems* **117**, 102250 (2025).

[36] Zhou, M., Zhou, J., Zhou, J., Lei, S. & Zhao, Z. Introducing social contacts into the node-place model: A case study of hong kong. *Journal of Transport Geography* **107**, 103532 (2023).

[37] Lucas, K. Transport and social exclusion: Where are we now? *Transport Policy* **20**, 105–113 (2012).

[38] Creutzig, F. *et al.* Adjust urban and rural road pricing for fair mobility. *Nature Climate Change* **10**, 591–594 (2020).

[39] Cramton, P., Geddes, R. R. & Ockenfels, A. Set road charges in real time to ease traffic. *Nature* **560**, 23 – 25 (2018).

[40] Wachs, M. U.S. transit subsidy policy: In need of reform. *Science* **244**, 1545–1549 (1989).

[41] Ding, D. *et al.* The co-benefits of active travel interventions beyond physical activity: a systematic review. *The Lancet Planetary Health* **8**, e790–e803 (2024).

[42] Timmons, S., Andersson, Y., McGowan, F. P. & Lunn, P. D. Active travel infrastructure design and implementation: Insights from behavioral science. *Wiley Interdisciplinary Reviews: Climate Change* **15**, e878 (2024).

[43] Ross, M. L., Hazlett, C. & Mahdavi, P. Global progress and backsliding on gasoline taxes and subsidies. *Nature Energy* **2**, 16201 (2017).

[44] Shoup, D. C. The high cost of free parking. *Journal of Planning Education and Research* **17**, 3–20 (1997).

[45] Eagle, N. & Pentland, A. Reality mining: sensing complex social systems. *Personal and Ubiquitous Computing* **10**, 255–268 (2006).

[46] Chetty, R. *et al.* Social capital I: measurement and associations with economic mobility. *Nature* **608**, 108–121 (2022).

[47] Chetty, R. *et al.* Social capital II: determinants of economic connectedness. *Nature* **608**, 122–134 (2022).

[48] Thrift, N. & Pred, A. Time-geography: a new beginning. *Progress in Human Geography* **5**, 277–286 (1981).

[49] Batty, M. The size, scale, and shape of cities. *Science* **319**, 769–771 (2008).

[50] Graham, S. & Marvin, S. *Splintering Urbanism: Networked Infrastructures, Technological Mobilities and the Urban Condition* 1st edn (Routledge, London, 2001).

[51] Moreno, C., Allam, Z., Chabaud, D., Gall, C. & Pratlong, F. Introducing the "15-minute city": Sustainability, resilience and place identity in future post-pandemic cities. *Smart Cities* **4**, 93–111 (2021).

[52] LianJia property database. https://bj.lianjia.com/xiaoqu/. Accessed on May 2024.

[53] Amap navigation application API. https://lbs.amap.com/. Accessed on May 2024.

[54] OpenStreetMap Contributors. OpenStreetMap Data. https://www.openstreetmap.org/. Accessed on May 2024.

[55] Zheng, Y., Zhang, L., Xie, X. & Ma, W.-Y. Mining interesting locations and travel sequences from GPS trajectories. *Proceedings of the 18th International Conference on World Wide Web* 791–800 (2009).

[56] Cai, Z. *et al.* MemDA: Forecasting urban time series with memory-based drift adaptation. *Proceedings of the 32nd ACM International Conference on Information and Knowledge Management* 193–202 (2023).

[57] Hariharan, R. & Toyama, K. Project lachesis: Parsing and modeling location histories. *Geographic Information Science* 106–124 (2004).

[58] National Bureau of Statistics of China. The seventh national population census (2021). https://www.stats.gov.cn/english/PressRelease/202105/t20210510_1817185.html.

[59] Shannon, C. E. A mathematical theory of communication. *The Bell System Technical Journal* **5**, 3–55 (2001).

[60] Zhao, P., Wang, H., Liu, Q., Yan, X.-Y. & Li, J. Unravelling the spatial directionality of urban mobility. *Nature Communications* **15**, 4507 (2024).

[61] Greene, W. H. *Econometric Analysis* 8th edn (Pearson, New York, NY, 2018).

[62] McFadden, D. *Conditional Logit Analysis of Qualitative Choice Behavior* (Academic Press, New York, 1974).

# Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- supplementary.pdf